Genetics of healthy ageing

Figarska, Sylwia

Genetics of healthy ageing

Sylwia Figarska

"Genetics of healthy aging", PhD thesis of S. Figarska
Department of Epidemiology, University Medical Center Groningen, University of Groningen, the Netherlands

The work described in this thesis was funded by Stichting Astma Bestrijding (grant number 2005/020), Netherlands Asthma Foundation (grant number 3.2.02.51) and GUIDE, University Medical Center Groningen, the Netherlands.

ISBN: 978-90-367-6994-5 (print)
978-90-367-6993-8 (e-book)

Cover: S. Figarska & Lovebird design & printing solutions
Layout and printing:  Lovebird design & printing solutions
www.lovebird-design.com

university of groningen

Genetics of healthy ageing

PhD Thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. E. Sterken
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on
Monday 26 May 2014 at 12:45 hours

by

Sylwia Marta Figarska
born on 5 November 1985
in Głowno, Poland

Supervisor:

Prof. H.M.Boezen

Co-Supervisor:

Dr. J.M. Vonk

Assessment committee:

Prof. J. L. Kirkland

Prof. P.E. Slagboom

Prof. F. Kuipers

CONTENTS

Chapter 1

General introduction

Healthy ageing: How to define and how to achieve it?

Over the past century improvements in public health, nutrition, education, living conditions, and medicine led to an increase in life expectancy. For instance, of the cohort born in 1910 in the United States, fewer than 3% survived to the age of 90, whereas of those born in 2000 at least 15% of men and 20% of women are expected to reach 90 years (1). However, there are big variations in health status and levels of physical condition among older people of the same age (2). Current medicine is effective in preventing death from age-related diseases without delaying their onset, thus increasing the number of people with age-related diseases and increasing the number of diseases affecting each elderly person. In addition, each age-related disease is now treated separately, which is costly (3). The ideal solution is to prevent or delay age-related diseases, in other words to extend the healthy years of life, and by this decrease the ever-rising costs of the medical care.

A few years ago successful ageing was defined as an integral part of three components: a) low probability of disease or disability; b) high cognitive and physical functioning; and c) active engagement with life (4). Subjects of the same age vary in their susceptibility to age-related diseases due to their genetic background. Furthermore, studies in twins have shown that approximately 30% of the variation in human lifespan is genetically determined (5). So far the majority of the genetic contribution to ageing and its accompanying phenotypes is still unknown. Thus, a part of the research on ageing should be dedicated to identifying genes that cause frailty phenotypes in order to provide a therapy for subjects at risk. The ultimate goal of this research is to provide targets for prevention and treatment of disabilities and diseases that occur with increasing age.

Why do we age?

Senescence (a term that is derived from the Latin word senescere meaning "to grow old") is the process of accumulative changes to molecular and cellular structures that affect the metabolism and with time lead to deterioration and death. According to the Hayflick limit, introduced in 1961 by Dr. Leonard Hayflick, each cell in the human body has limited ability to divide due to shortening of telomeres with each cell cycle. This is called replicative senescence (6). However the primary cause for the accumulation of senescent cells in human tissues over time is probably another form of senescence, called cellular senescence. In cellular senescence a variety of stimuli like ultraviolet light, reactive oxygen species, chemotherapeutics, ionizing radiation, and distortion of chromatin structure cause an irreversible cell cycle arrest before cells lose their proliferative capacity (7). Senescence that occurs on the cellular level may eventually contribute to the ageing of the whole body, generating age-related phenotypes.

Age-related diseases and mortality

Aging is a complex process and death is the ultimate endpoint. Nonetheless old age is not a cause of death itself. Death is due to diseases, which are manifestations of advanced age. Two of the common age-related diseases are Chronic Obstructive Pulmonary Disease (COPD) and cardiovascular disease (CVD, i.e. all disorders affecting the heart and blood vessels, like atherosclerosis, myocardial infarction or cerebrovascular disease). These diseases are the leading causes of death worldwide. According to WHO, in 2011 nearly 17 million people died due to CVD and 3 million people

died due to COPD. Thus, more effort to uncover their underlying risk factors, preferably shared risk factors, is needed to develop and implement prevention strategies against both diseases at once. Although there are some general risk factors of mortality, including high serum cholesterol, high blood pressure and smoking (8), in this project we were especially interested in strong and well-established respiratory-related conditions that increase the mortality risk, such as low lung function level and dyspnea (breathlessness). Lung function decreases with age and this is part of normal ageing when structural changes occur in the respiratory system. However, irrespective of age, gender, and smoking history, impaired lung function (9) is a major predictor of all-cause and CVD and COPD mortality (10;11). Heritability of lung function level, measured by spirometry, is as high as 40% (12), and for that reason more research on the genetics of reduced lung function level is of great importance. Dyspnea is a respiratory burden present in 32% of people aged 70 years and over (13). This respiratory symptom deserves more attention especially since it is self-reported, thus easy to monitor and therefore a simply, low-cost screening tool. Dyspnea affects healthy ageing not only by worsening quality of life, but is a well-established risk factor of CVD and COPD mortality (14-18).

Genetic approaches

Mutations are changes in the nucleotide sequence of DNA and some of them are deleterious to health if they reduce the effectiveness of certain proteins (enzymes, hormones, etc.). A single nucleotide polymorphism (SNP) is a DNA sequence variation occurring when a single nucleotide (A, C, G or T) in the genome differs between members of a biological species (Figure 1). SNPs can alter the protein

product through amino acid substitution or the introduction of a nonsense/truncation mutation. SNPs can also affect the expression by interrupting a regulatory region or by interfering with normal splicing (19). The common variant that occurs at a given locus is called wild type allele. Subjects carrying two wild type alleles have the wild type genotype, whereas the other subjects are heterozygous or homozygous for the minor (mutant) allele.

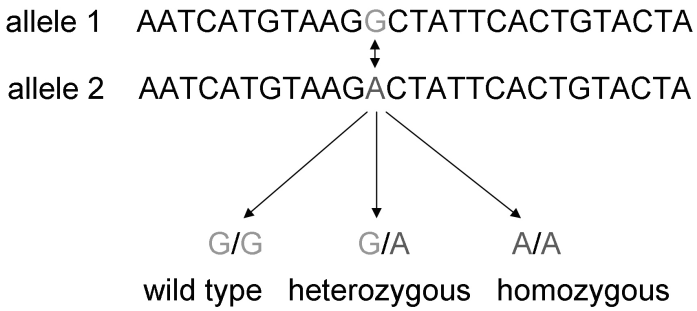


Figure 1 Single nucleotide polymorphism (SNP). Genotype is a combination of alleles, situated on the two corresponding chromosomes.

Human lifespan, as other complex traits, is a challenge for genetic studies. Two main approaches may be taken to investigate the genetic components related to human lifespan, i.e. a candidate gene study and a genome-wide association study (GWA study). Since both genetic approaches are based on association studies, it does not mean that SNPs found to be significant are indeed causative, they may only be in linkage disequilibrium with the actual causative variant. SNPs in candidate genes or SNPs identified in GWA studies may be also further investigated by combining them into a single score, so-called the Genetic Risk Score (GRS). Even if the individual genetic markers (SNPs) have only small effect and a borderline significant trend, the combined score could be a strong predictor of disease risk or level of a trait (20).

Candidate gene study

In the candidate gene study the association of SNPs in one gene or a few genes with the outcome of interest is investigated. The selection of genes is based on their a priori known biological function or physiological analogy with other disease, which makes them plausible candidates to be related to the outcome (21;22)

Over the past two decades, scientists have found that certain genes can play an important role in determining the rate of ageing in model organisms. Experiments in the most commonly studied models (budding yeast *Saccharomyces cerevisiae*, nematode worm *Caenorhabditis elegans*, fruit fly *Drosophila melanogaster* and the house mouse *Mus musculus*) have demonstrated that modifications in individual genes can dramatically increase lifespan (23-30). Human ageing is a very complex trait, and as such can be suspected to involve many genes, however so far most of them are still unknown. In The Ageing Gene Database (GenAge) (31) it can be seen that ageing-related human genes, tested as candidates, showed inconsistent associations across different human study populations. The most consistent evidence for association with longevity was found for the *Apolipoprotein E* (*APOE*) gene. The ε4 allele of the polymorphic *APOE* gene is associated with increased risk of Alzheimer's disease (AD), impaired cognitive function and cardiovascular complications, while the ε2 allele is protective (32-37). Another widely investigated and confirmed gene is the gene encoding a transcription factor called *forkhead box O3A* (*FOXO3A*), which is the human homolog of *daf-16* modulating lifespan in worms (38-41). Many other candidates from plausible pathways failed to replicate (42).

Candidate genes in this thesis

SIRT1

Sirtuins comprise an evolutionary conserved family of Silent Information Regulator 2 (*Sir2*) NAD-dependent histone deacetylases. Overexpression of *Sir2* has been shown to extend lifespan in model organisms (43-45). *SIRT1*, the closest human homolog of *S.cerevisiae Sir2* (46), has a wide range of substrates, including p53, NF-κB, FOXO transcription factors, and PGC-1α, and thus plays a role in multiple cellular processes ranging from energy metabolism to cell survival (47). Therefore *SIRT1* is an important candidate gene for (healthy) ageing and is a prominent therapeutic target.

ADAM33

A Desintegrin and Metalloproteinase (ADAM) is a family of membrane-anchored proteins implicated in cell-cell interactions, cell fusion, and cell signaling (48). *ADAM33*, that encodes a member of this family, was identified in 2002 as an asthma susceptibility gene (49). Since then it has been further investigated in relation to lung function level and decline and COPD (50-53), but so far the exact substrates cleaved by ADAM33 remain unknown (54). *ADAM33* is considered a gene associated with airway remodeling (55), however evidence linking it to the severity of atherosclerosis (56), suggests that *ADAM33*, probably due to pro-inflammatory properties, has an extensive function in human age-related diseases.

NFE2L2

Nuclear Factor (Erythroid-derived 2)-Like 2 (*NFE2L2* or *NRF2*) is a gene encoding a basic leucine zipper transcription factor that governs antioxidant and inflammatory responses (57). Given a broad role of *NFE2L2* which is the transcription activation of more than 200 genes crucial in the metabolism of drugs and toxins, protection against oxidative stress and inflammation, this gene is considered a guardian of health, protecting against age-related diseases (58). *NFE2L2* may be important in lifespan variation, since this gene was 6-fold overexpressed in the liver of the naked mole rats, a species with a relatively very long lifespan as compared to mice having a much shorter lifespan. Furthermore, silencing of *NFE2L2* leads to premature senescence in human fibroblasts, whereas treatment with the *NFE2L2* inducer leads to enhanced cell survival under oxidative stress conditions (59).

A genome-wide association study (GWA study)

GWA study, which is hypothesis free, is a second approach to perform. GWA study is based on genotyping a huge number of SNPs (i.e. markers) across the genome, in many individuals, to test whether they are associated with a trait. Since there is linkage disequilibrium not all the SNPs need to be genotyped; each participant is tested for selected markers of genetic variation that provide dense genome-wide coverage. If the SNP is related to the phenotype the minor allele frequency (MAF) will be different between cases and controls (dichotomous traits) or there will be an association between the level of the trait and the number of risk alleles present (quantitative traits). Since the number of tests

performed in such a study is large there is a chance of false positive findings. To reduce this risk multiple testing correction is applied, usually the Bonferroni correction, which determines the significance cut-off by simply dividing the nominal level of significance (α=0.05) by the number of tests performed. Additionally a replication study in independent samples should be performed, where the subset of significant SNPs is genotyped, to confirm the original findings of the identification cohort in other cohorts and/or populations, so called replication cohorts/populations.

The GWA studies on longevity were not successful to identify any additional autosomal susceptibility genes, they just confirmed the *APOE*-longevity association (60-62). In complex traits as ageing there are multiple loci with moderate effects and GWA study may have limited statistical power to detect them. In a meta-analysis of nine GWA studies performed on all-cause mortality no SNP was a genome-wide significant predictor, however SNPs that predict risk of death at $p<10^{-5}$ were found which fall in or near genes highly expressed in the brain (*HECW2, HIP1, BIN2*), genes involved in neural development and function (*KCNQ4, LMO4, NETO1*) and autophagy (*ATG4C*) (63). Thus GWA study may be a good tool to point out promising signals, however further functional studies may be needed to confirm the relevance of the findings (64).

Study population

All studies presented in this thesis were performed within the Vlagtwedde-Vlaardingen study. The Vlagtwedde-Vlaardingen study is based on a general population cohort of exclusively Caucasian individuals of Dutch descent, recruited from Vlagtwedde, a rural area,

Figure 2 Location of Vlagtwedde (the rural area) and Vlaardingen (the urban area) in the Netherlands

and Vlaardingen, an urban area in the Netherlands (see Figure 2). This study started in 1965, and participants had medical exams every 3 years until the last survey in 1989/1990. The Committee on Human Subjects in Research of the University of Groningen reviewed the study and affirmed the safety of the protocol and study design and specifically approved this study. All participants gave their written informed consent. At all surveys information on area of residence, age, sex, smoking habits and respiratory symptoms was collected by the Dutch version of the UK Medical Research Council standard questionnaire, and spirometry was performed. At the last survey (1989/1990) blood samples were collected and stored at –20°C. In 2003–2004 DNA was extracted from these samples and used for genotyping. In 1970/72, 1973/75 and 1976 an additional study was performed in males only by cardiologists with extra measurements of glucose tolerance, cholesterol and triglyceride levels. The vital status of all participants in the Vlagtwedde-Vlaardingen study on December 31, 2008 was assessed. Primary and secondary causes of death were obtained from Statistics Netherlands (CBS) and categorized using the International Classification of Diseases codes, version 9 and 10 (ICD-9 and ICD-10). The longitudinal design of the study gives us the unique opportunity to investigate the associations between polymorphisms in candidate genes and the whole genome (GWA study) and all-cause and cause-specific mortality.

Aim of the thesis

The general aim of this thesis was to investigate genetic background of ageing, defined as the risk of all-cause mortality. We were especially interested in identifying pleiotropic genes, i.e. genes that may affect human lifespan by influencing the risk of more than one disease. Since COPD and CVD are age-related diseases we focused also on mortality due to these disorders. Moreover, we studied genetics of lung function level and relation between mortality and dyspnea, a common respiratory symptom in elderly.

The thesis consists of eight chapters. In **Chapter 1** a general introduction is given.

In **Chapter 2** we assessed the associations of the *SIRT1* gene variants with long-term survival in the general population. In the light of recent doubts about the role of sirtuins in longevity, we support the importance of *SIRT1* in ageing.

Chapter 3 describes the *ADAM33* gene as a pleiotropic gene and an independent risk factor of all-cause mortality, COPD and CVD mortality.

In **Chapter 4** we investigated the relation between the *NFE2L2* gene polymorphisms and all-cause, CVD and COPD mortality and its associations with triglyceride and cholesterol levels.

In **Chapter 5** we described a GWA studies on all-cause and cardiovascular mortality performed in a general population. We identified novel common SNPs associated with mortality in the Vlagtwedde-Vlaardingen cohort, however a replication of the results in another population is needed.

In **Chapter 6** we examined whether combining multiple SNPs with modest effects into a genetic risk score (GRS) may improve identification of subjects with reduced lung function level.

Chapter 7 provides new insights into associations between a common respiratory symptom (i.e. dyspnea) and mortality. We investigated the effect of dyspnea severity (moderate and severe) and changes in dyspnea status (development, persistence and remission of dyspnea) on all-cause, CVD, and COPD mortality. The interesting and important results in this study show that dyspnea remission normalizes mortality risk.

In **Chapter 8** the results are summarized, main conclusions are drawn and future perspectives are discussed.

REFERENCES

1. Newman AB, Murabito JM. The Epidemiology of Longevity and Exceptional Survival. Epidemiol Rev 2013 Jan 31.
2. Van ML, Mets T, De VP, Gorus E. The Active Ageing-concept translated to the residential long-term care. Qual Life Res 2013 Jun;22(5):929-37.
3. Blagosklonny MV. How to save Medicare: the anti-aging remedy. Aging (Albany NY) 2012 Aug;4(8):547-52.
4. Rowe JW, Kahn RL. Successful aging. Gerontologist 1997 Aug;37(4):433-40.
5. 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 Mar;97(3):319-23.
6. Hayflick L, Moorhead PS. The serial cultivation of human diploid cell strains. Exp Cell Res 1961 Dec;25:585-621.
7. Naylor RM, Baker DJ, van Deursen JM. Senescent cells: a novel therapeutic target for aging and age-related diseases. Clin Pharmacol Ther 2013 Jan;93(1):105-16.
8. Stamler J, Dyer AR, Shekelle RB, Neaton J, Stamler R. Relationship of baseline major risk factors to coronary and all-cause mortality, and to longevity: findings from long-term follow-up of Chicago cohorts. Cardiology 1993;82(2-3):191-222.
9. Kalhan R, Tran BT, Colangelo LA, Rosenberg SR, Liu K, Thyagarajan B, et al. Systemic inflammation in young adults is associated with abnormal lung function in middle age. PLoS One 2010;5(7):e11431.
10. Hole DJ, Watt GC, Davey-Smith G, Hart CL, Gillis CR, Hawthorne VM. Impaired lung function and mortality risk in men and women: findings from the Renfrew and Paisley prospective population study. BMJ 1996 Sep 21;313(7059):711-5.
11. Sin DD, Wu L, Man SF. The relationship between reduced lung function and cardiovascular mortality: a population-based study and a systematic review of the literature. Chest 2005 Jun;127(6):1952-9.
12. Wilk JB, Walter RE, Laramie JM, Gottlieb DJ, O'Connor GT. Framingham Heart Study genome-wide association: results for pulmonary function measures. BMC Med Genet 2007;8 Suppl 1:S8.
13. Ho SF, O'Mahony MS, Steward JA, Breay P, Buchalter M, Burr ML. Dyspnoea and quality of life in older people at home. Age Ageing 2001 Mar;30(2):155-9.
14. Frostad A, Soyseth V, Haldorsen T, Andersen A, Gulsvik A. Respiratory symptoms and 30 year mortality from obstructive lung disease and pneumonia. Thorax 2006 Nov;61(11):951-6.
15. Frostad A, Soyseth V, Haldorsen T, Andersen A, Gulsvik A. Respiratory symptoms and long-term cardiovascular mortality. Respir Med 2007 Nov;101(11):2289-96.

16. Huijnen B, van der Horst F, van Amelsvoort L, Wesseling G, Lansbergen M, Aarts P, et al. Dyspnea in elderly family practice patients. Occurrence, severity, quality of life and mortality over an 8-year period. Fam Pract 2006 Feb;23(1):34-9.
17. Nishimura K, Izumi T, Tsukino M, Oga T. Dyspnea is a better predictor of 5-year survival than airway obstruction in patients with COPD. Chest 2002 May;121(5):1434-40.
18. Tessier JF, Nejjari C, Letenneur L, Filleul L, Marty ML, Barberger Gateau P, et al. Dyspnea and 8-year mortality among elderly men and women: the PAQUID cohort study. Eur J Epidemiol 2001;17(3):223-9.
19. Mooney S. Bioinformatics approaches and resources for single nucleotide polymorphism functional analysis. Brief Bioinform 2005 Mar;6(1):44-56.
20. Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet 2013 Mar;9(3):e1003348.
21. Zhu M, Zhao S. Candidate gene identification approach: progress and challenges. Int J Biol Sci 2007;3(7):420-7.
22. Wayne ML, McIntyre LM. Combining mapping and arraying: An approach to candidate gene identification. Proc Natl Acad Sci U S A 2002 Nov 12;99(23):14903-6.
23. Bartke A, Brown-Borg H. Life extension in the dwarf mouse. Curr Top Dev Biol 2004;63:189-225.
24. Friedman DB, Johnson TE. A mutation in the age-1 gene in Caenorhabditis elegans lengthens life and reduces hermaphrodite fertility. Genetics 1988 Jan;118(1):75-86.
25. 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 Jan 9;421(6919):182-7.
26. Kaeberlein M, McVey M, Guarente L. The SIR2/3/4 complex and SIR2 alone promote longevity in Saccharomyces cerevisiae by two different mechanisms. Genes Dev 1999 Oct 1;13(19):2570-80.
27. Kenyon C, Chang J, Gensch E, Rudner A, Tabtiang R. A C. elegans mutant that lives twice as long as wild type. Nature 1993 Dec 2;366(6454):461-4.
28. 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 Nov 14;278(5341):1319-22.
29. Schriner SE, Linford NJ, Martin GM, Treuting P, Ogburn CE, Emond M, et al. Extension of murine life span by overexpression of catalase targeted to mitochondria. Science 2005 Jun 24;308(5730):1909-11.
30. Tissenbaum HA, Guarente L. Increased dosage of a sir-2 gene extends lifespan in Caenorhabditis elegans. Nature 2001 Mar 8;410(6825):227-30.
31. de Magalhaes JP, Toussaint O. GenAge: a genomic and proteomic network map of human ageing. FEBS Lett 2004 Jul 30;571(1-3):243-7.

32. Eisenstein M. Centenarians: Great expectations. Nature 2012 Dec 6;492(7427):S6-S8.
33. Frisoni GB, Louhija J, Geroldi C, Trabucchi M. Longevity and the epsilon2 allele of apolipoprotein E: the Finnish Centenarians Study. J Gerontol A Biol Sci Med Sci 2001 Feb;56(2):M75-M78.
34. Jian-Gang Z, Yong-Xing M, Chuan-Fu W, Pei-Fang L, Song-Bai Z, Nui-Fan G, et al. Apolipoprotein E and longevity among Han Chinese population. Mech Ageing Dev 1998 Aug 14;104(2):159-67.
35. Kervinen K, Savolainen MJ, Salokannel J, Hynninen A, Heikkinen J, Ehnholm C, et al. Apolipoprotein E and B polymorphisms--longevity factors assessed in nonagenarians. Atherosclerosis 1994 Jan;105(1):89-95.
36. Louhija J, Miettinen HE, Kontula K, Tikkanen MJ, Miettinen TA, Tilvis RS. Aging and genetic variation of plasma apolipoproteins. Relative loss of the apolipoprotein E4 phenotype in centenarians. Arterioscler Thromb 1994 Jul;14(7):1084-9.
37. Stakias N, Liakos P, Tsiapali E, Goutou M, Koukoulis GN. Lower prevalence of epsilon 4 allele of apolipoprotein E gene in healthy, longer-lived individuals of Hellenic origin. J Gerontol A Biol Sci Med Sci 2006 Dec;61(12):1228-31.
38. 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 Apr;12(2):95-104.
39. 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 Feb 24;106(8):2700-5.
40. 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 Dec 15;18(24):4897-904.
41. 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 Sep 16;105(37):13987-92.
42. Christensen K, Johnson TE, Vaupel JW. The quest for genetic determinants of human longevity: challenges and insights. Nat Rev Genet 2006 Jun;7(6):436-48.
43. Kaeberlein M, McVey M, Guarente L. The SIR2/3/4 complex and SIR2 alone promote longevity in Saccharomyces cerevisiae by two different mechanisms. Genes Dev 1999 Oct 1;13(19):2570-80.
44. Rogina B, Helfand SL. Sir2 mediates longevity in the fly through a pathway related to calorie restriction. Proc Natl Acad Sci U S A 2004 Nov 9;101(45):15998-6003.
45. Tissenbaum HA, Guarente L. Increased dosage of a sir-2 gene extends lifespan in Caenorhabditis elegans. Nature 2001 Mar 8;410(6825):227-30.

46. Frye RA. Characterization of five human cDNAs with homology to the yeast SIR2 gene: Sir2-like proteins (sirtuins) metabolize NAD and may have protein ADP-ribosyltransferase activity. Biochem Biophys Res Commun 1999 Jun 24;260(1):273-9.

47. Leibiger IB, Berggren PO. Sirt1: a metabolic master switch that modulates lifespan. Nat Med 2006 Jan;12(1):34-6.

48. Gunn TM, Azarani A, Kim PH, Hyman RW, Davis RW, Barsh GS. Identification and preliminary characterization of mouse Adam33. BMC Genet 2002;3:2.

49. Van Eerdewegh P, Little RD, Dupuis J, Del Mastro RG, Falls K, Simon J, et al. Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. Nature 2002 Jul 25;418(6896):426-30.

50. Aierken H, Wushouer Q, Shayhidin E, Jing W, Wufuer D. Polymorphisms of the ADAM33 gene and Chronic Obstructive Pulmonary Disease Risk: a Meta-Analysis. Clin Respir J 2013 Aug 1.

51. Sadeghnejad A, Ohar JA, Zheng SL, Sterling DA, Hawkins GA, Meyers DA, et al. Adam33 polymorphisms are associated with COPD and lung function in long-term tobacco smokers. Respir Res 2009;10:21.

52. Simpson A, Maniatis N, Jury F, Cakebread JA, Lowe LA, Holgate ST, et al. Polymorphisms in a disintegrin and metalloprotease 33 (ADAM33) predict impaired early-life lung function. Am J Respir Crit Care Med 2005 Jul 1;172(1):55-60.

53. van Diemen CC, Postma DS, Vonk JM, Bruinenberg M, Schouten JP, Boezen HM. A disintegrin and metalloprotease 33 polymorphisms and lung function decline in the general population. Am J Respir Crit Care Med 2005 Aug 1;172(3):329-33.

54. Zou J, Zhu F, Liu J, Wang W, Zhang R, Garlisi CG, et al. Catalytic activity of human ADAM33. J Biol Chem 2004 Mar 12;279(11):9818-30.

55. Shaffiq A, Haitchi HM, Pang YY, Alangari AA, Jones MG, Marshall BG, et al. A disintegrin and metalloprotease (ADAM) 33 protein in patients with pulmonary sarcoidosis. Respirology 2012 Feb;17(2):342-9.

56. Holloway JW, Laxton RC, Rose-Zerilli MJ, Holloway JA, Andrews AL, Riaz Z, et al. ADAM33 expression in atherosclerotic lesions and relationship of ADAM33 gene variation with atherosclerosis. Atherosclerosis 2010 Feb 24.

57. Hybertson BM, Gao B, Bose SK, McCord JM. Oxidative stress in health and disease: the therapeutic potential of Nrf2 activation. Mol Aspects Med 2011 Aug;32(4-6):234-46.

58. Lewis KN, Mele J, Hayes JD, Buffenstein R. Nrf2, a guardian of healthspan and gatekeeper of species longevity. Integr Comp Biol 2010 Nov;50(5):829-43.

59. Kapeta S, Chondrogianni N, Gonos ES. Nuclear erythroid factor 2-mediated proteasome activation delays senescence in human fibroblasts. J Biol Chem 2010 Mar 12;285(11):8171-84.

60. 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 Aug;10(4):686-98.

61. 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 Jun;132 (6-7):324-30.

62. 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(1):e29848.

63. Walter S, Atzmon G, Demerath EW, Garcia ME, Kaplan RC, Kumari M, et al. A genome-wide association study of aging. Neurobiol Aging 2011 Nov;32(11):2109-28.

64. Ioannidis JP, Thomas G, Daly MJ. Validating, augmenting and refining genome-wide association signals. Nat Rev Genet 2009 May;10(5):318-29.

Chapter 2

SIRT1 polymorphism, long-term survival and glucose tolerance in the general population

Sylwia M Figarska, Judith M Vonk and H Marike Boezen

Department of Epidemiology, University of Groningen,
University Medical Center Groningen, the Netherlands

PLOS ONE 2013;8(3):e58636.

ABSTRACT

Mutations that increase activity of Sir2 (silent information regulator 2) are associated with extended lifespan of yeast, fruit flies and worms. *SIRT1*, the human homolog of Sir2, that controls numerous physiological processes including the glucose metabolism, is considered a candidate gene for predicting variation in human lifespan. Whereas the role of Sir2 has been extensively investigated in model organisms, less is known about the relation between *SIRT1* and lifespan in humans. In the current study we included 1,390 subjects from a general population-based cohort with 18 years of follow-up to investigate associations between variation in single nucleotide polymorphisms (SNPs) in the *SIRT1* gene and human survival. Additionally in 535 male subjects with available data we investigated associations between *SIRT1* and glucose tolerance. Carriers of the minor allele of rs12778366 had a significantly reduced mortality risk compared to the wild types: Hazard Ratio 0.69 (95% CI 0.50 to 0.96; p=0.025). The directions of the effect were the same in females and males, never and ever smokers and the effect was significantly protective in overweight/obese subjects. Carriers of the minor allele of SNP rs12778366 had better glucose tolerance indicated by 0.34 mmol/l lower glucose levels compared to wild type subjects (p=0.03).

This study shows that *SIRT1* affects human long-term survival and therefore may be an important factor in modulating lifespan not only in lower organisms, but also in humans.

INTRODUCTION

In the ongoing quest to uncover factors that increase longevity, sirtuins have attracted scientific and public interest for the past decades (1). Initially, overexpression of the silent information regulator Sir2, nicotinamide adenine dinucleotide (NAD^+)-dependent histone deacetylase, has shown a beneficial effect on lifespan in budding yeast (*Saccharomyces cerevisae*) (2). Subsequently, the experiments performed in worms (*Caenorhabditis elegans*) and flies (*Drosophila melanogaster*) (3;4) confirmed favorable properties of Sir2, signifying the importance of sirtuins as longevity genes. Since a more recent report showed an absence of effects of Sir2 overexpression on lifespan in *C.elegans* and *Drosophila* (5), a debate about the role of sirtuins in lifespan prolongation has arisen (1). Therefore more studies are needed to elucidate the impact of sirtuins on lifespan, especially in humans. Out of seven identified mammalian homologues, sirtuin 1 (*SIRT1*) is the most closely related to Sir2 (6). *SIRT1* influences the activity of various transcription factors, including forkhead-box transcription factors (FOXOs), peroxisome proliferator-activated receptor γ (PPAR γ) and nuclear factor-κB (NF-κB) in target tissues resulting in enhanced gluconeogenesis and repressed glycolysis in the liver, reduction of adipogenesis in adipose tissue, and increased release of insulin in pancreatic beta cells (7). *SIRT1* controls adiponectin levels, inflammatory processes, gluconeogenesis, and levels of reactive oxygen species that together may lead to the development of insulin resistance (8). Overexpression of *SIRT1* or using *SIRT1* activators improves glucose homeostasis and insulin sensitivity in mice (8-11). Therefore partly the effect of *SIRT1* on longevity may be exerted via its association with insulin signaling, which has been proven to extend lifespan by 18% in fat-specific insulin

receptor knockout (FIRKO) mice (12). Furthermore, *SIRT1* is required for a normal response to caloric restriction that causes many changes in glucose metabolism and increases lifespan (13).

In humans, during the last years polymorphisms in *SIRT1* have been investigated in a context of metabolism and have been associated with BMI and risk of obesity (14-17), acute insulin response in Pima Indians (18), body fat and blood pressure in Japanese (19), basal energy expenditure and respiratory quotient (20), and with diabetes risk in interaction with prenatal exposure to famine (21). The few studies that investigated SNPs in *SIRT1* in relation to human lifespan or mortality did not find any associations (22-25).

Given the fact that near 30% of the individual variance in life expectancy is genetically determined (26) and the specific genetic determinants of human lifespan still remain largely unknown, *SIRT1,* as a metabolic master switch (7), may be considered a candidate gene for predicting variation in human lifespan. The Vlagtwedde-Vlaardingen cohort offers the unique opportunity to investigate the role of *SIRT1* in long-term survival, because subjects included in the current study were followed up for 18 years. Since *SIRT1* modulates a range of cellular processes involved in maintaining glucose homeostasis (27), we additionally investigated *SIRT1* polymorphisms and glucose tolerance.

METHODS

Ethics Statement

The study protocol was approved by the local university medical hospital ethics committee, University of Groningen, University Medical Center Groningen, The Netherlands and all participants

gave their written informed consent. In 1984, the Committee on Human Subjects in Research of the University of Groningen reviewed the study and affirmed the safety of the protocol and study design.

Study population

We studied 1,390 subjects of the Vlagtwedde-Vlaardingen cohort participating in the last survey in 1989/1990 (28). This general population-based cohort of white individuals of Dutch descent started in 1965 and has been followed for 25 years. The main focus of the study was on respiratory health. Surveys (median number of 7 per subject, range 1-8) were performed every 3 years, during which information was collected on smoking status, age, sex and respiratory symptoms by the Dutch version of the British Medical Council standardized questionnaire, BMI was determined, spirometry was performed and and the number of eosinophils in peripheral blood was measured. The vital status of all participants in the study on December 31, 2008 was assessed. Causes of death were coded according to the International Classification of Diseases (ICD) and obtained from the Statistics Netherlands (The Hague). In order to avoid bias and provide true associations, the external causes of death (i.e. suicides, homicides, traffic accidents etc.) were excluded from the analyses. (ICD-9: codes ≥ 800 and in ICD-10: codes ≥ S00).

Blood samples

In 1989/1990 neutrophil depots from peripheral blood samples were collected and stored at –20°C. In 2003–2004 DNA was extracted from these samples with a QIAamp DNA blood mini kit (Qiagen, Hilden, Germany) and checked for purity and concentration with a NanoDrop ND-1000 UV–Vis spectrophotometer (NanoDrop Technologies, Wilmington, DE) (28).

SNP Selection and Genotyping

Four SNPs (rs12778366, rs10823108, rs7069102 and rs2273773), that tag all 21 SNPs in *SIRT1* and its 5kb up-/downstream region with r^2>0.8 and Minor Allele Frequency>5% (based on the HapMap release 23a/March 2008) were genotyped by K-Bioscience Ltd (UK) (29). Since rs10823108 and rs2273773 were in complete linkage disequilibrium (r^2=1.0, Supplementary Figure 1) in our study population, only rs2273773 was analyzed.

Oral glucose tolerance test

In 1970/1972, 1973 and 1976 male subjects underwent the oral glucose tolerance test (OGTT). They were given a drink of 100g glucose solution, and blood glucose was measured two hours later.

Statistical Analysis

Descriptive analyses of the subject characteristics were performed using χ^2 tests for categorical variables and Mann-Whitney U test for continuous variables (i.e. packyears in ever smokers and age). The genotype frequencies were tested for Hardy-Weinberg Equilibrium (HWE) by χ^2 analysis. Differences in genotype distribution between dead and alive subjects were tested using χ^2 tests. SNP rs10823108 was tested in a general genetic model. Due to the low frequency of individuals being homozygous for the minor allele for rs12778366 (n=14) and rs2273773 (n=9) heterozygotes and homozygotes for the minor allele were analyzed in a one group. Cox proportional hazards regression models adjusted for gender, age and packyears of smoking (all at the survey in 1989/1990) were used to evaluate the association between SNPs and all-cause mortality. Time was defined from the examination in 1989/1990 until death, end of follow-up in 2008 or last registration

if subjects were lost to follow-up. Survival curves are depicted based on these Cox models. Stratified analyses according to gender, smoking habits (never smokers vs ever smokers), BMI and age (dichotomized based on a median age at visit in 1989/1990, i.e. 52 yrs) were performed. Subjects with BMI ≥ 25 kg/m^2 were categorized into the overweight/obese group according to World Health Organization (WHO) criteria.

A linear regression model adjusted for age at the measurement was used to evaluate the associations between SNPs in *SIRT1* and glucose tolerance.

P values <0.05 were considered statistically significant (tested 2-sided). All statistical analyses were performed using SPSS version 18.0 for Windows.

RESULTS

Subjects with genetic data available and participating in the last survey in 1989/1990 were included in this study (n=1,390, see Table 1). After 18 years of follow-up, 78.2% (n=1,087) of the cohort was still alive and 284 deaths (20.4%) were recorded. Out of all deaths, 14 (4.9%) occurred due to external causes and these were excluded from the analyses. Of the participants who had died 207 (76.7%) were ever smokers and these had significantly higher numbers of packyears compared to participants still alive. It is important to note that our study had an excellent follow-up rate, since only 19 subjects (1.4%) could not be traced back.

The 3 tested SNPs (rs12778366, rs7069102 and rs2273773) were in Hardy-Weinberg equilibrium. Among subjects who died, 83% had the wild type genotype of rs12778366 which is significantly higher (p=0.015) than the 76% in the subjects who were still alive.

