Genetic and biomarker studies of human longevity

Joris Deelen

Genetic and biomarker studies of human longevity
J. Deelen MSc

The cover displays newly hatched sea turtles crawling towards the ocean. Turtles are a symbol for a long and healthy life. They age very slowly and thus have the propensity to become long-lived. However, only very few, or even none, of the newly hatched sea turtles on this cover will actually become long-lived. To a lesser extent, this also seems to apply to individuals from long-lived families. Hence, by studying these families, using both genetic and biomarker approaches, we may be able to find the key towards a long and healthy life.

Financial support for the printing of this thesis was provided by the Netherlands Consortium for Healthy Ageing (NGI 050-060-810).

PhD thesis with summary in Dutch
ISBN: 978-90-8891-906-0

Cover design:	Proefschriftmaken.nl || Uitgeverij BOXPress
Printed & Lay Out by:	Proefschriftmaken.nl || Uitgeverij BOXPress
Published by:	Uitgeverij BOXPress, ‘s-Hertogenbosch

Genetic and biomarker studies of human longevity

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 25 juni 2014
klokke 15.00 uur

door

Joris Deelen

geboren te ’s-Gravenhage
in 1985

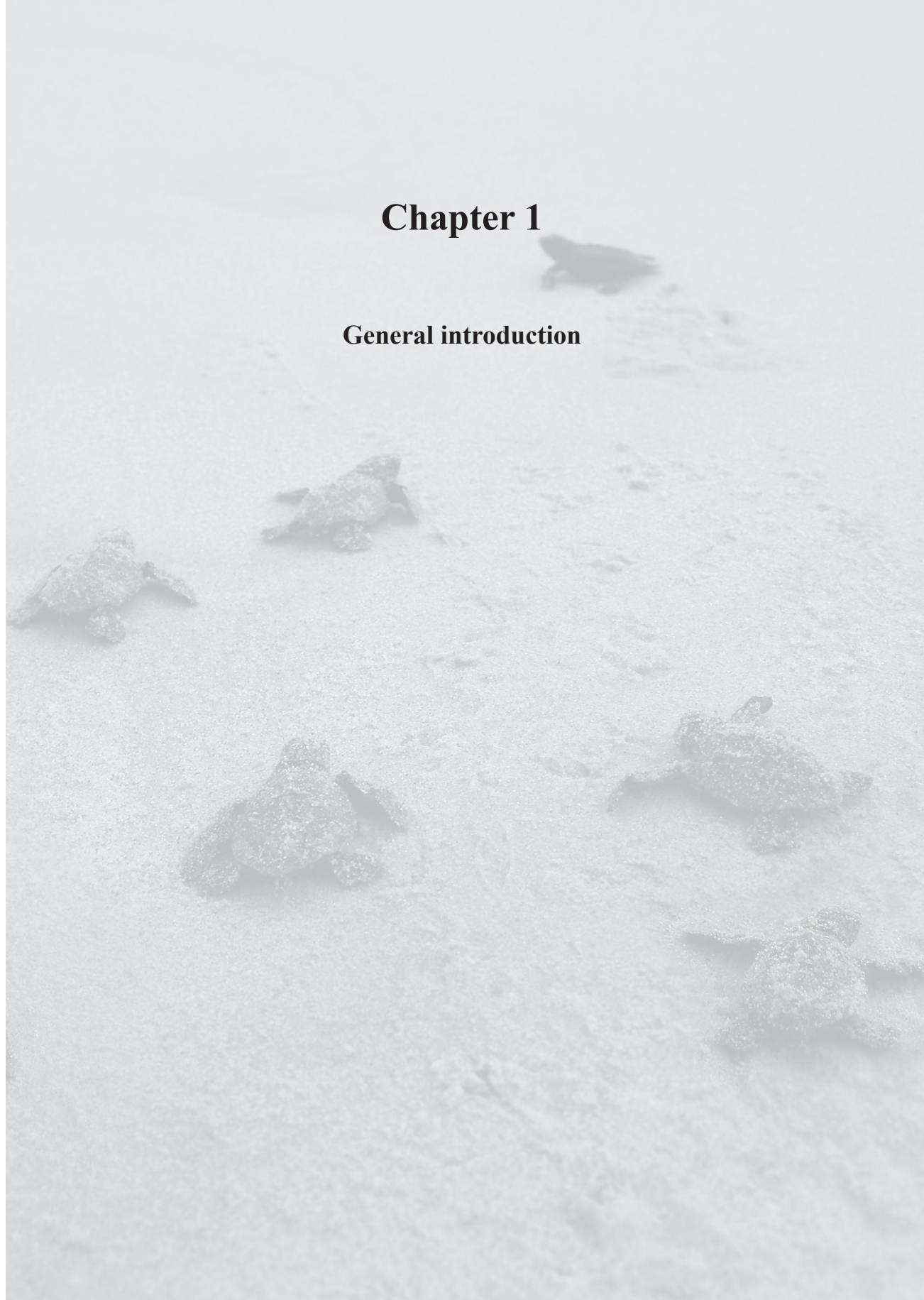
Promotiecommissie

Promotor:	Prof. Dr. P. Slagboom
	Prof. Dr. R.G.J. Westendorp
Co-promotor:	Dr. M. Beekman
Overige leden:	Dr. A.J.M. de Craen
	Prof. Dr. T.B.L. Kirkwood *Newcastle University*
	Prof. Dr. T.N. Wijmenga *University Medical Center Groningen*

Contents

Chapter 1

General introduction



Worldwide, life expectancy has shown a remarkable linear increase over the last two centuries [1]. However, the number of years of life spent in disability also increases and individuals from the European Union born in 2009 are expected to spend on average 25% (women) or 20% (men) of their life in poor health, i.e., experience limited or severe long-term limitation (> 6 months) in usual activity caused by ill-health (http://www.healthy-life-years.eu/) [2]. This stresses the importance of efforts aimed at increasing the disability-free life expectancy. The majority of disabilities are caused by diseases, such as cancer, cardiovascular disease, hypertension, osteoarthritis, and type 2 diabetes, for which chronological age is the main risk factor. Interestingly, part of the individuals that survive to exceptionally old ages do not display excessive levels of disability [3,4], indicating that reaching a high age does not necessarily result in an increase in age-related disability.

Use of family-based cohorts to study healthy aging and longevity

It is expected that the number of years spent in disability could be reduced by avoiding age-related diseases [5]. Remarkably, long-lived families indeed display a low prevalence of age-related diseases from middle age onwards [6-10]. In addition, they show beneficial or "youthful" profiles for numerous cognitive, metabolic, and immune-related parameters. Examples of these features are the low prevalence of cytomegalovirus infections, low free triiodothyronine and triglyceride serum levels, and preservation of insulin sensitivity in middle age [7,11-16]. Thus, by studying long-lived families (Table 1.1), one might be able to identify mechanisms driving healthy aging and protection from age-related diseases in middle and old age. Ultimately, this knowledge may be used to extend the disability-free life expectancy in the population.

One strategy to identify the mechanisms underlying lifespan regulation is by applying genetic approaches. The genetic component of longevity, as estimated from twin and family-based studies, is ~25% (Table 1.2) and the genetic contribution increases with age [17,18]. However, there is large heterogeneity in the genetic component estimates between studies, which could be caused by geographical or methodological differences [19]. The genetic component is most prominent in long-lived families [20,21], which makes them highly suitable for genomic approaches.

In addition to the genetic approach, research into biological and physiological phenotypes accompanying a long life may illuminate mechanisms of healthy aging. To this end, long-lived families are being studied for quantitative parameters or profiles that mark chronological and/or biological age, i.e., the age based on the molecular and psychological functioning of the individual, which could subsequently be investigated in large cohorts of middle-aged individuals. Thus, identifying the genetic component and/or biomarkers of longevity may contribute to the disclosure of mechanisms driving healthy aging and longevity.

Table 1.1 Overview of family-based longevity studies.

	Long-lived individuals		Offspring	Controls		
Study	***n***	**Type**	***n***	***n***	**Type**	**Reference**
Ashkenazi Jews	365	Centenarians	593	356	Spouses of offspring + population controls	22
European Challenge for Healthy Ageing	257	Centenarians	276	204	Cousins of offspring without centenarian parent	23
GEnetics of Healthy Ageing*	4,498	Nonagenarian siblings	~700	2,249	Population controls	24
Leiden Longevity Study	944	Nonagenarian siblings	1,671	744	Spouses of offspring	21
Long Life Family Study	1,373	Nonagenarian siblings	2,317	582	Spouses of offspring	7
New England Centenarian Study**	>1,800	Centenarians	>600	437	Population controls	25

*Offspring is recruited in the MARK-AGE project, **Recruitment still ongoing.

Table 1.2 Overview of studies that examined the genetic component of longevity or lifespan.

			Total		Men		Women		
Study	**Country**	**Type**	***n***	h^2	***n***	h^2	***n***	h^2	**Reference**
GenomEUtwin	Denmark / Finland / Italy / Sweden	Twins	9,334		4,598	0.120	4,736	0.260	26
Danish Twin Registry	Denmark	Twins	5,744		2,816	0.260	2,928	0.230	27
Swedish Twin Registry	Sweden	Twins	1,250		164	0.010	194	0.150	28
Utah Population Database	United States	Families	78,994	0.147					29
European royal and noble families	Europe	Families	12,150		8,409	0.180	3,741	0.200	17
MICROS study	Italy	Families	8,277	0.150	4,299	0.160	3,978	0.180	19
Genealogia Sursilliana CD-2000	Finland	Families	2,614		1,226	0.175	1,388	0.167	30
Old Order Amish	United States	Families	1,655	0.250					31
Valserine Valley XVIII-XXth Centuries	France	Families	1,102	0.270	586		516		32

h^2, genetic component estimate (heritability or comparable statistic).

Genetic research of aging and longevity in animal models

The first studies into the genetics of lifespan regulation were performed in animal models, such as yeast, worms, flies, and mice. In contrast to human longevity studies, which are mainly observational, animal-based studies benefit from genetic manipulation (mutagenesis) via RNA interference, knock-out or overexpression of single genes. Using these approaches, many genes have been identified that extend lifespan in these models (GenAge; http://genomics.senescence.info/genes/) [33]. The most interesting conserved pathways identified using animal models are the growth hormone (GH)/insulin/insulin-like growth factor 1 (IGF-1) signaling and mammalian target of rapamycin signaling pathways [34]. The limitation of the animal-based longevity studies in lower species, such as worms, is that they mainly focus on lifespan as an outcome and that the parameters that reflect the physiology and pathology of aging are not well defined or highly difficult to compare with their human counterparts. Nonetheless, these studies have been crucial for the identification of lifespan regulating pathways that also contribute to human longevity.

Application of GWAS for identification of novel human longevity loci

Most genetic research on human longevity has been focused on lifespan regulating loci involved in GH/insulin/IGF-1 signaling [35]. Although many of the GH/insulin/IGF-1 signaling genes have been investigated (see http://genomics.senescence.info/longevity/ [36] for an overview), the only gene associated with human longevity in multiple independent studies is *FOXO3A* [37-39]. *FOXO3A* encodes the protein forkhead box O3, which acts as a transcription factor for many different genes involved in, e.g., apoptosis and oxidative stress [40]. In addition, a study by van Heemst and colleagues showed that a composite pathway score based on 6 genetic variants in GH/insulin/IGF-1 signaling genes is associated with mortality in women, which further highlights the role of this pathway in lifespan regulation [41]. The other candidate gene that has consistently been associated with human longevity in multiple independent studies is *APOE* [35,42]. *APOE* encodes the protein apolipoprotein E (ApoE), which seems to be involved in, e.g., lipoprotein metabolism, cognitive function, and immune regulation [43]. The ApoE protein has three isoforms (ApoE ε2, ApoE ε3, and ApoE ε4) defined by two single nucleotide polymorphisms (SNPs), rs7412 (Arg136Cys; ε2) and rs429358 (Cys112Arg; ε4). Interestingly, ApoE ε4 has been associated with a decreased probability to become long-lived, while ApoE ε2 has an opposite effect. However, since the effect of ApoE ε4 seems to be most prominent, *APOE* is generally considered a "frailty gene" [44]. Thus, although candidate gene studies have shown to be useful, the number of human longevity genes identified by these studies is limited.

Instead of studying the genome using a hypothesis-based approach, hypothesis-free approaches could be performed. An example of such an approach is the genome-

wide association study (GWAS), aimed at identifying common genetic variants with, usually, small effects. In a GWAS, 300,000-2,500,000 SNPs are assessed for association with the trait of interest. This approach has successfully been applied to many diseases and traits (National Human Genome Research Institute GWAS Catalog; http://www.genome.gov/gwastudies/) [45]. In GWAS for longevity, genotype frequencies are compared between long-lived cases and shorter-lived or young controls. The genome of long-lived individuals is assumed to be characterized by a decreased prevalence of disease-promoting variants of considerable effect and an increased prevalence of variants promoting healthy aging. Since longevity is assumed to be determined by many genes with small effects, GWAS is expected to be a successful method to identify novel human longevity loci.

Genomic research might benefit from biomarker research

The number of long-lived individuals that can currently be included in genomic studies is limited (~30,000 individuals). Hence, it is almost impossible to reach a sufficient sample size required to identify genetic variants with relatively small effects, such as those identified for more common traits, like height and lipid levels, with sample sizes > 100,000 individuals. To overcome this problem, one might try to identify (combinations of) phenotypes that could be used as biomarkers of healthy aging in genomic studies of large cohorts of middle-aged individuals. We propose that a biomarker of healthy aging should; (1) show a change with chronological age, at least above 40 years, (2) discriminate individuals with a "youthful" or old level relative to their age category in the general population, (3) associate with known health parameters, and (4) associate with future morbidity and/or mortality in prospective studies (**Chapter 2**).

Aim and outline of the thesis

The drivers of human longevity may provide insight in the mechanisms that result in delay or avoidance of age-related diseases. Since knowledge of such mechanisms may contribute to the extension of disability-free lifespan, the aim of this thesis was to identify novel lifespan regulating loci that influence human longevity and population mortality. We performed our research in various cohorts of elderly individuals, including the family-based Leiden Longevity Study (LLS) and GEnetics of Healthy Ageing project (Table 1.1), the population-based Rotterdam Study, which includes individuals above 55 years that were followed-up for > 20 years, and the prospective Leiden 85-plus study and PROspective Study of Pravastatin in the Elderly at Risk, in which the association of a genetic variant with mortality can be tested.

To identify genetic drivers of human longevity by GWAS, we first compared unrelated nonagenarians from the LLS (Table 1.1) with young controls from the Rotterdam Study. The loci that showed suggestive evidence for association with survival ≥ 90 years were tested for replication in the Rotterdam Study, Leiden 85-plus study, and Danish 1905 cohort. Subsequently,

we performed a combined analysis of the discovery and replication cohorts (4,149 cases and 7,582 controls) (**Chapter 3**).

Due to the complexity of the longevity phenotype and the relatively small sample size, the LLS longevity GWAS turned out to have insufficient power to detect significant effects besides the well-established *TOMM40/APOE/APOC1* locus (**Chapter 3**). We therefore carried out an extended GWAS, in which we studied the genetics of long-lived cases (≥ 85 years) and younger controls (< 65 years of age) from all over Europe. The loci that showed suggestive evidence for association with survival ≥ 85 and/or ≥ 90 years were taken forward for replication in 6 additional cohorts and we performed a combined analysis of the discovery and replication cohorts (20,789 cases and 77,277 controls) (**Chapter 4**).

Instead of analyzing single SNPs, as was done in the LLS and EU longevity GWAS (**Chapter 3** and **4**), the combined effect of a SNP set, grouped per pathway or gene region, can be tested for association with longevity. The advantage of these tests is that they are very suitable for studies of polygenic complex traits with limited power for GWAS analysis, such as longevity [46], due to the low penalty for multiple testing as compared to single SNP analysis. Two candidate pathways for human longevity are the insulin/IGF-1 signaling (IIS) pathway and the telomere maintenance (TM) pathway. The IIS pathway is involved in the adaptation of the organism to its (changing) environment [47], while the TM pathway regulates telomere integrity [48,49]. Genetic variation in genes that play a role in IIS and TM has previously been associated with human longevity [37,39,41,50]. To determine if the combined effect of IIS and TM pathway SNPs is associated with human longevity, we performed gene set analysis with gene sets based on these pathways using the LLS longevity GWAS dataset (**Chapter 5**).

Since our genetic approaches delivered a limited number of longevity loci and pathways, we also performed a study on leukocyte telomere length (LTL), a potential biomarker of healthy aging that could be used for genomic studies in large cohorts of middle-aged individuals. Previous studies have shown that LTL is associated with multiple diseases and increased prospective mortality [51]. In addition, a study in an Ashkenazi Jewish population (Table 1.1) showed that offspring of centenarians have a longer mean LTL as compared to controls from the general population [50], indicating that mechanisms regulating LTL might also be involved in human lifespan regulation. Hence, to test the proposed criteria for biomarkers of healthy aging, we investigated LTL for association with chronological age, familial longevity, known health parameters, and prospective mortality in long-lived families from the LLS (**Chapter 6**). In addition, we performed a look-up of the LTL-associated genetic variants in our EU longevity GWAS results described in **Chapter 4** to determine the association with survival to ages beyond 90 years.

References

1. Oeppen J, Vaupel JW. Demography. Broken limits to life expectancy. *Science* 2002; **296**: 1029-31.
2. Jagger C, Gillies C, Moscone F, Cambois E, van Oyen H, Nusselder W, *et al*. Inequalities in healthy life years in the 25 countries of the European Union in 2005: a cross-national meta-regression analysis. *Lancet* 2008; **372**: 2124-31.
3. Christensen K, McGue M, Petersen I, Jeune B, Vaupel JW. Exceptional longevity does not result in excessive levels of disability. *Proc Natl Acad Sci U S A* 2008; **105**: 13274-9.
4. Terry DF, Sebastiani P, Andersen SL, Perls TT. Disentangling the roles of disability and morbidity in survival to exceptional old age. *Arch Intern Med* 2008; **168**: 277-83.
5. Jagger C, Matthews R, Matthews F, Robinson T, Robine JM, Brayne C. The burden of diseases on disability-free life expectancy in later life. *J Gerontol A Biol Sci Med Sci* 2007; **62**: 408-14.
6. Atzmon G, Schechter C, Greiner W, Davidson D, Rennert G, Barzilai N. Clinical phenotype of families with longevity. *J Am Geriatr Soc* 2004; **52**: 274-7.
7. Newman AB, Glynn NW, Taylor CA, Sebastiani P, Perls TT, Mayeux R, *et al*. Health and function of participants in the Long Life Family Study: A comparison with other cohorts. *Aging (Albany NY)* 2011; **3**: 63-76.
8. Bos SD, Beekman M, Maier AB, Karsdal MA, Kwok WY, Bay-Jensen AC, *et al*. Metabolic health in families enriched for longevity is associated with low prevalence of hand osteoarthritis and influences OA biomarker profiles. *Ann Rheum Dis* 2013; **72**: 1669-74.
9. Terry DF, Wilcox MA, McCormick MA, Pennington JY, Schoenhofen EA, Andersen SL, *et al*. Lower all-cause, cardiovascular, and cancer mortality in centenarians' offspring. *J Am Geriatr Soc* 2004; **52**: 2074-6.
10. Westendorp RG, van Heemst D, Rozing MP, Frolich M, Mooijaart SP, Blauw GJ, *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**: 1634-7.
11. Stijntjes M, de Craen AJ, van Heemst D, Meskers CG, van Buchem MA, Westendorp RG, *et al*. Familial longevity is marked by better cognitive performance at middle age: the leiden longevity study. *PLoS One* 2013; **8**: e57962.
12. Barral S, Cosentino S, Costa R, Matteini A, Christensen K, Andersen SL, *et al*. Cognitive function in families with exceptional survival. *Neurobiol Aging* 2012; **33**: 619-7.
13. Barzilai N, Atzmon G, Schechter C, Schaefer EJ, Cupples AL, Lipton R, *et al*. Unique lipoprotein phenotype and genotype associated with exceptional longevity. *JAMA* 2003; **290**: 2030-40.
14. Derhovanessian E, Maier AB, Beck R, Jahn G, Hahnel K, Slagboom PE, *et al*. Hallmark features of immunosenescence are absent in familial longevity. *J Immunol* 2010; **185**: 4618-24.
15. Slagboom PE, Beekman M, Passtoors WM, Deelen J, Vaarhorst AA, Boer JM, *et al*. Genomics of human longevity. *Philos Trans R Soc Lond B Biol Sci* 2011; **366**: 35-42.
16. Wijsman CA, Rozing MP, Streefland TC, le Cessie S, Mooijaart SP, Slagboom PE, *et al*. Familial longevity is marked by enhanced insulin sensitivity. *Aging Cell* 2011; **10**: 114-21.
17. Gavrilova NS, Gavrilov LA, Evdokushkina GN, Semyonova VG, Gavrilova AL, Evdokushkina NN, *et al*. Evolution, mutations, and human longevity: European royal and noble families. *Hum Biol* 1998; **70**: 799-804.
18. Hjelmborg JV, Iachine I, Skytthe A, Vaupel JW, McGue M, Koskenvuo M, *et al*. Genetic influence on human lifespan and longevity. *Hum Genet* 2006; **119**: 312-21.
19. Gogele M, Pattaro C, Fuchsberger C, Minelli C, Pramstaller PP, Wjst M. Heritability analysis of life span in a semi-isolated population followed across four centuries reveals the presence of pleiotropy between life span and reproduction. *J Gerontol A Biol Sci Med Sci* 2011; **66**: 26-37.

20. Perls T, Shea-Drinkwater M, Bowen-Flynn J, Ridge SB, Kang S, Joyce E, *et al*. Exceptional familial clustering for extreme longevity in humans. *J Am Geriatr Soc* 2000; **48**: 1483-5.
21. Schoenmaker M, de Craen AJ, de Meijer PH, Beekman M, Blauw GJ, Slagboom PE, *et al*. Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur J Hum Genet* 2006; **14**: 79-84.
22. Lai JY, Atzmon G, Melamed ML, Hostetter TH, Crandall JP, Barzilai N, *et al*. Family history of exceptional longevity is associated with lower serum uric acid levels in Ashkenazi Jews. *J Am Geriatr Soc* 2012; **60**: 745-50.
23. De Rango F, Dato S, Bellizzi D, Rose G, Marzi E, Cavallone L, *et al*. A novel sampling design to explore gene-longevity associations: the ECHA study. *Eur J Hum Genet* 2008; **16**: 236-42.
24. Skytthe A, Valensin S, Jeune B, Cevenini E, Balard F, Beekman M, *et al*. Design, recruitment, logistics, and data management of the GEHA (Genetics of Healthy Ageing) project. *Exp Gerontol* 2011; **46**: 934-45.
25. Perls TT, Bochen K, Freeman M, Alpert L, Silver MH. Validity of reported age and centenarian prevalence in New England. *Age Ageing* 1999; **28**: 193-7.
26. Skytthe A, Pedersen NL, Kaprio J, Stazi MA, Hjelmborg JV, Iachine I, *et al*. Longevity studies in GenomEUtwin. *Twin Res* 2003; **6**: 448-54.
27. Herskind AM, McGue M, Holm NV, Sorensen TI, Harvald B, Vaupel JW. The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870-1900. *Hum Genet* 1996; **97**: 319-23.
28. Ljungquist B, Berg S, Lanke J, McClearn GE, Pedersen NL. The effect of genetic factors for longevity: a comparison of identical and fraternal twins in the Swedish Twin Registry. *J Gerontol A Biol Sci Med Sci* 1998; **53**: M441-M446.
29. Kerber RA, O'Brien E, Smith KR, Cawthon RM. Familial excess longevity in Utah genealogies. *J Gerontol A Biol Sci Med Sci* 2001; **56**: B130-B139.
30. Pettay JE, Kruuk LE, Jokela J, Lummaa V. Heritability and genetic constraints of life-history trait evolution in preindustrial humans. *Proc Natl Acad Sci U S A* 2005; **102**: 2838-43.
31. Mitchell BD, Hsueh WC, King TM, Pollin TI, Sorkin J, Agarwala R, *et al*. Heritability of life span in the Old Order Amish. *Am J Med Genet* 2001; **102**: 346-52.
32. Cournil A, Legay JM, Schachter F. Evidence of sex-linked effects on the inheritance of human longevity: a population-based study in the Valserine valley (French Jura), 18-20th centuries. *Proc Biol Sci* 2000; **267**: 1021-5.
33. de Magalhaes JP, Budovsky A, Lehmann G, Costa J, Li Y, Fraifeld V, *et al*. The Human Ageing Genomic Resources: online databases and tools for biogerontologists. *Aging Cell* 2009; **8**: 65-72.
34. Fontana L, Partridge L, Longo VD. Extending healthy life span--from yeast to humans. *Science* 2010; **328**: 321-6.
35. Christensen K, Johnson TE, Vaupel JW. The quest for genetic determinants of human longevity: challenges and insights. *Nat Rev Genet* 2006; **7**: 436-48.
36. Tacutu R, Craig T, Budovsky A, Wuttke D, Lehmann G, Taranukha D, *et al*. Human Ageing Genomic Resources: integrated databases and tools for the biology and genetics of ageing. *Nucleic Acids Res* 2013; **41**: D1027-D1033.
37. Flachsbart F, Caliebe A, Kleindorp R, Blanche H, von Eller-Eberstein H, Nikolaus S, *et al*. Association of FOXO3A variation with human longevity confirmed in German centenarians. *Proc Natl Acad Sci U S A* 2009; **106**: 2700-5.
38. Kuningas M, Magi R, Westendorp RG, Slagboom PE, Remm M, van Heemst D. Haplotypes in the human Foxo1a and Foxo3a genes; impact on disease and mortality at old age. *Eur J Hum Genet* 2007; **15**: 294-301.
39. Willcox BJ, Donlon TA, He Q, Chen R, Grove JS, Yano K, *et al*. FOXO3A genotype is strongly associated with human longevity. *Proc Natl Acad Sci U S A* 2008; **105**: 13987-92.

40. van der Horst A, Burgering BM. Stressing the role of FoxO proteins in lifespan and disease. *Nat Rev Mol Cell Biol* 2007; **8**: 440-50.
41. van Heemst D, Beekman M, Mooijaart SP, Heijmans BT, Brandt BW, Zwaan BJ, *et al*. Reduced insulin/IGF-1 signalling and human longevity. *Aging Cell* 2005; **4**: 79-85.
42. Schachter F, Faure-Delanef L, Guenot F, Rouger H, Froguel P, Lesueur-Ginot L, *et al*. Genetic associations with human longevity at the APOE and ACE loci. *Nat Genet* 1994; **6**: 29-32.
43. Mahley RW, Rall SC, Jr. Apolipoprotein E: far more than a lipid transport protein. *Annu Rev Genomics Hum Genet* 2000; **1**: 507-37.
44. Gerdes LU, Jeune B, Ranberg KA, Nybo H, Vaupel JW. 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**: 202-10.
45. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, *et al*. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 2009; **106**: 9362-7.
46. Finch CE, Tanzi RE. Genetics of aging. *Science* 1997; **278**: 407-11.
47. Tatar M, Bartke A, Antebi A. The endocrine regulation of aging by insulin-like signals. *Science* 2003; **299**: 1346-51.
48. Collins K, Mitchell JR. Telomerase in the human organism. *Oncogene* 2002; **21**: 564-79.
49. de Lange T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev* 2005; **19**: 2100-10.
50. Atzmon G, Cho M, Cawthon RM, Budagov T, Katz M, Yang X, *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**: 1710-7.
51. Sanders JL, Newman AB. Telomere Length in Epidemiology: A Biomarker of Aging, Age-Related Disease, Both, or Neither? *Epidemiol Rev* 2014; *In press*.

Chapter 2

Identifying the genomic determinants of aging and longevity in human population studies; progress and challenges

Joris Deelen, Marian Beekman, Miriam Capri, Claudio Franceschi and P Eline Slagboom

Bioessays 2013; **35**: 386-96.

Abstract

Human lifespan variation is mainly determined by environmental factors, whereas the genetic contribution is 25-30% and expected to be polygenic. Two complementary fields go hand in hand in order to unravel the mechanisms of biological aging: genomic and biomarker research. Explorative and candidate gene studies of the human genome by genetic, transcriptomic, and epigenomic approaches have resulted in the identification of a limited number of interesting positive linkage regions, genes, and pathways that contribute to lifespan variation. The possibilities to further exploit these findings are rapidly increasing through the use of novel technologies, such as next-generation sequencing. Genomic research is progressively being integrated with biomarker studies on aging, including the application of (noninvasive) deep phenotyping and omics data – generated using novel technologies – in a wealth of studies in human populations. Hence, these studies may assist in obtaining a more holistic perspective on the role of the genome in aging and lifespan regulation.

Introduction

Human life expectancy has increased remarkably over the last two centuries worldwide [1], although it is still highly variable between countries [2]. This lifespan extension is mainly due to improvement of health care, hygiene, and nutrition. The healthy life expectancy, however, has not increased at the same rate; in Europe, men spend on average 20.5% and women 25.4% of their life dealing with disability caused through disease or injury (Healthy Life Years; http://www.healthy-life-years.eu/) [3]. Although age is the main risk factor for the majority of common diseases contributing to disability, reaching an old age does not necessarily result in a higher degree of age-related disability. This is illustrated by the presence of long-lived individuals from families expressing exceptional longevity that may reach high ages without major disabilities [4,5]. Moreover, their offspring – considered "decelerated" or "healthy agers" – have a lower prevalence of age-related diseases, such as cancer, cardiovascular disease, hypertension, and type 2 diabetes [6-9], compared to similar-aged controls. Concomitantly, they show beneficial or "youthful" profiles for many metabolic and immune-related parameters [10]. Most of the human aging studies are concentrated around long-lived families, including highly and middle-aged members, sporadic highly aged individuals from the general population or population-based cohorts containing different age groups.

Due to the different study designs (Box 2.1 and Figure 2.1), human aging cohorts provide complementary information and are intensively being studied from a biomarker and genomic perspective. The assumption is that, together, these studies will provide insight into the mechanisms that could (i) drive the biological aging rate, (ii) positively and negatively influence the risk for age-related disease, and (iii) explain the variation in lifespan between individuals. Genomic research, including genetic, epigenetic, and transcriptomic studies, is expected to provide both markers and determinants of aging. The search for biomarkers of human aging and longevity is aimed at identifying parameters and profiles that reflect the biological age of individuals and predict long-term morbidity and/or mortality [11].

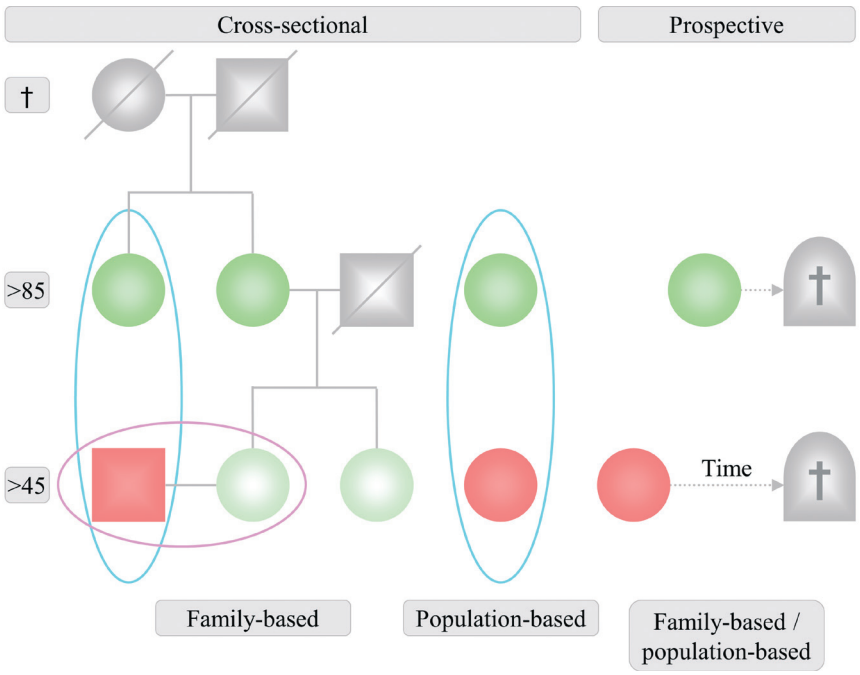
For most diseases, like osteoarthritis, osteoporosis, and type 2 diabetes, standardized phenotypes and diagnostic criteria are used for genomic research. No standardized phenotype or marker, however, is indicating biological aging rate. Hence, genomic studies into aging thus far focus on the determinants of human lifespan variation by using age at death, prospective survival, disease-free survival, or exceptional longevity as outcome. Biomarker research is therefore just as relevant for genomic studies of human aging as the analysis of the genome itself.

The possibility to study causal determinants and quantitative biomarkers of biological aging and longevity in humans strongly depends on the study designs that are available (Box 2.1 and Figure 2.1). Using these designs, we determined four relevant phases in aging studies in order to establish whether a quantitative parameter (or profile) is a biomarker of biological age; (i) Determine the change in a quantitative parameter with chronological age in cross-sectional studies

and, preferably, by repeated measures in longitudinal studies. Parameters reflecting biological age are expected to show an increased variance with age. (ii) Determine whether a marker of chronological age also discriminates individuals with a "youthful" or old level relative to their age category in the general population, which would indicate that the quantitative parameter potentially marks biological age (Figure 2.2). The comparison between offspring of long-lived individuals and age-matched population controls is also part of this phase. (iii) Determine whether the potential marker for biological age associates with known parameters of health, such as blood pressure, serum levels of glucose, insulin, and cholesterol. (iv) Determine whether the potential marker for biological age associates with morbidity (based on clinical endpoints) and/or mortality in prospective studies.

In this review we will give an overview of the main genomic approaches and discuss the concept of biomarker approaches used in the research field of human aging and longevity. In addition, we will discuss the progress and challenges of integration of data that has been generated using these approaches.

Figure 2.1 Study designs applied in studies of healthy aging and longevity. Family- or population-based cross-sectional designs usually compare highly aged individuals with younger controls (blue ovals). Alternatively, the offspring of long-lived individuals is compared to age-matched controls (their spouses or random population controls) (purple oval). Thirdly, prospective studies are performed in highly or middle-aged individuals (unrelated or from (long-lived) families) which are followed over time (ranging from 10 to 30 years, depending on the study). Highly aged individuals are depicted in green, their offspring in light green and middle-aged individuals in red.



Box 2.1

Study designs

The ultimate epidemiological study design to investigate markers and determinants of biological aging and longevity in humans would be to follow a large group of individuals during their entire lifetime. These individuals should be examined at different time points so that changes in markers could be related to the actual lifespan of the individual. However, since this design is not feasible, several other designs are being applied in human studies (Figure 2.1).

Cross-sectional study designs

***Population-based cohorts*:** Cross-sectional longevity studies typically compare unrelated highly aged individuals (nonagenarians/centenarians) with younger controls or evaluate differences between groups of unrelated individuals in categories of increasing age. Inclusion of individuals for these studies is relatively easy, which is reflected by the large sample sizes of population-based cross-sectional studies. The cross-sectional study usually provides the first level of observation that a parameter is correlated with chronological age or a health condition. However, causality of the genetic and/or genomic parameter on aging and longevity cannot be determined from a cross-sectional design. For cross-sectional studies the long-lived cases should be compared with controls originating from the same birth cohort. However, since these controls usually already died, controls are generally selected from other birth cohorts. Given that these cohorts have a different life expectancy, this could confound the studied association. In addition, structural differences between birth cohorts, caused by, e.g., migration, could also confound the results. Examples of longevity studies used for cross-sectional analysis in unrelated individuals are the New England Centenarian Study (NECS) [12], German long-lived individuals [13], French centenarians [14], and Southern Italian Centenarian Study (SICS) [15]. In addition, various cross-sectional studies are included in the MARK-AGE project, which consists of 2,320 randomly recruited volunteers from the general population (35-74 years).

***Family-based cohorts*:** Family-based longevity studies consist of nonagenarians/centenarians (siblings) and their middle-aged offspring. The controls used in these studies are either (age-matched) random individuals from the general population or spouses of the offspring of the long-lived individuals. Due to the common genetic background among family members, family-based longevity studies are enriched for familial and genetic effects on longevity and are more robust against population substructure. However, these studies generally have a small sample size, since it is quite difficult to collect long-lived families. To determine which age-related phenotypes associate with human familial longevity, the offspring of long-lived individuals, which are predisposed to longevity, can be compared to geographically- and age-matched population controls. This design allows analysis of molecular and clinical parameters specific for long-lived family members in multiple generations. Examples of family-based longevity studies are the Ashkenazi Jews cohort [16], GEnetics of Healthy Ageing (GEHA) project (of which the offspring is collected in the MARK-AGE project) [17], Long Life Family Study (LLFS) [7], and Leiden Longevity Study (LLS) [18].

Prospective studies

Most prospective longevity studies consist of highly (> 85 years of age) or middle-aged (> 55 years of age) individuals (related or unrelated) that are followed over time and sampled at multiple time points. This design is most often applied to provide more evidence for causality of determinants or markers detected in cross-sectional studies. In this design an (unbiased) baseline parameter may show to precede a functional aspect of aging. Several large population-based prospective

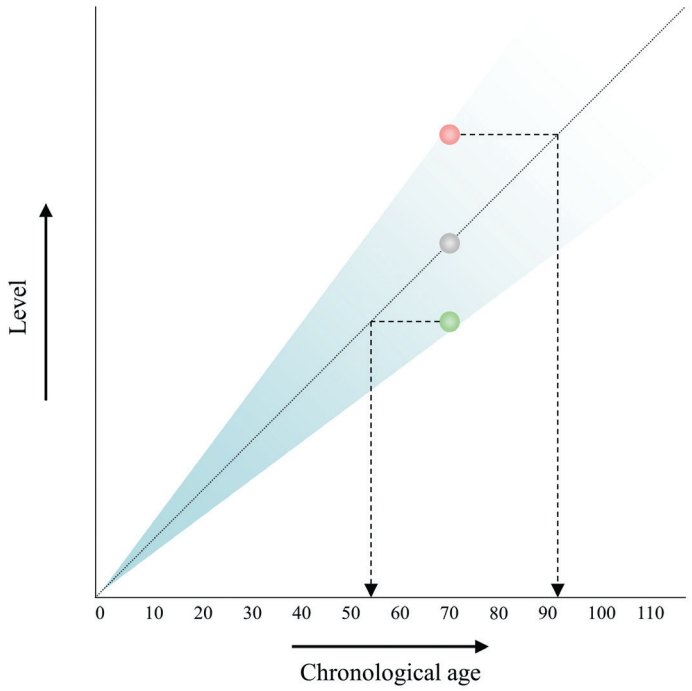
studies have been initiated. However, the main disadvantage of these studies is that the number of individuals that will become long-lived is usually very small. Examples of prospective longevity studies are the Leiden 85-plus study [19,20], Newcastle 85+ study [21,22], Danish 1905 cohort [23], the population-based Rotterdam Study [24], and Framingham Heart Study (FHS) (consisting of three generations) [25].

Genomic research

Human longevity is not just explained by the absence of disease-susceptibility alleles

Genomic research into human lifespan regulation could be subdivided into genetic, epigenetic, and transcriptomic research. Studies of mono- and dizygous twins have revealed that the genetic contribution to the variation in human lifespan is about 25-30% [26,27] and is most prominent in families clustered for longevity [18,28]. This genetic contribution is mainly apparent after the age of 60 years and seems to increase with

Figure 2.2 Interpretation of the potential relationship between a marker of chronological age and biological age using categories of increasing age. The blue zone indicates the increasing variance of the marker with age. Individuals can be assigned to having a marker level which matches (i) the expected level for their age in the population (gray dot, 75 years in this example), (ii) the level of a younger age group (green dot, biological age may be lower than chronological age), or (iii) the level of an older age group (red dot, biological age may be higher than chronological age).



age [27,29]. Furthermore, human lifespan is a complex trait, which is assumed to be determined by many genes with small individual effects [30], although the polygenic architecture still needs to be characterized [31,32]. The diverse health features of long-lived families illustrate that different age-related diseases have common determinants and implicate that pathways can be identified that attenuate aging and delay age-related disease. From a genomic perspective, individuals from long-lived families are assumed to be characterized by a decreased prevalence of disease-promoting variants (referred to as disease-susceptibility alleles) and an increased prevalence of variants conferring maintenance of health and protection from disease, when compared to population controls. In the last 5 years, many disease-susceptibility alleles have been identified (National Human Genome Research Institute genome-wide association study (GWAS) Catalog; http://www.genome.gov/gwastudies/) [33]. A first comparison between long-lived individuals, selected from both long-lived families (LLS) and the general population (Leiden 85-plus study), and young controls showed no difference in the distribution or frequency of disease-susceptibility alleles identified in cancer, coronary artery disease, and type 2 diabetes [34]. The search for lifespan regulating loci – contributing to longevity and population mortality – must therefore extend beyond a focus on disease-susceptibility alleles. We will first discuss the efforts to identify longevity loci by genetics approaches.

Candidate gene studies identified *APOE* and *FOXO3A* as human longevity genes

The first genetic longevity studies mainly focused on lifespan regulating loci that emerged from animal models [35]. Lifespan extension in animal models was obtained by applying caloric restriction or by modifying gene functions (mutagenesis) using RNA interference, knock-out or overexpression of single genes (GenAge; http://genomics.senescence.info/genes/) [36]. The most interesting pathways identified using these models are the growth hormone (GH)/insulin/insulin-like growth factor 1 (IGF-1) signaling and mammalian target of rapamycin (mTOR) signaling pathways [37]. Thus far, lifespan has been the main phenotype investigated in animal models. In order to make these models more translatable to human studies research should focus on defining the parameters that reflect the physiology and pathology of aging in both animals and humans [38,39].

Most of the human candidate gene studies were performed in cross-sectional designs (Box 2.1 and Figure 2.1), comparing allele frequencies of potential longevity loci between highly aged individuals and young controls. The candidate gene studies based on single genes have pointed a role for genes involved in, e.g., GH/insulin/IGF-1 signaling, immune regulation, and lipoprotein metabolism (Table S2.1), although most of these results have not (yet) been confirmed in sufficient independent studies. The most convincing human longevity loci today are *APOE* and *FOXO3A,* which have frequently been associated with longevity in cross-sectional studies (see for a review [39]) and survival in prospective studies [40-42]

(Figure 2.3). *APOE* encodes the protein apolipoprotein E (ApoE), which seems to play a role in e.g., lipoprotein metabolism, cognitive function, and immune regulation [43]. *FOXO3A* encodes the protein forkhead box O3 (FOXO3a), which acts as a transcription factor for many different genes involved in processes like apoptosis and oxidative stress [44].

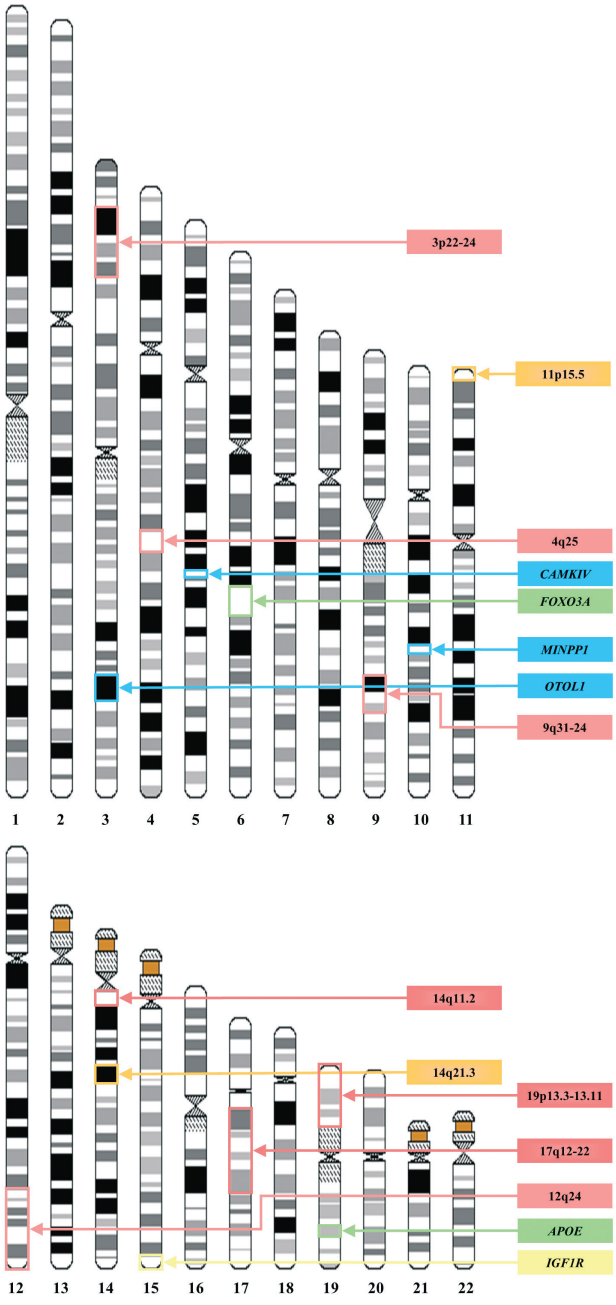
In addition to single gene studies, several candidate gene studies based on whole pathways have been performed. These pathway-based candidate gene studies showed a role for genes within the DNA damage signaling and repair, GH/insulin/IGF-1 signaling, immune regulation, pro/antioxidant, and telomere maintenance pathways [45-49] (Table S2.1). Most of these pathway-based studies tested for effects of individual single nucleotide polymorphisms (SNPs) on prospective mortality or longevity [45,47,48] and, so far, only a limited number of studies determined the joint effect of SNPs within a pathway [46,49].

Large meta-GWAS are required for identification of novel human longevity loci

As an alternative to hypothesis-based candidate gene studies, hypothesis-free or explorative approaches could be applied to studies of the genome. These methods should initially be aimed at prioritizing the location of regions linked to longevity and, subsequently, identifying the genetic variation causal to the trait. One example of an explorative approach is the GWAS. In this cross-sectional approach, in which long-lived individuals are compared with young or shorter-lived controls, the – usually small – effect of common variants can be identified. Typically, genotype distributions of 300,000-2,500,000 SNPs are assessed for association with the trait in GWAS. Since longevity is assumed to be determined by many genes with small effects, it could be a successful method to identify novel longevity loci. However, so far, GWAS for longevity in the LLS [50], Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) [51,52], NECS [53], German long-lived individuals [54], and SICS [55] have only identified one genome-wide significant ($P < 5 \times 10^{-8}$) locus: *APOE*, which has long been established as a longevity gene. Several other loci showed suggestive association with longevity ($P < 5 \times 10^{-6}$), namely *MINPP1* [51], *OTOL1* [52], and *CAMKIV* [55] (Figure 2.3). However, the effect of these loci on prospective mortality is not yet known. All GWAS-identified suggestive longevity loci are deleterious, i.e., the minor allele is associated with a decreased probability to become long-lived, and, as expected, their effects are small (odds ratio's > 0.5).

In general, to have sufficient power to detect significant effects, GWAS require much larger sample sizes than thus far accomplished for human longevity. One of the challenges of GWAS for longevity is that the lifespan variation induced by the genetic component is expected to be small relative to that induced by the environmental component (i.e., health care and nutrition). A large sample size, acquired through meta-analysis of GWAS (meta-GWAS), may cope with the so-called "phenocopies" and could potentially detect genome-wide significant loci besides *APOE*. Currently, two initiatives

Figure 2.3 Karyogram (adapted from http://hapmap.ncbi.nlm.nih.gov/karyogram/gwas.html) containing candidate genes whose association with longevity has been replicated in multiple association studies (green), candidate genes with interesting results from sequencing studies (yellow), interesting loci from linkage (logarithm (base 10) of odds ≥ 2.95) (red) and copy number variant (orange) studies, and loci that showed suggestive association with longevity ($P \leq 5 \times 10^{-6}$) in genome-wide association studies (blue).



for meta-GWAS for longevity are on-going. One consists of ~8,000 long-lived individuals (≥ 85 years of age) from all over Europe (EU longevity GWAS), while the other consists of ~6,000 long-lived individuals (≥ 90 years of age), collected in Northern America and Europe, from the CHARGE consortium. If these meta-GWAS lead to the identification of new loci that significantly associate with longevity, they should consequently be tested for an effect on prospective survival in middle and old age.

CNV studies identified potential longevity regions

Besides SNP analysis, several other methods have been applied to study the genetics of longevity, mainly using a prospective design (Box 2.1 and Figure 2.1). One study determined the effect of copy number variants (CNVs), which are deletions or duplications of stretches of DNA, on longevity in the Rotterdam Study and FHS. The meta-analysis of these cohorts showed an association between the burden of large (≥ 500 kb) CNVs and mortality at old age. In addition, they showed an association of common CNV regions on 11p15.5 and 14q21.3 [56] (Figure 2.3). However, to qualify them as longevity-regions, these associations still need to be replicated in several larger independent cross-sectional and prospective studies.

The same group also studied the effect of regions of homozygosity (ROHs), which are uninterrupted stretches of homozygous SNPs, on longevity in the Rotterdam Study and found no association between ROHs and survival into old age [57]. However, to rule out effects of ROHs on longevity, larger cross-sectional and prospective studies should be performed.