Cox regression showed that carriers of the minor allele of rs12778366 had a significantly reduced risk of mortality compared to wild types: HR 0.69 (95% CI: 0.50 to 0.96; p=0.025; see Table 2). Survival curves according to genotypes of rs12778366 clearly show the difference in mortality risk (Figure 1a). The same directions of the effect of rs12778366 were observed within groups that have different mortality risks, i.e. females (HR=0.82 (0.50-1.35)) and males (HR=0.63(0.41-0.96)), never smokers (HR=0.53 (0.25-1.11)) and ever smokers (HR=0.75 (0.52-1.08)), younger subjects (age ≤ 52 yrs) (HR=0.68 (0.30-1.54)) and older subjects (age > 52 yrs) (HR=0.69 (0.48-0.98)), (Table 3 and Figure 1 b-c, e). Remarkably, the protective effect of rs12778366 was observed in overweight/obese subjects (HR=0.62 (0.43-0.91)) but not in subjects with normal weight (HR=0.95 (0.50-1.80). The survival curves clearly show that

Table 1 Characteristics of participants at visit 1989/1990 by vital status on Dec 31st, 2008

Status on 31-12-2008	Alive	Dead	p value
Number (%)	1087 (78.2)	270 (19.4)	
Males, n (%)	525 (48.3)	166 (61.5)	0.000
Age, median (range)	49.4 (36.0 to 72.6)	61.8 (37.3 to 79.1)	0.000
Ever smokers, n (%)	711 (65.4)	207 (76.7)	0.000
Packyears in ever smokers, median (range)	17.2 (0.1 to 117.1)	26.0 (0.6 to 262.2)	0.000
BMI			
Normal weight, n (%)	284 (26.2)	65 (24.2)	
Overweight, n (%)	540 (49.8)	139 (51.6)	
Obese, n (%)	261 (24.0)	65 (24.2)	0.781

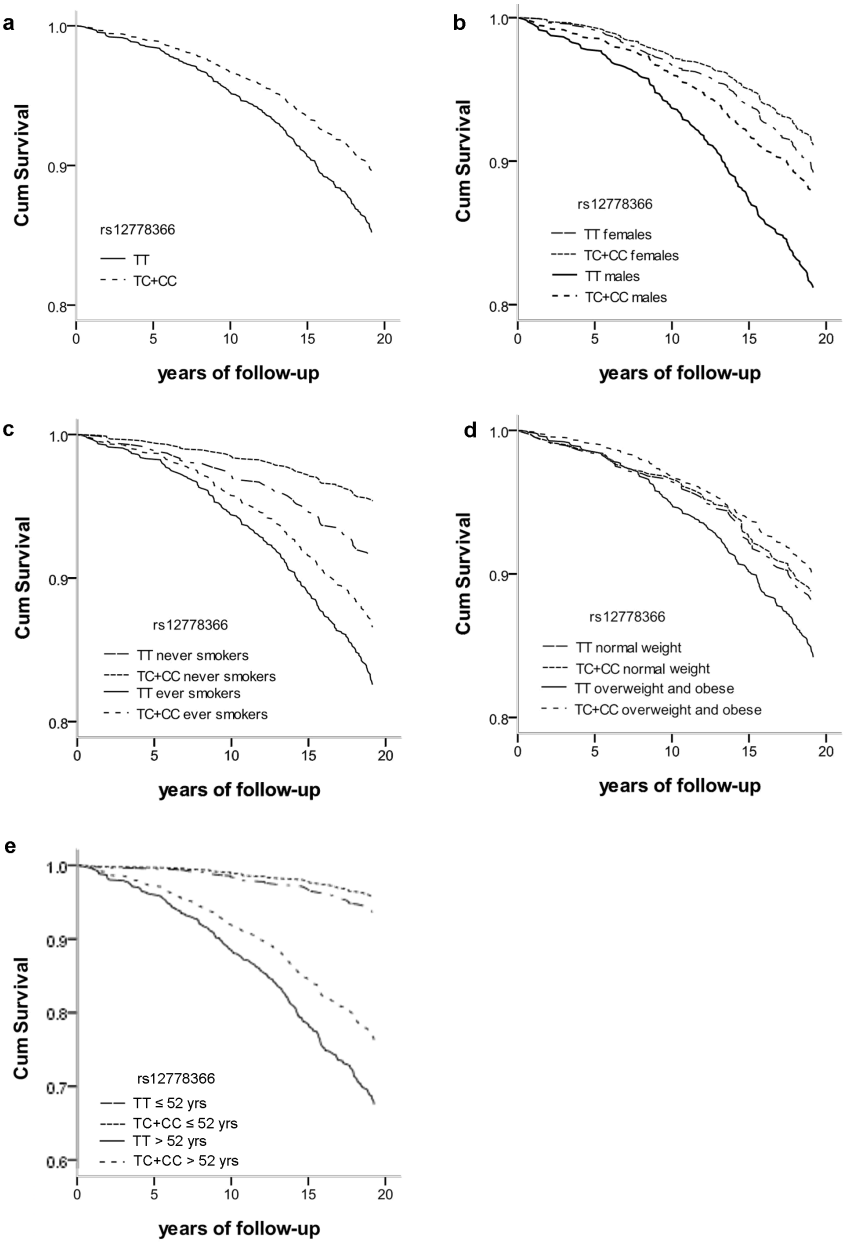


Figure 1 Survival curves for all-cause mortality according to SNP rs12778366

a. all subjects; **b.** stratified according to gender; **c.** stratified according to smoking habits; **d.** stratified according to BMI; **e.** stratified according to age (median age at visit in 1989/1900). *the Y axis scale in a figure 1e differs from Y axis scales in other figures

Table 2 Distribution of genotypes and hazard ratio (HR) for all-cause mortality

SNP	Genotype	Alive n=1,087	All-cause mortality n=270	p value**	HR (95% CI)***	p value
rs12778366*	TT	798 (76.0)	220 (83.0)		1	
	TC+CC	252 (24.0)	45 (17.0)	**0.015**	0.69 (0.50-0.96)	**0.025**
rs7069102	GG	449 (43.6)	124 (48.8)		1	
	GC	458 (44.5)	103 (40.6)		0.85 (0.65-1.11)	0.234
	CC	123 (11.9)	27 (10.6)	0.323	0.90 (0.59-1.37)	0.628
rs2273773*	TT	897 (81.1)	164 (76.6)		1	
	TC+CC	209 (18.9)	50 (23.4)	0.132	1.26 (0.92-1.73)	0.155

* Due to the low frequency of individuals being homozygous for the minor allele heterozygotes and homozygotes variants were combined

** Differences in genotype distribution between alive subjects and those who died (excluding external causes of death) tested with χ^2 test

*** Cox regression adjusted for age, gender and packyears at visit in 1989/90

overweight/obese minor allele carriers of rs12778366 had survival comparable to subjects with normal weight while overweight/obese subjects with the wildtype genotype had an increased mortality risk (Figure 1d). The 2 other SNPs did not show significant associations between genotypes and mortality risk (Table 2).

We analyzed the available data from 535 male subjects from the current study (aged at the measurement 18-61 years) who underwent the glucose tolerance test (OGTT). We found that the minor allele carriers of SNP rs12778366 had better glucose tolerance, since they had a 0.34 mmol/l lower glucose levels compared to wild type subjects (p=0.03, see Table 4). When this association was further

Table 3 HR for all-cause mortality for rs12778366 TC+CC genotypes in stratified analysis

Stratification	HR (95% CI)**	p value
a) gender		
Females, n*=653	0.82 (0.50-1.35)	0.424
Males, n=680	0.63 (0.41-0.96)	0.032
b) smoking habits		
Never smokers, n=424	0.53 (0.25-1.11)	0.092
Ever smokers, n=909	0.75 (0.52-1.08)	0.125
c) BMI		
Normal weight, n=344	0.95 (0.50-1.80)	0.870
Overweight and obese, n=987	0.62 (0.43-0.91)	0.014
d) age		
≤ 52 yrs, n=670	0.68 (0.30-1.54)	0.350
> 52 yrs, n=663	0.69 (0.48-0.98)	0.040

***** n=number of all subjects included in the analysis (excluding those who died due to external causes)

****** rs12778366 TT genotype as a reference

investigated in subjects with normal weight (n=249) and overweight/obese subjects (n=284) separately, we found significantly better glucose tolerance in overweight/obese carriers of the minor allele of rs12778366 (i.e. 0.60 mmol/l lower glucose levels (p=0.01)) compared to overweight/obese wild type subjects. In subjects with normal weight the same direction was observed (0.18 mmol/l lower glucose levels), but the effect was not significant (p=0.50).

Table 4 Glucose levels (mmol/l) measured in males after the oral glucose tolerance test (OGGT)

SNP	Genotype	n	Mean (SD)	B* (mmol/l)	SE	p value
rs12778366	TT	414	6.33 (1.55)			
	TC+CC	119	5.99 (1.22)	-0.34	0.16	**0.030**
rs7069102	GG	229	6.33 (1.59)			
	GC	232	6.02 (1.43)	-0.12	0.14	0.401
	CC	59	6.32 (1.44)	0.02	0.22	0.931
rs2273773	TT	446	6.27 (1.51)			
	TC+CC	89	6.22 (1.44)	-0.04	0.17	0.830

***** Regression coefficient (B), its standard error (SE) and p value obtained with linear regression analysis adjusted for age at the measurement

DISCUSSION

We found a 30% reduced mortality risk among minor allele carriers of SNP rs12778366 in *SIRT1*, during a 18 years follow-up study in the general population. Therefore, *SIRT1* appears to be an important candidate gene explaining individual differences in human lifespan. There is evidence linking overexpression of *Sir2* to extended lifespan in yeast, worms and flies (2-4). Despite the established role of mammalian *SIRT1* in metabolism, genome stability and stress response (30;31), polymorphisms in *SIRT1* were not associated with exceptional human longevity in a cross-sectional case-control study (22;24) nor with all-cause mortality in a general population-based cohort (25) and in a group of over 85 years who were followed up until they died (23). Whereas the last study included only pre-

selected old subjects the advantage of our current study is that we did not use a selection criteria based on age, but investigated the whole population in a longitudinal manner. Actually, up till now, only one study in humans showed associations between *SIRT1* variants and healthy aging, what the authors defined as being healthy (i.e. normal brain function and verbal fluency test; laboratory findings for hemogram, peripheral smear, urine, electrocytes, chest X-ray, kidney, pulmonary function, echocardiography and ECG were normal) at the age 60 or higher (32). In the light of recent uncertainty about the role of sirtuins in longevity (1) our results importantly provide new evidence in favor of a role of the gene in longevity.

We have shown an association between the *SIRT1* gene and long-term survival among minor allele carriers of SNP rs12778366 in *SIRT1* in the total population. Furthermore, the directions of the effect did not change in stratified analyses according to gender, smoking habits and age. Interestingly, stratification according to BMI showed the protective effect of rs1277836 only in overweight/obese subjects. Taking into account the increased mortality per se in obese subjects this finding may shed a new light on obesity-related burdens.

One of the physiological pathways through which *SIRT1* may affect longevity might be glucose homeostasis. Indications suggesting a role of this pathway were backed up by evidence that minor allele carriers of SNP rs12778366 had better glucose tolerance as determined by the oral glucose tolerance test. Interestingly, stratified analysis according to BMI showed that the effect was more pronounced in overweight/obese subjects, whereas in subjects with normal weight only the direction of the effect remained the same, but was not significant. In this light the better glucose tolerance in overweight/obese minor allele carriers of SNP rs12778366 could

be considered a condition leading to better survival in this group. This additional result emphasizes the relevance of rs12778366 and indicates its possible use as a screening tool for clinical purposes.

Previous studies indicate that transgenic mice that overexpress *SIRT1* appear to have beneficial phenotypes that may be relevant in human health, including better glucose tolerance (33;34). In contrast to the positive effects of increased *SIRT1* activity, *SIRT1* deficiency impairs metabolism (35). Therefore, we hypothesize that variants in rs12778366 may lead to overexpression of the protein, especially since this SNP is located in Transcription Factor Binding Site (TFBS). Although rs12778366 is not associated with the SIRT1 protein expression in adipose tissue, lymphoblastoid cell lines and skin (Genevar (GENe Expression VARiation) database) (36), this does not rule out a possible effect of the SNP on protein expression in other tissues i.e. in the lung, liver or heart, given the broad *SIRT1* expression in humans.

Strengths and limitations

The major strength of the current study is the longitudinal design. We were able to followparticipants for 18 years, which provided a wide time window for evaluating survival in the cohort. A strength of our study is also the number of subjects (n=1,390), sampled from the general population. Additionally, the high follow-up rate is a major strength of the study, since 98.6% of the included subjects could be traced back. A limitation of our study is the limited data on metabolic profile, because the oral glucose tolerance test was performed only in male subjects and was not accompanied by insulin measurements. Furthermore, the study population consisted of white individuals of Dutch descent which limits extrapolation to other ethnic groups.

In summary, this is the first study showing that *SIRT1* plays a role in human lifespan in a non-selected general population cohort. The importance of *SIRT1* is supported by the association of its polymorphism with long-term survival in the general population. Furthermore, linking *SIRT1* polymorphisms to improved glucose tolerance stresses the impact of *SIRT1* on metabolism in humans and identifies *SIRT1* as a possible candidate for therapeutic purposes.

Author contributions JMV prepared the mortality data for analysis. JMV and HMB supervised the study. SMF carried out the statistical analysis and drafted the manuscript. JMV and HMB obtained the grants for the study. All authors discussed results, proposed corrections and approved the final version of the manuscript.

REFERENCES

1. Accili D, de Cabo R, Sinclair DA. An unSIRTain role in longevity. Nat Med 2011;17(11):1350-1.
2. Kaeberlein M, McVey M, Guarente L. The SIR2/3/4 complex and SIR2 alone promote longevity in Saccharomyces cerevisiae by two different mechanisms. Genes Dev 1999 Oct 1;13(19):2570-80.
3. Rogina B, Helfand SL. Sir2 mediates longevity in the fly through a pathway related to calorie restriction. Proc Natl Acad Sci U S A 2004 Nov 9;101(45):15998-6003.
4. Tissenbaum HA, Guarente L. Increased dosage of a sir-2 gene extends lifespan in Caenorhabditis elegans. Nature 2001 Mar 8;410(6825):227-30.
5. Burnett C, Valentini S, Cabreiro F, Goss M, Somogyvari M, Piper MD, et al. Absence of effects of Sir2 overexpression on lifespan in C. elegans and Drosophila. Nature 2011 Sep 22;477(7365):482-5.
6. Frye RA. Characterization of five human cDNAs with homology to the yeast SIR2 gene: Sir2-like proteins (sirtuins) metabolize NAD and may have protein ADP-ribosyltransferase activity. Biochem Biophys Res Commun 1999 Jun 24;260(1):273-9.
7. Leibiger IB, Berggren PO. Sirt1: a metabolic master switch that modulates lifespan. Nat Med 2006 Jan;12(1):34-6.
8. Liang F, Kume S, Koya D. SIRT1 and insulin resistance. Nat Rev Endocrinol 2009 Jul;5(7):367-73.
9. Li Y, Xu S, Giles A, Nakamura K, Lee JW, Hou X, et al. Hepatic overexpression of SIRT1 in mice attenuates endoplasmic reticulum stress and insulin resistance in the liver. FASEB J 2011 May;25(5):1664-79.
10. Feige JN, Lagouge M, Canto C, Strehle A, Houten SM, Milne JC, et al. Specific SIRT1 activation mimics low energy levels and protects against diet-induced metabolic disorders by enhancing fat oxidation. Cell Metab 2008 Nov;8(5):347-58.
11. Lagouge M, Argmann C, Gerhart-Hines Z, Meziane H, Lerin C, Daussin F, et al. Resveratrol improves mitochondrial function and protects against metabolic disease by activating SIRT1 and PGC-1alpha. Cell 2006 Dec 15;127(6):1109-22.
12. Bluher M, Kahn BB, Kahn CR. Extended longevity in mice lacking the insulin receptor in adipose tissue. Science 2003 Jan 24;299(5606):572-4.
13. Boily G, Seifert EL, Bevilacqua L, He XH, Sabourin G, Estey C, et al. SirT1 regulates energy metabolism and response to caloric restriction in mice. PLoS One 2008;3(3):e1759.
14. Clark SJ, Falchi M, Olsson B, Jacobson P, Cauchi S, Balkau B, et al. Association of Sirtuin 1 (SIRT1) Gene SNPs and Transcript Expression Levels With Severe Obesity. Obesity (Silver Spring) 2012 Jan;20(1):178-85.

15. Peeters AV, Beckers S, Verrijken A, Mertens I, Roevens P, Peeters PJ, et al. Association of SIRT1 gene variation with visceral obesity. Hum Genet 2008 Nov;124(4):431-6.

16. Zheng J, Chen LL, Xiao F, Hu X, Deng X, Li H. Three single nucleotide variants of the SIRT1 gene are associated with overweight in a Chinese population: a case control study. Endocr J 2012;59(3):229-37.

17. Zillikens MC, van Meurs JB, Rivadeneira F, Amin N, Hofman A, Oostra BA, et al. SIRT1 genetic variation is related to BMI and risk of obesity. Diabetes 2009 Dec;58(12):2828-34.

18. Dong Y, Guo T, Traurig M, Mason CC, Kobes S, Perez J, et al. SIRT1 is associated with a decrease in acute insulin secretion and a sex specific increase in risk for type 2 diabetes in Pima Indians. Mol Genet Metab 2011 Dec;104(4):661-5.

19. Shimoyama Y, Suzuki K, Hamajima N, Niwa T. Sirtuin 1 gene polymorphisms are associated with body fat and blood pressure in Japanese. Transl Res 2011 Jun;157(6):339-47.

20. Weyrich P, Machicao F, Reinhardt J, Machann J, Schick F, Tschritter O, et al. SIRT1 genetic variants associate with the metabolic response of Caucasians to a controlled lifestyle intervention–the TULIP Study. BMC Med Genet 2008;9:100.

21. Botden IP, Zillikens MC, de Rooij SR, Langendonk JG, Danser AH, Sijbrands EJ, et al. Variants in the SIRT1 gene may affect diabetes risk in interaction with prenatal exposure to famine. Diabetes Care 2012 Feb;35(2):424-6.

22. Flachsbart F, Croucher PJ, Nikolaus S, Hampe J, Cordes C, Schreiber S, et al. Sirtuin 1 (SIRT1) sequence variation is not associated with exceptional human longevity. Exp Gerontol 2006 Jan;41(1):98-102.

23. 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 Sep;62(9):960-5.

24. 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 Sep 16;105(37):13987-92.

25. Zillikens MC, van Meurs JB, Sijbrands EJ, Rivadeneira F, Dehghan A, van Leeuwen JP, et al. SIRT1 genetic variation and mortality in type 2 diabetes: interaction with smoking and dietary niacin. Free Radic Biol Med 2009 Mar 15;46(6):836-41.

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 Mar;97(3):319-23.

27. Houtkooper RH, Pirinen E, Auwerx J. Sirtuins as regulators of metabolism and healthspan. Nat Rev Mol Cell Biol 2012 Apr;13(4):225-38.

28. van Diemen CC, Postma DS, Vonk JM, Bruinenberg M, Schouten JP, Boezen HM. A disintegrin and metalloprotease 33 polymorphisms and lung function decline in the general population. Am J Respir Crit Care Med 2005 Aug 1;172(3):329-33.

29. Siedlinski M, Boer JM, Smit HA, Postma DS, Boezen HM. Dietary factors and lung function in the general population: wine and resveratrol intake. Eur Respir J 2012 Feb;39(2):385-91.

30. Guarente L. Franklin H. Epstein Lecture: Sirtuins, aging, and medicine. N Engl J Med 2011 Jun 9;364(23):2235-44.

31. Chua KF, Mostoslavsky R, Lombard DB, Pang WW, Saito S, Franco S, et al. Mammalian SIRT1 limits replicative life span in response to chronic genotoxic stress. Cell Metab 2005 Jul;2(1):67-76.

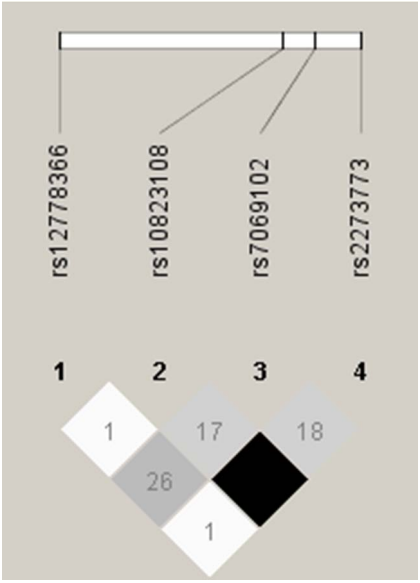
32. Zhang WG, Bai XJ, Chen XM. SIRT1 variants are associated with aging in a healthy Han Chinese population. Clin Chim Acta 2010 Nov 11;411(21-22):1679-83.

33. Bordone L, Cohen D, Robinson A, Motta MC, van VE, Czopik A, et al. SIRT1 transgenic mice show phenotypes resembling calorie restriction. Aging Cell 2007 Dec;6(6):759-67.

34. Moynihan KA, Grimm AA, Plueger MM, Bernal-Mizrachi E, Ford E, Cras-Meneur C, et al. Increased dosage of mammalian Sir2 in pancreatic beta cells enhances glucose-stimulated insulin secretion in mice. Cell Metab 2005 Aug;2(2):105-17.

35. Baur JA, Ungvari Z, Minor RK, Le Couteur DG, de CR. Are sirtuins viable targets for improving healthspan and lifespan? Nat Rev Drug Discov 2012;11(6):443-61.

36. 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 Oct 1;26(19):2474-6.



Supplementary figure 1 *SIRT1* linkage disequilibrium plot ($100 \cdot r^2$) in the Vlagtwedde-Vlaardingen cohort

Chapter 3

ADAM33 gene polymorphisms and mortality. A prospective cohort study

Sylwia M Figarska[1,4], Judith M Vonk[1,4], Cleo C van Diemen[2], Dirkje S Postma[3,4] and H Marike Boezen[1,4]

[1] University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, The Netherlands

[2] University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, The Netherlands

[3] University of Groningen, University Medical Center Groningen, Department of Pulmonology, Groningen, The Netherlands

[4] University Medical Center Groningen, Groningen Research Institute for Asthma and COPD, Groningen, The Netherlands

PLOS ONE 2013 Jul 4;8(7):e67768.

ABSTRACT

The *ADAM33* gene is associated with the pathophysiology of Chronic Obstructive Pulmonary Disease (COPD) and atherosclerosis. In this study we investigated all-cause, COPD and cardiovascular mortality, in relation to single nucleotide polymorphisms (SNPs) in *ADAM33* (Q_1, S_1, S_2, T_1 and T_2) that were genotyped in 1,390 subjects from the Vlagtwedde-Vlaardingen cohort. Participants were examined at entry in 1989/1990 and followed up till evaluation of the vital status on December 31st, 2008. Using Cox proportional hazards regression we estimated the risk of the SNPs in relation to mortality, adjusting for gender, age, FEV_1, height, place of residence and packyears of smoking. Additionally, we performed stratified analyses according to gender and smoking habits.

After 18 years, 284 (20.4%) subjects had died (107 due to cardiovascular disease and 20 due to COPD). Individuals homozygous for the minor allele of SNP T_2 had an increased risk of all-cause and cardiovascular mortality compared to wild types: hazard ratio 3.6 (95% confidence interval 2.0 to 6.7) and 3.4 (1.2 to 9.5) respectively. Individuals homozygous for the minor allele of S_1, S_2, T_2 or Q_1 had a significantly increased risk of COPD mortality. In stratified analyses the risk of all-cause mortality associated with SNP T_2 did not change: females 3.5 (1.5 to 8.3), males 3.1 (1.2 to 7.6), never smokers 3.8 (0.9 to 16.3), ever smokers 3.6 (1.8 to 7.2).

This study shows for the first time that *ADAM33* is a pleiotropic gene that is associated with all-cause, COPD and cardiovascular mortality, independent of potential confounders.

INTRODUCTION

Human lifespan has increased over the years almost worldwide (1). Therefore the concept of healthy ageing, defined as a high quality of life into later stages of life with an absence of age-related disease, is becoming increasingly important (2). So far the mechanisms explaining individual differences in lifespan and susceptibility to disease are not well understood. Thirty percent of the individual variance in life expectancy is genetically determined (3), yet the specific genetic determinants of human lifespan still remain largely unknown. One of the main objectives in research on ageing is to identify people at higher risk to developing early onset pathologies commonly associated with ageing and contributing to premature death (3). There is an unmet need for studies that increase our knowledge about determinants of the variation in human lifespan, morbidity and mortality and that highlight potential targets for prevention. One of the goals is to identify pleiotropic genes that may lead to premature death by influencing the risk of one, or more than one, disease.

A family of proteins that may be important in explaining the individual differences in lifespan is the ADAM (A Desintegrin and Metalloproteinase) family. ADAMs are membrane-anchored proteins belonging to the zinc protease superfamily (4;5). They play a role in cell adhesion, cell migration and proteolysis (6) and thus are fundamental to many control processes in development and homeostasis (7). *ADAM33* might be associated with overall mortality through its link to "inflamm-ageing". This phenomenon refers to the fact that ageing is associated with chronic, low grade inflammatory activity leading to long-term tissue damage and systemic chronic inflammation (8), which contribute to increased

mortality in elderly individuals (8;9). ADAM proteinases can release and activate cytokines, and if a single nucleotide polymorphism (SNP) in the *ADAM33* gene would promote a pro-inflammatory or tissue damaging activity of the transcribed protein, this may contribute to early mortality events.

In 2002, Van Eerdewegh et al. identified *ADAM33* as a susceptibility gene for asthma and airway hyperresponsiveness (5). Subsequent studies have linked polymorphisms in *ADAM33* to airway hyperresponsiveness and airway inflammation in Chronic Obstructive Pulmonary Disease (COPD), and to accelerated lung function decline and COPD development in the general population (10;11). Moreover, recently *ADAM33* was linked to cardiovascular disease (CVD), emphasizing its potential pleiotropic role in age-related diseases (6).

Given the physiological importance of *ADAM33* in pulmonary and cardiovascular diseases, we hypothesize that *ADAM33* has an impact on mortality due to these disorders.

The objective of the current study was to investigate whether SNPs in the *ADAM33* gene are associated with all-cause, COPD and cardiovascular mortality.

METHODS

Ethics statement

The study protocol was approved by the local university medical hospital ethics committee, University of Groningen, University Medical Center Groningen, The Netherlands and all participants gave their written informed consent. In 1984, the Committee on Human

Subjects in Research of the University of Groningen reviewed the study and affirmed the safety of the protocol and study design.

Study population

We studied 1,390 subjects of the Vlagtwedde-Vlaardingen cohort participating in the last survey in 1989/1990 (11). This general population-based cohort of white individuals of Dutch descent started in 1965 and has been followed for 25 years. Surveys were performed every 3 years, in which the Dutch version of the British Medical Council standardized questionnaire was filled in, and spirometry was performed (11). The vital status of all participants in the Vlagtwedde-Vlaardingen study on December 31, 2008 was assessed. We evaluated three mortality outcomes, i.e. all-cause mortality (excluding external causes of death), and COPD and cardiovascular mortality (either as primary or secondary cause of death). The causes of death were coded according to the International Classification of Diseases (ICD-9 and ICD-10, Table S1). Analyses on cause specific mortality were performed at Statistics Netherlands (The Hague).

Blood samples and DNA extraction

In 1989/1990 neutrophil depots from peripheral blood samples were collected and stored at –20°C. In 2003–2004 DNA was extracted from these samples with a QIAamp DNA blood mini kit (Qiagen, Hilden, Germany) and checked for purity and concentration with a NanoDrop ND-1000 UV–Vis spectrophotometer (NanoDrop Technologies, Wilmington, DE)

SNP selection and genotyping

Five SNPs in *ADAM33*, previously linked to asthma, airway hyperresponsiveness, COPD, or accelerated decline in Forced Expiratory Volume in 1 second (FEV_1) (5;11) were genotyped: rs612709 (Q_1, (C/T)), rs3918396 (S_1, (Val-Ile)), rs528557 (S_2, (G/C)), rs2280091 (T_1, (Met-Thr)) and rs2280090 (T_2, (Pro-Ser)). Details on genotyping and probes for genotyped SNPs were published previously (11). Since SNPs T_1 and T_2 are in high linkage disequilibrium (r^2=0.97, Figure 1) only SNP T_2 was analyzed. Figure 2 shows the position of genotyped SNPs in the *ADAM33* gene.

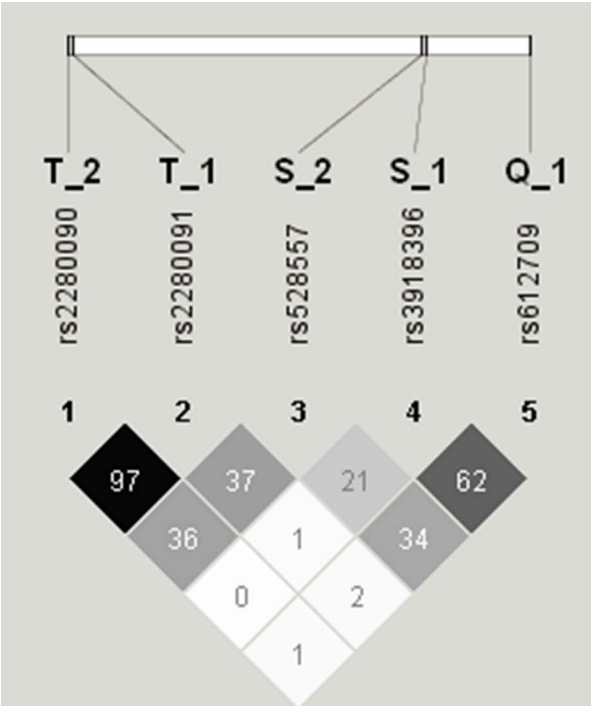


Figure 1 *ADAM33* linkage disequilibrium plot ($100 \cdot r^2$) in the Vlagtwedde-Vlaardingen cohort

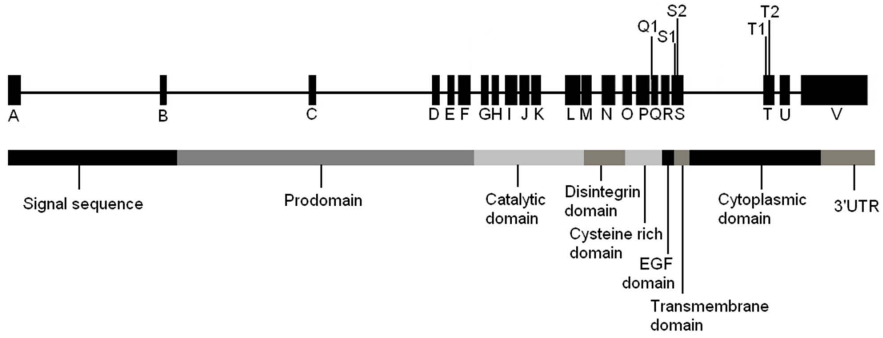


Figure 2 Position of genotyped SNPs in the *ADAM33* gene and the domain organization of *ADAM33* (adapted from (10))

Statistical analysis

Hardy-Weinberg Equilibrium was tested using the χ^2 test (cut-off value $p<0.05$). First, descriptive analyses were performed. Differences in genotype distribution between dead and alive subjects were tested using χ^2 tests. Cox proportional hazards regression adjusted for gender, age, FEV_1, height, place of residence and packyears of smoking (all at the 1989/1990 survey) was used to evaluate the association between SNPs and all-cause and cause-specific (COPD and cardiovascular) mortality. Time was defined from the examination in 1989/1990 until death, end of follow-up in 2008 or last registration if subjects were lost to follow-up. Survival curves were calculated using Cox regression models. In addition, stratified analyses according to gender and smoking habits were performed.

Logistic regression adjusted for the same covariates as in the Cox regression was used to calculate odds ratios and 95% confidence intervals for the chance of survival to the age of 75 and 85 years respectively, in relation to genotypes for every SNP separately.

Table 1 Characteristics of participants at visit 1989/1990 by vital status on Dec 31st, 2008

Status on 31-12-2008	Alive	Dead	p value
N (%)	1,087 (78.2)	284 (20.4)	
Males	525 (48.3)	178 (62.7)	<0.001
Age	49.4 (36.0 to 72.6)	61.9 (35.8 to 79.1)	<0.001
Ever smokers	711 (65.4)	219 (77.1)	<0.001
Packyears in ever smokers	17.2 (0.1-117.1)	27.1 (0.6-262.2)	<0.001
FEV_1, liters	2.96 (0.75)	2.48 (0.71)	<0.001
FEV_1, % predicted	94.6 (13.9)	84.3 (18.0)	<0.001
Causes of death			
COPD*		20 (7.0)	
Cardiovascular disease*		107 (37.7)	
External causes**		14 (4.9)	

All variables are expressed as number (%) or mean (SD) or median (range) as appropriate

* Either primary or secondary cause of death, number (% of all deaths)

** Suicides, homicides, traffic accidents *etc.*

P values <0.05 were considered statistically significant (tested 2-sided). All statistical analyses were performed using SPSS version 16.0 for Windows.

RESULTS

Table 1 shows the population characteristics at the survey in 1989/1990, according to vital status on December 31st, 2008. After 18 years of follow-up 78.2% (n=1,087) of the cohort was still alive.

Table 2 Distribution of genotypes according to all-cause and cause-specific mortality

SNP	Genotype	Alive n=1,087	All-cause mortality n=284	p value*	COPD mortality n=20	p value**	CVD mortality n=107	p value***
Q_1	CC	831 (77.2)	203 (77.5)		13 (65.0)		81 (78.6)	
	CT	226 (21.0)	55 (21.0)	0.965	5 (25.0)	0.024	21 (20.4)	0.822
	TT	19 (1.8)	4 (1.5)		2 (10.0)		1 (1.0)	
S_1	GG	913 (84.2)	228 (85.4)		12 (63.2)		91 (86.7)	
	GA	165 (15.2)	37 (13.9)	0.846	6 (31.6)	0.008	13 (12.4)	0.698
	AA	7 (0.6)	2 (0.7)		1 (5.2)		1 (0.9)	
S_2	GG	609 (57.3)	143 (54.2)		7 (35.0)		52 (50.5)	
	GC	386 (36.3)	99 (37.5)	0.446	9 (45.0)	0.023	44 (42.7)	0.399
	CC	68 (6.4)	22 (8.3)		4 (20.0)		7(6.8)	
T_2	GG	805 (76.9)	183 (69.3)		11 (57.9)		70 (67.3)	
	GA	230 (22.0)	70 (26.5)	0.001	5 (26.3)	<0.001	30 (28.9)	0.018
	AA	12 (1.1)	11 (4.2)		3 (15.8)		4 (3.8)	

***** Differences between alive subjects and those who died (excluding external causes of death) tested with χ^2 test

****** Differences between alive subjects and those who died due to COPD tested with χ^2 test

******* Differences between alive subjects and those who died due to CVD tested with χ^2 test

We had an almost perfect follow-up, since only 19 (1.4%) of the genotyped participants were lost to follow-up. Among all 284 deaths, 20 (7.0%) occurred due to COPD and 107 (37.7%) due to CVD. All tested SNPs were in Hardy-Weinberg equilibrium.

Table 3 Hazard ratio (95% CI) of all-cause, COPD and cardiovascular mortality

SNP	Genotype	All-cause mortality*	COPD mortality**	CVD mortality**
		HR (95% CI)	HR (95% CI)	HR (95% CI)
Q_1	CT	1.0 (0.8-1.4)	1.6 (0.5-5.3)	1.0 (0.6-1.7)
	TT	0.7 (0.2-2.2)	7.6 (1.6-37.2)***	0.5 (0.1-3.9)
S_1	GA	0.8 (0.5-1.2)	2.3 (0.8-7.1)	0.7 (0.4-1.3)
	AA	1.6 (0.4-6.6)	38.4 (3.8-389.3)***	2.3 (0.3-16.9)
S_2	GC	1.1 (0.9-1.5)	1.4 (0.4-4.6)	1.4 (0.9-2.2)
	CC	1.6 (1.0-2.6)	6.1 (1.6-23.1)***	1.5 (0.7-3.3)
T_2	GA	1.3 (1.0-1.7)	0.9 (0.2-3.2)	1.4 (0.9-2.3)
	AA	3.6 (2.0-6.7)***	13.8 (3.3-58.3)***	3.4 (1.2-9.5)***

Cox regression adjusted for gender, age, FEV_1, height, place of residence and packyears smoking (all at the last survey 1989/1990)

***** Excluding external causes of death

****** Primary or secondary causes of death

******* P value < 0.05

All-cause mortality

Table 2 shows the genotype distributions of alive subjects and those who had died during 18 years of follow-up. The distribution of SNP T_2 was significantly different between alive and dead subjects. Furthermore, individuals homozygous for the minor allele of SNP T_2 had a significantly increased hazard ratio for all-cause mortality compared to wild types, 3.6 (95% confidence interval 2.0 to 6.7) (Table 3). SNP T_2 showed increased all-cause mortality among those with the AA genotype (Figure 3). The other investigated SNPs

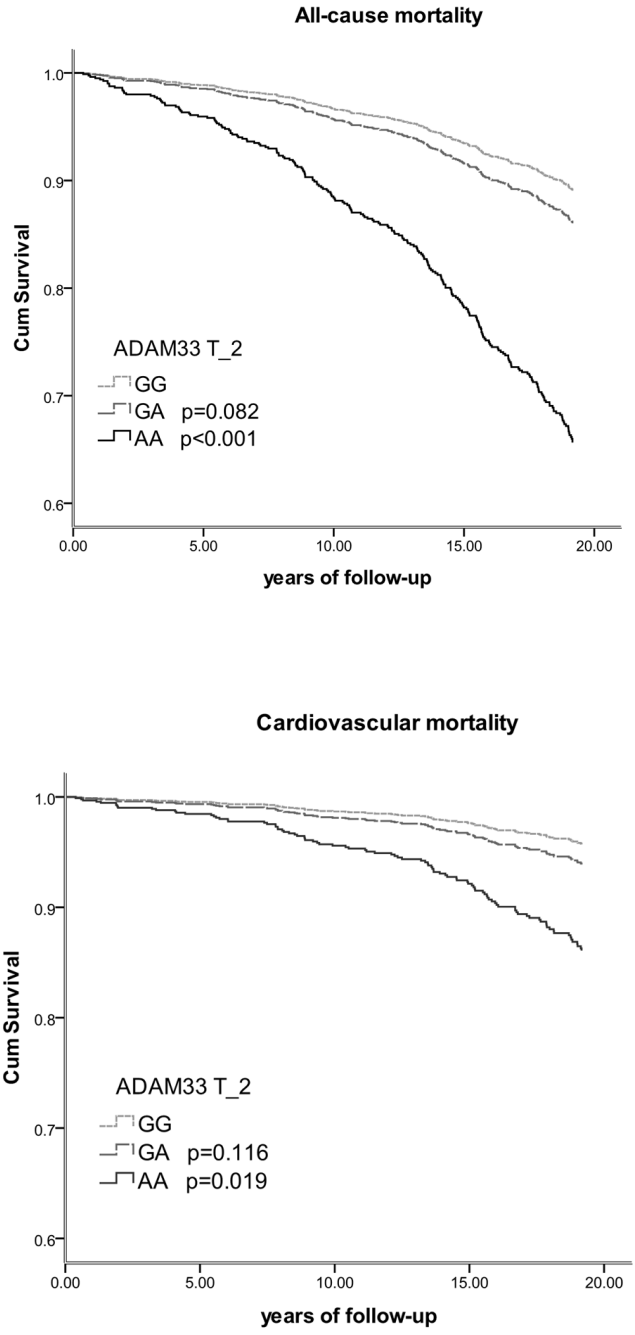


Figure 3 Survival curves for all-cause and CVD mortality according to SNP T_2

in *ADAM33* were not significantly associated with all-cause mortality. Table 4 presents the stratified analyses. The risk of all-cause mortality associated with SNP T_2 was similar in females (3.5, 1.5 to 8.3) and males (3.1, 1.2 to 7.6), as well as in never smokers (3.8, 0.9 to 16.3) and ever smokers (3.6, 1.8 to 7.2). Never smoking individuals, homozygous for the minor allele of SNP S_1 had a significantly increased all-cause mortality risk.

Table 4 Risk of all-cause mortality according to gender and smoking habits

SNP	Genotype	Gender		Smoking status	
		Females	Males	Never smokers	Ever smokers
		HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)
Q_1	CT	1.0 (0.6-1.7)	1.0 (0.6-1.7)	1.3 (0.7-2.5)	0.9 (0.7-1.4)
	TT	1.0 (0.1-7.2)	1.0 (0.1-7.2)	1.2 (0.2-9.0)	0.6 (0.1-2.2)
S_1	GA	0.6 (0.3-1.2)	0.6 (0.3-1.2)	0.7 (0.3-1.7)	0.8 (0.5-1.2)
	AA	4.2 (0.6-30.7)	4.2 (0.6-30.7)	7.9 (1.0-61.4)*	0.9 (0.1-6.8)
S_2	GC	1.1 (0.7-1.8)	1.1 (0.7-1.8)	0.9 (0.5-1.6)	1.2 (0.9-1.7)
	CC	1.6 (0.8-3.5)	1.6 (0.8-3.5)	1.7 (0.6-4.8)	1.6 (0.9-2.7)
T_2	GA	1.0 (0.6-1.6)	1.0 (0.6-1.6)	0.7 (0.4-1.4)	1.5 (1.1-2.1)*
	AA	3.5 (1.5-8.4)*	3.5 (1.5-8.4)*	3.8 (0.9-16.3)**	3.6 (1.8-7.1)*

Females n=676 (103 deaths); **Males** n=714 (166 deaths); **Never smokers** n=445 (62 deaths); **Ever smokers** n=945 (207 deaths); n=14 deaths due to external causes are excluded

* P value < 0.05

** P=0.07

COPD mortality

There were significant differences in genotype distribution for SNPs Q_1, S_1, S_2 and T_2 between alive subjects and those who died due to COPD (Table 2). Carriers of minor alleles of SNPs Q_1 S_1, S_2 and T_2 had a significantly increased COPD mortality risk compared to non-carriers (Table 3). Figure 4 shows a clear trend for higher COPD mortality for subjects homozygous for SNPs Q_1 S_1, S_2 and T_2.

Cardiovascular mortality

Individuals homozygous for the minor allele of SNP T_2 had a significantly increased risk of cardiovascular mortality compared to wild types (3.4, 1.2 to 9.5) (Table 3, Figure 3). Stratified analyses according to gender and smoking status showed that the risk of cardiovascular mortality among subjects homozygous for the minor allele of SNP T_2 was increased in all strata (all borderline significant; see Table S2). Also, females and never smokers who were homozygous for the minor allele of SNP S_1 had significantly increased cardiovascular mortality risk.

Chance of survival to ages 75 and 85 yrs

At age of 75 years, subjects with the AA genotype for SNP T_2 were more likely to have died than wild types (p=0.017, Table S3). Remarkably, none of the subjects with the AA genotype for SNP T_2 survived to the age of 85 years (Table S4).

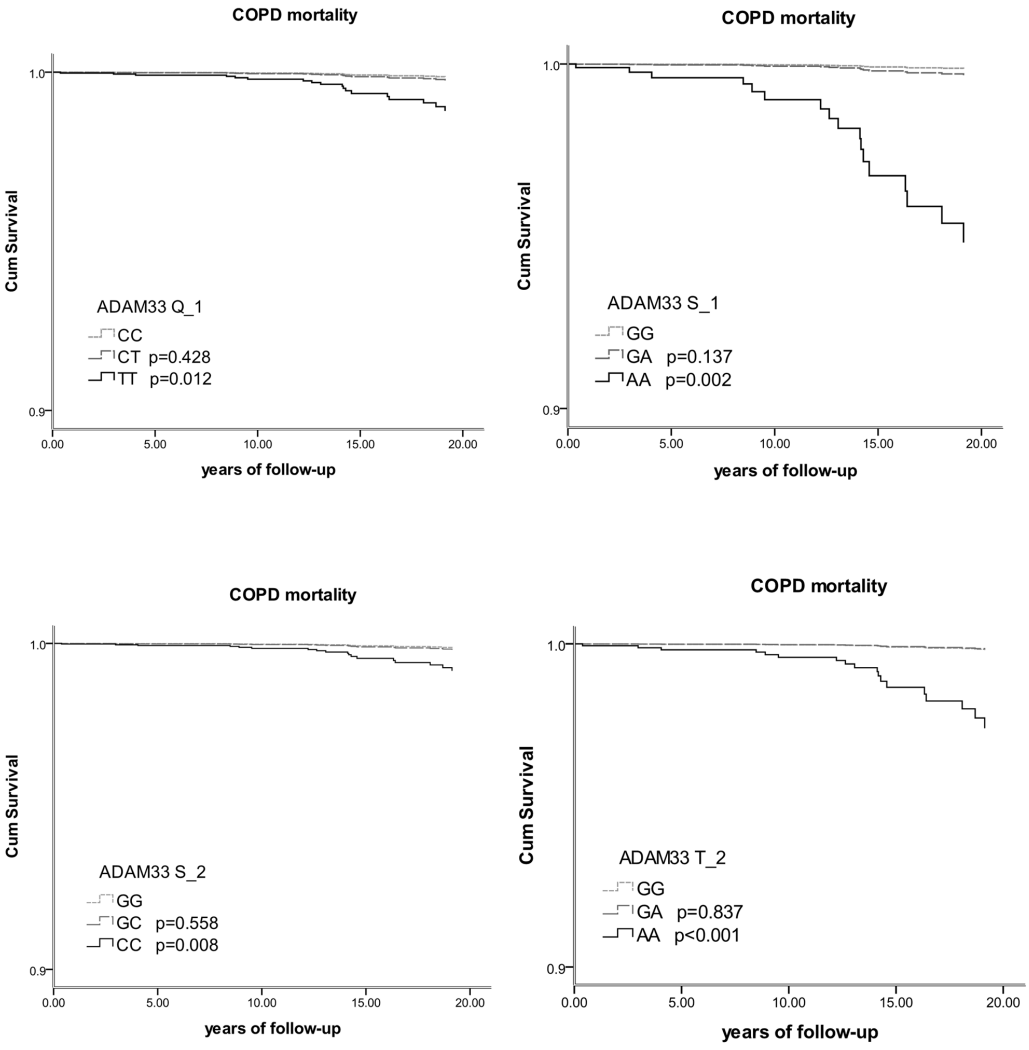


Figure 4 Survival curves for COPD mortality according to SNPs Q_1, S_1, S_2 and T_2

DISCUSSION

The present study shows for the first time that polymorphisms in *ADAM33* are associated with all-cause, COPD and cardiovascular mortality. Thus, *ADAM33* appears to constitute an important candidate gene explaining individual differences in human lifespan. So far, *ADAM33* has been related to pulmonary diseases and lung function decline (11-17). Since associations in our study are independent of lung function, these findings put a new light on the role of *ADAM33*. Of importance, effects of the SNPs were observed both in females and males, and in never and ever smokers, indicating the robustness of the associations between *ADAM33* and mortality.

ADAM33 is preferentially expressed in smooth muscle cells, myofibroblasts, and fibroblasts (5), suggesting that this protein may be important for the functionality of the whole human organism, and likely for the lungs and the cardiovascular system.

SNP T_2 showed the broadest range of associations, since carriers of the minor alleles had an increased mortality risk for every investigated cause of death. Interestingly, the all-cause mortality risk remained significantly increased when stratified analyses were performed for gender and smoking habits. Thus SNP T_2 is associated with reduced survival, independent of other risk factors.

ADAM33-null mice that do not express ADAM33 at all, do not exhibit morphological or behavioral abnormalities compared to wild type mice (18). These findings provide suggestive evidence that over-expression rather than under expression of the ADAM33 protein contributes to morbidity and in turn to mortality events.

Minor alleles of SNPs Q_1, S_1, S_2 and T_2 had a higher prevalence in subjects who died due to COPD than alive subjects, consistent with our previous findings showing a higher prevalence of the minor allele of these SNPs in subjects with COPD than in healthy controls (11). The latter study also reported an additional association between polymorphisms in *ADAM33* and accelerated lung function decline in the general population (11). The current study showed a higher risk of COPD mortality for individuals with polymorphisms in *ADAM33* independently of their lung function. This suggests that *ADAM33* plays a role not only in local airway events leading to impaired lung function, but also to disease progression or more extensive physiological processes which can contribute to poorer survival.

Overproduction of ADAM33 may lead to excessive shedding of inflammatory mediators and growth factors, which induce pathological states like proliferation of smooth muscle cells and fibroblasts observed in pulmonary and cardiovascular disorders (6;11).

ADAM33 protein isoforms occur in human embryonic lungs, suggesting a role in airway development (19). SNPs in *ADAM33* predict poor lung function in early childhood (20), thus it is plausible that ADAM33 plays a role in tissue development, and that minor alleles of the *ADAM33* SNPs lead to pathological conditions in lungs.

So far the role of *ADAM33* in cardiovascular disease is poorly understood. Holloway et al showed that *ADAM33* expression was higher in atherosclerotic lesions than in the normal vascular wall and found an association between an intronic polymorphism (rs574174 ST_7) in *ADAM33* and atherosclerosis severity (6). Moreover ADAM12, a member of the same subfamily and closely related to ADAM33, is involved in development of cardiac hypertrophy that leads to sudden cardiac death (21).

Taking all results into account, we suggest that SNPs in *ADAM33* can be considered a risk factor for all-cause and disease specific mortality. Furthermore, since we found that subjects with the AA genotype for SNP T_2 had a lower chance to reach the age of 75, and all carriers of this genotype had died before the age of 85 we believe that the current study is an important step towards identifying genes influencing human lifespan. This may suggest that screening for this SNP, probably in conjunction with other SNPs in genes, may identify subjects who are at risk for premature death. Additionally screening for SNP T_2 may allow direct identification of subjects at risk for COPD or cardiovascular mortality. Given the increased *ADAM33* expression in smooth muscle cells in atherosclerosis (6) and following our hypothesis that overexpression of *ADAM33* may lead to the pathological events, subjects at risk may receive tailored therapy with a special target on *ADAM33* levels or activity. SNP T_2 is located in the T-exon encoding a cytoplasmic domain. In this light it is interesting to note that loss of the membrane anchor and regulatory cytoplasmic domain of *ADAM33* results in a disease-related gain of function and release from cell membrane a soluble ADAM33 form. This form in turn induces endothelial cells differentiation and promotes angiogenesis, a process important in tissue inflammation and remodeling (22). Therefore if in subjects homozygous for the minor allele of T_2 a disease-related gain of function occurs, these may receive a special therapy. Our findings, that link *ADAM33* to the main leading diseases worldwide, reveal potential novel therapeutic targets. Hypothetically, a new drug that controls the unfavorable ADAM33 activity could prevent development of both COPD and CVD via regulation of pathological neovascularization. However to this end more studies are needed.

A major strength of this research is an excellent follow-up rate, since only 1.4% of genotyped participants could not be traced back after 18 years.

The small number of deaths due to COPD (i.e. 20) could be considered a limitation of our study. However, all SNPs which showed associations with COPD mortality were associated with COPD development in previous studies.

We decided to not correct our results for multiple testing, since our hypotheses were stated a priori and based on previous evidence indicating a role of *ADAM33* polymorphisms in pathophysiology of COPD and CVD, thus following Steiner's advice: "If the primary outcomes have been specified beforehand, then correcting for multiplicity may be too conservative and should be avoided" (23). Although adjustment for multiple testing will decrease the chance of type I error, it will also increase the likelihood of type II errors and potentially useful observations may be prematurely discarded (23;24).

In summary, this study implicates that *ADAM33* is involved in all-cause mortality and in mortality due to both COPD and cardiovascular disease and these associations are independent of level of lung function, gender and smoking habits. Our findings highlight the importance of *ADAM33* as a pleiotropic gene involved not only in pulmonary disease, but in cardiovascular disease as well. Since polymorphisms in this gene are associated with increased mortality risk and with a reduced chance of survival to age of 75, we believe that *ADAM33* may affect human lifespan. Future studies should focus on the functionality of the various SNPs in this gene to further unravel its role in ageing.

REFERENCES

1. Blagosklonny MV. Why human lifespan is rapidly increasing: solving „longevity riddle" with „revealed-slow-aging" hypothesis. Aging (Albany NY) 2010 Apr;2(4):177-82.
2. Swindell WR, Ensrud KE, Cawthon PM, Cauley JA, Cummings SR, Miller RA. Indicators of „Healthy Aging" in older women (65-69 years of age). A data-mining approach based on prediction of long-term survival. BMC Geriatr 2010;10:55.
3. Lao JI, Montoriol C, Morer I, Beyer K. Genetic contribution to aging: deleterious and helpful genes define life expectancy. Ann N Y Acad Sci 2005 Dec;1057:50-63.
4. Seals DF, Courtneidge SA. The ADAMs family of metalloproteases: multidomain proteins with multiple functions. Genes Dev 2003 Jan 1;17(1):7-30.
5. Van Eerdewegh P, Little RD, Dupuis J, Del Mastro RG, Falls K, Simon J, et al. Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. Nature 2002 Jul 25;418(6896):426-30.
6. Holloway JW, Laxton RC, Rose-Zerilli MJ, Holloway JA, Andrews AL, Riaz Z, et al. ADAM33 expression in atherosclerotic lesions and relationship of ADAM33 gene variation with atherosclerosis. Atherosclerosis 2010 Feb 24.
7. Edwards DR, Handsley MM, Pennington CJ. The ADAM metalloproteinases. Mol Aspects Med 2008 Oct;29(5):258-89.
8. Butcher SK, Lord JM. Stress responses and innate immunity: aging as a contributory factor. Aging Cell 2004 Aug;3(4):151-60.
9. Naumova E, Mihaylova A, Ivanova M, Michailova S, Penkova K, Baltadjieva D. Immunological markers contributing to successful aging in Bulgarians. Exp Gerontol 2004 Apr;39(4):637-44.
10. Gosman MM, Boezen HM, van Diemen CC, Snoeck-Stroband JB, Lapperre TS, Hiemstra PS, et al. A disintegrin and metalloprotease 33 and chronic obstructive pulmonary disease pathophysiology. Thorax 2007 Mar;62(3):242-7.
11. van Diemen CC, Postma DS, Vonk JM, Bruinenberg M, Schouten JP, Boezen HM. A disintegrin and metalloprotease 33 polymorphisms and lung function decline in the general population. Am J Respir Crit Care Med 2005 Aug 1;172(3):329-33.
12. Blakey JD, Sayers I, Ring SM, Strachan DP, Hall IP. Positionally cloned asthma susceptibility gene polymorphisms and disease risk in the British 1958 Birth Cohort. Thorax 2009 May;64(5):381-7.
13. Tripathi P. Haplotypic association of ADAM33 (T+1, S+1 and V - 3) gene variants in genetic susceptibility to asthma in Indian population. 2012 Nov.

14. Korytina GF, Tselousova OS, Akhmadishina LZ, Victorova EV, Zagidullin S, Victorova TV. [Association of the MMP3, MMP9, ADAM33 and TIMP3 genes polymorphic markers with development and progression of chronic obstructive pulmonary disease]. Mol Biol (Mosk) 2012 May;46(3):487-99.

15. Lee YH, Song GG. Association between ADAM33 T1 polymorphism and susceptibility to asthma in Asians. Inflamm Res 2012 Dec;61(12):1355-62.

16. Holgate ST. ADAM metallopeptidase domain 33 (ADAM33): identification and role in airways disease. Drug News Perspect 2010 Jul;23(6):381-7.

17. Sadeghnejad A, Ohar JA, Zheng SL, Sterling DA, Hawkins GA, Meyers DA, et al. Adam33 polymorphisms are associated with COPD and lung function in long-term tobacco smokers. Respir Res 2009;10:21.

18. Chen C, Huang X, Sheppard D. ADAM33 is not essential for growth and development and does not modulate allergic asthma in mice. Mol Cell Biol 2006 Sep;26(18):6950-6.

19. Haitchi HM, Powell RM, Shaw TJ, Howarth PH, Wilson SJ, Wilson DI, et al. ADAM33 expression in asthmatic airways and human embryonic lungs. Am J Respir Crit Care Med 2005 May 1;171(9):958-65.

20. Simpson A, Maniatis N, Jury F, Cakebread JA, Lowe LA, Holgate ST, et al. Polymorphisms in a disintegrin and metalloprotease 33 (ADAM33) predict impaired early-life lung function. Am J Respir Crit Care Med 2005 Jul 1;172(1):55-60.

21. Asakura M, Kitakaze M, Takashima S, Liao Y, Ishikura F, Yoshinaka T, et al. Cardiac hypertrophy is inhibited by antagonism of ADAM12 processing of HB-EGF: metalloproteinase inhibitors as a new therapy. Nat Med 2002 Jan;8(1):35-40.

22. Puxeddu I, Pang YY, Harvey A, Haitchi HM, Nicholas B, Yoshisue H, et al. The soluble form of a disintegrin and metalloprotease 33 promotes angiogenesis: implications for airway remodeling in asthma. J Allergy Clin Immunol 2008 Jun;121(6):1400-6, 1406.

23. Streiner DL, Norman GR. Correction for multiple testing: is there a resolution? Chest 2011 Jul;140(1):16-8.

24. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology 1990 Jan;1(1):43-6.

SUPPORTING INFORMATION

Table S1 ICD-codes for the investigated causes of death

Cause of death	ICD-9	ICD-10
External causes*	≥ 800	S, T, V, W, X, Y
COPD	490-492, 494, 496	J40-J44, J47
Cardiovascular disease	390-398, 401-405, 410-417, 420-438, 440-448, 451-459, 785.4	G45-G46, I00-I15, I20-I28, I30-I52, I60-I69, I70-I79, I80-I89, I95-I97, I98.2, I98.8, I99, M30-M31, N28.0, R02, R58

* Suicides, homicides, traffic accidents *etc.*

Table S2 Risk of cardiovascular mortality according to gender and smoking habits

SNP	Genotype	Gender		Smoking status	
		Females	Males	Never smokers	Ever smokers
		HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)
Q_1	CT	0.9 (0.4-2.1)	1.1 (0.6-2.0)	1.1 (0.4-3.2)	1.0 (0.6-1.8)
	TT	1.8 (0.2- 3.1)	-	2.4 (0.3-18.6)	-
S_1	GA	0.5 (0.1-1.5)	0.8 (0.4-1.7)	0.3 (0.0-2.3)	0.8 (0.4-1.5)
	AA	10.2 (1.3-78.7)*	-	17.4 (2.0-148.6)*	-
S_2	GC	2.0 (1.0-4.0)	1.2 (0.7-2.0)	1.4 (0.6-3.4)	1.4 (0.9-2.3)
	CC	1.6 (0.4-5.8)	1.3 (0.5-3.7)	1.9 (0.4-8.9)	1.3 (0.5-3.3)
T_2	GA	1.8 (0.9-3.6)	1.2 (0.7-2.2)	1.1 (0.4-3.0)	1.5 (0.9-2.5)
	AA	4.0 (0.9-18.3)	2.8 (0.7-11.7)	6.5 (0.8-51.7)	2.8 (0.9-9.3)

Females n=676 (40 deaths); **Males** n=714 (67 deaths); **Never smokers** n=445 (26 deaths); **Ever smokers** n=945 (81 deaths)

* P value < 0.05

Table S3 Distribution of genotypes according to being alive or dead at the age of 75, and chance of survival to this age

SNP	Genotype	Dead at the age of 75	Alive at the age of 75	p value*	Chance of survival to age of 75 OR (95% CI)
Q_1	CC	126 (77.3)	202 (77.4)		1
	CT	35 (21.5)	54 (20.7)	0.853	1.1 (0.6-2.2)
	TT	2 (1.2)	5 (1.9)		2.2 (0.2-22.8)
S_1	GG	142 (85.5)	226 (85.3)		1
	GA	23 (13.9)	37 (14.0)	0.982	1.4 (0.6-3.1)
	AA	1 (0.6)	2 (0.7)		0.5 (0.0-6.7)
S_2	GG	89 (54.6)	146 (55.7)		1
	GC	57 (35.0)	101 (38.6)	0.190	1.3 (0.8-2.4)
	CC	17 (10.4)	15 (5.7)		0.7 (0.3-1.8)
T_2	GG	111 (67.3)	192 (75.9)		1
	GA	45 (27.3)	58 (22.9)	0.017	1.0 (0.6-1.9)
	AA	9 (5.4)	3 (1.2)		0.2 (0.0-0.9)**

Logistic regression adjusted for gender, age, FEV_1, height, place of residence and packyears of smoking at survey in 1989/90

***** Differences between subjects alive at the age of 75 and those who died before this age tested with χ^2 test

****** P value=0.040

Table S4 Distribution of genotypes according to being alive or dead at the age of 85, and chance of survival to this age

SNP	Genotype	Dead at the age of 85	Alive at the age of 85	p value*	Chance of survival to age of 85 OR (95% CI)
Q_1	CC	197 (77.9)	44 (75.9)		1
	CT	52 (20.6)	13 (22.4)	0.947	1.2 (0.4-3.1)
	TT	4 (1.6)	1 (1.7)		1.0 (0.1-15.3)
S_1	GG	222 (86.0)	49 (81.7)		1
	GA	34 (13.2)	11 (18.3)	0.475	2.2 (0.8-6.5)
	AA	2 (0.8)	0 (0.0)		-
S_2	GG	138 (54.1)	36 (61.0)		1
	GC	96 (37.6)	19 (32.2)	0.629	0.9 (0.4-2.2)
	CC	21 (8.2)	4 (6.8)		1.1 (0.2-6.2)
T_2	GG	174 (68.2)	44 (78.6)		1
	GA	70 (27.5)	12 (21.4)	0.154	1.1 (0.4-2.8)
	AA	11 (4.3)	0 (0.0)		-

Logistic regression adjusted for gender, age, FEV_1, height, place of residence and packyears of smoking at survey in 1989/90

***** Differences between subjects alive at the age of 85 and those who died before this age tested with χ^2 test

Chapter 4

NFE2L2 polymorphisms, mortality and metabolism in the general population

Sylwia M Figarska, Judith M Vonk and H Marike Boezen

University of Groningen, University Medical Center Groningen, Department of Epidemiology, Hanzeplein 1, P.O. Box 30001, 9700 RB Groningen, The Netherlands

Submitted

ABSTRACT

The Nuclear Factor (Erythroid-derived 2)-Like 2 (*NFE2L2* or *NRF2*) gene regulates transcription of enzymes involved in cellular detoxification and lipids homeostasis. *NFE2L2* is associated with pathophysiology of atherosclerosis and Chronic Obstructive Pulmonary Disease (COPD). Therefore we studied the relation between *NFE2L2* and all-cause, cardiovascular and COPD mortality and its associations with triglyceride and cholesterol levels. We genotyped five tagging SNPs (rs4243387, rs2364723, rs13001694, rs1806649 and rs6726395) in *NFE2L2* in 1,390 subjects from the Vlagtwedde-Vlaardingen cohort. Participants were examined in 1989/1990 and followed up till the vital status evaluation on December 31st, 2008. Associations between SNPs and mortality were estimated using Cox proportional hazards regression and associations between SNPs and triglyceride and cholesterol levels were tested with linear regression.

After 18 years, 284 (20.4%) subjects had died, 107 due to cardiovascular disease and 20 due to COPD. Minor allele carriers of rs13001694 had a significantly reduced risk of all-cause mortality compared to wild types: Hazard Ratio 0.8 (95% Confidence Interval 0.6 to 1.0). Minor allele carriers of rs2364723 had significantly reduced risk of cardiovascular mortality: HR=0.5, (95% CI: 0.3-0.7). This result was consistent in stratified analyses: females 0.4 (0.2-0.7), males 0.6 (0.3-0.9), never smokers 0.5 (0.2-1.1), ever smokers 0.5 (0.3-0.8). Minor allele carriers of rs1806649 had a markedly reduced COPD mortality: HR=0.3 (95% CI: 0.1-0.9). Rs2364723 was associated with lower triglyceride levels. None of the SNPs was associated with cholesterol levels.

This study shows for the first time that *NFE2L2* is associated with reduced risk of all-cause, cardiovascular and COPD mortality in humans.

INTRODUCTION

Nearly 30% of the individual variance in life expectancy is genetically determined (1), however the specific genetic determinants of human lifespan still remain largely unknown. A candidate gene for predicting variation in human lifespan is the Nuclear Factor (Erythroid-derived 2)-Like 2 (*NFE2L2* or *NRF2*) gene. *NFE2L2* is a master regulator of antioxidant-related genes and genes that control immune and inflammatory responses and those involved in tissue remodeling (2), and thus has an important role in cytoprotection in the whole organism.

NFE2L2 is a basic leucine zipper transcription factor and regulates expression of genes, via direct binding to the antioxidant responsive element (ARE) in the target gene. The targets include genes encoding glutathione S-transferases, γ-glutamylcysteine ligases, heme oxygenase 1 and NADPH quinione oxidoreductase (3;4). *NFE2L2* is expressed in all tissues, with highest levels in the key detoxification organs being the kidney and liver (5). Despite the well-documented relationship between *NFE2L2* and cellular protective mechanisms relevant to aging, very few studies have evaluated the role of *NFE2L2* in mediating rates of aging and longevity (6-9). One of these studies indicated that *NFE2L2* may be important in longevity, since this gene was 6-fold overexpressed in the liver of the naked mole rat, a species with a very long lifespan compared to mice having a much shorter lifespan (7). In human fibroblasts it has been demonstrated that *NFE2L2* function declines in senescence, whereas silencing of *NFE2L2* leads to premature senescence (6). Furthermore, the broad role of *NFE2L2* in age-related diseases has been indicated by its association with the development of atherosclerosis (10-13) and COPD (14;15). Atherosclerosis is characterized by lipid deposition in the artery

wall, and the impact of *NFE2L2* on atherosclerosis development may be mediated by its role in lipid homeostasis (16). The previous studies, performed in mice, have shown that the expression of *NFE2L2* regulates the expression of lipogenic genes and affects lipid accumulation and deposition in aortic lesions (10).

The Vlagtwedde-Vlaardingen cohort offers the unique opportunity to investigate the role of *NFE2L2* in long-term survival, since subjects included in this study were followed for 18 years. Besides all-cause mortality, we evaluated the association between *NFE2L2* and cardiovascular and COPD mortality and we also tested the associations between SNPs in *NFE2L2* and triglyceride and cholesterol levels.

METHODS

Study population

We studied subjects of the Vlagtwedde-Vlaardingen cohort, a general population-based cohort of exclusively Caucasian individuals of Dutch descent, recruited from Vlagtwedde, a rural area, and Vlaardingen, an urban area in the Netherlands (17). This cohort started in 1965, and participants had medical exams every 3 years until the last survey in 1989/1990. In each survey the Dutch version of the British Medical Council standardized questionnaire was filled in, and spirometry was performed. In this study we included 1,390 subjects (16.4% of the original cohort) out of 2,467 subjects of whom DNA was collected in the final survey in 1989/1990 (those with DNA samples contained more than 1500ng isolated DNA). There were no differences in characteristics between the selected (n=1390) and non-selected group (17). The Vlagtwedde-Vlaardingen cohort was set up to study respiratory health in the general population.

During survey 2 and 3 (in 1970/72 and 1973/75) an add-on study was performed by cardiologists. At that time the prevalence of cardiovascular problems was much higher in males than in females and therefore only males were included in this cardiovascular add-on study. In the surveys of 1970/72 and 1973/75, fasting serum triglyceride and total cholesterol levels were measured in 493 males out of the 1,390 genotyped subjects. The vital status of all participants in the Vlagtwedde-Vlaardingen study was assessed on December 31, 2008. We evaluated three mortality outcomes, i.e. all-cause mortality (excluding external causes of death) and cardiovascular and COPD mortality (either as primary or secondary cause of death). The causes of death were coded according to the International Classification of Diseases (ICD-9 and ICD-10, Table 1). Analyses on cause specific mortality were performed at Statistics Netherlands (The Hague). The study protocol was approved by the local university medical hospital ethics committee, University of Groningen, University Medical Center Groningen, The Netherlands and all participants gave their written informed consent.

Table 1 ICD-codes for the investigated causes of death

Cause of death	ICD-9	ICD-10
External causes*	≥ 800	S, T, V, W, X, Y
Cardiovascular disease	390-398, 401-405, 410-417, 420-438, 440-448, 451-459, 785.4	G45-G46, I00-I15, I20-I28, I30-I52, I60-I69, I70-I79, I80-I89, I95-I97, I98.2, I98.8, I99, M30-M31, N28.0, R02, R58
COPD	490-492, 494, 496	J40-J44, J47

*suicides, homicides, traffic accidents *etc.*

Samples collection, DNA extraction and genotyping

In 1989–1990 neutrophil depots from peripheral blood samples were collected and stored at –20°C. In 2003–2004 DNA was extracted from these samples with a QIAamp DNA blood mini kit (Qiagen, Hilden, Germany) and checked for purity and concentration with a NanoDrop ND-1000 UV–Vis spectrophotometer (NanoDrop Technologies, Wilmington, DE) (17). We pairwise tagged *NFE2L2* with five SNPs according to the HapMap CEU genotype data (23a) with r^2 threshold of 0.8 and Minor Allele Frequency (MAF) >5%. SNPs were genotyped at K-Bioscience Ltd (UK) using their patent-protected competitive allele specific PCR system (KASPar) (18). Figure 1 shows the position of genotyped SNPs in the *NFE2L2* gene and linkage disequilibrium (LD) between genotyped SNPs in the Vlagtwedde-Vlaardingen cohort. Hardy-Weinberg Equilibrium was tested using the χ^2 test (cut-off value $P<0.05$).

Statistical analysis

First, descriptive analyses were performed. Differences in genotype distribution between deceased and alive subjects were tested using χ^2 tests. Cox proportional hazards regression adjusted for gender, age, Forced Expiratory Volume in 1 second (FEV_1) %predicted and packyears of smoking (all at the 1989/1990 survey) was used to evaluate the association between SNPs and all-cause, cardiovascular and COPD mortality. Time was defined from the examination in 1989/1990 until death, end of follow-up in 2008 or last registration if subjects were lost to follow-up. Survival curves were depicted based on these Cox models. In addition, stratified analyses according to gender and smoking habits were performed.

Linear regression adjusted for age at the measurement was used to evaluate the associations between lipid profile i.e. triglyceride and total

cholesterol levels and SNPs. The triglyceride and cholesterol levels were logarithmically transformed to obtain a normal distribution. P values <0.05 were considered statistically significant (tested 2-sided). All statistical analyses were performed using IMB SPSS 20 software.

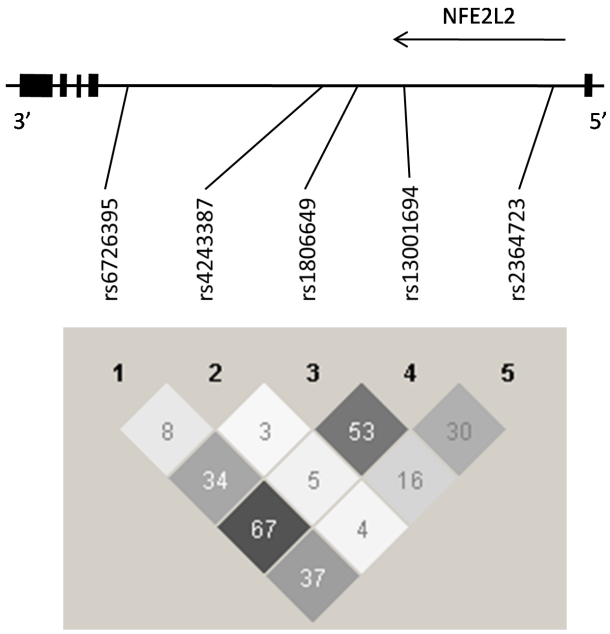


Figure 1 Position of genotyped SNPs in the *NFE2L2* gene and linkage disequilibrium plot ($100 \cdot r^2$) in the Vlagtwedde-Vlaardingen cohort
The black boxes represent exons.

RESULTS

Table 2 shows the population characteristics at the survey in 1989/1990, according to vital status on December 31st, 2008. After 18 years of follow-up 78.2% (n=1,087) of the cohort was still alive. We had an almost perfect follow-up, since only 19 (1.4%) of the genotyped participants were lost to follow-up. Among all 284

deaths, 107 (37.7%) occurred due to CVD, and 20 (14.2%) due to COPD. The 14 (4.9%) deaths due to external causes were excluded from the analyses. All tested SNPs were in Hardy-Weinberg equilibrium. A comparison of the characteristics of the genotyped subjects at baseline (survey in 1989/90) stratified by genotype for each SNP, (data not shown) shows that, except for a slightly higher FEV_1 % predicted in subjects homozygous for the minor allele of rs1806649, no differences between the genotypes were seen.

Table 2 Characteristics of participants at visit 1989/1990 by vital status on Dec 31st, 2008

Status on 31-12-2008	Alive	Deceased	p value
N (%)	1,087 (78.2)	284 (20.4)	
Males	525 (48.3)	178 (62.7)	<0.001
Vlagtwedde	981 (90.2)	258 (90.8)	0.761
Age at last status	68.4 (54.2-90.8)	72.6 (41.4-87.4)	<0.001
Ever smokers	711 (65.4)	219 (77.1)	<0.001
Packyears in ever smokers	17.2 (0.1-117.1)	27.1 (0.6-262.2)	<0.001
FEV_1 % predicted	94.6 (13.9)	84.3 (18.0)	<0.001
Causes of death			
Cardiovascular disease*		107 (37.7)	
COPD*		20 (7.0)	
External causes		14 (4.9)	
Other causes		149 (52.5)	

All variables are expressed as number (%) or mean (SD) or median (range) as appropriate

* Either as primary or secondary cause of death, 6 subjects had both cardiovascular disease and COPD as primary and secondary cause of death

Table 3 Distribution of genotypes and Hazard Ratios for all-cause, cardiovascular and COPD mortality

SNP	Genotype	Alive	Deceased	CVD deceased	COPD deceased	**All-cause mortality**	**Cardiovascular mortality**	**COPD mortality**
						HR (95% CI)	**HR (95% CI)**	**HR (95% CI)**
rs4243387	TT	898 (85.0)	238 (85.3)	84 (80.0)	19 (95.0)			
	TC+CC	158 (15.0)	41 (14.7)	21 (20.0)	1 (5.0)	0.96 (0.67-1.37)	1.52 (0.93-2.47)	0.42 (0.06-3.23)
rs2364723	GG	481 (46.1)	117 (41.9)	60 (56.6)	6 (30.0)			
	GC+CC	562 (53.9)	162 (58.1)	46 (43.4)*	14 (70.0)	0.96 (0.74-1.25)	0.49 (0.33-0.74)*	1.46 (0.45-4.72)
rs13001694	AA	354 (33.7)	117 (42.4)*	40 (38.5)	9 (45.0)			
	AG+GG	697 (66.3)	159 (57.6)	64 (61.5)	11 (55.0)	0.77 (0.59-1.00)*	0.89 (0.58-1.34)	0.47 (0.18-1.24)
rs1806649	CC	549 (53.8)	153 (57.5)	55 (54.5)	12 (66.7)			
	CT+TT	471 (46.2)	113 (42.5)	46 (45.5)	6 (33.3)	0.94 (0.72-1.22)	0.96 (0.63-1.46)	0.26 (0.08-0.89)*
rs6726395	GG	305 (28.9)	95 (34.4)	29 (27.4)	9 (45.0)			
	GA+AA	751 (71.1)	181 (65.6)	77 (72.6)	11 (55.0)	0.89 (0.67-1.17)	1.23 (0.78-1.94)	0.48 (0.18-1.28)

* P value < 0.05

All-cause, cardiovascular and COPD mortality

Table 3 shows the genotype distributions and the hazard ratios of alive subjects and those who had died during 18 years of follow-up. Among subjects who died, the carriers of the minor allele of SNP rs13001694 were significantly less common (58%) than in alive subjects (66%). Furthermore, individuals carrying the minor allele of rs13001694 had a significantly reduced hazard ratio for all-cause mortality compared to wild types (Hazard Ratio 0.77 (95% Confidence Interval (0.59 to 1.00), *P*=0.049)), (Table 3). Stratified analysis according to gender showed no significant HR in females and males, (0.83 (0.54-1.28) and 0.77 (0.55-1.07), respectively), (Table 4). The stratified analysis according to smoking habits showed that the protective effect of rs13001694 was observed in ever smokers only: HR=0.71 (95% CI (0.53-0.95), *P*=0.023), (Table 5). Carriers of the minor allele of rs2364723 had a significantly reduced risk of cardiovascular mortality: HR=0.49, (0.33-0.74). In stratified analyses the risk of cardiovascular mortality associated with SNP rs2364723 showed to be robust: females 0.35 (0.17-0.70), males 0.56 (0.33-0.93), never smokers 0.46 (0.20-1.07), ever smokers 0.50 (0.31-0.79), (Table 4 and Table 5). Carriers of the minor allele of rs1806649 had a reduced risk of COPD mortality: HR=0.26 (0.08-0.89). Survival curves according to genotypes of rs13001694, rs2364723 and rs1806649 clearly show the differences in mortality risk (Figure 2, 3 and 4 respectively). Dominant genetic models (homozygous wild type individuals versus heterozygous/homozygous minor allele individuals) were used for all analyses, since the number of individuals homozygous for the minor allele was low, especially within subjects who died due to CVD or COPD. Results of co-dominant models (data not shown) indicate that, in general, the HRs for heterozygotes and homozygotes for the minor allele are similar.