Linkage studies have discovered chromosomal regions linked to human longevity

The explorative studies of the genome for longevity effects actually started with linkage analysis in family-based designs (Box 2.1 and Figure 2.1). For this approach, the excess sharing of alleles between siblings identical by descent at 6,000-12,000 loci not in linkage disequilibrium over sharing by chance provides a likelihood for the presence of a longevity locus in any region on the genome. There have been several small-scale genome-wide linkage studies of long-lived sibling pairs ($n_{cases} < 300$) that showed inconsistent results [58-61] (Figure 2.3). Recently, a large linkage analysis for longevity has been performed in 2,118 nonagenarian Caucasian sibling pairs from the GEHA project. In this study, linkage with longevity was observed at chromosome 14q11.2 (logarithm (base 10) of odds (LOD) = 3.47), chromosome 17q12-22 (LOD = 2.95), chromosome 19p13.3-13.11 (LOD = 3.76), and chromosome 19q13.11-13.32 (LOD = 3.57) (Figure 2.3), of which the latter was explained by the ApoE ε4 and ApoE ε2 alleles [62]. Since the linkage at the remaining loci could not be explained by association of common variants, human familial longevity at these loci may be explained by rare variants.

Next-generation sequencing studies may reveal rare longevity-associated variants

Rare variants can be identified by applying next-generation (whole-genome or exome)

sequencing. In the case of Mendelian disorders and strong familial traits, next-generation sequencing of a limited number of well-selected individuals may reveal relevant alleles with functional consequences. Analysis of sequencing data is a bioinformatic challenge and good sample selection is therefore extremely important. The most informative individuals for next-generation sequencing in longevity research would be individuals from long-lived families with a long family history of longevity. One candidate gene study analyzed the complete coding region of *IGF1* and *IGF1R* using 2D gene scanning and DNA sequencing in centenarians and their offspring. Two rare nonsynonymous SNPs in *IGF1R* associated with both longevity and decreased IGF-1 signaling. This further indicates a role for GH/insulin/IGF-1 signaling genes in human longevity [63] (Figure 2.3).

For exploratory analyses, the whole genome can be analyzed. Up to now, this has been published for one female and one male supercentenarian [64]. To identify variants relevant for longevity, analysis on the genomes of many more of such individuals must be performed. Various initiatives are ongoing in which larger numbers of genomes of population and family-based centenarians are being sequenced, e.g., the Wellderly Study (consisting of ~1,000 individuals ≥ 80 years of age) and the LLS (consisting of 220 individuals ≥ 90 years of age).

Explorative studies identify transcriptomic profiles marking longevity

Since the genetic approaches have thus far provided little robust evidence for loci contributing to human aging and longevity, attempts have been made to identify such loci by exploration of the human transcriptome. The transcriptome of an individual reflects the influence of genetic variation, as well as the response to the environment. As an approach to find determinants of aging and longevity, transcriptomic studies require specific designs to disentangle primary and causal changes in gene expression from the consequences of aging.

Most studies of the transcriptome try to identify genes that show a differential change with chronological age and mainly use cross-sectional designs (Box 2.1 and Figure 2.1). In these designs, highly aged individuals are compared to young controls or categories of increasing age are examined. The larger studies are performed in whole blood, since this is the most accessible tissue. However, whole blood contains different cell populations, which may confound observed differences in gene expression. If possible, observations of differential gene expression should thus be adjusted for proportions of blood cell subsets, which is not always done. One study partly circumvented this problem by investigating the transcriptome of T cells from healthy individuals with ages ranging from 25 to over 95 years and highlighted similarities in gene expression profiles between young and "successfully aged" individuals [65]. This illustrates that cross-sectional transcriptome studies may be used to identify genes potentially indicative of the biological age of an individual by comparing the expression level of the gene for an individual to the average expression of individuals of his/her chronological age.

The transcriptomic studies focused on chronological age revealed that genes

and microRNAs involved in many different processes, e.g., oxidative phosphorylation, complement activation, and synaptic transmission, change with age [66-71]. The pathways that have been associated with chronological age include peroxisome proliferator-activated receptor, glucose and glutathione metabolism, and mTOR signaling [65]. The relevance of mTOR pathway genes for human aging has been further illustrated by associations of gene expression changes with chronological age in a candidate gene study of two independent human cohorts [69]. Most of the gene expression associations with chronological age in human populations have not yet been validated and replicated with comparable technology platforms in independent studies. In addition, transcriptomic studies on chronological age cannot rate which changes are causal and which are consequential to aging.

One way to overcome (part of) this problem is by using a family-based study design (Box 2.1 and Figure 2.1), in which the offspring of long-lived individuals – representing "healthy agers" – are compared to similar-aged controls from the general population. The differential gene expression profiles identified using this design may represent markers of healthy aging and familial longevity. This approach has been applied in the LLS to explore the transcriptome in whole blood for association with human familial longevity. Genes belonging to the mTOR pathway, as well as *ASF1A* and *IL7R*, were differentially expressed between offspring and controls [72,73]. In addition, the expression of mTOR genes in blood associated to prevalent diabetes and serum glucose. However, the association with familial longevity was not dependent on this. Thus, gene expression profiles in blood mark human longevity in middle age and potentially provide information on the pathways that contribute to healthy aging and longevity.

Epigenomic studies are at hand

Another molecular level that could provide additional insight in the processes of aging is the epigenome, the intermediate layer of genomic information between the genome and transcriptome. Epigenetic regulation of transcription is mediated by histon modification, DNA methylation, and microRNAs. Changes in the epigenome with chronological age have been explored and show that methylation patterns of genes involved in, e.g., development and morphogenesis, DNA binding and regulation of transcription [74-76] tend to change with age. A recent remarkable finding in a small study sample, confirmed in a cohort of 501 individuals ranging from birth to 99 years, was the progressive linear increase in methylation with age at the *ELOVL2* gene [77]. Because the epigenomic field recently became more accessible for the screening of large study populations, the identification of a new range of epigenetic biomarkers is at hand. To consider such epigenetic measures as markers for biological age, confounding of cell type distributions should be accounted for – like in transcriptomic studies – and effects should be established using various study designs.

In conclusion, up to now, genomic research to identify drivers of healthy aging and

longevity in humans has not yet delivered many robust longevity loci and pathways. However, larger studies, new methodologies, and the consistent use of different study designs to follow up results might help to unravel the genomic component of healthy aging and longevity.

Phenotypes that reflect biological aging

In addition to focusing on lifespan as primary phenotype, genomic studies into aging may profit from insights into phenotypes that reflect biological age. One can think of parameters or profiles reflecting immunosenescence or metabolic health established as pre-clinical measures in middle-aged individuals. In addition, phenotyping by novel noninvasive technologies, such as imaging (e.g., functional magnetic resonance imaging) and longitudinal and ambulatory measurements using electronic devices (e.g., gait speed, 24-hour glucose, and blood pressure), will improve the monitoring of the physiology of aging in epidemiological studies. Such research is often referred to as biomarker research and is aimed at finding parameters and profiles predicting long-term morbidity and/or mortality. Classical examples are blood pressure and hypertension as markers for clinical events in cardiovascular disease, joint-space width as marker for osteoarthritis, and bone mineral density and risk of fracture as markers for osteoporosis. Comparable to the genomic research of the transcriptome and epigenome, the main problem with biomarker research is that it is hard to disentangle the changes causal to aging and longevity from those that are a consequence of normative aging. For classical (e.g., leukocyte telomere length (LTL)) and novel potential biomarker of aging the four relevant phases to establish whether a quantitative parameter (or profile) is a biomarker of biological aging should be taken into account (see Introduction section).

Clinical biomarkers of biological age hint at metabolic processes

Several prospective studies investigated the effect of clinical, physical, and cognitive parameters on mortality. Many different parameters have been shown to influence mortality after 55 years of age in the general population [78-83]. To determine whether these parameters potentially contribute to longevity from middle age onwards, family-based studies have been performed (Box 2.1 and Figure 2.1), whereby the offspring of long-lived individuals is compared with similar-aged controls from the general population. Of the parameters that associate with mortality after 55 years of age, cortisol levels, digit symbol substitution test score, fasting glucose levels, free triiodothyronine levels, and gait speed also mark familial longevity in middle age [7,16,84-90] (Table 2.1). Together, these biomarkers of biological age suggest the involvement of metabolic processes in healthy aging and longevity.

Metabolic profiles seem promising predictive biomarkers

Instead of testing single quantitative parameters from a clinical perspective, the development of novel technologies and methodologies has made it possible to study age-related changes in the whole

glycome and metabolome [91,92]. These novel explorative omics studies could potentially be much more informative on physiological aspects of aging than the single parameters studied so far, since a single-point measurement contains a wealth of information. A cross-sectional comparison of "healthy agers" and similar-aged controls has shown that decreased levels of bisecting GlcNAc glycoforms of IgG and higher levels of specific *N*-glycan features mark healthy aging and familial longevity [93,94]. Datasets generated by metabolomic platforms provide information on biogenic amines, central metabolism, and lipids and can give insight into their relevance for morbidity and/or mortality, as was previously shown for cardiovascular disease [95]. In a recent study, using a prospective design, it was shown that a single-point nuclear magnetic resonance measurement could also predict incident risk of coronary heart disease, comparable to the gold standard, i.e., the Framingham risk score (unpublished results). However, additional prospective studies into morbidity and/or mortality, preferably on the basis of repeated measures, need to be performed to provide more information about the usefulness of metabolomic and glycomic profiles as biomarkers of biological age and longevity.

Integrating genomics and biomarker research

Once the use of established biomarkers of biological age is standardized, the biomarker information can be integrated into studies aimed at finding causal determinants of aging and longevity. An example of an integrated approach to identify lifespan regulating loci is represented by testing whether genetic variants associated with potential biomarkers also associate with longevity. To date, GWAS have identified many genetic variants that associate with age-related traits, such as LTL and features from glycome and metabolome profiles [96-98]. The joint effect of the majority of these variants on aging and longevity still needs to be determined. One study identified a haplotype in the *TERT* gene that was associated with increased LTL and longevity, which indicates that genetic variants associated with telomere length regulation might also play a role in longevity [99].

Conclusions and prospects

Over the past two decades the human aging field has built up the necessary resources to study the biology of aging and longevity by establishing human populations with a diversity of designs. Meta-analyses integrating genetic and phenotypic datasets have successfully identified variants associated with a range of age-related traits and diseases. Despite these accomplishments, the number of novel leads contributing to human lifespan regulation is limited. Although positive regions of linkage and suggestive GWAS hits have been reported, the field has not yet identified the loci that explain the clustering of longevity in families and the variation in biological aging rate in the population. As for animal models, down-signaling of the GH/insulin/IGF-1 and mTOR signaling pathways appears to be relevant in humans. These findings are being

Table 2.1 Potential biomarkers that have been shown to be associated with mortality in prospective studies of different age categories and their association with familial longevity in middle age.

Parameter	Effect on mortality	References	Effect on familial longevity	References
Ankle-arm index	>65 years	79	No	7
Systolic blood pressure	>55 years / >90 years	78,82,83	No	7
Diastolic blood pressure	>55 years / >90 years	78,79,82,83	No	7
Body mass index	>55 years / >90 years	82,83	No	7,89
High-density lipoprotein cholesterol	>85 years	78,82,83	No	7,16,89
Total cholesterol	>55 years / >90 years	78,82,83	No	7,16,89
Cortisol	>85 years	78	Yes	85
C-reactive protein	>55 years	78,79,83	No	88
Creatinine	>55 years / >65 years	78,79,83	No	7,90
Digit symbol substitution test score	>65 years	79	Yes	7
Fasting glucose	>65 years	78,79	Yes	7,86
Forced vital capacity	>65 years	79	No	7
Free triiodothyronine	>85 years	78	Yes	87
Gait speed	>65 years	81	Yes	7
Grip strength	>92 years	78,80	No	7,84
Instrumental activities of daily living impairment	>65 years	79,83	No	7

followed up by molecular and physiological profiling using skin, fat, and muscle tissue of long-lived family members and controls. Human studies now also include the response of nutrient sensing systems to the application of dietary and physical challenges.

The ongoing whole genome sequencing of centenarians and their families may provide novel genes contributing to longevity. Relevant variations may include gain-of-function mutations or heterozygous loss-of-function mutations in genes with deleterious effect late in life. Novel biomarkers represented by omics profiles and ambulatory measures to establish the biological aging rate (such as 24-hour glucose [100]) will be used in integrated analyses. It has already become feasible to study the integrative personal omics profiles, the combination of the genetic, transcriptomic, proteomic, metabolomic, and autoantibody profile of individuals [101].

In conclusion, novel methodologies, comprehensively applied to multiple studies of well-phenotyped (middle and highly aged) individuals from long-lived families and large prospective cohort studies, will help to connect human molecular epidemiology and biology in aging research. Ultimately, this will provide leads that can be followed up in animal studies.

Acknowledgements

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2011) under grant agreement nº. 259679. This study was supported by a grant from the Innovation-Oriented Research Program on Genomics (SenterNovem IGE05007), the Centre for Medical Systems Biology, and the Netherlands Consortium for Healthy Ageing (grant 050-060-810), all in the framework of the Netherlands Genomics Initiative, Netherlands Organization for Scientific Research (NWO), and by Unilever Colworth.

References

1. Oeppen J, Vaupel JW. Demography. Broken limits to life expectancy. *Science* 2002; **296**: 1029-31.
2. World Health Organization. World Health Statistics 2012. *Geneva: WHO* 2012.
3. Jagger C, Gillies C, Moscone F, Cambois E, van Oyen H, Nusselder W, *et al*. Inequalities in healthy life years in the 25 countries of the European Union in 2005: a cross-national meta-regression analysis. *Lancet* 2008; **372**: 2124-31.
4. Christensen K, McGue M, Petersen I, Jeune B, Vaupel JW. Exceptional longevity does not result in excessive levels of disability. *Proc Natl Acad Sci U S A* 2008; **105**: 13274-9.
5. Terry DF, Sebastiani P, Andersen SL, Perls TT. Disentangling the roles of disability and morbidity in survival to exceptional old age. *Arch Intern Med* 2008; **168**: 277-83.
6. Atzmon G, Schechter C, Greiner W, Davidson D, Rennert G, Barzilai N. Clinical phenotype of families with longevity. *J Am Geriatr Soc* 2004; **52**: 274-7.
7. Newman AB, Glynn NW, Taylor CA, Sebastiani P, Perls TT, Mayeux R, *et al*. Health and function of participants in the Long Life Family Study: A comparison with other cohorts. *Aging (Albany NY)* 2011; **3**: 63-76.
8. Terry DF, Wilcox MA, McCormick MA, Pennington JY, Schoenhofen EA, Andersen SL, *et al*. Lower all-cause, cardiovascular, and cancer mortality in centenarians' offspring. *J Am Geriatr Soc* 2004; **52**: 2074-6.
9. Westendorp RG, van Heemst D, Rozing MP, Frolich M, Mooijaart SP, Blauw GJ, *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**: 1634-7.
10. Slagboom PE, Beekman M, Passtoors WM, Deelen J, Vaarhorst AA, Boer JM, *et al*. Genomics of human longevity. *Philos Trans R Soc Lond B Biol Sci* 2011; **366**: 35-42.
11. Sprott RL. Biomarkers of aging and disease: introduction and definitions. *Exp Gerontol* 2010; **45**: 2-4.
12. Perls TT, Bochen K, Freeman M, Alpert L, Silver MH. Validity of reported age and centenarian prevalence in New England. *Age Ageing* 1999; **28**: 193-7.
13. Nebel A, Croucher PJ, Stiegeler R, Nikolaus S, Krawczak M, Schreiber S. No association between microsomal triglyceride transfer protein (MTP) haplotype and longevity in humans. *Proc Natl Acad Sci U S A* 2005; **102**: 7906-9.
14. Blanche H, Cabanne L, Sahbatou M, Thomas G. A study of French centenarians: are ACE and APOE associated with longevity? *C R Acad Sci III* 2001; **324**: 129-35.
15. Anselmi CV, Malovini A, Roncarati R, Novelli V, Villa F, Condorelli G, *et al*. Association of the FOXO3A locus with extreme longevity in a southern Italian centenarian study. *Rejuvenation Res* 2009; **12**: 95-104.
16. Barzilai N, Atzmon G, Schechter C, Schaefer EJ, Cupples AL, Lipton R, *et al*. Unique lipoprotein phenotype and genotype associated with exceptional longevity. *JAMA* 2003; **290**: 2030-40.
17. Skytthe A, Valensin S, Jeune B, Cevenini E, Balard F, Beekman M, *et al*. Design, recruitment, logistics, and data management of the GEHA (Genetics of Healthy Ageing) project. *Exp Gerontol* 2011; **46**: 934-45.
18. Schoenmaker M, de Craen AJ, de Meijer PH, Beekman M, Blauw GJ, Slagboom PE, *et al*. Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur J Hum Genet* 2006; **14**: 79-84.
19. Bootsma-van der Wiel A, van Exel E, de Craen AJ, Gussekloo J, Lagaay AM, Knook DL, *et al*. A high response is not essential to prevent selection bias: results from the Leiden 85-plus study. *J Clin Epidemiol* 2002; **55**: 1119-25.
20. Weverling-Rijnsburger AW, Blauw GJ, Lagaay AM, Knook DL, Meinders AE, Westendorp RG. Total cholesterol and risk of mortality in the oldest old. *Lancet* 1997; **350**: 1119-23.
21. Collerton J, Barrass K, Bond J, Eccles M, Jagger C, James O, *et al*. The Newcastle 85+ study: biological, clinical and psychosocial factors associated with healthy ageing: study protocol. *BMC Geriatr* 2007; **7**: 14.

22. Collerton J, Davies K, Jagger C, Kingston A, Bond J, Eccles MP, *et al*. Health and disease in 85 year olds: baseline findings from the Newcastle 85+ cohort study. *BMJ* 2009; **339**: b4904.
23. Nybo H, Gaist D, Jeune B, Bathum L, McGue M, Vaupel JW, *et al*. The Danish 1905 cohort: a genetic-epidemiological nationwide survey. *J Aging Health* 2001; **13**: 32-46.
24. Hofman A, Breteler MM, van Duijn CM, Janssen HL, Krestin GP, Kuipers EJ, *et al*. The Rotterdam Study: 2010 objectives and design update. *Eur J Epidemiol* 2009; **24**: 553-72.
25. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol* 1979; **110**: 281-90.
26. Herskind AM, McGue M, Holm NV, Sorensen TI, Harvald B, Vaupel JW. The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870-1900. *Hum Genet* 1996; **97**: 319-23.
27. Hjelmborg JV, Iachine I, Skytthe A, Vaupel JW, McGue M, Koskenvuo M, *et al*. Genetic influence on human lifespan and longevity. *Hum Genet* 2006; **119**: 312-21.
28. Perls T, Shea-Drinkwater M, Bowen-Flynn J, Ridge SB, Kang S, Joyce E, *et al*. Exceptional familial clustering for extreme longevity in humans. *J Am Geriatr Soc* 2000; **48**: 1483-5.
29. Gavrilova NS, Gavrilov LA, Evdokushkina GN, Semyonova VG, Gavrilova AL, Evdokushkina NN, *et al*. Evolution, mutations, and human longevity: European royal and noble families. *Hum Biol* 1998; **70**: 799-804.
30. Finch CE, Tanzi RE. Genetics of aging. *Science* 1997; **278**: 407-11.
31. Stahl EA, Wegmann D, Trynka G, Gutierrez-Achury J, Do R, Voight BF, *et al*. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat Genet* 2012; **44**: 483-9.
32. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, *et al*. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet* 2011; **43**: 519-25.
33. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, *et al*. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 2009; **106**: 9362-7.
34. Beekman M, Nederstigt C, Suchiman HE, Kremer D, van der Breggen R, Lakenberg N, *et al*. Genome-wide association study (GWAS)-identified disease risk alleles do not compromise human longevity. *Proc Natl Acad Sci U S A* 2010; **107**: 18046-9.
35. Christensen K, Johnson TE, Vaupel JW. The quest for genetic determinants of human longevity: challenges and insights. *Nat Rev Genet* 2006; **7**: 436-48.
36. de Magalhaes JP, Budovsky A, Lehmann G, Costa J, Li Y, Fraifeld V, *et al*. The Human Ageing Genomic Resources: online databases and tools for biogerontologists. *Aging Cell* 2009; **8**: 65-72.
37. Fontana L, Partridge L, Longo VD. Extending healthy life span--from yeast to humans. *Science* 2010; **328**: 321-6.
38. Pettan-Brewer C, Treuting PM. Practical pathology of aging mice. *Pathobiol Aging Age Relat Dis* 2011; **1**.
39. Special Issue on Genetics of Aging. *J Gerontol A Biol Sci Med Sci* 2012; **67**: 467-519.
40. Soerensen M, Dato S, Christensen K, McGue M, Stevnsner T, Bohr VA, *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**: 1010-7.
41. Soerensen M, Dato S, Tan Q, Thinggaard M, Kleindorp R, Beekman M, *et al*. Evidence from case-control and longitudinal studies supports associations of genetic variation in APOE, CETP, and IL6 with human longevity. *Age (Dordr)* 2013; **35**: 487-500.
42. Zeng Y, Cheng L, Chen H, Cao H, Hauser ER, Liu Y, *et al*. Effects of FOXO genotypes on longevity: a biodemographic analysis. *J Gerontol A Biol Sci Med Sci* 2010; **65**: 1285-99.

43. Mahley RW, Rall SC, Jr. Apolipoprotein E: far more than a lipid transport protein. *Annu Rev Genomics Hum Genet* 2000; **1**: 507-37.
44. van der Horst A, Burgering BM. Stressing the role of FoxO proteins in lifespan and disease. *Nat Rev Mol Cell Biol* 2007; **8**: 440-50.
45. Broer L, Demerath EW, Garcia ME, Homuth G, Kaplan RC, Lunetta KL, *et al*. Association of heat shock proteins with all-cause mortality. *Age (Dordr)* 2013; **35**: 1367-76.
46. Deelen J, Uh HW, Monajemi R, van Heemst D, Thijssen PE, Bohringer S, *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**: 235-49.
47. Pawlikowska L, Hu D, Huntsman S, Sung A, Chu C, Chen J, *et al*. Association of common genetic variation in the insulin/IGF1 signaling pathway with human longevity. *Aging Cell* 2009; **8**: 460-72.
48. Soerensen M, Dato S, Tan Q, Thinggaard M, Kleindorp R, Beekman M, *et al*. Human longevity and variation in GH/IGF-1/insulin signaling, DNA damage signaling and repair and pro/antioxidant pathway genes: Cross sectional and longitudinal studies. *Exp Gerontol* 2012; **47**: 379-87.
49. van Heemst D, Beekman M, Mooijaart SP, Heijmans BT, Brandt BW, Zwaan BJ, *et al*. Reduced insulin/IGF-1 signalling and human longevity. *Aging Cell* 2005; **4**: 79-85.
50. Deelen J, Beekman M, Uh HW, Helmer Q, Kuningas M, Christiansen L, *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**: 686-98.
51. Newman AB, Walter S, Lunetta KL, Garcia ME, Slagboom PE, Christensen K, *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**: 478-87.
52. Walter S, Atzmon G, Demerath EW, Garcia ME, Kaplan RC, Kumari M, *et al*. A genome-wide association study of aging. *Neurobiol Aging* 2011; **32**: 2109-28.
53. Sebastiani P, Solovieff N, Dewan AT, Walsh KM, Puca A, Hartley SW, *et al*. Genetic signatures of exceptional longevity in humans. *PLoS One* 2012; **7**: e29848.
54. Nebel A, Kleindorp R, Caliebe A, Nothnagel M, Blanche H, Junge O, *et al*. A genome-wide association study confirms APOE as the major gene influencing survival in long-lived individuals. *Mech Ageing Dev* 2011; **132**: 324-30.
55. Malovini A, Illario M, Iaccarino G, Villa F, Ferrario A, Roncarati R, *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**: 283-91.
56. Kuningas M, Estrada K, Hsu YH, Nandakumar K, Uitterlinden AG, Lunetta KL, *et al*. Large common deletions associate with mortality at old age. *Hum Mol Genet* 2011; **20**: 4290-6.
57. Kuningas M, McQuillan R, Wilson JF, Hofman A, van Duijn CM, Uitterlinden AG, *et al*. Runs of homozygosity do not influence survival to old age. *PLoS One* 2011; **6**: e22580.
58. Beekman M, Blauw GJ, Houwing-Duistermaat JJ, Brandt BW, Westendorp RG, Slagboom PE. 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**: 355-62.
59. Boyden SE, Kunkel LM. High-density genomewide linkage analysis of exceptional human longevity identifies multiple novel Loci. *PLoS One* 2010; **5**: e12432.
60. Puca AA, Daly MJ, Brewster SJ, Matise TC, Barrett J, Shea-Drinkwater M, *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**: 10505-8.
61. Reed T, Dick DM, Uniacke SK, Foroud T, Nichols WC. Genome-wide scan for a healthy aging phenotype provides support for a locus near D4S1564 promoting healthy aging. *J Gerontol A Biol Sci Med Sci* 2004; **59**: 227-32.
62. Beekman M, Blanche H, Perola M, Hervonen A, Bezrukov V, Sikora E, et al

Genome-wide linkage analysis for human longevity: Genetics of Healthy Ageing Study. *Aging Cell* 2013 ; **12**: 184-93.

63. Suh Y, Atzmon G, Cho MO, Hwang D, Liu B, Leahy DJ, *et al*. Functionally significant insulin-like growth factor I receptor mutations in centenarians. *Proc Natl Acad Sci U S A* 2008; **105**: 3438-42.
64. Sebastiani P, Riva A, Montano M, Pham P, Torkamani A, Scherba E, *et al*. Whole genome sequences of a male and female supercentenarian, ages greater than 114 years. *Front Genet* 2011; **2**: 90.
65. Remondini D, Salvioli S, Francesconi M, Pierini M, Mazzatti DJ, Powell JR, *et al*. Complex patterns of gene expression in human T cells during in vivo aging. *Mol Biosyst* 2010; **6**: 1983-92.
66. de Magalhaes JP, Curado J, Church GM. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics* 2009; **25**: 875-81.
67. Elsharawy A, Keller A, Flachsbart F, Wendschlag A, Jacobs G, Kefer N, *et al*. Genome-wide miRNA signatures of human longevity. *Aging Cell* 2012; **11**: 607-16.
68. Harries LW, Hernandez D, Henley W, Wood AR, Holly AC, Bradley-Smith RM, *et al*. Human aging is characterized by focused changes in gene expression and deregulation of alternative splicing. *Aging Cell* 2011; **10**: 868-78.
69. Harries LW, Fellows AD, Pilling LC, Hernandez D, Singleton A, Bandinelli S, *et al*. Advancing age is associated with gene expression changes resembling mTOR inhibition: Evidence from two human populations. *Mech Ageing Dev* 2012; **133**: 556-62.
70. Hong MG, Myers AJ, Magnusson PK, Prince JA. Transcriptome-wide assessment of human brain and lymphocyte senescence. *PLoS One* 2008; **3**: e3024.
71. Passtoors WM, Beekman M, Gunn D, Boer JM, Heijmans BT, Westendorp RG, *et al*. Genomic studies in ageing research: the need to integrate genetic and gene expression approaches. *J Intern Med* 2008; **263**: 153-66.
72. Passtoors WM, Boer JM, Goeman JJ, van den Akker EB, Deelen J, Zwaan BJ, *et al*. Transcriptional profiling of human familial longevity indicates a role for ASF1A and IL7R. *PLoS One* 2012; **7**: e27759.
73. Passtoors WM, Beekman M, Deelen J, van der Breggen R, Maier AB, Guigas B, *et al*. Gene expression analysis of mTOR pathway: association with human longevity. *Aging Cell* 2013; **12**: 24-31.
74. Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, *et al*. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet* 2012; **8**: e1002629.
75. Hernandez DG, Nalls MA, Gibbs JR, Arepalli S, van der Brug M, Chong S, *et al*. Distinct DNA methylation changes highly correlated with chronological age in the human brain. *Hum Mol Genet* 2011; **20**: 1164-72.
76. Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H, *et al*. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res* 2010; **20**: 434-9.
77. Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D, *et al*. Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell* 2012; **12**: 184-93.
78. Martin-Ruiz C, Jagger C, Kingston A, Collerton J, Catt M, Davies K, *et al*. Assessment of a large panel of candidate biomarkers of ageing in the Newcastle 85+ study. *Mech Ageing Dev* 2011; **132**: 496-502.
79. Newman AB, Sachs MC, Arnold AM, Fried LP, Kronmal R, Cushman M, *et al*. Total and cause-specific mortality in the cardiovascular health study. *J Gerontol A Biol Sci Med Sci* 2009; **64**: 1251-61.
80. Oksuzyan A, Maier H, McGue M, Vaupel JW, Christensen K. Sex differences in the level and rate of change of physical function and grip strength in the Danish 1905-cohort study. *J Aging Health* 2010; **22**: 589-610.
81. Studenski S, Perera S, Patel K, Rosano C, Faulkner K, Inzitari M, *et al*. Gait speed and survival in older adults. *JAMA* 2011; **305**: 50-8.
82. van Vliet P, Oleksik AM, van Heemst D, de Craen AJ, Westendorp RG. Dynamics of traditional metabolic risk factors associate with specific causes of death in old age. *J*

Gerontol A Biol Sci Med Sci 2010; **65**: 488-94.
83. Walter S, Mackenbach J, Voko Z, Lhachimi S, Ikram MA, Uitterlinden AG, *et al*. Genetic, physiological, and lifestyle predictors of mortality in the general population. *Am J Public Health* 2012; **102**: e3-10.
84. Ling CH, de Craen AJ, Slagboom PE, Westendorp RG, Maier AB. Handgrip strength at midlife and familial longevity : The Leiden Longevity Study. *Age (Dordr)* 2012; **34**: 1261-8.
85. Noordam R, Gunn DA, Tomlin CC, Rozing MP, Maier AB, Slagboom PE, *et al*. Cortisol serum levels in familial longevity and perceived age: The Leiden Longevity Study. *Psychoneuroendocrinology* 2012; **37**: 1669-75
86. Rozing MP, Westendorp RG, de Craen AJ, Frolich M, de Goeij MC, Heijmans BT, *et al*. Favorable glucose tolerance and lower prevalence of metabolic syndrome in offspring without diabetes mellitus of nonagenarian siblings: the Leiden longevity study. *J Am Geriatr Soc* 2010; **58**: 564-9.
87. Rozing MP, Westendorp RG, de Craen AJ, Frolich M, Heijmans BT, Beekman M, *et al*. Low serum free triiodothyronine levels mark familial longevity: the Leiden Longevity Study. *J Gerontol A Biol Sci Med Sci* 2010; **65**: 365-8.
88. Rozing MP, Mooijaart SP, Beekman M, Wijsman CA, Maier AB, Bartke A, *et al*. C-reactive protein and glucose regulation in familial longevity. *Age (Dordr)* 2011; **33**: 623-30.
89. Vaarhorst AA, Beekman M, Suchiman HE, van Heemst D, Houwing-Duistermaat JJ, Westendorp RG, *et al*. Lipid metabolism in long-lived families: the Leiden Longevity Study. *Age (Dordr)* 2011; **33**: 219-27.
90. Wijsman CA, van Heemst D, Rozing MP, Slagboom PE, Beekman M, de Craen AJ, *et al*. Homocysteine and familial longevity: the Leiden Longevity Study. *PLoS One* 2011; **6**: e17543.
91. Dall'olio F, Vanhooren V, Chen CC, Slagboom PE, Wuhrer M, Franceschi C. N-glycomic biomarkers of biological aging and longevity: A link with inflammaging. *Ageing Res Rev* 2012; **12**: 685-98.
92. Yu Z, Zhai G, Singmann P, He Y, Xu T, Prehn C, *et al*. Human serum metabolic profiles are age dependent. *Aging Cell* 2012; 11: 960-7.
93. Ruhaak LR, Uh HW, Beekman M, Koeleman CA, Hokke CH, Westendorp RG, *et al*. Decreased levels of bisecting GlcNAc glycoforms of IgG are associated with human longevity. *PLoS One* 2010; **5**: e12566.
94. Ruhaak LR, Uh HW, Beekman M, Hokke CH, Westendorp RG, Houwing-Duistermaat J, *et al*. Plasma protein N-glycan profiles are associated with calendar age, familial longevity and health. *J Proteome Res* 2011; **10**: 1667-74.
95. Shah SH, Bain JR, Muehlbauer MJ, Stevens RD, Crosslin DR, Haynes C, *et al*. Association of a peripheral blood metabolic profile with coronary artery disease and risk of subsequent cardiovascular events. *Circ Cardiovasc Genet* 2010; **3**: 207-14.
96. Codd V, Nelson CP, Albrecht E, Mangino M, Deelen J, Buxton JL, et al. Identification of seven loci affecting mean telomere length and their association with disease. *Nat Genet* 2013; **45**: 422-7.
97. Huffman JE, Knezevic A, Vitart V, Kattla J, Adamczyk B, Novokmet M, *et al*. Polymorphisms in B3GAT1, SLC9A9 and MGAT5 are associated with variation within the human plasma N-glycome of 3533 European adults. *Hum Mol Genet* 2011; **20**: 5000-11.
98. Kettunen J, Tukiainen T, Sarin AP, Ortega-Alonso A, Tikkanen E, Lyytikainen LP, *et al*. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat Genet* 2012; **44**: 269-76.
99. Atzmon G, Cho M, Cawthon RM, Budagov T, Katz M, Yang X, *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**: 1710-7.
100. Wijsman CA, van Heemst D, Hoogeveen ES, Slagboom PE, Maier AB, de Craen AJ, *et al*. Ambulant 24-hour glucose rhythms mark calendar and biological age in

apparently healthy individuals. *Aging Cell* 2013; **12**: 207-13.
101. Chen R, Mias GI, Li-Pook-Than J, Jiang L, Lam HY, Chen R, *et al*. Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* 2012; **148**: 1293-307.
102. Nebel A, Flachsbart F, Till A, Caliebe A, Blanche H, Arlt A, *et al*. A functional EXO1 promoter variant is associated with prolonged life expectancy in centenarians. *Mech Ageing Dev* 2009; **130**: 691-9.
103. Flachsbart F, Caliebe A, Kleindorp R, Blanche H, von Eller-Eberstein H, Nikolaus S, *et al*. Association of FOXO3A variation with human longevity confirmed in German centenarians. *Proc Natl Acad Sci U S A* 2009; **106**: 2700-5.
104. Kleindorp R, Flachsbart F, Puca AA, Malovini A, Schreiber S, Nebel A. Candidate gene study of FOXO1, FOXO4, and FOXO6 reveals no association with human longevity in Germans. *Aging Cell* 2011; **10**: 622-8.
105. Kuningas M, Magi R, Westendorp RG, Slagboom PE, Remm M, van Heemst D. Haplotypes in the human Foxo1a and Foxo3a genes; impact on disease and mortality at old age. *Eur J Hum Genet* 2007; **15**: 294-301.
106. Kuningas M, Putters M, Westendorp RG, Slagboom PE, van Heemst D. SIRT1 gene, age-related diseases, and mortality: the Leiden 85-plus study. *J Gerontol A Biol Sci Med Sci* 2007; **62**: 960-5.
107. Lescai F, Blanche H, Nebel A, Beekman M, Sahbatou M, Flachsbart F, *et al*. Human longevity and 11p15.5: a study in 1321 centenarians. *Eur J Hum Genet* 2009; **17**: 1515-9.
108. Li Y, Wang WJ, Cao H, Lu J, Wu C, Hu FY, *et al*. Genetic association of FOXO1A and FOXO3A with longevity trait in Han Chinese populations. *Hum Mol Genet* 2009; **18**: 4897-904.
109. Nygaard M, Soerensen M, Flachsbart F, Mengel-From J, Tan Q, Schreiber S, *et al*. AKT1 fails to replicate as a longevity-associated gene in Danish and German nonagenarians and centenarians. *Eur J Hum Genet* 2013; **21**: 574-7.
110. Willcox BJ, Donlon TA, He Q, Chen R, Grove JS, Yano K, *et al*. FOXO3A genotype is strongly associated with human longevity. *Proc Natl Acad Sci U S A* 2008; **105**: 13987-92.
111. Capurso C, Solfrizzi V, D'introno A, Colacicco AM, Capurso SA, Semeraro C, *et al*. Interleukin 6 variable number of tandem repeats (VNTR) gene polymorphism in centenarians. *Ann Hum Genet* 2007; **71**: 843-8.
112. Di Bona D, Vasto S, Capurso C, Christiansen L, Deiana L, Franceschi C, *et al*. Effect of interleukin-6 polymorphisms on human longevity: a systematic review and meta-analysis. *Ageing Res Rev* 2009; **8**: 36-42.
113. Li J, Niu W, Qi Y, Mayila W, Zhu P, Muhuyati, *et al*. Interactive association of heat shock protein 70 genes variants with natural longevity in Xinjiang Hetian Uygur ethnicity. *Transl Res* 2009; **154**: 257-64.
114. Nebel A, Flachsbart F, Schafer A, Nothnagel M, Nikolaus S, Mokhtari NE, *et al*. Role of the toll-like receptor 4 polymorphism Asp299Gly in longevity and myocardial infarction in German men. *Mech Ageing Dev* 2007; **128**: 409-11.
115. Singh R, Kolvraa S, Bross P, Christensen K, Bathum L, Gregersen N, *et al*. Anti-inflammatory heat shock protein 70 genes are positively associated with human survival. *Curr Pharm Des* 2010; **16**: 796-801.
116. Caliebe A, Kleindorp R, Blanche H, Christiansen L, Puca AA, Rea IM, *et al*. No or only population-specific effect of PON1 on human longevity: a comprehensive meta-analysis. *Ageing Res Rev* 2010; **9**: 238-44.
117. Lescai F, Marchegiani F, Franceschi C. PON1 is a longevity gene: results of a meta-analysis. *Ageing Res Rev* 2009; **8**: 277-84.
118. Novelli V, Anselmi CV, Roncarati R, Guffanti G, Malovini A, Piluso G, *et al*. Lack of replication of genetic associations with human longevity. *Biogerontology* 2008; **9**: 85-92.
119. Soerensen M, Christensen K, Stevnsner T, Christiansen L. The Mn-superoxide dismutase single nucleotide polymorphism rs4880 and the glutathione peroxidase 1 single nucleotide polymorphism rs1050450 are associated with aging and longevity in the oldest old. *Mech Ageing Dev* 2009; **130**: 308-14.

120. Soerensen M, Thinggaard M, Nygaard M, Dato S, Tan Q, Hjelmborg J, *et al*. Genetic variation in TERT and TERC and human leukocyte telomere length and longevity: a cross-sectional and longitudinal analysis. *Aging Cell* 2012; **11**: 223-7.
121. Jacobsen R, Martinussen T, Christiansen L, Jeune B, Andersen-Ranberg K, Vaupel JW, *et al*. Increased effect of the ApoE gene on survival at advanced age in healthy and long-lived Danes: two nationwide cohort studies. *Aging Cell* 2010; **9**: 1004-9.
122. Maruszak A, Peplonska B, Safranow K, Chodakowska-Zebrowska M, Barcikowska M, Zekanowski C. TOMM40 rs10524523 polymorphism's role in late-onset Alzheimer's disease and in longevity. *J Alzheimers Dis* 2012; **28**: 309-22.
123. McKay GJ, Silvestri G, Chakravarthy U, Dasari S, Fritsche LG, Weber BH, *et al*. Variations in apolipoprotein E frequency with age in a pooled analysis of a large group of older people. *Am J Epidemiol* 2011; **173**: 1357-64.
124. Napolioni V, Gianni P, Carpi FM, Predazzi IM, Lucarini N. APOE haplotypes are associated with human longevity in a Central Italy population: evidence for epistasis with HP 1/2 polymorphism. *Clin Chim Acta* 2011; **412**: 1821-4.
125. Schachter F, Faure-Delanef L, Guenot F, Rouger H, Froguel P, Lesueur-Ginot L, *et al*. Genetic associations with human longevity at the APOE and ACE loci. *Nat Genet* 1994; **6**: 29-32.
126. Sebastiani P, Montano M, Puca A, Solovieff N, Kojima T, Wang MC, *et al*. RNA editing genes associated with extreme old age in humans and with lifespan in C. elegans. *PLoS One* 2009; **4**: e8210.
127. Conneely KN, Capell BC, Erdos MR, Sebastiani P, Solovieff N, Swift AJ, *et al*. Human longevity and common variations in the LMNA gene: a meta-analysis. *Aging Cell* 2012; **11**: 475-81.
128. Flachsbart F, Caliebe A, Nothnagel M, Kleindorp R, Nikolaus S, Schreiber S, *et al*. Depletion of potential A2M risk haplotype for Alzheimer's disease in long-lived individuals. *Eur J Hum Genet* 2010; **18**: 59-61.
129. Khabour OF, Abdelhalim ES, bu-Wardeh A. Association between SOD2 T-9C and MTHFR C677T polymorphisms and longevity: a study in Jordanian population. *BMC Geriatr* 2009; **9**: 57.
130. Yang JK, Gong YY, Xie L, Lian SG, Yang J, Xu LY, *et al*. Lack of genetic association between the angiotensin-converting enzyme gene insertion/deletion polymorphism and longevity in a Han Chinese population. *J Renin Angiotensin Aldosterone Syst* 2009; **10**: 115-8.
131. Zajc Petranovic M, Skaric-Juric T, Smolej Narancic N, Tomas Z, Krajacic P, Milicic J, *et al*. Angiotensin-converting enzyme deletion allele is beneficial for the longevity of Europeans. *Age (Dordr)* 2012; 34: 583-95.

Supplementary Information

Table S2.1 Identified candidate genes involved in longevity.

Pathway	Analysis	Genes	References
DNA damage signaling and repair	Pathway-based	*EXO1*, *POLB*, *NTLH1*, *RAD23B*, *RAD52*, ***WRN***	48
	Single gene-based	*EXO1*	102
GH/insulin/IGF-1 signaling	Pathway-based	*AKT1*, *AKT3*, *FOXO3A*, *FOXO4*, *GHRHR*, *GHSR*, *IGF2*, *IGF2R*, *INS*, ***KL***, *PIK3CA*, *SGK1*, *SGK2*, *YWHAG*	46-49
	Single gene-based	*FOXO1*, *FOXO3A*, ***IGF1R***, ***SIRT3***	15,40,41,103-110
Immune regulation	Pathway-based	*HSF2*	45
	Single gene-based	***HSPA1A***, ***HSPA1L***, *HSPA14*, *HSPA1B*, ***IL6***, *TLR4*	41,111-115
Pro/antioxidant	Pathway-based	*GSR*	48
	Single gene-based	*GPX1*, ***PON1***, ***SOD2***	116-119
Telomere maintenance	Pathway-based	*POT1*	46
	Single gene-based	*TERC*, *TERT*	99.120
Lipoprotein metabolism	Single gene based	***APOE***, ***CETP***	41,118,121-125
RNA editing	Single gene-based	*ADARB1, ADARB2*	126
Other	Single gene-based	*A2M*, ***ACE***, *LMNA*	41,118,127-131

GH, growth hormone; *IGF-1*, insulin-like growth factor 1. Genes depicted in **bold** were previously reported in the review of Christensen and colleagues [35].

Chapter 3

Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited

Joris Deelen*, Marian Beekman*, Hae-Won Uh, Quinta Helmer, Maris Kuningas, Lene Christiansen, Dennis Kremer, Ruud van der Breggen, H Eka D Suchiman, Nico Lakenberg, Erik B van den Akker, Willemijn M Passtoors, Henning Tiemeier, Diana van Heemst, Anton J M de Craen, Fernando Rivadeneira, Eco J de Geus, Markus Perola, Frans J van der Ouderaa, David A Gunn, Dorret I Boomsma, André G Uitterlinden, Kaare Christensen, Cornelia M van Duijn, Bastiaan T Heijmans, Jeanine J Houwing-Duistermaat, Rudi G J Westendorp and P Eline Slagboom

*These authors contributed equally to this work.

Aging Cell 2011; **10**: 686-98.

Abstract

By studying the loci that contribute to human longevity, we aim to identify mechanisms that contribute to healthy aging. To identify such loci, we performed a genome-wide association study (GWAS) comparing 403 unrelated nonagenarians from long-living families included in the Leiden Longevity Study (LLS) and 1,670 younger population controls. The strongest candidate single nucleotide polymorphisms (SNPs) from this GWAS have been analyzed in a meta-analysis of nonagenarian cases from the Rotterdam Study, Leiden 85-plus study, and Danish 1905 cohort. Only one of the 62 prioritized SNPs from the GWAS analysis ($P < 1 \times 10^{-4}$) showed genome-wide significance with survival into old age in the meta-analysis of 4,149 nonagenarian cases and 7,582 younger controls (OR = 0.71 (95% CI 0.65 – 0.77), $P = 3.39 \times 10^{-17}$). This SNP, rs2075650, is located in *TOMM40* at chromosome 19q13.32 close to the *APOE* gene. Although there was only moderate linkage disequilibrium between rs2075650 and the apolipoprotein E (ApoE) ε4 defining SNP rs429358, we could not find an ApoE-independent effect of rs2075650 on longevity, either in cross-sectional or in longitudinal analyses. As expected, rs429358 associated with metabolic phenotypes in the offspring of the nonagenarian cases from the LLS and their partners. In addition, we observed a novel association between this locus and serum levels of insulin-like growth factor 1 in women ($P = 0.005$). In conclusion, the major locus determining familial longevity up to high age as detected by GWAS was marked by rs2075650, which tags the deleterious effects of the ApoE ε4 allele. No other major longevity locus was found.

Introduction

Worldwide human populations have shown an increase in mean life expectancy in the past two centuries [1]. This is mainly because of environmental factors, such as improved hygiene, nutrition, and health care. The large variation in healthy lifespan among the elderly has prompted research into the determinants of aging and lifespan regulation. The genetic contribution to human lifespan variation was estimated at 25-30% in twin studies [2-4]. The most prominent genetic influence is observed in families in which the capacity to attain a long lifespan clusters [5,6]. Exceptional longevity can be reached with a low degree of age-related disability [7,8], raising the question whether protective mechanisms against disease exist in long-lived subjects.

In most experimentally modified animal model systems, single-gene mutations in many different genes have major life extension effects [9,10]. However, natural human and animal longevity is presumed to be a complex trait [11]. In humans, both candidate gene and genome-wide genetic association approaches have been applied in an attempt to identify longevity loci. The frequency of genetic variants has been typically compared between nonagenarian cases and young controls, revealing loci at which genetic variants may contribute to a higher or lower probability of survival into old age. The initial candidate gene studies aimed at finding human longevity genes were dominated by contradictory results [12]. The more consistent evidence obtained by repeated observation in independent cohort studies for association with longevity has so far only been observed for three loci, the *APOE* locus [12,13], the *FOXO3A* locus [14-17], and the *AKT1* locus [15]. Thus, despite the expectation that longevity would be influenced by many genetic variants with small effect sizes, the effect of variants has consistently been shown in only three genes.

Hypothesis-free genome-wide approaches have also been undertaken. Genome-wide linkage scans reported evidence for linkage with longevity on chromosome 4q25 [18], 3p24-22, 9q31-34, and 12q24 [19]. However, the evidence for these loci is still very weak, as the results, obtained in centenarians and their families, could not be replicated in nonagenarian sibling pairs [20] or have yet to be tested in other studies. A meta-genome-wide association study (GWAS) for survival to 90 years or older in 1,836 cases and 1,955 controls did not find any significant genome-wide associations [21]. Thus far, hypothesis-free approaches have not identified any loci involved in longevity.

In a few studies, such as the Ashkenazi Jewish Centenarian Study and the Leiden Longevity Study (LLS), different generations of long-lived families are being investigated for parameters and pathways contributing to the longevity phenotype [6,22]. The survival benefit of the LLS families is marked by a 30% decreased mortality risk in the survival analysis of three generations, i.e., the parents of the probands in this study (nonagenarian sibling pairs), their unselected additional siblings, and their offspring [6]. As compared to their partners, the offspring of nonagenarians siblings have a lower prevalence of type 2 diabetes, myocardial infarction, and hypertension

[23], a beneficial glucose, lipid, and thyroid metabolism, and a preservation of insulin sensitivity with age [24-28]. Hence, in middle age, these families display beneficial metabolic profiles.

Because the longevity phenotype is inherited in the LLS families, they offer a route to identify genetic variants that influence human longevity. Previously, we tested whether the absence of GWAS-identified alleles promoting common diseases might explain their familial longevity [29]. Longevity was not easily explained by the absence of disease-susceptibility alleles. More likely therefore, the genome of the long-lived harbors longevity-promoting alleles. To identify such loci, we performed a GWAS comparing nonagenarian siblings from the LLS and younger population controls. We subsequently investigated emerging candidate single nucleotide polymorphisms (SNPs) in nonagenarian cases from the Rotterdam Study (RS), the Leiden 85-plus study, and the Danish 1905 cohort.

Results

GWAS

A GWAS was performed in nonagenarian participants from the LLS and middle-aged controls from the RS. Genotype data for 516,721 SNPs that passed quality control thresholds were analyzed in a comparison of 403 unrelated nonagenarians (94 years on average) and 1,670 controls (58 years on average). A flow chart of the consecutive

Figure 3.1 Flow chart of experimental work.