Table 4 Distribution of genotypes and Hazard Ratios for all-cause and cardiovascular mortality stratified according to gender

SNP	Geno-type	Females					Males				
		Alive	Deceased	CVD deceased	All-cause mortality HR (95% CI)	CVD mortality HR (95% CI)	Alive	Deceased	CVD deceased	All-cause mortality HR (95% CI)	CVD mortality HR (95% CI)
rs4243387	TT	467 (86.5)	93 (88.6)	33 (82.5)			431 (83.5)	145 (83.3)	51 (78.5)		
	TC+CC	73 (13.5)	12 (11.4)	7 (17.5)	0.96 (0.51-1.81)	1.71 (0.75-3.91)	85 (16.5)	29 (16.7)	14 (21.5)	0.98 (0.63-1.51)	1.47 (0.81-2.70)
rs2364723	GG	257 (47.9)	44 (41.9)	26 (65.0)			224 (44.2)	73 (42.0)	34 (51.5)		
	GC+CC	279 (52.1)	61 (58.1)	14 (35.0)*	0.91 (0.59-1.39)	0.35 (0.17-0.70)	283 (55.8)	101 (58.0)	32 (48.5)	0.95 (0.68-1.33)	0.56 (0.33-0.93)
rs13001694	AA	180 (33.1)	40 (39.2)	12 (31.6)			174 (34.3)	77 (44.3)	28 (42.4)		
	AG+GG	364 (66.9)	62 (60.8)	26 (68.4)	0.83 (0.54-1.28)	1.25 (0.60-2.61)	333 (65.7)	97 (55.7)*	38 (57.6)	0.77 (0.55-1.07)	0.78 (0.47-1.31)
rs1806649	CC	290 (54.7)	54 (54.0)	20 (52.6)			259 (52.9)	99 (59.6)	35 (55.6)		
	CT+TT	240 (45.3)	46 (46.0)	18 (47.4)	1.25 (0.82-1.92)	1.23 (0.62-2.44)	231 (47.1)	67 (40.4)	28 (44.4)	0.82 (0.58-1.15)	0.89 (0.53-1.50)
rs6726395	GG	162 (29.5)	35 (34.0)	10 (25.0)			143 (28.2)	60 (34.7)	19 (28.8)		
	GA+AA	387 (70.5)	68 (66.0)	30 (75.0)	0.86 (0.55-1.33)	1.44 (0.67-3.09)	364 (71.8)	113 (65.3)	47 (71.2)	0.93 (0.65-1.32)	1.17 (0.66-2.07)

* P value < 0.05

Table 5 Distribution of genotypes and Hazard Ratios for all-cause and cardiovascular mortality stratified according to smoking habits

SNP	Geno-type	Never Smokers					Ever Smokers				
		Alive	Deceased	CVD deceased	All-cause mortality HR (95% CI)	CVD mortality HR (95% CI)	Alive	Deceased	CVD deceased	All-cause mortality HR (95% CI)	CVD mortality HR (95% CI)
rs4243387	TT	313 (87.2)	55 (87.3)	21 (84.0)			585 (83.9)	183 (84.7)	63 (78.8)		
	TC+CC	46 (12.8)	8 (12.7)	4 (16.0)	0.93 (0.40-2.19)	1.59 (0.53-4.71)	112 (16.1)	33 (15.3)	17 (21.2)	0.99 (0.66-1.47)	1.53 (0.89-2.66)
rs2364723	GG	161 (45.2)	29 (45.3)	16 (61.5)			320 (46.6)	88 (40.9)	44 (55.0)		
	GC+CC	195 (54.8)	35 (54.7)	10 (38.5)	0.88 (0.51-1.53)	0.46 (0.20-1.07)	367 (53.4)	127 (59.1)	36 (45.0)	0.96 (0.71-1.29)	0.50 (0.31-0.79)*
rs13001694	AA	128 (35.9)	18 (28.6)	6 (25.0)			226 (32.6)	99 (46.5)	34 (42.5)		
	AG+GG	229 (64.1)	45 (71.4)	18 (75.0)	1.10 (0.61-2.00)	1.57 (0.57-4.33)	468 (67.4)	114 (53.5)*	46 (57.5)	0.71 (0.53-0.95)*	0.80 (0.50-1.27)
rs1806649	CC	199 (56.2)	32 (51.6)	12 (48.0)			350 (52.6)	121 (59.3)	43 (56.6)		
	CT+TT	155 (43.8)	30 (48.4)	13 (52.0)	1.47 (0.84-2.55)	1.44 (0.62-3.33)	316 (47.4)	83 (40.7)	33 (43.4)	0.84 (0.62-1.14)	0.86 (0.53-1.39)
rs6726395	GG	112 (30.9)	16 (25.8)	6 (23.1)			193 (27.8)	79 (36.9)	23 (28.8)		
	GA+AA	250 (69.1)	46 (74.2)	20 (76.9)	1.12 (0.60-2.09)	1.49 (0.55-4.04)	501 (72.2)	135 (63.1)*	57 (71.2)	0.85 (0.62-1.15)	1.19 (0.71-2.00)

* P value < 0.05

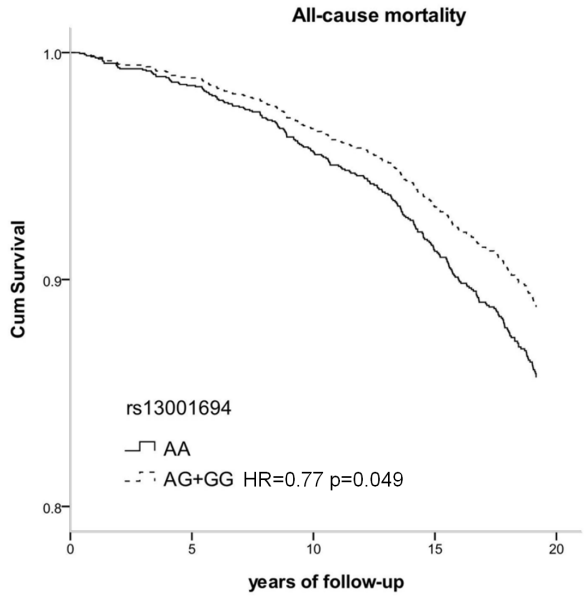


Figure 2 Survival curves for all-cause mortality according to SNP rs13001694

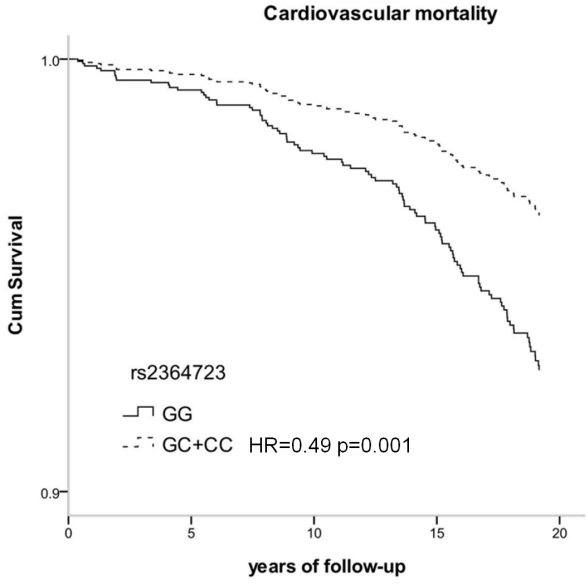


Figure 3 Survival curves for cardiovascular mortality according to SNP rs2364723

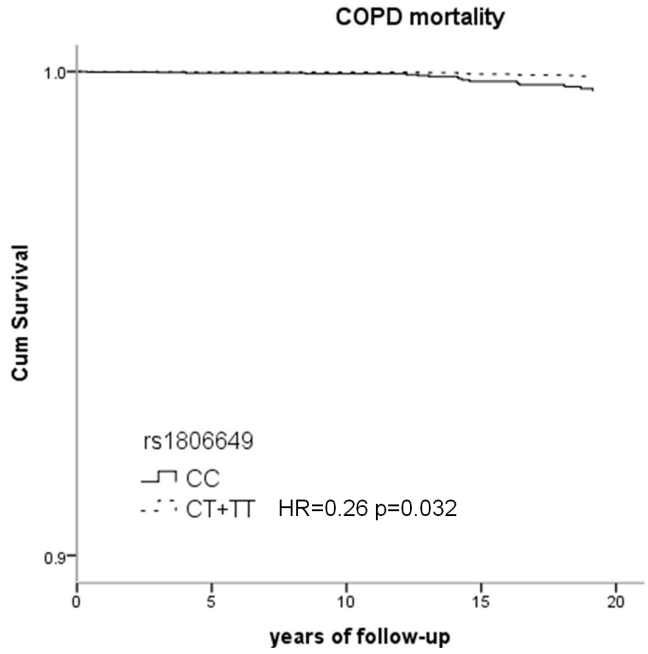


Figure 4 Survival curves for COPD mortality according to SNP rs1806649

Triglyceride and cholesterol levels

Heterozygotes and homozygotes for the minor allele of rs2364723 had 0.084 lower ln(triglyceride levels) compared to the wild types, whereas heterozygotes and homozygotes for the minor allele of rs13001694, rs1806649 and rs6726395 had respectively 0.110, 0.113 and 0.107 higher ln(triglyceride levels) compared to their wild types.

None of the SNPs was significantly associated with cholesterol levels (Table 6). Results of co-dominant models (data not shown) indicate that, in general, the effect estimates for heterozygotes and homozygotes for the minor allele are comparable.

Table 6 Estimated effects of the NFE2L2 genetic variation on Ln-transformed triglyceride and cholesterol levels

SNP	Genotype	Triglyceride levels				Cholesterol levels			
		N	B*	SE	p value	N	B*	SE	p value
rs4243387	TT	400				345			
	TC+CC	83	0.015	0.056	0.783	76	-0.020	0.023	0.397
rs2364723	GG	218				184			
	GC+CC	259	-0.084	0.043	0.050	231	-0.006	0.018	0.756
rs13001694	AA	176				164			
	AG+GG	299	0.110	0.044	0.013	249	0.026	0.019	0.172
rs1806649	CC	254				227			
	CT+TT	205	0.113	0.043	0.009	174	0.003	0.019	0.875
rs6726395	GG	142				130			
	GA+AA	336	0.107	0.047	0.022	288	0.014	0.020	0.486

* Regression coefficient (B), its standard error (SE) and p value obtained with linear regression analysis adjusted for age at the measurement

Cigarette smoking has an adverse effect on blood lipids, and indeed in the Vlagtwedde-Vlaardingen cohort smoking (packyears) was associated with triglyceride and cholesterol levels. However, smoking does not seem to be a confounder of the associations between lipids and SNPs, since in the analysis additionally adjusted for smoking (packyears) the effects of the SNPs on triglycerides and cholesterol levels remained similar (data not shown).

DISCUSSION

We found a 20% reduced mortality risk among minor allele carriers of SNP rs13001694 in *NFE2L2*, during 18 years of follow-up in the general population. Recently, a study that compared expression of oxidoreduction genes in the naked mole rat, the longest-living rodent known, to expression in mice, has shown that *NFE2L2* was 6-fold higher expressed in the rats (26). Here we show evidence to support the hypothesis that *NFE2L2* may be one of the genes contributing to individual differences in human lifespan. In the further analysis of survival in the Vlagtwedde-Vlaardingen cohort, we showed that the effect is gender-independent, since the effect was similar in females and males. In stratification according to smoking habits we found the significant protective effect of rs13001694 in ever smokers only, so it seems that the effect of the A/G substitution in rs13001694 may be exerted only under oxidative stress conditions. Indeed, a previous study showed that cigarette smoke exposure, a potent source of oxidative stress in the human lungs (19), may influence *NFE2L2* activity in alveolar macrophages by inducing *NFE2L2* nuclear accumulation and up-regulation of *NFE2L2* target genes (20). Furthermore, SNP rs10183914, which is highly correlated with rs13001694 (r^2=0.97; 1000 Genomes project) showed a protective

effect on lung function decline in smokers in the Lung Health Study (LHS) (15). However this result was not replicated when rs13001694 was tested in smokers from the Vlagtwedde-Vlaardingen cohort (15). Sandford et al. explain that the lack of replication may be due to differences in recruitment between the two studies, as the LHS selected mild to moderate COPD patients and the Vlagtwedde-Vlaardingen cohort was from the general population and the genetic factors associated with lung function decline in COPD patients and in the general population could be different (15). In our study, we adjusted the analysis for FEV_1 % predicted, thus the association of rs13001694 with better survival is independent of lung function level.

Another important finding in our study is that SNP rs2364723 is associated with a reduced risk of cardiovascular mortality. Even more interesting, the protective effect of this SNP was consistent within groups that have different mortality risks, i.e. females and males and never and ever smokers. This finding is of high importance since cardiovascular disease is a main cause of morbidity and a leading cause of death in elderly. A previous study in the Vlagtwedde-Vlaardingen cohort on the *NFE2L2* gene investigated its relation to lung function level and decline (18). Interestingly SNP rs2364723 was associated with a lower FEV_1 level and the same direction of the effect was observed in the British 1958 Birth cohort (18;21). Based on our finding showing the protective effect of this SNP on cardiovascular mortality, different pathways via which *NFE2L2* acts in pulmonary and cardiovascular events might be involved. There is evidence that *NFE2L2* plays a key role in preserving a healthy endothelial phenotype and maintaining the functional integrity of vasculature (22). Furthermore *NFE2L2* dysfunction may lead to functional impairment of arteries, increasing susceptibility of blood

vessels to injury in metabolic diseases (23). *NFE2L2* is essential for normal endothelial angiogenic processes and its dysfunction may be a potential mechanism underlying impaired angiogenesis and decreased microvascular density in aging (22). Several studies suggest that *NFE2L2* alters susceptibility to atherosclerosis (10-13). One of these studies indicated that activation of *NFE2L2* may exert antiatherogenic effects in vascular endothelium by suppressing inflammation (12), whereas another showed *NFE2L2* expression unexpectedly promotes atherosclerotic lesions formation (10). *NFE2L2* may affect atherosclerosis development since it regulates hepatic lipid homeostasis via activation of lipogenic genes expression (10;16;24). We found that rs2364723 was associated with lower triglyceride levels in male subjects. Triglyceride level is highly related to cardiovascular mortality, therefore it is plausible that rs2364723 contributes to the observed, reduced cardiovascular mortality via its favorable effect on triglyceride levels. The publicly available data of the British 1958 Birth cohort (21) showed that rs2364723 is also associated with lower cholesterol levels and it gives us another intermediate phenotype by which rs2364723 exerts its overall protective effect on cardiovascular mortality. This SNP is in LD (r^2=0.99) with promoter polymorphism rs35652124 (-653A/G) in the Vlagtwedde-Vlaardingen cohort (18). Marzec et al. have reported that rs35652124 impedes the transcriptional activity of *NFE2L2 in vitro* (25). Hence, it is possible that rs2364723 is not the causative variant, but that its association with cardiovascular mortality, due to almost complete LD, represents the effect of rs35652124. *NFE2L2* knockout mice exhibit around 50% reduction in the degree of aortic atherosclerosis compared to the wild type littermates (10), thus in this light a SNP that attenuates *NFE2L2* expression or activity is indeed likely to have the protective potential for cardiovascular

morbidity and mortality in humans. Furthermore, it is known that decreased expression of *NFE2L2* in vitro is associated with acute lung injury, characterized by pulmonary edema and inflammation (25). Therefore it seems that *NFE2L2* may have different local vascular and pulmonary effects, what could explain the opposite effects of rs2364723 on cardiovascular mortality and triglyceride levels on the one hand and on lung function level (FEV_1) on the other hand.

The other investigated SNPs (rs13001694, rs1806649 and rs6726395) were associated with increased triglyceride levels and interestingly these SNPs were also associated with increased cholesterol levels in the British 1958 Birth cohort (21), which confirms that these SNPs indeed affect metabolism of lipoproteins in humans.

We found a reduced risk of COPD mortality in carriers of the T allele of rs1806649 and we thus confirm the relevance of this SNP in COPD. Canova et al. have already introduced the wild type allele (C) of rs1806649 as a factor increasing the risk of air pollution-induced asthma/COPD hospital admissions (26). In this light indeed the T allele of rs1806649 may have protective potential as compared to the C allele. It would be interesting to further investigate whether this effect differs depending on smoking habits, but during 18 years of follow-up only 20 deaths due to COPD occurred and stratification according to smoking habits was not feasible due to lack of study power. However, all these subjects who died due to COPD were ever smokers and we showed that the protective effect is observed in this group.

The functionally important SNP rs6721961 (-617C/A) located in the *NFE2L2* promoter region significantly impedes *NFE2L2* function and is associated with increased risk of lung cancer (27) and acute lung

injury (25). In our study we tested rs4243387, which is in perfect LD (r^2=1.0) with rs6721961, but we did not observe an association with mortality risk. With regards to the longevity phenotype, according to the GWAS Central database (http://www.gwascentral.org) none of the previous GWA studies reported any SNP in *NFE2L2* to be associated with longevity in humans.

Polymorphisms in *NFE2L2* may alter expression of *NFE2L2*, or its ability to translocate from cytoplasm to the nuclear binding sites (25). Based on the data presented by Marzec et al. two functional SNPs (rs6721961 and rs6706649) impede the activity of the promoter region by more than 50% (indicated by lower luciferase activity *in vitro*) and one (rs35652124) by around 50% (25). Thus in humans we would not expect changes close to 6-fold higher expression or activity, as it was shown in the naked mole rat (7). However, it is hard to predict to what extent the expression of *NFE2L2* might be altered by SNPs. It may also differ between different organs, such as lungs, liver, brain or heart and their exposure to oxidative stress.

Strengths & limitations

The major strength of the current study is the longitudinal design. The follow-up period of 18 years provided a wide time window for evaluating survival of subjects, who were sampled from the general population. Also the high follow-up rate should be mentioned, since 98.6% of the included subjects could be traced back.

We could evaluate the associations between triglyceride and cholesterol levels in males only, since these measurements were not performed in females in the Vlagtwedde-Vlaardingen cohort. Nonetheless in the previous studies most of the effects of *NFE2L2* on

lipids levels were observed in the male mice (10) and we were able to investigate the relations, previously shown in mice, in humans.

The small number of deaths due to COPD (i.e. 20) could be considered a limitation of our study. However, SNP rs1806649 which showed association with reduced COPD mortality was previously associated with asthma/COPD hospital admissions in the same direction (26) making our finding consistent with the results of previous studies.

In summary this is the first study showing that *NFE2L2* plays an important role in human survival, as shown by its associations with reduced all-cause, cardiovascular and COPD mortality in the general population.

ACKNOWLEDGEMENTS

We acknowledge use of genotype data from the British 1958 Birth Cohort DNA collection, funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02.

GRANTS

This work was supported y Graduate School for Drug Exploration (GUIDE) University of Groningen, University Medical Center Groningen, the Netherlands; and the Netherlands Asthma Foundation (grant number 3.2.02.51).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

REFERENCES

1. 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 Mar;97(3):319-23.
2. Hybertson BM, Gao B, Bose SK, McCord JM. Oxidative stress in health and disease: the therapeutic potential of Nrf2 activation. Mol Aspects Med 2011 Aug;32(4-6):234-46.
3. Nguyen T, Yang CS, Pickett CB. The pathways and molecular mechanisms regulating Nrf2 activation in response to chemical stress. Free Radic Biol Med 2004 Aug 15;37(4):433-41.
4. Chorley BN, Campbell MR, Wang X, Karaca M, Sambandan D, Bangura F, et al. Identification of novel NRF2-regulated genes by ChIP-Seq: influence on retinoid X receptor alpha. Nucleic Acids Res 2012 Aug;40(15):7416-29.
5. Lewis KN, Mele J, Hayes JD, Buffenstein R. Nrf2, a guardian of healthspan and gatekeeper of species longevity. Integr Comp Biol 2010 Nov;50(5):829-43.
6. Kapeta S, Chondrogianni N, Gonos ES. Nuclear erythroid factor 2-mediated proteasome activation delays senescence in human fibroblasts. J Biol Chem 2010 Mar 12;285(11):8171-84.
7. Yu C, Li Y, Holmes A, Szafranski K, Faulkes CG, Coen CW, et al. RNA sequencing reveals differential expression of mitochondrial and oxidation reduction genes in the long-lived naked mole-rat when compared to mice. PLoS One 2011;6(11):e26729.
8. Suh JH, Shenvi SV, Dixon BM, Liu H, Jaiswal AK, Liu RM, et al. Decline in transcriptional activity of Nrf2 causes age-related loss of glutathione synthesis, which is reversible with lipoic acid. Proc Natl Acad Sci U S A 2004 Mar 9;101(10):3381-6.
9. Tomobe K, Shinozuka T, Kuroiwa M, Nomura Y. Age-related changes of Nrf2 and phosphorylated GSK-3beta in a mouse model of accelerated aging (SAMP8). Arch Gerontol Geriatr 2012 Mar;54(2):e1-e7.
10. Barajas B, Che N, Yin F, Rowshanrad A, Orozco LD, Gong KW, et al. NF-E2-related factor 2 promotes atherosclerosis by effects on plasma lipoproteins and cholesterol transport that overshadow antioxidant protection. Arterioscler Thromb Vasc Biol 2011 Jan;31(1):58-66.
11. Freigang S, Ampenberger F, Spohn G, Heer S, Shamshiev AT, Kisielow J, et al. Nrf2 is essential for cholesterol crystal-induced inflammasome activation and exacerbation of atherosclerosis. Eur J Immunol 2011 Jul;41(7):2040-51.
12. Zakkar M, Van der Heiden K, Luong lA, Chaudhury H, Cuhlmann S, Hamdulay SS, et al. Activation of Nrf2 in endothelial cells protects arteries from exhibiting a proinflammatory state. Arterioscler Thromb Vasc Biol 2009 Nov;29(11):1851-7.

13. Sussan TE, Jun J, Thimmulappa R, Bedja D, Antero M, Gabrielson KL, et al. Disruption of Nrf2, a key inducer of antioxidant defenses, attenuates ApoE-mediated atherosclerosis in mice. PLoS One 2008;3(11):e3791.

14. Malhotra D, Thimmulappa R, Navas-Acien A, Sandford A, Elliott M, Singh A, et al. Decline in NRF2-regulated antioxidants in chronic obstructive pulmonary disease lungs due to loss of its positive regulator, DJ-1. Am J Respir Crit Care Med 2008 Sep 15;178(6):592-604.

15. Sandford AJ, Malhotra D, Boezen HM, Siedlinski M, Postma DS, Wong V, et al. NFE2L2 pathway polymorphisms and lung function decline in chronic obstructive pulmonary disease. Physiol Genomics 2012 Aug 1;44(15):754-63.

16. Huang J, Tabbi-Anneni I, Gunda V, Wang L. Transcription factor Nrf2 regulates SHP and lipogenic gene expression in hepatic lipid metabolism. Am J Physiol Gastrointest Liver Physiol 2010 Dec;299(6):G1211-G1221.

17. van Diemen CC, Postma DS, Vonk JM, Bruinenberg M, Schouten JP, Boezen HM. A disintegrin and metalloprotease 33 polymorphisms and lung function decline in the general population. Am J Respir Crit Care Med 2005 Aug 1;172(3):329-33.

18. Siedlinski M, Postma DS, Boer JM, van der Steege G, Schouten JP, Smit HA, et al. Level and course of FEV1 in relation to polymorphisms in NFE2L2 and KEAP1 in the general population. Respir Res 2009;10:73.

19. Carnevali S, Petruzzelli S, Longoni B, Vanacore R, Barale R, Cipollini M, et al. Cigarette smoke extract induces oxidative stress and apoptosis in human lung fibroblasts. Am J Physiol Lung Cell Mol Physiol 2003 Jun;284(6):L955-L963.

20. Suzuki M, Betsuyaku T, Ito Y, Nagai K, Nasuhara Y, Kaga K, et al. Down-regulated NF-E2-related factor 2 in pulmonary macrophages of aged smokers and patients with chronic obstructive pulmonary disease. Am J Respir Cell Mol Biol 2008 Dec;39(6):673-82.

21. http://www.b58cgene.sgul.ac.uk/, accessed (19-11-2013).

22. Valcarcel-Ares MN, Gautam T, Warrington JP, Bailey-Downs L, Sosnowska D, de CR, et al. Disruption of Nrf2 signaling impairs angiogenic capacity of endothelial cells: implications for microvascular aging. J Gerontol A Biol Sci Med Sci 2012 Aug;67(8):821-9.

23. Ungvari Z, Bailey-Downs L, Gautam T, Jimenez R, Losonczy G, Zhang C, et al. Adaptive induction of NF-E2-related factor-2-driven antioxidant genes in endothelial cells in response to hyperglycemia. Am J Physiol Heart Circ Physiol 2011 Apr;300(4):H1133-H1140.

24. Tanaka Y, Aleksunes LM, Yeager RL, Gyamfi MA, Esterly N, Guo GL, et al. NF-E2-related factor 2 inhibits lipid accumulation and oxidative stress in mice fed a high-fat diet. J Pharmacol Exp Ther 2008 May;325(2):655-64.

25. Marzec JM, Christie JD, Reddy SP, Jedlicka AE, Vuong H, Lanken PN, et al. Functional polymorphisms in the transcription factor NRF2 in humans increase the risk of acute lung injury. FASEB J 2007 Jul;21(9):2237-46.

26. Canova C, Dunster C, Kelly FJ, Minelli C, Shah PL, Caneja C, et al. PM10-induced hospital admissions for asthma and chronic obstructive pulmonary disease: the modifying effect of individual characteristics. Epidemiology 2012 Jul;23(4):607-15.

27. Suzuki T, Shibata T, Takaya K, Shiraishi K, Kohno T, Kunitoh H, et al. Regulatory nexus of synthesis and degradation deciphers cellular Nrf2 expression levels. Mol Cell Biol 2013 Jun;33(12):2402-12.

Chapter 5

Genome-wide association study on all-cause and cardiovascular mortality

Sylwia M Figarska, Judith M Vonk and H Marike Boezen

University of Groningen, University Medical Center Groningen,
Department of Epidemiology, Groningen, The Netherlands

ABSTRACT

Although approximately 30% of the variation in human lifespan has a genetic background, the specific contributing genetic factors still remain to be discovered. We therefore performed a genome-wide association (GWA) study on all-cause mortality and on the leading cause of death worldwide, i.e. cardiovascular mortality. We used the Vlagtwedde-Vlaardingen general population-based cohort consisting of Caucasian individuals. In the survey in 1989/90 blood samples were drawn. The vital status of all subjects was assessed on December 31, 2008. During 18 years of follow-up, out of 1546 included subjects, 356 had died, 154 due to cardiovascular disease. We performed Cox regression on time to event (i.e. death due to any cause or due to cardiovascular disease, respectively), adjusted for age and sex. Genome-wide significance was $p=2.06x10^{-7}$. SNP rs10237193 in *VWC2* was genome-wide significantly associated with all-cause mortality. We identified 20 SNPs at $p<5x10^{-5}$. Of these 10 were annotated to genes: *RASAL2* (3 SNPs), *SAMD4A, MSI2, MBTPS1, CSF1R, NAP5, GAP43, TNKS1BP1*. We found no genome-wide associations for cardiovascular mortality, but 10 SNPs were associated at $p<5x10^{-5}$ and 4 of them were intragenic (*AMPH, LRMP, RASAL2, NELL1*). Some of the SNPs identified in our study have been associated with mortality in previous GWA studies at $p\leq0.05$, which emphasizes the relevance of our results. However further replication is needed to confirm our findings.

INTRODUCTION

The heritability of human lifespan is estimated to be approximately 30% (1), however, the specific contributing genetic factors still remain to be discovered. Several candidate genes involved in plausible biological pathways have been investigated, mostly in relation to a very long lifespan, but only two genes, namely apolipoprotein E (*APOE*) and forkhead box O3A (*FOXO3*), were consistently associated with longevity across different human populations (2-8). A genome-wide association (GWA) study, which is a hypothesis free approach of scanning the whole genome, may identify new susceptibility loci, including also noncoding variants, which would never be selected as a candidate otherwise (9).

A few GWA studies interested in variants contributing to extremely long lifespan confirmed the *APOE* association and emphasized the importance of the *APOE* ε4 allele, but were not successful in identifying any additional contributing genes (10-12).

Another way to investigate the genetics of human lifespan is studying time till death, which may identify loci associated with premature death. In a meta-analysis of nine GWA studies on all-cause mortality, genes highly expressed in the brain (*HECW2, HIP1, BIN2*), genes involved in neural development and function (*KCNQ4, LMO4, NETO1*) and autophagy (*ATG4C*) were identified (13).

Human lifespan is affected by the development of age-related diseases and among these cardiovascular disease (CVD) is the leading cause of morbidity and mortality worldwide (14). Genetic factors play an important role in cardiovascular mortality, since heritability of death from coronary heart disease is estimated to be as high as 60% in males and 40% in females (15). Initially, candidate genes known to influence traditional cardiovascular risk factors, such as increased

blood cholesterol, triglycerides or hypertension, were investigated. Genes associated with cardiovascular mortality include *APOE*, angiotensin I-converting enzyme (*ACE*), methylenetetrahydrofolate reductase (*MTHFR*), C-reactive protein (*CRP*), and endothelial nitric oxide synthase (*eNOS, NOS3*) (16-20). So far, only one GWA study has been performed on cardiovascular-related mortality (i.e. sudden cardiac death) and this study identified the *BAZ2B* locus (21). CVD encompasses a broad range of diseases of the heart and blood vessels like stroke or atherosclerosis, thus a GWA study on death due to any type of CVD may identify SNPs that are associated with CVD through multiple mechanisms. Such an identification of SNPs with a broad role in CVD may provide new targets for treatment in personalized medicine in the future. The aim of the current study was to scan the whole genome to identify SNPs associated with all-cause mortality and cardiovascular mortality in a general population.

METHODS

Study population

We studied participants of the Vlagtwedde-Vlaardingen cohort (22). This is a general population cohort of exclusively Caucasian individuals of Dutch descent. This study started in 1965 and participants had medical exams every 3 years until the last survey in 1989/1990. Information on area of residence, age, sex, smoking habits and respiratory symptoms was collected by the Dutch version of the UK Medical Research Council standard questionnaire (22). The vital status of all participants in the Vlagtwedde-Vlaardingen study was assessed on December 31, 2008. Mortality records were provided by Statistics Netherlands (CBS), where causes of death were coded using the

International Classification of Diseases version 9 and 10 (ICD-9 and ICD-10). Codes applicable to this study are presented in Table 1.

Table 1 ICD-codes for the investigated causes of death

Cause of death	ICD-9	ICD-10
External causes*	≥ 800	S, T, V, W, X, Y
Cardiovascular disease	390-398, 401-405, 410-417, 420-438, 440-448, 451-459, 785.4	G45-G46, I00-I15, I20-I28, I30-I52, I60-I69, I70-I79, I80-I89, I95-I97, I98.2, I98.8, I99, M30-M31, N28.0, R02, R58

* Suicides, homicides, traffic accidents *etc.*

Genotyping

In 1989–1990 neutrophil depots from peripheral blood samples were collected and stored at –20°C. In 2003–2004 DNA was extracted from these samples with a QIAamp DNA blood mini kit (Qiagen, Hilden, Germany) and checked for purity and concentration with a NanoDrop ND-1000 UV–Vis spectrophotometer (NanoDrop Technologies, Wilmington, DE). Genotyping was performed with Illumina CytoSNP-12 arrays, comprising 301,232 SNPs. The SNPs were called using the Illumina algorithm (Genome Studio, version 2.10.1).

Quality Control

Quality Control (QC) was performed using PLINK (23). SNPs with call rates less than 95%, a minor allele frequency (MAF) of less than 1% or deviation from Hardy–Weinberg equilibrium ($p<0.001$) and samples with call rates below 95% were excluded.

Statistical analysis

Cox proportional hazards models on time to death, adjusted for age at visit in 1989/90 and gender, were used to calculate the Hazard Ratio (HR) for each SNP using an additive genetic model. Time was defined as the time from the examination in 1989/1990 until death (from all-causes or CVD, respectively), until end of follow-up in 2008 for subjects who did not die or until last registration if subjects were lost to follow-up. External causes of death, such as suicide or homicide, were excluded in the analysis on all-cause mortality and subjects alive or lost to follow-up were censored. In the second analysis CVD was considered as a primary or secondary cause of death and subjects who did not experience the event of interest, thus these who died due to other cause than CVD, subjects who were alive and subjects lost to follow-up, were censored. Quantile-quantile (Q-Q) plots for the observed versus expected p values for all SNPs were generated from Cox models and the genomic inflation factor (λ) was calculated. All tests were performed using GenABEL, an R package that enables GWA analysis of time to event traits (24). The threshold p value for a genome-wide significance of 2.06×10^{-7} was calculated by dividing $\alpha=0.05$ by number of test performed (n=242,925).

Gene information and eQTL search

We used the SCAN database to assign top associated SNPs to genes and to check whether they are expression quantitative trait loci (eQTLs) in Caucasians ($p<0.0001$). The SCAN database provides expression data in lymphoblastoid cell lines (25). We searched for more information about the genes using the GeneCards database (www.genecards.org). We also checked on GWAS Central (http://www.gwascentral.org)

Table 2 Characteristics of the Vlagtwedde-Vlaardingen cohort at visit in 1989/90

	Alive	All-cause mortality[1]	p value[2]	CVD mortality	Causes of death other than CVD	p value[3]
N	1174	340		154	201	
Males, n (%)	575 (49)	221 (65)	< 0.001	99 (64)	135 (67)	0.571
Age (median, range)	50 (35-78)	63 (37-80)	< 0.001	64 (37-80)	62 (37-77)	0.030
Age at last status on Dec 31, 2008[4]	69 (54-97)	74 (41-94)	< 0.001	75 (41-94)	73 (48-92)	0.043
Ever smokers, n (%)	788 (67)	254 (75)	0.006	114 (74)	154 (77)	0.574
Vlagtwedde, n (%)	871 (74)	219 (64)	0.001	92 (60)	140 (70)	0.052

1. Excluding external causes of death (n=16)

2. Difference between alive and deceased subjects

3. Difference between subjects who died due to CVD and those who died due to other causes

4. For those who died this is age at death

whether the top SNPs associated with all-cause and cardiovascular mortality in our study were associated with other traits or diseases in other GWA studies at a p value<0.05.

RESULTS

After QC we excluded 58,307 SNPs, leaving 242,925 SNPs to be tested. We included 1546 subjects, 820 (53%) of them were males, 1071 (69%) were ever smokers and the median age of all the included subjects in 1989/90 was 53 years (range 35-80). Characteristics of the subjects stratified by vital status are presented in Table 2. In total, 356 subjects had died during the 18 years of follow-up, 154 of them due to CVD. Deaths due to external causes (n=16) were excluded from the analysis of all-cause mortality.

All-cause mortality

The Q-Q plot (Figure 1) and the genomic inflation factor (λ=1.02) indicated appropriate quality control and no population stratification. We found one SNP, rs10237193 in Von Willebrand factor C domain containing 2 (*VWC2*) located on chromosome 7, genome-wide significantly associated with all-cause mortality ($p=7.05x10^{-8}$). Furthermore, we identified 20 SNPs associated with all-cause mortality with a $p<5x10^{-5}$ (Figure 2). More details about these SNPs are given in Table 3.