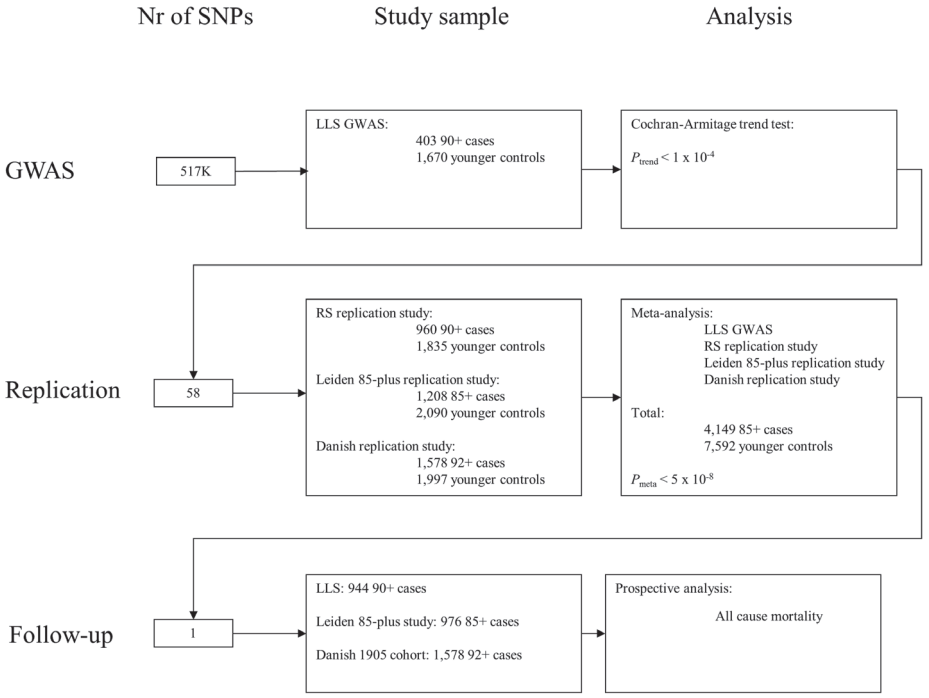


Table 3.1 Characteristics of the genotyped samples used for analysis.

Study	SNPs	Samples	*n*	Mean age	Age range	Men/women
LLS GWAS	517K	Cases	403	94	89 - 102	137/266
	517K	Controls	1,670	58	55 - 59	745/925
RS replication study	58	Cases	960	94	90 - 106	217/743
	58	Controls	1,835	62	60 - 65	809/1,026
Leiden 85-plus replication study	58	Cases	1,208	92	85 - 109	372/836
	58	Controls	2,090	35	15 - 70	743/1,347
Danish replication study	58	Cases	1,578	93	92 - 93	430/1,148
	58	Controls	1,997	57	46 - 68	900/1,097

SNPs, single nucleotide polymorphisms; *LLS*, Leiden Longevity Study; *GWAS*, genome-wide association study; *RS*, Rotterdam Study.

analysis steps is depicted in Figure 3.1 and a description of the population samples investigated in the GWAS and subsequent replication studies is given in Table 3.1. Results of the association analysis of stage 1 are depicted in Figure S3.1. None of the SNPs reached genome-wide significance ($P < 5 \times 10^{-8}$).

Replication studies

We prioritized the SNPs that had the most significant association with survival into old age according to the analysis of stage 1 ($P < 1 \times 10^{-4}$, Table S3.1). For 58 of the 62 selected SNPs, successful genotyping was obtained in the replication cohorts. In stage 2, these 58 SNPs were tested for association comparing 960 RS replication cases (mean age of 93 years), 1,208 Leiden 85-plus replication cases (mean age of 92 years), and 1,578 Danish replication cases (mean age of 93 years) with appropriate middle-aged population controls (Table 3.1). Meta-analysis for the 58 SNPs, comprising a total of 4,149 nonagenarian cases and 7,582 younger controls (from the LLS GWAS, RS replication, Leiden 85-plus replication, and Danish replication studies), was performed.

Rs2075650 on chromosome 19 was the only SNP that was associated with survival into old age at the genome-wide significance level ($P = 3.39 \times 10^{-17}$) (Table S3.2A). The minor allele was underrepresented among the older cases as compared to middle-aged controls, hence associated with the decreased probability of carriers surviving into old age, corresponding to an odds ratio (OR) below unity (OR = 0.71 (95% CI 0.65 – 0.77)). This effect is observed in both sexes (Tables S3.2B and S3.2C). The remaining 57 SNPs did not show genome-wide significant effects on longevity either in men or women (Tables S3.2B, for men, and S3.2C, for women). The association of rs2075650 with survival did show some heterogeneity across the four studies ($P = 0.0495$), which is mainly because of the RS replication study.

Rs2075650 and the ApoE ε2/ε3/ε4 polymorphism

Rs2075650 is located in the *TOMM40* gene, next to the *APOE* gene (Figure S3.2). *APOE*

was previously associated with longevity [12,13]. The apolipoprotein E (ApoE) protein has three isoforms (ApoE ε2, ApoE ε3, and ApoE ε4), which are defined by two SNPs, rs7412 (Arg136Cys; ε2) and rs429358 (Cys112Arg; ε4). A meta-analysis of rs7412 and rs429358 in the LLS GWAS, the Leiden 85-plus replication study, and the Danish replication study samples (3,189 cases and 5,757 controls), showed a significant association of rs429358 with longevity (OR = 0.62 (95% CI 0.56 – 0.68), $P = 1.33 \times 10^{-23}$), which was comparable to rs2075650 (OR = 0.67 (95% CI 0.61 – 0.74), $P = 9.15 \times 10^{-17}$). Rs7412 also showed an association with longevity, with a higher prevalence of the minor allele in nonagenarians (OR = 1.31 (95% CI 1.17 – 1.46), $P = 1.35 \times 10^{-6}$).

We observed only moderate linkage disequilibrium (LD) between rs2075650 and rs429358 ($r^2 = 0.553$) and low LD between rs2075650 and rs7412 ($r^2 = 0.014$) when analyzing all samples with genotype data of rs2075650, rs429358, and rs7412 ($n = 8,946$). Nevertheless, in a conditional analysis with rs429358 and rs7412 (Model 1, described in the Materials and methods section), rs2075650 was no longer associated with longevity (OR = 0.93 (95% CI 0.81 – 1.07), $P = 0.337$). The OR increased from 0.67 to 0.93, i.e., the deleterious effect of rs2075650 on longevity diminishes and is statistically non-significant. However, the deleterious effect of rs429358 (OR = 0.64 (95% CI 0.56 – 0.74), $P = 2.68 \times 10^{-9}$) and the protective effect of rs7412 (OR = 1.20 (95% CI 1.07 – 1.36), $P = 0.002$) on longevity remained significant.

To determine whether there was an ApoE-independent effect of rs2075650 on survival after 90 years, prospective analysis of rs2075650, adjusted for rs429358 and rs7412, was performed. This analysis showed that carriers of the minor allele of rs2075650 displayed no increased mortality, i.e., a significant hazard ratio (HR) above 1, after 90 years of age independently of ApoE in two of the three cohorts analyzed (LLS, HR = 0.99 (95% CI 0.78 – 1.25), $P = 0.914$; Leiden 85-plus study, HR = 1.06 (95% CI 0.89 – 1.27), $P = 0.521$; Danish 1905 cohort, HR = 1.21 (95% CI 1.01 – 1.44), $P = 0.036$, Table S3.3A and Figure S3.3).

Overall, our results suggest that the association of rs2075650 with longevity is most likely a reflection of the effects of rs429358, caused by the moderate LD between the loci.

Association of rs429358 (ε4) and rs2075650 with serum parameters

As previous studies showed that rs429358 was associated with several metabolic phenotypes [30-32], association of this SNP with relevant serum parameters was determined in the offspring of the elderly LLS cases and their partners ($n = 2,324$, Model 2 described in the Materials and methods section). We replicated the previously reported associations of rs429358 with plasma levels of ApoE ($P = 7.42 \times 10^{-28}$), total cholesterol ($P = 0.001$), low-density lipoprotein (LDL) cholesterol ($P = 4.91 \times 10^{-5}$), high-density lipoprotein (HDL) cholesterol ($P = 0.062$), and C-reactive protein (CRP) ($P = 0.028$) and with HDL ($P = 0.061$) and LDL particle size ($P = 0.062$) (Table 3.2). In addition, we detected a minor effect on Insulin-like growth factor 1 (IGF-1) ($P = 0.025$) and insulin-like growth

Table 3.2 Association analysis of serum parameters between carriers and non-carriers of rs429358.

Serum parameter	*n*	β	95% CI	*P*	*n*	β	95% CI	*P*	*n*	β	95% CI	*P*
ApoE (mg/dL)*	2,222	0.83	0.80 - 0.86	7.42×10^{-28}	1,015	0.85	0.80 - 0.89	2.74×10^{-11}	1,207	0.81	0.78 - 0.85	1.13×10^{-22}
Total cholesterol (mmol/L)	2,229	0.18	0.07 - 0.29	0.001	1,019	0.18	0.04 - 0.32	0.011	1,210	0.18	0.02 - 0.33	0.024
HDL cholesterol (mmol/L)	2,228	-0.04	-0.07 - 0.00	0.062	1,018	-0.04	-0.09 - 0.00	0.064	1,210	-0.03	-0.08 - 0.02	0.286
LDL cholesterol (mmol/L)	2,168	0.20	0.10 - 0.29	4.91×10^{-5}	978	0.19	0.07 - 0.31	0.002	1,190	0.20	0.07 - 0.33	0.003
HDL size (nm)	2,219	-0.04	-0.08 - 0.00	0.061	1,011	-0.04	-0.10 - 0.02	0.159	1,208	-0.04	-0.09 - 0.02	0.165
LDL size (nm)	2,219	-0.06	-0.13 - 0.00	0.062	1,011	-0.08	-0.19 - 0.02	0.117	1,208	-0.05	-0.14 - 0.03	0.246
CRP (mg/L)*	2,216	0.90	0.81 - 0.99	0.028	1,014	0.84	0.73 - 0.94	0.005	1,202	0.94	0.83 - 1.08	0.399
IGF-1 (nmol/L)	2,223	-0.49	-0.92 - -0.06	0.025	1,015	-0.10	-0.74 - 0.53	0.748	1,208	-0.80	-1.36 - -0.24	0.005
IGFBP3 (mg/L)	2,223	-0.09	-0.17 - 0.00	0.042	1,015	-0.06	-0.18 - 0.06	0.281	1,208	-0.10	-0.21 - 0.01	0.062
IGF-1/IGFBP3	2,223	-0.03	-0.11 - 0.04	0.384	1,015	0.04	-0.08 - 0.15	0.504	1,208	-0.09	-0.19 - 0.01	0.065

95% CI, 95% confidence interval; *ApoE*, apolipoprotein E; *HDL*, high-density lipoprotein; *LDL*, low-density lipoprotein; *CRP*, C-reactive protein; *IGF-1*, insulin-like growth factor 1; *IGFBP3*, insulin-like growth factor binding protein 3.*Natural log transformed serum parameter was used in the association analysis.

Table 3.3 Association analysis of serum parameters previously associated with familial longevity in middle age in the Leiden Longevity Study families between carriers and non-carriers of rs429358.

Serum parameter	*n*	β	95% CI	*P*	*n*	β	95% CI	*P*	*n*	β	95% CI	*P*
Glucose (mmol/L)	2,234	-0.05	-0.17 - 0.07	0.388	1,021	-0.16	-0.36 - 0.04	0.116	1,213	0.03	-0.11 - 0.18	0.660
Insulin (mU/L)*	2,163	0.95	0.88 - 1.02	0.123	990	0.93	0.84 - 1.03	0.158	1,173	0.96	0.87 - 1.06	0.400
HDL cholesterol (mmol/L)	2,228	-0.04	-0.07 - 0.00	0.062	1,018	-0.04	-0.09 - 0.00	0.064	1,210	-0.03	-0.08 - 0.02	0.286
Triglycerides (mmol/L)*	2,229	1.03	0.98 - 1.08	0.203	1,016	1.07	0.99 - 1.15	0.095	1,208	1.01	0.95 - 1.07	0.834
HDL Size (nm)	2,219	-0.04	-0.08 - 0.00	0.061	1,011	-0.04	-0.10 - 0.02	0.159	1,208	-0.04	-0.09 - 0.02	0.165
LDL Size (nm)	2,219	-0.06	-0.13 - 0.00	0.062	1,011	-0.08	-0.19 - 0.02	0.117	1,208	-0.05	-0.14 - 0.03	0.246
fT3 (pmol/L)	2,223	0.05	-0.02 - 0.12	0.141	1,015	0.07	-0.02 - 0.16	0.127	1,208	0.03	-0.06 - 0.13	0.470

95% CI, 95% confidence interval; *HDL*, high-density lipoprotein; *LDL*, low-density lipoprotein; *fT3*, free triiodothyronine.*Natural log transformed serum parameter was used in the association analysis.

factor binding protein 3 (IGFBP3) levels (P = 0.042) (Table 3.2). The effect on IGF-1 seems to be female-specific (P = 0.005 and P = 0.748, in women and men, respectively) and is still significant after correction for multiple testing. We observed no ApoE-independent effect of rs2075650 on these traits, except for an increase of 0.18 mmol/L total cholesterol (P = 0.017) and 0.14 mmol/L LDL cholesterol (P = 0.014) with each minor allele of rs2075650 (using Model 3 described in the Materials and methods section).

No significant effects of rs429358 were observed on glucose (P = 0.388), insulin (P = 0.123), triglyceride (P = 0.203), and free triiodothyronine (fT3) (P = 0.141) levels (Table 3.3); the phenotypes that have previously been associated, in middle age, with familial longevity in the LLS families [24-28].

Analysis of Alzheimer's disease SNPs

Rs2075650 has consistently been associated with an increased risk of Alzheimer's disease in several independent GWAS [33-35]. Therefore, we studied the effect of SNPs present in the AlzGene database (http://www.alzgene.org/) [36], on survival into old age in the LLS GWAS. Apart from rs2075650, none of the 751 measured Alzheimer's disease SNPs showed a significant association after adjustment for multiple testing (Table S3.4).

Analysis of *FOXO3A* and *AKT1* SNPs

Apart from *APOE*, two other genes have shown consistent evidence for association with longevity, *FOXO3A* [14-17] and *AKT1* [15]. For the longevity-promoting *FOXO3A* SNPs previously reported with centenarian longevity, we observed no association with survival into old age in our nonagenarians (Table S3.5). For *AKT1*, one of the two measured SNPs, rs2498804, showed a significant association with survival into old age (OR = 0.75 (95% CI 0.63 – 0.89), P = 0.001) (Table S3.5).

Discussion

To identify common SNPs contributing to longevity, GWAS analysis of 403 nonagenarian cases and 1,670 population controls was performed. Of the 62 top associating SNPs, 58 were tested in a meta-analysis of 4,149 nonagenarian cases and 7,582 younger controls and we identified one SNP, rs2075650, that associated significantly with survival into old age (P = 3.39 x 10^{-17}). Carriers of the minor allele had a 29% decreased probability of reaching 90 years on average. Although cases and controls originate from different generations, we concluded that there was no substructure to an extent that would affect the observations.

Rs2075650 is located in the *TOMM40* gene at chromosome 19q13.32 close to and centromeric of the *APOE* gene (Figure S3.2), which has shown consistent evidence for association with longevity [12,13]. The ApoE protein has three isoforms (ApoE ε2, ApoE ε3, and ApoE ε4) that are defined by two SNPs, rs7412 (Arg136Cys; ε2) and rs429358 (Cys112Arg; ε4). ApoE ε4 carriers have an increased risk of cardiovascular disease and Alzheimer's disease, while ApoE ε2 carriers are protected from these diseases [12,37,38]. Although we detected only moderate LD (r^2 = 0.553) between rs2075650 and the ApoE ε4-defining SNP rs429358, we could

not detect a significant effect of rs2075650 on longevity independent of rs429358. Several prospective studies, including one with the Danish 1905 cohort [39], reported increased mortality for ApoE ε4 carriers, even though there is still much debate about *APOE* being a "frailty gene" or a "longevity gene" [12,39-41]. The prospective data in the LLS and Leiden 85-plus study support the "frailty gene" hypothesis, as rs429358 affects mortality after 85 years and continues the effect after 90 years (HR = 1.08 (95% CI 1.03 – 1.13), $P = 0.001$ and HR = 1.08 (95% CI 1.03 – 1.12), $P = 0.001$, respectively, Table S3.3B and Figure S3.4). In these prospective studies, carriers of the minor allele of rs2075650 showed no increased mortality independent of rs429358, which indicates that the association of rs2075650 with longevity is most likely due to variation in the *APOE* gene. Although GWAS have reported significant associations between rs2075650 and Alzheimer's disease, brain imaging, total cholesterol, and CRP plasma levels [35,42-44], no analyses were performed to determine whether these associations are ApoE-independent. We observed no ApoE-independent effect on the phenotypes investigated in the LLS offspring and partners except for total and LDL cholesterol.

Previously, rs429358 had been associated with several metabolic phenotypes, such as ApoE, total cholesterol, HDL cholesterol, LDL cholesterol, and CRP levels, as well as HDL and LDL particle size [30-32] and, here, we have confirmed these findings using serum measurements of the offspring and partners from the LLS. Because the insulin/IGF-1 signaling (IIS) pathway has a lifespan regulating effect in several model organisms [9,10] and humans [45], we also investigated the effect of rs429358 on serum levels of IGF-1 and IGFBP3, which both play a role in this pathway. Both proteins are involved in the etiology of several age-related diseases. However, up till now, it is not clear whether higher or lower serum levels are beneficial for longevity. Low IGF-1 serum levels associate with a decreased risk of cancer, but an increased risk of cardiovascular disease and neurodegenerative disease [46]. Previously, we showed in the Leiden 85-plus study cohort that genetic variants known to associate with lower IIS activity and IGF-1 serum levels at younger age associated with better survival at ages above 85 years [47]. However, the effect of these genetic variants on IGF-1 serum levels was not tested in the Leiden 85-plus study cohort. In addition, we showed previously that neither IGF-1 and IGFBP3 levels nor their ratio differed between partners and offspring from the LLS [24], which indicates that IGF-1 serum levels are, in middle age, not a marker for longevity, whereas a decreased risk of metabolic diseases was evident at that age in long-lived families [23]. In the current study, we found that the minor allele of rs429358 associates with lower IGF-1 levels in middle-aged women, which, to our knowledge, has not previously been reported. Like low IGF-1 levels, ApoE ε4 was previously associated with an increased risk of developing cardiovascular disease and neurodegenerative disease [12,37,38]. Thus, the mechanism behind the increased risk of female ApoE ε4 carriers of developing cardiovascular and/or neurodegenerative

diseases might involve serum levels of IGF-1 or other aspects of IIS activity reflected by these levels. Apart from lipid metabolism, the parameters determining the longevity phenotype in middle age in the LLS, such as glucose metabolism, insulin sensitivity, and thyroid hormone metabolism [24-28], were not influenced by the presence of the minor allele of rs429358. This indicates that it is likely that other loci could explain the differences in these phenotypes between LLS offspring and partners.

The strength of this study is that, by using a GWAS, we were able to replicate the previously reported association of the *APOE* locus with longevity [12,13] as the major locus. This was not observed in the previously published meta-GWAS of Newman *et al*. [21], possibly because of differences in the study design and population control selection between the studies. While Newman *et al*. used nonagenarian cases in a population-based design, we made use of a family-based design in which the families are genetically enriched for longevity. In addition, Newman *et al*. used population controls from the same cohort which had died before the age of 80. Between 60 and 80 years, however, there might already have been a selection on survival, decreasing the frequency of ApoE ε4 carriers in the control group. In contrast, we made comparisons to a younger population group (55-60 years) from a different cohort (RS).

As we previously reported that long-lived individuals carry the same number of disease risk alleles for cardiovascular disease, cancer, and type 2 diabetes as young controls [29], we expected to primarily find longevity-promoting alleles. However, although most of the 58 prioritized SNPs (n = 43) from the LLS GWAS showed a longevity-promoting effect ranging from 36 to 168%, none of them could be replicated in additional study populations of nonagenarian singletons. The only replicated locus is *APOE*, which is a mortality locus that has previously been reported to be the major locus responsible for Alzheimer's disease [33-35], a well-known age-related disease. Nevertheless, none of the other Alzheimer's disease loci showed an association with survival to 90 years, which indicates that the remaining genetic variation in longevity in the LLS could not be explained by the genetic variation which contributes to Alzheimer's disease. In addition to *APOE*, we also observed evidence for association at the previously reported *AKT1* locus [15] with survival into old age in the LLS GWAS, although the effect of this SNP is relatively small (25% decreased probability of becoming 90 years) compared to the effect of rs429358 (51%). The previously reported longevity-promoting effect of the *FOXO3A* locus could not be replicated in this study. This is probably due to the relatively low number of centenarians in the LLS GWAS case group, in which the effect of SNPs in *FOXO3A* on longevity seems to be most prominent. The still unexplained genetic variation in longevity might be attributable to rare variants or variants with small effects, which has previously been reported for other complex traits, such as Alzheimer's disease. These loci could not be identified in this study because of the relatively small number of cases in the LLS GWAS, the heterogeneity of factors influencing lifespan within populations, and the difference in the design of the studies used for replication.

One way to identify variants with small effects would be to increase the initial sample size of the GWAS and perform replication in other studies of nonagenarians. Given the higher heritability of longevity at older ages [48], one may also limit the study population to centenarians or supercentenarians. In addition to common variants with small effects, rare variants with large effects might play a role in longevity. By whole-genome/exome sequencing of long-lived subjects and their families, rare variants can be identified and associated with human longevity.

In conclusion, we have shown that the deleterious effect of the ApoE ε4 allele, tagged by rs2075650, is the single major hit in our GWAS for longevity, indicating that no other major longevity locus was present among these nonagenarians. We confirmed the previously reported associations of the ApoE ε4 allele with lipid metabolism parameters and report an additional effect on IGF-1 signaling in women. To identify genetic variants with smaller and protective effects on human lifespan, a meta-GWAS for longevity with a larger sample size is merited.

Materials and methods

Study Populations

Leiden Longevity Study

For the LLS, long-lived siblings of European 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 men and 91 years or older for women, representing < 0.5% of the Dutch population in 2001 [6]. 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). DNA from the LLS was extracted from samples at baseline using conventional methods [20]. For the GWAS, 403 unrelated LLS siblings (one sibling from each sibling pair) were included (LLS GWAS cases).

Rotterdam Study

The RS is a prospective population-based study of people aged 55 years and older, which was designed to study neurological, cardiovascular, locomotor, and ophthalmological diseases [49]. The study consists of 7,983 participants from the baseline cohort (RS-I) and 3,011 participants from an independent extended cohort formed in 1999 (RS-II) from which DNA was isolated between 1990 and 1993 (RS-I) or between 2000 and 2001 (RS-II). For the GWAS, 1,731 participants from the combined cohort who were below 60 years of age and for whom GWAS data were available were included as controls (RS GWAS controls). For the replication study, 960 cases above 90 years at time of recruitment (RS replication cases) and 1,825 controls between 60 and 65 years at baseline (RS replication controls) from the combined cohorts, for whom GWAS data were also available, were included.

Leiden 85-plus study

In the Leiden 85-plus study, two prospective population-based cohorts were recruited from inhabitants of Leiden [50,51]. Between 1987 and 1989, 673 subjects aged 85 years and older were enrolled in a prospective

study (Cohort 1). Between 1997 and 1999, 563 subjects were enrolled in the month of their 85th birthday with follow-up (Cohort 2). Subjects were visited at their home and there were no exclusion criteria related to health. DNA was available from the combined cohorts consisting of 1,208 subjects aged 85 years and older (Leiden 85-plus replication cases).

Netherlands Twin Registry

From the Netherlands Twin Registry (NTR), 2,090 unrelated participants of European descent for whom DNA was available were selected as control samples [52] (Leiden 85-plus replication controls). The substructure in the NTR has been reported before [53], and in this study, we included samples aged between 15 and 70 years at the time of blood sampling, without known family relations (i.e., those without any substructure).

Danish 1905 cohort

The participants in this study are from the Danish 1905 birth cohort recruited in 1998 [54] when they were aged 92-93 years. From this cohort, 3,600 subjects were still alive, of whom 2,262 participated in the study. Participants were subjected to a home-based interview on health and lifestyle parameters, physical and cognitive function tests, and the collection of biological material. The current genetic study comprises a total of 1,578 of these individuals (Danish replication cases). Survival was followed up until January 2010. Ninety-nine percent (1,561 subjects) of subjects died in the 12 years of follow-up. Control samples were 1997 twins (one twin for each pair) between 46 and 68 years of age collected from all over Denmark (Danish replication controls).

The cases in all three replication cohorts originate from population-based cohort studies from a genetic background similar to the LLS [55]. All the participants in these studies have signed an informed consent.

Genotyping

Genome-wide association study

LLS GWAS cases were genotyped using Illumina Infinium HD Human660W-Quad BeadChips (Illumina, San Diego, CA, USA). The RS-I and RS-II cohorts were genotyped using Illumina Infinium II HumanHap 550K Beadchips and Illumina Infinium II HumanHap550-Duo BeadChips (Illumina), respectively [49].

For the GWAS, we selected 551,606 SNPs for analysis because these were genotyped in both the LLS GWAS cases and (some of) the RS GWAS controls. Of these 551,606 SNPs, 34,885 SNPs were excluded on the basis of the following criteria: SNP call rate < 0.95 or minor allele frequency < 0.01 in RS GWAS controls or LLS GWAS cases (n = 8,908 and n = 24,586, respectively), and $P_{HWE} < 10^{-4}$ in RS GWAS controls (n = 1,355). In addition, SNPs with a between-chip effect in the RS GWAS controls were removed using a genotype trend test comparing the RS GWAS controls from RS-I with RS-II (n = 36), leaving 516,721 SNPs for statistical analysis. The Illumina clusterplots of the SNPs with $P < 1 \times 10^{-4}$ (n = 71) were visually inspected to confirm high-quality genotyping and 9 SNPs were excluded on the basis of bad clustering in the LLS GWAS cases or RS GWAS controls.

Genotype data were used to confirm gender and family relationships. Two RS GWAS control samples were excluded because of abnormalities in the sex chromosomes (both samples had Triple X Syndrome). Latent clustering of genotypes because of population substructure was assessed by pairwise identity-by-state (IBS) distance using Graphical Relationship Representation (http://bioinformatics.well.ox.ac.uk/GRR) [56]. LLS GWAS cases showed no relationship errors. From the RS GWAS controls, 59 samples were excluded because of high IBS. In total, 403 LLS GWAS cases and 1,670 RS GWAS control samples with a sample call rate > 0.95 were analyzed. Because cases and controls originate from different generations, we investigated whether substructure in these cohorts could have influenced the observed associations. IBS estimates for all pairs of subjects in the data set were computed on a randomly selected set of 10% of the SNPs that passed quality control thresholds, using the --genome, --cluster, and --mds-plot 4 commands in PLINK (http://pngu.mgh.harvard.edu/purcell/plink) [57]. The first two resulting principal components (C1 and C2) were plotted against each other, which gives a representation of the data in two dimensions. In the resulting scatter plot, each point represents an individual (green = LLS GWAS case and blue = RS GWAS control, Figure S3.5). If there had been substructure, one would see multiple clusters in one plot. However, because all our samples seem to be in one cluster, we concluded that there was no substructure to an extent that would affect the observations.

Replication studies

For the RS replication study, we used the existing GWAS data in the Rotterdam Study after the quality control screening described by Teichert *et al*. [49]. For the Leiden 85-plus and Danish replication studies, genotyping was performed using the Sequenom MassARRAY iPLEX Gold and TaqMan SNP Genotyping assays. Of the 62 prioritized SNPs, 58 could be designed for replication studies using Sequenom, of which 56 were successfully genotyped in > 95% of the samples displayed in Table 3.1. The average genotype call rate for SNPs genotyped with Sequenom was 98.40%, and the average concordance rate with GWAS data among the LLS GWAS cases was 99.97%. For 2 of the 6 SNPs that could not be genotyped with Sequenom, rs2075650 and rs642990, pre-designed TaqMan SNP genotyping assays (C____3084828_20 and C____2206314_20, respectively) were used for genotyping, following the manufacturer's instructions. The average genotype call rate for the SNPs genotyped with TaqMan was 99.04%, and the average concordance rate with GWAS data among the LLS GWAS cases was 100%.

ApoE ε2/ε3/ε4 polymorphism

The ApoE ε2/ε3/ε4 defining SNPs, rs429358 (Cys112Arg; ε4) and rs7412 (Arg136Cys; ε2), were genotyped in the LLS GWAS cases, Leiden 85-plus replication study, and Danish replication study controls using pre-designed TaqMan SNP genotyping assays (C___3084793_20 and C____904973_10, respectively). For the RS GWAS controls and Danish replication study cases, previously measured data were used [39,58].

Measurement of serum parameters

All standard serum measurements were performed using fully automated equipment.

Glucose, total cholesterol, HDL cholesterol, and triglyceride levels were measured using the Hitachi Modular P 800 (Roche, Almere, the Netherlands) [24]. LDL cholesterol was calculated using the Friedewald formula [59].

LDL and HDL particle sizes were measured using proton NMR spectroscopy (LipoScience Inc, Raleigh, NY, USA) [60].

IGF-1, IGFBP3, and insulin levels were measured using the Immulite 2500 (DPC, Los Angeles, CA, USA) [24].

fT3 was measured using the Modular E170 and high-sensitivity CRP was measured using Cobas Integra 800 (both from Roche) [26].

The level of ApoE was determined in serum samples using a human ApoE-specific sandwich ELISA [61,62].

Statistical analysis

GWAS and replication studies

For the association analysis of the GWAS data, we applied a Cochran-Armitage trend test [63,64]. For X-linked SNPs, the genotypes of the men were considered as homozygous genotypes. SNPs with a *P*-value < 1 x 10^{-4} (n = 62) were selected for replication. Odds ratios were estimated and the corresponding 95% confidence intervals were computed. For meta-analyses, a fixed effect approach was used. Scores and their variances were computed within each study and combined across the four studies to obtain a single meta-statistic. *P*-values below 5 x 10^{-8} were considered as genome-wide significant [65]. The between-study variance was calculated to determine heterogeneity across the four studies. All these analysis were performed using Bioconductor R (http://www.bioconductor.org) [66].

The quantile–quantile plot (Figure S3.6), constructed using Bioconductor R (http://www.bioconductor.org) [66], showed that the *P*-value distribution of stage 1 conformed to a null distribution at all but the extreme tail. The genomic inflation factor (λ), which measures over-dispersion of test statistics from association tests indicating population stratification, was 1.027 and we therefore decided not to adjust for population stratification.

Linkage Disequilibrium between rs2075650 and the ApoE ε2/ε3/ε4 polymorphism

Pairwise linkage disequilibrium (LD) between rs2075650 and the ApoE ε2/ε3/ε4 polymorphism determining SNPs rs7412 and rs429358 was calculated in 8,946 individuals using the --ld command in PLINK (http://pngu.mgh.harvard.edu/purcell/plink) [57].

ApoE-independent association of rs2075650 with longevity

To determine whether the association of rs2075650 with longevity was independent of the ApoE ε2/ε3/ε4 polymorphism, a logistic regression model with adjustment for rs429358, rs7412 and an interaction term for ε2/ε3 with ε3/ε4 was tested [67]:

$$Logit\ (P_{status} = 1) = \beta 0 + \beta 1 * rs2075650 + \beta 2 * rs429358 + \beta 3 * rs7412 + \beta 4 * (rs429358 * rs7412) + \beta 5 * study \quad (Model\ 1)$$

status was coded as 0 (control) or 1 (long-lived case), *study* was coded as 0 (LLS

GWAS), 1 (Leiden 85-plus replication study) or 2 (Danish replication study) and the genotypes of *rs2075650*, *rs429358* and *rs7412* were coded as 0 (homozygous for the common allele), 1 (heterozygous) or 2 (homozygous for the rare allele). STATA/SE 11.1 (StataCorp LP, TX, USA) was used for this analysis.

Prospective analysis

Prospective analysis of rs2075650 and rs429358 was performed with 944 nonagenarian siblings from the LLS, 976 octogenarians and nonagenarians from the Leiden 85-plus study, and 1,578 nonagenarians from the Danish 1905 cohort.

After a mean follow-up time of 5.7 years (LLS), 14.8 years (Leiden 85-plus study), and 11.4 years (Danish 1905 cohort), 73.2% (n = 691) (LLS), 84.8% (n = 828) (Leiden 85-plus study), and 98.9% (n = 1,561) (Danish 1905 cohort) of the individuals had died.

Mortality analyses were performed with STATA/SE 11.1 (StataCorp LP) using a gender-adjusted, left-truncated Cox proportional hazards model to adjust for late entry into the data set according to age.

Association of rs429358 (ε4) and rs2075650 with serum parameters

To determine the association of rs429358 and the ApoE-independent association of rs2075650 with serum parameters in the offspring and their partners from the LLS the following regression models were tested:

$$Serum\ parameter = \beta0 + \beta1*age + \beta2*gender + \beta3*(age*gender) + \beta4*group + \beta5*rs429358\ (Model\ 2)$$

$$Serum\ parameter = \beta0 + \beta1*age + \beta2*gender + \beta3*(age*gender) + \beta4*group + \beta5*rs2075650 + \beta6*rs429358 + \beta7*rs7412 + \beta8*(rs429358*rs7412)\ (Model\ 3)$$

age was coded in years, *gender* was coded as 1 (male) or 2 (female), *group* was coded as 0 (partner) or 1 (offspring) and the genotypes of *rs2075650*, *rs429358* and *rs7412* were coded as 0 (homozygous for the common allele), 1 (heterozygous) or 2 (homozygous for the rare allele). Robust standard errors were used to account for sibship relations. STATA/SE 11.1 (StataCorp LP) was used for these analyses.

Acknowledgments

We thank all participants of the Leiden Longevity Study, Rotterdam Study, Leiden 85-plus study, Netherlands Twin Register, and the Danish 1905 cohort. This study was supported by a grant from the Innovation-Oriented Research Program on Genomics (SenterNovem IGE05007), the Centre for Medical Systems Biology, and the Netherlands Consortium for Healthy Ageing (grant 050-060-810), all in the framework of the Netherlands Genomics Initiative/Netherlands Organization for Scientific Research (NWO) and by Unilever Colworth. The generation and management of GWAS genotype data for the Rotterdam study is supported by the Netherlands Organisation for Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012). This study is funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2) and the Netherlands Genomics

Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) project nr. 050-060-810. The Rotterdam Study is funded by the Erasmus Medical Center and Erasmus University, Rotterdam, the 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. The data collection of the Danish 1905 cohort is supported by the US National Institute on Aging, research grant NIA-P01-AG08761, and the Danish Aging Research Center is supported by a grant from the VELUX foundation. The data collection of the Netherlands Twin Register is supported by NWO (904-61-090; 904-61-193; 480-04-004; 400-05-717; SPI 56-464-14192), Center for Medical Systems Biology (NWO Genomics; Centre for Neurogenomics and Cognitive Research (CNCR-VU); the EU (EU/QLRT-2001-01254); Genome scan for neuroticism (NIMH R01 MH059160); the Geestkracht program of ZonMW (10-000-1002), and institutes involved in NESDA (VU University Medical Center, Leiden University Medical Center, GGZ Buitenamstel-Geestgronden, Rivierduinen, University Medical Center Groningen, GGZ Lentis, GGZ Friesland, GGZ Drenthe). The genotyping of the samples was provided through the Genetic Association Information Network (GAIN).

References

1. Oeppen J, Vaupel JW. Demography. Broken limits to life expectancy. *Science* 2002; **296**: 1029-31.
2. Gudmundsson H, Gudbjartsson DF, Frigge M, Gulcher JR, Stefansson K. Inheritance of human longevity in Iceland. *Eur J Hum Genet* 2000; **8**: 743-9.
3. Hjelmborg JV, Iachine I, Skytthe A, Vaupel JW, McGue M, Koskenvuo M, *et al*. Genetic influence on human lifespan and longevity. *Hum Genet* 2006; **119**: 312-21.
4. Skytthe A, Pedersen NL, Kaprio J, Stazi MA, Hjelmborg JV, Iachine I, *et al*. Longevity studies in GenomEUtwin. *Twin Res* 2003; **6**: 448-54.
5. Perls T, Shea-Drinkwater M, Bowen-Flynn J, Ridge SB, Kang S, Joyce E, *et al*. Exceptional familial clustering for extreme longevity in humans. *J Am Geriatr Soc* 2000; **48**: 1483-5.
6. Schoenmaker M, de Craen AJ, de Meijer PH, Beekman M, Blauw GJ, Slagboom PE, *et al*. Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur J Hum Genet* 2006; **14**: 79-84.
7. Christensen K, McGue M, Petersen I, Jeune B, Vaupel JW. Exceptional longevity does not result in excessive levels of disability. *Proc Natl Acad Sci U S A* 2008; **105**: 13274-9.
8. Terry DF, Sebastiani P, Andersen SL, Perls TT. Disentangling the roles of disability and morbidity in survival to exceptional old age. *Arch Intern Med* 2008; **168**: 277-83.
9. Fontana L, Partridge L, Longo VD. Extending healthy life span--from yeast to humans. *Science* 2010; **328**: 321-6.
10. Kenyon CJ. The genetics of ageing. *Nature* 2010; **464**: 504-12.
11. Finch CE, Tanzi RE. Genetics of aging. *Science* 1997; **278**: 407-11.
12. Christensen K, Johnson TE, Vaupel JW. The quest for genetic determinants of human longevity: challenges and insights. *Nat Rev Genet* 2006; **7**: 436-48.
13. Schachter F, Faure-Delanef L, Guenot F, Rouger H, Froguel P, Lesueur-Ginot L, *et al*. Genetic associations with human longevity at the APOE and ACE loci. *Nat Genet* 1994; **6**: 29-32.
14. Flachsbart F, Caliebe A, Kleindorp R, Blanche H, von Eller-Eberstein H, Nikolaus S, *et al*. Association of FOXO3A variation with human longevity confirmed in German centenarians. *Proc Natl Acad Sci U S A* 2009; **106**: 2700-5.
15. Pawlikowska L, Hu D, Huntsman S, Sung A, Chu C, Chen J, *et al*. Association of common genetic variation in the insulin/IGF1 signaling pathway with human longevity. *Aging Cell* 2009; **8**: 460-72.
16. Soerensen M, Dato S, Christensen K, McGue M, Stevnsner T, Bohr VA, *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**: 1010-7.
17. Willcox BJ, Donlon TA, He Q, Chen R, Grove JS, Yano K, *et al*. FOXO3A genotype is strongly associated with human longevity. *Proc Natl Acad Sci U S A* 2008; **105**: 13987-92.
18. Puca AA, Daly MJ, Brewster SJ, Matise TC, Barrett J, Shea-Drinkwater M, *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**: 10505-8.
19. Boyden SE, Kunkel LM. High-density genomewide linkage analysis of exceptional human longevity identifies multiple novel Loci. *PLoS One* 2010; **5**: e12432.
20. Beekman M, Blauw GJ, Houwing-Duistermaat JJ, Brandt BW, Westendorp RG, Slagboom PE. 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**: 355-62.
21. Newman AB, Walter S, Lunetta KL, Garcia ME, Slagboom PE, Christensen K, *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**: 478-87.
22. Atzmon G, Schechter C, Greiner W, Davidson D, Rennert G, Barzilai N. Clinical

phenotype of families with longevity. *J Am Geriatr Soc* 2004; **52**: 274-7.
23. Westendorp RG, van Heemst D, Rozing MP, Frolich M, Mooijaart SP, Blauw GJ, *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**: 1634-7.
24. Rozing MP, Westendorp RG, Frolich M, de Craen AJ, Beekman M, Heijmans BT, *et al*. Human insulin/IGF-1 and familial longevity at middle age. *Aging (Albany NY)* 2009; **1**: 714-22.
25. Rozing MP, Westendorp RG, de Craen AJ, Frolich M, de Goeij MC, Heijmans BT, *et al*. Favorable glucose tolerance and lower prevalence of metabolic syndrome in offspring without diabetes mellitus of nonagenarian siblings: the Leiden longevity study. *J Am Geriatr Soc* 2010; **58**: 564-9.
26. Rozing MP, Westendorp RG, de Craen AJ, Frolich M, Heijmans BT, Beekman M, *et al*. Low serum free triiodothyronine levels mark familial longevity: the Leiden Longevity Study. *J Gerontol A Biol Sci Med Sci* 2010; **65**: 365-8.
27. Vaarhorst AA, Beekman M, Suchiman HE, van Heemst D, Houwing-Duistermaat JJ, Westendorp RG, *et al*. Lipid metabolism in long-lived families: the Leiden Longevity Study. *Age (Dordr)* 2011; **33**: 219-27.
28. Wijsman CA, Rozing MP, Streefland TC, le Cessie S, Mooijaart SP, Slagboom PE, *et al*. Familial longevity is marked by enhanced insulin sensitivity. *Aging Cell* 2011; **10**: 114-21.
29. Beekman M, Nederstigt C, Suchiman HE, Kremer D, van der Breggen R, Lakenberg N, *et al*. Genome-wide association study (GWAS)-identified disease risk alleles do not compromise human longevity. *Proc Natl Acad Sci U S A* 2010; **107**: 18046-9.
30. Boerwinkle E, Utermann G. Simultaneous effects of the apolipoprotein E polymorphism on apolipoprotein E, apolipoprotein B, and cholesterol metabolism. *Am J Hum Genet* 1988; **42**: 104-12.
31. Hubacek JA, Peasey A, Pikhart H, Stavek P, Kubinova R, Marmot M, *et al*. APOE polymorphism and its effect on plasma C-reactive protein levels in a large general population sample. *Hum Immunol* 2010; **71**: 304-8.
32. Topic A, Spasojevic K, V, Zeljkovic A, Vekic J, Jelic IZ. Gender-related effect of apo E polymorphism on lipoprotein particle sizes in the middle-aged subjects. *Clin Biochem* 2008; **41**: 361-7.
33. Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, *et al*. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* 2009; **41**: 1088-93.
34. Lambert JC, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, *et al*. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* 2009; **41**: 1094-9.
35. Seshadri S, Fitzpatrick AL, Ikram MA, DeStefano AL, Gudnason V, Boada M, *et al*. Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* 2010; **303**: 1832-40.
36. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat Genet* 2007; **39**: 17-23.
37. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, *et al*. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 1993; **261**: 921-3.
38. Eichner JE, Kuller LH, Orchard TJ, Grandits GA, McCallum LM, Ferrell RE, *et al*. Relation of apolipoprotein E phenotype to myocardial infarction and mortality from coronary artery disease. *Am J Cardiol* 1993; **71**: 160-5.
39. Jacobsen R, Martinussen T, Christiansen L, Jeune B, Andersen-Ranberg K, Vaupel JW, *et al*. Increased effect of the ApoE gene on survival at advanced age in healthy and long-lived Danes: two nationwide cohort studies. *Aging Cell* 2010; **9**: 1004-9.
40. Ewbank DC. Differences in the association between apolipoprotein E genotype and mortality across populations. *J Gerontol A Biol Sci Med Sci* 2007; **62**: 899-907.

41. Gerdes LU, Jeune B, Ranberg KA, Nybo H, Vaupel JW. 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**: 202-10.
42. Aulchenko YS, Ripatti S, Lindqvist I, Boomsma D, Heid IM, Pramstaller PP, *et al*. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet* 2009; **41**: 47-55.
43. Reiner AP, Barber MJ, Guan Y, Ridker PM, Lange LA, Chasman DI, *et al*. Polymorphisms of the HNF1A gene encoding hepatocyte nuclear factor-1 alpha are associated with C-reactive protein. *Am J Hum Genet* 2008; **82**: 1193-201.
44. Shen L, Kim S, Risacher SL, Nho K, Swaminathan S, West JD, *et al*. Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: A study of the ADNI cohort. *Neuroimage* 2010; **53**: 1051-63.
45. Suh Y, Atzmon G, Cho MO, Hwang D, Liu B, Leahy DJ, *et al*. Functionally significant insulin-like growth factor I receptor mutations in centenarians. *Proc Natl Acad Sci U S A* 2008; **105**: 3438-42.
46. Juul A. Serum levels of insulin-like growth factor I and its binding proteins in health and disease. *Growth Horm IGF Res* 2003; **13**: 113-70.
47. van Heemst D, Beekman M, Mooijaart SP, Heijmans BT, Brandt BW, Zwaan BJ, *et al*. Reduced insulin/IGF-1 signalling and human longevity. *Aging Cell* 2005; **4**: 79-85.
48. Tan Q, Zhao JH, Iachine I, Hjelmborg J, Vach W, Vaupel JW, *et al*. Power of non-parametric linkage analysis in mapping genes contributing to human longevity in long-lived sib-pairs. *Genet Epidemiol* 2004; **26**: 245-53.
49. Teichert M, Eijgelsheim M, Rivadeneira F, Uitterlinden AG, van Schaik RH, Hofman A, *et al*. A genome-wide association study of acenocoumarol maintenance dosage. *Hum Mol Genet* 2009; **18**: 3758-68.
50. Bootsma-van der Wiel A, van Exel E, de Craen AJ, Gussekloo J, Lagaay AM, Knook DL, *et al*. A high response is not essential to prevent selection bias: results from the Leiden 85-plus study. *J Clin Epidemiol* 2002; **55**: 1119-25.
51. Weverling-Rijnsburger AW, Blauw GJ, Lagaay AM, Knook DL, Meinders AE, Westendorp RG. Total cholesterol and risk of mortality in the oldest old. *Lancet* 1997; **350**: 1119-23.
52. Boomsma DI, Willemsen G, Sullivan PF, Heutink P, Meijer P, Sondervan D, *et al*. Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. *Eur J Hum Genet* 2008; **16**: 335-42.
53. Sullivan PF, de Geus EJ, Willemsen G, James MR, Smit JH, Zandbelt T, *et al*. Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Mol Psychiatry* 2009; **14**: 359-75.
54. Nybo H, Petersen HC, Gaist D, Jeune B, Andersen K, McGue M, *et al*. Predictors of mortality in 2,249 nonagenarians--the Danish 1905-Cohort Survey. *J Am Geriatr Soc* 2003; **51**: 1365-73.
55. Heath SC, Gut IG, Brennan P, McKay JD, Bencko V, Fabianova E, *et al*. Investigation of the fine structure of European populations with applications to disease association studies. *Eur J Hum Genet* 2008; **16**: 1413-29.
56. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. GRR: graphical representation of relationship errors. *Bioinformatics* 2001; **17**: 742-3.
57. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559-75.
58. Slooter AJ, Cruts M, Kalmijn S, Hofman A, Breteler MM, Van Broeckhoven C, *et al*. Risk estimates of dementia by apolipoprotein E genotypes from a population-based incidence study: the Rotterdam Study. *Arch Neurol* 1998; **55**: 964-8.

59. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem* 1972; **18**: 499-502.
60. Heijmans BT, Beekman M, Houwing-Duistermaat JJ, Cobain MR, Powell J, Blauw GJ, *et al*. Lipoprotein particle profiles mark familial and sporadic human longevity. *PLoS Med* 2006; **3**: e495.
61. Mooijaart SP, Berbee JF, van Heemst D, Havekes LM, de Craen AJ, Slagboom PE, *et al*. ApoE plasma levels and risk of cardiovascular mortality in old age. *PLoS Med* 2006; **3**: e176.
62. van Vlijmen BJ, van den Maagdenberg AM, Gijbels MJ, van der Boom H, HogenEsch H, Frants RR, *et al*. Diet-induced hyperlipoproteinemia and atherosclerosis in apolipoprotein E3-Leiden transgenic mice. *J Clin Invest* 1994; **93**: 1403-10.
63. Armitage P. Tests for linear trends in proportions and frequencies. *Biometrics* 1955; **11**: 375-86.
64. Cochran WG. Some methods for strengthening the common X^2 tests. *Biometrics* 1954; **10**: 417-51.
65. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* 2008; **32**: 381-5.
66. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, *et al*. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004; **5**: R80.
67. Ronald J, Rajagopalan R, Ranchalis JE, Marshall JK, Hatsukami TS, Heagerty PJ, *et al*. Analysis of recently identified dyslipidemia alleles reveals two loci that contribute to risk for carotid artery disease. *Lipids Health Dis* 2009; **8**: 52.

Supplementary Information

The supplementary information belonging to this chapter can be found at: http://onlinelibrary.wiley.com/doi/10.1111/j.1474-9726.2011.00705.x/suppinfo.

Figure S3.1 Manhattan plot presenting the -$\log_{10}$ *P*-value from the Cochran-Armitage trend test for the 516,721 SNPs that passed the quality control thresholds in the LLS GWAS.

Figure S3.2 Genomic region surrounding rs2075650 (obtained from the UCSC genome browser (http://genome.ucsc.edu/). The physical distances between rs2075650 and rs429358 and between rs2075650 and rs7412 are 16.32 kb and 16.46 kb, respectively.

Figure S3.3 Kaplan-Meier curves showing the survival rate over years to follow-up of ApoE ε3ε3 carriers with zero (solid line) or one (large dashed line) minor allele(s) of rs2075650 in the LLS, Leiden 85-plus study and Danish 1905 cohort.

Figure S3.4 Kaplan-Meier curves showing the survival rate over years to follow-up of carriers of zero (solid line), one (large dashed line), or two (small dashed line) ε4 allele(s) of rs429358 in the LLS and Leiden 85-plus study.

Figure S3.5 C1 values plotted against the C2 values, both resulting from the multidimensional scaling analysis, of the 403 LLS GWAS cases (green) and the 1,670 RS GWAS controls (blue).

Figure S3.6 Quantile–quantile plot of expected vs. observed chi-square values for the test statistic from the Cochran-Armitage trend test for 516,721 SNPs that passed the quality control thresholds in the LLS GWAS. The slope of the dashed line represents the genomic inflation factor ($\lambda = 1.027$). The shaded region represents the 95% confidence band.

Table S3.1 SNPs ($n = 62$) selected for replication analysis, associating at $P < 1 \times 10^{-4}$ with survival into old age in the analysis of the LLS GWAS.

Table S3.2A Results of the association analysis with survival into old age of the 58 prioritized SNPs from the LLS GWAS in the RS replication study, Leiden 85-plus replication study, Danish replication study, and the meta-analysis.

Table S3.2B Results of the meta-association analysis with survival into old age of the 58 prioritized SNPs in male cases compared to all controls.

Table S3.2C Results of the meta-association analysis with survival into old age of the 58 prioritized SNPs in female cases compared to all controls.

Table S3.3A Results of the prospective analysis of rs2075650 adjusted for rs429358 (ε4) and rs7412 (ε2).

Table S3.3B Results of the prospective analysis of rs429358 (ε4).

Table S3.4 Association of LLS GWAS SNPs selected from the AlzGene database (http://www.alzgene.org/) [36] with survival into old age.

Table S3.5 Association of LLS GWAS SNPs within a 10-kb window around *FOXO3A* and *AKT1* with survival into old age.

Chapter 4

Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age

Joris Deelen, Marian Beekman, Hae-Won Uh, Linda Broer, Kristin L Ayers, Qihua Tan, Yoichiro Kamatani, Anna M Bennet, Riin Tamm, Stella Trompet, Daníel F Guðbjartsson, Friederike Flachsbart, Giuseppina Rose, Alexander Viktorin, Krista Fischer, Marianne Nygaard, Heather J Cordel, Paolina Crocco, Erik B van den Akker, Stefan Böhringer, Quinta Helmer, Christopher P Nelson, Gary I Saunders, Maris Alver, Karen Andersen-Ranberg, Marie E Breen, Ruud van der Breggen, Amke Caliebe, Miriam Capri, Elisa Cevenini, Joanna C Collerton, Serena Dato, Karen Davies, Ian Ford, Jutta Gampe, Paolo Garagnani, Eco J C de Geus, Jennifer Harrow, Diana van Heemst, Bastiaan T Heijmans, Femke-Anouska Heinsen, Jouke-Jan Hottenga, Albert Hofman, Bernard Jeune, Palmi V Jonsson, Mark Lathrop, Doris Lechner, Carmen Martin-Ruiz, Susan E Mcnerlan, Evelin Mihailov, Alberto Montesanto, Simon P Mooijaart, Anne Murphy, Ellen A Nohr, Lavinia Paternoster, Iris Postmus, Fernando Rivadeneira, Owen A Ross, Stefano Salvioli, Naveed Sattar, Stefan Schreiber, Hreinn Stefánsson, David J Stott, Henning Tiemeier, André G Uitterlinden, Rudi G J Westendorp, Gonneke Willemsen, Nilesh J Samani, Pilar Galan, Thorkild I A Sørensen, Dorret I Boomsma, J Wouter Jukema, Irene Maeve Rea, Giuseppe Passarino, Anton J M de Craen, Kaare Christensen, Almut Nebel, Kári Stefánsson, Andres Metspalu, Patrik Magnusson, Hélène Blanché, Lene Christiansen, Thomas B L Kirkwood, Cornelia M van Duijn, Claudio Franceschi, Jeanine J Houwing-Duistermaat and P Eline Slagboom

Hum Mol Genet 2014; In press.

Abstract

The genetic contribution to the variation in human lifespan is ~25%. Despite the large number of identified disease-susceptibility loci, it is not known which loci influence population mortality. We performed a genome-wide association meta-analysis of 7,729 long-lived individuals of European descent (≥ 85 years) and 16,121 younger controls (< 65 years) followed by replication in an additional set of 13,060 long-lived individuals and 61,156 controls. In addition, we performed a subset analysis in cases aged ≥ 90 years. We observed genome-wide significant association with longevity, as reflected by survival to ages beyond 90 years, at a novel locus, rs2149954, on chromosome 5q33.3 (OR = 1.10, $P = 1.74 \times 10^{-8}$). We also confirmed association of rs4420638 on chromosome 19q13.32 (OR = 0.72, $P = 3.40 \times 10^{-36}$), representing the *TOMM40*/*APOE*/*APOC1* locus. In a prospective meta-analysis (n = 34,103), the minor allele of rs2149954 (T) on chromosome 5q33.3 associates with increased survival (HR = 0.95, P = 0.003). This allele has previously been reported to associate with low blood pressure in middle age. Interestingly, the minor allele (T) associates with decreased cardiovascular mortality risk, independent of blood pressure. We report on the first genome-wide association study-identified longevity locus on chromosome 5q33.3 influencing survival in the general European population. The minor allele of this locus associates with low blood pressure in middle age, although the contribution of this allele to survival may be less dependent on blood pressure. Hence, the pleiotropic mechanisms by which this intragenic variation contributes to lifespan regulation have to be elucidated.

Introduction

Worldwide, human life expectancy has increased remarkably over the last two centuries [1], although the healthy life expectancy lags behind. Citizens of the European Union, for example, spend only 75-80% of their lifespan in good health [2]. Families in which longevity clusters form an exception in this sense, by showing beneficial or "youthful" profiles for many metabolic and immune-related parameters [3-7] and a low prevalence of common diseases from middle age onwards [5,8,9]. Therefore, the genome of long-lived individuals is investigated to identify variants that promote healthy aging and protect against age-related disease. This is a major challenge because the genetic component of lifespan variation in the population at large has been estimated to be only ~25% [10,11] and is assumed to be determined by many, still uncharacterized, genes [12,13]. Genetic influences on human longevity are expected to reflect longevity assurance mechanisms acting across species [14], as well as more heterogeneous population-specific effects. Although numerous genome-wide association studies (GWAS) have successfully identified loci involved in common age-related diseases [15], the corresponding susceptibility loci do not explain the genetic component of human longevity [16]. GWAS for human longevity have thus far failed to identify genome-wide significant loci, besides the well-known *TOMM40/APOE/APOC1* locus [17-19].

In this paper we conducted a large genome-wide association meta-analysis of human longevity in 14 studies with long-lived cases (≥ 85 years) and younger controls (< 65 years) from European descent. In addition, we performed a subset analysis in cases aged ≥ 90 years. The novel longevity locus we identified was tested for association with prospective (cause-specific) mortality in a meta-analysis of 11 European cohorts and examined for association with various metabolic traits that may explain the mechanism by which the locus contributes to survival to high ages.

Results

Genome-wide association analysis

In order to identify novel loci involved in lifespan regulation, we conducted a meta-analysis on GWAS data of 7,729 long-lived cases (≥ 85 years) and 16,121 younger controls (< 65 years) from 14 studies originating from 7 European countries (Table S4.1). For each study, cases and controls originated from the same country. Given the higher heritability of longevity at older ages [11,20], we performed a subset analysis in which we compared cases aged ≥ 90 years (n = 5,406) with 15,112 controls (< 65 years) from the corresponding control cohorts. Replication was performed in 13,060 cases aged ≥ 85 years (of which 7,330 were ≥ 90 years) and 61,156 controls from 6 additional studies, of which 3 originated from European countries not represented in the discovery phase meta-analysis (Table S4.1). Analysis of each study was performed using a logistic regression-based method and results were adjusted for study-specific genomic inflation factors (λ) (Table S4.2). Meta-analysis was performed on 2,480,356 (≥ 85 years) and 2,470,825 (≥ 90 years) imputed single

nucleotide polymorphisms (SNPs) using a fixed-effect approach and results were further adjusted for the overall genomic inflation factor ($\lambda = 1.019$) (Figure S4.1). A flow chart of the consecutive analysis steps is depicted in Figure 4.1.

The discovery phase meta-analyses of the cases aged ≥ 85 years (n = 7,729) showed genome-wide significant association with survival into old age at one locus, the previously identified *TOMM40/APOE/APOC1* locus [17,21] (rs4420638 (G); odds ratio (OR) = 0.71, $P = 6.14 \times 10^{-19}$, Table 4.1). No gender-dependent effects were observed in the sex-stratified analysis of the cases aged ≥ 85 years (Table S4.4). The discovery-phase meta-analysis of the cases aged ≥ 90 years (n = 5,406) showed a similar result, i.e., the *TOMM40/APOE/APOC1* locus was the only genome-wide significant locus (OR = 0.64, $P = 4.09 \times 10^{-21}$, Figure 4.2 and Table 4.2). The regional association plot and forest plot for the *TOMM40/APOE/APOC1* locus are depicted in Figures 4.3 and 4.4, respectively.

Although several SNPs on chromosome 19q13.32, which are in moderate linkage disequilibrium (LD) with rs4420638, show additional association with survival into old age, meta-analysis conditional on rs4420638 showed no independent associations among these SNPs (Figure S4.2 and Table S4.3).

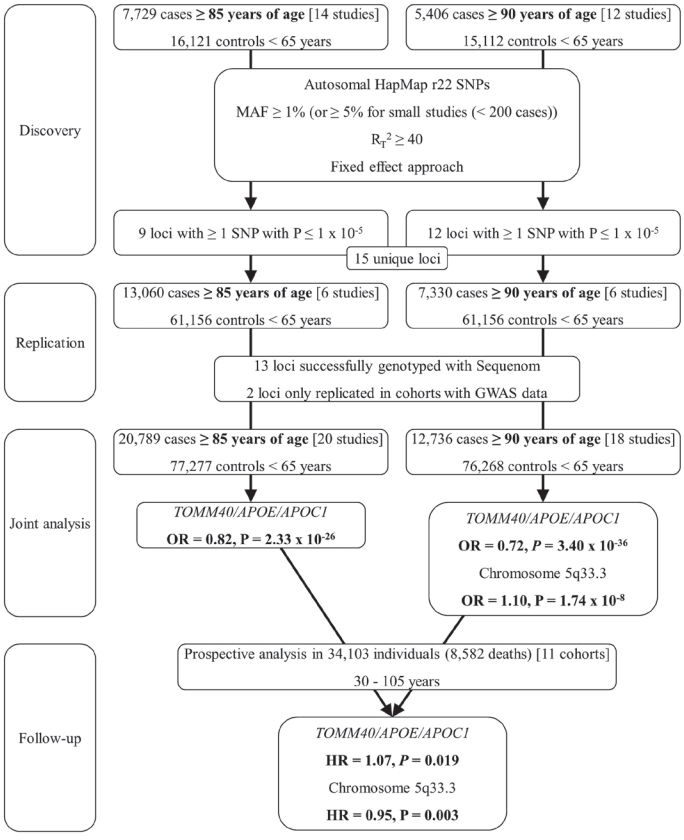
Replication

In addition to the *TOMM40/APOE/APOC1* locus, we found eight loci that showed suggestive evidence for association in the discovery-phase meta-analysis of cases aged ≥ 85 years ($P \leq 1 \times 10^{-5}$, Table 4.1), whereas six additional SNPs met this criterion in the meta-analysis of cases aged ≥ 90 years (Table 4.2). The most or (when not successfully measured) second most significant SNPs from these 14 loci and the *TOMM40/APOE/APOC1* locus were taken forward for replication in 13,060 cases aged ≥ 85 years (of which 7,330 were also ≥ 90 years) and 61,156 controls from 6 additional studies. In the joint analysis of the discovery and replication phase of the cases aged ≥ 85 years (9 loci), the *TOMM40/APOE/APOC1* locus remained the only genome-wide significant locus (Table 4.1). The joint analysis of the discovery and replication phase of the cases aged ≥ 90 years (12 loci), however, showed an additional genome-wide significant locus, rs2149954 (T), on chromosome 5q33.3 (OR = 1.10, $P = 1.74 \times 10^{-8}$, Table 4.2). Although the association of this SNP with survival up to 85 years is not genome-wide significant (OR = 1.07, $P = 4.34 \times 10^{-6}$, Table 4.1), the locus likely affects survival from middle age onwards. The regional association plot (based on the discovery phase only) and forest plot of this locus are depicted in Figures 4.3 and 4.4, respectively. Conditional analysis of rs4420638 in the discovery phase studies showed that the association of rs2149954 (T) with survival is independent of the *TOMM40/APOE/APOC1* locus ($P = 7.20 \times 10^{-6}$ instead of $P = 5.98 \times 10^{-6}$ in the analysis of survival up to 85 years).

Prospective analysis

To determine the association of rs4420638 (*TOMM40/APOE/APOC1* locus) and rs2149954 (chromosome 5q33.3 locus) with longitudinal survival, we performed a prospective meta-analysis of the 2 SNPs in 34,103 individuals aged 30-105 years

Figure 4.1 Flow chart of experimental work. The analysis in the cases aged ≥ 90 years is a subset analysis of the analysis in the cases aged ≥ 85 years. Twelve out of 14 studies used for the discovery phase analysis of cases aged ≥ 85 years contained at least 100 cases over 90 years of age and were thus analyzed in the subset analysis of cases aged ≥ 90 years.



from 11 different cohorts, of which 8,582 had died after a mean follow-up time ranging from 2.2 to 17.4 years (Table S4.5). Carriers of the minor allele of rs4420638 (G) showed significantly higher all-cause mortality (hazard ratio (HR) = 1.07, P = 0.019), whereas carriers of the minor allele of rs2149954 (T) demonstrated significantly lower all-cause mortality (HR = 0.95, P = 0.003, Table S4.6).

Association with cardiovascular disease and blood pressure

To gain insight into the mechanism by which the chromosome 5q33.3 locus might promote human longevity, we analyzed the cause-specific mortality of rs2149954. Carriers of the minor allele of rs2149954 have a lower mortality risk for cardiovascular disease (CVD) (HR = 0.86, P = 0.004), which mainly appeared to be caused by protection from stroke (HR = 0.60, P = 2.27 x 10^{-7}). In

Table 4.1 Results of the discovery phase, replication phase and joint analysis of cases aged ≥ 85 years.

							n		EAF						
Locus	Lead SNP	Chr	Position (bp)	Candidate / closest gene	EA	Analysis	Cases	Controls	Cases	Controls	OR	95% CI	*P*	I^2 (%)	P_{het}
1q43	rs1625040	1	235,213,002	*MTR, RYR2*	A	Discovery	7,729	16,121	0.170	0.150	1.16	1.09 - 1.23	3.36×10^{-6}		
						Replication	13,027	60,914	0.178	0.182	1.02	0.98 - 1.07	0.216		
						Joint	20,756	77,035			1.07	1.03 - 1.10	3.50×10^{-4}	31.0	0.093
2q24.3	rs6432832	2	166,079,072	*CSRNP3*	A	Discovery	7,729	16,121	0.344	0.321	1.12	1.07 - 1.17	2.79×10^{-6}		
						Replication	13,019	60,824	0.346	0.339	1.03	1.00 - 1.07	0.029		
						Joint	20,748	76,945			1.06	1.03 - 1.09	8.73×10^{-6}	0.0	0.467
4q27	rs13114426	4	120,942,533	*PDE5A, MAD2L1*	T	Discovery	7,729	16,121	0.387	0.405	0.90	0.87 - 0.95	2.20×10^{-5}		
						Replication	13,024	60,932	0.364	0.351	1.00	0.97 - 1.04	0.711		
						Joint	20,753	77,053			0.97	0.94 - 0.99	0.033	46.5	0.012
5q33.3	rs2149954	5	157,753,180	*EBF1*	T	Discovery	7,729	16,121	0.388	0.360	1.12	1.07 - 1.17	5.98×10^{-6}		
						Replication	12,973	60,262	0.365	0.352	1.04	1.01 - 1.07	0.013		
						Joint	20,702	76,383			1.07	1.04 - 1.09	4.34×10^{-6}	28.2	0.118
8q13.3	rs10957550*	8	72,457,142	*EYA1*	A	Discovery	7,727	16,093	0.268	0.285	0.88	0.84 - 0.93	3.61×10^{-6}		
						Replication	10,056	56,262	0.236	0.244	0.95	0.92 - 0.99	0.012		
						Joint	17,783	72,355			0.92	0.90 - 0.95	1.41×10^{-6}	29.4	0.130
10q23.33	rs4466755	10	96,622,243	*CYP2C19, CYP2C9*	T	Discovery	7,729	16,121	0.454	0.443	1.12	1.07 - 1.16	2.72×10^{-6}		
						Replication	13,051	61,105	0.488	0.508	0.98	0.95 - 1.01	0.129		
						Joint	20,780	77.226			1.03	1.00 - 1.05	0.161	65.6	2.15×10^{-5}
17q23.3	rs17760362	17	58,772,399	*TANC2*	A	Discovery	7,729	16,121	0.252	0.233	1.13	1.07 - 1.19	5.38×10^{-6}		
						Replication	13,007	60,679	0,252	0.249	1.04	1.00 - 1.07	0.033		
						Joint	20,736	76,800			1.07	1.04 - 1.10	1.56×10^{-5}	0.0	0.473

Locus	Lead SNP	Chr	Position (bp)	Candidate / closest gene	EA	Analysis	*n* Cases	*n* Controls	EAF Cases	EAF Controls	OR	95% CI	*P*	I^2 (%)	P_{het}
19q13.32	rs4420638*	19	50,114,786	*APOE*	G	Discovery	7,728	16,111	0.157	0.195	0.71	0.67 - 0.77	6.14×10^{-19}		
						Replication	10,165	57,126	0.180	0.202	0.87	0.83 - 0.91	2.12×10^{-12}		
						Joint	17,893	73,237			0.82	0.79 - 0.85	2.33×10^{-26}	80.2	4.35×10^{-10}
20q13.2	rs8126377	20	51,590,254	*TSHZ2, ZNF217*	G	Discovery	7,532	15,902	0.059	0.069	0.79	0.71 - 0.87	1.35×10^{-5}		
						Replication	12,974	60,647	0.058	0.054	1.01	0.94 - 1.08	0.901		
						Joint	20,506	76,549			0.93	0.88 - 0.99	0.020	51.1	0,006

Chr, chromosome according to NCBI build 36; *Position*, (bp), position of the lead SNP according to NCBI build 36; *EA*, effect allele; *EAF*, effect allele frequency after pooling the data of all analyzed individuals; *OR*, odds ratio for the effect allele; *95% CI*, 95% confidence interval; I^2, heterogeneity statistic; P_{het}, *P*-value for heterogeneity. *Genotyping of these SNPs with the Sequenom MassARRAY system for the replication phase was unsuccessful. The SNPs in **bold** overlap with Table 4.2.

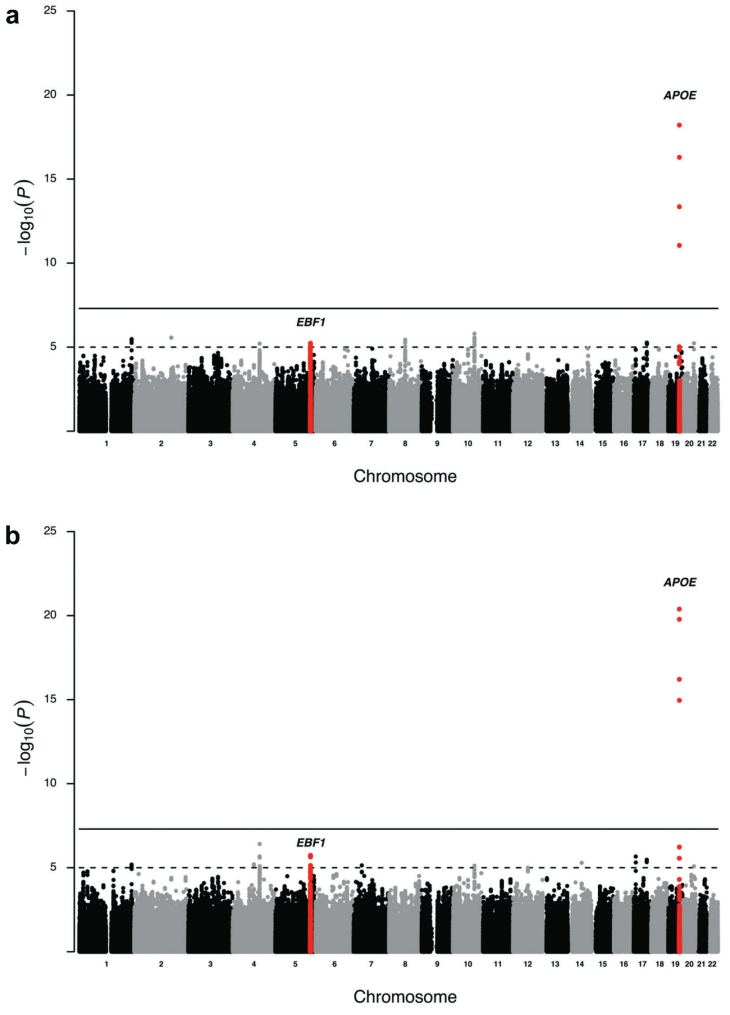
Table 4.2 Results of the discovery phase, replication phase and joint analysis of cases aged ≥ 90 years.

							n		EAF							
Locus	Lead SNP	Chr	Position (bp)	Candidate / closest gene	EA	Analysis	Cases	Controls	Cases	Controls	OR	95% CI	*P*	I^2 (%)	P_{het}
1q43	rs1625040	1	235,213,002	*MTR, RYR2*	A	Discovery	5,406	15,112	0.176	0.150	1.18	1.10 - 1.26	6.53×10^{-6}		
						Replication	7,310	60,914	0.175	0.182	1.05	0.99 - 1.10	0.065		
						Joint	12,716	76,026			1.10	1.05 - 1.14	2.60×10^{-5}	9.3	0.343
4q22.2	rs4693331	4	94,760,609	*GRID2*	C	Discovery	5,406	15,112	0.416	0.444	0.89	0.84 - 0.93	6.63×10^{-6}		
						Replication	7,267	60,324	0.449	0.440	1.03	0.99 - 1.07	0.095		
						Joint	12,673	75,436			0.97	0.94 - 1.00	0.139	61.3	3.51×10^{-4}
4q27	rs13114426	4	120,942,533	*PDE5A, MAD2L1*	T	Discovery	5,406	15,112	0.381	0.405	0,88	0.84 - 0.92	2.11×10^{-6}		
						Replication	7,305	60,932	0.369	0.351	0.98	0.94 - 1.02	0.336		
						Joint	12,711	76,044			0.94	0.91 - 0.97	1.95×10^{-4}	32.5	0.090
5q33.3	rs2149954	5	157,753,180	*EBF1*	T	Discovery	5,406	15,112	0.396	0.360	1.14	1.09 - 1.21	1.85×10^{-6}		
						Replication	7,298	60,262	0.374	0.352	1.07	1.03 - 1.12	5.98×10^{-4}		
						Joint	12,704	75,374			1.10	1.06 - 1.14	1.74×10^{-8}	28.5	0.125
7p14.2	rs11977641	7	36,761,949	*AOAH, ELMO1*	C	Discovery	5,406	15,112	0.062	0.076	0.78	0.70 - 0.87	7.31×10^{-6}		
						Replication	3,049	4,805	0.071	0.073	0.93	0.82 - 1.06	0.226		
						Joint	8,455	19,917			0.84	0.77 - 0.91	1.57×10^{-5}	50.2	0.010
10q23.33	rs4466755	10	96,622,243	*CYP2C19, CYP2C9*	T	Discovery	5,406	15,112	0.455	0.445	1.13	1.07 - 1.18	1.30×10^{-5}		
						Replication	7,326	61,105	0.477	0.508	0.98	0.94 - 1.02	0.208		
						Joint	12,732	76,217			1.03	1.00 - 1.07	0.087	55.4	0.002
12q15	rs11834614	12	67,197,344	*MDM1, RAP1B*	C	Discovery	5,406	15,112	0.138	0.155	0.85	0.79 - 0.91	9.94×10^{-6}		
						Replication	7,272	60,210	0.165	0.173	1.01	0.96 - 1.07	0.603		
						Joint	12,678	75,322			0.95	0.91 - 0.99	0.023	43.9	0.024
14q23.2	rs2784505	14	61,501,766	*SYT16*	G	Discovery	5,406	15,112	0.080	0.067	1.23	1.11 - 1.35	8.87×10^{-5}		
						Replication	7,323	60,979	0.070	0.066	1.10	1.02 - 1.19	0.012		
						Joint	12,729	76,091			1.15	1.08 - 1.22	9.47×10^{-6}	28.3	0.127
17p13.1	rs940850	17	8,870,805	*NTN1*	T	Discovery	5,405	15,112	0.072	0.093	0.78	0.70 - 0.87	4.93×10^{-6}		
						Replication	7,276	60,146	0.109	0.118	1.03	0.97 - 1.10	0.318		
						Joint	12,681	75,258			0.95	0.90 - 1.01	0.111	63.7	1.32×10^{-4}

Locus	Lead SNP	Chr	Position (bp)	Candidate / closest gene	EA	Analysis	*n* Cases	*n* Controls	EAF Cases	EAF Controls	OR	95% CI	*P*	I^2 (%)	P_{het}
17q23.2	rs2109265	17	58,307,001	*MARCH10, TANC2*	A	Discovery	5,406	15,112	0.443	0.420	1.13	1.08 - 1.19	3.34×10^{-6}		
						Replication	7,307	60,672	0.453	0.465	1.01	0.97 - 1.05	0.671		
						Joint	12,713	75,784			1.06	1.02 - 1.09	0.001	34.7	0.074
19q13.32	rs4420638*	19	50,114,786	*APOE*	G	Discovery	5,405	15,102	0.145	0.195	0.64	0.59 - 0.70	4.09×10^{-21}		
						Replication	4,861	57,126	0.165	0.202	0.77	0.72 - 0.82	2.95×10^{-18}		
						Joint	10,266	72,228			0.72	0.68 - 0.76	3.40×10^{-36}	70.1	3.69×10^{-5}
20q13.2	rs8126377	20	51,590,254	*TSHZ2, ZNF217*	G	Discovery	5,209	14,893	0.057	0.068	0.75	0.66 - 0.85	3.38×10^{-5}		
						Replication	7,278	60,647	0.063	0.054	1.04	0.95 - 1.13	0.309		
						Joint	12,487	75,540			0.94	0.87 - 1.00	0,117	58.1	0.001

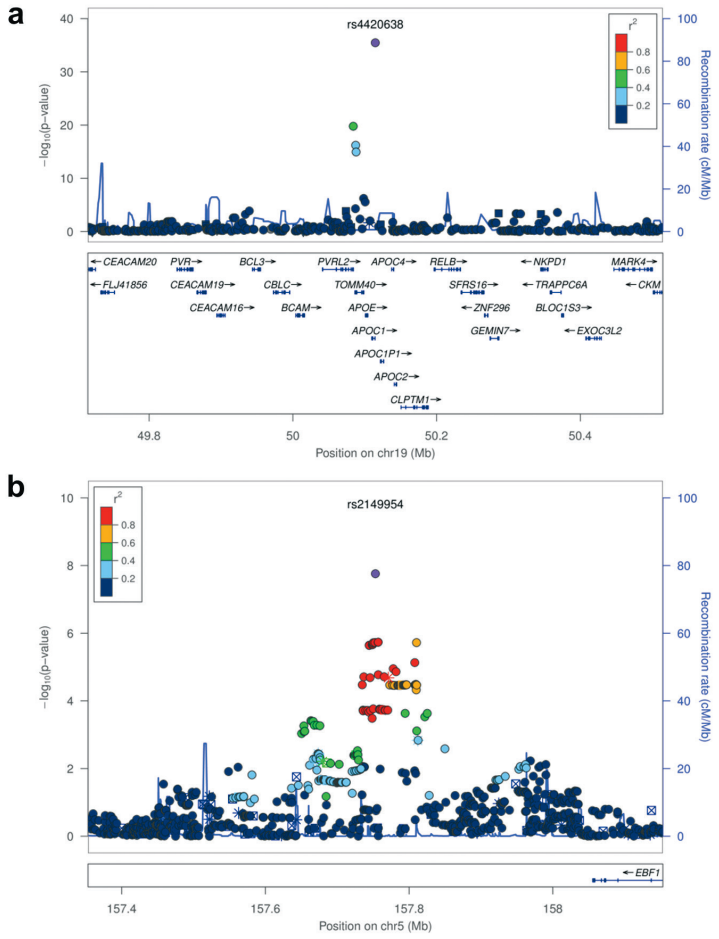
Chr, chromosome according to NCBI build 36; *Position* (bp); position of the lead SNP according to NCBI build 36; *EA*, effect allele; *EAF*, effect allele frequency after pooling the data of all analyzed individuals; *OR*, odds ratio for the effect allele; *95% CI*, 95% confidence interval; I^2, heterogeneity statistic; P_{het}, *P*-value for heterogeneity. *Genotyping of this SNP with the Sequenom MassARRAY system for the replication phase was unsuccessful. The SNPs in **bold** overlap with Table 4.1.

Figure 4.2 Results of the discovery phase analysis. Manhattan plot presenting the -$\log_{10}$ *P*-values from the discovery phase analysis of cases aged ≥ 85 years (A) and ≥ 90 years (B). The loci that showed a genome-wide significant association after the joint analysis of the discovery and replication phase (chromosome 19q13.32 and 5q33.3) are shown in red.



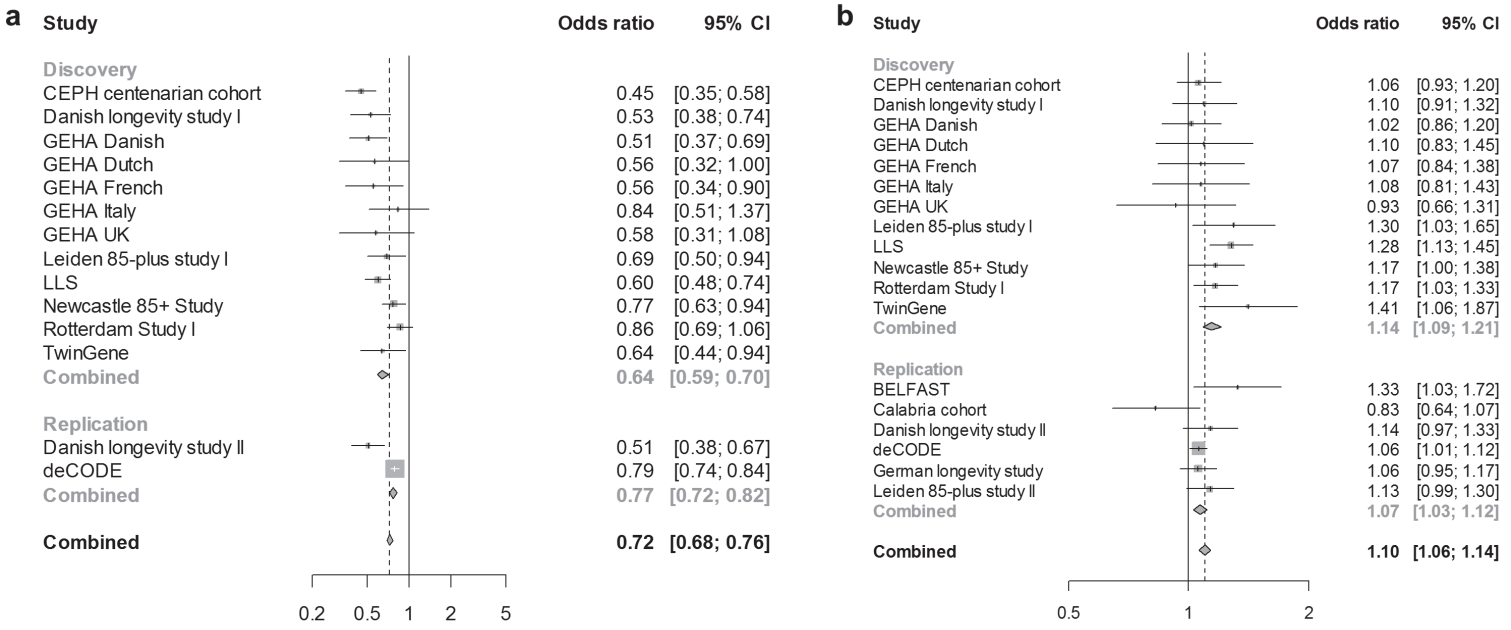
addition, we observed an effect of this SNP on non-CVD mortality (HR = 0.86, *P* = 0.002, Table S4.7). We also examined the Coronary ARtery DIsease Genome-Wide Replication And Meta-Analysis (CARDIoGRAM) GWAS [23], which showed a significant association of rs2149954 with a decreased risk for coronary artery disease (CAD) (OR = 0.96, *P* = 0.011, Table S4.8). In addition, two SNPs on chromosome 5q33.3 in high LD with rs2149954, rs9313772 (r^2 = 0.928) and rs11953630 (r^2 = 0.854) have previously been reported to associate with blood pressure and hypertension [24,25]. As expected, examining rs2149954 in the International Consortium for Blood Pressure

Figure 4.3 Regional association plots for the chromosome 19q13.32 and 5q33.3 loci. Results of the discovery-phase analysis of chromosome 19q13.32 (A) and 5q33.3 (B) in cases aged ≥ 90 years, generated using LocusZoom (http://csg.sph.umich.edu/locuszoom/) [22]. For the two single nucleotide polymorphisms (SNPs) taken forward to the replication phase (rs4420638 and rs2149954), the results of the joint analysis are plotted. The color of the SNPs is based on the linkage disequilibrium with the lead SNP (shown in purple). The blue peaks represent the recombination rates based on HapMap Phase I+II CEU release 22 (hg18/build36), and the RefSeq genes in the region are shown in the lower panel.



GWAS [24] showed a significant association of the minor allele with lower diastolic ($P = 3.46 \times 10^{-5}$) and systolic ($P = 6.55 \times 10^{-6}$) blood pressure (DBP/SBP) (Table S4.9). Despite the highly interesting association of the minor allele of rs2149954 with low blood pressure and a decreased risk for CAD, stroke, and mortality, its association with decreased all-cause mortality was not influenced by blood pressure in two studies of participants aged ≥ 75 years (the PROspective Study of Pravastatin in the Elderly at Risk (PROSPER) and Leiden 85-plus study Cohort II, Table S4.10). This may indicate that at higher ages, this locus influences longevity via pathways

Figure 4.4 Forest plots for rs4420638 and rs2149954. Forest plots representing the odds ratios with 95% confidence interval of rs4420638 (A) and rs2149954 (B) for the cohorts analyzed in the discovery and replication phase (≥ 90 years). The size of the boxes represents the sample size of the cohort.



additional to those involved in blood pressure regulation.

Phenotypic characterization and pathway analysis

In an attempt to identify the underlying mechanism by which this novel longevity locus at chromosome 5q33.3 could influence human longevity, we examined rs2149954 in the published data of several large GWAS consortia for association with metabolic traits in generally middle-aged individuals. None of the investigated traits, i.e., 2-h glucose (OGTT), Hb1Ac, fasting glucose, fasting insulin, insulin resistance (HOMA-IR), β-cell activity (HOMA-B), total/high-density lipoprotein/low-density lipoprotein cholesterol, triglycerides, and type 2 diabetes [26-32], demonstrated evidence for association (all $P > 0.05$) with rs2149954 (Tables S4.8 and S4.9).

Gene set enrichment analysis (GSEA) of the meta-analysis results of the discovery-phase analysis of survival ≥ 90 years of age using Meta-Analysis Gene-set Enrichment of variaNT Associations (MAGENTA) [33], as well as examination of interconnectivity of implicated genes using Gene Relationships Across Implicated Loci (GRAIL) [34] (Figure S4.3 and Table S4.11), provided no firm clues for potential pathways involved in human longevity.

Fine mapping and functional characterization

The newly identified longevity locus on chromosome 5q33.3 is located in an intergenic region on chromosome 5q33.3, 302 kb downstream of the *EBF1* gene. To determine the functional impact of this

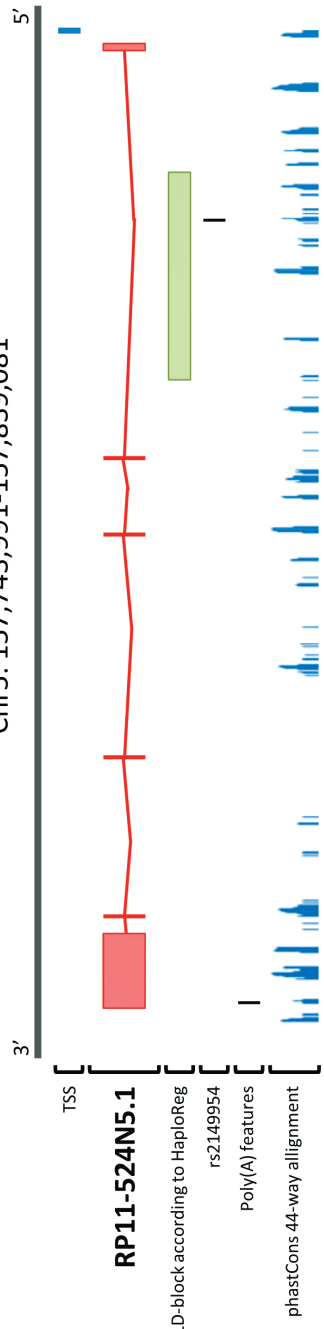


Figure 4.5 Chromosomal region around rs2149954. The region contains a long intronic non-coding RNA (RP11-524N5.1) for which the poly(A) features are supported by PolyA-seq reads from liver, muscle and testis. RP11-524N5.1 is transcribed from the negative strand, and the phastCons 44-way alignment supports conservation of the transcription start site, 3' untranslated region and the third, fifth and last exon of the transcript. Rs2149954 and the 25 single nucleotide polymorphisms in high linkage disequilibrium ($r^2 \geq 0.8$, according to HaploReg v2 [35]) are located in the first intron of RP11-524N5.1.

locus, we first identified the SNPs in LD with rs2149954 ($r^2 \geq 0.8$) using the 1000 Genomes CEU Phase 1 data implemented in HaploReg v2 (http://www.broadinstitute.org/mammals/haploreg/haploreg.php) [35]. In total, we identified 25 SNPs, spanning a region of ~22.3 kb (Table S4.12). Subsequently, we examined the potential effects of these SNPs on gene expression using several expression quantitative trait locus (eQTL) databases. None of the SNPs showed an association with gene expression in the various examined tissues, so it is still unclear in which tissue(s) the locus exert its longevity-promoting effect. We did, however, find some promising functional implication of this locus, i.e., the presence of multiple DNase I hypersensitivity sites, transcription factor binding sites and enhancer histone marks, by exploring ENCODE data using HaploReg v2 [35] and RegulomeDB (http://www.regulomedb.org/) [36] (Table S4.12). Very recently, a large intergenic non-coding RNA (lincRNA), RP11-524N5.1, has been annotated right on top of our locus. The poly(A) features of this lincRNA are supported by PolyA-seq reads from liver, muscle, and testis. PhastCons 44-way alignment supports conservation of the transcription start site, 3' untranslated region and the third, fifth, and last exon of the lincRNA transcript (Figure 4.5). The transcript does not align to the mouse genome, but orthologous transcripts are found in other primate genome sequences, suggesting that this is a primate-specific lincRNA.

Discussion

We have performed the largest genome-wide association meta-analysis for human longevity, in which a novel locus on chromosome 5q33.3 associating with survival beyond 90 years was identified.

The minor allele of rs2149954 (T) promotes human longevity by reducing the risk of mortality owing to stroke and non-cardiovascular causes. In addition, this allele has previously been associated with low blood pressure, which may explain the protection from CVD mortality risk in middle age. At ages above 80 years, however, low SBP associates with increased mortality [37,38]. Hence, the observed blood pressure-independent association of the minor allele with mortality $\geq$ 75 years may be due to pleiotropic effects on other mortality-related clinical parameters. Examination of publically available data of several large GWAS consortia for association of the locus with parameters related to glucose and fat metabolism provided as yet no clues for other potentially involved mechanisms.

Rs2149954 is located in an intergenic region on chromosome 5q33.3 between *CLINT1* and *EBF1*. The presence of several regulatory elements in this region implies that transcription factor binding and/or expression of (nearby) genes could be influenced. The currently available eQTL databases did not provide evidence for such effects, which might be due to the limited tissue diversity of the databases. The effects of the chromosome 5q33.3 locus on human longevity might be exerted through the lincRNA, which has recently been annotated right on top of our locus (RP11-524N5.1) and shows evidence for expression in liver, muscle and testis. LincRNAs are involved in chromatin modification and transcriptional regulation [39] and seem to play a role in

human disease [40]. However, the newly annotated lincRNA is not yet available in the large eQTL databases, and the effect of SNPs in the chromosome 5q33.3 locus on expression of this transcript still needs to be determined. Hence, further functional studies are required to illuminate the mechanism by which this locus influences human longevity.

GWAS has thus far not been a successful approach to identify genome-wide significant hits for human longevity or mortality besides the well-known *TOMM40/APOE/APOC1* locus [17-19]. The *FOXO3A* locus, for which the longevity effect is most prominent in individuals aged ≥ 100 years [41], showed only moderate evidence for association with survival ≥ 90 years in the discovery phase of our GWAS (lowest $P = 1.35 \times 10^{-4}$ (rs1268161)). Sebastiani and colleagues suggested that human longevity might be explained by a signature consisting of 281 SNPs [42]. However, none of the SNPs (except the already known SNP rs2075650 in *TOMM40*) was significant after adjustment for multiple testing ($P = 1.78 \times 10^{-4}$ (0.05/281)). In addition, we did not observe an enrichment of significant SNPs from their signature in our data ($\lambda = 1.004$, Figure S4.4). Because the association of SNPs other than the *TOMM40/APOE/APOC1* locus could not be replicated in this, much larger, GWAS, we have doubts that these signature SNPs are indeed candidate SNPs influencing human longevity. Although we detected merely one novel genome-wide significant locus, the current GWAS had sufficient power, based on our results, to detect lifespan regulating loci with relatively small effects (OR's < 0.9 and > 1.1).

The genetic component of human longevity is small (~25%) [10,11] and is assumed to be determined by many genes [12,13]. Furthermore, the genetic heterogeneity in aging and lifespan regulation is expected to be high, because individual genes may contribute by a diversity of late acting deleterious stochastic (germline) variation resulting in a genetic component that is hard to disentangle [13]. GWAS for complex late-onset diseases, such as osteoarthritis and Alzheimer's disease, with sample sizes comparable to our current study [43-45], have identified more loci compared with GWAS for longevity. This most likely reflects the greater inherent complexity of the longevity trait, with its diverse spectrum of biological pathways subject to intrinsic and extrinsic (environmental) interactions. Hence, even larger GWAS (> 50,000 long-lived individuals) may be required to identify additional longevity loci, preferably in the most stringent phenotype, i.e., the oldest old.

As survival to ages ≥ 85 or 90 years is relatively common in Western populations, the human longevity trait suffers from etiological heterogeneity. Lifespan extension in the past generations owing to non-genetic factors likely created phenocopies diluting the genetic component of survival to ages ≥ 85 years. The genetic contribution to survival to ages ≥ 100 years is higher but will render smaller sample sizes for GWAS. This may explain why the novel locus on chromosome 5q33.3 was only genome-wide significant in the subset analysis of cases aged ≥ 90 years. For the same reason, a large number of individuals from the control groups (up to 50%, depending on the gender and year of birth of the individuals and demography of

the cohort) will live to ages ≥ 85 years. In 2011, the mean life expectancy at age 65 in Europe was 21.3 years for women and 17.8 years for men (http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/dataset?P_product_code=TSDDE210), which makes selection of proper controls a challenging issue. The most ideal controls would be individuals from the same birth cohort as the long-lived cases that survived to the mean age of death of that birth cohort. However, for most of these individuals there is no DNA available. Alternatively, we selected controls that have not yet reached the age of 65 years at inclusion to represent the frequency of variants in the general population and minimize selection owing to mortality. Hence, the low contrast between cases and controls likely has reduced our probability of identifying longevity loci.

In addition, there will be differences between case and control cohorts that may have had an impact on our results. An example of a potential confounder is smoking behavior, which was not adequately measured in most elderly cohorts. However, none of the SNPs that were previously associated with smoking behavior in cohorts from European descent (according to the National Human Genome Research Institute GWAS Catalog (http://www.genome.gov/gwastudies/)), namely rs1051730, rs1329650 and rs4105144, show differences between cases (≥ 85 years) and controls in the joint analysis of the discovery and replication phase (all $P > 0.05$). We have to note that these SNPs only explain a small proportion of the variance observed in smoking behavior. However, as the frequency of these proxy SNPs for smoking behavior is similar between cases and controls, we expect no obvious differences in smoking behavior between the groups.

In conclusion, besides the previously implicated *TOMM40/APOE/APOC1* locus, we identified a novel locus on chromosome 5q33.3 that associates with survival beyond 90 years. Although rs2149954 is associated with survival beyond 90 years at a genome-wide significant level in our study, replication in additional cohorts from European as well as non-European descent is warranted. The minor allele of the lead SNP at this locus, rs2149954, promotes human longevity in a prospective meta-analysis by lowering the risk of mortality owing to stroke and non-cardiovascular causes. The locus harbors a lincRNA and is implicated in blood pressure regulation, but the mechanism by which it influences longevity likely also involves other traits.

Materials and methods

Study populations

The discovery analysis was performed in 7,729 cases that survived to ages ≥ 85 years (of which 5,406 also survived to ages ≥ 90 years) and 16,121 controls below 65 years at baseline, from 14 studies. Replication was performed in 13,060 cases that survived to ages ≥ 85 years (of which 7,330 also survived to ages ≥ 90 years) and 61,156 controls below 65 years at baseline, from 6 additional studies. All individuals were of European descent. The details of the discovery and replication studies can be found in Tables S4.1 and S4.2. Some cohorts only provided controls (GOYA, NTR,

SU.VI.MAX, TwinsUK and WTCCC2) or only cases (BELFAST, CEPH centenarian cohort, Danish longevity study I/II, Leiden 85-plus study I/II and Newcastle 85+ Study), whereas others contained both (Calabria cohort, deCODE, EGCUT, GEHA project, German longevity study, Leiden Longevity Study, Rotterdam Study I/II and TwinGene). The names of the studies in the tables and figures are based on the names of the cohorts containing the cases. The cases and controls used for each study originated from the same country (Table S4.1). The only exception is BELFAST (Northern Ireland), for which we used controls from the NTR (Netherlands). A check in the PROSPER study, which includes individuals from Northern Ireland and the Netherlands, showed that the allele frequencies in control individuals from both countries are similar for our SNPs (data not shown). All participants provided written informed consent, and the study was approved by the relevant institutional review boards.

Genotyping, imputation and genome-wide association analysis

All discovery studies were genotyped using Illumina genotyping arrays, and pre-imputation quality control was performed for each study separately. Imputation was performed using IMPUTE or MACH with reference HapMap Phase I+II CEU release 22 (hg18/build36). Further details about the genotyping, quality control and imputation of each study are summarized in Table S4.2.

Two replication studies (deCODE and the Danish longevity study II) were also genotyped using Illumina genotyping arrays and imputed using IMPUTE with reference HapMap Phase I+II CEU release 22 (hg18/build36) (Danish longevity study II) or deCODE software (deCODE). The other replication studies were genotyped with the Sequenom MassARRAY system using iPLEX Gold genotyping assays (Sequenom, San Diego, CA, USA). More information about the studies used in the replication phase can be found in Tables S4.1 and S4.2. Of the 15 SNPs measured with the Sequenom MassARRAY system, 13 were successfully genotyped in at least 95% of the samples and the average genotyping call rate was 99.80%. We also checked the concordance between the SNPs measured with the Sequenom MassARRAY system and (imputed) GWAS data of the Leiden 85-plus study I cases, and the average concordance rate was 99.07%. The two SNPs that were not successfully genotyped with the Sequenom MassARRAY system (rs10957550 and rs4420368) were only analyzed in the replication studies, which had imputed GWAS data available (deCODE and the Danish longevity study II).

All studies were analyzed separately using CC-assoc (https://www.msbi.nl/dnn/Research/Genetics/Software/TestsforGWASinrelatedindividuals(cc_assoc).aspx), which is based on a modified version of the score test that takes into account imputation uncertainty and familial relatedness [46]. SNPs with a low imputation quality ($R_T^2 \leq 40$) and a minor allele frequency of ≤ 1 or $\leq 5\%$ (if $n_{cases} < 200$) were excluded from analysis in the discovery phase. Adjustment for population stratification of the discovery studies was performed by multiplying the R_T^2-adjusted variances of the score statistic

with the genomic inflation factor (λ_{range} = 0.97 – 1.08, Table S4.2) of the study.

Meta-analyses

For the meta-analyses, a fixed-effect approach was used. Scores and variances of the studies were combined to obtain a single meta-statistic, which was adjusted using the genomic inflation factor ($\lambda = 1.019$, discovery phase only) (Figure S4.1). For each analysis, we only used studies with at least 100 cases (Table S4.1). *P*-values < 5 x 10^{-8} were considered genome-wide significant [47]. To determine heterogeneity across the studies, the between-study variance was calculated.

Conditional analysis

To ascertain independent signals at the chromosome 19q13.32 locus, we performed a meta-analysis conditional on rs4420638 in all studies used for the discovery phase analysis in cases aged ≥ 85 years. The results are depicted in Figure S4.2 and Table S4.3.

Sex-stratified analysis

Sex-stratified analysis of the cases aged ≥ 85 years (n_{women} = 5,400 and n_{men} = 1,865) was performed to investigate the presence of gender-dependent associations. In addition, the 15 loci that showed (suggestive) evidence for association with survival ≥ 85 and/or ≥ 90 years were tested for differences between sexes using the formula: $(\beta_{women} - \beta_{men}) / \sqrt{(SE_{women}^2 + SE_{men}^2)}$. The results of this analysis are depicted in Table S4.4.

Prospective analysis

Prospective analysis of rs2149954 and rs4420638 was performed using a Cox proportional hazards model adjusted for age at baseline, sex and study-specific covariates. The details about each of the analyzed cohorts are summarized in Table S4.5.