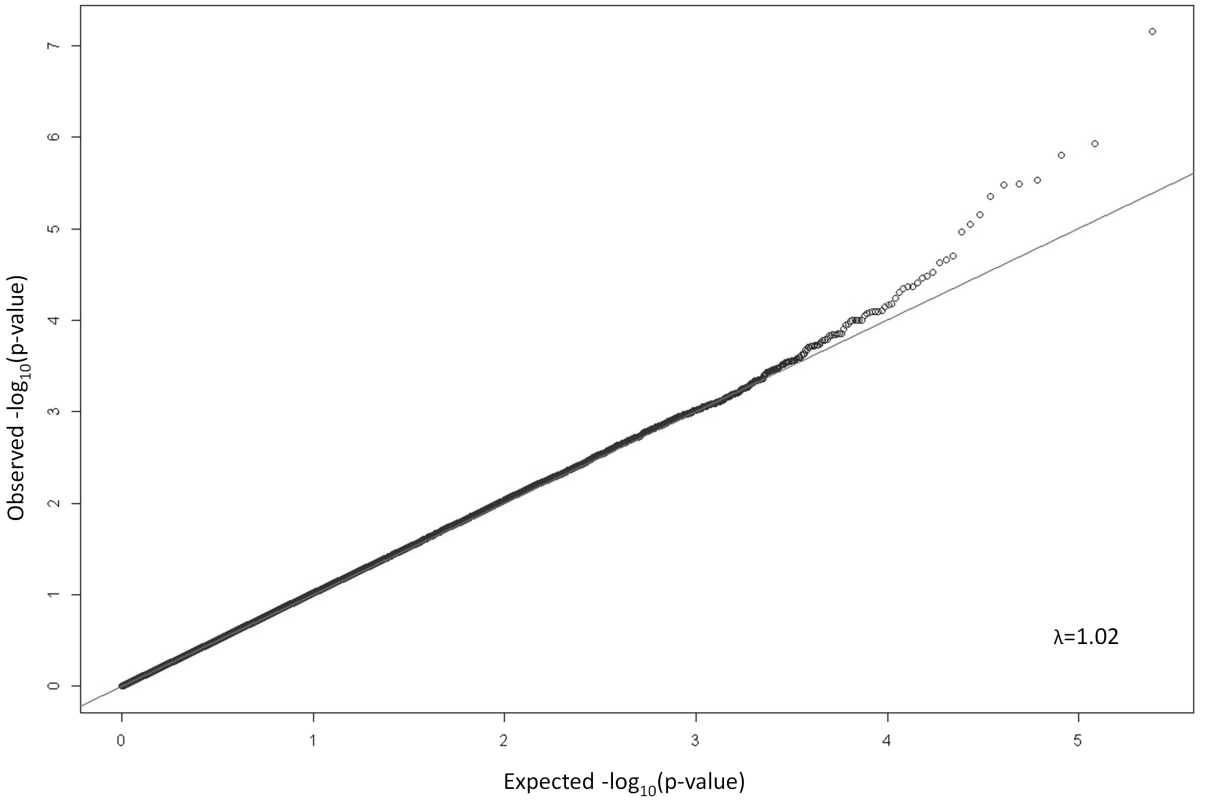


Figure 1 Quantile-Quantile (Q-Q) plot of the expected against observed -log p values from analyses of all-cause mortality.

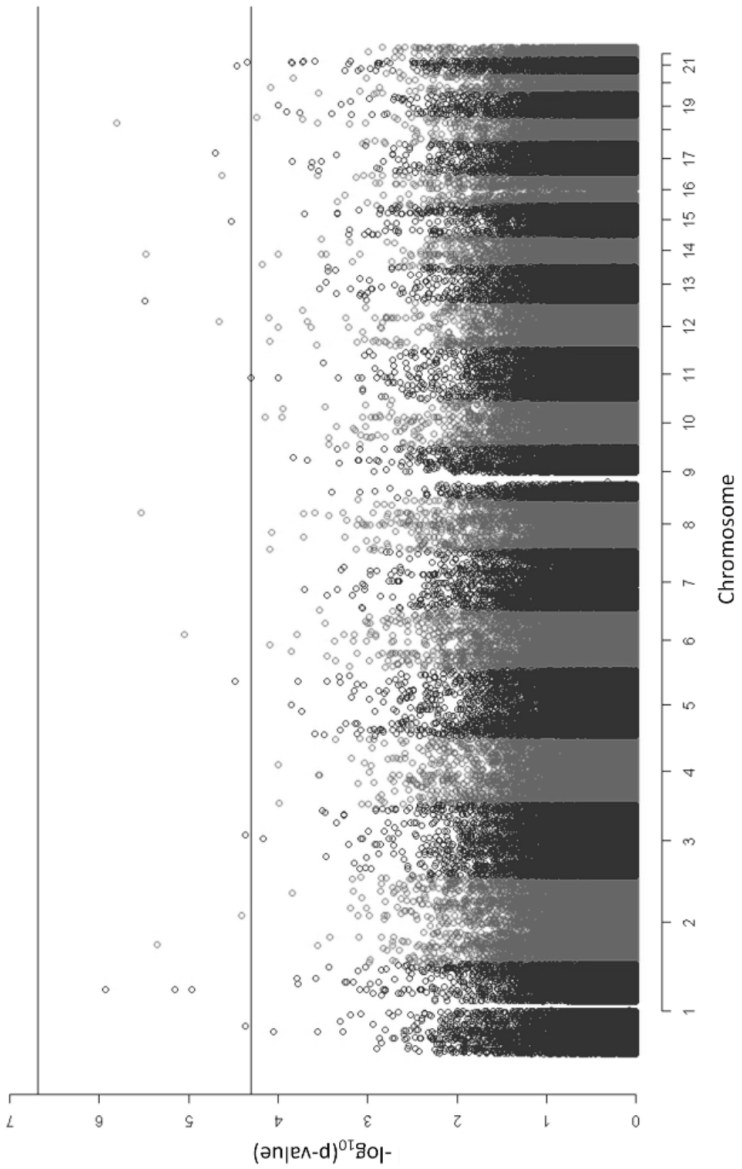


Figure 2 Manhattan plot presenting -log p value for the association with all-cause mortality in the additive genetic model for each SNP according to location in the 22 autosomal chromosomes. Horizontal lines indicate the 2.06×10^{-7} threshold of genome-wide significance and the threshold of $p<5X10^{-5}$.

Table 3 SNPs associated with all-cause mortality in the Vlagtwedde-Vlaardingen cohort with a p value <5x10^{-5}

SNP	Chr	Position	Gene[1]	eQTL[2]	A1[3]	A2[4]	HR	p value
rs10237193	7	49836460	VWC2	ABAT, MUTYH, A1BG	G	A	1.87	7.05E-08
rs2048333	1	178368180	RASAL2	NA	A	C	3.09	1.18E-06
rs489310	18	58183931	MC4R/ MRPS5P4	NA	A	G	1.47	1.56E-06
rs284489	8	105958020	LOC100128418/ LOC100129377	NA	G	A	1.52	2.91E-06
rs2152686	13	19913270	LOC100101938/ LOC100129883	C15orf31	G	A	2.46	3.27E-06
rs7159029	14	55073288	SAMD4A	NA	A	G	1.58	3.36E-06
rs2287511	2	53772561	LOC129656/ ASB3	ILK, ILK-2, TIMM9	A	G	1.96	4.39E-06
rs35215865	1	178313756	RASAL2	NA	A	G	2.84	6.98E-06
rs9320768	6	98342812	C6orf167/ LOC100129158	FGF11	A	G	1.42	8.86E-06
rs4342816	1	176640740	RASAL2	NA	A	G	2.85	1.08E-05
rs17761485	17	52849159	MSI2	NA	A	C	1.44	1.97E-05
rs771620	12	76090241	E2F7/ NAV3	NA	G	A	1.38	2.19E-05
rs2280024	16	82650563	MBTPS1	NA	G	A	0.69	2.35E-05
rs421705	15	56429879	LOC441726/ LIPC	NA	G	A	1.60	2.97E-05
rs2569075	5	149439022	CSF1R	NA	G	A	1.60	3.33E-05
rs4817193	21	28179020	RPL10P1/ C21orf94	NA	G	A	1.41	3.45E-05
rs1437903	2	133488128	NAP5	C1orf19, USP13	A	G	0.70	3.93E-05
rs12126169	1	77063610	LOC100133118/ ST6GALNAC5	NA	A	G	0.72	4.26E-05
rs2028248	3	116849656	GAP43	NHLRC1, IRF8, RNF32	A	G	1.42	4.33E-05
rs2836651	21	39025821	ERG/ NCRNA00114	NA	G	A	1.77	4.55E-05
rs10896602	11	56834040	TNKS1BP1	NA	A	G	1.83	4.95E-05

1. For extragenic SNPs the flanking left/right genes are given; 2. Expression in lymphoblastoid cell lines p< 0.0001 in CEU based on SCAN database; 3. Allele 1, for which HR is given; 4. Allele 2, reference allele

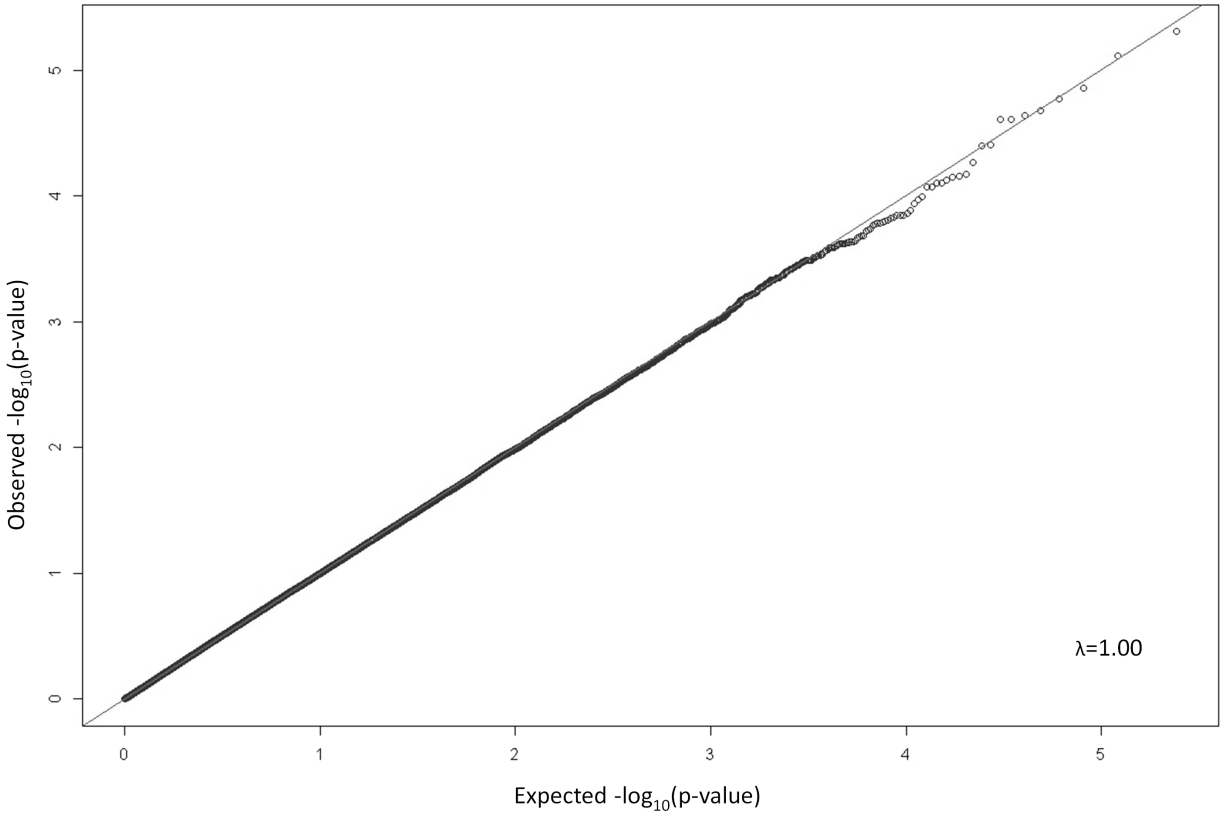


Figure 3 Quantile-Quantile (Q-Q) plot of the expected against observed -log p values from analyses of cardiovascular mortality.

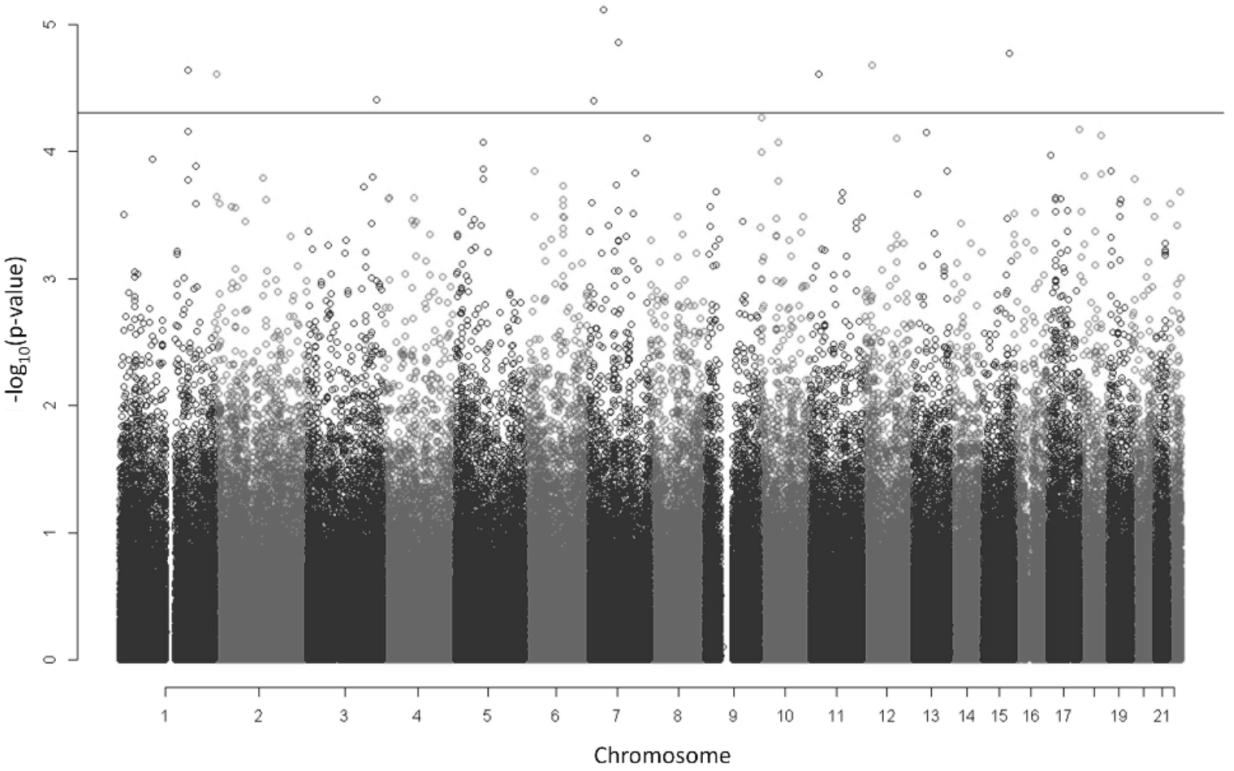


Figure 4 Manhattan plot presenting -log p value for the association with cardiovascular mortality in the additive genetic model for each SNP according to location in the 22 autosomal chromosomes. Horizontal line indicates the threshold of $p<5x10^{-5}$.

Cardiovascular mortality

The Q-Q plot (Figure3) and λ for cardiovascular mortality equaled 1.00, indicated appropriate quality control and no population stratification. In the analysis of cardiovascular mortality none of the tested SNPs reached genome-wide significance ($p=2.06x10^{-7}$), however 10 SNPs were associated at $p<5x10^{-5}$ (Table 4, Figure 4).

Gene information and eQTL analysis

Table 5 and 6 provide information on genes related to the SNPs associated at $p<5x10^{-5}$ with all-cause and cardiovascular mortality, respectively.

Whether identified SNPs are eQTLs is indicated in Table 3 and 4, for SNPs associated with all-cause or cardiovascular mortality, respectively. Our genome-wide significant hit for all-cause mortality, rs10237193, is an intronic SNP in the *VWC2* gene. This gene encodes a protein that is possibly involved in neural function and development and may have a role in cell adhesion. The identified SNP is significantly associated with the expression of 4-aminobutyrate aminotransferase (*ABAT*), mutY homolog E. coli (*MUTYH*) and alpha-1-B glycoprotein (*A1BG*) in Caucasians.

Regarding cardiovascular mortality, SNP rs998994 is an eQTL for cystathionine-beta-synthase (CBS), an enzyme involved in homocysteine metabolism. SNP rs11026076 is associated with the expression of arrestin beta 2 (*ARRB2*), which activates signaling of the beta-2-adrenergic receptor (26).

Interestingly we also found that the top associated SNPs in our study were associated with other traits or diseases at $p\leq0.05$ in previously performed GWA studies (Table 7 and 8). These traits include fasting plasma glucose, systolic blood pressure, BMI, serum cholesterol, and lung function.

Table 4 SNPs associated with cardiovascular mortality in the Vlagtwedde-Vlaardingen cohort with a p value <5x10^{-5}

SNP	Chr	Position	Gene[1]	eQTL[2]	A1[3]	A2[4]	HR	p value
rs9969851	9	12054659	LOC646114/ LOC100049717	NA	A	G	1.86	4.96E-06
rs11760351	7	38652687	AMPH	NA	A	G	2.56	7.71E-06
rs998994	7	77062679	PION/ PTPN12	ZNF259, CBS, GEM	A	G	1.69	1.39E-05
rs1036806	15	88844814	NTRK3/ MRPL46	NA	G	A	3.69	1.68E-05
rs1497253	12	25246048	LRMP	HAAO	A	G	1.67	2.11E-05
rs2048333	1	178368180	RASAL2	NA	A	C	3.82	2.28E-05
rs17291890	2	5560403	LOC727982/ SOX11	NA	A	G	1.63	2.45E-05
rs11026076	11	21436232	NELL1	ZER1, IFT52, LOC440396, SRR, RHOG, ARRB2, TBCB	C	A	0.50	2.47E-05
rs9840545	3	181484270	SOX2OT/ LOC100132918	NA	A	G	0.60	3.96E-05
rs3735182	7	12511037	VWDE/ TAS2R2	YY1AP1, TSC22D2, CSAD, C1QTNF6, GSTM1, GSTM2, GSTM4, EXOSC10, CAPZB, LOC100130193, LOC644075	A	G	1.70	4.02E-05

1. For extragenic SNPs the flanking left/right genes are given
2. Expression in lymphoblastoid cell lines p< 0.0001 in CEU based on SCAN database
3. Allele 1, for which HR is given
4. Allele 2, reference allele

Table 5 Gene information for intragenic SNPs and left/right genes for extragenic SNPs identified in the analysis on all-cause mortality (based on the GeneCards database).

SNP ID	Gene	Left gene	Right gene	Full name and function	
rs10237193	VWC2	LOC100130122	LOC100128734	**Von Willebrand Factor C Domain Containing 2.** This gene encodes a secreted bone morphogenic protein antagonist. The protein is possibly involved in neural function and development and may have a role in cell adhesion.	
rs2048333	RASAL2	LOC100131700	NCRNA00083	**RAS Protein Activator-Like 2.** Ablation of the RASAL2 is an alternative mechanism by which Ras becomes activated in breast cancer.	
rs489310	NA	MC4R	MRPS5P4	**Melanocortin Receptor 4** is a membrane-bound receptor and member of the melanocortin receptor family. The encoded protein interacts with adrenocorticotropic and MSH hormones and is mediated by G proteins. Defects in this gene are a cause of autosomal dominant obesity.	**Mitochondrial Ribosomal Protein S5 Pseudogene 4,** NA
rs284489	NA	LOC100128418	LOC100129377	NA	NA
rs2152686	NA	LOC100101938	LOC100129883	NA	NA
rs7159029	SAMD4A	LOC645602	LOC729451	**Sterile Alpha Motif Domain Containing 4A.** Sterile alpha motifs (SAMs) in proteins such as SAMD4A are part of an RNA-binding domain that functions as a posttranscriptional regulator by binding to an RNA sequence motif known as the Smaug recognition element.	
rs2287511	NA	LOC129656	ASB3	**CREB Regulated Transcription Coactivator 1 Pseudogene,** NA	**Ankyrin Repeat And SOCS Box-Containing 3.** The SOCS box serves to couple suppressor of cytokine signalling (SOCS) proteins and their binding partners with the elongin B and C complex, possibly targeting them for degradation.
rs35215865	RASAL2	LOC100131700	NCRNA00083	**RAS Protein Activator-Like 2,** described above for rs2048333.	

rs9320768	NA	C6orf167 (MMS22L)	LOC100129158	**Chromosome 6 Open Reading Frame 167 (MMS22-Like, DNA Repair Protein).** Component of the MMS22L-TONSL complex, a complex that stimulates the recombination-dependent repair of stalled or collapsed replication forks. The MMS22L-TONSL complex is required to maintain genome integrity during DNA replication by promoting homologous recombination-mediated repair of replication fork-associated double-strand breaks.	NA
rs4342816	RASAL2	LOC100131700	NCRNA00083	**RAS Protein Activator-Like 2**, described above for rs2048333.	
rs17761485	MSI2	LOC100128719	MRPS23	**Musashi RNA-Binding Protein 2** encodes a protein containing two conserved tandem RNA recognition motifs. Musashi, a translational repressor of Numb, regulates cell fate via Notch signaling by maintaining a pool of self-renewing stem cells.	
rs771620	NA	E2F7	NAV3	**E2F Transcription Factor 7** factor that participates to various processes such as angiogenesis, polyploidization of specialized cells and DNA damage response. E2F7 as the only E2F transcription factor potently up-regulated during oncogene-induced senescence, a setting where it acts in response to p53 as a direct transcriptional target. Acts as a promoter of sprouting angiogenesis and as a negative regulator of keratinocyte differentiation.	**Neuron navigator 3:** belongs to the neuron navigator family and is expressed predominantly in the nervous system. NAV3 may contribute to the growth, differentiation, and apoptosis of cutaneous T-cell lymphoma.
rs2280024	MBTPS1	SLC38A8	LOC729887	**Membrane-bound transcription factor peptidase, site 1.** Has a central role in the regulation of lipid metabolism in cells. It is a sterol-regulated subtilisin-like serine protease that cleaves ER membrane-bound sterol regulatory element-binding proteins (SREBPs), a reaction that initiates the two-step proteolytic process by which transcriptionally active fragments of SREBPs are released from the membrane for translocation to the nucleus.	

rs421705	NA	LOC441726	LIPC	NA	**Hepatic Lipase** is expressed in liver. **LIPC** has the dual functions of triglyceride hydrolase and ligand/ bridging factor for receptor-mediated lipoprotein uptake.
rs2569075	CSF1R	KIAA0194	RPL7P	**Colony stimulating factor 1 receptor,** The protein encoded by this gene is the receptor for colony stimulating factor 1, a cytokine which controls the production, differentiation, and function of macrophages. This receptor mediates most if not all of the biological effects of this cytokine.	
rs4817193	NA	RPL10P1	C21orf94 (NCRNA00314)	**Ribosomal Protein L10 Pseudogene,** NA	**Chromosome 21 Open Reading Frame 94,** NA
rs1437903	NAP5	LYPD1	LOC100130315	**Nck-associated protein 5,** its expression was detected in fetal and adult brain, leukocytes, and fetal fibroblasts.	
rs12126169	NA	LOC100133118	ST6GALNAC5	NA	**ST6 (alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3)-N-acetylgalactosaminide alpha-2,6-sialyltransferase 5**, belongs to a family of sialyltransferases that modify proteins and ceramides on the cell surface to alter cell-cell or cell-extracellular matrix interactions.
rs2028248	GAP43	LOC645207	LSAMP	**Growth-Associated Protein 43,** ‚growth' or ‚plasticity' protein because it is expressed at high levels in neuronal growth cones during development and axonal regeneration. This protein is considered a crucial component of an effective regenerative response in the nervous system.	
rs2836651	NA	ERG (p55)	NCRNA00114	**V-Ets Erythroblastosis Virus E26 Oncogene Homolog (Avian)** is a member of the erythroblast transformation-specific (ETS) family of transcriptions factors. All members of this family are key regulators of embryonic development, cell proliferation, differentiation, angiogenesis, inflammation, and apoptosis. The protein encoded by this gene is mainly expressed in the nucleus.	**Long Intergenic Non-Protein Coding RNA 114,** NA
rs10896602	TNKS1BP1	APLNR	SSRP1	**Tankyrase 1 binding protein 1**, Tankyrase 1 poly(ADP-ribosyl)ates TRF1 and releases it from telomeres, allowing access of telomerase to telomeres.	

Table 6 Gene information for intragenic SNPs and left/right genes for extragenic SNPs identified in the analysis on cardiovascular mortality (based on the GeneCards database).

SNP ID	Gene	Left gene	Right gene	Full name and function	
rs9969851	NA	LOC646114	LOC100049717	**A Kinase (PRKA) Anchor Protein 8 Pseudogene,** NA	**JNK1/MAPK8-associated membrane protein pseudogene,** NA
rs11760351	AMPH	TRG@	KRT8P20	**Amphiphysin** is a protein associated with the cytoplasmic surface of synaptic vesicles. A subset of patients with stiff-man syndrome who were also affected by breast cancer are positive for autoantibodies against this protein.	
rs998994	NA	PION	PTPN12	**Pigeon Homolog (Drosophila)** is a regulator of gamma-secretase activity, which specifically activates the production of beta-amyloid protein (beta-amyloid protein 40 and beta-amyloid protein 42), without affecting the cleavage of other gamma-secretase targets such has Notch. The gamma-secretase complex is an endoprotease complex that catalyzes the intramembrane cleavage of integral membrane proteins such as Notch receptors and APP (beta-amyloid precursor protein).	**Protein Tyrosine Phosphatase, Non-Receptor Type 12,** member of the protein tyrosine phosphatase (PTP) family. PTPs are signaling molecules that regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation. Protein tyrosine phosphatase actively contributes to the cellular apoptotic response.
rs1036806	NA	NTRK3	MRPL46	**Neurotrophic tyrosine kinase, receptor, type 3,** a member of the neurotrophic tyrosine receptor kinase (**NTRK**) family. This kinase is a membrane-bound receptor that, upon neurotrophin binding, phosphorylates itself and members of the MAPK pathway. Signalling through this kinase leads to cell differentiation and may play a role in the development of proprioceptive neurons that sense body position. Mutations in this gene have been associated with medulloblastomas, secretory breast carcinomas and other cancers. Associated with early-onset major depressive disorder.[1] Decreased expression of NTRK3 is associated with the outflow tract defect of human tetralogy of Fallot.[2]	**Mitochondrial Ribosomal Protein L46.** Mammalian mitochondrial ribosomal proteins are encoded by nuclear genes and help in protein synthesis within the mitochondrion. Mitochondrial ribosomes (mitoribosomes) consist of a small 28S subunit and a large 39S subunit.

rs1497253	LRMP	LOC645177	CASC1	**Lymphoid-Restricted Membrane Protein** is expressed in a developmentally regulated manner in lymphoid cell lines and tissues. The protein is localized to the cytoplasmic face of the endoplasmic reticulum. One locus close to LRMP was linked to higher resting heart rate (proxy to mortality risk) but is not in LD with our SNP.	
rs2048333	RASAL2	LOC100131700	NCRNA00083	**RAS Protein Activator-Like**. Ablation of the RASAL2, is an alternative mechanism by which Ras becomes activated in breast cancer.	
rs17291890	NA	LOC727982	SOX11	NA	**SRY (Sex Determining Region Y)-Box 11,** This intronless gene encodes a member of the **SOX** (SRY-related HMG-box) family of transcription factors involved in the regulation of embryonic development and in the determination of the cell fate. The encoded protein may act as a transcriptional regulator after forming a protein complex with other proteins. The protein may function in the developing nervous system and play a role in tumorgenesis. *Sox11*-deficient mice died at birth from congenital cyanosis, likely resulting from heart defects. These included ventricular septation defects and outflow tract malformations that ranged from arterial common trunk to a condition known as double outlet right ventricle. Study in *Sox11*-deficient mice identified a prime function for Sox11 in tissue remodeling.[3]
rs11026076	NELL1	LOC100130160	ANO5	**NEL-Like 1** is a cytoplasmic protein that contains epidermal growth factor (EGF)-like repeats. The encoded heterotrimeric protein may be involved in cell growth regulation and differentiation. A similar protein in rodents is involved in craniosynostosis.	

rs9840545	NA	SOX2OT	LOC100132918	**SOX2 Overlapping Transcript** (Non-Protein Coding). Close correlation between the expression pattern of SOX2OT variants and master regulators of pluripotency (OCT4 and SOX2) suggests a role in similar regulatory pathways including: self-renewal, pluripotency, differentiation in tumor initiation and/or progression.[4]	**Eukaryotic Translation Initiation Factor 4E Pseudogene,** NA
rs3735182	NA	VWDE	TAS2R2	**Von Willebrand Factor D And EGF Domains,** NA	**Taste Receptor, Type 2, Member 2 Pseudogene,** NA

1. Verma et al. Linkage disequilibrium mapping of a chromosome 15q25–26 major depression linkage region and sequencing of NTRK3. Biol Psychiatry. 2008

2. Tetralogy of Fallot is a congenital heart defect , low oxygenation of blood due to the mixing of oxygenated and deoxygenated blood in the left ventricle.

3. Sock et al. Gene Targeting Reveals a Widespread Role for the High-Mobility-Group Transcription Factor Sox11 in Tissue Remodeling. Mol Cell Biol. 2004

4. Shahryari et al. Two Novel Splice Variants of SOX2OT, SOX2OT-S1, and SOX2OT-S2 are Coupregulated with SOX2 and OCT4 in Esophageal Squamous Cell Carcinoma. Stem Cells. 2014

Table 7 Top 21 SNPs associated with all-cause mortality in the Vlagtwedde-Vlaardingen cohort and their associations with other outcomes (based on the GWAS Central database)

SNP	Trait/disease in other GWA studies	p value
rs10237193	Asthma	0.0308
rs2048333	-	
rs489310	Rheumatoid arthritis	0.0022
rs284489	Proinsulin levels	0.0120
	Fasting plasma glucose	0.0524
	Ulcerative colitis	0.0414
	Birth weight	0.0077
	Bipolar disorder	0.0105
rs2152686	Serum cholesterol	0.0080
	Fasting insulin	0.0495
rs7159029	-	
rs2287511	Age-related macular degeneration	0.0496
rs35215865 proxy: rs11805590 (r^2=1.0)	Type II diabetes	0.0349
	Systolic blood presure	0.0379
rs9320768	Pulmonary function (FEV1/FVC)	0.0024
	Body mass index	0.0217
rs4342816	Alzheimer's disease	0.0209
	Brain glutamate concentrations in patients with MS	0.0493
rs17761485	Forced expiratory volume	0.0001
	Pulmonary function (FEV1)	0.0029
rs771620	Rheumatoid arthritis	0.0301
	Brain glutamate concentrations	0.0488
rs2280024	Fasting plasma glucose	0.0080
	sporadic Creutzfeldt-Jakob disease (sCJD)	0.0114
	Waist-hip ratio	0.0420
rs421705	Amyotrophic lateral sclerosis	0.0138
	Psoriasis	0.0279
rs2569075	Flucloxacillin-induced liver injury	0.0251
rs4817193	Glycated hemoglobin levels	0.0141
rs1437903	-	
rs12126169	Body mass index	0.0142
	Proinsulin levels	0.0525
rs2028248	Serum cholesterol	0.0098
	Anti-cyclic citrullinated peptide-positive rheumatoid arthritis	0.0175
	Amyotrophic lateral sclerosis	0.0103
	Kuru	0.0335

rs2836651	Height-adjusted highest forced expiratory volume	0.0270
	Verbal recall	2.69e-05
	Homeostatic model assessment of insulin resistance	0.0156
	Fasting insulin	0.0127
rs10896602 proxy: rs10792085 (r^2=1.0)	Schizophrenia	0.0277

Table 8 Top 10 SNPs associated with cardiovascular mortality in the Vlagtwedde-Vlaardingen cohort and their associations with other outcomes (based on the GWAS Central database)

SNP	Trait/disease in other GWA studies	p value
rs9969851	Two-hour glucose challenge	0.0566
	Fasting plasma glucose	0.0316
rs11760351	Ulcerative colitis	0.0495
rs998994	Type II diabetes	0.0014
	Bipolar disorder	0.0303
	Waist-hip ratio	0.0310
	Response to antipsychotic therapy, extrapyramidal side effects: Barnes Akathisia rating scale	0.0230
rs1036806 proxy: rs11632476 (r^2=0.85)	Body mass index	0.0153
	Log10 serum total immunoglobulin E concentration	0.0241
rs1497253	Ulcerative colitis	0.0111
proxy: rs7303669 (r^2=1.0)	Birth weight	0.0343
proxy: rs7977670 (r^2=1.0)	Serum cholesterol	0.0507
rs2048333	-	
rs17291890	sporadic Creutzfeldt-Jakob disease (sCJD)	0.0398
rs11026076	Partial epilepsies	0.0173
	Asthma	0.0136
	Pulmonary function (FEV1/FVC)	0.0593
rs9840545	Type II diabetes	0.0570
	Height-adjusted highest forced expiratory volume	0.0567
rs3735182	Log10 serum total immunoglobulin E concentration	0.0278
	Fasting plasma glucose	0.0479

DISCUSSION

In the analysis of 1546 subjects followed for 18 years we identified one locus associated with all-cause mortality at the genome-wide significance level. In the GWA study on cardiovascular mortality, we did not identify genome-wide significant SNPs.

Some of the SNPs we found to be associated with $p<5x10^{-5}$ are associated with gene expression, even when the specific SNPs are not located in a gene. Showing evidence that the SNP is an expression quantitative trait locus (eQTL) can improve the ability to discover a true association and clarify the nature of the mechanism driving the association (27). The genome-wide significant hit for all-cause mortality, rs10237193, is an intronic SNP in the *VWC2* gene. This gene encodes a protein that is possibly involved in neural function and development and may have a role in cell adhesion. Rs10237193 is significantly associated with expression of *ABAT*, a gene responsible for the catabolism of gamma-aminobutyric acid (GABA), an important, mostly inhibitory neurotransmitter in the central nervous system, with the expression of *MUTYH*, which encodes a DNA glycosylase involved in oxidative DNA damage repair, and with the expression of alpha-1-B glycoprotein (*A1BG*), which encodes a plasma glycoprotein of unknown function. In addition, we found three SNPs (rs2048333, rs35215865 and rs4342816) in three distinct introns of the RAS protein activator like 2 (*RASAL2*). Ras proteins are very important in cell growth, differentiation and survival. Another interesting finding is the SNP rs2287511. The closest right gene to this SNP is ankyrin repeat and SOCS box-containing 3 (*ASB3*). ASB proteins are involved in mediating protein ubiquitination and degradation, using the ankyrin (ANK) repeats to recruit substrates (28). Interestingly, p16INK4a, a biomarker of

human ageing also interacts with its substrates via ANK repeats (29;30). Our results on cardiovascular mortality also identified SNPs with a possible biological relevance. For example, rs998994 is an eQTL for cystathionine-beta-synthase (CBS), an enzyme involved in homocysteine metabolism. Increased CBS expression may lead to hyperhomocysteinaemia, which is a risk factor for CVD (31). Additionally, SNP rs11026076 is associated with the expression of arrestin beta 2 (*ARRB2*), which activates signaling of the beta-2-adrenergic receptor (26). This activation is anti-apoptotic and contributes to myocardial protection, thus the risk allele of this SNP may decrease expression of *ARRB2* (32).

Interestingly some of our top SNPs were associated with conditions increasing the risk of death, such as elevated fasting plasma glucose, systolic blood pressure, BMI, serum cholesterol, or decreased lung function ($p \leq 0.05$) in previous GWA studies. This additionally strengthens the relevance of our findings and suggests that these SNPs constitute predictors of all-cause and cardiovascular mortality via their impact on these intermediate phenotypes.

Both all-cause and cardiovascular mortality are complex and heterogeneous traits. Therefore, they may involve many loci with moderate effects that contribute to these events, but they are difficult to be identified in GWA studies. However, genome-wide non-significant results should not be discarded and further functional studies on these SNPs may help to better understand such findings. In the GWA study performed in long-living individuals from Southern Italy the top SNP located in *CAMKIV* did not reach genome-wide significance (p value=2.9×10^{-5}), but its relevance was further highlighted by showing its association with expression. The same study showed that *CAMKIV* activates the proteins involved

in cell survival: *AKT, SIRT1*, and *FOXO3A* (33). This is an example how further studies may pinpoint functionally relevant loci, even if they do not reach the genome-wide significance.

Strengths and limitations

The main strength of our study is the longitudinal design with 18 years of follow-up. Such a study is less prone to bias than a longevity case-control study, where two groups from different generations are compared and it is difficult to predict how many control subjects will become very old or centenarians. In GWA study hundreds of thousand SNPs are tested and to reach the pre-specified, very low, thresholds of genome-wide significance, a large sample size is required (34). In this context the relatively low number of 154 subjects who died due to CVD may be considered a limitation of the study, since indeed none of the SNPs reached the genome-wide significance threshold in our GWA study on cardiovascular mortality. Nonetheless, analyzing all-cause mortality with 356 subjects who died during the 18 years of follow-up, we found one genome-wide significant SNP. At the moment the lack of replication is a limitation of our study, however replication of all SNPs associated at p value $<5x10^{-5}$ in another cohort is planned.

In summary, with a hypothesis free approach we identified SNPs related to all-cause and cardiovascular mortality in the general Caucasian population. Replication of the results in another cohort/ population should confirm their importance. Furthermore, with these complex heterogeneous outcomes, studying gene-gene (epistasis) and gene-environment interactions may provide more clues on the pathways involved in human lifespan.