Pathway analysis

For the pathway analysis, we used GSEA implemented in MAGENTA (http://www.broadinstitute.org/mpg/magenta/) [33]. In short, each SNP is mapped to a gene considering a window of 110 kb upstream and 40 kb downstream around the genes. Subsequently, each gene is assigned a gene association score based on the SNP with the lowest *P*-value, which is mapped to that gene and this score is adjusted for confounding factors like gene size and the number of SNPs per kb. Genes within the HLA region were removed from analysis owing to high LD and high gene density in that region. The GSEA algorithm tests for over-representation of adjusted gene scores in a given pathway using a pre-defined score rank cutoff (in our case, the 95th and 75th percentile). The generated statistic is then compared with 10,000 – 1,000,000 gene sets of identical size randomly sampled from the genome to generate an empirical *P*-value for each pathway.

In total, 3,216 pathways from Gene Ontology, PANTHER, Ingenuity, KEGG, REACTOME, and BIOCARTA were tested. Pathways were considered significant if the FDR-adjusted *P*-value (the 95th or 75th percentile) was ≤ 0.05.

To determine the relationship between loci associated with survival ≥ 90 years, we used GRAIL (http://www.broadinstitute.org/mpg/grail/) [34]. In short, this program maps SNPs to genes and subsequently uses a text-mining algorithm on PubMed abstracts to determine connections between these

genes. Genes from independent loci, which share informative words, receive a high GRAIL similarity score and are more likely to be functionally related. As we only had a limited number of loci with at least one SNP with a *P*-value $\leq 1 \times 10^{-5}$ ($n = 12$, Table 4.2), we decided to perform GRAIL analysis on all loci with at least one SNP with a *P*-value $\leq 1 \times 10^{-4}$ ($n = 65$).

eQTL analysis

To determine whether rs2149954 or SNPs in LD ($r^2 \geq 0.8$ based on 1000 Genomes CEU Phase 1 data) influenced gene expression, we searched several eQTL databases, namely (1) the Gutenberg Heart Study database [48], which is based on expression data of monocytes; (2) the Genotype-Tissue Expression eQTL database (http://www.ncbi.nlm.nih.gov/gtex/GTEX2/gtex.cgi), which is based on expression data of brain (cerebellum, frontal cortex, temporal cortex, and pons), liver, and lymphoblastoid cell lines; (3) the GENe Expression VARiation database (http://www.sanger.ac.uk/resources/software/genevar/), which is based on expression data of adipose tissue, fibroblasts, T cells, skin, and lymphoblastoid cell lines [49], and (4) the Blood eQTL browser (http://genenetwork.nl/bloodeqtlbrowser/) [50].

Acknowledgments

We thank Laurens Wilming from the Wellcome Trust Sanger Institute for the annotation of the lincRNA RP11-524N5.1. This study was undertaken within the framework of European Union's Seventh Framework Programme (FP7/2007-2011) under grant agreement n° 259679 (IDEAL). A full list of acknowledgments, including support for each study, is provided in the Supplementary Information.

Funding

This work was supported by the Augustinus Foundation; Avera Institute for Human Genetics (AIHG); AXA Research Fund; Belfast City Hospital Trust Fund, Research and Education into Ageing-0153; Biobanking and Biomolecular Resources Research Infrastructure (BBMRI –NL, NWO 184.021.007); Biotechnology and Biological Sciences Research Council (BBSRC); Bristol-Myers Squibb; Center for Inherited Disease Research (CIDR); Centre for Medical Systems Biology (CMSB); CERA Foundation; Commissariat à L'Energie Atomique (CEA)-Centre National de Génotypage (CNG); Danish Agency for Science, Technology and Innovation (DASTI)/The Danish Council for Independent Research (DCIR, grant 11-107308); Danish National Research Foundation (DNRF); Department of Health and Social Services (Northern Ireland); DFG-Cluster of Excellence 'Inflammation at Interfaces'; Dunhill Medical Trust (grant R124/0509); Egmont Foundation; Estonian Science Foundation (grant 7859); Estonian Government (grant SF0180142s08); European Research Council (ERC, advanced grant 230374); European Science Foundation (ESF, EU/QLRT-2001-01254); European Union's Fifth/Sixth/Seventh Framework Programmes (FP5-QLK6-CY-2001-00128,

FP6-LIFESCIHEALTH-36894, FP6-LSHM-CT-2004-503270, FP7-HEALTH-2007-B-223004, FP7-HEALTH-F4-2007-201413, FP7-HEALTH-F4-2008-202047, FP7-HEALTH-2009-single-stage-242244 and FP7-HEALTH-2010-two-stage-259679); Fondation Caisse d'Epargne Rhône-Alpes Lyon CERAL (2004–2007); Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health (NIMH, grant MH081802); GenomEUtwin (EU/QLRT-2001-01254; QLG2-CT-2002-01254); Guy's & St Thomas' NHS Foundation Trust; Health Foundation; Heart and Lung foundation (grant 20070481); Innovation-Oriented Research Program on Genomics (SenterNovem, grant IGE05007); Institute for Ageing and Health; Institut National de la Recherche Agronomique (INRA); Institut National de la Santé et de la Recherche Médicale (INSERM); INTERREG 4A programme Syddanmark-Schleswig-K.E.R.N (with EU funds from the European Regional Development Fund); King's College London; Medical Research Council (MRC, grant G0500997 and G0601333); Ministére de l'Enseignement supérieur et de la Recherche (MESR); National Institutes of Health (NIH)/National Institute of Aging (NIA, P01AG08761,R01D0042157-01A and U01DK066134); National Institute for Health Research (NIHR) Newcastle Biomedical Research Centre; NBIC BioAssist (NWO-NBIC/BioAssist/RK/2008.024); Netherlands Consortium for Healthy Ageing (NCHA, grant 050-060-810); Netherlands Genomics Initiative (NGI); Netherlands Heart Foundation (NHF, grant 2001 D 032); Netherlands Organization for Scientific Research (NWO, MagW/ZonMW grant 904-61-090, 904-61-193, 480-04-004, 400-05-717, Spinozapremie 56-464-14192, 175.010.2005.011, 911-03-012, 985-10-002, Addiction-31160008 and Middelgroot-911-09-032); Netspar – Living longer for a good health; NHS North of Tyne (Newcastle Primary Care Trust); Pharmacy Foundation; Regione Autonoma della Sardegna; Rutgers University Cell and DNA Repository (NIMH U24 MH068457-06); Swedish Research Council (grant M-2005-1112); The Competitive Research Funding of the Tampere University Hospital and Academy of Finland; The Danish Interdisciplinary Research Council; The Health Foundation (Helsefonden); The Ministry for Higher Education; The National Program for Research Infrastructure 2007 (grant 09-063256); The March of Dimes Birth Defects Foundation; The Swedish Foundation for Strategic Research (SSF); Unilever Discover Colworth; Université Paris 13; University of Calabria; University of Tartu (grant SP1GVARENG); Velux Foundation; VU University's Institute for Health and Care Research (EMGO+) and Neuroscience Campus Amsterdam (NCA); Wellcome Trust (grant 084762, 085475 and 087436). Funding to pay the Open Access publication charges for this article was provided by IDEAL (FP7-HEALTH-2010-two-stage-259679).

References

1. Oeppen J, Vaupel JW. Demography. Broken limits to life expectancy. *Science* 2002; **296**: 1029-31.
2. Jagger C, Gillies C, Moscone F, Cambois E, van Oyen H, Nusselder W, *et al*. Inequalities in healthy life years in the 25 countries of the European Union in 2005: a cross-national meta-regression analysis. *Lancet* 2008; **372**: 2124-31.
3. Barzilai N, Atzmon G, Schechter C, Schaefer EJ, Cupples AL, Lipton R, *et al*. Unique lipoprotein phenotype and genotype associated with exceptional longevity. *JAMA* 2003; **290**: 2030-40.
4. Derhovanessian E, Maier AB, Beck R, Jahn G, Hahnel K, Slagboom PE, *et al*. Hallmark features of immunosenescence are absent in familial longevity. *J Immunol* 2010; **185**: 4618-24.
5. Newman AB, Glynn NW, Taylor CA, Sebastiani P, Perls TT, Mayeux R, *et al*. Health and function of participants in the Long Life Family Study: A comparison with other cohorts. *Aging (Albany NY)* 2011; **3**: 63-76.
6. Slagboom PE, Beekman M, Passtoors WM, Deelen J, Vaarhorst AA, Boer JM, *et al*. Genomics of human longevity. *Philos Trans R Soc Lond B Biol Sci* 2011; **366**: 35-42.
7. Wijsman CA, Rozing MP, Streefland TC, Le Cessie, Mooijaart SP, Slagboom PE, *et al*. Familial longevity is marked by enhanced insulin sensitivity. *Aging Cell* 2011; **10**: 114-21.
8. Atzmon G, Schechter C, Greiner W, Davidson D, Rennert G, Barzilai N. Clinical phenotype of families with longevity. *J Am Geriatr Soc* 2004; **52**: 274-7.
9. Westendorp RG, van Heemst D, Rozing MP, Frolich M, Mooijaart SP, Blauw GJ, *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**: 1634-7.
10. Herskind AM, McGue M, Holm NV, Sorensen TI, Harvald B, Vaupel JW. The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870-1900. *Hum Genet* 1996; **97**: 319-23.
11. Hjelmborg JV, Iachine I, Skytthe A, Vaupel JW, McGue M, Koskenvuo M, *et al*. Genetic influence on human lifespan and longevity. *Hum Genet* 2006; **119**: 312-21.
12. Finch CE, Tanzi RE. Genetics of aging. *Science* 1997; **278**: 407-11.
13. Kirkwood TB, Cordell HJ, Finch CE. Speed-bumps ahead for the genetics of later-life diseases. *Trends Genet* 2011; **27**: 387-8.
14. Sacher GA. 1976 Robert W. Kleemeier Award Lecture: Longevity, aging, and death: an evolutionary perspective. *Gerontologist* 1978; **18**: 112-9.
15. Ganna A, Rivadeneira F, Hofman A, Uitterlinden AG, Magnusson PK, Pedersen NL, *et al*. Genetic determinants of mortality. Can findings from genome-wide association studies explain variation in human mortality? *Hum Genet* 2013; **132**: 553-61.
16. Beekman M, Nederstigt C, Suchiman HE, Kremer D, van der Breggen R, Lakenberg N, *et al*. Genome-wide association study (GWAS)-identified disease risk alleles do not compromise human longevity. *Proc Natl Acad Sci U S A* 2010; **107**: 18046-9.
17. Deelen J, Beekman M, Uh HW, Helmer Q, Kuningas M, Christiansen L, *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**: 686-98.
18. Newman AB, Walter S, Lunetta KL, Garcia ME, Slagboom PE, Christensen K, *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**: 478-87.
19. Deelen J, Beekman M, Capri M, Franceschi C, Slagboom PE. Identifying the genomic determinants of aging and longevity in human population studies: progress and challenges. *Bioessays* 2013; **35**: 386-96.
20. Gavrilova NS, Gavrilov LA, Evdokushkina GN, Semyonova VG, Gavrilova AL, Evdokushkina NN, *et al*. Evolution, mutations, and human longevity: European

royal and noble families. *Hum Biol* 1998; **70**: 799-804.
21. Nebel A, Kleindorp R, Caliebe A, Nothnagel M, Blanche H, Junge O, *et al*. A genome-wide association study confirms APOE as the major gene influencing survival in long-lived individuals. *Mech Ageing Dev* 2011; **132**: 324-30.
22. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, *et al*. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 2010; **26**: 2336-7.
23. Schunkert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, *et al*. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* 2011; **43**: 333-8.
24. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, *et al*. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 2011; **478**: 103-9.
25. Wain LV, Verwoert GC, O'Reilly PF, Shi G, Johnson T, Johnson AD, *et al*. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat Genet* 2011; **43**: 1005-11.
26. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, *et al*. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 2010; **42**: 105-16.
27. Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, *et al*. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet* 2012; **44**: 659-69.
28. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, *et al*. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* 2012; **44**: 981-90.
29. Saxena R, Hivert MF, Langenberg C, Tanaka T, Pankow JS, Vollenweider P, *et al*. Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet* 2010; **42**: 142-8.
30. Soranzo N, Sanna S, Wheeler E, Gieger C, Radke D, Dupuis J, *et al*. Common variants at 10 genomic loci influence hemoglobin A(1)(C) levels via glycemic and nonglycemic pathways. *Diabetes* 2010; **59**: 3229-39.
31. Strawbridge RJ, Dupuis J, Prokopenko I, Barker A, Ahlqvist E, Rybin D, *et al*. Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* 2011; **60**: 2624-34.
32. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, *et al*. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010; **466**: 707-13.
33. Segre AV, Groop L, Mootha VK, Daly MJ, Altshuler D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet* 2010; **6**: e1001058.
34. Raychaudhuri S, Plenge RM, Rossin EJ, Ng AC, Purcell SM, Sklar P, *et al*. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet* 2009; **5**: e1000534.
35. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012; **40**: D930-D934.
36. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, *et al*. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012; **22**: 1790-7.
37. Molander L, Lovheim H, Norman T, Nordstrom P, Gustafson Y. Lower systolic blood pressure is associated with greater mortality in people aged 85 and older. *J Am Geriatr Soc* 2008; **56**: 1853-9.
38. Oates DJ, Berlowitz DR, Glickman ME, Silliman RA, Borzecki AM. Blood pressure and survival in the oldest old. *J Am Geriatr Soc* 2007; **55**: 383-8.

39. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet* 2009; **10**: 155-9.
40. Esteller M. Non-coding RNAs in human disease. *Nat Rev Genet* 2011; **12**: 861-74.
41. Flachsbart F, Caliebe A, Kleindorp R, Blanche H, von Eller-Eberstein H, Nikolaus S, *et al*. Association of FOXO3A variation with human longevity confirmed in German centenarians. *Proc Natl Acad Sci U S A* 2009; **106**: 2700-5.
42. Sebastiani P, Solovieff N, Dewan AT, Walsh KM, Puca A, Hartley SW, *et al*. Genetic signatures of exceptional longevity in humans. *PLoS One* 2012; **7**: e29848.
43. Hollingworth P, Harold D, Sims R, Gerrish A, Lambert JC, Carrasquillo MM, *et al*. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* 2011; **43**: 429-35.
44. Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J, *et al*. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* 2011; **43**: 436-41.
45. Zeggini E, Panoutsopoulou K, Southam L, Rayner NW, Day-Williams AG, Lopes MC, *et al*. Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. *Lancet* 2012; **380**: 815-23.
46. Uh HW, Deelen J, Beekman M, Helmer Q, Rivadeneira F, Hottenga JJ, *et al*. How to deal with the early GWAS data when imputing and combining different arrays is necessary. *Eur J Hum Genet* 2012; **20**: 572-6.
47. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* 2008; **32**: 381-5.
48. Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R, *et al*. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One* 2010; **5**: e10693.
49. Yang TP, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, Stranger BE, *et al*. Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics* 2010; **26**: 2474-6.
50. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, *et al*. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 2013; **45**: 1238-43.

Supplementary Information

The supplementary information belonging to this chapter can be found at: http://hmg.oxfordjournals.org/content/early/2014/04/15/hmg.ddu139/suppl/DC1.

Figure S4.1 Quantile-quantile plots. Quantile-quantile plots of the expected versus (unadjusted) observed X^2 values for the discovery phase analysis of cases aged ≥ 85 years (A) and ≥ 90 years (B). The shaded region represents the 95% confidence band.

Figure S4.2 Regional association plot for the chromosome 19q13.32 locus after conditional analysis for rs4420638. Results of the discovery phase analysis of chromosome 19q13.32 after conditional analysis in cases aged ≥ 85 years (generated using LocusZoom (http://csg.sph.umich.edu/locuszoom/) [22]. The color of the SNPs is based on the LD with the lead SNP (shown in purple). The blue peaks represent the recombination rates based on HapMap Phase I + II CEU release 22 (hg18/build36) and the RefSeq genes in the region are shown in the lower panel.

Figure S4.3 Graphical representation of GRAIL connections. Results of the GRAIL analysis using the loci with at least one SNP with a $P \leq 1 \times 10^{-4}$ in the discovery analysis of cases aged ≥ 90 years ($n = 65$). The thickness of the red line represents the strength of the literature-based connection.

Figure S4.4 Quantile-quantile plot of the expected versus (unadjusted) observed X^2 values for the 281 SNPs from the signature of Sebastiani et al. [42] in the analysis of cases aged ≥ 90 years. The shaded region represents the 95% confidence band. The red line indicates the threshold for significance after adjustment for multiple testing ($P < 1.78 \times 10^{-4}$ (0.05/281)).

Table S4.1 Cohort demographics of the studies of European descent included in the discovery and replication phase.

Table S4.2 Details of genotyping, quality control and imputation of the studies of European descent included in the discovery and replication phase.

Table S4.3 Results of the conditional analysis to test for independent signals at chromosome 19q13.32.

Table S4.4 Results of the sex-stratified analysis of the cases ≥ 85 years of age for the 15 loci taken forward to the replication phase.

Table S4.5 Details of the cohorts used for the prospective analysis of rs4420638 and rs2149954.

Table S4.6 Results of the prospective analysis of rs4420638 and rs2149954 (all-cause mortality).

Table S4.7 Results of the prospective analysis of rs2149954 (cause-specific mortality).

Table S4.8 Association of rs2149954 and rs4420638 with CAD and type 2 diabetes.

Table S4.9 Association of rs4420638 and rs2149954 with blood pressure and metabolic traits.

Table S4.10 Results of the prospective analysis of rs2149954 adjusted for blood pressure.

Table S4.11 Results of the GRAIL analysis of SNPs that showed moderate evidence for association ($P \leq 1 \times 10^{-4}$) with survival ≥ 90 years of age in the discovery phase analysis.

Table S4.12 Functional implications of SNPs on chromosome 5q33.3.

Chapter 5

Gene set analysis of GWAS data for human longevity highlights the relevance of the insulin/IGF-1 signaling and telomere maintenance pathways

Joris Deelen, Hae-Won Uh, Ramin Monajemi, Diana van Heemst, Peter E Thijssen, Stefan Böhringer, Erik B van den Akker, Anton J M de Craen, Fernando Rivadeneira, André G Uitterlinden, Rudi G J Westendorp, Jelle J Goeman, P Eline Slagboom, Jeanine J Houwing-Duistermaat and Marian Beekman

Age (Dordr) 2013; **35**: 235-49.

Abstract

In genome-wide association studies (GWAS) for complex traits, single single nucleotide polymorphism (SNP) analysis is still the most applied approach. However, the identified SNPs have small effects and provide limited biological insight. A more appropriate approach to interpret GWAS data of complex traits is to analyze the combined effect of a SNP set grouped per pathway or gene region. We used this approach to study the joint effect on human longevity of genetic variation in two candidate pathways, the insulin/insulin-like growth factor 1 (IGF-1) signaling (IIS) pathway and the telomere maintenance (TM) pathway. For the analyses we used genotyped GWAS data of 403 unrelated nonagenarians from long-lived sibships collected in the Leiden Longevity Study (LLS) and 1,670 younger population controls. We analyzed 1,021 SNPs in 68 IIS pathway genes and 88 SNPs in 13 TM pathway genes using four self-contained pathway tests (PLINK set-based test, Global test, GRASS and SNP ratio test). Although we observed small differences between the results of the different pathway tests, they showed consistent significant association of the IIS and TM pathway SNP sets with longevity. Analysis of gene SNP sets from these pathways indicates that the association of the IIS pathway is scattered over several genes (*AKT1*, *AKT3*, *FOXO4*, *IGF2*, *INS*, *PIK3CA*, *SGK*, *SGK2*, and *YWHAG*), while the association of the TM pathway seems to be mainly determined by one gene (*POT1*). In conclusion, this study shows that genetic variation in genes involved in the IIS and TM pathways is associated with human longevity.

Introduction

Genome-wide association studies (GWAS) using single single nucleotide polymorphism (SNP) analysis have been very successful in identifying loci for various quantitative traits and diseases [1]. It became apparent that complex traits are usually determined by many genes with small effects and that results from single SNP analysis provide limited biological insight and only partly explain the genotypic variation of the studied trait. Instead of analyzing single SNPs, the combined effect of a SNP set, grouped per pathway or gene region, can be tested for association with the trait of interest. Such SNP set analysis could be used as an alternative approach for GWAS analysis and, since the composition of SNP sets is often based on pathways, should be able to provide additional biological insight into the studied trait.

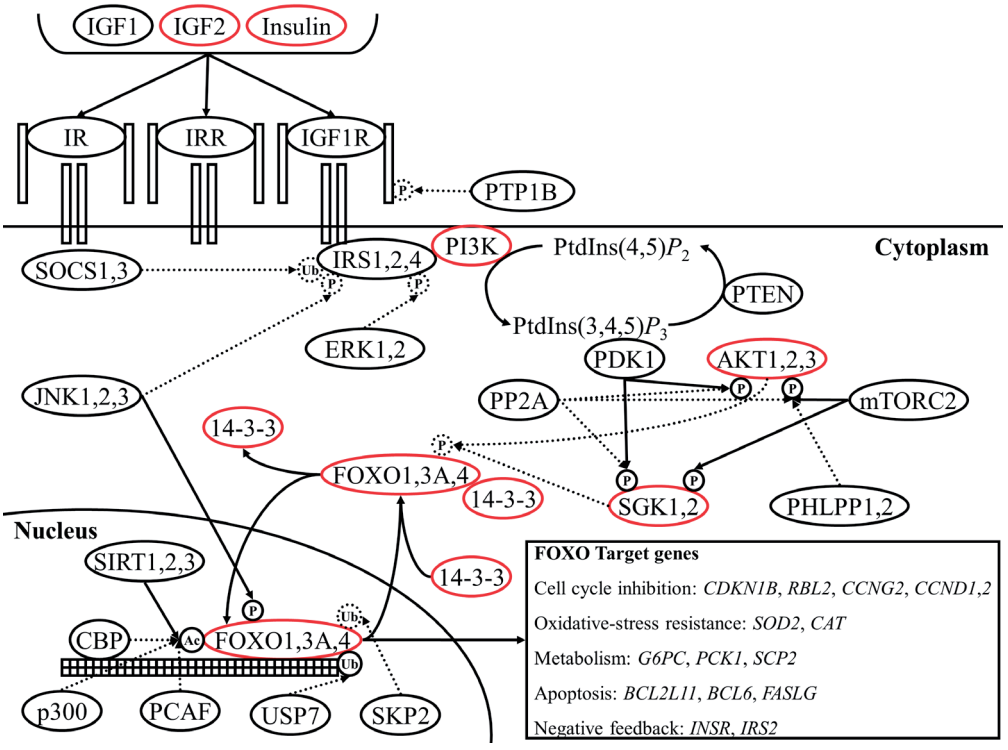
Since the number of tests in SNP set analysis is low compared to single SNP analysis, it requires a lower penalty for multiple testing. Therefore, SNP set analysis is also very suitable in studies with low power for GWAS analysis. The last couple of years, several methods have been developed to perform SNP set analysis on GWAS data [2-4]. There are two main types of methods, the competitive and the self-contained tests. The competitive tests compare the association between a SNP set and trait to a standard defined by the genotyped SNPs outside the SNP set (complement), while the self-contained tests compare the SNP set to a fixed standard that does not depend on the complement [5].

Human longevity is a complex trait that is assumed to be determined by variation in many genes with small effects. Previous GWAS, in which single SNP analyses were performed [6,7], have identified only one genome-wide significant locus contributing to survival into old age; *APOE*. However, the genetic contribution to human lifespan variation, determined in twin studies, is estimated at 25-30% [8-10] and, although the effect of genetic variation in *APOE* is relatively large, the heritability of longevity is only partially explained by this variation [6]. Part of the remaining heritability might be explained by functionally related SNPs with small effects, of which the joint effect could not be detected in a single SNP analysis. Testing of SNP sets of candidate pathways for association with longevity would therefore be valuable.

The insulin/insulin-like growth factor 1 (IGF-1) signaling (IIS) pathway is considered as a candidate pathway for studying human longevity. It is involved in the adaptation of the organism to its (changing) environment [11]. When experimentally induced in model organisms like worms, flies, and mice, mutations in genes that play a role in IIS, e.g., homologues of human *IGF1R*, *INSR*, *IRS1*, *PI3K*, and *FOXO*, were shown to have a considerable effect on lifespan [12-24]. Although the IIS pathway is evolutionarily conserved, the complexity of the human IIS pathway (Figure 5.1) is much larger compared to that of model organisms. Several studies have investigated associations between single SNPs in genes from the IIS pathway and human longevity. The most prominent results came from *FOXO3A* [25-30] and *AKT1* [28], which showed associations with longevity in several independent cohort studies.

Another candidate pathway for studying human longevity is the mechanism of telomere maintenance (TM). Telomeres are structures at the end of chromosomes, consisting of TTAGGG tandem repeats [31], which protect chromosomes from degradation or rearrangement [32]. In normal human cells, telomere length declines with every cell division [33] and when a critical length is reached, the cell will enter replicative senescence [34]. In human epidemiological studies in blood, increased telomere length has been associated with longevity [35], while decreased telomere length has been associated with increased mortality [36-38], although some studies showed contradictory results [39,40]. Telomere integrity is essentially regulated by two protein networks, telomerase and its associated factors, which regulate telomere length, and the shelterin complex, which covers the telomeres [41,42] (Figure 5.2). Several studies have investigated associations between single SNPs in telomerase and shelterin genes and telomere length. The most promising results came from *TERC* and *TERT* [35,43-46], of which the latter has also been associated with human longevity [35].

Figure 5.1 Insulin/insulin-like growth factor 1 (IGF-1) signaling pathway. The insulin/IGF-1 signaling pathway consists of the core components IGF1R/IR/IRR, IRS, PI3K, AKT/SGK, FOXO, and SIRT and proteins that have a direct activating or inhibiting effect on these proteins. The *small closed circles* (containing Ac, P, or Ub) indicate an activating effect of the posttranslational modification on the protein, while the *small dashed circles* indicate an inhibiting effect. The *straight arrows* pointing to these *small circles* indicate an activating effect on the posttranslational modification, while the *dashed arrows* indicate an inhibiting effect. *Ac* acetylation, *P* phosphorylation, *Ub* ubiquitylation.

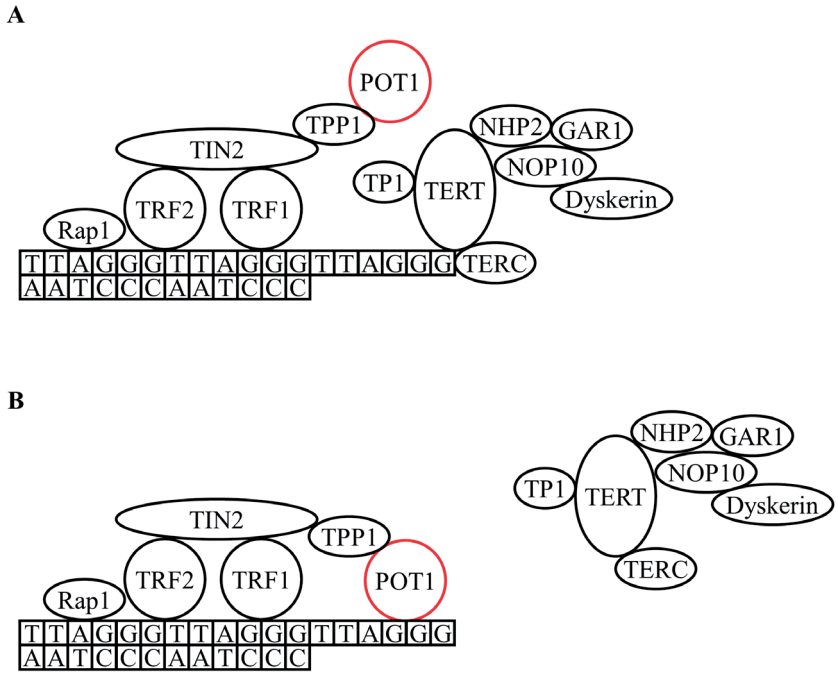


In this study, we used four self-contained tests (PLINK set-based test [47], GRASS [48], Global test [49], and SNP ratio test [50]) and one competitive test (the comparative approach of Global test) to study the joint effect of genetic variation in the IIS and TM pathways on human longevity. For the analyses we used genotyped GWAS data of nonagenarian siblings from the Leiden Longevity Study (LLS) and younger population controls from the Rotterdam Study (RS) [6].

Results

For the IIS pathway, we selected genes encoding proteins that belong to the well-described core of the pathway, consisting of IGF1R/IR/IRR, IRS, PI3K, AKT/SGK, FOXO, and SIRT, or that had a direct activating or inhibiting effect on these core components [51,52]. In addition, we selected several FOXO target genes that play a role in cell-cycle inhibition, oxidative-stress resistance, metabolism, and apoptosis [52] (Figure 5.1). For the TM pathway, we selected genes encoding proteins that were specifically associated with telomeres and belonged to telomerase and its associated factors or to the shelterin complex [42,53,54] (Figure 5.2). We analyzed SNPs within a 10-kb window around the selected genes (based on [28]) from genotyped GWAS data of 403 unrelated nonagenarian participants from the LLS and 1,670 middle-aged controls from the RS [6]. A description of the investigated

Figure 5.2 Telomere maintenance pathway. The telomere maintenance pathway consists of proteins belonging to telomerase and its associated factors or to the shelterin complex. Telomere elongation is performed by telomerase after binding to the telomere (a). However, binding of the shelterin protein POT1 to the telomere blocks this process (b).



samples is given in Table S5.1. In total, 1,021 SNPs in 68 IIS pathway genes and 88 SNPs in 13 TM pathway genes were analyzed (Tables 5.1, 5.2, S5.3A, and S5.3B).

Four methods, PLINK set-based test, Global test, GRASS, and SNP ratio test (Table S5.2), were used to investigate the association of the SNP sets from the IIS and TM pathways with longevity. As a biological negative control, we also analyzed a SNP set of 223 SNPs in 9 genes previously associated with eye and hair color [55] (Tables 5.3 and S5.3C). Both candidate pathways were consistently associated with longevity across all four tests (Table 5.4). We applied Bonferroni correction to adjust for the number of tested pathways (i.e., 2, so for significance $P < 0.025$). After Bonferroni correction, the IIS pathway SNP set remained significant in GRASS and Global test, while the TM pathway SNP set remained significant in the PLINK set-based test, GRASS, and Global test. Using the comparative approach in Global test as a competitive test, we also showed that the probability to find a random SNP set with the same number of genes as the IIS or TM pathway and a comparable or lower *P*-value is less than 5% (2.11% for the IIS and 2.95% for the TM pathway).

To determine which genes are mainly responsible for the observed association of the pathway SNP sets from the IIS and TM pathways with longevity, we also investigated the association of gene SNP sets from these pathways. Although the power to detect an association using gene SNP set analysis is lower than for pathway SNP set analysis, due to the larger number of tests, it provides a ranking of genes based on the contribution to the observed associations of the pathways. To analyze the gene SNP sets, we used the PLINK set-based test, Global test, and SNP ratio test. GRASS was not used, since the underlying statistical method of this test is less suitable for analysis of gene SNP sets. Nine of the 68 IIS pathway gene SNP sets (*AKT1*, *AKT3*, *FOXO4*, *IGF2*, *INS*, *PIK3CA*, *SGK1*, *SGK2*, and *YWHAG*) and 1 of the 13 TM pathway gene SNP sets (*POT1*) showed an association ($P < 0.05$) with longevity in at least two tests (Tables 5.5 and 5.6).

Discussion

To study the effect of the IIS and TM pathways on longevity, SNP set analysis on GWAS data of 403 nonagenarian cases and 1,670 population controls was performed. Both pathway SNP sets associated significantly with longevity. The gene SNP sets analysis showed that the association of the IIS pathway was scattered over several genes (*AKT1*, *AKT3*, *FOXO4*, *IGF2*, *INS*, *PIK3CA*, *SGK1, SGK2* and *YWHAG*), while the association of the TM pathway seems to be mainly determined by one gene (*POT1*).

The proteins encoded by the IIS gene SNP sets that associate with longevity are involved in several parts of the IIS pathway (Figure 5.1). Akt1, Akt3, Foxo4, Igf2, Ins2, Pik3ca, and Sgk1 knockout mice all show abnormalities in growth and/or increased mortality (www.informatics.jax.org) [56], which indicates that these genes are indeed responsible for the growth and lifespan regulating effects of the IIS pathway. Previously, SNPs in several of the significant IIS pathway genes (*AKT1*, *FOXO4*, *INS*, and *PIK3CA*) were studied by single SNP analysis

Table 5.1 Characteristics of the insulin/insulin-like growth factor 1 signaling pathway proteins.

Protein	Gene	ID	Chr	Start (bp)	End (bp)	Size (kb)	SNPs	Coverage
AKT1	*AKT1*	207	14	104,306,732	104,333,125	26.39	2	25.00%
AKT2	*AKT2*	208	19	45,428,064	45,483,142	55.08	6	45.45%
AKT3	*AKT3*	10000	1	241,718,158	242,073,509	355.35	25	50.00%
BIM	*BCL2L11*	10018	2	111,594,962	111,642,493	47.53	12	47.06%
BCL-6	*BCL6*	604	3	188,921,859	188,946,207	24.35	6	23.81%
CAT	*CAT*	847	11	34,417,048	34,450,183	33.14	18	80.00%
Cyclin D1	*CCND1*	595	11	69,165,054	69,178,423	13.37	3	27.27%
Cyclin D2	*CCND2*	894	12	4,253,163	4,284,783	31.62	20	45.00%
Cyclin G2	*CCNG2*	901	4	78,297,381	78,310,237	12.86	4	33.33%
p27kip	*CDKN1B*	1027	12	12,761,569	12,766,572	5.00	9	75.00%
CBP	*CREBBP*	1387	16	3,715,057	3,870,122	155.07	15	50.00%
Deptor (mTORC2)	*DEPDC6*	64798	8	120,955,081	121,132,338	177.26	47	61.90%
p300	*EP300*	2033	22	39,818,560	39,906,027	87.47	6	50.00%
Fas ligand	*FASLG*	356	1	170,894,808	170,902,635	7.83	7	45.45%
FOXO1	*FOXO1*	2308	13	40,027,801	40,138,734	110.93	19	65.38%
FOXO3a	*FOXO3A*	2309	6	108,987,719	109,112,664	124.95	21	68.75%
FOXO4	*FOXO4*	4303	X	70,232,724	70,240,109	7.39	3	NA
G6P	*G6PC*	2538	17	38,306,340	38,319,976	13.63	5	83.33%
IGF-1	*IGF1*	3479	12	101,313,775	101,398,508	84.73	20	47.06%
IGF1R	*IGF1R*	3480	15	97,010,284	97,325,282	315.00	102	56.34%
IGF2	*IGF2*	3481	11	2,106,923	2,127,409	20.49	7	63.64%
Insulin	*INS*	3630	11	2,137,585	2,139,015	1.43	4	80.00%
IR	*INSR*	3643	19	7,063,266	7,245,011	181.75	52	50.00%
IRR	*INSRR*	3645	1	155,077,289	155,095,336	18.05	6	37.50%
IRS1	*IRS1*	3667	2	227,304,277	227,371,750	67.47	11	53.33%
IRS2	*IRS2*	8660	13	109,204,185	109,236,915	32.73	15	59.09%
IRS4	*IRS4*	8471	X	107,862,383	107,866,263	3.88	2	NA
PCAF	*KAT2B*	8850	3	20,056,528	20,170,900	114.37	44	62.96%
ERK2	*MAPK1*	5594	22	20,443,947	20,551,970	108.02	12	58.33%
ERK1	*MAPK3*	5595	16	30,032,927	30,042,131	9.20	2	33.33%
JNK1	*MAPK8*	5599	10	49,279,661	49,317,408	37.75	6	35.71%
JNK2	*MAPK9*	5601	5	179,593,201	179,651,677	58.48	17	40.00%
JNK3	*MAPK10*	5602	4	87,155,300	87,593,307	438.01	71	55.81%
mSIN1 (mTORC2)	*MAPKAP1*	79109	9	127,239,494	127,509,334	269.84	26	64.00%
mLST8 (mTORC2)	*MLST8*	64223	16	2,195,179	2,199,419	3.97	4	80.00%
mTOR (mTORC2)	*MTOR*	2475	1	11,089,175	11,245,195	156.02	11	50.00%

Protein	Gene	ID	Chr	Start (bp)	End (bp)	Size (kb)	SNPs	Coverage
PEPCK	*PCK1*	5105	20	55,569,542	55,574,919	5.38	10	33.33%
PDK1	*PDPK1*	5170	16	2,527,966	2,593,192	65.23	0	0.00%
PHLPP1	*PHLPP1*	23239	18	58,533,652	58,798,656	265.00	41	64.29%
PHLPP2	*PHLPP2*	23035	16	70,236,353	70,306,205	69.85	4	15.38%
PI3K	*PIK3CA*	5290	3	180,349,005	180,435,191	86.19	10	44.44%
	PIK3CB	5291	3	139,854,230	139,960,891	106.66	8	42.86%
	PIK3CD	5293	1	9,634,377	9,711,759	77.38	8	42.11%
	PIK3R1	5295	5	67,547,340	67,633,405	86.07	31	65.12%
	PIK3R2	5296	19	18,124,988	18,142,343	17.36	5	66.67%
	PIK3R3	8503	1	46,278,399	46,371,295	92.90	10	60.00%
PP2A	*PPP2R5B*	5526	11	64,448,719	64,458,526	9.81	3	37.50%
Protor-1 (mTORC2)	*PRR5*	55615	22	43,443,091	43,512,225	69.13	32	50.00%
PTEN	*PTEN*	5728	10	89,613,175	89,718,512	105.34	8	47.06%
PTP1B	*PTPN1*	5770	20	48,560,265	48,634,706	74.44	17	44.44%
p130Rb2	*RBL2*	5934	16	52,025,852	52,083,061	57.21	3	33.33%
RICTOR (mTORC2)	*RICTOR*	253260	5	38,973,780	39,110,258	136.48	9	30.77%
SCP2	*SCP2*	6342	1	53,165,489	53,289,877	124.39	18	50.00%
SGK1	*SGK1*	6446	6	134,532,077	134,680,889	148.81	38	46.75%
SGK2	*SGK2*	10110	20	41,621,049	41,647,687	26.64	9	34.62%
SIRT1	*SIRT1*	23411	10	69,314,433	69,348,153	33.72	4	33.33%
SIRT2	*SIRT2*	22933	19	44,061,035	44,082,342	21.31	7	38.89%
SIRT3	*SIRT3*	23410	11	205,030	226,362	21.33	17	60.00%
SKP2	*SKP2*	6502	5	36,187,902	36,219,899	32.00	15	51.72%
SOCS1	*SOCS1*	8651	16	11,255,775	11,257,540	1.77	4	50.00%
SOCS3	*SOCS3*	9021	17	73,864,453	73,867,755	3.30	4	50.00%
MnSOD	*SOD2*	6648	6	160,020,139	160,034,343	14.20	4	44.44%
USP7	*USP7*	7874	16	8,893,452	8,964,842	71.39	12	42.11%
14-3-3	*YWHAB*	7529	20	42,947,654	42,970,587	22.93	6	50.00%
	YWHAE	7531	17	1,194,584	1,250,306	55.72	16	70.00%
	YWHAG	7532	7	75,794,044	75,826,278	32.23	5	35.71%
	YWHAH	7533	22	30,670,479	30,683,590	13.11	9	43.75%
	YWHAQ	10971	2	9,641,547	9,688,635	47.08	10	58.33%
	YWHAZ	7534	8	101,999,980	102,034,799	34.82	6	40.00%
Total							**1,023**	

ID, Entrez Gene ID; *Chr,* chromosome according to NCBI Build 36; *Start (bp),* start position according to NCBI Build 36; *End (bp),* end position according to NCBI Build 36; *Coverage,* coverage based on HapMap II release 22 CEU; *NA* not available.

Table 5.2 Characteristics of the telomere maintenance pathway proteins.

Protein	Gene	ID	Chr	Start (bp)	End (bp)	Size (kb)	SNPs	Coverage
TPP1 (Shelterin)	*ACD*	65057	16	66,248,916	66,252,219	3.30	2	50.00%
Dyskerin (Telomerase)	*DKC1*	1736	X	153,644,225	153,659,158	14.93	1	NA
GAR1 (Telomerase)	*GAR1*	54433	4	110,956,115	110,965,342	9.23	1	14.29%
NHP2 (Telomerase)	*NHP2*	55651	5	177,509,072	177,513,567	4.50	2	33.33%
NOP10 (Telomerase)	*NOP10*	55505	15	32,421,209	32,422,654	1.45	7	45.45%
POT1 (Shelterin)	*POT1*	25913	7	124,249,676	124,357,273	107.60	25	55.56%
TP1	*TEP1*	7011	14	19,903,666	19,951,419	47.75	21	40.00%
TERC (Telomerase)	*TERC*	7012	3	170,965,092	170,965,542	0.45	1	25.00%
TERF1 (Shelterin)	*TERF1*	7013	8	74,083,651	74,122,541	38.89	10	60.00%
TERF2 (Shelterin)	*TERF2*	7014	16	67,946,965	67,977,392	30.43	6	57.14%
RAP1 (Shelterin)	*TERF2IP*	54386	16	74,239,136	74,248,842	9.71	4	50.00%
TERT (Telomerase)	*TERT*	7015	5	1,306,287	1,348,162	41.88	7	41.18%
TIN2 (Shelterin)	*TINF2*	26277	14	23,778,691	23,781,720	3.03	1	14.29%
Total							**88**	

ID, Entrez Gene ID; *Chr,* chromosome according to NCBI Build 36; *Start (bp),* start position according to NCBI Build 36; *End (bp),* end position according to NCBI Build 36; *Coverage,* coverage based on HapMap II release 22 CEU; *NA,* not available.

Table 5.3 Characteristics of the eye and hair color pathway proteins.

Protein	Gene	ID	Chr	Start (bp)	End (bp)	Size (kb)	SNPs	Coverage
ASIP	*ASIP*	434	20	32,311,832	32,320,809	8.98	5	50.00%
HERC2	*HERC2*	8924	15	26,029,778	26,240,893	211.12	9	41.67%
IRF4	*IRF4*	3662	6	336,739	356,443	19.70	14	65.00%
MC1R	*MC1R*	4157	16	88,511,788	88,514,886	3.10	3	33.33%
OCA2	*OCA2*	4948	15	25,673,616	26,018,053	344.44	82	58.00%
SLC24A4	*SLC24A4*	123041	14	91,858,678	92,037,578	178.90	62	53.68%
SLC45A2	*SLC45A2*	51151	5	33,980,478	34,020,537	40.06	15	44.83%
TYR	*TYR*	7299	11	88,550,688	88,668,575	117.89	22	56.00%
TYRP1	*TYRP1*	7306	9	12,683,386	12,700,266	16.88	11	50.00%
Total							**223**	

ID, Entrez Gene ID; *Chr,* chromosome according to NCBI Build 36; *Start (bp),* start position according to NCBI Build 36; *End (bp),* end position according to NCBI Build 36; *Coverage,* coverage based on HapMap II release 22 CEU.

Table 5.4 Results of gene set analysis of insulin/insulin-like growth factor 1 signaling, telomere maintenance, and eye and hair color pathway single nucleotide polymorphism sets.

Pathway test	Insulin/IGF-1 signaling	Telomere maintenance	Eye and hair color
PLINK set-based test*	0.064	0.019	0.340
GRASS*	0.010	0.023	0.540
Global test*	0.011	0.023	0.362
SNP ratio test*	0.044	0.034	0.337

IGF-1, insulin-like growth factor 1. *Permutation (n = 10,000) P-value.

Table 5.5 Results of gene set analysis of insulin/insulin-like growth factor 1 signaling pathway gene single nucleotide polymorphism sets.

Gene	PLINK set-based test*	Global test*	SNP ratio test*
AKT1	0.003	0.002	0.099
AKT2	0.193	0.461	0.197
AKT3	0.101	0.023	0.043
BCL2L11	1	0.678	1
BCL6	1	0.539	1
CAT	1	0.661	1
CCND1	1	0.471	1
CCND2	0.248	0.073	0.073
CCNG2	1	0.528	1
CDKN1B	1	0.675	1
CREBBP	1	0.495	1
DEPDC6	1	0.378	1
EP300	1	0.823	1
FASLG	1	0.219	1
FOXO1	1	0.688	1
FOXO3A	0.181	0.138	0.180
FOXO4	0.023	0.023	0.055
G6PC	0.156	0.172	0.173
IGF1	0.342	0.042	0.148
IGF1R	0.054	0.373	0.491
IGF2	0.028	0.019	0.084
INS	0.022	0.049	0.188
INSR	0.154	0.217	0.286
INSRR	0.139	0.247	0.224
IRS1	1	0.873	1
IRS2	1	0.569	1
IRS4	1	0.605	1
KAT2B	1	0.905	1
MAPK1	1	0.248	1
MAPK3	1	0.132	1

Gene	PLINK set-based test*	Global test*	SNP ratio test*
MAPK8	0.185	0.531	0.215
MAPK9	1	0.198	1
MAPK10	0.191	0.885	0.068
MAPKAP1	1	0.372	1
MLST8	1	0.593	1
MTOR	1	0.722	1
PCK1	1	0.547	1
PHLPP1	0.113	0.398	0.200
PHLPP2	1	0.364	1
PIK3CA	0.003	9.36×10^{-4}	0.022
PIK3CB	1	0.726	1
PIK3CD	1	0.828	1
PIK3R1	1	0.666	1
PIK3R2	1	0.722	1
PIK3R3	1	0.263	1
PPP2R5B	1	0.363	1
PRR5	0.355	0.163	0.257
PTEN	1	0.855	1
PTPN1	1	0.982	1
RBL2	1	0.061	1
RICTOR	1	0.343	1
SCP2	1	0.729	1
SGK1	0.091	0.007	0.016
SGK2	0.027	0.042	0.349
SIRT1	1	0.941	1
SIRT2	1	0.282	1
SIRT3	0.241	0.232	0.326
SKP2	1	0.898	1
SOCS1	1	0.349	1
SOCS3	1	0.996	1
SOD2	1	0.692	1
USP7	0.025	0.101	0.103
YWHAB	1	0.223	1
YWHAE	0.067	0.124	0.196
YWHAG	0.090	0.032	0.018
YWHAH	1	0.236	1
YWHAQ	0.228	0.175	0.293
YWHAZ	1	0.756	1

*Permutation (n = 10,000) P-value.

Table 5.6 Results of gene set analysis of telomere maintenance pathway gene single nucleotide polymorphism sets.

Gene	PLINK set-based test*	Global test*	SNP ratio test*
ACD	1	0.491	1
DKC1	1	0.642	1
GAR1	1	0.281	1
NHP2	1	0.759	1
NOP10	1	0.208	1
POT1	0.007	0.014	0.019
TEP1	1	0.525	1
TERC	1	0.202	1
TERF1	1	0.821	1
TERF2	0.018	0.160	0.164
TERF2IP	1	0.825	1
TERT	1	0.471	1
TINF2	1	0.587	1

*Permutation (n = 10,000) P-value.

and only one SNP, rs3803304 in *AKT1*, which was not measured in our study, showed an association with longevity [28]. However, gene set analysis, which could have detected association of additional genes containing SNPs with many small effects, was not applied in that study. Most signaling cascades require cooperation of several genes in multiple branches of the cascade. This indicates that, for signaling pathways, mutations in different genes could result in similar downstream effects, which would explain the scattered association in the IIS pathway.

Although SNPs in *FOXO3A* have previously been associated with longevity in several independent studies [25-30], the gene SNP set showed no effect in our study in the PLINK set-based test, Global test, and SNP ratio test (P = 0.181, P = 0.138, and P = 0.180, respectively, Table 5.5). This might be due to the fact that the effects of *FOXO3A* on longevity are most prominent in centenarians. As was previously reported by Flachsbart *et al*., centenarians represent a highly selected phenotype, even among nonagenarians [26]. In addition, the genetic contribution to longevity in general is increased at higher ages [9] and the small effects of longevity-promoting gene variants, relative to other factors, may be larger in centenarians [57] and not detectable in nonagenarians. The cases in our study, which are from long-lived families, have a mean age of 94 years. Yet, we had only 11 individuals > 100 years, which may explain the absence of significance of the *FOXO3A* association in our population.

POT1 is part of the shelterin complex and is responsible for the binding of this complex to the TTAGGG repeats of telomeres. Binding of POT1 to the telomere leads to decreased elongation by telomerase [42]. Reduction of POT1 in human fibroblasts by RNAi leads to induction of apoptosis, chromosomal instability, and senescence

[58]. The same effects are observed in Pot1b knockout mice [59,60]. In addition, telomerase-deficient Pot1b knockout mice show a reduction in lifespan compared to "normal" telomerase-deficient mice [60], which stresses the importance of TM in lifespan regulation. Most protein complexes contain one or several proteins essential for specific functions of the complex, e.g., binding, transport, or activation/repression activity. This indicates that, for pathways containing a protein complex, mutations in a single gene, encoding such an essential protein, could be sufficient to alter the function of the complex, which would explain the single-gene association in the TM pathway.

There are two main kinds of pathway analyses, explorative and candidate-based. Since we want to focus on two pathways, the IIS and TM pathways, we performed candidate-based pathway analysis. The advantage of testing candidate pathways instead of explorative analysis is the decreased penalty for multiple testing, due to the limited number of tests performed. For information about pathways, several databases are available, e.g., Gene Ontology [61] and Kyoto Encyclopedia of Genes and Genomes (KEGG) [62], which are particularly useful for explorative studies [4]. However, to our knowledge, the IIS and TM pathways are not described in sufficient detail in these databases and we therefore assembled these pathways based on literature. Although the IIS pathway is available in KEGG (hsa04910; insulin-signaling pathway), only four of the nine IIS pathway genes that were associated with longevity, *AKT1*, *AKT3*, *INS*, and *PIK3CA*, were part of this pathway, which indicates that the pathway definition used in this study could have had a large influence on the results of the analysis.

Different pathway tests could show contradictory results, even when analyzing the same GWAS data [4]. These discrepancies are caused by differences in, for example, the underlying statistical methods of the tests. Therefore, we used several pathway tests in parallel for our analysis. Some of the available pathway tests require SNP *P*-values as input data, while others require raw genotypes [4]. Given that we have GWAS data available, we selected pathway tests that make use of raw genotypes. All four selected pathway tests are self-contained tests that deal with the complexity of SNP set analysis by permuting the case-control status. While, the PLINK set-based test, Global test, and SNP ratio test do not completely incorporate linkage disequilibrium (LD) information, GRASS employs principal component analysis to deal with correlations within each gene. A simulation study showed that in general, GRASS was more powerful than the PLINK set-based test [48]. Simulation studies for Global test or SNP ratio test are not yet available. However, despite the differences between the methods, they all showed similar results for the IIS and TM pathways in this study.

SNP set analysis could have power to detect significant association, even if the power to detect associations in single SNP analysis is low [2], as was previously shown in the Welcome Trust Case Control Consortium [63]. Our study has a power < 1% to detect single SNP associations of the tested SNPs with an OR of 1.2 and a minor

allele frequency of 0.25 (the mean frequency of the tested SNPs). However, because the small (non-significant) effects of the SNPs are jointly tested, the pathway SNP set analysis is able to detect a significant association of the IIS and TM pathway. This indicates that SNP set analysis could be a useful approach for studies which showed no significant associations in single SNP analysis.

There is still much debate about the optimal size of the window used in SNP set analysis [2-4] and we therefore choose a fixed window of 10 kb to take into account effects of SNPs in regulatory regions surrounding the genes. The same window was also used in a previous study of the IIS pathway [28]. Although there is a chance that we will miss some functional SNPs, increasing the window would increase the chance that SNPs are included with no functional relationship to the tested gene.

The number and diversity of SNPs measured per gene/pathway is highly variable between genotyping platforms used for GWAS. In addition, there is a large variety in allele frequencies and presence of SNPs between populations. For single SNP analysis, one is dependent on association of the same SNP (or a SNP in high LD) for replication. However, when, due to varying allele frequencies, different SNPs associate in different populations, SNP set analysis determines the combined effect of SNPs within a gene and is able to overcome this problem. Therefore, replication of SNP set analysis is assumed to be more reproducible between genotyping platforms and populations [4,64]. To support these assumptions, our findings should be replicated in other cohorts.

In conclusion, we have shown that genetic variation in genes involved in the IIS and TM pathways is associated with human longevity. In addition, we provide evidence that different self-contained tests show similar results when applied to candidate-based pathway analysis.

Materials and methods

Study Populations

Leiden Longevity Study

For the LLS, long-lived siblings of European 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 men and 91 years or older for women, representing less than 0.5% of the Dutch population in 2001 [65]. 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). DNA from the LLS was extracted from samples at baseline using conventional methods [66]. For the GWAS, 403 unrelated LLS siblings (one sibling from each sibling pair) were included (LLS GWAS cases) [6].

Rotterdam Study

The RS is a prospective population-based study of people aged 55 years and older, which was designed to study neurological, cardiovascular, locomotor, and ophthalmological diseases [67]. The study consists of 7,983 participants from the baseline cohort (RS-I) and 3,011 participants from an independent extended cohort formed

in 1999 (RS-II) from which DNA was isolated between 1990 and 1993 (RS-I) or between 2000 and 2001 (RS-II). For the GWAS, 1,731 participants from the combined cohort who were below 60 years of age and for whom GWAS data were available were included as controls (RS GWAS controls) [6].

Population substructure

Multidimensional scaling analysis in PLINK (http://pngu.mgh.harvard.edu/purcell/plink) [47] showed that there was no substructure in the GWAS data to an extent that would affect the observations [6].

Genotyping and SNP selection

For the SNP set analyses we used the genotype data from the GWAS described by Deelen *et al*. [6]. The LLS GWAS cases were genotyped using Illumina Infinium HD Human660W-Quad BeadChips (Illumina, San Diego, CA, USA). The RS GWAS controls were genotyped using Illumina Infinium II Human-Hap 550K Beadchips and Illumina Infinium II HumanHap550-Duo BeadChips (Illumina), respectively [67]. Of the 551,606 SNPs measured in both the LLS GWAS cases and RS GWAS controls, 516,712 SNPs passed quality control using the following criteria: SNP call rate ≥ 0.95 or minor allele frequency ≥ 0.01 in RS GWAS controls and LLS GWAS cases, $P_{HWE} \geq 10^{-4}$ and no between-chip effect in the RS GWAS controls, and good cluster plots in the LLS GWAS cases and RS GWAS controls if $P < 1 \times 10^{-4}$ [6].

We analyzed SNPs within a 10-kb window around genes encoding proteins that belonged to the IIS (Figure 5.1) and TM pathway (Figure 5.2). A gene was defined as an NCBI Entrez Gene (mRNA or RNA) cluster, corresponding to a set of transcripts (RefSeq) for which the alignments can be obtained from the UCSC genome browser (http://genome.ucsc.edu/), in which all transcripts within a cluster agree on strand and overlap. Due to an overlap of the 10-kb windows around *IGF2* and *INS*, two SNPs, rs4320932 and rs7924316, were assigned to both genes.

Statistical analysis

PLINK set-based test

In the PLINK set-based test (--set-test, http://pngu.mgh.harvard.edu/purcell/plink) [47], a single SNP analysis (in our case, a trend test) of the original pathway or gene SNP set is performed. For each SNP set, a mean SNP statistic is calculated from the single SNP statistics of a maximum number (--setmax) of independent SNPs below a certain *P*-value threshold (--set-p). If SNPs are not independent, i.e., in case LD (r^2) is above a certain threshold (--set-r2), the SNP with the lowest *P*-value in the single SNP analysis is selected. The same analysis is performed with a certain number (--mperm) of simulated SNP sets in which the phenotype status of the individuals is permuted. An empirical *P*-value for the SNP set is computed by calculating the number of times the test statistic of the simulated SNP sets exceeds that of the original SNP set. For the analysis in this study, the parameters were set to --setp 0.05 --set-r2 0.5, --set-max 99,999, and --mperm 10,000.

GRASS

GRASS (http://linchen.fhcrc.org/grass.html) [48] calculates "eigenSNPs" for each gene in the pathway SNP set by summarizing the

variation of a gene using principal component analysis. Subsequently, one or more of these "eigenSNPs" per gene are selected using regularized logistic regression to calculate a test statistic for each pathway SNP set. The same analysis is performed with simulated SNP sets in which the phenotype status of the individuals is permuted. The *P*-value per pathway SNP set is calculated by comparing the test statistic of the original pathway SNP set with that of the combined simulated pathway SNP sets. For the analysis in this study, the number of simulated pathway SNP sets was 10,000.

Global test

In this study, we used a modified version of the Global test (http://www.bioconductor.org/help/bioc-views/release/bioc/html/globaltest.html) [49], which is capable and powerful for analyzing GWAS data [68,69]. This test is based on a multiple logistic regression model that uses the phenotype as the response variable and the SNPs in the SNP set as covariates and automatically takes the correlations between SNPs into account. The null hypothesis is tested that none of the SNPs in the SNP set are associated with the phenotype. *P*-values are calculated using a permutation test based on 10,000 permutations. For the comparative approach, 10,000 random SNP sets per pathway SNP set were generated and tested to determine the chance to find a similarsized SNP set with a comparable or lower *P*-value as compared to the original pathway SNP set.

SNP ratio test

The SNP ratio test (http://sourceforge.net/projects/snpratiotest/) [50] performs a single SNP analysis (in our case, a trend test) of the original pathway or gene SNP set and of similar-sized SNP sets in which the phenotype status of the individuals is permuted. An empirical *P*-value of the SNP set is computed by calculating the ratio between the proportion of SNPs that shows an association below a certain *P*-value threshold (*p*) in the original GWAS dataset and in the simulated GWAS datasets. The number of significant SNPs in the simulated GWAS datasets is defined as the top *n* SNPs with the lowest *P*-values, where *n* is the number of SNPs with an association below *p* in the original GWAS dataset. For the analysis in this study, we made use of the scripts described in "SRT_documentation_090310.pdf" (http://sourceforge.net/projects/snpratiotest/). For the analysis in this study, *p* was set to 0.05 and the number of simulated datasets used was 10,000.

Statistical significance

To adjust for multiple testing, the significance level was set at the Bonferroni-corrected nominal *P*-value (which is 0.05/(number of pathway or gene SNP sets tested)).

Acknowledgments

We thank all participants of the Leiden Longevity Study and Rotterdam Study. The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2011) under grant agreement nº 259679. This study was supported by a grant from the Innovation-Oriented Research Program on Genomics (SenterNovem IGE05007), the

Centre for Medical Systems Biology, and the Netherlands Consortium for Healthy Ageing (grant 050-060-810), all in the framework of the Netherlands Genomics Initiative, Netherlands Organization for Scientific Research (NWO), and by Unilever Colworth. The generation and management of GWAS genotype data for the Rotterdam study are supported by the Netherlands Organisation for Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012). This study is funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2) and the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) project nr. 050-060-810. The Rotterdam Study is funded by the Erasmus Medical Center and Erasmus University, Rotterdam; the 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.

References

1. Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. *J Clin Invest* 2008; **118**: 1590-605.
2. Fridley BL, Biernacka JM. Gene set analysis of SNP data: benefits, challenges, and future directions. *Eur J Hum Genet* 2011; **19**: 837-43.
3. Holmans P. Statistical methods for pathway analysis of genome-wide data for association with complex genetic traits. *Adv Genet* 2010; **72**: 141-79.
4. Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. *Nat Rev Genet* 2010; **11**: 843-54.
5. Goeman JJ, Buhlmann P. Analyzing gene expression data in terms of gene sets: methodological issues. *Bioinformatics* 2007; **23**: 980-7.
6. Deelen J, Beekman M, Uh HW, Helmer Q, Kuningas M, Christiansen L, *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**: 686-98.
7. Newman AB, Walter S, Lunetta KL, Garcia ME, Slagboom PE, Christensen K, *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**: 478-87.
8. Gudmundsson H, Gudbjartsson DF, Frigge M, Gulcher JR, Stefansson K. Inheritance of human longevity in Iceland. *Eur J Hum Genet* 2000; **8**: 743-9.
9. Hjelmborg JV, Iachine I, Skytthe A, Vaupel JW, McGue M, Koskenvuo M, *et al*. Genetic influence on human lifespan and longevity. *Hum Genet* 2006; **119**: 312-21.
10. Skytthe A, Pedersen NL, Kaprio J, Stazi MA, Hjelmborg JV, Iachine I, *et al*. Longevity studies in GenomEUtwin. *Twin Res* 2003; **6**: 448-54.
11. Tatar M, Bartke A, Antebi A. The endocrine regulation of aging by insulin-like signals. *Science* 2003; **299**: 1346-51.
12. Bluher M, Kahn BB, Kahn CR. Extended longevity in mice lacking the insulin receptor in adipose tissue. *Science* 2003; **299**: 572-4.
13. Clancy DJ, Gems D, Harshman LG, Oldham S, Stocker H, Hafen E, *et al*. Extension of life-span by loss of CHICO, a Drosophila insulin receptor substrate protein. *Science* 2001; **292**: 104-6.
14. Friedman DB, Johnson TE. A mutation in the age-1 gene in Caenorhabditis elegans lengthens life and reduces hermaphrodite fertility. *Genetics* 1988; **118**: 75-86.
15. Giannakou ME, Goss M, Junger MA, Hafen E, Leevers SJ, Partridge L. Long-lived Drosophila with overexpressed dFOXO in adult fat body. *Science* 2004; **305**: 361.
16. Holzenberger M, Dupont J, Ducos B, Leneuve P, Geloen A, Even PC, *et al*. IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. *Nature* 2003; **421**: 182-7.
17. Hwangbo DS, Gershman B, Tu MP, Palmer M, Tatar M. Drosophila dFOXO controls lifespan and regulates insulin signalling in brain and fat body. *Nature* 2004; **429**: 562-6.
18. Kenyon C, Chang J, Gensch E, Rudner A, Tabtiang R. A C. elegans mutant that lives twice as long as wild type. *Nature* 1993; **366**: 461-4.
19. Kimura KD, Tissenbaum HA, Liu Y, Ruvkun G. daf-2, an insulin receptor-like gene that regulates longevity and diapause in Caenorhabditis elegans. *Science* 1997; **277**: 942-6.
20. Lin K, Dorman JB, Rodan A, Kenyon C. daf-16: An HNF-3/forkhead family member that can function to double the life-span of Caenorhabditis elegans. *Science* 1997; **278**: 1319-22.
21. Morris JZ, Tissenbaum HA, Ruvkun G. A phosphatidylinositol-3-OH kinase family member regulating longevity and diapause in Caenorhabditis elegans. *Nature* 1996; **382**: 536-9.
22. Ogg S, Paradis S, Gottlieb S, Patterson GI, Lee L, Tissenbaum HA, *et al*. The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in C. elegans. *Nature* 1997; **389**: 994-9.

23. Selman C, Partridge L, Withers DJ. Replication of extended lifespan phenotype in mice with deletion of insulin receptor substrate 1. *PLoS One* 2011; **6**: e16144.
24. Tatar M, Kopelman A, Epstein D, Tu MP, Yin CM, Garofalo RS. A mutant Drosophila insulin receptor homolog that extends life-span and impairs neuroendocrine function. *Science* 2001; **292**: 107-10.
25. Anselmi CV, Malovini A, Roncarati R, Novelli V, Villa F, Condorelli G, *et al*. Association of the FOXO3A locus with extreme longevity in a southern Italian centenarian study. *Rejuvenation Res* 2009; **12**: 95-104.
26. Flachsbart F, Caliebe A, Kleindorp R, Blanche H, von Eller-Eberstein H, Nikolaus S, *et al*. Association of FOXO3A variation with human longevity confirmed in German centenarians. *Proc Natl Acad Sci U S A* 2009; **106**: 2700-5.
27. Li Y, Wang WJ, Cao H, Lu J, Wu C, Hu FY, *et al*. Genetic association of FOXO1A and FOXO3A with longevity trait in Han Chinese populations. *Hum Mol Genet* 2009; **18**: 4897-904.
28. Pawlikowska L, Hu D, Huntsman S, Sung A, Chu C, Chen J, *et al*. Association of common genetic variation in the insulin/IGF1 signaling pathway with human longevity. *Aging Cell* 2009; **8**: 460-72.
29. Soerensen M, Dato S, Christensen K, McGue M, Stevnsner T, Bohr VA, *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**: 1010-7.
30. Willcox BJ, Donlon TA, He Q, Chen R, Grove JS, Yano K, *et al*. FOXO3A genotype is strongly associated with human longevity. *Proc Natl Acad Sci U S A* 2008; **105**: 13987-92.
31. Moyzis RK, Buckingham JM, Cram LS, Dani M, Deaven LL, Jones MD, *et al*. A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. *Proc Natl Acad Sci U S A* 1988; **85**: 6622-6.
32. Blackburn EH. Structure and function of telomeres. *Nature* 1991; **350**: 569-73.
33. Harley CB, Futcher AB, Greider CW. Telomeres shorten during ageing of human fibroblasts. *Nature* 1990; **345**: 458-60.
34. Allsopp RC. Models of initiation of replicative senescence by loss of telomeric DNA. *Exp Gerontol* 1996; **31**: 235-43.
35. Atzmon G, Cho M, Cawthon RM, Budagov T, Katz M, Yang X, *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**: 1710-7.
36. Bakaysa SL, Mucci LA, Slagboom PE, Boomsma DI, McClearn GE, Johansson B, *et al*. Telomere length predicts survival independent of genetic influences. *Aging Cell* 2007; **6**: 769-74.
37. Cawthon RM, Smith KR, O'Brien E, Sivatchenko A, Kerber RA. Association between telomere length in blood and mortality in people aged 60 years or older. *Lancet* 2003; **361**: 393-5.
38. Kimura M, Hjelmborg JV, Gardner JP, Bathum L, Brimacombe M, Lu X, *et al*. Telomere length and mortality: a study of leukocytes in elderly Danish twins. *Am J Epidemiol* 2008; **167**: 799-806.
39. Bischoff C, Petersen HC, Graakjaer J, Andersen-Ranberg K, Vaupel JW, Bohr VA, *et al*. No association between telomere length and survival among the elderly and oldest old. *Epidemiology* 2006; **17**: 190-4.
40. Martin-Ruiz CM, Gussekloo J, van Heemst D, von Zglinicki T, Westendorp RG. 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**: 287-90.
41. Collins K, Mitchell JR. Telomerase in the human organism. *Oncogene* 2002; **21**: 564-79.
42. de Lange T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev* 2005; **19**: 2100-10.
43. Codd V, Mangino M, van der Harst P, Braund PS, Kaiser M, Beveridge AJ, *et al*. Common variants near TERC are associated with mean telomere length. *Nat Genet* 2010; **42**: 197-9.

44. Levy D, Neuhausen SL, Hunt SC, Kimura M, Hwang SJ, Chen W, *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**: 9293-8.
45. Mirabello L, Yu K, Kraft P, De Vivo I, Hunter DJ, Prescott J, *et al*. The association of telomere length and genetic variation in telomere biology genes. *Hum Mutat* 2010; **31**: 1050-8.
46. Rafnar T, Sulem P, Stacey SN, Geller F, Gudmundsson J, Sigurdsson A, *et al*. Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nat Genet* 2009; **41**: 221-7.
47. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559-75.
48. Chen LS, Hutter CM, Potter JD, Liu Y, Prentice RL, Peters U, *et al*. Insights into colon cancer etiology via a regularized approach to gene set analysis of GWAS data. *Am J Hum Genet* 2010; **86**: 860-71.
49. Goeman JJ, van de Geer SA, de Kort F, van Houwelingen HC. A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics* 2004; **20**: 93-9.
50. O'Dushlaine C, Kenny E, Heron EA, Segurado R, Gill M, Morris DW, *et al*. The SNP ratio test: pathway analysis of genome-wide association datasets. *Bioinformatics* 2009; **25**: 2762-3.
51. Taniguchi CM, Emanuelli B, Kahn CR. Critical nodes in signalling pathways: insights into insulin action. *Nat Rev Mol Cell Biol* 2006; **7**: 85-96.
52. van der Horst A, Burgering BM. Stressing the role of FoxO proteins in lifespan and disease. *Nat Rev Mol Cell Biol* 2007; **8**: 440-50.
53. Harrington L, McPhail T, Mar V, Zhou W, Oulton R, Bass MB, *et al*. A mammalian telomerase-associated protein. *Science* 1997; **275**: 973-7.
54. Vulliamy T, Beswick R, Kirwan M, Marrone A, Digweed M, Walne A, *et al*. Mutations in the telomerase component NHP2 cause the premature ageing syndrome dyskeratosis congenita. *Proc Natl Acad Sci U S A* 2008; **105**: 8073-8.
55. Eriksson N, Macpherson JM, Tung JY, Hon LS, Naughton B, Saxonov S, *et al*. Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet* 2010; **6**: e1000993.
56. Blake JA, Bult CJ, Kadin JA, Richardson JE, Eppig JT. The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics. *Nucleic Acids Res* 2011; **39**: D842-D848.
57. Perls T, Kunkel LM, Puca AA. The genetics of exceptional human longevity. *J Mol Neurosci* 2002; **19**: 233-8.
58. Yang Q, Zheng YL, Harris CC. POT1 and TRF2 cooperate to maintain telomeric integrity. *Mol Cell Biol* 2005; **25**: 1070-80.
59. He H, Wang Y, Guo X, Ramchandani S, Ma J, Shen MF, *et al*. Pot1b deletion and telomerase haploinsufficiency in mice initiate an ATR-dependent DNA damage response and elicit phenotypes resembling dyskeratosis congenita. *Mol Cell Biol* 2009; **29**: 229-40.
60. Hockemeyer D, Palm W, Wang RC, Couto SS, de Lange T. Engineered telomere degradation models dyskeratosis congenita. *Genes Dev* 2008; **22**: 1773-85.
61. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, *et al*. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**: 25-9.
62. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000; **28**: 27-30.
63. Torkamani A, Topol EJ, Schork NJ. Pathway analysis of seven common diseases assessed by genome-wide association. *Genomics* 2008; **92**: 265-72.
64. Luo L, Peng G, Zhu Y, Dong H, Amos CI, Xiong M. Genome-wide gene and pathway analysis. *Eur J Hum Genet* 2010; **18**: 1045-53.
65. Schoenmaker M, de Craen AJ, de Meijer PH, Beekman M, Blauw GJ, Slagboom PE, *et al*. Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur J Hum Genet* 2006; **14**: 79-84.

66. Beekman M, Blauw GJ, Houwing-Duistermaat JJ, Brandt BW, Westendorp RG, Slagboom PE. 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**: 355-62.
67. Teichert M, Eijgelsheim M, Rivadeneira F, Uitterlinden AG, van Schaik RH, Hofman A, *et al*. A genome-wide association study of acenocoumarol maintenance dosage. *Hum Mol Genet* 2009; **18**: 3758-68.
68. Chapman J, Whittaker J. Analysis of multiple SNPs in a candidate gene or region. *Genet Epidemiol* 2008; **32**: 560-6.
69. Pan W. Asymptotic tests of association with multiple SNPs in linkage disequilibrium. *Genet Epidemiol* 2009; **33**: 497-507.

Supplementary Information

The supplementary information belonging to this chapter can be found at: http://link.springer.com/article/10.1007%2Fs11357-011-9340-3.

Table S5.1 Characteristics of the genotyped samples used for analysis.

Table S5.2 Characteristics of the pathway tests used for analysis.

Table S5.3A Measured SNPs in the insulin/IGF-1 signaling pathway.

Table S5.3B Measured SNPs in the telomere maintenance pathway.

Table S5.3C Measured SNPs in the eye and hair color pathway.

Chapter 6

Leukocyte telomere length associates with prospective mortality independent of immune-related parameters and known genetic markers

Joris Deelen, Marian Beekman, Veryan Codd, Stella Trompet, Linda Broer, Sara Hägg, Krista Fischer, Peter E Thijssen, H Eka D Suchiman, Iris Postmus, André G Uitterlinden, Albert Hofman, Anton J M de Craen, Andres Metspalu, Nancy L Pedersen, Cornelia M van Duijn, J Wouter Jukema, Jeanine J Houwing-Duistermaat, Nilesh J Samani and P Eline Slagboom

Int J Epidemiol 2014; *In press.*

Abstract

Human leukocyte telomere length (LTL) decreases with age and shorter LTL has previously been associated with increased prospective mortality. However, it is not clear whether LTL merely marks the health status of an individual by its association with parameters of immune function, for example, or whether telomere shortening also contributes causally to lifespan variation in humans. We measured LTL in 870 nonagenarian siblings (mean age 93 years), 1,580 of their offspring, and 725 spouses thereof (mean age 59 years) from the Leiden Longevity Study (LLS). We found that shorter LTL is associated with increased prospective mortality in middle (30-80 years; hazard ratio (HR) = 0.75, $P = 0.001$) and highly advanced age ($\geq$ 90 years; HR = 0.92, $P = 0.028$) and show that this association cannot be explained by the association of LTL with the immune-related markers insulin-like growth factor 1 to insulin-like growth factor binding protein 3 molar ratio, C-reactive protein, interleukin 6, cytomegalovirus serostatus, or white blood cell counts. We found no difference in LTL between the middle-aged LLS offspring and their spouses ($\beta = 0.006$, $P = 0.932$). Neither did we observe an association of LTL-associated genetic variants with mortality in a prospective meta-analysis of multiple cohorts (n = 8,165). We confirm LTL to be a marker of prospective mortality in middle and highly advanced age and additionally show that this association could not be explained by the association of LTL with various immune-related markers. Furthermore, the approaches performed here do not further support the hypothesis that LTL variation contributes to the genetic propensity for longevity.

Introduction

Telomeres are TTAGGG tandem repeat structures at the end of chromosomes that protect chromosomes from degradation and rearrangement [1]. In somatic cells, telomere length declines with every cell division and, accordingly, human leukocyte telomere length (LTL) decreases with age [2,3]. In addition, LTL differs between sexes, women have a longer LTL than men of the same age [4]. Shorter LTL has been associated with an increased risk of several age-related diseases, such as cardiovascular disease [5], hypertension [6], and cancer [7]. Likewise, several studies, although not all, have shown an association of shorter LTL with prospective mortality, mainly through infectious- and cardiovascular-related causes of death [8].

The shortening of telomeric DNA is mainly caused by incomplete DNA replication during the cell cycle S phase [9], but also by oxidative stress [10], which plays a role in the pathogenesis of viral infections [11]. An association of shorter LTL with increased prospective mortality is thus not necessarily explained by a causal effect of LTL on health conditions. The association of LTL with prospective mortality risk might be confounded by immune functions influencing prospective mortality, which may be investigated by immune-related markers. Previous studies showed that shorter LTL is associated with a decrease in serum levels of insulin-like growth factor 1 (IGF-1) [12,13], which is an important regulator of cell replication and, in addition, seems to play a role in the regulation of immunity and inflammation [14]. Furthermore, shorter LTL has been shown to associate with increased levels of the inflammatory markers C-reactive protein (CRP) and interleukin 6 (IL-6), although not in all studied populations [8]. During inflammation, IL-6 and other cytokines are secreted by T cells and macrophages and trigger the synthesis of CRP by the liver, ultimately resulting in clearance of necrotic and apoptotic cells. However, an increased level of IL-6 or CRP is not necessarily the result of increased inflammation [15,16].

Whether LTL is associated with familial longevity in middle age is not extensively studied. One study showed that offspring of Ashkenazi Jewish centenarians (n = 175) have a longer mean LTL as compared with controls from the general population (n = 93) [17]. Since the centenarians in this study (n = 86) and their offspring did not show a decline in LTL with age as observed in controls, the authors suggested that better LTL maintenance may be a feature in long-lived families. These interesting observations in a relatively small study warrant replication in larger populations.

LTL is a highly heritable trait [3,18]. Insights into the causal effects of LTL on human lifespan might be obtained by testing genetic variants influencing LTL for their association with prospective mortality. Recently, nine loci have been identified that influence LTL variation in Western populations [19,20]. These loci include the known telomere biology genes *CTC1*, *NAF1*, *OBFC1, RTEL1, TERC,* and *TERT*, explaining ~1% of the variance in LTL. In addition, genetic variation in two genes involved in telomere maintenance, *TERC* and *POT1*, was found to be associated with human longevity [17,21,22].

In this study, we explored the data of the Leiden Longevity Study (LLS) in which we measured LTL in nonagenarian siblings (n = 870), their offspring (n = 1,580) and the spouses thereof (n = 725; serving as controls). The survival benefit of LLS families is marked by a 30% decreased mortality risk in three generations, i.e., the parents of the nonagenarian siblings, their unselected additional siblings and their offspring, when compared with the general Dutch population [23]. In addition, the LLS offspring, who are considered "decelerated" or "healthy agers", have a lower prevalence of age-related diseases, such as type 2 diabetes, cardiovascular disease, and hypertension, as compared with their spouses, and show beneficial or "youthful" profiles for many metabolic parameters [24,25]. We first investigated whether the association of shorter LTL with increased prospective mortality, which is observed in both generations, could be explained by lymphocyte counts, serum CRP levels, serum IGF-1 to insulin-like growth factor binding protein 3 (IGFBP3) molar ratio (IGF-1/IGFBP3), serum IL-6, or the presence of cytomegalovirus (CMV) infection as immune-related markers. Next, we examined whether the LLS offspring have a longer LTL and a different association of LTL with age as compared with their spouses. Finally, we determined the effect of genetic variants associated with LTL on prospective mortality. We performed a prospective meta-analysis of multiple cohorts (n = 8,165) in which known LTL-associated single nucleotide polymorphisms (SNPs) were investigated separately and in combination as a genetic risk score (GRS).

Results

LTL and prospective mortality in two generations

The characteristics of the LLS nonagenarians, their offspring and the spouses thereof for demographic variables, LTL, immune-related markers, and mortality analysis are depicted in Tables 6.1, 6.2, and 6.3. We first analyzed the association between LTL and prospective mortality in the middle-aged and nonagenarian generations. We found that shorter LTL is associated with increased prospective all-cause mortality in the combined group of middle-aged LLS offspring and their spouses (30-80 years of age), i.e., per unit longer LTL there is a 25% decrease in mortality risk (hazard ratio (HR) = 0.75 (95% CI 0.64 – 0.88), P = 0.001). In addition, we observed a similar association in the LLS nonagenarians (≥ 90 years of age, HR = 0.92 (95% CI 0.86 – 0.99), P = 0.028). Since it has previously been reported that LTL declines with decreasing serum IGF-1 levels and increasing serum CRP levels, the effect of LTL on prospective mortality might be explained by the association of LTL with immune functions, as reflected by immune-related markers such as serum IGF-1/IGFBP3, serum CRP levels, serum IL-6 levels, presence of CMV infection and white blood cell (WBC) counts. We previously showed that long-lived family members from the LLS have a lower prevalence of CMV infection as compared to controls from the general population [26,27]. The levels of the other markers did not differ between long-lived family members and controls [28,29]. Most of these markers associated with LTL in the LLS (Table S6.1). Therefore,

Table 6.1 Characteristics of the Leiden Longevity Study samples used for the linear regression and prospective analysis of leukocyte telomere length.

	Offspring				Spouses				Nonagenarians			
	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)	734	59.33 (6.51)	846	59.47 (6.52)	309	61.23 (7.40)	416	57.02 (6.85)	333	92.24 (2.72)	537	93.98 (2.34)
LTL (T/S ratio)*	734	1.43 (0.25)	846	1.48 (0.26)	309	1.42 (0.25)	416	1.49 (0.28)	333	1.25 (0.21)	537	1.29 (0.22)
IGF-1/IGFBP3 (molar ratio)	709	0.11 (0.02)	814	0.10 (0.02)	302	0.11 (0.02)	400	0.10 (0.02)	320	0.11 (0.02)	506	0.09 (0.02)
CRP (mg/L)	713	2.12 (2.94)	813	2.51 (3.73)	301	2.09 (2.83)	403	2.61 (3.56)	321	6.27 (10.39)	512	5.66 (10.00)
IL-6 (pg/ml)	667	0.60 (0.69)	758	0.54 (0.63)	280	0.57 (0.60)	384	0.57 (0.69)	NA	NA	NA	NA
Lymphocyte count (%)	695	27.74 (6.85)	797	30.07 (7.01)	301	28.39 (7.11)	392	30.37 (6.52)	317	21.11 (7.27)	492	21.98 (7.70)
Neutrophil count (%)	695	60.75 (7.74)	797	59.87 (7.59)	301	60.45 (8.14)	392	59.44 (7.18)	317	66.69 (8.31)	492	66.96 (8.91)
Monocyte count (%)	695	5.61 (1.34)	797	5.06 (1.29)	301	5.61 (1.32)	392	5.04 (1.31)	317	6.34 (1.79)	492	5.82 (1.63)
Eosinophil count (%)	695	2.77 (1.44)	797	2.35 (1.24)	301	2.77 (1.37)	392	2.48 (1.33)	317	3.12 (1.73)	492	2.65 (1.49)
Basophil count (%)	695	0.74 (0.35)	797	0.69 (0.31)	301	0.72 (0.35)	392	0.71 (0.34)	317	0.65 (0.36)	492	0.69 (0.39)

SD, standard deviation; *IGF-1*, insulin-like growth factor 1; *IGFBP3*, insulin-like growth factor binding protein 3; *CRP*, C-reactive protein; *IL-6*, interleukin 6. *A decrease in one T/S unit represents a decrease of 5,000 telomeric base pairs.

Table 6.2 Characteristics of the Leiden Longevity Study samples used for the linear regression and prospective analysis of leukocyte telomere length.

	Offspring				Spouses				Nonagenarians			
	Men		Women		Men		Women		Men		Women	
	n	Yes/no	*n*	Yes/no	*n*	Yes/no	*n*	Yes/no	*n*	Yes/no	*n*	Yes/no
CMV infection	713	276/437	815	414/401	301	155/146	407	212/195	326	178/148	514	322/192

CMV, cytomegalovirus.

we adjusted for IGF-1/IGFBP3, CRP, IL-6, CMV infection, and lymphocyte counts in the Cox proportional hazard model used for prospective analysis of mortality. This showed that the association of LTL with all-cause mortality in both generations of the LLS was independent from these immune-related markers (HR = 0.68 (95% CI 0.56 – 0.82), $P = 9.23 \times 10^{-5}$ (offspring and spouses) and HR = 0.90 (95% CI 0.84 – 0.97), P = 0.006 (nonagenarians), even though all markers showed an association (in one or both generations) with mortality (Table S6.2). Since both a low and/or high level of WBC counts and CRP could be detrimental, we also performed the analyses without individuals with low ($< 4 \times 10^9$, $n_{offspring/spouses} = 36$, $n_{nonagerarians} = 16$) and high ($> 10 \times 10^9$, $n_{offspring/spouses} = 96$, $n_{nonagerarians} = 58$) WBC counts and high CRP levels (> 30 mg/L, $n_{offspring/spouses} = 2$, $n_{nonagerarians} = 24$). However, these analyses provide similar results (data not shown). In addition, there was no interaction between LTL and immune-related markers. This indicates that the effect of LTL on prospective mortality could not be explained by its association with these immune-related markers.

LTL and familial longevity in middle age

Next, we compared LTL between the LLS offspring (n = 1,580) and their spouses (n = 725). We found no evidence for a difference in mean LTL between the groups considering age and gender as covariates in our linear regression model (β = 0.006 (95% CI -0.125 – 0.136), $P = 0.932$, Figure 6.1A). In addition, we found no evidence that the association of LTL with age is different among long-lived families, since the estimated decline of LTL (in T/S ratio units) per calendar year in the LLS offspring (β = -0.009) and their spouses (β = -0.006) was similar to other studies with participants of middle age [20] (Figure 6.1B). This indicates that LTL does not explain the propensity for familial longevity in middle age.

Prospective meta-analysis of LTL-associated genetic variants

To determine whether the genetic component of LTL contributes to prospective mortality, we investigated whether the lead SNPs from the 7 loci that showed association with LTL variation ($P < 5 \times 10^{-8}$) in the largest genome-wide association study (GWAS) up to now [20], as well as a GRS based on these SNPs, also associate with prospective mortality. To this end, we performed a prospective meta-analysis of mortality in 8,165 individuals above 75 years from 6 different cohorts, of whom 3,893 had died (Table S6.3). This analysis showed no association of the LTL SNPs, nor of the GRS, with all-cause, cardiovascular or cancer mortality after Bonferroni correction to adjust for multiple testing ($P_{adjusted} > 0.0056$, Tables 6.4, S6.4, and S6.5), although we had an 80% power (α = 0.05) to detect HR's below 0.91.

Discussion

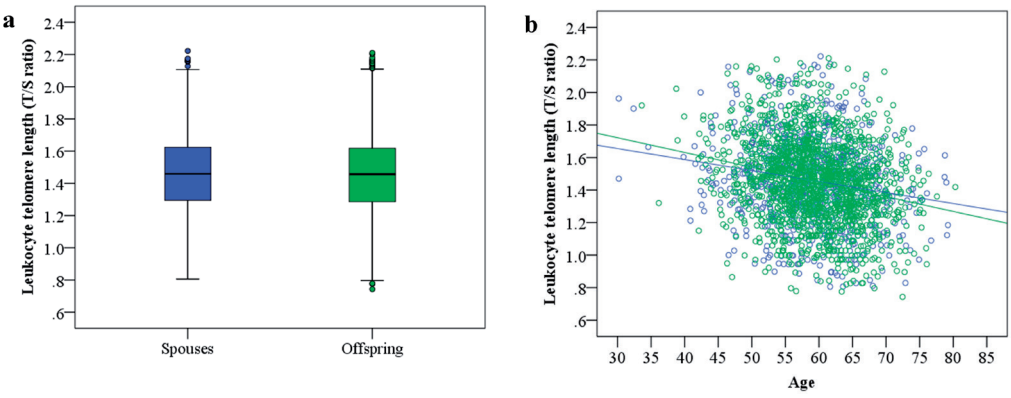
To examine the association between telomeres and human lifespan we studied LTL in nonagenarians and their middle-aged offspring from the LLS for association with prospective mortality and familial longevity in middle age. Interestingly, carriers of long telomeres, as compared to those with shorter

Table 6.3 Characteristics of the Leiden Longevity Study samples used for the prospective analysis of leukocyte telomere length.

	n	n_{deaths}	Mean age (SD)	Age range	Men/women	Mean follow-up time (SD)
LLS offspring + spouses	2,294	106	59.18 (6.78)	30 - 80	1,037/1,257	7.56 (0.95)
LLS nonagenarians	870	751	93.31 (2.63)	89 - 103	333/537	7.57 (0.84)

SD, standard deviation; *LLS*, Leiden Longevity Study.

Figure 6.1 Characteristics of leukocyte telomere length (LTL) in the Leiden Longevity Study (LLS) offspring and their spouses. Mean LTL (A) and the age-related decline of LTL (B) in the LLS offspring (n = 1580, green) and their spouses (n = 725, blue).



telomeres, have a clear survival benefit, which is independent of immune-related markers associated with LTL. We found no association of LTL with familial longevity in middle age. Neither did we observe an association of LTL-associated genetic variants with mortality in a prospective meta-analysis of multiple cohorts (n = 8,165). This confirms the study of monozygotic twins, in which LTL predicts prospective mortality in the absence of genetic differences between the twins [30].

The observed association of shorter LTL with increased prospective mortality in two generations of the LLS is in accordance with previous studies [8]. However, it is unclear what biological phenomenon telomere length in blood reflects. Shorter LTL has previously been associated with decreased serum levels of IGF-1 [12,13] and increased levels of CRP and IL-6 [8], which are known markers of inflammation [31,32]. Hence, the established association between LTL and prospective mortality might be explained by confounding factors such as immune functions. These factors could on the one hand associate with LTL, by affecting replication of specific cell populations, and on the other hand with prospective mortality, reflecting the health status of an individual. To test this hypothesis, we investigated several immune-related markers, namely serum IGF-1/IGFBP3, which is a marker for the amount of biologically active IGF-1, serum CRP levels, serum IL-6 levels, WBC counts, and seropositivity for CMV

Table 6.4 Association between leukocyte telomere length-associated genetic variants and all-cause mortality.

SNP/GRS	*n*	n_{deaths}	HR	SE	95% CI	*P*
rs11125529	8,165	3,893	1.01	0.04	0.94 - 1.08	0.863
rs10936599	8,165	3,893	1.00	0.03	0.95 - 1.06	0.966
rs7675998	8,165	3,893	0.99	0.03	0.94 - 1.05	0.728
rs2736100	8,165	3,893	1.03	0.02	0.99 - 1.08	0.159
rs9420907	8,165	3,893	1.01	0.03	0.94 - 1.08	0.823
rs8105767	8,165	3,893	0.97	0.03	0.92 - 1.02	0.226
rs755017	8,165	3,893	0.92	0.03	0.86 - 0.98	0.009
Unweighted GRS	8,165	3,893	1.00	0.01	0.98 - 1.02	0.985
Weigthed GRS	8,165	3,893	0.97	0.02	0.93 - 1.01	0.133

SNP, single nucleotide polymorphism; *GRS,* genetic risk score; *HR*, hazard ratio; *SE*, standard error; *95% CI*, 95% confidence interval.

infection. Whereas most of these markers were indeed associated with LTL and showed an individual effect on prospective mortality, shorter LTL remained independently associated with prospective mortality in two generations. We should note that we did not have data available regarding other relevant immune-related markers such as erythrocyte sedimentation rate and fibrinogen. The associations of LTL with prospective mortality could also be confounded by non-cell-autonomous senescence as a consequence of viral infection [33]. Another explanation for the association between LTL and prospective mortality might be found in the association of LTL with metabolic parameters associated with cardiovascular disease risk and/or mortality, such as fasting insulin and homocysteine [8]. Whether these markers explain LTL-related prospective mortality still needs to be determined.

Previous studies have shown that telomere length dynamics is age-related, i.e., the rate of LTL shortening during adulthood is much slower than during early life [34]. However, the age range of our samples within the middle-aged and highly advanced aged groups is relatively small and all our individuals are in adulthood (above 30 years of age). Hence, our finding that shorter LTL is associated with increased mortality is less likely to be confounded by this phenomenon.

Our finding that LTL is not associated with familial longevity in middle age is in contradiction to the observations in long-lived Ashkenazi Jewish families [17]. This discrepancy could be explained by natural variation, such as differences in the demography or age of the samples (mean age 68 and 72 years (among Ashkenazi offspring and controls, respectively) versus 59 years (LLS offspring and spouses)) or by differences in the selection criteria which may have an effect on the genetic component of the longevity trait (offspring of centenarians (mean age 97 years) versus offspring of nonagenarian siblings (mean age 94 years)). Another possible explanation is the small sample size of the study of Atzmon and colleagues ($n_{offspring}$ = 175 and $n_{controls}$

= 93) in comparison to the current study ($n_{offspring}$ = 1,580 and $n_{spouses}$ = 725), which may have led to a non-random selection of individuals from the population, resulting in a false positive association.

The GRS composed of the 7 genetic variants associated with LTL variation in the largest GWAS reported so far is associated with coronary artery disease risk [20], but does not associate with prospective mortality in our study of 8,165 individuals of whom 3,893 died during follow-up. Since the 7 genetic variants only explain ~1% of the variation in LTL [20], the sample size of the current study might be insufficient to detect their effect on prospective mortality. Hence, on the basis of our data we cannot exclude a causative role for genetic variants in LTL related genes in prospective mortality.

Critical telomere length in tissues may be causally involved in lifespan regulation and our results further highlight the role of telomere length in blood as marker for prospective mortality. The lack of association of LTL with familial longevity in middle age and of the LTL-associated genetic variants with prospective mortality provides thus far no support for LTL causally contributing to lifespan variation in humans. However, LTL does reflect environmental effects, as demonstrated by the observation that there is a correlation in LTL between spouses [18]. Assortative mating may thus have obscured a difference between the LLS offspring and their spouses. However, the LLS offspring do have a more "youthful" metabolic profile and a lower prevalence of age-related diseases compared with their spouses, whereas LTL is not different between the groups. This indicates that LTL seems to associate with mortality independent of the familial trait that influences the metabolic health in these families in middle age. LTL meets three of the four criteria we proposed for a biomarker of healthy aging in a recent review [35], i.e., LTL associates with chronological age and with morbidity and mortality in prospective studies. However, LTL cannot be used to discriminate individuals in middle age according to their genetic propensity for longevity. Other potential biomarkers of healthy aging, such as fasting glucose and free triiodothyronine, did meet all criteria in studies of various human cohorts.

LTL could reflect the compartment of vital haematopoietic stem cells (HSCs) in individuals. Leukocytes consist of different subsets of cells, namely lymphocytes, monocytes and granulocytes (neutophils, basophils and eosinophils), which all originate from the HSC. Telomere length differs between leukocyte subsets [36]. However, since there is synchrony between the different subsets, an individual's LTL likely reflects the telomere length of the HSCs [37]. In this study, we show that LTL is associated with several leukocyte subset counts, namely lymphocyte, neutrophil and basophil counts (Table 6.1), which indicates that mean LTL is influenced by the frequency of the different leukocyte subsets. However, when we adjusted the prospective analysis of mortality for these counts, the effect of LTL on prospective mortality remained unchanged, indicating that the proportion of LTL variation caused by the frequency of the different leukocyte subsets does not influence prospective mortality. Nevertheless, this leaves the possibility open that LTL reflects the available HSC population.

Telomere dysfunction was found to be determined by the frequency of critically short telomeres. A recent study showed that the rate of increase in the frequency of these critically short telomeres and not the rate of telomere length shortening determines longevity in mice [38]. Since in the current study we only determined the mean LTL of an individual, we could not discriminate between individuals according to the frequency of dysfunctional telomeres. It would therefore be interesting to use quantitative fluorescence *in situ* hybridization, a method that is able to quantify critically short telomeres in subsets of cells [39], to determine the influence of the frequency of dysfunctional telomeres on longevity and prospective mortality in humans.

In conclusion, we confirmed LTL to be a marker of prospective mortality in middle and highly advanced age and additionally show that this association could not be explained by the association of LTL with the immune-related markers IGF-1/IGFBP3, CRP, IL-6, CMV serostatus, or WBC counts or by the currently known genetic variants contributing to LTL variation. Furthermore, we have shown that LTL is not associated with familial longevity in middle age. Hence, the approaches followed here do not further support the hypothesis that LTL contributes to the genetic propensity for longevity. Further studies need to be performed to determine which other environmental or novel genetic effects could underlie the association of LTL with prospective mortality.

Material and methods

Study populations

Leiden Longevity Study

For the LLS, long-lived siblings of European descent were recruited together with their offspring and the spouses 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 men and 91 years or older for women, representing less than 0.5% of the Dutch population in 2001 [23]. In total, 944 long-lived proband siblings with a mean age of 94 years (range, 89-104), 1,671 offspring (61 years, 39-81), and 744 spouses thereof (60 years, 36-79) were included. DNA from the LLS was extracted from samples at baseline using conventional methods [40] and genotyping was performed with Illumina Human660W-Quad and OmniExpress BeadChips (Illumina, San Diego, CA, USA). Imputation was performed using IMPUTE2 with reference HapMap Phase I + II CEU release 22 (hg18/build36).

A description of the cohorts used for the prospective meta-analysis of LTL-associated genetic variants is provided in the Supplementary Information.

Measurement of leukocyte telomere length

Mean LTL was measured as a ratio (T/S) of telomere repeat length (T) to the copy number (S) of the single-copy gene *36B4*, as previously described [20]. The inter- and intra-run coefficients of variation were 2.73% and 2.73% for the LLS nonagenarians and 3.74% and 2.85% for the LLS offspring and spouses, respectively. LTL was obtained

in 3,194 samples from the LLS, of which 19 were removed due to a deviation from the mean > 3 SD, leaving 3,175 samples for the analysis (Tables 6.1, 6.2, and 6.3). If we consider an LTL attrition rate of 30 telomeric base pairs per year and a decline of 0.006 T/S units per year [20], a decrease of one T/S unit reflects a decrease of 5,000 telomeric base pairs in our study. This LTL attrition rate is based on several studies that have used DNA blotting to measure LTL [20]. A recent review indicated that the LTL attrition rate is most likely somewhere in the range of 20-30 base pairs/year [41], so, the estimated telomeric base pairs representing one T/S ratio may vary between 3,333 and 5,000. In the analyses of LTL, one LTL unit represents 1 kb telomeric base pairs.

Measurement of immune-related parameters

In the LLS all standard serum measurements were performed using fully automated equipment. WBC counts were measured using the Sysmex XE-2100 (TOA Medical Electronics, Kobe, Japan). IGF-1 and IGFBP3 were measured using the Immulite 2500 (DPC, Los Angeles, CA, USA) [28] and high-sensitivity CRP was measured using the Hitachi Modular P800 (Roche, Almere, the Netherlands) [29]. Since the CRP levels were not normally distributed the log transformed values were used for analysis. IL-6 was measured with the PeliKine Compact human IL-6 ELISA kit (Sanquin Reagents, Amsterdam, the Netherlands) [42]. For calculation of IGF-1/IGFBP3 we used the following formula:

*IGF-1/IGFBP3 = IGF-1 (ng/ml)*0.130 / IGFBP3 (ng/ml)*0.036*

For all serum parameters, measurements with a deviation from the mean > 3 SD were removed. CMV serostatus was determined on blinded samples using the CMV-IgG-ELISA PKS assay (Medac GmbH, Wedel, Germany) [26]. The characteristics of the measured parameters are depicted in Tables 6.1 and 6.2.

Statistical analysis

Prospective analysis

Prospective analysis of LTL and LTL-associated genetic variants was performed using a Cox proportional hazards model adjusted for age, gender, population stratification, and study specific covariates. The number of individuals and (cause-specific) deaths for every cohort, as well as the follow-up times, are depicted in Tables 6.3 and S6.3. To determine whether the association of LTL with mortality was independent of immune-related markers, we fitted a model with and without adjustment for immune-related markers and determined whether the association of LTL with mortality remained ($P < 0.05$).

Association of LTL with immune-related markers

To determine the association of LTL with serum parameters in the LLS, we performed linear regression, adjusted for age, gender, and familial relationships, using the following model in STATA/SE 11.2 (StataCorp LP, College Station, TX, USA).

*T/S ratio ~ β0 + β1*age+ β2*gender + β3*(age*gender) + β4*group + β5*immune-related marker*

age was coded in years, *gender* was coded as 1 (male) or 2 (female), and *group* was coded as 0 (LLS spouse) or 1 (LLS offspring). Robust standard errors were used to account for sibship relations.

We assumed a linear association between LTL and the different immune-related markers since the augmented partial residual plots of the markers showed no visual deviation from linearity.

Association of LTL with familial longevity in middle age

To determine the association of LTL with familial longevity in middle age in the LLS offspring (n = 1,580) and their spouses (n = 725), linear regression, adjusted for age, gender, and familial relationships, was performed using the following model in STATA/SE 11.2 (StataCorp LP):

*T/S ratio ~ β0 + β1*age+ β2*gender + β3*(age*gender) + β4*group*

age was coded in years, *gender* was coded as 1 (male) or 2 (female), and *group* was coded as 0 (LLS spouse) or 1 (LLS offspring). Robust standard errors were used to account for sibship relations.