REFERENCES

1. 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 Mar;97(3):319-23.
2. 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 Apr;12(2):95-104.
3. 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 Feb 24;106(8):2700-5.
4. Frisoni GB, Louhija J, Geroldi C, Trabucchi M. Longevity and the epsilon2 allele of apolipoprotein E: the Finnish Centenarians Study. J Gerontol A Biol Sci Med Sci 2001 Feb;56(2):M75-M78.
5. Jian-Gang Z, Yong-Xing M, Chuan-Fu W, Pei-Fang L, Song-Bai Z, Nui-Fan G, et al. Apolipoprotein E and longevity among Han Chinese population. Mech Ageing Dev 1998 Aug 14;104(2):159-67.
6. Kervinen K, Savolainen MJ, Salokannel J, Hynninen A, Heikkinen J, Ehnholm C, et al. Apolipoprotein E and B polymorphisms--longevity factors assessed in nonagenarians. Atherosclerosis 1994 Jan;105(1):89-95.
7. 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 Dec 15;18(24):4897-904.
8. 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 Sep 16;105(37):13987-92.
9. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science 2012 Sep 7;337(6099):1190-5.
10. 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 Aug;10(4):686-98.
11. 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 Jun;132(6-7):324-30.
12. 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(1):e29848.

13. Walter S, Atzmon G, Demerath EW, Garcia ME, Kaplan RC, Kumari M, et al. A genome-wide association study of aging. Neurobiol Aging 2011 Nov;32(11):2109-28.

14. Guilbert JJ. The world health report 2. Educ Health (Abingdon) 2003 Jul;16(2):230.

15. Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, De FU. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. J Intern Med 2002 Sep;252(3):247-54.

16. Kalina A, Czeizel AE. The methylenetetrahydrofolate reductase gene polymorphism (C677T) is associated with increased cardiovascular mortality in Hungary. Int J Cardiol 2004 Nov;97(2):333-4.

17. Lange LA, Carlson CS, Hindorff LA, Lange EM, Walston J, Durda JP, et al. Association of polymorphisms in the CRP gene with circulating C-reactive protein levels and cardiovascular events. JAMA 2006 Dec 13;296(22):2703-11.

18. Rossi GP, Maiolino G, Zanchetta M, Sticchi D, Pedon L, Cesari M, et al. The T(-786)C endothelial nitric oxide synthase genotype predicts cardiovascular mortality in high-risk patients. J Am Coll Cardiol 2006 Sep 19;48(6):1166-74.

19. Sayed-Tabatabaei FA, Schut AF, Arias VA, Bertoli-Avella AM, Hofman A, Witteman JC, et al. Angiotensin converting enzyme gene polymorphism and cardiovascular morbidity and mortality: the Rotterdam Study. J Med Genet 2005 Jan;42(1):26-30.

20. Stengard JH, Weiss KM, Sing CF. An ecological study of association between coronary heart disease mortality rates in men and the relative frequencies of common allelic variations in the gene coding for apolipoprotein E. Hum Genet 1998 Aug;103(2):234-41.

21. Arking DE, Junttila MJ, Goyette P, Huertas-Vazquez A, Eijgelsheim M, Blom MT, et al. Identification of a sudden cardiac death susceptibility locus at 2q24.2 through genome-wide association in European ancestry individuals. PLoS Genet 2011 Jun;7(6):e1002158.

22. van Diemen CC, Postma DS, Vonk JM, Bruinenberg M, Schouten JP, Boezen HM. A disintegrin and metalloprotease 33 polymorphisms and lung function decline in the general population. Am J Respir Crit Care Med 2005 Aug 1;172(3):329-33.

23. 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 Sep;81(3):559-75.

24. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics 2007 May 15;23(10):1294-6.

25. Gamazon ER, Zhang W, Konkashbaev A, Duan S, Kistner EO, Nicolae DL, et al. SCAN: SNP and copy number annotation. Bioinformatics 2010 Jan 15;26(2):259-62.

26. Patel PA, Tilley DG, Rockman HA. Physiologic and cardiac roles of beta-arrestins. J Mol Cell Cardiol 2009 Mar;46(3):300-8.

27. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet 2010 Apr;6(4):e1000888.

28. Chung AS, Guan YJ, Yuan ZL, Albina JE, Chin YE. Ankyrin repeat and SOCS box 3 (ASB3) mediates ubiquitination and degradation of tumor necrosis factor receptor II. Mol Cell Biol 2005 Jun;25(11):4716-26.

29. Byeon IJ, Li J, Ericson K, Selby TL, Tevelev A, Kim HJ, et al. Tumor suppressor p16INK4A: determination of solution structure and analyses of its interaction with cyclin-dependent kinase 4. Mol Cell 1998 Feb;1(3):421-31.

30. Liu Y, Sanoff HK, Cho H, Burd CE, Torrice C, Ibrahim JG, et al. Expression of p16(INK4a) in peripheral blood T-cells is a biomarker of human aging. Aging Cell 2009 Aug;8(4):439-48.

31. Wald NJ, Watt HC, Law MR, Weir DG, McPartlin J, Scott JM. Homocysteine and ischemic heart disease: results of a prospective study with implications regarding prevention. Arch Intern Med 1998 Apr 27;158(8):862-7.

32. Patterson AJ, Zhu W, Chow A, Agrawal R, Kosek J, Xiao RP, et al. Protecting the myocardium: a role for the beta2 adrenergic receptor in the heart. Crit Care Med 2004 Apr;32(4):1041-8.

33. 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 Jun;14(3):283-91.

34. Ioannidis JP, Thomas G, Daly MJ. Validating, augmenting and refining genome-wide association signals. Nat Rev Genet 2009 May;10(5):318-29.

Chapter 6

Genetic Risk Score for prediction of reduced lung function level in the general population: The Vlagtwedde-Vlaardingen study

Sylwia M Figarska, Judith M Vonk and H Marike Boezen

Department of Epidemiology, University of Groningen, University Medical Center Groningen, the Netherlands

ABSTRACT

Background: Reduced Forced Expiratory Volume in one second (FEV_1) is a predictor of premature death. Combining multiple loci with modest effects into a genetic risk score (GRS) might improve identification of subjects at risk for reduced FEV_1.

Methods: By testing associations between reduced lung function (i.e. FEV_1<80% predicted) and 75 single nucleotide polymorphisms (SNPs) in 24 candidate genes in 1,390 subjects from the general population-based Vlagtwedde-Vlaardingen cohort, 11 SNPs were selected to construct unweighted and weighted genetic risk (uGRS and wGRS respectively). We estimated the association, goodness of fit (Akaike Information Criterion: AIC) and increase in discriminative ability of uGRS, wGRS and 11 SNPs jointly for FEV_1<80% predicted, compared to the discriminative ability of packyears, a well-established risk factor of reduced lung function.

Results: The odds ratios (95% CI) per unit were 1.8 (1.6 to 2.2) and 1.5 (1.3 to 1.7) for uGRS and wGRS respectively ($p<0.001$). 11 SNPs tested jointly were associated with reduced lung function at the p value ≤0.10 for each, giving ORs between 1.5 (rs1048290) to 4.7 (rs8187858). The uGRS, wGRS or 11 SNPs included in the model next to packyears increased the AUC to 0.735, 0.740 and 0.749 respectively (packyears AUC=0.646). The model with wGRS best fits the data.

Conclusions: This study shows for the first time that a GRS composed of 11 SNPs associated with FEV_1<80% predicted is a better independent predictor of reduced lung function than the 11 SNPs separately.

INTRODUCTION

Forced Expiratory Volume in one second (FEV_1) is a very useful index for clinical monitoring and assessment of health. Reduced FEV_1 is not only a measure of airflow limitation, but also a predictor of Chronic Obstructive Pulmonary Disease (COPD), lung cancer, coronary artery disease and stroke, collectively accounting for 70-80% of premature death in smokers (1). The level of lung function is both genetically and environmentally determined. The most important factors explaining the interindividual variation in lung function are sex, height and age. Therefore, in clinical practice an absolute value of FEV_1 is commonly converted into a percentage of the predicted value (FEV_1%predicted) using a formula that takes into account the sex, age and height of the person (2). Values above 80% of the predicted value are considered normal, whereas FEV_1 less than 80% of predicted indicates reduced level of lung function (3).

Twin studies in European and US populations estimated that FEV_1 heritability might be as high as 0.77 (4;5). Importantly, longitudinal studies in families have shown that heritability was consistent through time, whereas common familial environmental effects on FEV_1 explained only 1-4% of the variability in children and 11-28% in adults (6). One of the approaches to determine genetic components contributing to a reduced level of lung function is studying single nucleotide polymorphisms (SNPs) in candidate genes, chosen for their plausible biological role in the pathological processes in the lung tissue. However, SNPs in candidate genes when studied separately only confer modestly increased risk and are of limited value in disease prediction (7). Previous studies have shown that combining multiple loci with modest effects into a global genetic risk score (GRS) might improve identification of subjects at risk for a trait (7-14).

In the current study, first we tested 74 SNPs in 24 genes, that previously have been shown to play an important role in the lung tissue composition, protease-antiprotease balance, antioxidant response in the lungs, xenobiotics transport or chronic inflammation, to detect those with modest effects on FEV_1%predicted, and then use these in composing the GRS. The main purpose of this study was to examine the discriminatory and predictive ability of a lung function GRS and comparing this to a model containing packyears only. Furthermore, we also compared the GRS to a model containing all selected SNPs separately, thus not combined in the GRS.

METHODS

Participants & measurements

We studied 1,390 subjects of the Vlagtwedde-Vlaardingen cohort participating in the last survey in 1989/1990. This general population-based cohort of exclusively Caucasian individuals of Dutch descent started in 1965 and has been followed for 25 years (15). The Committee on Human Subjects in Research of the University of Groningen reviewed the study and affirmed the safety of the protocol and study design and all participants gave their written informed consent. In surveys, that took place every 3 years, information on respiratory symptoms, smoking status, age, and sex by the Dutch version of the UK Medical Research Council standard questionnaire was collected and spirometry was performed.

Spirometry

Pulmonary function measurements were performed with a water-sealed spirometer (Lode Spirograph, Lode Instruments, Groningen, The Netherlands). Measurement of inspiratory vital capacity (IVC) after a deep

expiration was followed by measurement of forced expiratory volume in one second (FEV_1). The highest of the values obtained in two technically satisfactory tracings was taken as long as the difference between the two IVC values was less than 150 ml and the difference between two FEV_1 values was less than 100 ml. We used the Quanjer prediction equations to calculate FEV_1%predicted value for each participant (2).

Table 1 74 SNPs in 24 a priori candidate genes tested for association with FEV_1 less than 80% of predicted

Gene symbol	Gene name	SNP ID	SNP name
ADAM33	A Desintegrin and Metalloproteinase	rs2787094	V_4
		rs528557	S_2
		rs2280090	T_2
		rs612709	Q_1
		rs511898	F_1
		rs3918396	S_1
		rs597980	ST_5
MMP1	Matrix Metalloproteinase 1	rs1799750	G-1607GG
MMP2	Matrix Metalloproteinase 2	rs243865	C-1306T
MMP9	Matrix Metalloproteinase 9	rs6065912	
		rs3918278	
		rs8113877	
MMP12	Matrix Metalloproteinase 12	rs2276109	A-82G
		rs652438	Asn357Ser
TIMP1	Tissue Inhibitor Of Metalloproteinases 1	rs4898	Phe124Phe
		rs11551797	Ile158Ile
TGF-β1	Transforming Growth Factor, Beta 1	rs6957	3'UTR
		rs1800469	C-509T
		rs1800470	Leu10Pro
DCN	Decorin	rs1803343	
		rs11106030	
		rs741212	
		rs516115	
GSTM1	Glutathione S-Transferase Mu 1	deletion	
GSTT1	Glutathione S-Transferase Theta 1	deletion	
GSTP1	Glutathione S-Transferase Pi 1	rs1695	Ile105Val
		rs1138272	Ala114Val
SOD2	Superoxide Dismutase 2	rs4880	Ala16Val
		rs2842958	C7693T

SOD3	Superoxide Dismutase 3	rs2536512	Ala40Thr
		rs1799895	Arg213Gly
		rs8192288	G(-4466)T
SFTPA1	Surfactant Protein A1	rs1059047	Ala19Val
		rs1136450	Leu50Val
		rs4253527	Arg219Trp
SFTPA2	Surfactant Protein A2	rs1059046	Asn9Thr
		rs17886395	Pro91Ala
		rs1965707	Ser140Ser
SFTPB	Surfactant Protein B	rs1130866	Ile131Thr
SFTPD	Surfactant Protein D	rs721917	Met11Thr
		rs2243639	Thr160Ala
GCLC	Glutamate-Cysteine Ligase, Catalytic Subunit	rs17883901	C(-129)T
GCLM	Glutamate-Cysteine Ligase, Modifier Subunit	rs41303970	C(-588)T
SIRT1	Sirtuin 1	rs12778366	
		rs7069102	
		rs2273773	
ABCC1 (*MRP1*)	ATP-Binding Cassette, Sub-Family C (CFTR/ MRP), Member 1	rs4148382	
		rs212093	
		rs504348	
		rs35621	
		rs4781699	
		rs215100	
		rs35597	
		rs212083	
		rs35610	
		rs3743527	
		rs215101	
		rs8187858	
		rs3851710	
		rs2074087	
		rs2283515	
		rs215095	
		rs12448760	
		rs4148330	
		rs3819552	
NFE2L2 (*NRF2*)	Nuclear Factor (Erythroid-derived 2)-Like 2	rs6726395	
		rs4243387	
		rs1806649	
		rs13001694	
		rs2364723	
KEAP1	Kelch-Like ECH-Associated Protein 1	rs1048290	
		rs11085735	
		rs1048287	
HO-1	Heme Oxygenase 1	rs2071747	Asp7His

DNA collection and genotyping

In 1989–1990 neutrophil depots from peripheral blood samples were collected and stored at –20°C. In 2003–2004 DNA was extracted from these samples with a QIAamp DNA blood mini kit (Qiagen, Hilden,Germany) and checked for purity and concentration with a NanoDropND-1000 UV–Vis spectrophotometer (NanoDrop Technologies, Wilmington, DE). We genotyped DNA samples of those subjects with more than 1,500ng of isolated DNA available (n=1,390) (15). 74 SNPs in 24 candidate genes were genotyped (Table 1). The 24 candidate genes were selected based on their possible importance in lung tissue composition (*SFTPA1, SFTPA2, SFTPB, SFPTD*), protease-antiprotease balance (*MMP1, MMP2, MMP9, MMP12, TIMP1, DCN, TGF-β*$_1$), antioxidant response in the lungs (*HMOX1, SOD2, SOD3, NFE2L2, KEAP1, GCLC, GCLM, GSTM1, GSTP1, GSTT1*), xenobiotics transport (*ABCC1*) or chronic inflammation (*SIRT1, ADAM33*).

Statistical analysis

A χ^2 test was used to assess whether the SNPs were in Hardy-Weinberg Equilibrium (p value <0.05). Values below 80% of FEV_1 predicted indicated reduced lung function level thus this value was used as a threshold and descriptives of the subject characteristics are presented stratified by lung function level: FEV_1 less than 80% predicted and FEV_1 greater or equal to 80% predicted.
Logistic regression adjusted for packyears was used to estimate odds ratios and 95% confidence intervals (OR (95% CI)) for the risk of having an FEV_1 less than 80% predicted, in relation to the genotypes. Correction for additional covariates was not performed given thatFEV_1%predicted takes age, sex and height into account.

SNPs associated with p values less than or equal to 0.10 were further considered for inclusion in the GRS, with a coding value of 0 for the non risk genotype and 1 for the risk genotype. Using this set of SNPs we followed a modified forward multiple regression approach with a cut-off p value ≤0.10 to decrease the number of variants included in the GRS. This approach has higher power than the Bonferroni and false discovery rate (FDR) procedure for detecting moderate and weak genetic effects (16).

Two approaches were used to calculate the GRS: a simple risk genotypes summing method (unweighted GRS: uGRS) and a weighted method (weighted GRS: wGRS). The wGRS was obtained by multiplying each risk genotype by the OR for that genotype from the univariate model, summing the products and dividing by the number of incorporated SNPs. In order to reduce the bias caused by missing data, we did not include any subjects who had any missing genotype in generating the GRS (7). The t-test was used to compare the mean GRS between subjects with FEV_1 greater or equal to 80% of predicted and those with FEV_1 less than 80% of predicted. To evaluate the association between risk factors and reduced lung function we used four separate logistic regression models:

model 1) contained only a well-established predictor, i.e. packyears;

model 2) uGRS + packyears;

model 3) wGRS + packyears, and

model 4) the selected SNPs jointly but not combined in one variable + packyears.

We compared the relative goodness of fit of these models using the Akaike information criterion (AIC). The lowest AIC value indicates the best fitting model. As a sensitivity analysis we additionally evaluated

the performance of the uGRS, wGRS or the selected SNPs jointly in a linear regression on FEV_1%predicted as a continuous variable. We used c-statistics, i.e. the area under the receiver operating characteristic (ROC) curve (AUC) to quantify the discriminative accuracy of a prediction model (17). The AUC is the probability that the test correctly separates two groups and ranges from 0.5 (total lack of discrimination) to 1.0 (perfect discrimination). All statistical analyses were performed using SPSS version 18.0 for Windows.

RESULTS

Table 2 shows the population characteristics at the survey in 1989/1990, stratified by lung function level. Out of 1,390 genotyped subjects, 228 (17%) had a lung function lower than 80% of predicted. In the group with a lower lung function 153 subjects (67%) were males and 182 subjects (80%) were ever smokers.

Table 2 Characteristics of participants at visit 1989/1990 by lung function level

Lung function (FEV_1 % predicted)	Greater or equal to 80% of predicted	Less than 80% of predicted	p value
N (%)	1,102 (82.9)	228 (17.1)	
Males, n (%)	534 (48.5)	153 (67.1)	<0.001
Age, median (range)	50.4 (35.8-79.1)	57.5 (36.0-76.7)	<0.001
Ever smokers, n (%)	728 (66.1)	182 (79.8)	<0.001
Packyears in ever smokers, median (range)	17.3 (0.1-129.1)	28.3 (0.1-262.2)	<0.001
Height in cm, mean (SD)	170.0 (9.2)	172.0 (8.8)	0.004
Unweighted GRS*,mean (SD)	4.0 (1.2)	4.7 (1.1)	<0.001
Weighted GRS*, mean (SD)	6.7 (1.8)	7.9 (1.8)	<0.001

***** Valid for 1,037 of the genotyped subjects and based on 11 SNPs

Analyses of SNPs

All tested SNPs were in Hardy-Weinberg equilibrium ($p<0.05$). Of the 74 analyzed SNPs 16 showed associations with lung function lower than 80% of predicted at the p value ≤0.10. Nine SNPs (rs4148382, GSTT1 deletion, rs12448760, rs2364723, rs1048290, rs1800470, rs511898, rs528557 and rs1800469) were associated in a dominant model and seven SNPs (rs2842958, rs8187858, rs1695, rs2787094, rs215100, rs504348 and rs1136450) in a recessive model (Table 3). Out of these SNPs, 11 were selected for inclusion in the uGRS and wGRS based on the forward multiple regression approach ($p\leq0.10$). SNPs included in the GRS (rs2842958, rs511898, GSTT1 deletion (rs226633 (18)), rs215100, rs504348, rs4148382, rs1800470, rs1695, rs1048290, rs8187858 and rs1136450) were not in linkage disequilibrium ($r^2<0.17$ in the Vlagtwedde-Vlaardingen cohort).

Genetic Risk Score

The uGRS and wGRS were calculated for 1,037 subjects and normally distributed. As shown in Table 2, the mean number of risk genotypes in subjects with reduced lung function (4.7 (SD 1.1)) was significantly higher compared to the mean in subjects with FEV_1 greater or equal to 80% of predicted (4.0 (SD 1.2)).

Table 4 shows the results of four separate logistic regression models. In the first one packyears was a statistically significant predictor of reduced lung function. The association of the uGRS and wGRS with lung function level was tested in a logistic regression model containing packyears, for each GRS separately. The uGRS and wGRS were significant predictors of reduced lung function level ($p<0.001$). The ORs per unit were 1.8 (95% CI 1.6 to 2.2) and 1.5 (95%

Table 3 SNPs associated with FEV_1 less than 80% of predicted

Gene symbol	SNP ID	Model	OR	p value
ABCC1	rs4148382	dominant	1.91	0.004
GSTT1	deletion	dominant	1.67	0.006
ABCC1	rs12448760	dominant	1.53	0.007
SOD2	rs2842958	recessive	2.26	0.007
NFE2L2	rs2364723	dominant	1.50	0.01
ABCC1	rs8187858	recessive	4.24	0.016
GSTP1	rs1695	recessive	1.72	0.025
KEAP1	rs1048290	dominant	1.38	0.041
ADAM33	rs2787094	recessive	1.79	0.045
ABCC1	rs215100	recessive	1.85	0.045
TGF-β_1	rs1800470	dominant	1.38	0.047
ABCC1	rs504348	recessive	2.13	0.048
ADAM33	rs511898	dominant	1.33	0.064
ADAM33	rs528557	dominant	1.31	0.079
TGF-β_1	rs1800469	dominant	1.31	0.085
SFTPA1	rs1136450	recessive	1.75	0.091

CI 1.3 to 1.7) for uGRS and wGRS respectively. 11 SNPs tested jointly in one model were associated with reduced lung function level at the p value ≤0.10 for each, giving ORs between 1.5 (rs1048290) to 4.7 (rs8187858). As shown in Table 4, the model with wGRS was the preferred one since it had the lowest AIC value among all models tested. Based on the adjusted R^2 obtained from the linear regression on FEV_1%predicted, the observed goodness of fit was comparable between models, ranging from 0.101 for uGRs and for wGRS to 0.103 for 11 SNPs included jointly, but not combined (Table 4).

Table 4 Factors associated with FEV_1

Model*	FEV_1 lower than 80% of predicted				FEV_1% predictedlinear		
	OR (95% CI)	p value	AIC§	AUC±	B (SE)‡	p value	Adjusted R^2
1. packyears	1.0 (1.0-1.0)	<0.001	812.552	0.646	-0.24 (0.02)	<0.001	0.080
2. uGRS	1.8 (1.6-2.2)	<0.001	759.567	0.735	-1.89 (0.39)	<0.001	0.101
3.wGRS	1.5 (1.3-1.7)	<0.001	756.977	0.740	-1.18 (0.24)	<0.001	0.101
4. SNPs tested in multiple regression							
*SOD2*rs2842958	3.2 (1.5-6.6)	0.002			-3.70 (2.31)	0.109	
*ADAM33*rs511898	1.7 (1.2-2.6)	0.004			-2.23 (0.91)	0.014	
GSTT1 deletion	1.9 (1.2-2.9)	0.007			-2.35 (1.21)	0.052	
*ABCC1*rs215100	3.2 (1.3-7.7)	0.009			0.51 (1.60)	0.750	
*ABCC1*rs504348	4.2 (1.4-12.4)	0.011			-1.57 (2.94)	0.594	
*ABCC1*rs4148382	2.0 (1.2-3.3)	0.012			-3.70 (1.14)	0.001	
*TGF-β_1*rs1800470	1.6 (1.1-2.3)	0.015			-2.07 (0.95)	0.029	
*GSTP1*rs1695	1.9 (1.1-3.5)	0.029	770.480	0.749	-0.91 (1.24)	0.464	0.103
*KEAP1*rs1048290	1.5 (1.0-2.2)	0.036			-0.52 (0.95)	0.585	
*ABCC1*rs8187858	4.7 (1.1-20.5)	0.038			-10.32 (4.52)	0.022	
*SFTPA1*rs1136450	2.1 (0.9-4.7)	0.079			-1.77 (2.29)	0.441	

*In Model 2, 3 and 4 indicated variables were included next to packyears

§The Akaike information criterion

±The area under the curve, referring to the model containing packyears and the indicated explanatory variable

‡Linear regression on FEV_1%predicted as a continuous variable,referring to the model containing packyears and the indicated explanatory variable

Next we compared the discriminative power for FEV_1 less than 80% predicted of the analyzed models. The AUC for packyears + uGRS, packyears + wGRS and packyears + 11 SNPs were compared with the AUC for the model that tested packyears. Figure 1 shows that inclusion of the uGRS or wGRS in the model improved the AUC (0.735 and 0.740 respectively) over that observed when ROC curve was based on packyears only (0.646). This improvement was statistically significant ($p<0.001$) for both uGRS and wGRS. The discriminatory power of the model containing packyears and 11 SNPs included jointly was the highest among all models (AUC=0.749), the achieved improvement in the AUC was 0.103 ($p<0.001$) compared to the AUC obtained using packyears (Figure 1).

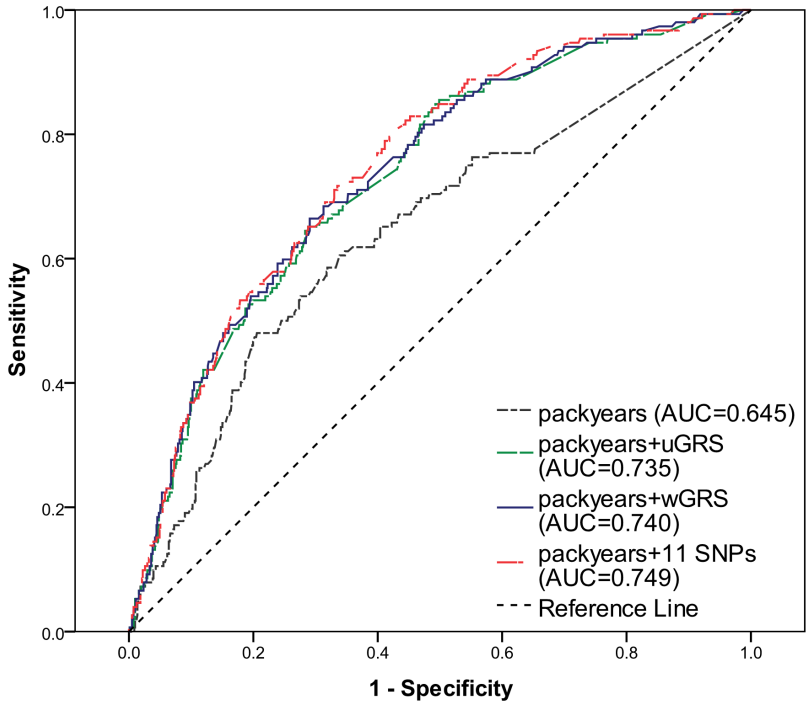


Figure 1 Receiver-operating characteristic curves for models predicting reduced lung function level.

DISCUSSION

The results of this study show that GRS weighted by the odds ratio of each risk genotype, better discriminates subjects with reduced lung function than GRS simply summing risk genotypes (i.e. uGRS). This is in agreement with previous studies showing that a weighted GRS can capture considerably more genetic risk than an unweighted GRS (uGRS) (7;19). Piccolo *et al.* claim that aggregating data from many markers identified in genome-wide association (GWA) study into a single GRS variable allows genetic and biological information related to phenotype to be condensed into one dimension (19). However, so far none of the previous studies attempted to proof this statement by investigating the effect of the selected SNPs jointly, but not combined in one variable. In the current study we additionally tested the joint effect of all SNPs used for the GRS, when not combined in one variable. Using the AIC, we found that this model, although it showed the highest increase in the AUC, does not fit the data as well as the model with wGRS.

Pulmonary function, measured by spirometry, is a heritable trait reflecting the physiological state of the airways and lungs (20). Nearly 20% of the adult population has reduced FEV_1 values, indicating impaired lung function (21). There is overwhelming evidence from large scale epidemiological studies demonstrating that reduced FEV_1 is a powerful predictor of future morbidity and mortality (1;22-24). FEV_1 is not only a way to quantify airflow limitation, but it can also be used as a predictor of premature death with a broad clinical use, and possibly for prevention of both respiratory (COPD and lung cancer) and cardiovascular (coronary artery disease and stroke) diseases (1).

Apart from smoking (25), lung function is largely influenced by genetics (4;5) Therefore it is not enough to merely promote smoking

cessation and a greater effort should be made to identify genetically susceptible subjects (1).

We found 11 loci to be associated with reduced lung function in the general population and the accuracy to discriminate subjects with reduced lung function improved significantly when these 11 SNPs were combined into a GRS. Given the importance of reduced lung function in pulmonary and cardiovascular morbidity and mortality, a prevention strategy is needed. Therefore we might consider the clinical utility of a GRS to identify subjects at increased risk for reduced lung function in early adulthood. Nonetheless, the clinical usefulness of this GRS should be further confirmed in other general population studies.

Strengths and limitations

The major strength of the current study is a number of tested candidate genes i.e. 24 that previously have been shown to play a role in lung tissue composition, protease-antiprotease balance, antioxidant response in the lungs, xenobiotics transport and chronic inflammation. In contrast most of the previous studies that composed a GRS used SNPs associated with a disease in GWA studies. In GWA studies genes with a small effect are more likely to be missed since modest association signals are filtered out in correction for multiple testing. Here we followed a candidate gene approach, of which an advantage is that we were able to test the best genetic model for the associations (in this study it was dominant or recessive) and obtain risk genotypes, whereas in screening by GWA studies just an additive model is assumed.

This study has some limitations. First, most of the genotyped SNPs were chosen based on results presented in previous literature. This may cause a deficiency in our study, since these SNPs not always cover the complete genetic variation in the gene, and hence we can miss associations and cannot pinpoint the causal genetic variant. However, for four tested genes (*SIRT1, ABCC1, NFE2L2* and *KEAP1)* tagging SNPs were selected aiming to cover the whole gene. It would be of interest to test whether inclusion of more SNPs, also those identified in GWA studies, can improve prediction. Second, the GRS models were investigated in the same data on which association testing and construction of the models were performed. Future studies in other population cohorts should indicate whether these results can be generalized to other populations.

In conclusion, we found that a GRS composed of 11 SNPs is associated with reduced lung function level. Appropriate weighing of the risk genotypes included in the GRS resulted in a better discriminatory tool than the simple count of risk genotypes. Therefore weighted GRS may be useful for identifying subjects with reduced lung function. Further studies, which validate the discriminatory accuracy of these SNPs, need to be undertaken.

ACKNOWLEGMENTS

Author contributions: JM Vonk and HM Boezen planned the project and supervised its execution, and analysis. SM Figarska carried out the data analysis and drafted the paper. All authors contributed to the writing and preparation of the paper.

Funding: Stichting Astma Bestrijding (grantnumber 2005/020). GUIDE, UMCG, the Netherlands.

Role of the funding source: The funding source had no role in the study design, in the collection, analysis, and interpretation of the data, in the writing of the report, and in the decision to submit the paper for publication.

Competing interests: None to declare.

Ethnical approval: The Committee on Human Subjects in Research of the University of Groningen reviewed the study and affirmed the safety of the protocol and study design and all participants gave their written informed consent.

REFERENCES

1. Young RP, Hopkins R, Eaton TE. Forced expiratory volume in one second: not just a lung function test but a marker of premature death from all causes. Eur Respir J 2007 Oct;30(4):616-22.
2. Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yernault JC. Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. Eur Respir J Suppl 1993 Mar;16:5-40.
3. Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med 2007 Sep 15;176(6):532-55.
4. Hubert HB, Fabsitz RR, Feinleib M, Gwinn C. Genetic and environmental influences on pulmonary function in adult twins. Am Rev Respir Dis 1982 Apr;125(4):409-15.
5. McClearn GE, Svartengren M, Pedersen NL, Heller DA, Plomin R. Genetic and environmental influences on pulmonary function in aging Swedish twins. J Gerontol 1994 Nov;49(6):264-8.
6. Lewitter FI, Tager IB, McGue M, Tishler PV, Speizer FE. Genetic and environmental determinants of level of pulmonary function. Am J Epidemiol 1984 Oct;120(4):518-30.
7. Chen H, Poon A, Yeung C, Helms C, Pons J, Bowcock AM, et al. A genetic risk score combining ten psoriasis risk loci improves disease prediction. PLoS One 2011;6(4):e19454.
8. Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. N Engl J Med 2008 Nov 20;359(21):2208-19.
9. Weedon MN, McCarthy MI, Hitman G, Walker M, Groves CJ, Zeggini E, et al. Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. PLoS Med 2006 Oct;3(10):e374.
10. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. Genome Res 2007 Oct;17(10):1520-8.
11. Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet 2013 Mar;9(3):e1003348.
12. Ganna A, Magnusson PK, Pedersen NL, de FU, Reilly M, Arnlov J, et al. Multilocus Genetic Risk Scores for Coronary Heart Disease Prediction. Arterioscler Thromb Vasc Biol 2013 Sep;33(9):2267-72.

13. Hughes MF, Saarela O, Stritzke J, Kee F, Silander K, Klopp N, et al. Genetic markers enhance coronary risk prediction in men: the MORGAM prospective cohorts. PLoS One 2012;7(7):e40922.

14. Kwak SH, Choi SH, Kim K, Jung HS, Cho YM, Lim S, et al. Prediction of type 2 diabetes in women with a history of gestational diabetes using a genetic risk score. Diabetologia 2013 Sep 22.

15. van Diemen CC, Postma DS, Vonk JM, Bruinenberg M, Schouten JP, Boezen HM. A disintegrin and metalloprotease 33 polymorphisms and lung function decline in the general population. Am J Respir Crit Care Med 2005 Aug 1;172(3):329-33.

16. Gu X, Frankowski RF, Rosner GL, Relling M, Peng B, Amos CI. A modified forward multiple regression in high-density genome-wide association studies for complex traits. Genet Epidemiol 2009 Sep;33(6):518-25.

17. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation 2007 Feb 20;115(7):928-35.

18. Zhao Y, Marotta M, Eichler EE, Eng C, Tanaka H. Linkage disequilibrium between two high-frequency deletion polymorphisms: implications for association studies involving the glutathione-S transferase (GST) genes. PLoS Genet 2009 May;5(5):e1000472.

19. Piccolo SR, Abo RP, Allen-Brady K, Camp NJ, Knight S, Anderson JL, et al. Evaluation of genetic risk scores for lipid levels using genome-wide markers in the Framingham Heart Study. BMC Proc 2009;3 Suppl 7:S46.

20. Soler Artigas M, Loth DW, Wain LV, Gharib SA, Obeidat M, Tang W, et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. Nat Genet 2011 Nov;43(11):1082-90.

21. Mannino DM, Ford ES, Redd SC. Obstructive and restrictive lung disease and markers of inflammation: data from the Third National Health and Nutrition Examination. Am J Med 2003 Jun 15;114(9):758-62.

22. Hole DJ, Watt GC, Davey-Smith G, Hart CL, Gillis CR, Hawthorne VM. Impaired lung function and mortality risk in men and women: findings from the Renfrew and Paisley prospective population study. BMJ 1996 Sep 21;313(7059):711-5.

23. Knuiman MW, James AL, Divitini ML, Ryan G, Bartholomew HC, Musk AW. Lung function, respiratory symptoms, and mortality: results from the Busselton Health Study. Ann Epidemiol 1999 Jul;9(5):297-306.

24. Schunemann HJ, Dorn J, Grant BJ, Winkelstein W, Jr., Trevisan M. Pulmonary function is a long-term predictor of mortality in the general population: 29-year follow-up of the Buffalo Health Study. Chest 2000 Sep;118(3):656-64.

25. Lange P, Groth S, Nyboe GJ, Mortensen J, Appleyard M, Jensen G, et al. Effects of smoking and changes in smoking habits on the decline of FEV1. Eur Respir J 1989 Oct;2(9):811-6.

Chapter 7

Dyspnea severity, changes in dyspnea status and mortality in the general population: The Vlagtwedde-Vlaardingen study

SM Figarska[1,2], HM Boezen[1,2] and JM Vonk[1,2]

[1] Department of Epidemiology, University of Groningen, University Medical Center Groningen, Hanzeplein 1, P.O. Box 30001, 9700 RB Groningen, The Netherlands

[2] GRIAC reseach institute, University Medical Center Groningen, Hanzeplein 1, P.O. Box 30001, 9700 RB Groningen, The Netherlands

Eur J Epidemiol 2012 Nov;27(11):867-76.

ABSTRACT

Dyspnea is a predictor of mortality. The effects of dyspnea severity and changes in dyspnea status on all-cause and cause-specific mortality remain unclear. The Vlagtwedde-Vlaardingen study started in 1965 and subjects were re-examined every three years until 1989/1990. Vital status of all 8,465 subjects on December 31st, 2008 was assessed. Associations between mortality and dyspnea severity and changes in dyspnea status were investigated using Cox regression adjusted for gender, age, FEV_1%predicted, place of residence, smoking and BMI. After 43 years of follow-up, 2,883 (39%) of 7,360 subjects examined for dyspnea severity had died, 1,386 (19%) due to cardiovascular disease, 267 (4%) due to chronic obstructive pulmonary disease (COPD). Subjects with moderate and severe dyspnea had increased all-cause and cardiovascular mortality [moderate: HR=1.3 (95% CI 1.2-1.5) and 1.4 (1.1-1.6), severe: 1.5 (1.1-2.0) and 1.9 (1.3-2.6) respectively] compared to asymptomatics. Severe dyspnea was significantly associated with COPD mortality [3.3 (2.0-5.2)]. Subjects who lost dyspnea had hazard ratios for all-cause and cause-specific mortality comparable to asymptomatics. Persistent dyspnea and dyspnea development were risk factors for all-cause, cardiovascular and COPD mortality [persistent: 2.0 (1.4-2.8), 1.9 (1.2-3.3) and 3.3 (1.2-8.9), development: 1.5 (1.2-1.8), 2.0 (1.5-2.6) and 3.8 (2.3-6.3) respectively]. Additionally, dyspnea effects on mortality were more pronounced in overweight/obese and older subjects and in subjects with better lung function. These results show that dyspnea is associated with mortality in a severity-dependent manner. Furthermore this study is the first showing that dyspnea remission normalizes mortality risk. Having or developing dyspnea is a risk factor for mortality.

Keywords *dyspnea, dyspnea severity, dyspnea remission, mortality, risk factor, longitudinal studies*

INTRODUCTION

Dyspnea is a subjective experience of breathing discomfort that consists of qualitatively distinct sensations that vary in intensity (1). Dyspnea on exertion is a common symptom not only in patients with lung and heart diseases, but it is also fairly prevalent (i.e. 24%) and associated with poor health outcome in people with no apparent pre-existing disease (2;3). Therefore dyspnea is of public health importance and needs more attention, especially since this symptom is easily recognized and detected. One of the tools to quantify the intensity of dyspnea on exertion is the self-reported Medical Research Council (MRC) questionnaire, with a scale developed by Fletcher and coworkers (4). The MRC is an excellent instrument for categorizing patients according to the severity of their dyspnea (5;6). A positive response to a simple question about effort-related dyspnea can predict subsequent mortality, independently of other risk factors (3) . Furthermore, dyspnea is a better predictor of 5-year survival than airway obstruction in patients with COPD (6).

Several studies have established associations between dyspnea and all-cause and cardiovascular mortality (7-14). However, little is known about the association between severity of dyspnea and all-cause mortality. Although two previous studies have investigated the threshold at which dyspnea is a determinant of all-cause mortality (8;13), those were not controlled for lung function level, an important long-term predictor of mortality (15-17). Since the level of lung function is associated with dyspnea (2), lung function should be considered as an influential factor that potentially modifies the relationship between scores of dyspnea and mortality. Besides lung function, gender, weight and age are important factors determining dyspnea intensity (18) and need attention when association between dyspnea and mortality are studied.

Respiratory symptoms are often transient, with a high remission rate (19). Xu and colleagues (20) have shown that dyspnea is likely to develop or remit during lifetime. Whether these changes in dyspnea status modify mortality risk has not been previously examined. We have the unique opportunity to study this relation in a large population-based cohort.

The Vlagtwedde-Vlaardingen cohort, which has been followed up for over 40 years, offers us the unique possibility to investigate the presence and severity of respiratory symptoms over lifetime and their influence on all-cause and cause-specific mortality. Since lung function has also been measured in this cohort, we are able to investigate whether the associations between dyspnea and mortality are independent of lung function level. To our knowledge there is no study showing dyspnea-related mortality in groups of subjects categorized based on their gender, body mass index (BMI), age and lung function level. In the current study we evaluated long-term effects of dyspnea severity and changes in dyspnea status on all-cause, cardiovascular, and COPD mortality in this large general population-based cohort. Additionally we studied these effect taking gender, BMI, age and level of lung function into account, which are possible modifiers of these associations.

METHODS

Participants and measurements

The Vlagtwede/Vlaardingen study is based on a general population cohort of exclusively Caucasian individuals of Dutch descent. This study started in 1965 and participants had medical examinations every 3 years until the last survey in 1989/1990. In Vlaardingen, only

participants who were included at baseline (1965 or 1969) were approached for follow-up, whereas in Vlagtwedde new subjects aged between 20 and 65 years were invited to participate at every survey. The Committee on Human Subjects in Research of the University of Groningen reviewed the study and affirmed the safety of the protocol and study design and all participants gave their written informed consent. Information on area of residence, age, sex, smoking habits and respiratory symptoms was collected by the Dutch version of the UK Medical Research Council standard questionnaire, and spirometry was performed (20).

Level of dyspnea on exertion at baseline was evaluated by asking the subject whether he/she felt shortness of breath in the following circumstances: never except on strenuous exercise (dyspnea grade I), when hurrying or walking up a slight hill (grade II), while walking with other people of the same age on level ground (grade III), after walking about 100 m or a few minutes on level ground (grade IV), during everyday activities such as dressing or undressing (grade V), or at rest (grade VI). Using dyspnea grade at entry, we defined severity of dyspnea: no dyspnea (grade I and II) (15;21), moderate (III and IV) and severe (V and VI). For subjects who had at least three available surveys we investigated changes in dyspnea status creating five groups: never dyspnea, persistent dyspnea, development of dyspnea, remission of dyspnea and inconsistent dyspnea. In this analysis moderate and severe dyspnea was taken together indicating the presence of dyspnea.

We updated the vital status of all participants in the Vlagtwedde-Vlaardingen study on December 31st, 2008, and evaluated three main mortality outcomes, i.e. all-cause mortality (excluding external causes of death such as accidents and suicides), cardiovascular and

COPD mortality (as primary or secondary cause of death). Analyses on cause-specific mortality were performed at Statistics Netherlands (The Hague, the Netherlands). The causes of death were coded according to the International Classification of Diseases (Table S1).

Statistical methods

At first, a descriptive analysis of the subject characteristics was performed. The results are shown as median (range) or number (percentage) as appropriate. Mann-Whitney U tests, Kruskal-Wallis tests and χ^2 tests were used to compare the subject characteristics between groups of dyspnea severity and between groups of changes in dyspnea status. To evaluate the long-term effect of dyspnea severity and changes in dyspnea status on all-cause and cause-specific mortality, we performed Cox proportional hazard regression and produced Cox regression curves to compare survival of the subjects according to their dyspnea severity and change in dyspnea status. The dependent variable was time to mortality. Individuals who had not died or died from a cause other than the one of interest were considered censored. External causes of death were excluded from the analyses. Dyspnea was introduced as the independent (explanatory) variable. We adjusted for potential confounders: gender, age, place of residence, smoking habits and BMI at baseline. Since the perception of dyspnea is strongly associated with lung function level (2;22), which is a predictor of mortality (15;16), we first evaluated the effect of dyspnea in models not corrected for lung function level and subsequently adjusted the analyses for FEV_1 % predicted.

Time was defined from the initial examination until death, end of follow-up in 2008 or last registration if subjects were lost to follow-

up. Smoking at baseline was treated as a categorical variable: never smokers, former smokers and current smokers. Smokers of pipes and cigars were classified separately. BMI at baseline was estimated by dividing the weight by height^2 and categorized into four groups according to World Health Organization (WHO) criteria: underweight (BMI < 18.5 kg/m^2), normal weight (BMI 18.5 to 25 kg/m^2), overweight (BMI 25 to 30 kg/m^2) and obese (BMI > 30 kg/m^2). To investigate the effects of possible modifiers of the observed associations, i.e. gender, BMI, age and lung function level, we performed stratified analyses. Every potential factor was dichotomized using cut points (for BMI: 25 kg/m^2, for age: 75th percentile of all the subjects and for lung function level: median FEV_1 % predicted). P value <0.05 tested 2-sided was considered statistically significant for all analyses. All statistical analyses were performed using SPSS version 18.0 for Windows.

RESULTS

Table 1 shows the characteristics of subjects examined for dyspnea severity at their initial survey. Of a total of 8,465 subjects, information on dyspnea was available for 7,360 subjects (87%). Among them moderate dyspnea was reported by 388 subjects (5%) and severe dyspnea by 67 (1%) subjects.

Table 2 shows the characteristics of 3,991 subjects available for evaluation of changes in dyspnea status. A minority of subjects (1%) had persistent dyspnea, 6% of subjects developed dyspnea and 2% indicated remission of dyspnea during the follow-up. Most subjects (87%) never reported having dyspnea. Since answers of 153 subjects (4%) were not consistent, these subjects were treated as a separate group (inconsistent).

Table 1 Characteristics of subjects at baseline by dyspnea severity

	No dyspnea n=6,905 (93.8%)	Moderate n=388 (5.3%)	Severe n=67 (0.9%)	p value[a]	p value[b]	p value[c]
Sex						
Male, n (%)	3,634 (52.6)	147 (37.9)	32 (47.8)			
Female, n (%)	3,271 (47.4)	241 (62.1)	35 (52.2)	< 0.001	0.427	0.127
Area of residence						
Vlagtwedde, n (%)	4,392 (63.6)	178 (45.9)	32 (47.8)	< 0.001	0.007	0.775
Vlaardingen, n (%)	2,513 (36.4)	210 (54.1)	35 (52.2)			
Median age						
Age at baseline, yr (range)	35 (14-79)	46 (14-67)	52 (19-64)	< 0.001	< 0.001	0.001
Age on Dec 31 2008, yr (range)	68 (16-103)	73 (42-97)	71 (50-91)	< 0.001	0.048	0.641
Smoking status at baseline						
Never, n (%)	2,442 (35.5)	175 (45.7)	23 (34.9)			
Former, n (%)	718 (10.4)	34 (8.9)	8 (12.1)			
Pipe/cigar, n (%)	277 (4.1)	10 (2.6)	3 (4.5)	< 0.001	0.967	0.351
Current, n (%)	3,441 (50.0)	164 (42.8)	32 (48.5)			
Lung function at baseline						
FEV_1 % predicted, median (range)	89.0 (26.0-140.9)	80.3 (20.7-123.7)	71.6 (21.6-99.9)	< 0.001	< 0.001	< 0.001
BMI at baseline						
Underweight, n (%)	109 (1.6)	5 (1.3)	2 (3.0)			
Normal, n (%)	3,487 (50.5)	128 (33.0)	18 (26.9)			
Overweight, n (%)	2,610 (37.8)	168 (43.3)	30 (44.8)	< 0.001	< 0.001	0.576
Obese, n (%)	699 (10.1)	87 (22.4)	17 (25.4)			

Status on Dec 31st,2008						
Alive, n (%)	4,215 (61.1)	134 (34.6)	17 (25.4)			
Dead, n (%)	2,584 (37.4)	250 (64.4)	49 (73.1)			
Lost to follow-up, n (%)	106 (1.5)	4 (1.0)	1 (1.5)	< 0.001	< 0.001	0.330
Causes of death						
CVD-primary or secondary, n (%)[d]	1,217 (47.1)	135 (54.0)	34 (69.4)	0.042[e]	0.002	0.047
COPD-primary or secondary, n (%)[d]	219 (8.5)	26 (10.4)	22 (44.9)	0.310[f]	< 0.001	<0.001
External causes, n (%)	108 (4.2)	11 (4.4)	2 (4.1)	0.878[g]	0.968	0.920

[a]Differences between subjects without dyspnea and subjects with moderate dyspnea, [b]Differences between subjects without dyspnea and subjects with severe dyspnea, [c]Differences between subjects with moderate dyspnea and subjects with severe dyspnea, [d]% of all deaths in the group, 137 subjects have CVD as primary and COPD as secondary cause of death or *vice versa*, [e]Compared to mortality not due to CVD, [f]Compared to mortality not due to COPD, [g]Compared to mortality not due to external causes.

Table 2 Characteristics of subjects at baseline by changes in dyspnea status

	Never	Development		Persistent	Remission		Inconsistent	
	n=3,478 (87.1%)	n=218 (5.5%)	p value[a]	n=48 (1.2%)	n=94 (2.4%)	p value[b]	n=153 (3.8%)	p value[c]
Sex								
Male, n (%)	1,871 (53.8)	114 (52.3)	0.666	16 (33.3)	29 (30.9)	0.764	78 (51.0)	< 0.001
Female, n (%)	1,607 (46.2)	104 (47.7)		32 (66.7)	65 (69.1)		75 (49.0)	
Area of residence								
Vlagtwedde, n (%)	2,186 (62.9)	109 (50.0)	< 0.001	11 (22.9)	32 (34.0)	0.172	72 (47.1)	< 0.001
Vlaardingen, n (%)	1,292 (37.1)	109 (50.0)		37 (77.1)	62 (66.0)		81 (52.9)	
Median age								
Age at baseline, yr (range)	32 (14-58)	40 (15-63)	< 0.001	41 (18-58)	33 (15-54)	< 0.001	39 (15-53)	< 0.001
Age at Dec 31 2008, yr (range)	68 (26-97)	72 (42-96)	< 0.001	73 (44-91)	71 (51-93)	0.900	74 (42-92)	< 0.001
Smoking status at baseline								
Never smokers, n (%)	1,202 (34.7)	70 (32.3)		20 (41.6)	42 (45.2)		44 (29.0)	
Former smokers, n (%)	368 (10.5)	20 (9.2)	0.557	2 (4.2)	5 (5.4)	0.954	11 (7.2)	0.153
Pipe/cigar smokers, n (%)	112 (3.2)	5 (2.3)		1 (2.1)	2 (2.1)		3 (2.0)	
Current smokers, n (%)	1,789 (51.6)	122 (56.2)		25 (52.1)	44 (47.3)		94 (61.8)	
Lung function at baseline								
FEV_1% predicted, median (range)	89.9 (37.1-138.7)	83.5 (35.9-120.6)	< 0.001	76.2 (23.7-96.2)	88.5 (55.0-116.6)	< 0.001	83.4 (28.0-121.4)	< 0.001
BMI at baseline								
Underweight, n (%)	57 (1.6)	1 (0.5)		2 (4.2)	1 (1.1)		0 (0.0)	

Normal, n (%)	1,843 (53.1)	89 (41.0)	< 0.001	13 (27.1)	38 (40.4)	0.031	67 (44.1)	< 0.001
Overweight, n (%)	1,316 (37.9)	95 (43.8)		19 (39.6)	44 (46.8)		65 (42.8)	
Obese, n (%)	256 (7.4)	32 (14.7)		14 (29.2)	11 (11.7)		20 (12.2)	
Status on Dec 31st,2008								
Alive, n (%)	2,505 (72.0)	100 (45.9)		11 (22.9)	65 (69.1)		81 (52.9)	
Dead, n (%)	944 (27.2)	118 (54.1)	< 0.001	37 (77.1)	29 (30.9)	< 0.001	71 (46.4)	< 0.001
Lost to follow-up, n (%)	29 (0.8)	0 (0.0)		0 (0.0)	0 (0.0)		1 (0.7)	
Causes of death								
CVD-primary or secondary, n (%)[d]	387 (41.0)	71 (60.2)	0.001	17 (45.9)	14 (48.3)	0.851	38 (53.5)	0.001
COPD-primary or secondary, n (%)	61 (6.4)	24 (20.3)	< 0.001	6 (16.2)	1 (3.4)	0.095	7 (9.9)	< 0.001
External causes, n (%)	40 (4.2)	3 (2.5)	0.372	4 (10.8)	0 (0.0)	0.068	1 (1.4)	0.103

[a]Differences between never dyspnea and dyspnea development, [b]Differences between persistent dyspnea and dyspnea remission, [c]Differences among all groups d % of all deaths in the group

Table 3 Hazard ratio (HR) with 95% confidence interval (CI) for all-cause, cardiovascular and COPD mortality by severity of dyspnea

Dyspnea severity	All-cause mortality[a]		Cardiovascular mortality		COPD mortality	
	HR[b] (95% CI)	HR[c] (95% CI)	HR[b] (95% CI)	HR[c] (95% CI)	HR[b] (95% CI)	HR[c] (95% CI)
no dyspnea N=6,905	1	1	1	1	1	1
moderate N=388	1.5 (1.3-1.7)	1.3 (1.2 - 1.5)	1.6 (1.3-1.9)	1.4 (1.1 - 1.6)	2.3 (1.5-3.6)	1.1 (0.7 - 1.7)
severe N=67	1.9 (1.4-2.5)	1.5 (1.1 - 2.0)	2.5 (1.8-3.6)	1.9 (1.3 - 2.6)	9.3 (5.9-14.6)	3.3 (2.0 - 5.2)

HR[c] (95% CI) for possible modifiers (no dyspnea as a reference category)

(a) gender	Females n=3,547		Males n=3,813		Females		Males		Females		Males	
	n/N[d]	HR[3] (95% CI)	n/N	HR[c] (95%CI)	n	HR[c] (95% CI)	n	HR[c] (95% CI)	n	HR[c] (95% CI)	n	HR[c] (95% CI)
moderate	152/241	1.4 (1.1-1.6)	87/147	1.3 (1.1-1.7)	89	1.4 (1.1-1.8)	46	1.3 (0.9-1.7)	9	1.2 (0.5-2.5)	17	1.1 (0.7-1.9)
severe	20/35	1.5 (0.9-2.4)	27/32	1.5 (1.0-2.2)	16	2.2 (1.3-3.8)	18	1.8 (1.1-2.9)	4	2.6 (0.8-8.1)	18	3.9 (2.2-6.6)
(b) BMI	< 25 n=3,749		≥ 25 n=3,611		< 25		≥ 25		< 25		≥ 25	
moderate	63/133	1.4 (1.1-1.8)	175/255	1.4 (1.2-1.6)	32	1.5 (1.0-2.2)	103	1.4 (1.1-1.8)	6	0.5 (0.2-1.3)	20	1.8 (1.1-2.9)
severe	12/20	1.3 (0.7-2.4)	35/47	1.5 (1.1-2.2)	7	1.6 (0.7-3.4)	27	2.1 (1.4-3.1)	9	2.7 (1.3-5.8)	13	4.6 (2.4-8.6)

(c) age	≤ 46 yrs n=5,568		>46 yrs n=1,792		≤ 46 yrs		>46 yrs		≤ 46 yrs		>46 yrs	
moderate	75/201	1.4 (1.1-1.7)	164/187	1.3 (1.1-1.5)	29	1.1 (0.7-1.6)	106	1.5 (1.2-1.8)	9	1.2 (0.6-2.5)	17	1.1 (0.7-1.9)
severe	8/23	1.2 (0.6-2.4)	39/44	1.5 (1.1-2.1)	6	1.8 (0.8-4.0)	28	1.9 (1.3-2.9)	2	1.2 (0.3-5.3)	20	3.8 (2.2-6.4)

(d) lung function	< 88.5 %predicted n=3,673		≥ 88.5 %predicted n=3,687		< 88.5 %predicted		≥ 88.5 %predicted		< 88.5 %predicted		≥ 88.5 %predicted	
moderate	179/276	1.3 (1.1-1.5)	60/112	1.5 (1.1-1.9)	100	1.2 (1.0-1.5)	35	1.9 (1.3-2.6)	26	1.2 (0.8-1.8)	0	-
severe	39/55	1.2 (0.9-1.7)	8/12	3.2 (1.6-6.5)	27	1.4 (1.0-2.1)	7	5.6 (2.6-11.9)	21	3.1 (1.9-5.0)	1	13.9 (1.8-108.0)

[a]Excluding external causes of death, [b]Adjusted for age, gender, place of residence, smoking habits and BMI at baseline, [c]Adjusted for age, gender, place of residence, smoking habits, BMI and FEV_1 % predicted at baseline, [d]n=number of deaths in the group/ N=number of all subjects in the group.

Dyspnea severity and all-cause, cardiovascular and COPD mortality

Moderate dyspnea and severe dyspnea were significantly associated with all-cause, cardiovascular and COPD mortality in a severity-dependent manner. Although the risks of all the investigated mortality outcomes associated with moderate and severe dyspnea were lower when corrected for lung function, dyspnea remained a significant independent predictor of mortality. Only the association between moderate dyspnea and COPD mortality was no longer significant after adjustment for FEV_1 % predicted (Table 3). Survival curves according to severity of dyspnea clearly show that all-cause and cardiovascular mortality risks were significantly increased among dyspneic subjects and the highest risks were observed among subjects with severe dyspnea (Figure 1). Regarding COPD mortality, only the survival curve that represents subjects with severe dyspnea clearly shows an increased risk. Stratification for gender did not change the observed associations in all-cause mortality. The effect of severe dyspnea on cardiovascular mortality was more pronounced in females and on COPD mortality more pronounced in males. The mortality risk associated with severe dyspnea was more prominent in subjects with BMI ≥ 25 and in the older group. Interestingly, in the analysis stratified according to lung function, the mortality risk associated with dyspnea was highest for subjects with an above median FEV_1 % predicted. As a sensitivity analysis we tested the risk of all-cause mortality for six original dyspnea grades (Table S2). To ensure that grade II included in a reference category did not affect associations described below we performed analysis, where grade II was tested separately from grade I, as mild dyspnea (Table S2).

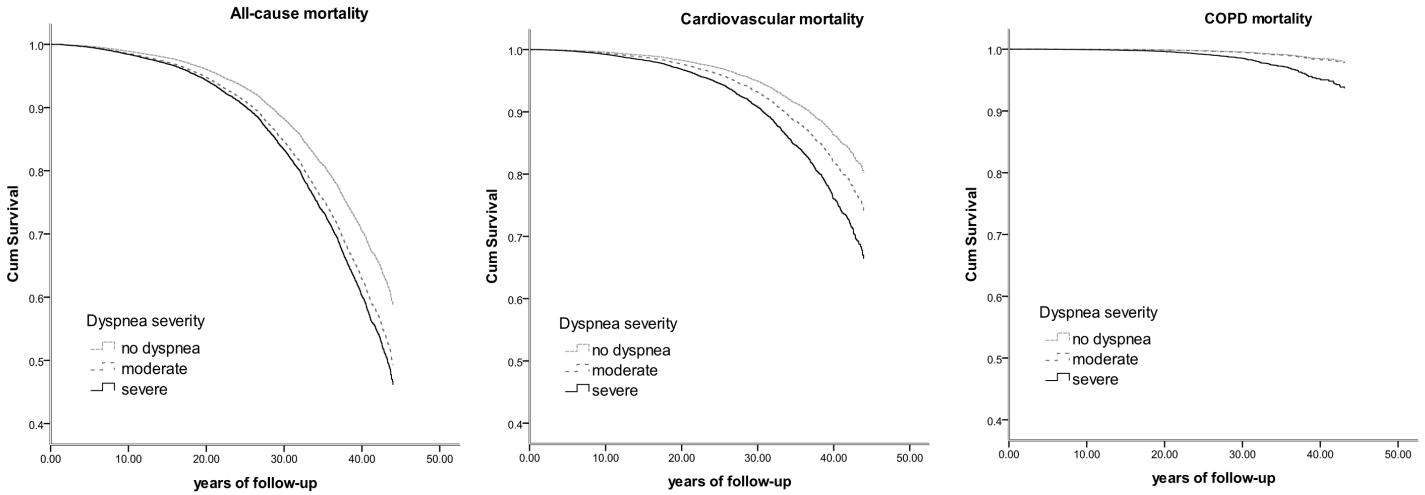


Figure 1 Cox proportional hazard regression survival curves showing the relation between dyspnea severity and all-cause, cardiovascular and COPD mortality

Table 4 Hazard ratio (HR) with 95% confidence interval (CI) for all-cause, cardiovascular and COPD mortality by changes in dyspnea

Change in dyspnea status	All-cause mortality[a]		Cardiovascular mortality		COPD mortality	
	HR[b] (95% CI)	HR[c] (95% CI)	HR[b] (95% CI)	HR[c] (95% CI)	HR[b] (95% CI)	HR[c] (95% CI)
never N=3,478	1	1	1	1	1	1
persistent N=48	2.4 (1.7-3.5)	2.0 (1.4 - 2.8)	2.5 (1.5-4.1)	1.9 (1.2 - 3.3)	8.6 (3.6-20.7)	3.3 (1.2 - 8.9)
development N=218	1.6 (1.3-2.0)	1.5 (1.2 - 1.8)	2.2 (1.7-2.8)	2.0 (1.5 - 2.6)	5.3 (3.3-8.7)	3.8 (2.3 - 6.3)
remission N=94	1.0 (0.7-1.5)	1.0 (0.7 - 1.4)	1.2 (0.7-2.1)	1.1 (0.6 - 1.9)	0.6 (0.1-4.5)	0.6 (0.1 - 4.0)
inconsistent N=153	1.2 (0.9-1.5)	1.1 (0.9 - 1.4)	1.5 (1.1-2.1)	1.4 (1.0 - 1.9)	1.9 (0.8-4.1)	1.3 (0.6 - 2.8)

HR[c] (95% CI) for possible modifiers (never dyspnea as a reference category)

(a) gender	Females n=1,883		Males n=2,108		Females		Males		Females		Males	
	n/N[d]	HR[3] (95% CI)	n/N	HR[c] (95%CI)	n	HR[c] (95% CI)	n	HR[c] (95% CI)	n	HR[c] (95% CI)	n	HR[c] (95% CI)
persistent	22/32	1.8 (1.1-2.9)	11/16	2.3 (1.3-4.4)	13	2.4 (1.3-4.5)	4	1.5 (0.5-4.1)	3	5.2 (1.1-24.3)	3	4.2 (1.0-17.0)
development	50/104	1.5 (1.1-2.1)	65/114	1.5 (1.1-1.9)	30	2.1 (1.4-3.3)	41	1.9 (1.3-2.7)	6	4.1 (1.4-11.7)	18	3.8 (2.1-6.8)
remission	19/65	0.9 (0.5-1.4)	10/29	1.3 (0.7-2.4)	10	1.1 (0.6-2.1)	4	1.1 (0.4-3.0)	0	-	1	1.7 (0.2-12.6)
(b) BMI	< 25 n=2,111		≥ 25 n=1,872		< 25		≥ 25		< 25		≥ 25	
persistent	11/15	2.5 (1.3-4.6)	22/33	2.0 (1.3-3.2)	4	2.0 (0.7-5.6)	13	2.3 (1.3-4.2)	1	0.7 (0.1-6.1)	5	7.1 (2.5-20.3)
development	43/90	1.4 (1.0-1.9)	72/127	1.5 (1.2-2.0)	23	1.6 (1.0-2.5)	48	2.2 (1.6-3.0)	11	2.0 (0.9-4.3)	13	5.4 (2.7-10.5)

remission	8/39	0.9 (0.4-1.9)	21/55	1.0 (0.6-1.6)	2	0.6 (0.1-2.4)	12	1.3 (0.7-2.3)	1	0.9 (0.1-7.1)	0	-
(c) age	≤ 46 yrs n=3,584		>46 yrs n=399		≤ 46 yrs		>46 yrs		≤ 46 yrs		>46 yrs	
persistent	19/30	2.6 (1.6-4.1)	14/18	1.8 (1.0-3.3)	8	2.0 (0.9-4.1)	9	2.1 (1.0-4.4)	5	3.5 (1.1-10.9)	1	3.2 (0.4-27.1)
development	70/165	1.8 (1.4-2.3)	45/52	1.7 (1.2-2.3)	39	2.2 (1.6-3.1)	20	2.5 (1.6-3.8)	12	3.3 (1.7-6.4)	12	7.2 (3.0-16.9)
remission	18/81	0.9 (0.5-1.4)	11/13	1.4 (0.7-2.5)	9	1.0 (0.5-1.9)	8	1.1 (0.4-2.7)	1	0.6 (0.1-4.6)	0	-
(d) lung function	< 88.5 %predicted n=1,901		≥ 88.5 %predicted n=2,082		< 88.5 %predicted		≥ 88.5 %predicted		< 88.5 %predicted		≥ 88.5 %predicted	
persistent	26/38	1.7 (1.1-2.7)	7/10	3.5 (1.7-7.6)	11	1.3 (0.7-2.5)	6	6.5 (2.8-15.2)	6	2.9 (1.0-8.2)	0	-
development	82/137	1.4 (1.1-1.8)	33/80	1.7 (1.2-2.4)	52	2.0 (1.5-2.8)	19	1.9 (1.1-3.0)	18	2.9 (1.6-5.2)	6	14.3 (4.5-44.7)
remission	11/47	0.6 (0.3-1.1)	18/47	1.5 (0.9-2.5)	6	0.8 (0.3-1.7)	8	1.5 (0.7-3.2)	1	0.6 (0.1-4.1)	0	-

[a]Excluding external causes of death, [b]Adjusted for age, gender, place of residence, smoking habits and BMI at baseline, [c]Adjusted for age, gender, place of residence, smoking habits, BMI and FEV_1 % predicted at baseline, [d]n=number of deaths in the group/ N=number of all subjects in the group.

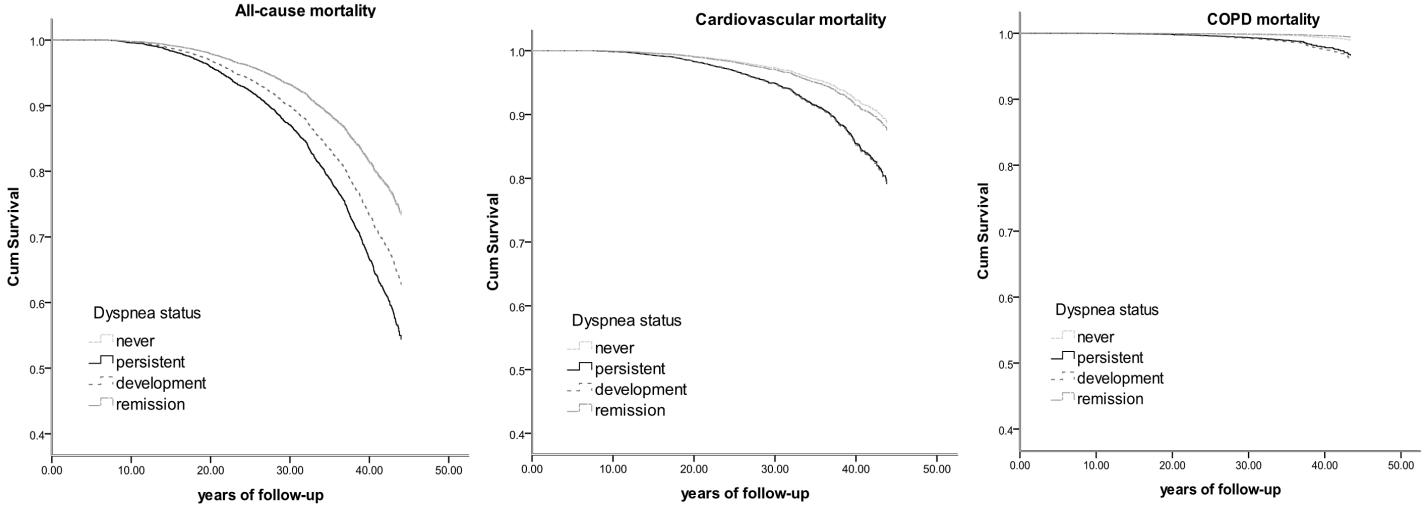


Figure 2 Cox proportional hazard regression survival curves showing the relation between changes in dyspnea status and all-cause, cardiovascular and COPD mortality

Changes in dyspnea status and all-cause, cardiovascular and COPD mortality

Subjects who reported remission of dyspnea had hazard ratios for all-cause, cardiovascular and COPD mortality comparable to the asymptomatic group. Subjects with persistent dyspnea and those who developed dyspnea had increased risks of all-cause, cardiovascular and COPD mortality compared to the asymptomatic group. Risks of all the investigated mortality outcomes associated with change in dyspnea were lower when corrected for lung function, however persistent dyspnea and development of dyspnea remained significant independent predictors of mortality (Table 4).

We additionally stratified the analysis by presence of dyspnea at baseline and compared development of dyspnea vs never dyspnea and remission vs persistent dyspnea, (Table S3). We found that development of dyspnea compared to never dyspnea was significantly associated with all-cause, cardiovascular and COPD mortality [1.5 (1.2-1.8), 1.9 (1.4-2.5) and 3.6 (2.1-6.1), respectively]. Furthermore analysis of dyspnea remission vs persistent showed that remission of dyspnea significantly reduces all-cause and cardiovascular mortality [0.4 (0.2-0.7) and 0.3 (0.1-0.7), respectively]. Remission was not significantly associated with COPD mortality [0.1 (0.0-1.8)]. When the analysis was further adjusted for severity at baseline (in 6 grades scale) the associations did not change and remained significant.

In survival curves there was a clear trend of increased mortality risks for subjects who developed dyspnea, whereas no significant difference between asymptomatic subjects and subjects who reported remission was observed (Figure 2). Effects of persistent dyspnea and development of dyspnea on cardiovascular and COPD

mortality were more pronounced in females and in subjects with BMI ≥ 25. In the analysis stratified according to age, risk for all-cause mortality associated with persistent dyspnea was higher in the younger than in the older subjects. The risk of cardiovascular and COPD mortality associated with development of dyspnea was higher in the older group. Furthermore, having dyspnea or developing this symptom was a significant risk factor for all-cause and cardiovascular mortality showing higher risks in the group with better lung function. Remission of dyspnea was not associated with mortality in any of the investigated groups.

DISCUSSION

This study has shown that dyspnea on exertion is associated with mortality in a severity-dependent manner in a general population cohort followed up for 43 years, even so after adjustment for potential risk factors. In particular, mortality due to cardiovascular disease shows a clear pattern: the higher the severity of dyspnea, the higher mortality risk. With regard to COPD mortality, only severe dyspnea was a significant predictor of mortality. It is interesting to note that the associations were present in both males and females. Furthermore, a novel finding in the current study is that remission of dyspnea normalises the risk of all-cause, cardiovascular and COPD mortality, whereas persistent dyspnea and development of dyspnea is a predictor of all-cause, cardiovascular and COPD mortality.

Our results support previous findings that dyspnea on exertion is related to all-cause mortality. In a previous study with 8 years of follow-up, overall mortality risk was significant only for grade III (dyspnea when walking with other people of the same age on level ground) and

over, and therefore this level has been suggested to use as a threshold (13). The same study showed that mortality risk significantly increases according to grade of dyspnea, being six-fold higher for the most severe grade of dyspnea (13). However, the authors of the study did not control for lung function level, which is a well-established predictor of mortality potentially modifying the results (2). In our study, dyspneic subjects at baseline had significantly worse lung function compared to subjects without dyspnea. Therefore, in order to exclude the possibility that pulmonary function drives these observed associations, we adjusted for lung function level. We showed that the hazard ratio of overall mortality for subjects who had reported dyspnea was higher when we did not adjust for lung function level but remained significant when lung function level was taken into account (HR: 1.3 and 1.5 for moderate and severe dyspnea, respectively). Interestingly, moderate and severe dyspnea were significant predictors of all-cause and cardiovascular mortality in subjects with lung function above the median of the population distribution. This emphasizes the importance of this self-reported symptom, even if an objective parameter of lung health (i.e. FEV_1) was in the healthy range. Dyspnea on exertion may be a manifestation of left ventricular hypertrophy and diastolic abnormalities (23). Myocardial ischemia may cause transient episodes of increased left ventricular end-diastolic as well as left atrial pressure with subsequent transient or long-term pulmonary engorgement (3). This may lead to increased airway resistance prior to any changes in pulmonary compliance. Contrary to severe forms of dyspnea observed in heart failure (such as orthopnea, paroxysmal nocturnal dyspnea or pulmonary edema) dyspnea on exertion can be an early indicator of coronary artery disease, even in the absence of either angina or electrocardiographic evidence of ischemia (24).

Dyspnea is a result of complex and multifactorial mechanisms, including abnormalities in the respiratory control system, neurochemical receptors, ventilation, respiratory muscles and gas exchange (1;25). In most conditions that cause breathlessness, several mechanisms are s involved (25). Dyspnea-associated mortality is probably often the consequence of both cardiac and respiratory disease (13). With regard to all-cause mortality, it has been suggested that respiratory ageing due to hypo-oxygenation could lead to functioning degeneration of other organs, thus leading to death as a result of the cumulative effect of successive pathologies (13).

We found that moderate and severe dyspnea significantly increase the risk of cardiovascular death. This finding supports the conclusion of Frostad's study, which showed that both moderate and severe dyspnea are predictors of mortality due to ischaemic heart disease and due to stroke (10). For COPD mortality, only severe dyspnea was a significant long-term predictor. This is in line with Nishimura's study that demonstrated that only dyspnea grade IV and V (i.e. severe) were significant predictors of mortality among COPD patients (6).

Dyspnea could be an indication of other associated conditions, such as respiratory and heart diseases. In the current study we found that the effects of persistent dyspnea and development of dyspnea on cardiovascular and COPD mortality were more pronounced in overweight and obese subjects (BMI ≥ 25). This may suggest a need to find efficacious weight-loss strategies for obese patients with respiratory symptoms, since weight loss can help reduce dyspnea, and perhaps underlying pathological conditions. Nonetheless, an accurate diagnosis in this group is important because dyspnea related to other mechanisms or diseases may require a different therapeutic strategy (26).

An important novel finding in our study is that remission of dyspnea normalises the risk of all-cause, cardiovascular and COPD mortality, since subjects who lost the symptom had hazard ratios comparable to the asymptomatic controls. Therefore we stress the importance of early detection of dyspnea as an important and simple indicator for mortality risk, which is cheap and applicable to most people (10). This early detection should of course be followed by identifying the underlying causative condition, and treatment that can be initiated. This is especially important since remission of dyspnea may also improve daily functioning and quality of life (19). According to our current findings, having persistent dyspnea or developing this symptom is an independent risk factor for all-cause, cardiovascular and COPD mortality. Since relatively few people who experience shortness of breath visit their family practitioners for this complaint, Huijnen *et al.* recommend that dyspnea should be added to the structured inventory of patients' problems in order to facilitate early detection (11). Pulmonary rehabilitation programs, that relieve dyspnea and reduce hospitalization (1), might be potential tools that subsequently can improve survival. This may be especially important in the light of our findings that persistent dyspnea or development of dyspnea increase all-cause mortality risk in subjects who were 46 years or younger at baseline, stressing the importance of early detection and intervention at relatively young age.