Genetic risk score

To determine the joint effect of LTL-associated genetic variants on all-cause, cardiovascular, and cancer mortality, we created a GRS using a previously described approach [20,43]. The unweighted GRS of a subject was defined as the combined number of risk alleles associated with shorter LTL in a previous GWAS [20]. For the weighted GRS, the β for each SNP in this GWAS was added as weight and the total score was divided by the sum of all weights.

Acknowledgments

This work was supported by funds from the European Community's Seventh Framework Programme (FP7/2007–2013), ENGAGE Consortium, grant agreement HEALTH-F4-2007-201413.

For the LLS, the research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2011) under grant agreement n° 259679. This study was financially supported by the Innovation-Oriented Research Program on Genomics (SenterNovem IGE05007), the Centre for Medical Systems Biology and the Netherlands Consortium for Healthy Ageing (grant 050-060-810), all in the framework of the Netherlands Genomics Initiative, Netherlands Organization for Scientific Research (NWO), by Unilever Colworth and by BBMRI-NL, a Research Infrastructure financed by the Dutch government (NWO 184.021.007).

CardioMetabolic Twins was supported by grants from the National Institutes of Health (AG028555, AG08861, AG04563, AG10175, and AG08724), the Swedish Council on Social Research, the MacArthur Foundation Research Network on Successful Aging, the Axel and Margaret Ax:son Johnsons Foundation, the Swedish

Foundation for Health Care Sciences and Allergy Research, and the Swedish Research Council (2007-2722).

The work by co-authors on behalf of the EGCUT study was supported by the OPENGENE grant from the European Commission Seventh Framework Programme, targeted funding from the Estonian Government (SF0180142s08), the Estonian Research Roadmap through the Estonian Ministry of Education and Research, the Center of Excellence in Genomics (EXCEGEN), the University of Tartu (SP1GVARENG) and the Estonian Science Foundation (ETF9353).

The PROSPER was supported by an investigator-initiated grant obtained from Bristol-Myers Squibb. J.W.J. is an established clinical investigator for the Netherlands Heart Foundation (grant 2001 D 032). Support for genotyping was provided by the seventh framework program of the European commission (grant 223004) and by the Netherlands Genomics Initiative (Netherlands Consortium for Healthy Aging, grant 050-060-810).

The Rotterdam Study was supported by Netherlands Genomics Initiative/ Netherlands Consortium for Healthy Aging (050-060-810); Netherlands Organisation for Scientific Research (NWO) (904-61-090, 904-61-193, 480-04-004, 400-05-717, SPI 56-464-1419, 175.010.2005.011 and 911-03-012); Netspar – Living longer for a good health.

TwinGene was supported by the Swedish Research Council (M-2005-1112), GenomEUtwin (EU/QLRT-2001-01254; QLG2-CT-2002-01254), NIH DK U01-066134, the Swedish Foundation for Strategic Research (SSF) and the Heart and Lung foundation no. 20070481.

V.C. and N.J.S. are supported by the British Heart Foundation and N.J.S. is an NIHR Senior Investigator.

References

1. Blackburn EH. Structure and function of telomeres. *Nature* 1991; **350**: 569-73.
2. Harley CB, Futcher AB, Greider CW. Telomeres shorten during ageing of human fibroblasts. *Nature* 1990; **345**: 458-60.
3. Slagboom PE, Droog S, Boomsma DI. Genetic determination of telomere size in humans: a twin study of three age groups. *Am J Hum Genet* 1994; **55**: 876-82.
4. Benetos A, Okuda K, Lajemi M, Kimura M, Thomas F, Skurnick J, *et al*. Telomere length as an indicator of biological aging: the gender effect and relation with pulse pressure and pulse wave velocity. *Hypertension* 2001; **37**: 381-5.
5. Fitzpatrick AL, Kronmal RA, Gardner JP, Psaty BM, Jenny NS, Tracy RP, *et al*. Leukocyte telomere length and cardiovascular disease in the cardiovascular health study. *Am J Epidemiol* 2007; **165**: 14-21.
6. Demissie S, Levy D, Benjamin EJ, Cupples LA, Gardner JP, Herbert A, *et al*. Insulin resistance, oxidative stress, hypertension, and leukocyte telomere length in men from the Framingham Heart Study. *Aging Cell* 2006; **5**: 325-30.
7. Willeit P, Willeit J, Kloss-Brandstatter A, Kronenberg F, Kiechl S. Fifteen-year follow-up of association between telomere length and incident cancer and cancer mortality. *JAMA* 2011; **306**: 42-4.
8. Sanders JL, Newman AB. Telomere Length in Epidemiology: A Biomarker of Aging, Age-Related Disease, Both, or Neither? *Epidemiol Rev* 2014; *In press*.
9. Levy MZ, Allsopp RC, Futcher AB, Greider CW, Harley CB. Telomere end-replication problem and cell aging. *J Mol Biol* 1992; **225**: 951-60.
10. von Zglinicki T. Oxidative stress shortens telomeres. *Trends Biochem Sci* 2002; **27**: 339-44.
11. Schwarz KB. Oxidative stress during viral infection: a review. *Free Radic Biol Med* 1996; **21**: 641-9.
12. Barbieri M, Paolisso G, Kimura M, Gardner JP, Boccardi V, Papa M, *et al*. Higher circulating levels of IGF-1 are associated with longer leukocyte telomere length in healthy subjects. *Mech Ageing Dev* 2009; **130**: 771-6.
13. Kaplan RC, Fitzpatrick AL, Pollak MN, Gardner JP, Jenny NS, McGinn AP, *et al*. Insulin-like growth factors and leukocyte telomere length: the cardiovascular health study. *J Gerontol A Biol Sci Med Sci* 2009; **64**: 1103-6.
14. Heemskerk VH, Daemen MA, Buurman WA. Insulin-like growth factor-1 (IGF-1) and growth hormone (GH) in immunity and inflammation. *Cytokine Growth Factor Rev* 1999; **10**: 5-14.
15. Naugler WE, Karin M. The wolf in sheep's clothing: the role of interleukin-6 in immunity, inflammation and cancer. *Trends Mol Med* 2008; **14**: 109-19.
16. Pepys MB, Hirschfield GM. C-reactive protein: a critical update. *J Clin Invest* 2003; **111**: 1805-12.
17. Atzmon G, Cho M, Cawthon RM, Budagov T, Katz M, Yang X, *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**: 1710-7.
18. Broer L, Codd V, Nyholt DR, Deelen J, Mangino M, Willemsen G, *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; **21**: 1163-8.
19. Mangino M, Hwang SJ, Spector TD, Hunt SC, Kimura M, Fitzpatrick AL, *et al*. Genome-wide meta-analysis points to CTC1 and ZNF676 as genes regulating telomere homeostasis in humans. *Hum Mol Genet* 2012; **21**: 5385-94.
20. Codd V, Nelson CP, Albrecht E, Mangino M, Deelen J, Buxton JL, *et al*. Identification of seven loci affecting mean telomere length and their association with disease. *Nat Genet* 2013; **45**: 422-7.
21. Deelen J, Uh HW, Monajemi R, van Heemst D, Thijssen PE, Bohringer S, *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; **35**: 235-49.
22. Soerensen M, Thinggaard M, Nygaard M, Dato S, Tan Q, Hjelmborg J, *et al*. Genetic

variation in TERT and TERC and human leukocyte telomere length and longevity: a cross-sectional and longitudinal analysis. *Aging Cell* 2012; **11**: 223-7.
23. Schoenmaker M, de Craen AJ, de Meijer PH, Beekman M, Blauw GJ, Slagboom PE, *et al*. Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur J Hum Genet* 2006; **14**: 79-84.
24. Slagboom PE, Beekman M, Passtoors WM, Deelen J, Vaarhorst AA, Boer JM, *et al*. Genomics of human longevity. *Philos Trans R Soc Lond B Biol Sci* 2011; **366**: 35-42.
25. Westendorp RG, van Heemst D, Rozing MP, Frolich M, Mooijaart SP, Blauw GJ, *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**: 1634-7.
26. Derhovanessian E, Maier AB, Beck R, Jahn G, Hahnel K, Slagboom PE, *et al*. Hallmark features of immunosenescence are absent in familial longevity. *J Immunol* 2010; **185**: 4618-24.
27. Mortensen LH, Maier AB, Slagboom PE, Pawelec G, Derhovanessian E, Petersen I, *et al*. Early-life environment influencing susceptibility to cytomegalovirus infection: evidence from the Leiden Longevity Study and the Longitudinal Study of Aging Danish Twins. *Epidemiol Infect* 2012; **140**: 835-41.
28. Rozing MP, Westendorp RG, Frolich M, de Craen AJ, Beekman M, Heijmans BT, *et al*. Human insulin/IGF-1 and familial longevity at middle age. *Aging (Albany NY)* 2009; **1**: 714-22.
29. Rozing MP, Mooijaart SP, Beekman M, Wijsman CA, Maier AB, Bartke A, *et al*. C-reactive protein and glucose regulation in familial longevity. *Age (Dordr)* 2011; **33**: 623-30.
30. Bakaysa SL, Mucci LA, Slagboom PE, Boomsma DI, McClearn GE, Johansson B, *et al*. Telomere length predicts survival independent of genetic influences. *Aging Cell* 2007; **6**: 769-74.
31. Rajpathak SN, McGinn AP, Strickler HD, Rohan TE, Pollak M, Cappola AR, *et al*. Insulin-like growth factor-(IGF)-axis, inflammation, and glucose intolerance among older adults. *Growth Horm IGF Res* 2008; **18**: 166-73.
32. Sesmilo G, Biller BM, Llevadot J, Hayden D, Hanson G, Rifai N, *et al*. Effects of growth hormone administration on inflammatory and other cardiovascular risk markers in men with growth hormone deficiency. A randomized, controlled clinical trial. *Ann Intern Med* 2000; **133**: 111-22.
33. Robinson MW, McGuinness D, Swann R, Barclay S, Mills PR, Patel AH, *et al*. Non cell autonomous upregulation of CDKN2 transcription linked to progression of chronic hepatitis C disease. *Aging Cell* 2013; **12**: 1141-3.
34. Aviv A. The epidemiology of human telomeres: faults and promises. *J Gerontol A Biol Sci Med Sci* 2008; **63**: 979-83.
35. Deelen J, Beekman M, Capri M, Franceschi C, Slagboom PE. Identifying the genomic determinants of aging and longevity in human population studies: progress and challenges. *Bioessays* 2013; **35**: 386-96.
36. Rufer N, Brummendorf TH, Kolvraa S, Bischoff C, Christensen K, Wadsworth L, *et al*. Telomere fluorescence measurements in granulocytes and T lymphocyte subsets point to a high turnover of hematopoietic stem cells and memory T cells in early childhood. *J Exp Med* 1999; **190**: 157-67.
37. Kimura M, Gazitt Y, Cao X, Zhao X, Lansdorp PM, Aviv A. Synchrony of telomere length among hematopoietic cells. *Exp Hematol* 2010; **38**: 854-9.
38. Vera E, Bernardes de JB, Foronda M, Flores JM, Blasco MA. The Rate of Increase of Short Telomeres Predicts Longevity in Mammals. *Cell Rep* 2012; **2**: 732-7.
39. Canela A, Vera E, Klatt P, Blasco MA. High-throughput telomere length quantification by FISH and its application to human population studies. *Proc Natl Acad Sci U S A* 2007; **104**: 5300-5.
40. Beekman M, Blauw GJ, Houwing-Duistermaat JJ, Brandt BW, Westendorp RG, Slagboom PE. 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**: 355-62.

41. Muezzinler A, Zaineddin AK, Brenner H. A systematic review of leukocyte telomere length and age in adults. *Ageing Res Rev* 2013; **12**: 509-19.
42. Stijntjes M, de Craen AJ, van Heemst D, Meskers CG, van Buchem MA, Westendorp RG, *et al*. Familial longevity is marked by better cognitive performance at middle age: the Leiden Longevity Study. *PLoS One* 2013; **8**: e57962.
43. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, *et al*. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 2011; **478**: 103-9.

Supplementary Information

The supplementary information belonging to this chapter can be found at: http://ije.oxfordjournals.org/content/early/2014/01/14/ije.dyt267/suppl/DC1.

Table S6.1 Results regression analysis LTL and immune-related markers.

Table S6.2 Association between LTL and mortality adjusted for immune-related markers.

Table S6.3 Characteristics of the samples used for the prospective analysis of LTL-associated genetic variants.

Table S6.4 Association between LTL-associated genetic variants and cardiovascular mortality.

Table S6.5 Association between LTL-associated genetic variants and cancer mortality.

Chapter 7

General discussion

The aim of this thesis was to identify novel lifespan regulating loci that influence human longevity and population mortality. The genetic component of longevity is expected to be small (~25%, Table 1.2). However, it is more prominent in families in which longevity clusters [1,2], which makes individuals from such families very suitable for genetic research. Since long-lived family members show a low prevalence of common diseases from middle age onwards [3-7], the genome of long-lived individuals is expected to harbor genetic variants that promote healthy aging and protect against age-related disease. We previously showed that longevity is not easily explained by the absence of susceptibility loci involved in common age-related diseases [8]. Therefore, we performed a genome-wide association study (GWAS) of long-lived individuals from the family-based Leiden Longevity Study (LLS) to identify genetic variants associated with increased survival into old age and extended the analysis by including individuals from other family-based and population-based cohorts of European descent. In addition, we performed gene set analysis on the LLS longevity GWAS dataset to determine the combined effect of genetic variation in two candidate pathways on longevity. We additionally investigated whether leukocyte telomere length (LTL) could be used as a biomarker of healthy aging in genomic studies of large cohorts of middle-aged individuals and whether the genetic component of LTL may be involved in human lifespan regulation.

Main findings

In **Chapter 2** we give an overview of the different genomic approaches that have thus far been used to identify mechanisms underlying healthy aging and longevity. Up till the start of this project, the number of identified genes and pathways contributing to human lifespan regulation had been limited.

As a first attempt to identify novel longevity loci, we performed a GWAS for longevity in long-lived families (**Chapter 3**), in which we identified one locus, the previously implicated *TOMM40/APOE/APOC1* locus [9,10], which associates with a decreased probability to survive to ages beyond 85 years. Through a prospective analysis, we additionally showed that the ApoE ε4 allele associates with increased mortality after 90 years, while we observed the opposite effect for the ApoE ε2 allele, although the latter was not significant. We confirmed the previously reported associations of the locus with metabolic and immune-related parameters and found a novel association with insulin-like growth factor 1 (IGF-1) signaling in women. Hence, the mechanism underlying the association of the *TOMM40/APOE/APOC1* locus with increased mortality likely involves a complex interaction between multiple physiological processes.

As our LLS longevity GWAS (**Chapter 3**), as well as those performed by other groups [11-15], had limited power, we substantially increased the sample size, thereby potentially enabling the identification of loci with smaller effects (odds ratio (OR) < 0.9 and > 1.1). Hence, in this extended GWAS in individuals from all over Europe

(**Chapter 4**), we identified a novel locus on chromosome 5q33.3 that associates with an increased probability to survive to ages beyond 90 years. In addition, prospective analysis showed that genetic variation at this locus also associates with decreased mortality. The locus has previously been reported to associate with low blood pressure in middle age, although we show that the mortality effects of the locus above 75 years seem to be independent from blood pressure, at least in the PROspective Study of Pravastatin in the Elderly at Risk and Leiden 85-plus study Cohort II. Thus, although the locus is implicated in blood pressure regulation, the mechanism by which genetic variation at chromosome 5q33.3 influences longevity likely also involves other traits.

The genetic component of longevity is expected to be small (~25%, Table 1.2) and assumed to be determined by many genes with small effects [16], which might explain the limited number of GWAS-identified longevity loci. Moreover, the increase in human life expectancy over the last two centuries due to environmental factors has resulted in the presence of so-called long-lived "phenocopies" in the population, i.e., individuals that survived to high ages independent of their genetic background. Although GWA analysis has successfully been applied to identify common genetic variants with small effects for several traits and diseases [17-19], the main problem of performing GWAS for longevity is the relatively low number of long-lived individuals with GWA data. The EU longevity GWAS described in **Chapter 4**, which is the largest GWAS for longevity up to date, contained ~18,000 long-lived individuals with GWA data. By combining the data of all currently available longevity cohorts with GWA data worldwide (~30,000 individuals above 85 years of age), we might be able to identify some additional longevity loci, although this sample size will still be insufficient to identify common genetic variants with relatively small effects (OR's between 0.9 and 1.1). Thus, instead of focusing on common genetic variants, genetic research of longevity should move towards genetic approaches in which the effect of high-impact private, i.e., observed in a single family, and rare genetic variants can be investigated, using, for example, next-generation sequencing.

Another approach is to determine the combined effect of single nucleotide polymorphisms (SNPs) on a trait, which may reflect the involvement of specific networks on aging. Hence, we performed candidate pathway-based SNP set analysis (**Chapter 5**) using the genotypes from the dataset described in **Chapter 3**. Based on results from previous studies in humans and animal models, we selected two candidate pathways for human longevity, the insulin/IGF-1 signaling (IIS) and telomere maintenance (TM) pathways. We showed that genetic variation in both these pathways is indeed associated with human longevity, at least in the LLS, which is mainly caused by the IIS genes *AKT1*, *AKT3*, *FOXO4*, *IGF2*, *INS*, *PIK3CA*, *SGK*, *SGK2*, and *YWHAG* and the TM gene *POT1*. In addition, we performed gene-set enrichment analysis on the summary data from the EU longevity GWAS described in **Chapter 4** using Meta-Analysis Gene-set Enrichment of variaNT Associations (http://www.broadinstitute.org/mpg/magenta/)

[20]. However, in this larger dataset, we were unable to find an enrichment of the loci within the IIS and TM pathways (P = 0.656 and P = 1.000, respectively), nor in any of the SNP sets from Kyoto Encyclopedia of Genes and Genomes and Gene Ontology. The difference with the results from the analysis described in **Chapter 5** might, for example, be due to the use of summary data instead of "raw" genotypes, although the observed associations within the IIS and TM could also be specific to individuals from long-lived families, like the LLS, or be false positives. Thus, SNP set analysis may be a useful method, that can be applied in addition to GWAS, to determine the combined effect of genetic variation in (known) genes and pathways on longevity.

A possibility to increase the sample size and, thus, the power of genetic approaches is by using biomarkers of healthy aging as a standardized phenotype for genetic studies. In **Chapter 2**, we discuss the concept of biomarker approaches and we propose four criteria for quantitative parameters (or profiles) that should be fulfilled before consideration as biomarkers of healthy aging. In short, a biomarker of healthy aging must (1) show a change with chronological age, (2) discriminate individuals based on their biological age and/or genetic propensity for longevity, and associate with (3) known health parameters and (4) morbidity and/or mortality in prospective studies.

A potential biomarker of healthy aging is LTL, since it has previously been associated with multiple diseases and increased prospective mortality [21]. We therefore investigated whether LTL satisfies the proposed criteria for biomarkers of healthy aging (**Chapter 6**). We showed that LTL indeed changes with chronological age and is associated with known health parameters and (immune-independent) prospective mortality. However, LTL was unable to discriminate individuals based on their genetic propensity for longevity (criterion 2). To determine whether LTL could nevertheless be used as a standardized phenotype for genetic studies of healthy aging and longevity, we performed a look-up of the previously identified LTL-associated genetic variants [22] in our EU longevity GWAS results described in **Chapter 4**. Interestingly, two of these variants, rs10936599 (*TERC*) and rs2736100 (*TERT*), were located near or in genes that we also analyzed in the gene set analysis of the TM pathway described in **Chapter 5**. However, none of the LTL-associated variants showed an association with survival to ages above 90 years (Table 7.1). Thus, although LTL meets three of the four proposed criteria for a biomarker of healthy aging, it could not be used as a standardized phenotype for genetic studies of healthy aging and longevity. Hence, we need to search for parameters that meet all four proposed criteria for biomarkers of healthy aging.

Functional characterization of longevity loci

Once novel longevity loci have been identified through genetic approaches, one of the challenges that lies ahead is the functional characterization of such loci, since quite a few of them will be mapped to non-protein-coding regions of which the

Table 7.1 Association of leukocyte telomere length-associated genetic variants with survival to ages above 90 years.

SNP	Chr	Position (bp)	Candidate / closest gene	EA	*n*		EAF		*P*
					Cases	Controls	Cases	Controls	
rs11125529	2	54,329,370	*ACYP2*	C	5,406	15,112	0.864	0.861	0.872
rs10936599	3	170,974,795	***TERC***	T	5,406	15,112	0.248	0.250	0.467
rs7675998	4	164,227,270	*NAF1*	G	5,406	15,112	0.212	0.217	0.385
rs2736100	5	1,339,516	***TERT***	C	5,024	9,996	0.474	0.485	0.452
rs9420907	10	105,666,455	*OBFC1*	C	5,406	15,112	0.855	0.872	0.140
rs8105767	19	22,007,281	*ZNF208*	G	5,406	15,112	0.719	0.714	0.702
rs755017	20	61,892,066	*RTEL1*	G	5,406	15,108	0.880	0.869	0.320

SNP, single nucleotide polymorphism; *Chr*, chromosome according to NCBI build 36; *Position (bp)*, position according to NCBI build 36; *EA*, effect allele (allele associated with shorter LTL); *EAF*, effect allele frequency; *P*, *P*-value for the association with survival to ages above 90 years. Genes in **bold** were also analyzed in the gene set analysis of the telomere maintenance pathway described in **Chapter 5**.

functional consequences are still unclear. An example is the chromosome 5q33.3 locus we identified in **Chapter 4**. The functional characterization of longevity loci consist of several steps (Figure 7.1), of which many overlap with the steps proposed for other traits [23,24].

The first step is genotypic fine-mapping, i.e., to identify the causal variant(s) by, for example, targeted resequencing based on the linkage disequilibrium (LD) structure within the locus. Since targeted resequencing is expensive, one could first browse the publically available data of the 1000 Genomes Project, which is aimed at capturing all common and low-frequency genetic variation (minor allele frequency > 1%) in diverse ethnic populations [25], to fine-map the region of interest based on the haplotypes of the individuals from the same ethnicity. Alternatively, one could use population specific reference panels, such as the ones that will be created in the Singapore Sequencing Malay Project [26] and the Genome of the Netherlands project [27]. We performed genotypic fine-mapping for the chromosome 5q33.3 locus using the publically available 1000 Genomes Project data (**Chapter 4**) and were able to fine-map our locus to a ~22.3 kb region. However, we have thus far not identified the causal variant(s), although several of the variants in high LD with our lead SNP (r^2 > 0.8) are, according to the ENCODE data implemented in the UCSC genome browser, located in functional elements, such as DNase I hypersensitivity sites, transcription factor binding sites, and enhancer histone marks (Figure 7.2). In addition, the ~22.3 kb region seems to contain a long intergenic non-coding RNA, RP11-524N5.1, which has recently been annotated by the GENCODE consortium.

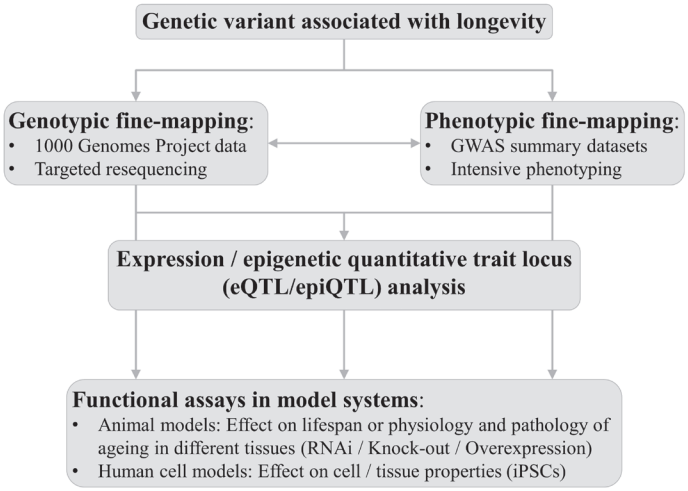
The second step is phenotypic fine-mapping, i.e., to identify other (combinations of) metabolic phenotypes, clinical endpoints,

and diseases associating with the locus of interest that could shed light on the mechanism underlying the association of interest. A helpful intermediate step is to browse the large publically available GWAS summary datasets, such as those for cholesterol levels [19], blood pressure [28], and type 2 diabetes [18]. One has to note, however, that these sets only contain data on HapMap imputed SNPs (~2,500,000), although several large GWAS initiatives based on 1000 Genomes imputation are ongoing. Another approach that may be helpful in identifying other traits and diseases associating with a locus of interest is to perform a PheWAS, i.e., to determine the association of a SNP with thousands of different phenotypes at once using, for example, International Classification of Diseases codes in large population-based studies. Up till now, phenotypic fine-mapping of the chromosome 5q33.3 locus using the publically available GWAS summary datasets has not resulted in identification of phenotypes that may shed light on the mechanisms by which the locus influences longevity (**Chapter 4**). Furthermore, application of the PheWAS approach using the available phenotypic data in the LLS was unsuccessful. However, we have, thus far, not performed the PheWAS approach in a large population-based study containing thousands of phenotypes, such as the Rotterdam Study.

The third step is expression/ epigenetic quantitative trait locus (eQTL/ epiQTL) analysis, i.e., to determine whether there is an effect of the causal variant(s) on expression and/or methylation of (nearby) genes. The pathophysiology of aging and longevity involves many different tissues. Hence, eQTL/epiQTL effects of longevity loci could be present in tissues for which gene expression or methylation data is not (yet) available. In addition, eQTL/epiQTL effects are expected to be small, so large datasets will be required to achieve sufficient power to detect them. There are several publically available databases containing eQTL data for multiple tissue, such as adipose tissue, brain (cerebellum, frontal cortex, temporal cortex, and pons), fibroblasts, liver, skin, and lymphoblastoid cell lines [29,30]. In addition, the ongoing Genotype-Tissue Expression project (http://www.broadinstitute.org/gtex/) will provide publically available eQTL data for around 30 different tissues. Thus far, there is no publically available database containing epiQTL data. However, novel platforms, such as Infinium HumanMethylation450 BeadChips and reduced representation bisulfite sequencing, have made it possible to determine epigenetic effects on the whole genome, which will aid to the identification of epiQTL effects in large datasets. We performed a look-up of all SNPs in high LD with our lead SNP at chromosome 5q33.3 ($r^2 > 0.8$) in several of these eQTL databases (**Chapter 4**). However, none of the SNPs showed an association with gene expression, so it is still unclear on which gene(s) and in which tissue(s) our locus exert its effects.

When a candidate susceptibility gene or region is identified (through step 1-3), the final step is to perform functional assays in model systems (animals/cell models). There are several animals that are routinely used in research of healthy aging and longevity, namely worms, flies, and mice. In these animals lifespan regulating effects could be studied by modifying gene functions

Figure 7.1 Functional characterization of longevity loci.



(mutagenesis) via RNA interference, knock-out, or overexpression. In addition, mice could also be used to study the effect of genes on the physiology and pathology of aging in different tissues [31]. However, before a gene or region can be studies in animal models it is important to determine the conservation. The chromosome 5q33.3 region, for example, is only conserved in primates, so for this region studies in animal models seem not very useful. To study the effects of the gene or region of interest in humans, one could create cell lines of different tissues by differentiation of induced pluripotent stem cells obtained by de-differentiation of fibroblast from carriers and non-carriers of the locus of interest.

Reducing heterogeneity in the healthy aging phenotype

Our genetic analyses illustrate that it is very difficult to identify human longevity loci, which may be due to the complexity of the phenotype along with the low number of long-lived individuals available for genetic research. In addition, analyses might be confounded by environmental factors that give rise to long-lived "phenocopies", which could even be present within long-lived families. Hence, to reduce phenotypic heterogeneity, additional selection criteria are required to select the most optimal individuals for genetic research, which could, for example, be based on the age (centenarians or even supercentenarians), zygosity (monozygotic twins), and/or family characteristics (families with the highest number of long-lived individuals or best family history for longevity, i.e., the longest survival among their parents) of an individual.

In addition, genetic studies may profit from biomarker studies that are aimed to identify phenotypes that reflect biological age. Up till now, several potential biomarkers of biological age have been

identified, such as fasting glucose levels, free triiodothyronine (fT3) levels, and gait speed (see **Chapter 2** for an overview). The next step is to determine whether these biomarkers, which should preferably be combined into one multimarker score, could be used as standardized phenotype for genetic studies. Therefore, the joint effect of (GWAS-identified) genetic variants associated with this multimarker score should be tested for their effect on longevity in using, for example, genetic risk scores. Ideally, one would perform a GWAS for this multimarker score in individuals from long-lived families, since identified loci are expected to be involved in the mechanism underlying their longevity as well. However, there is large heterogeneity between long-lived family studies and the number of individuals with GWA data (< 10,000 individuals) is insufficient to identify common genetic variants with small effects. Hence, instead one could use the loci identified through large GWAS of population-based cohorts (> 100,000 individuals). Thus far, however, the only potential biomarker of biological age for which multiple GWAS-identified loci have been reported is fasting glucose, although the currently identified genetic variants only explain 4.8% of the variance in fasting glucose levels [32]. Hence, larger GWAS are required to identify genetic variants explaining the remaining heritability of fasting glucose, fT3, and gait speed, which could subsequently be tested for their association with mortality and longevity. Interestingly, a look-up of the fasting glucose-associated variants in our EU longevity GWAS results described in **Chapter 4** showed that several of these variants also seem to associate with survival

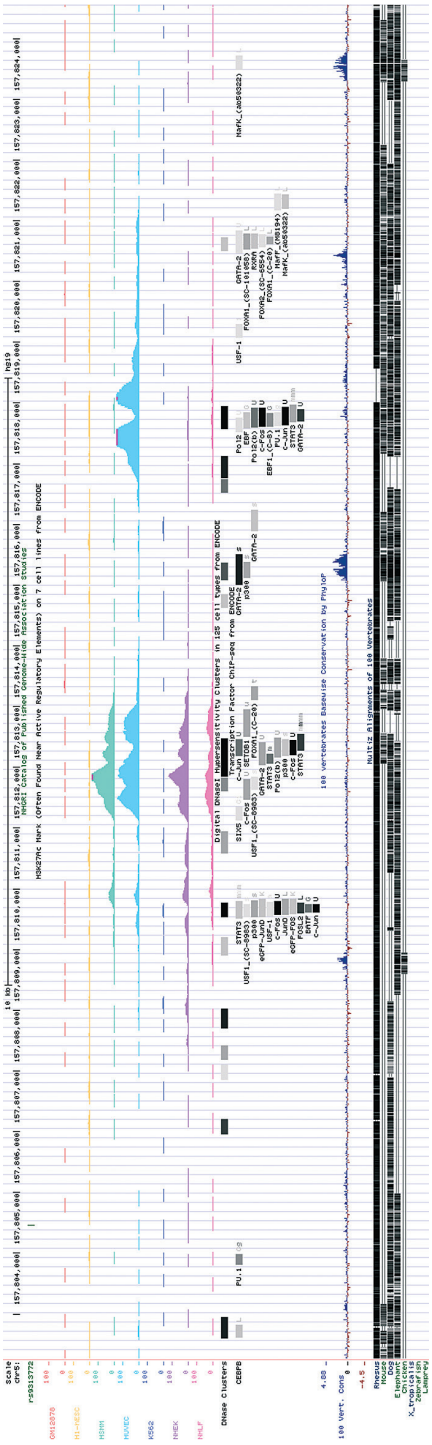


Figure 7.2 UCSC plot of the ~22.3 kb intergenic region on chromosome 5q33.3.

to ages above 90 years (Table 7.2), which is more promising than what we observed for LTL-associated genetic variants (Table 7.1).

Combining study designs for biomarker research

In addition to the lack of a well-defined phenotype for healthy aging, there is currently no study that allows testing of all the proposed criteria for a biomarker of healthy aging. The most optimal study design would be a population-based study in which a large group of families is followed during their entire lifetime and examined at multiple time points. An example of such a study is LifeLines (https://lifelines.nl/), which currently contains ~146.000 individuals from the Northern part of the Netherlands. However, this study is still in the recruitment phase and at the moment the best alternative for studies of healthy aging and longevity is to combine family-based studies with large prospective population-based studies.

The advantage of the study design of the LLS, as compared to other long-lived family-based studies, is that individuals have been followed-up for over 10 years. Hence, the LLS allows testing of most of the proposed criteria for a biomarker of healthy aging, although replication of results in larger family-based and prospective studies with longer follow-up times is still required. The association of a marker with chronological age could be determined using all individuals included in the study, although one has to take into account that the age range in the LLS is limited due to the family-based design of the study. The association of a marker with biological age could be determined by comparing the LLS offspring (considered as "healthy agers") with their spouses (controls). The strength of this comparison is that the offspring and their spouses share the same environment, so observed difference are most likely caused by differences in the genetic background. However, since approximately 50% (for a dominant inherited locus) or 75% (for a recessive inherited locus) of the offspring will not have inherited the genes responsible for the long-lived phenotype in their parents, phenotypic differences might be diluted due to the presence of individuals in the offspring group without the genetic background to become long-lived. In addition, the effects of a marker on biological age might only be present at older ages. Hence, these effects might not be detected in the middle-aged offspring and spouses. The association of a marker with known health parameters could be determined in the combined group of offspring and controls, for which data on numerous phenotypes is available. The association of a marker with mortality could be determined in the LLS nonagenarians (highly advanced age) and the combined group of LLS offspring and controls (middle age). In addition, the latter group could be used to determine the association of a marker with morbidity.

Novel methods and technologies plea for data integration

Research into human lifespan may also benefit from novel technologies and methodologies that have (recently) become available.

Table 7.2 Association of fasting glucose-associated genetic variants with survival to ages above 90 years.

					n		EAF		
SNP	Chr	Position (bp)	Candidate / closest gene	EA	Cases	Controls	Cases	Controls	*P*
rs340874	1	212,225,879	*PROX1*	C	5,406	15,112	0.548	0.553	0.728
rs780094	2	27,594,741	*GCKR*	C	5,406	15,111	0.607	0.626	0.509
rs560887	2	169,471,394	*G6PC2*	C	5,406	15,112	0.698	0.703	0.515
rs11715915	3	49,430,334	*AMT*	C	5,406	15,104	0.692	0.701	0.024
rs11708067	3	124,548,468	*ADCY5*	A	5,406	15,111	0.765	0.767	0.491
rs1280	3	172,195,984	*SLC2A2*	T	5,406	15,112	0.858	0.867	0.685
rs7651090	3	186,996,086	*IGF2BP2*	G	5,406	15,112	0.288	0.302	0.241
rs7708285	5	76,461,623	*ZBED3*	G	5,406	15,105	0.290	0.280	0.505
rs4869272	5	95,565,204	*PCSK1*	T	5,406	15,111	0.689	0.686	0.915
rs17762454	6	7,158,199	*RREB1*	T	5,406	15,112	0.258	0.253	0.493
rs9368222	6	20,794,975	*CDKAL1*	A	5,406	15,107	0.280	0.272	0.527
rs2191349	7	15,030,834	*DGKB / TMEM195*	T	5,406	15,112	0.536	0.532	0.334
rs2908289	7	44,190,467	*GCK*	A	5,406	15,112	0.178	0.175	0.570
rs6943153	7	50,759,073	*GRB10*	T	5,406	15,112	0.314	0.315	0.043
rs983309	8	9,215,142	*PPP1R3B*	T	5,406	15,111	0.108	0.114	0.749
rs11558471	8	118,254,914	*SLC30A8*	A	5,406	15,112	0.687	0.691	0.027
rs10814916	9	4,283,150	*GLIS3*	C	5,406	15,112	0.497	0.494	0.375
rs10811661	9	22,124,094	*CDKN2B*	T	5,406	15,112	0.820	0.828	0.803
rs16913693	9	110,720,180	*IKBKAP*	T	4,417	10,445	0.969	0.972	0.800
rs3829109	9	138,376,587	*DNLZ*	G	5,406	15,112	0.702	0.711	0.091
rs11195502	10	113,029,657	*ADRA2A*	C	5,406	15,112	0.916	0.911	0.434
rs7903146	10	114,748,339	*TCF7L2*	T	5,406	15,111	0.283	0.282	0.028
rs11607883	11	45,796,285	*CRY2*	G	5,406	15,112	0.473	0.479	0.475
rs11039182	11	47,303,299	*MADD*	T	5,406	15,112	0.707	0.724	0.854
rs174576	11	61,360,086	*FADS1*	C	5,406	15,112	0.650	0.660	0.004
rs11603334	11	72,110,633	*ARAP1*	G	5,406	15,110	0.844	0.834	0.755
rs10830963	11	92,348,358	*MTNR1B*	G	5,406	15,111	0.283	0.288	0.046
rs2657879	12	55,151,605	*GLS2*	G	5,406	15,112	0.203	0.178	0.111
rs10747083	12	131,551,691	*P2RX2*	A	5,406	15,112	0.683	0.690	0.921
rs11619319	13	27,385,599	*PDX1*	G	5,406	15,108	0.211	0.217	0.568
rs576674	13	32,452,302	*KL*	G	5,406	15,112	0.159	0.160	0.747
rs3783347	14	99,909,014	*WARS*	G	5,406	15,112	0.776	0.763	0.367
rs4502156	15	60,170,447	*VPS13C / C2CD4A/B*	T	5,406	15,112	0.573	0.561	0.799
rs2302593	19	50,888,474	*GIPR*	C	5,406	15,112	0.495	0.488	0.877
rs6113722	20	22,505,099	*FOXA2*	G	4,997	11,529	0.960	0.963	0.218
rs6072275	20	39,177,319	*TOP1*	A	5,406	15,112	0.151	0.155	0.850

SNP, single nucleotide polymorphism; *Chr*, chromosome according to NCBI build 36; *Position (bp)*, position according to NCBI build 36; *EA*, effect allele (allele associated with higher fasting glucose); *EAF*, effect allele frequency; *P*, *P*-value for the association with survival to ages above 90 years.

For genetic research, next-generation (whole-genome or exome) sequencing and multigenerational linkage may be used, since these require a limited number of individuals to identify novel longevity-associated loci.

Next-generation sequencing can be used to identify high-impact private and rare genetic variants associated with the trait of interest. This method allows hypothesis-based, such as regions identified through linkage analysis, as well as explorative studies of the genome and has successfully been applied to detect novel genetic variants associated with, for example, Alzheimer's disease [33] and bone mineral density [34]. We recently finished whole-genome sequencing of 220 nonagenarian individuals from the LLS with the best family history for longevity, i.e., the longest survival among their parents, to reduce heterogeneity in the phenotype due to "phenocopies". We will compare the genome of these individuals with that of younger controls to identify genetic variants that could explain the long-lived phenotype in their families.

Linkage analysis takes advantage of the sharing of alleles between siblings identical by descent and/or parents and their offspring to identify genomic regions associated with the trait of interest. The most optimal linkage study would be multigenerational, i.e., containing data on multiple generations within families. However, the main problem with the use of the multigenerational design for longevity research is that there is currently no (combination of) phenotype(s) that is able to predict which middle-aged individuals will become long-lived. Hence, up till now, linkage analysis for longevity has only been performed using long-lived siblings. Nevertheless, the use of fasting glucose levels, fT3 levels, and gait speed, or a multimarker score based on all three, would be a good starting point for multigenerational linkage analysis.

Biomarker research has, thus far, mostly been focussed on single quantitative parameters that are also used in the clinic. However, several technologies have recently become available that made it possible to study age-related changes in a large part of the human transcriptome [35], epigenome [36], metabolome [37], and glycome [38], Due to the wealth of information obtained using a single-point measurement these omics-based technologies could potentially be much more informative than the single quantitative parameters studied so far. For most of the omics-based technologies one or more of the proposed criteria for biomarkers of healthy aging have already been tested and the most interesting potential biomarkers identified using these platforms are the genes *RPTOR*, *ASF1A*, *IL7R* (transcriptomics) [39,40], and *ELOVL2* (epigenomics) [41], the *N*-glycan features LC-7 and LC-8 [42], bisecting GlcNAc glycoforms of IgG (glycomics) [43], and several lipid species (lipidomics) [44]. However, it still needs to be determined whether these features also associate with known health parameters and morbidity and/or mortality before they can be considered as biomarkers of healthy aging.

Instead of testing single parameters and/or profiles for association with longevity one could try to combine data to create a multimarker prediction score. An example of a multimarker prediction score that is highly informative for the prediction of coronary

heart disease is the Framingham risk score [45]. This score is a combination of age, gender, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, blood pressure, diabetes, and smoking. We are currently working on a multimarker prediction score for longevity by combining all clinical measurements available in the LLS. This multimarker prediction score, which, in the future, may also take into account omics-based measurements, should be able to discriminate individuals based on their biological age, i.e., classify individuals as member of long-lived families or controls. Subsequently, this score could be used in middle-aged cohorts to identify individuals suitable for genetic studies of longevity, even before these individuals have reached a high age.

Since the use of genome-wide omics-based measurements often leads to novel findings which are hard to interpret biologically, multilevel data integration may add to the interpretability of research into healthy aging and longevity. Alternatively, data may be integrated over species to identify conserved pathways. In contrast to human studies, animal-based studies are being used to investigate the effect of genetic manipulation and gene-environment interactions on life history traits and lifespan regulation. An example of a project which makes use of a data integration approach is the Integrated research on Developmental determinants of Ageing and Longevity project (http://www.ideal-ageing.eu/), in which late effects of early adverse exposures are being studied in various organisms simultaneously. Optimistically, data integration approaches over species contribute to the identification of novel conserved pathways involved in healthy aging and longevity. Not all the loci relevant for human aging, however, obtain attention in animal-based studies. The novel identified chromosome 5q33.3 region, for example, is a primate-specific locus involved in blood pressure regulation. Hence, for this locus, as well as the phenotype, animal-based studies of mice and lower species may not be very useful.

On the other hand, omics-based measurements may be integrated using a systems biology approach. This approach covers the study of the complex interactions within biological systems, which requires both data-driven modelling and hypothesis-driven experimental studies [46]. The extensive systems biology animal and human-based studies into the effects of aging on metabolism of cells and tissues requires perturbations and careful measurement of system responses. This will contribute to a deeper understanding of metabolism and will open possibilities for interpretation of human data. An example of this approach in humans is to analyze integrative personal omics profiles, the combination of the genetic, transcriptomic, proteomic, metabolomic, and autoantibody profile of individuals [47], for association with phenotypes of interest. This results in a model for the etiology of the phenotype, which may be tested in other individuals. Hence, a systems biology data integration approach may provide insight into the complex mechanisms underlying lifespan regulation.

Conclusions

The past couple of years large genome-wide association meta-analyses have successfully identified genetic variants associated with age-related diseases and traits [18,19,28]. However, the number of GWAS-identified genetic variants associated with human lifespan, thus far, has been limited to *TOMM40/APOE/APOC1* locus and our novel identified locus on chromosome 5q33.3. In addition, pathway analysis showed that there seems to be a role for genes involved in IIS and TM.

A better definition of the healthy aging phenotype, combining study designs, as well as the use of novel methods and technologies, such as next-generation sequencing, may help to identify novel loci contributing to longevity. In addition, biomarker approaches using omics-based technologies and multimarker prediction scores applied to individuals from long-lived families and large prospective study populations can help to identify parameters and/or profiles that can be used as standardized phenotype for genetic research. The data created using these approaches may subsequently be integrated over different species or in a systems biology approach to recognize the most relevant profiles and pathways involved in healthy aging and longevity.

References

1. Perls T, Shea-Drinkwater M, Bowen-Flynn J, Ridge SB, Kang S, Joyce E, *et al*. Exceptional familial clustering for extreme longevity in humans. *J Am Geriatr Soc* 2000; **48**: 1483-5.
2. Schoenmaker M, de Craen AJ, de Meijer PH, Beekman M, Blauw GJ, Slagboom PE, *et al*. Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur J Hum Genet* 2006; **14**: 79-84.
3. Atzmon G, Schechter C, Greiner W, Davidson D, Rennert G, Barzilai N. Clinical phenotype of families with longevity. *J Am Geriatr Soc* 2004; **52**: 274-7.
4. Newman AB, Glynn NW, Taylor CA, Sebastiani P, Perls TT, Mayeux R, *et al*. Health and function of participants in the Long Life Family Study: A comparison with other cohorts. *Aging (Albany NY)* 2011; **3**: 63-76.
5. Bos SD, Beekman M, Maier AB, Karsdal MA, Kwok WY, Bay-Jensen AC, *et al*. Metabolic health in families enriched for longevity is associated with low prevalence of hand osteoarthritis and influences OA biomarker profiles. *Ann Rheum Dis* 2013; **72**: 1669-74
6. Terry DF, Wilcox MA, McCormick MA, Pennington JY, Schoenhofen EA, Andersen SL, *et al*. Lower all-cause, cardiovascular, and cancer mortality in centenarians' offspring. *J Am Geriatr Soc* 2004; **52**: 2074-6.
7. Westendorp RG, van Heemst D, Rozing MP, Frolich M, Mooijaart SP, Blauw GJ, *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**: 1634-7.
8. Beekman M, Nederstigt C, Suchiman HE, Kremer D, van der Breggen R, Lakenberg N, *et al*. Genome-wide association study (GWAS)-identified disease risk alleles do not compromise human longevity. *Proc Natl Acad Sci U S A* 2010; **107**: 18046-9.
9. Christensen K, Johnson TE, Vaupel JW. The quest for genetic determinants of human longevity: challenges and insights. *Nat Rev Genet* 2006; **7**: 436-48.
10. Schachter F, Faure-Delanef L, Guenot F, Rouger H, Froguel P, Lesueur-Ginot L, *et al*. Genetic associations with human longevity at the APOE and ACE loci. *Nat Genet* 1994; **6**: 29-32.
11. Malovini A, Illario M, Iaccarino G, Villa F, Ferrario A, Roncarati R, *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**: 283-91.
12. Nebel A, Kleindorp R, Caliebe A, Nothnagel M, Blanche H, Junge O, *et al*. A genome-wide association study confirms APOE as the major gene influencing survival in long-lived individuals. *Mech Ageing Dev* 2011; **132**: 324-30.
13. Newman AB, Walter S, Lunetta KL, Garcia ME, Slagboom PE, Christensen K, *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**: 478-87.
14. Sebastiani P, Solovieff N, Dewan AT, Walsh KM, Puca A, Hartley SW, *et al*. Genetic signatures of exceptional longevity in humans. *PLoS One* 2012; **7**: e29848.
15. Walter S, Atzmon G, Demerath EW, Garcia ME, Kaplan RC, Kumari M, *et al*. A genome-wide association study of aging. *Neurobiol Aging* 2011; **32**: 2109-28.
16. Finch CE, Tanzi RE. Genetics of aging. *Science* 1997; **278**: 407-11.
17. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, *et al*. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 2010; **467**: 832-8.
18. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, *et al*. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* 2012; **44**: 981-90.
19. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M,

et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010; **466**: 707-13.
20. Segre AV, Groop L, Mootha VK, Daly MJ, Altshuler D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet* 2010; **6:** e1001058.
21. Sanders JL, Newman AB. Telomere Length in Epidemiology: A Biomarker of Aging, Age-Related Disease, Both, or Neither? *Epidemiol Rev* 2014; *In press*.
22. Codd V, Nelson CP, Albrecht E, Mangino M, Deelen J, Buxton JL. Identification of seven loci affecting mean telomere length and their association with disease. *Nat Genet* 2013; **45**: 422-7
23. Freedman ML, Monteiro AN, Gayther SA, Coetzee GA, Risch A, Plass C, *et al*. Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet* 2011; **43**: 513-8.
24. McCarthy MI, Hirschhorn JN. Genome-wide association studies: potential next steps on a genetic journey. *Hum Mol Genet* 2008; **17**: R156-R165.
25. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, *et al*. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; **491**: 56-65.
26. Wong LP, Ong RT, Poh WT, Liu X, Chen P, Li R, *et al*. Deep whole-genome sequencing of 100 southeast Asian Malays. *Am J Hum Genet* 2013; **92**: 52-66.
27. Boomsma DI, Wijmenga C, Slagboom PE, Swertz MA, Karssen LC, Abdellaoui A, *et al*. The Genome of the Netherlands: design, and project goals. *Eur J Hum Genet* 2014; **22**: 221-7.
28. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, *et al*. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 2011; **478**: 103-9.
29. Yang TP, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, Stranger BE, *et al*. Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics* 2010; **26**: 2474-6.
30. Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R, *et al*. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One* 2010; **5**: e10693.
31. Pettan-Brewer C, Treuting PM. Practical pathology of aging mice. *Pathobiol Aging Age Relat Dis* 2011; **1**.
32. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, *et al*. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet* 2012; **44**: 991-1005.
33. Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson PV, Bjornsson S, *et al*. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* 2012; **488**: 96-9.
34. Styrkarsdottir U, Thorleifsson G, Sulem P, Gudbjartsson DF, Sigurdsson A, Jonasdottir A, *et al*. Nonsense mutation in the LGR4 gene is associated with several human diseases and other traits. *Nature* 2013; **497**: 517-20.
35. de Magalhaes JP, Curado J, Church GM. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics* 2009; **25**: 875-81.
36. Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, *et al*. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet* 2012; **8**: e1002629.
37. Yu Z, Zhai G, Singmann P, He Y, Xu T, Prehn C, *et al*. Human serum metabolic profiles are age dependent. *Aging Cell* 2012; **11**: 960-7.
38. Dall'olio F, Vanhooren V, Chen CC, Slagboom PE, Wuhrer M, Franceschi C. N-glycomic biomarkers of biological aging and longevity: A link with inflammaging. *Ageing Res Rev* 2012 ; **12**: 685-98.
39. Passtoors WM, Boer JM, Goeman JJ, van den Akker EB, Deelen J, Zwaan BJ, *et al*. Transcriptional profiling of human familial longevity indicates a role for ASF1A and IL7R. *PLoS One* 2012; **7**: e27759.

40. Passtoors WM, Beekman M, Deelen J, van der Breggen R, Maier AB, Guigas B, *et al*. Gene expression analysis of mTOR pathway: association with human longevity. *Aging Cell* 2013; **12**: 24-31.
41. Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D, *et al*. Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell* 2012; **11**: 1132-4.
42. Ruhaak LR, Uh HW, Beekman M, Hokke CH, Westendorp RG, Houwing-Duistermaat J, *et al*. Plasma protein N-glycan profiles are associated with calendar age, familial longevity and health. *J Proteome Res* 2011; **10**: 1667-74.
43. Ruhaak LR, Uh HW, Beekman M, Koeleman CA, Hokke CH, Westendorp RG, *et al*. Decreased levels of bisecting GlcNAc glycoforms of IgG are associated with human longevity. *PLoS One* 2010; **5**: e12566.
44. Gonzalez-Covarrubias V, Beekman M, Uh HW, Dane A, Troost J, Paliukhovich I, *et al*. Lipidomics of familial longevity. *Aging Cell* 2013; **12**: 426-34.
45. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998; **97**: 1837-47.
46. Kirkwood TB. Systems biology of ageing and longevity. *Philos Trans R Soc Lond B Biol Sci* 2011; **366**: 64-70.
47. Chen R, Mias GI, Li-Pook-Than J, Jiang L, Lam HY, Chen R, *et al*. Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* 2012; **148**: 1293-307.

Chapter 8

Summary

One of the main problems in the Western world is the increase in years of life individuals spent in disability, which is mainly caused by the increased prevalence of diseases with increasing age. However, by identification of mechanisms driving healthy aging and protection from age-related diseases, we might be able to extend the disability-free life expectancy. In this thesis, we mainly focused on the genetic component of longevity, which has been estimated to explain ~25% of the variation in human lifespan.

In **Chapter 2** we have reviewed the different genetic studies that have thus far been performed to study healthy aging and longevity. In addition, we discussed several of the family-based and prospective studies that have been initiated to identify biomarkers of healthy aging. We propose that quantitative parameters (or profiles) must (1) show a change with chronological age, (2) discriminate individuals based on their biological age and/or genetic propensity for longevity, and associate with (3) known health parameters and (4) morbidity and/or mortality in prospective studies before consideration as biomarkers of healthy aging.

In **Chapter 3**, **4,** and **5** we tried to identify (novel) lifespan regulating loci using a genetic approach. Hence, we used two different methods, namely single single nucleotide polymorphism (SNP)-based genome-wide association study (GWAS) analysis (**Chapter 3** and **4**) and gene set analysis, which is able to determine the combined effect of SNPs on a trait (**Chapter 5**) In our first GWAS (**Chapter 3**), we identified one locus that associates with a decreased probability to survive to ages beyond 85 years, which is the previously implicated *TOMM40/APOE/APOC1* locus. In our extended GWAS, in individuals from all over Europe (**Chapter 4**), we confirmed the association of the *TOMM40/APOE/APOC1* locus with decreased survival to ages beyond 85 and 90 years (Table 8.1). In addition, we identified a novel locus that associates with an increased probability to survive to ages beyond 90 years (Table 8.1), which is located in an intergenic region on chromosome 5q33.3. As expected, prospective analysis showed that the minor allele of the lead SNP at the *TOMM40/APOE/APOC1* locus (rs4420638) associates with increased mortality, while the minor allele of the lead SNP at the chromosome 5q33.3 locus (rs2149954) associates with decreased mortality (Table 8.1). In **Chapter 3** we showed that the association at the *TOMM40/APOE/APOC1* locus is caused by the ApoE ε4 defining SNP rs429358, which has previously been associated with an increased risk of cardiovascular disease and Alzheimer's disease and unfavorable levels of several metabolic phenotypes, such as total/high-density lipoprotein/low-density lipoprotein cholesterol and C-reactive protein. We additionally show an effect of ApoE ε4 on insulin-like growth factor 1 (IGF-1) signaling in women. The locus on chromosome 5q33.3, on the other hand, has previously been associated with lower systolic and diastolic blood pressure in middle age. However, we showed that the association of the locus with decreased mortality above 75 years is not explained by its relation with blood pressure and, most likely, also involves other traits (**Chapter**

4). In the gene set analysis (**Chapter 5**) we showed that genetic variation in genes involved in the insulin/IGF-1 signaling (IIS) and telomere maintenance (TM) pathways is associated with human longevity. Hence, gene set analysis may be used, in addition to GWAS, to study the combined effect of genetic variation in (known) genes and pathways on longevity.

Since our genetic studies identified a limited number of longevity loci, we additionally examined whether leukocyte telomere length (LTL) could be used as a biomarker of healthy aging in genomic studies of large cohorts of middle-aged individuals (**Chapter 6**). We showed that LTL meets three of the four criteria for a biomarker of healthy aging in the Leiden Longevity Study (LLS), i.e., LTL changes with chronological age and is associated with prospective mortality and immune-related parameters (Table 8.2). However, this is still insufficient to use LTL as a standardized phenotype for genetic studies of healthy aging and longevity. Thus, we need to search for parameters that meet all four proposed criteria for biomarkers of healthy aging.

When novel longevity loci, such as the chromosome 5q33.3 locus, have been identified, functional characterization needs to be performed to determine the mechanism underlying the association with healthy aging and longevity. This process consist of several steps, namely (1) genotypic fine-mapping, (2) phenotypic fine-mapping, (3) expression/epigenetic quantitative trait locus analysis, and (4) functional assays in model systems (animals/cell models). Functional characterization of the chromosome 5q33.3 locus showed that the locus encompasses a

Table 8.1 Longevity-associated genetic variants identified through genome-wide association studies.

						Survival ≥ 90 years		Mortality		
Locus	Lead SNP	Chr	Position (bp)	Causal SNP(s)	Candidate / closest gene	OR	*P*	HR	*P*	Previously reported associations
5q33.3	rs2149954	2	157,753,180	Unknown	*EBF1*	1.10	1.74×10^{-8}	0.95	0.003	Blood pressure
19q13.32	rs4420638	19	50,114,786	rs429358 (ApoE ε4)	*APOE*	0.72	3.40×10^{-36}	1.07	0.019	Cardiovascular disease, Alzheimer's disease, metabolic phenotypes

Chr, chromosome according to NCBI Build 36; *Position (bp)*, position of the lead SNP according to NCBI Build 36; *OR*, odds ratio; *HR*, hazard ratio.

Table 8.2 Association of leukocyte telomere length with chronological age, prospective mortality, familial longevity and immune-related parameters in the Leiden Longevity Study (LLS) offspring and partners and LLS nonagenarians.

	LLS offspring + partners		LLS nonagenarians	
	β / HR	*P*	β / HR	*P*
Age (years)	-0.040	0.002	-0.044	0.003
Prospective mortality	0.75	0.001	0.92	0.028
Familial longevity	0.006	0.932	NA	NA
IGF-1/IGFBP3 (molar ratio)*	0.052	1.19×10^{-5}	-0.004	0.800
CRP (mg/L)**	-0.007	0.802	0.008	0.821
IL-6 (pg/ml)	0.099	0.010	NA	NA
CMV infection	-0.168	0.005	-0.305	1.57×10^{-4}
Lymphocyte count (%)	-0.015	1.84×10^{-4}	-0.013	0.011
Neutrophil count (%)	0.014	1.51×10^{-4}	0.013	0.006
Monocyte count (%)	-0.020	0.383	-0.016	0.502
Eosinophil count (%)	-0.007	0.727	-0.038	0.123
Basophil count (%)	0.353	3.63×10^{-5}	-0.263	0.019

HR, hazard ratio; *IGF-1*, insulin-like growth factor 1; *IGFBP3*, insulin-like growth factor binding protein 3; *CRP*, C-reactive protein; *IL-6*, interleukin 6; *CMV*, cytomegalovirus. The LTL outcome used for this analysis represents the number of 1 kb telomeric base pair units. *The outcome represents the effect of a 0.01 increase in the parameter. **Natural log transformed parameter was used in the analysis.

~22.3 kb region containing several functional elements, such as DNase I hypersensitivity sites, as well as a long intergenic noncoding RNA, which is conserved in primates. However, we have thus far not identified diseases or traits, additional to blood pressure, associating with the locus. In addition, it is still unclear on which gene(s) and in which tissue(s) the locus exerts its effects.

To identify novel longevity loci, future studies should use stricter criteria for selection of individuals for genetic research to reduce the phenotypic heterogeneity. In addition, genetic studies may use biomarkers of healthy aging, preferable combined into one multimarker score, as a standardized phenotype for genetic studies. Identification of such biomarkers may profit from combining family-based studies, like the LLS, and population-based prospective studies with long follow-up times, since this will allow testing of all four proposed criteria for biomarkers of healthy aging. Furthermore, research into human lifespan may benefit from novel technologies and methodologies, such as next-generation sequencing, multigenerational linkage analysis, omics-based measurements, and multimarker prediction scores. Subsequent integration of the created data over different species, or in a system biology approach, may provide insight into the complex mechanisms underlying lifespan regulation.

Chapter 9

Nederlandse samenvatting

Introductie

De afgelopen 200 jaar is de levensverwachting wereldwijd sterk toegenomen, wat voornamelijk komt door verbeterde hygiëne, voeding en gezondheidszorg. De toename in levensduur bestaat echter niet alleen uit gezonde jaren. We spenderen ongeveer 25% van ons leven in slechte gezondheid, voornamelijk veroorzaakt door leeftijd gerelateerde ziekten zoals kanker, hart- en vaatziekten en diabetes mellitus type II. Er zijn families waarin personen generaties achtereen langer leven in vergelijking tot personen uit hetzelfde geboortecohort. Op middelbare leeftijd ontwikkelen personen uit zulke langlevende families minder leeftijd gerelateerde ziekten en hebben ze daarnaast gezondere bloedspiegels voor vet- (cholesterol) en suikerhuishouding (glucose) en afweer (immuun) gerelateerde eigenschappen. Het lijkt er dus op dat personen uit langlevende families op een natuurlijke manier bescherming genieten tegen leeftijd gerelateerde ziekten, waarbij het met name gaat om aandoeningen die verband houden met veranderingen in de stofwisseling die optreden tijdens het ouder worden. Door het bestuderen van langlevende families hopen we mechanismen te kunnen vinden die bescherming bieden tegen leeftijd gerelateerde ziekten en daarmee bijdragen aan gezonde veroudering.

Om die mechanismen te ontrafelen, kunnen verschillende strategieën worden toegepast. Ongeveer 25% van de variatie in levensduur kan verklaard worden door genetische verschillen tussen personen (variatie), wat impliceert dat er een erfelijke basis is voor langlevendheid. Eén van de strategieën is dus om op zoek te gaan naar genetische varianten die samengaan met het bereiken van een hoge leeftijd, waarbij het erfelijk materiaal van groepen langlevende personen wordt vergeleken met dat van jongere controlegroepen. Omdat het aantal beschikbare studies met langlevende deelnemers echter beperkt is, kan men er ook voor kiezen om in veel grotere bevolkingsgroepen biomarkers voor gezonde veroudering te meten en vervolgens te kijken welke genetische varianten samengaan met deze biomarker. Omdat er nog geen betrouwbare biomarkers bekend zijn die langlevendheid voorspellen, is een deel van het onderzoek gericht op het identificeren van zulke biomarkers. Dat gebeurt door biologische eigenschappen en karakteristieken van fysiologische gezondheid (biologische leeftijd) van individuen te koppelen aan de leeftijd op basis van geboortedatum (kalenderleeftijd).

Verschillende soorten studie designs worden toegepast voor het genetisch en biomarker onderzoek naar gezonde veroudering en langlevendheid. Allereerst worden langlevende personen ($\geq$ 85 jaar) of hun kinderen (die door hun genetische achtergrond een verhoogde kans hebben om langlevend te worden) vergeleken met jonge ($<$ 65 jaar), nog levende, personen uit niet-langlevende families. Dit soort studies wordt ook wel cross-sectionele studies genoemd. Daarnaast worden langlevende personen of personen van middelbare leeftijd ($\geq$ 55 jaar) voor langere tijd gevolgd, idealiter tot hun overlijden, in zogeheten prospectieve studies. In dit proefschrift wordt gebruik gemaakt van diverse cross-sectionele en prospectieve studies.

Overzicht eerder genetisch onderzoek naar gezonde veroudering en langlevendheid

Hoofdstuk 2 geeft een overzicht van cross-sectionele studies waarin is gekeken naar de genetische basis van gezonde veroudering en langlevendheid. De meeste van deze studies hebben onderzoek gedaan naar DNA variaties, waarbij één bouwsteen van het DNA (nucleotide) bij een deel van de bevolking is vervangen door een andere. Dit soort genetische variaties worden ook wel 'single nucleotide polymorphisms (SNPs)' genoemd. Personen kunnen (1) de veel voorkomende variant van een SNP dragen op beide chromosomen, (2) de veel voorkomende variant van een SNP dragen op één chromosoom en de zeldzame variant op de andere of (3) de zeldzame variant van een SNP dragen op beide chromosomen. Als de zeldzame variant van de SNP vaker voorkomt in langlevende personen (cases) in vergelijking tot jonge personen (controles), dan zeggen we dat deze SNP samengaat (geassocieerd is) met langlevendheid. Komt de zeldzame variant juist minder vaak voor in cases dan in controles dan is de SNP mogelijk geassocieerd met mortaliteit.

Uit eerdere studies kwam naar voren dat verschillende isovormen van het apolipoproteïne E (ApoE) eiwit associëren met langlevendheid. Dit eiwit speelt onder andere een rol bij cholesterol huishouding, cognitief functioneren en het ontstaan van hart- en vaatziekten. Er bestaan drie isovormen van dit eiwit (ApoE ε2, ApoE ε3 en ApoE ε4). De ApoE ε4 isovorm is geassocieerd met een verlaagde kans om oud te worden, terwijl de ApoE ε2 isovorm juist geassocieerd is met een verhoogde kans om oud te worden. Uit genetische studies bleek vervolgens dat deze isovormen bepaald worden door twee variaties (SNPs) in het *APOE* gen, rs7412 (ApoE ε2) en rs429358 (ApoE ε4).

Naast de studies naar variatie in *APOE* waren de eerste genetische studies naar langlevendheid vooral gericht op de menselijke versies (homologen) van genen die na manipulatie in wormen, fruitvliegen en/of knaagdieren (muizen/ratten) leiden tot een verlengde levensduur. Voor één zo'n 'kandidaatgen', namelijk *FOXO3A*, bleek in meerdere onafhankelijke studies dat SNPs in dit gen ook samengaan met humane langlevendheid, vooral bij heel oude mensen. *FOXO3A* codeert voor het eiwit forkhead box O3, dat de transcriptie reguleert van genen die onder andere betrokken zijn bij geprogrammeerde celdood (apoptose).

Genoomwijde associatie studies voor langlevendheid

Naast het bestuderen van 'kandidaatgenen' werden er ook genoomwijde associatie studies (GWAS) naar langlevendheid verricht om in het hele genoom genen op te sporen die de menselijke levensverwachting beïnvloeden. In zo'n GWAS wordt voor 300.000 tot 2.500.000 SNPs, verspreid over alle chromosomen, getest of de zeldzame variant vaker, of juist minder vaak, voorkomt bij langlevende cases in vergelijking met jongere controles.

In **hoofdstuk 3** werd dit onderzocht voor ongeveer 520.000 SNPs in negentigjarigen cases uit de Leiden

LangLeven Studie en controles (< 60 jaar) uit de Rotterdam Studie. Vervolgens werden de 62 SNPs waarvan de zeldzame variant vaker, of juist minder vaak, voorkwam in de cases ook bestudeerd in een groep langlevende cases (≥ 85 jaar) uit de Rotterdam Studie, Leiden 85-plus studie en het Deense 1905 cohort en controles (< 65 jaar) uit het Nederlands Tweeling Register, Deense Tweeling Register en, wederom, de Rotterdam Studie. In de gecombineerde analyse van al deze studies (4.149 cases ≥ 85 jaar en 7.582 controles < 65 jaar) bleek er één SNP te zijn, rs2075650, waarvan de zeldzame variant minder vaak voorkomt in langlevende cases. Dragers van de zeldzame variant van rs2075650 bleken een 29% lagere kans te hebben om 85 jaar te worden en bleken na de 85 jaar nog steeds sneller te overlijden in vergelijking met niet-dragers. De SNP is gepositioneerd op de lange arm van chromosoom 19 in het gen *TOMM40*, direct naast het *APOE* gen. Na bestudering van deze SNP bleek echter dat het effect op langlevendheid eigenlijk bepaald wordt door de al bekende SNP rs429358 (ε4) in *APOE*, die niet gemeten was in deze GWAS.

Omdat er in **hoofdstuk 3** voor slechts één genetische variant bewijs werd gevonden voor associatie met langlevendheid, rees het vermoeden dat van andere genetische varianten die samengaan met langlevendheid het effect kleiner is. Om zulke genetische varianten te kunnen detecteren, hebben we daarom ook een groter aantal SNPs (~2.500.000) bestudeerd in een veel grotere groep cases (≥ 85 jaar) en controles (< 65 jaar) uit zesentwintig studies verspreid over Europa (**hoofdstuk 4**). Uiteindelijk bleek uit de gecombineerde analyse van al deze studies (20.789 cases en 77.277 controles) dat er een nieuwe SNP was, rs2149954, die associeerde met langlevendheid. Deze SNP is gepositioneerd op de lange arm van chromosoom 5 onder een zogeheten 'lincRNA', een stukje RNA dat andere genen aanstuurt. Dragers van de zeldzame variant van rs2149954 bleken een 10% hogere kans te hebben om 90 jaar te worden en bleken daarnaast minder snel te overlijden in vergelijking met niet-dragers. Uit eerder gepubliceerd onderzoek bleek dat het dragen van deze variant is geassocieerd met een verlaagde bloeddruk op middelbare leeftijd. Dit benadrukt het belang van het vermijden van een hoge bloeddruk op middelbare leeftijd voor het bereiken van een hoge leeftijd, bijvoorbeeld door middel van een gezonde leefstijl of medicatie. Het is nog niet duidelijk welke fysiologische processen de variant verder nog beïnvloedt.

Ondanks dat we twee SNPs gevonden hebben die van invloed lijken te zijn op langlevendheid is dit aantal vrij klein vergeleken met studies van vergelijkbare grootte naar erfelijke oorzaken van leeftijd gerelateerde ziekten als artrose en de ziekte van Alzheimer. Dit zou kunnen komen doordat slechts een klein deel van de levensduur bepaald wordt door de genetische achtergrond van een individu. Omgevingsfactoren hebben in feite een grotere invloed op de levensduur dan erfelijke factoren. De sterk toegenomen levensverwachting veroorzaakt door verbeterde hygiëne, voeding en medische zorg heeft er toe geleid dat er ook veel personen zijn die een hoge leeftijd bereiken zonder dat dit direct wordt veroorzaakt door hun genetische achtergrond. Daarnaast

lijkt het effect van genetische variatie op langlevendheid veroorzaakt te worden door een groot aantal varianten met kleine effecten. Om het effect van genetische variatie op langlevendheid beter te kunnen bestuderen, zouden daarom nog meer langlevende personen onderzocht moeten worden, met name uit langlevende families waarin de invloed van erfelijke factoren groter is dan in de algemene bevolking. Het aantal langlevende personen dat voor een GWAS in aanmerking komt, is echter beperkt. Er zal dus gezocht moeten worden naar andere methoden om nieuwe genetische varianten voor langlevendheid te vinden.

Genetische variatie in genen betrokken bij insuline/IGF-1 signaaltransductie en regulering van telomeerlengte

Zoals uit **hoofdstuk 3** en **4** blijkt, is het effect van genetische varianten die samengaan met langlevendheid waarschijnlijk te klein om te kunnen onderzoeken met behulp van GWAS. Een mogelijkheid om deze kleine effecten toch te kunnen detecteren, is door het gezamenlijk effect van SNPs gegroepeerd per gen of proces te bestuderen. In **hoofdstuk 5** werd deze methode toegepast om te kijken naar het gecombineerde effect van SNPs in 'kandidaatgenen' die betrokken zijn bij twee processen waarvan wordt verondersteld dat ze een rol spelen in langlevendheid. Het eerste proces is insuline/insulin-like growth factor 1 (IGF-1) signaaltransductie (IIS), dat betrokken is bij groei en stofwisseling. Het tweede proces is de regulering van de lengte van telomeren (TM). Telomeren zijn de uiteinden van chromosomen die er voor zorgen dat het DNA beschermd wordt. Met iedere celdeling neemt de lengte van het telomeer af en op het moment dat de lengte te kort wordt, stopt de cel met delen en gaat hij uiteindelijk dood. Indien de lengte van een telomeer constant kan worden gehouden, kan een cel in potentie dus vaker delen en langer leven of zich tot een kankercel ontpoppen. Uit de analyse, waarbij we gebruik hebben gemaakt van de dataset zoals beschreven in **hoofdstuk 3**, bleek inderdaad dat genetische variatie in beide processen van invloed lijkt te zijn op langlevendheid. Dit lijkt vooral veroorzaakt te worden door gecombineerde genetische variatie in de genen *AKT1*, *AKT3*, *FOXO4*, *IGF2*, *INS*, *PIK3CA*, *SGK*, *SGK2*, *YWHAG* (IIS) en *POT1* (TM). Analyse in grotere datasets met gebruik van meer geavanceerde methoden zal moeten uitwijzen of de complexiteit van langlevendheid kan worden ontrafeld door het onderzoeken van het gecombineerde effect van SNPs met kleine effecten.

Biomarkers voor gezonde veroudering

Zoals bediscussieerd in **hoofdstuk 3** en **4** zouden genetische varianten met heel kleine effecten op langlevendheid mogelijk nog gevonden kunnen worden in grotere genetische studies, maar helaas is het aantal beschikbare langlevende personen voor dit soort studies beperkt. Een mogelijkheid om toch grotere genetische studies te kunnen uitvoeren, is door het gebruik van een biomarker voor gezonde veroudering die gemeten kan worden in grote populatie

studies. In **hoofdstuk 2** worden vier criteria beschreven waaraan een dergelijke biomarker voor gezonde veroudering in onze ogen zou moeten voldoen. Ten eerste moet de waarde van de biomarker toe- of afnemen met de kalenderleeftijd. Ten tweede moet de biomarker onderscheid kunnen maken tussen personen op basis van hun biologische leeftijd of genetische aanleg voor langlevendheid. Ten derde moet de biomarker gerelateerd zijn aan eigenschappen die de verandering in gezondheid van een individu weergeven. Ten slotte moet de biomarker van invloed zijn op het risico op ziekten en mortaliteit in prospectieve studies.

Voor het testen van de verschillende criteria voor een biomarker voor gezonde veroudering zou het ideaal zijn om een studie te hebben waarin meerdere generaties familieleden van een groot aantal personen uit de bevolking voor langere tijd gevolgd wordt, idealiter tot aan hun dood. Een voorbeeld van zo'n studie is LifeLines (https://lifelines.nl/). Ondanks dat deze studie al een paar jaar geleden begonnen is, zal het echter nog vele jaren duren voordat de data gebruikt kunnen worden voor biomarker onderzoek. Het beste alternatief voor biomarker studies is daarom om cross-sectionele studies (het liefst gebaseerd op families) te combineren met grote populatie-gebaseerde prospectieve studies, zoals het Erasmus Rotterdam Gezondheid Onderzoek. Voor de biomarker studie beschreven in dit proefschrift is gebruik gemaakt van de Leiden LangLeven Studie. Deze studie bestaat uit langlevende personen (> 90 jaar), hun kinderen en de partners van die kinderen die inmiddels gedurende ongeveer 10 jaar gevolgd zijn. Aangezien de Leiden LangLeven Studie een combinatie is van een cross-sectionele en prospectieve studie kunnen de meeste criteria voor een biomarker voor gezonde veroudering getest worden.

Studie naar telomeerlengte als biomarker voor gezonde veroudering

Een potentiële biomarker voor gezonde veroudering is de lengte van telomeren gemeten in leukocyten (LTL), een bepaald type bloedcellen. In eerdere studies is gevonden dat LTL afneemt met kalenderleeftijd en dat verkorte LTL samengaat met slechtere waardes voor gezondheid gerelateerde eigenschappen en een verhoogde kans op ziekten en overlijden in prospectieve studies. Om te kijken of LTL als biomarker voor gezonde veroudering zou kunnen worden gebruikt, hebben wij getest of LTL ook aan het tweede criterium voor een biomarker voldoet (**hoofdstuk 6**). Met andere woorden, of LTL ook onderscheid kan maken tussen personen op basis van hun biologische leeftijd of genetische aanleg voor langlevendheid. Hiervoor werd LTL van langlevende personen uit de Leiden LangLeven Studie bestudeerd, evenals van hun kinderen en de partners van die kinderen. Zoals verwacht is LTL in personen uit de Leiden LangLeven Studie inderdaad gerelateerd aan de kalenderleeftijd, oudere mensen hebben een verkorte LTL. Tevens is een verkorte LTL geassocieerd met ongunstige bloedspiegels voor immuun gerelateerde eigenschappen en een verhoogde kans op overlijden. Echter, LTL is niet verschillend tussen de kinderen van

de langlevende personen en hun partners. Dit geeft aan dat LTL geen onderscheid kan maken tussen personen op basis van hun genetische aanleg voor langlevendheid. Verder bleek dat SNPs die van invloed lijken te zijn op LTL niet associëren met langlevendheid in de GWAS beschreven in **hoofdstuk 4**. Dus ondanks dat LTL aan drie van de vier criteria voor een biomarker van gezonde veroudering voldoet, lijkt dit onvoldoende te zijn om LTL als voorspeller van langlevendheid te kunnen gebruiken in grote genetische studies bij personen van middelbare leeftijd. Er zal dus gezocht moeten worden naar eigenschappen die aan alle vier de criteria voldoen.

Van genetische variant naar mechanisme

Nadat nieuwe genetische varianten voor langlevendheid zijn gevonden, zoals rs2149954 op chromosoom 5, moet worden bepaald hoe deze varianten gezonde veroudering en langlevendheid beïnvloeden. **Hoofdstuk 7** geeft een overzicht van de stappen die hiervoor nodig zijn. Allereerst zal worden gekeken welke genetische variant in de buurt van de gevonden SNP de causale variant is, oftewel de variant die daadwerkelijk het effect op langlevendheid veroorzaakt. Hiervoor zal de complete DNA volgorde in het gebied worden bepaald en zullen de zogenaamde 'epigenetische' signalen worden bekeken. Tevens zal bekeken worden voor welke 'kandidaatgenen' de expressie in gekweekte cellen door de variant wordt beïnvloedt. Verder zal in andere epidemiologische studies onderzocht worden of de causale variant gerelateerd is aan andere eigenschappen of ziekten om te ontdekken welke processen worden beïnvloed door deze variant. Daarnaast zal onderzocht worden hoe veranderingen in de expressie van de 'kandidaatgenen' bijdragen aan het ontstaan van ziekten en de kans op overlijden beïnvloeden door gebruik te maken van diermodellen.

Conclusie en toekomst

Het onderzoek beschreven in dit proefschrift laat zien dat het lastig is om genetische varianten te ontdekken die invloed hebben op langlevendheid. Desondanks hebben wij een nieuwe SNP ontdekt op chromosoom 5 waarvan de zeldzame variant associeert met een verhoogde kans om 90 jaar te worden en een verlaagde kans om te overlijden en op middelbare leeftijd bijdraagt aan een lage bloeddruk. Om nieuwe genetische varianten te vinden die samengaan met langlevendheid worden in **hoofdstuk 7** enkele mogelijkheden voor toekomstig onderzoek gegeven. Allereerst moeten er betere criteria komen voor het selecteren van personen voor genetische studies naar langlevendheid waardoor personen waarvan duidelijk is dat hun hoge leeftijd voornamelijk veroorzaakt wordt door omgevingsfactoren geëxcludeerd zullen worden. Daarnaast kunnen toekomstige genetische studies gebruik maken van biomarkers voor gezonde veroudering, het liefst samengevoegd tot een patroon gebaseerd op meerdere eigenschappen. Het combineren van verschillende soorten cross-sectionele familie studies, zoals de

Leiden LangLeven Studie, en prospectieve studies kan helpen bij het zoeken naar zulke biomarkers, aangezien we op deze manier alle vier de criteria voor een biomarker kunnen testen. In dit proefschrift hebben wij aan de hand van LTL laten zien hoe de vier criteria voor een potentiële biomarker getest kunnen worden. LTL bleek echter niet verschillend te zijn tussen kinderen van de langlevende personen en hun partners, waardoor LTL niet als biomarker voor langlevendheid kan worden gebruikt. Tot slot kan toekomstig onderzoek naar gezondere veroudering en langlevendheid gebruik maken van nieuwe technologieën waarbij de nucleotidevolgorde van het hele genoom van een individu wordt bepaald, zodat de relevante genetische variatie kan worden onderzocht.

Appendix

List of publications
Curriculum Vitae
Dankwoord

List of publications

1. Slagboom PE, Beekman M, Passtoors WM, **Deelen J**, Vaarhorst AA, Boer JM, van den Akker EB, van Heemst D, de Craen AJ, Maier AB, Rozing M, Mooijaart SP, Heijmans BT, Westendorp RG. Genomics of human longevity. *Philos Trans R Soc Lond B Biol Sci* 2011; **366**: 35-42.

2. **Deelen J**, Beekman M, Uh HW, Helmer Q, Kuningas M, Christiansen L, Kremer D, van der Breggen R, Suchiman HE, Lakenberg N, van den Akker EB, Passtoors WM, Tiemeier H, van Heemst D, de Craen AJ, Rivadeneira F, de Geus EJ, Perola M, van der Ouderaa FJ, Gunn DA, Boomsma DI, Uitterlinden AG, Christensen K, van Duijn CM, Heijmans BT, Houwing-Duistermaat JJ, Westendorp RG, Slagboom PE. Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. *Aging Cell* 2011; **10**: 686-98.

3. Sampietro ML, Trompet S, Verschuren JJ, Talens RP, **Deelen J**, Heijmans BT, de Winter RJ, Tio RA, Doevendans PA, Ganesh SK, Nabel EG, Westra HJ, Franke L, van den Akker EB, Westendorp RG, Zwinderman AH, Kastrati A, Koch W, Slagboom PE, de Knijff P, Jukema JW. A genome-wide association study identifies a region at chromosome 12 as a potential susceptibility locus for restenosis after percutaneous coronary intervention. *Hum Mol Genet* 2011; **20**: 4748-57.

4. Uh HW, **Deelen J**, Beekman M, Helmer Q, Rivadeneira F, Hottenga JJ, Boomsma DI, Hofman A, Uitterlinden AG, Slagboom PE, Böhringer S, Houwing-Duistermaat JJ. How to deal with the early GWAS data when imputing and combining different arrays is necessary. *Eur J Hum Genet* 2012; **20**: 572-6.

5. Passtoors WM, Boer JM, Goeman JJ, van den Akker EB, **Deelen J**, Zwaan BJ, Scarborough A, van der Breggen R, Vossen RH, Houwing-Duistermaat JJ, van Ommen GJ, Westendorp RG, van Heemst D, de Craen AJ, White AJ, Gunn DA, Beekman M, Slagboom PE. Transcriptional profiling of human familial longevity indicates a role for ASF1A and IL7R. *PLoS One* 2012; **7**: e27759.

6. Huffman DM, **Deelen J**, Ye K, Bergman A, Slagboom PE, Barzilai N, Atzmon G. Distinguishing between longevity and buffered-deleterious genotypes for exceptional human longevity: the case of the MTP gene. *J Gerontol A Biol Sci Med Sci* 2012; **67**: 1153-60.

7. **Deelen J**, Uh HW, Monajemi R, van Heemst D, Thijssen PE, Böhringer S, van den Akker EB, de Craen AJ, Rivadeneira F, Uitterlinden AG, Westendorp RG, Goeman JJ,

Slagboom PE, Houwing-Duistermaat JJ, Beekman M. 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**: 235-49.

8. Passtoors WM, Beekman M, **Deelen J**, van der Breggen R, Maier AB, Guigas B, Derhovanessian E, van Heemst D, de Craen AJ, Gunn DA, Pawelec G, Slagboom PE. Gene expression analysis of mTOR pathway: association with human longevity. *Aging Cell* 2013; **12**: 24-31.

9. Verschuren JJW, Trompet S, **Deelen J**, Stott DJ, Sattar N, Buckley BM, Ford I, Heijmans BT, Guchelaar HJ, Houwing-Duistermaat JJ, Slagboom PE, Jukema JW. Non-Homologous End-Joining Pathway Associated with Occurrence of Myocardial Infarction: Gene Set Analysis of Genome-Wide Association Study Data. *PLoS One* 2013; **8**: e56262.

10. Porcu E, Medici M, Pistis G, Volpato CB, Wilson SG, Cappola AR, Bos SD, **Deelen J**, den Heijer M, Freathy RM, Lahti J, Liu C, Lopez LM, Nolte IM, O'Connell JR, Tanaka T, Trompet S, Arnold A, Bandinelli S, Beekman M, Böhringer S, Brown SJ, Buckley BM, Camaschella C, de Craen AJ, Davies G, de Visser MC, Ford I, Forsen T, Frayling TM, Fugazzola L, Gögele M, Hattersley AT, Hermus AR, Hofman A, Houwing-Duistermaat JJ, Jensen RA, Kajantie E, Kloppenburg M, Lim EM, Masciullo C, Mariotti S, Minelli C, Mitchell BD, Nagaraja R, Netea-Maier RT, Palotie A, Persani L, Piras MG, Psaty BM, Räikkönen K, Richards JB, Rivadeneira F, Sala C, Sabra MM, Sattar N, Shields BM, Soranzo N, Starr JM, Stott DJ, Sweep FC, Usala G, van der Klauw MM, van Heemst D, van Mullem A, H Vermeulen S, Visser WE, Walsh JP, Westendorp RG, Widen E, Zhai G, Cucca F, Deary IJ, Eriksson JG, Ferrucci L, Fox CS, Jukema JW, Kiemeney LA, Pramstaller PP, Schlessinger D, Shuldiner AR, Slagboom PE, Uitterlinden AG, Vaidya B, Visser TJ, Wolffenbuttel BH, Meulenbelt I, Rotter JI, Spector TD, Hicks AA, Toniolo D, Sanna S, Peeters RP, Naitza S. A meta-analysis of thyroid-related traits reveals novel Loci and gender-specific differences in the regulation of thyroid function. *PLoS Genet* 2013; **9**: e1003266.

11. Beekman M, Blanché H, Perola M, Hervonen A, Bezrukov V, Sikora E, Flachsbart F, Christiansen L, de Craen AJ, Kirkwood TB, Rea IM, Poulain M, Robine JM, Valensin S, Stazi MA, Passarino G, Deiana L, Gonos ES, Paternoster L, Sørensen TI, Tan Q, Helmer Q, van den Akker EB, **Deelen J**, Martella F, Cordell HJ, Ayers KL, Vaupel JW, Törnwall O, Johnson TE, Schreiber S, Lathrop M, Skytthe A, Westendorp RG, Christensen K, Gampe J, Nebel A, Houwing-Duistermaat JJ, Slagboom PE, Franceschi C; the GEHA consortium. Genome-wide linkage analysis for human longevity: Genetics of Healthy Aging Study. *Aging Cell* 2013; **12**: 184-93.

12. **Deelen J**, Beekman M, Capri M, Franceschi C, Slagboom PE. Identifying the genomic determinants of aging and longevity in human population studies: Progress and challenges. *Bioessays* 2013; **35**: 386-96.

13. Codd V, Nelson CP, Albrecht E, Mangino M, **Deelen J**, Buxton JL, Hottenga JJ, Fischer K, Esko T, Surakka I, Broer L, Nyholt DR, Mateo Leach I, Salo P, Hägg S, Matthews MK, Palmen J, Norata GD, O'Reilly PF, Saleheen D, Amin N, Balmforth AJ, Beekman M, de Boer RA, Böhringer S, Braund PS, Burton PR, de Craen AJ, Denniff M, Dong Y, Douroudis K, Dubinina E, Eriksson JG, Garlaschelli K, Guo D, Hartikainen AL, Henders AK, Houwing-Duistermaat JJ, Kananen L, Karssen LC, Kettunen J, Klopp N, Lagou V, van Leeuwen EM, Madden PA, Mägi R, Magnusson PK, Männistö S, McCarthy MI, Medland SE, Mihailov E, Montgomery GW, Oostra BA, Palotie A, Peters A, Pollard H, Pouta A, Prokopenko I, Ripatti S, Salomaa V, Suchiman HE, Valdes AM, Verweij N, Viñuela A, Wang X, Wichmann HE, Widen E, Willemsen G, Wright MJ, Xia K, Xiao X, van Veldhuisen DJ, Catapano AL, Tobin MD, Hall AS, Blakemore AI, van Gilst WH, Zhu H, CARDIoGRAM consortium, Erdmann J, Reilly MP, Kathiresan S, Schunkert H, Talmud PJ, Pedersen NL, Perola M, Ouwehand W, Kaprio J, Martin NG, van Duijn CM, Hovatta I, Gieger C, Metspalu A, Boomsma DI, Jarvelin MR, Slagboom PE, Thompson JR, Spector TD, van der Harst P, Samani NJ. Identification of seven loci affecting mean telomere length and their association with disease. *Nat Genet* 2013; **45**: 422-7.

14. Broer L, Codd V, Nyholt DR, **Deelen J**, Mangino M, Willemsen G, Albrecht E, Amin N, Beekman M, de Geus EJ, Henders A, Nelson CP, Steves CJ, Wright MJ, de Craen AJ, Isaacs A, Matthews M, Moayyeri A, Montgomery GW, Oostra BA, Vink JM, Spector TD, Slagboom PE, Martin NG, Samani NJ, van Duijn CM, Boomsma DI. 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; **21**: 1163-8.

15. 't Hart LM, Fritsche A, Nijpels G, van Leeuwen N, Donnelly LA, Dekker JM, Alssema M, Fadista J, Carlotti F, Gjesing AP, Palmer CN, van Haeften TW, Herzberg-Schäfer SA, Simonis-Bik AM, Houwing-Duistermaat JJ, Helmer Q, **Deelen J**, Guigas B, Hansen T, Machicao F, Willemsen G, Heine RJ, Kramer MH, Holst JJ, de Koning EJ, Häring HU, Pedersen O, Groop L, de Geus EJ, Slagboom PE, Boomsma DI, Eekhoff EM, Pearson ER, Diamant M. The CTRB1/2 locus affects diabetes susceptibility and treatment via the incretin pathway. *Diabetes* 2013; **62**: 3275-81.

16. Ye K, Beekman M, Lameijer EW, Zhang Y, Moed MH, van den Akker EB, **Deelen J**, Houwing-Duistermaat JJ, Kremer D, Anvar SY, Laros JF, Jones D, Raine K, Blackburne B, Potluri S, Long Q, Guryev V, van der Breggen R, Westendorp RG, 't Hoen PA, den Dunnen J, van Ommen GJ, Willemsen G, Pitts SJ, Cox DR, Ning Z, Boomsma DI,

Slagboom PE. Aging as Accelerated Accumulation of Somatic Variants: Whole-Genome Sequencing of Centenarian and Middle-Aged Monozygotic Twin Pairs. *Twin Res Hum Genet* 2013; **16**: 1026-32.

17. **Deelen J**, Beekman M, Codd V, Trompet S, Broer L, Hägg S, Fischer K, Thijssen PE, Suchiman HE, Postmus I, Uitterlinden AG, Hofman A, de Craen AJ, Metspalu A, Pedersen NL, van Duijn CM, Jukema JW, Houwing-Duistermaat JJ, Samani NJ, Slagboom PE. Leukocyte telomere length associates with prospective mortality independent of immune-related parameters and known genetic markers. *Int J Epidemiol* 2014; *In press*.

18. van Leeuwen N, Beekman M, **Deelen J**, van den Akker EB, de Craen AJ, Slagboom PE, 't Hart LM. Low mitochondrial DNA content associates with familial longevity: the Leiden Longevity Study. *Age (Dordr)* 2014; *In press*.

19. **Deelen J**, Beekman M, Uh HW, Broer L, Ayers KL, Tan Q, Kamatani Y, Bennet AM, Tamm R, Trompet S, Guðbjartsson DF, Flachsbart F, Rose G, Viktorin A, Fischer K, Nygaard M, Cordell HJ, Crocco P, van den Akker EB, Böhringer S, Helmer Q, Nelson CP, Saunders GI, Alver M, Andersen-Ranberg K, Breen ME, van der Breggen R, Caliebe A, Capri M, Cevenini E, Collerton JC, Dato S, Davies K, Ford I, Gampe J, Garagnani P, de Geus EJ, Harrow J, van Heemst D, Heijmans BT, Heinsen FA, Hottenga JJ, Hofman A, Jeune B, Jonsson PV, Lathrop M, Lechner D, Martin-Ruiz C, McNerlan SE, Mihailov E, Montesanto A, Mooijaart SP, Murphy A, Nohr EA, Paternoster L, Postmus I, Rivadeneira F, Ross OA, Salvioli S, Sattar N, Schreiber S, Stefánsson H, Stott DJ, Tiemeier H, Uitterlinden AG, Westendorp RG, Willemsen G, Samani NJ, Galan P, Sørensen TI, Boomsma DI, Jukema JW, Rea IM, Passarino G, de Craen AJ, Christensen K, Nebel A, Stefánsson K, Metspalu A, Magnusson P, Blanché H, Christiansen L, Kirkwood TB, van Duijn CM, Franceschi C, Houwing-Duistermaat JJ, Slagboom PE. Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. *Hum Mol Genet* 2014; *In press*.

20. Guigas B, de Leeuw van Weenen JE, van Leeuwen N, Simonis-Bik AM, van Haeften TW, Nijpels G, Houwing-Duistermaat JJ, Beekman M, **Deelen J**, Havekes LM, Penninx BW, Vogelzangs N, van 't Riet E, Dehghan A, Hofman A, Witteman JC, Uitterlinden AG, Grarup N, Jørgensen T, Witte DR, Lauritzen T, Hansen T, Pedersen O, Hottenga J, Romijn JA, Diamant M, Kramer MH, Heine RJ, Willemsen G, Dekker JM, Eekhoff EM, Pijl H, de Geus EJ, Slagboom PE, 't Hart LM. Sex-specific effects of naturally occurring variants in the dopamine receptor D2 locus on insulin secretion and Type 2 diabetes susceptibility. *Diabet Med* 2014; *In press*.

Curriculum Vitae

Joris Deelen was born on the 1st of February 1985 in ‘s-Gravenhage. In 2003 he graduated at the Erasmus College in Zoetermeer and started the undergraduate program of Biomedical Sciences at Leiden University. His first internship, which was part of the Bachelor’s program, was at the Evolutionary Biology lab which is part of the Institute of Biology Leiden. During this first internship he studied the influence of different pre-adult and adult food conditions on longevity and fat content in *Drosophila melanogaster*. His second internship, which was part of the Master’s program, was at the department of Nephrology at the Leiden University Medical Center (LUMC). During this second internship he investigated the cross-talk between miRNAs expressed in primary endothelial cells and their predicted targets *GATA2* and *ERF*. His third and main internship, which was also part of the Master’s program, was at the department of Molecular Epidemiology at the LUMC. During this third internship he performed quantitative polymerase chain reaction experiments to validate oligonucleotide microarray data from the Leiden Longevity Study (LLS). In addition, he combined the results from a genome-wide association, linkage and transcriptomic analysis performed in samples from the LLS to identify genes associated with longevity. After graduating in October 2008, he started a PhD at the same department in November 2008 under supervision of Prof. Dr. Slagboom and Dr. Beekman. His PhD research was performed within the framework of the Netherlands Consortium for Healthy Ageing and was focused on finding genetic determinants for healthy ageing and longevity. The results of this research are outlined in this thesis. Currently he is employed as a post-doctoral researcher at the same department and he is involved in the Growing Old Together study. The main objective of this study is to determine the effect of a 3-month (13 weeks) intervention with 25% lowered energy expenditure (12.5% caloric restriction and 12.5% increased physical activity) on metabolic profiles in humans.

Dankwoord

Na ruim 5 jaar hard werken is hij hier dan eindelijk, mijn proefschrift! Natuurlijk had ik dit niet kunnen bereiken zonder hulp van anderen en het is mij daarom een genoegen om iedereen te bedanken die heeft bijgedragen aan de totstandkoming van dit boekje.

Allereerst wil ik hierbij mijn promotor, Professor Slagboom, bedanken voor de kans die zij mij geboden heeft om op haar afdeling te kunnen promoveren en natuurlijk voor de vele interessante discussies en inzichten die tot een aantal mooie publicaties hebben geleid die deel uitmaken van dit proefschrift. Ook mijn 2de promotor, Professor Westendorp, wil ik bedanken voor de, voornamelijk klinische en maatschappelijke, inzichten die hij mij gegeven heeft in het verouderingsonderzoek. Natuurlijk kan ik ook mijn co-promotor en dagelijks begeleider Dr. Beekman niet vergeten, met wie ik al die jaren heel erg fijn heb kunnen samenwerken. Beste Marian, dankzij jouw steun, vertrouwen en advies heb ik zowel mijn stage als mijn promotie tot een goed einde kunnen brengen.

En waar ben je als promovendus zonder goede analisten? Dennis, Eka, Nico, Ruud, Wesley en Joke, bedankt voor jullie hulp in het lab! Zonder jullie was er waarschijnlijk weinig terechtgekomen van alle metingen die beschreven staan in dit proefschrift. Eka wil ik hierbij nog in het speciaal bedanken voor het optreden als mijn "back-up" paranimf, waardoor mijn beide broers mij tijdens de plechtigheid kunnen flankeren. Ook heb ik gemerkt hoe belangrijk het is om goede secretaresses op de afdeling te hebben die altijd voor je klaar staan. Daarom wil ik hierbij Inge, Lieke en Lies bedanken voor alle administratieve hulp.

Verder wil ik ook alle overige collega's van de afdeling Moleculaire Epidemiologie (ook wel Malle Epi's genoemd) en de studenten die ik tijdens mijn promotie begeleid heb (Frans, Juliëtte, Madeline en Martijn) bedanken voor de interessante en gezellige discussies. Ik wil hier in het speciaal nog graag mijn dank betuigen aan Elmar (voor de vele interessante papers en de hulp bij de procedure die ertoe geleid heeft dat dit proefschrift nu in jullie handen ligt), Erik (voor alle hulp bij de analyses en het optreden als sparringpartner), Peter (voor de inzichten die hij mij gegeven heeft in het telomeeronderzoek) en Willemijn (voor alle hulp en begeleiding tijdens mijn stage).

Natuurlijk wil ik ook mijn gewaardeerde LUMC-collega's van de afdelingen Medische Statistiek en Ouderengeneeskunde bedanken voor de fijne samenwerking. Professor Houwing-Duistermaat, Hae-Won, Jelle, Quinta, Ramin en Stefan, bedankt voor al jullie hulp bij de statistische analyses die beschreven staan in dit proefschrift. Ik wil jullie met name bedanken voor het maken van de pipeline en het cc-assoc pakket zodat we alle GWAS data van de verschillende cohorten op dezelfde manier hebben kunnen analyseren. Professor Jukema, Dr.

de Craen, Diana, Iris en Stella, bedankt voor jullie feedback tijdens onze meetings en hulp bij de analyses van de LLS, Leiden 85-plus studie en PROSPER.

Daarnaast zijn er ook nog vele nationale en internationale collega's die ik niet kan vergeten te noemen. Met name aan de samenwerking met de Nederlandse collega's uit Rotterdam en Amsterdam heb ik veel te danken. Professors Uitterlinden, van Duijn en Boomsma, Dr. Rivadeneira, Dr. Willemsen, Linda en Maris bedankt voor al jullie hulp bij de analyses van de Rotterdam Studie en NTR. I am also very grateful for all the nice collaborations with my international colleagues. I would especially like to thank Professor Samani and Dr. Codd for their help with the telomere length measurements and all the IDEAL collaborators for their help with the EU longevity GWAS analyses.

Natuurlijk gaat een promotie ook gepaard met een hoop gezelligheid. Daarom wil ik hierbij ook nog graag Arjan, Angela, Kasper, Mark, Patrick, Steven en alle andere promovendi waar ik een gezellige tijd mee gehad heb tijdens de MGC workshops bedanken. Daarnaast wil ik al mijn basketbalvrienden, studievrienden, oud-huisgenoten en natuurlijk Ivan en Joris bedanken voor alle gezelligheid buiten het werk.

Ik wil graag afsluiten met het bedanken van mijn familie. Lieve papa, mama, opa en oma, bedankt voor alle kansen die jullie mij gegeven hebben en jullie onvoorwaardelijke steun. Natuurlijk kan ik ook mijn twee broers, Dirk en Giel, niet vergeten. Ondanks dat we in veel opzichten van elkaar verschillen heb ik altijd op jullie kunnen bouwen en ik ben dan ook blij dat jullie als mijn paranimfen zullen optreden.

Als laatste wil ik natuurlijk Kirsten bedanken, mijn allergrootste steun en toeverlaat. Ik kijk uit naar onze toekomst samen, die hopelijk nog lang mag gaan duren!