Strengths and weaknesses

The major strength of the current study is the longitudinal design. We were able to follow participants for over 40 years, which provided a unique wide time window for evaluating the risk of dyspnea severity and especially changes in dyspnea status on mortality.

A strength of our study is also the large number of subjects, sampled from the general population. Also the high follow-up rate should be mentioned, since 98.5% of the included subjects could be traced back. Other strengths of our study are that the groups for the changes in dyspnea status were created based on at least three surveys, and we investigated repeatability of the answers, treating subjects who had given inconsistent answers as a separate group. To investigate changes in dyspnea status we included only subjects with at least three available surveys, i.e.54% of the subjects from the first analysis, and automatically the number of deaths due to COPD (i.e. 99) decreased, leading to a low study power.

In conclusion, this study confirms an effect of dyspnea on all-cause and cause-specific mortality and additionally shows that dyspnea affects mortality in a severity-dependent manner. Moreover, this study is the first to show that remission of dyspnea normalises the risk of all-cause and cause-specific mortality, whereas persistent dyspnea or developing dyspnea is a risk factor for mortality. We additionally show that the effects of dyspnea on mortality are more pronounced in overweight and obese subjects, in the older group and in subjects with better lung function. The challenge for future research will be to identify the mechanisms underlying the development and remission of dyspnea. Only after the identification of these mechanisms, which may be driven by both genetic and environmental factors, a proper treatment of dyspnea can be established.

ACKNOWLEGMENTS

Author contributions: JM Vonk and HM Boezen planned the project and supervised its execution, analysis and writing. SM Figarska carried out the data analysis and drafted the paper. All authors contributed to the writing and preparation of the paper.

Funding: This work was supported by Stichtig Astma Bestrijding (grant number 2005/020) and Research Institute for Drug Exploration (GUIDE), University Medical Center Groningen, the Netherlands. The funding source had no role in the study design, in the collection, analysis, and interpretation of the data, in the writing of the report, and in the decision to submit the paper for publication.

Ethnical approval: The Committee on Human Subjects in Research of the University of Groningen reviewed the study and affirmed the safety of the protocol and study design and all participants gave their written informed consent.

Conflict of interest: none to declare.

REFERENCES

1. Dyspnea. Mechanisms, assessment, and management: a consensus statement. American Thoracic Society. Am J Respir Crit Care Med 1999 Jan;159(1):321-40.
2. Boezen HM, Rijcken B, Schouten JP, Postma DS. Breathlessness in elderly individuals is related to low lung function and reversibility of airway obstruction. Eur Respir J 1998 Oct;12(4):805-10.
3. Rosengren A, Wilhelmsen L. Respiratory symptoms and long-term risk of death from cardiovascular disease, cancer and other causes in Swedish men. Int J Epidemiol 1998 Dec;27(6):962-9.
4. Fletcher CM, Elmes PC, Fairbairn AS, Wood CH. The significance of respiratory symptoms and the diagnosis of chronic bronchitis in a working population. Br Med J 1959 Aug 29;2(5147):257-66.
5. Mahler DA. Mechanisms and measurement of dyspnea in chronic obstructive pulmonary disease. Proc Am Thorac Soc 2006 May;3(3):234-8.
6. Nishimura K, Izumi T, Tsukino M, Oga T. Dyspnea is a better predictor of 5-year survival than airway obstruction in patients with COPD. Chest 2002 May;121(5):1434-40.
7. Carpenter L, Beral V, Strachan D, Ebi-Kryston KL, Inskip H. Respiratory symptoms as predictors of 27 year mortality in a representative sample of British adults. BMJ 1989 Aug 5;299(6695):357-61.
8. Frostad A, Soyseth V, Andersen A, Gulsvik A. Respiratory symptoms as predictors of all-cause mortality in an urban community: a 30-year follow-up. J Intern Med 2006 May;259(5):520-9.
9. Frostad A, Soyseth V, Haldorsen T, Andersen A, Gulsvik A. Respiratory symptoms and 30 year mortality from obstructive lung disease and pneumonia. Thorax 2006 Nov;61(11):951-6.
10. Frostad A, Soyseth V, Haldorsen T, Andersen A, Gulsvik A. Respiratory symptoms and long-term cardiovascular mortality. Respir Med 2007 Nov;101(11):2289-96.
11. Huijnen B, van der Horst F, van Amelsvoort L, Wesseling G, Lansbergen M, Aarts P, et al. Dyspnea in elderly family practice patients. Occurrence, severity, quality of life and mortality over an 8-year period. Fam Pract 2006 Feb;23(1):34-9.
12. Stavem K, Sandvik L, Erikssen J. Breathlessness, phlegm and mortality: 26 years of follow-up in healthy middle-aged Norwegian men. J Intern Med 2006 Oct;260(4):332-42.
13. Tessier JF, Nejjari C, Letenneur L, Filleul L, Marty ML, Barberger Gateau P, et al. Dyspnea and 8-year mortality among elderly men and women: the PAQUID cohort study. Eur J Epidemiol 2001;17(3):223-9.

14. Ebi-Kryston KL. Respiratory symptoms and pulmonary function as predictors of 10-year mortality from respiratory disease, cardiovascular disease, and all causes in the Whitehall Study. J Clin Epidemiol 1988;41(3):251-60.

15. Knuiman MW, James AL, Divitini ML, Ryan G, Bartholomew HC, Musk AW. Lung function, respiratory symptoms, and mortality: results from the Busselton Health Study. Ann Epidemiol 1999 Jul;9(5):297-306.

16. Mannino DM, Buist AS, Petty TL, Enright PL, Redd SC. Lung function and mortality in the United States: data from the First National Health and Nutrition Examination Survey follow up study. Thorax 2003 May;58(5):388-93.

17. Schunemann HJ, Dorn J, Grant BJ, Winkelstein W, Jr., Trevisan M. Pulmonary function is a long-term predictor of mortality in the general population: 29-year follow-up of the Buffalo Health Study. Chest 2000 Sep;118(3):656-64.

18. Killian KJ, Watson R, Otis J, St Amand TA, O'Byrne PM. Symptom perception during acute bronchoconstriction. Am J Respir Crit Care Med 2000 Aug;162(2 Pt 1):490-6.

19. Voll-Aanerud M, Eagan TM, Wentzel-Larsen T, Gulsvik A, Bakke PS. Changes in respiratory symptoms and health-related quality of life. Chest 2007 Jun;131(6):1890-7.

20. Xu X, Rijcken B, Schouten JP, Weiss ST. Airways responsiveness and development and remission of chronic respiratory symptoms in adults. Lancet 1997 Nov 15;350(9089):1431-4.

21. Vestbo J, Knudsen KM, Rasmussen FV. Should we continue using questionnaires on breathlessness in epidemiologic surveys? Am Rev Respir Dis 1988 May;137(5):1114-8.

22. Gibson GJ. Obesity, respiratory function and breathlessness. Thorax 2000 Aug;55 Suppl 1:S41-S44.

23. Caidahl K, Eriksson H, Hartford M, Wikstrand J, Wallentin I, Svardsudd K. Dyspnoea of cardiac origin in 67 year old men: (1). Relation to systolic left ventricular function and wall stress. The study of men born in 1913. Br Heart J 1988 Mar;59(3):319-28.

24. Cook DG, Shaper AG. Breathlessness, angina pectoris and coronary artery disease. Am J Cardiol 1989 Apr 15;63(13):921-4.

25. Manning HL, Schwartzstein RM. Pathophysiology of dyspnea. N Engl J Med 1995 Dec 7;333(23):1547-53.

26. Poulain M, Doucet M, Major GC, Drapeau V, Series F, Boulet LP, et al. The effect of obesity on chronic respiratory diseases: pathophysiology and therapeutic strategies. CMAJ 2006 Apr 25;174(9):1293-9.

TableS 1 ICD-codes for the causes of death

ICD-version	ICD-7	ICD-8	ICD-9	ICD-10
Years of use	1965-1968	1969-1978	1979-1995	1996-2008
External causes of death	≥800	≥800	≥800	S, T, V, W, X, Y
Death due to cardiovascular disease	330-334, 400-416, 420-422, 430-434, 440-447, 450-456, 460-468, 782.4	390-398, 400-404, 410-414, 420-429, 430-438, 440-448 (excl.444.2), 450-458, 782.4	390-398, 401-405, 410-417, 420-438, 440-448, 451-459, 785.4	G45-G46, I00-I15, I20-I28, I30-I52, I60-I69, I70-I79, I80-I89, I95-I97, I98.2, I98.8, I99, M30-M31, N28.0, R02, R58
Death due to COPD	501, 502, 526, 527.1	490-492, 518	490-492, 494, 496	J40-J44, J47

Table S2 Hazard ratio (HR) with 95% confidence interval (CI) for all-cause mortality according to dyspnea original grades, only grade I a refernce

Dyspnea grades[a]	All-cause mortality			
	n/N[b]	HR[c] (95% CI)	Grades	HR[c] (95% CI)
I	1842/5529	reference	I	reference
II	624/1258	1.0 (1.0-1.2)	II (mild)	1.0 (1.0-1.2)
III	195/316	**1.3 (1.1-1.5)**	III (moderate)	**1.3 (1.2-1.6)**
IV	44/61	**1.5 (1.1-2.4)**	IV	
V	23/34	**1.6 (1.0-2.4)**	V (severe)	**1.5 (1.1-2.0)**
VI	24/31	1.4 (0.9-2.1)	VI	

[a] grade I as a reference, [b] n=number of deaths in the group/N=number of all subjects in the group, excluding those who died due to external causes, [c] Adjusted for age, gender, place of residence, smoking habits, BMI and FEV_1 % predicted at baseline

Table S3 Hazard ratio (HR) with 95% confidence interval (CI) for all-cause, cardiovascular mortality in stratification for dyspnea presence at baseline

Dyspnea presence at baseline	Dyspnea on exertion	All-cause mortality	CVD mortality	COPD mortality
		HR (95% CI)	HR (95% CI)	HR (95% CI)
No	Never	reference	reference	reference
	Development	**1.5 (1.2-1.8)**	**1.9 (1.4-2.5)**	**3.6 (2.1-6.1)**
Yes	Persistent	reference	reference	reference
	Remission	**0.4 (0.2-0.7)**	**0.3 (0.1-0.7)**	0.1 (0.0-1.8)
	Remission [a]	**0.4 (0.2-0.8)**	**0.3 (0.1-0.9)**	0.1 (0.0-1.7)

[a] additionally adjusted for dyspnea severity at baseline

Chapter 8

Summary, general discussion and future perspectives

Summary

Ageing is defined as a decline of physiological functions. Decline of physiological functions ultimately lead to the development of age-related diseases, such a cardiovascular disease (CVD) or Chronic Obstructive Pulmonary Disease (COPD), and in consequence to death (1;2). Humans age at different pace; they develop a disease and die at different ages. Among factors influencing human lifespan, genetics seem to play an important role, since the variation in human lifespan is heritable for approximately 30% (3). This thesis focuses on the genetic factors influencing human all-cause and cause-specific mortality, i.e. cardiovascular and COPD mortality in the general population.

In **Chapter 2** the role of the *Sirtuin 1* (*SIRT1*) gene in human lifespan is described. Sirtuins have become an important topic in biomedical research during the past decade and currently are under debate regarding their functions and as potential targets for therapeutic applications in humans (4). Mutations that increase activity of *Sir2* (silent information regulator 2) are associated with extended lifespan of yeast, fruit flies and worms (5-7). *SIRT1*, the human homolog of *Sir2*, that controls numerous physiological processes including the glucose metabolism, is considered a candidate gene for predicting variation in human lifespan. We studied 1,390 subjects from a general population-based cohort with 18 years of follow-up to investigate the associations between 3 SNPs tagging *SIRT1* and human survival. Our study provided new evidence in favor of a role of *SIRT1*, since carriers of the minor allele of rs12778366 had a significantly reduced mortality risk compared to the wild types indicated by a Hazard Ratio (HR) of 0.69 (95% CI 0.50 to 0.96; p=0.025). The direction of the effects were the same in females and males, never and ever smokers and in overweight/obese subjects (BMI ≥ 25). Previous studies indicate

that transgenic mice that overexpress *SIRT1* appear to have beneficial phenotypes that may be relevant in human health, including better glucose tolerance (8;9). In contrast to the positive effects of increased *SIRT1* activity, *SIRT1* deficiency impairs metabolism (10). Therefore, additionally in 535 male subjects, we investigated associations between *SIRT1* and glucose tolerance. We found that carriers of the minor allele of SNP rs12778366 had better glucose tolerance indicated by 0.34 mmol/l lower glucose levels compared to wild type subjects (p=0.03). When this association was further investigated in subjects with normal weight (n=249) and overweight/obese subjects (n=284) separately, we found significantly better glucose tolerance in overweight/obese carriers of the minor allele of rs12778366 (i.e. 0.60 mmol/l lower glucose levels (p=0.01)) compared to overweight/obese wild type subjects. In subjects with normal weight the same direction of effect was observed (0.18 mmol/l lower glucose levels), but it was not significant (p=0.50). Interestingly, a very recent study has shown that SIRT1 protein levels and SIRT1 activity, measured in the vasculature of Zucker rats, were greater in obese compared to lean animals. This suggests that the increased expression and activity of *SIRT1* may be a vascular adaptive mechanism in obesity (11), possibly preventing from hyperglycemia-induced vascular cell senescence (12). In the light of our findings, it is of future interest to study the functionality of rs12778366.

The study described in **Chapter 3** investigated SNPs in A Desintegrin and Metalloproteinase 33 (*ADAM33)* in relation to all-cause, COPD and cardiovascular mortality in the general population. The *ADAM33* gene was initially identified as an asthma susceptibility gene (13). However, since it has been linked to the pathophysiology of age-related diseases such as Chronic Obstructive Pulmonary Disease

(COPD) and atherosclerosis (14-18), it has become a plausible candidate gene for this thesis, as we were interested in a genetic determinant associated with more than one age-related disease. We selected four SNPs in *ADAM33* (Q_1, S_1, S_2 and T_2) previously linked to asthma, COPD, or accelerated decline in Forced Expiratory Volume in 1 second (FEV_1) for genotyping. Since we were interested in the independent effect of the SNPs, the analyses were adjusted for lung function level and other potential confounders. We found that individuals homozygous for the minor allele of SNP T_2 had an increased risk of all-cause and cardiovascular mortality compared to wild types: HR=3.6 (95% CI 2.0 to 6.7) and 3.4 (1.2 to 9.5) respectively. Individuals homozygous for the minor allele of Q_1, S_1, S_2 or T_2 had a significantly increased risk of COPD mortality. Importantly, in stratified analyses according to gender and smoking habits the risk of all-cause mortality associated with SNP T_2 did not change (females 3.5 (1.5 to 8.3), males 3.1 (1.2 to 7.6), never smokers 3.8 (0.9 to 16.3), ever smokers 3.6 (1.8 to 7.2)), which emphasizes the robustness of the association and indicates that SNP T_2 is a general risk factor for mortality. Moreover, SNP T_2 was associated with a reduced chance of survival to the age of 75, and therefore we believe that *ADAM33* may affect human lifespan. Our findings highlight the importance of *ADAM33* as a pleiotropic gene involved not only in pulmonary disease, but in cardiovascular disease as well. Future studies should focus on the functionality of the various SNPs in this gene to further unravel its role in ageing.

In the study described in **Chapter 4** we investigated the relation between the Nuclear Factor (Erythroid-derived 2)-Like 2 (*NFE2L2* or *NRF2*) gene polymorphisms and all-cause, cardiovascular and COPD mortality. *NFE2L2* is a master regulator of antioxidant-related genes and

genes controlling inflammatory responses, as well as of genes involved in tissue remodeling (19). Furthermore, the broad role of *NFE2L2* in age-related diseases has been indicated by its association with the development of atherosclerosis (20-23) and COPD (24;25). *NFE2L2* regulates transcription of enzymes involved in lipids homeostasis and previous studies, performed in mice, have shown that the expression of *NFE2L2* regulates the expression of lipogenic genes and affects lipid accumulation and deposition in aortic lesions (20). Therefore, we were additionally interested in *NFE2L2*'s relation with triglyceride and total cholesterol levels. Five SNPs tagging *NFE2L2* were selected and genotyped in 1,390 subjects. Carriers of the minor allele of rs13001694 had a significantly reduced risk of all-cause mortality compared to wild types: HR=0.8 (95% CI 0.6 to 1.0). Moreover, we found that carriers of the minor allele of rs2364723 had a significantly reduced risk of cardiovascular mortality: HR=0.5, (95% CI 0.3 to 0.7). This result remained in stratified analyses: females 0.4 (0.2 to 0.7), males 0.6 (0.3 to 0.9), never smokers 0.5 (0.2 to 1.1), ever smokers 0.5 (0.3 to 0.8). We also found that SNP rs2364723 was associated with lower triglyceride levels, which may in part explain its association with reduced cardiovascular mortality. We did not see a significant association of rs2364723 with cholesterol in our Vlagtwedde-Vlaardingen cohort, but this SNP was associated with a lower cholesterol level in the British 1958 Birth Cohort, which highlights the relevance of this SNP in cardiovascular disease. SNP rs2364723 is in high LD (r2=0.99) with the promoter polymorphism rs35652124 (-686A/G) in the Vlagtwedde-Vlaardingen cohort (26), suggesting a role in regulation of transcription. In this study we also found that carriers of the minor allele of rs1806649 had a markedly reduced COPD mortality: HR=0.3 (0.1 to 0.9). Interestingly, this SNP previously has been associated with reduced air pollution-induced asthma/COPD hospital admissions (27).

In **Chapter 5** a genome-wide association (GWA) study on all-cause and cardiovascular mortality is presented. We investigated Caucasian individuals in the Vlagtwedde-Vlaardingen, of whom blood samples were drawn in 1989/90 and the vital status was assessed on December 31 2008. During 18 years of follow-up, out of 1546 included subjects, 356 had died, of whom 154 due to cardiovascular disease. In both analyses we performed Cox regression on time to event, adjusted for age and sex. Genome-wide significance was set at $p<2.06x10^{-7}$ by dividing the classical significance threshold of $\alpha=0.05$ by the number of SNPs tested (n=242,926). We found SNP rs10237193 in *VWC2* to be associated with all-cause mortality at genome-wide significance level. This gene encodes a protein that is possibly involved in neural function and development and may have a role in cell adhesion. We also identified 20 SNPs associated with all-cause mortality at $p<5x10^{-5}$. Out of these, 10 are annotated to genes: *RASAL2* (3 SNPs), *SAMD4A, MSI2, MBTPS1, CSF1R, NAP5, GAP43, TNKS1BP1*.

For cardiovascular mortality we did not identify any genome-wide significant SNP, but we identified 10 SNPs associated at $p<5x10^{-5}$ and 4 of these were intragenic (*AMPH, LRMP, RASAL2, NELL1*). Using publicly available data (SCAN database, http://www.scandb.org) we found that some of the top hits in this study are expression quantitative trait loci (eQTLs; for example rs10237193 is associated with the expression of *MUTYH*, which encodes a DNA glycosylase involved in oxidative DNA damage repair).

Interestingly, the top associated SNPs in our GWA studies on all-cause and cardiovascular mortality were associated with other traits or diseases at $p\leq0.05$ in previously performed GWA studies (http://www.gwascentral.org). These traits include fasting plasma glucose, systolic blood pressure, BMI, serum cholesterol and lung function,

suggesting that the SNPs we found are predictors of all-cause and cardiovascular mortality via these intermediate phenotypes. Replication is needed to confirm our findings.

Chapter 6 describes the use of a Genetic Risk Score (GRS) for prediction of reduced lung function level in the general population. Forced Expiratory Volume in one second (FEV_1), measured by spirometry, reflects the physiological state of the airways and lungs (28), and reduced FEV_1 is a powerful indicator of future morbidity and mortality (29-32). FEV_1 is highly heritable, but SNPs in candidate genes when studied separately only confer modest risk and are of limited value in prediction (33). However, combining multiple loci with modest effects into a GRS might improve identification of subjects at risk for reduced FEV_1. This is of great importance since this may in turn prevent premature deaths. By testing associations between reduced lung function (i.e. FEV_1<80% predicted) and 74 SNPs in 24 candidate genes in 1,390 subjects from the Vlagtwedde-Vlaardingen cohort, 11 SNPs were selected to be included in an unweighted (uGRS) and weighted (wGRS) GRS. The results of this study showed that the GRS weighted by the odds ratio of each risk genotype, is a slightly better discrimination tool (area under the curve AUC=0.740) of the subjects with reduced lung function than the unweighted GRS that simply sums risk genotypes (AUC=0.735). This is in agreement with previous studies showing that a weighted GRS can capture considerably more genetic risk compared to an uGRS (33;34). Aggregating data from many markers into a single GRS variable allows genetic and biological information related to phenotype to be condensed into one dimension (34). However, so far none of the previous studies attempted to proof this statement by investigating the effect of the selected SNPs jointly, but not combined into one variable. Thus, we

additionally tested this joint effect of 11 SNPs and although it showed the highest discriminative power (AUC=0.749), based on the Akaike Information Criterion (AIC), it does not fit the data as good as the model with wGRS. Further studies, validating the discriminatory accuracy of the wGRS need to be undertaken.

In **Chapter 7** a study on the effects of dyspnea severity and changes in dyspnea status on all-cause and cause-specific mortality is described. The Vlagtwedde-Vlaardingen study started in 1965 and participants in this study were re-examined every three years until 1989/1990. This offered us the unique possibility to investigate the presence and severity of dyspnea during lifetime and their influence on all-cause and cause-specific mortality. In this study we included all 8,465 subjects of whom the vital status was assessed on December 31st, 2008. After 43 years of follow-up, 2,883 (39%) of 7,360 subjects examined for dyspnea severity had died; 1,386 (19%) due to cardiovascular disease and 267 (4%) due to Chronic Obstructive Pulmonary Disease (COPD). We found that subjects with moderate and severe dyspnea had increased all-cause and cardiovascular mortality in a severity-dependent manner, compared to asymptomatics. Moreover, we show that severe dyspnea was significantly associated with COPD mortality. Persistent dyspnea and dyspnea development were risk factors for all-cause, cardiovascular and COPD mortality. Additionally, effects of dyspnea on mortality were more pronounced in overweight/obese and older subjects, but also in subjects with better lung function. Respiratory symptoms are often transient, with a high remission rate (35) and dyspnea is likely to develop or remit during lifetime (36). The important finding in this study is that subjects who lost dyspnea had HR for all-cause and cause-specific mortality comparable to asymptomatics. Thus, this study is the first showing that dyspnea

remission normalizes mortality risk. Furthermore analysis of dyspnea remission vs persistent dyspnea showed that remission of dyspnea significantly reduces all-cause and cardiovascular mortality (HR=0.4 (95% CI 0.2 to 0.7) and 0.3 (0.1 to 0.7), respectively). When this analysis was further adjusted for dyspnea severity at baseline the associations did not change and remained significant. Additionally, our findings show that persistent dyspnea or development of dyspnea increase all-cause mortality risk especially in subjects who were 46 years or younger at baseline, which stresses the importance of early detection and intervention at a relatively young age. Since relatively few people who experience shortness of breath visit their general practitioners for this complaint, dyspnea should be added to the structured inventory of patients' problems in order to facilitate early detection (37).

Main findings

The chapters in this thesis describe research on SNPs from different genes and the effect of dyspnea in a general population-based cohort in relation to all-cause and cardiovascular and COPD mortality, and to lung function level.

The main conclusions of this thesis are:

1. A SNP in *SIRT1* is associated with reduced all-cause mortality risk and with better glucose tolerance.
2. SNPs in *ADAM33* are associated with all-cause, cardiovascular and COPD mortality, independent of their effect on lung function.
3. SNPs in *NFE2L2* are associated with all-cause, cardiovascular and COPD mortality and with triglyceride levels.

4. A SNP in *VWC2* is a genome-wide significant predictor of all-cause mortality and SNPs in genes *RASAL2, SAMD4A, MSI2, MBTPS1, CSF1R, NAP5, GAP43, TNKS1BP1* are associated with all-cause mortality at $p<5x10^{-5}$.
5. No genome-wide significant SNPs are associated with cardiovascular mortality, but SNPs in *AMPH, LRMP, RASAL2, NELL1* are associated with a $p<5x10^{-5}$.
6. Weighted GRS (wGRS) composed of 11 SNPs is a better tool to discriminate between subjects with normal and reduced lung function level than unweighted GRS (uGRS) and than 11 SNPs analyzed jointly but not combined in a score.
7. Dyspnea is a severity-dependent predictor of all-cause and cardiovascular mortality, but only severe dyspnea is associated with COPD mortality.
8. Persistent dyspnea or development of dyspnea increases all-cause mortality, but remission of dyspnea normalizes the mortality risk.

General discussion and future perspectives

All humans age and with ageing they develop age-related diseases. Among elderly people, co-morbidity is often observed, in which case age-related diseases are treated separately. When appropriate treatment is applied, death, a natural consequence of ageing, may be delayed. This would result in an extension of life and suffering from disease(s), whereas a perfect anti-ageing strategy would rely on identifying people prone to develop age-related diseases and postponing the onset of these diseases, thus, in practice, on increasing the healthy years of life. Deeper insight into the genetic background of age-related changes and diseases and the increased mortality will enrich the knowledge about ageing and may provide targets for effective interventions to live not only longer, but also in good health. In this project we were interested in all-cause mortality in the general population, but also in cause-specific mortality, i.e. cardiovascular and COPD mortality.

Genetic approaches

At the start of this project in 2009 there were some interesting candidate genes to be investigated, based on previous findings in animal models or their biological function. Additionally, the genome-wide association (GWA) studies era has developed giving the possibility to identify novel variants related to the trait in a hypothesis free manner. Since most single nucleotide polymorphisms (SNPs) identified by GWA studies fall in the 95% of non-coding region of the genome (38) this approach may reveal completely novel genes or regions that would not have been studied as candidate genes. Given the advantages and disadvantages of both approaches

(Table 1) we felt that both approaches could yield complementary information related to the genetics of healthy ageing and therefore we performed both types of studies.

Table 1 Comparison of candidate gene study and genome wide association study approaches

Candidate gene study	**Genome-wide association study**
Advantages	Advantages
• Focuses on genes with known biological function	• An unbiased picture of the genome
• Capable of identifying polymorphisms with low allele frequency	• May reveal unexpected SNPs associated with ageing. GWA studies present novel associations and complex traits can be explored
• A functional SNP that leads to alternation in protein function may be a relevant target in pharmacogenetics	• No prior hypothesis required
Disadvantages	Disadvantages
• Potentially miss important genes and may lead to limited explanation of variance in a complex and polygenic trait as ageing	• Some problems may arise regarding the discrepancy between type I errors (false positive results) and subsequently adjusted type II errors (false negative results) in detecting associated SNPs
• A priori knowledge of candidate gene is required	• Large number of samples required
	• Lack of information about gene function
	• Not capable of indentifying polymorphisms with low allele frequency e.g. rare variants

Genetic studies on ageing and longevity

In the research on human ageing different phenotypes are studied. There is a branch of studies on longevity, where the main goal is to uncover genetic factors contributing to extremely long lifespan (39-42). In the studies on healthy ageing, the phenotype is defined in different ways. For example, Zhang et al. defined healthy ageing as being 60 years or older and having a normal brain function and verbal fluency test, normal laboratory findings for hemogram, normal peripheral smear, urine, electrolytes chest X-ray, kidney, pulmonary function, echocardiography and ECG (43). Another study investigated time to develop a disease or time to death to identify risk factors for these events (44). Searching for genetic risk factors contributing to age-related phenotypes is promising, since it may provide novel therapeutic targets for the future. In this thesis on human ageing, using the data available in the Vlagtwedde-Vlaardingen cohort, the studied outcomes include all-cause mortality, cardiovascular and COPD mortality and reduced lung function level.

Candidate gene studies

It has been demonstrated that changes in single genes can significantly increase lifespan of model organisms, such as worms, fruit flies and mice (45-52). However, many of their human homologues failed to be replicated across different populations (49;53). Studies in model organisms have been performed under laboratory conditions, where temperature, presence of pathogens, food availability and population density are tightly controlled. In most cases these conditions poorly mimic the evolutionary niche in which the genes come to expression and it remains unknown

to what extent single mutations affecting lifespan in laboratory conditions would affect the lifespan in natural conditions. Thus, it is difficult to translate results from model organisms to humans and results obtained in model organisms should be interpreted with caution (54). In this thesis we were interested in two longevity genes that were identified in studies on model organisms: *SIRT1* and *NFE2L2*. We highlighted their relevance in humans by showing the associations between SNPs in these genes and human survival. In addition to the longevity genes confirmed in model organisms, pleiotropic genes involved in pathogenesis of more than one human age-related disease deserve more attention, since they may also influence survival in the population. The research on ageing should not necessarily aim to increase human lifespan, but extend the healthy years of life. In other words the research strategy should be focused on prevention of age-related diseases. For instance, we found that SNP T_2 in *ADAM33* is a risk factor for COPD and cardiovascular mortality which are both diseases that may lead to premature death (i.e. before the age of 75). This suggests that screening for this SNP, probably in conjunction with other SNPs in other genes, may identify subjects who are at risk for premature death. Eventually, in the future, for these subjects an individualized pharmacological therapy may be developed to prevent or treat a wide range of age-related diseases.

GWA Studies

In the last decade GWA studies became a popular and actually leading approach in genetic studies on complex diseases or traits and allowed the successful identification of hundreds of common susceptibility variants, even though they were able to explain only

a small fraction of genetic contribution leaving a problem of missing heritability (55).

To identify new loci related to human ageing several GWA studies were performed using different phenotypes. One of the investigated phenotypes was longevity. These studies could not find any new loci and merely confirmed the contribution of *APOE* (56-58) to longevity. Another GWA study outcome was the phenotype of survival to age of 90 years or older, but no genome-wide significant associations were found in this study (59). Interestingly, one study investigated the number of alleles increasing the risk for coronary artery disease, cancer, and type 2 diabetes, all loci identified previously in GWA studies. Surprisingly, it turned out that centenarians carry the same number of disease susceptibility alleles as young controls, thus longevity cannot be explained by the absence of age-related disease risk alleles (60). In the meta-analysis of nine GWA studies on all-cause mortality, genes highly expressed in the brain (*HECW2, HIP1, BIN2*), genes involved in neural development and function (KCNQ4, LMO4, NETO1) and autophagy (*ATG4C*) were highlighted (44). It is possible that some of the variations affect the trait not in the, commonly assumed, additive mode, but in a dominant or recessive mode and therefore were not detected in GWA studies performed so far. In one of the chapters we describe GWA studies performed in the Vlagtwedde-Vlaardingen cohort and show that SNPs in *VWC2, RASAL2, SAMD4A, MSI2, MBTPS1, CSF1R, NAP5, GAP43, TNKS1BP1* are associated with all-cause mortality and SNPs in *AMPH, LRMP, RASAL2, NELL1* are associated with cardiovascular mortality. Nevertheless, our findings need to be replicated in an independent cohort and functional studies for the associated SNPs could give more clues about their role in ageing.

Study design and methodological issues

Human studies of ageing face numerous theoretical, methodological and logistical challenges, since the determinants of lifespan are very complex (53).

Most of the genetic association studies on longevity have been performed in case-control settings where allele frequencies between centenarians and average-aged subjects were compared (61). In such a study the suitability of the matched young controls is of course questionable since they come from a different birth cohort and the environment influencing cases and controls may be different. Furthermore, it is not possible to predict who and how many of these young controls will become centenarians. Longitudinal studies, where a cohort of individuals is followed over time, like the Vlagtwedde-Vlaardingen cohort, are less prone than case-control studies to these cohort effects. The Vlagtwedde-Vlaardingen study is a general population-based cohort, in which the vital status of all 8,465 participants was assessed and the median age at death was 76 years. No selection bias toward centenarians is present in Vlagtwedde-Vlaardingen since the age of death is a continuous outcome variable here. In the described studies we assessed the vital status of subjects participating in the survey in 1989/90 after 18 years of follow-up. Studies focused on the overall survival in general population-based cohorts may provide results that are better generalizable to and realistic for subjects with an average lifespan.

Another methodological issue is the coverage of genetic variance. In the candidate gene study approach, the SNPs selected for testing (i.e. the tagging SNPs) should cover all the genetic variation in

a gene, otherwise associations can be missed. Additionally, based on prior knowledge selection of SNPs, which may be causal variants for disease itself, is possible. In GWA studies commercially available microarray platforms with hundreds of thousands SNPs providing a very good coverage of the whole genome are usually used. Thus, SNP may be associated with disease, only because of the close association (LD) with the real, causal variant. This is also the case in candidate gene studies, if variants with unknown function are studied.

When multiple tests are performed associations may be found due to chance. A possible strategy to deal with potential false positive findings are multiple testing corrections. In the candidate studies described in this thesis the correction for multiple testing has not been made. A policy of not adjusting for multiple testing will decrease the likehood of type 2 errors and therefore potentially useful observations will not be prematurely discarded (62). Our hypotheses have been stated a priori, based on the previous evidence indicating roles of *SIRT1*, *ADAM33* and *NFE2L2* in longevity or in the pathophysiology of COPD and atherosclerosis, thus correcting for multiple testing may be too conservative and should be avoided (63). Regarding the GWA studies described in this thesis we used Bonferroni correction, thus the threshold p-value for a genome-wide significance was calculated by dividing $\alpha=0.05$ by number of test performed (n=242,925) and determined to be 2.06×10^{-7}.

A very good method to provide more evidence that observed associations are true findings is a replication of the results in an independent cohort. In some cases, due to differences in study design, sample size, genotyping errors, population stratifications or different environmental modifiers, results cannot be replicated, but the lack of replication in a different study population does not

necessarily refute the original reported association (64). Another solution to strengthen findings is a systematic meta-analysis, a quantitative approach for combining the results on the same outcome from different study cohorts, and this allows estimation of population-wide effects of genetic risk factors in human disease (65;66). Furthermore, ideally, negative findings require replication like positive ones, to make sure that a gene is not discarded as a further candidate for testing due to, for example a poor study design or methodological issues.

Future perspectives

Ageing, age-related diseases, and death are consequences of multiple cellular and physiological processes, and therefore likely under the regulation of a wide and diverse set of genes and the whole set of them is very difficult to be discovered.

Functional studies

The candidate genes from our studies still need more investigation regarding their functionality in humans. For instance, we have found that polymorphisms in *ADAM33* are associated with all-cause, COPD and cardiovascular mortality in the general population and even with premature death in humans. A study in mice has provided suggestive evidence that overexpression rather than underexpression of the ADAM33 protein may contribute to morbidity and in turn to mortality events. Thus, in this light the next research question is whether overexpression of *ADAM33* leads to the senescence phenotype. This may be tested using the cellular markers of senescence (Ki67, p21 (CDKN1A), γH2AX, and Sen-β-Gal)

in human endothelium where *ADAM33* is highly expressed. Furthermore, the cellular senescence may be evaluated in other human cells in vitro given that *ADAM33* is broadly expressed in the human organism. If indeed overexpression of *ADAM33* is contributing to cellular senescence, next anti-inflammatory drugs may be tested to check whether overexpression can be diminished and the senescent phenotype can be delayed. Functional studies should be performed for the SNPs identified in GWA studies even for the top hits which are not genome-wide significant. For instance, Malovini et al. in a GWA study performed in long-living individuals from Southern Italy pointed out the calcium/calmodulin-dependent protein kinase IV (*CAMKIV*) (67). Although the top SNP located in *CAMKIV* did not reach genome-wide significance (p-value=$2.9x10^{-5}$), they highlighted its relevance by showing its association with the expression of *CAMKIV*. Moreover, in the same study they showed that *CAMKIV* activates the proteins involved in cell survival: *AKT, SIRT1*, and *FOXO3A* (67). This is an example how further studies may indicate functionally relevant loci, even if they do not meet the genome-wide significance threshold.

Pleiotropic candidate genes

Cardiovascular disease and COPD, being age-related diseases and the main leading causes of death worldwide, need to be studied simultaneously to uncover their shared genetic background. This may provide novel targets for possible early detection and early intervention to prevent both diseases simultaneously. This could in the future decrease the cost of treating these diseases separately and increase the healthy years of life. This can be done by testing candidate genes involved in the pathogenesis of both diseases.

Also genes with yet unknown function and their pleiotropic effects should be assessed comprehensively in mouse models (68). Such a search may provide targets for therapies, for instance with use of small molecules that activate or deactivate these genes or their protein products, which would slow down the ageing process and extend human lifespan.

Sequencing to identify rare variants

Common SNPs that occur in the population with a minor allele frequency (MAF) above 5% have not accounted for the missing heritability of human lifespan. Therefore, rare variants (MAF<1%) are often suggested to potentially explain the missing heritability. Rare variants have for example been identified to contribute to several complex psychological disorders (69). To discover genes related to ageing, especially to discover such rare variants in relevant genes, whole-genome sequencing technologies are needed. Nowadays next generation sequencing platforms allow sequencing of entire genomes. Although it is still costly, widespread use of sequencing is expected in the next few years both in research and clinics (70). This approach could be promising for studies in families, where siblings of long living subjects have not only lower prevalence of morbidity (71;72), but also a reduced mortality (73;74). Such an analysis may allow determination of important, rare variants protecting from mortality due to common age-related diseases. Alternatively, studies in families with a history of co-morbidities and early mortality (i.e. mortality below the average age) may be performed to identify causal variants leading to increased physiological decline with ageing.

Epigenetics

Since the identification of genetic variants does not fully explain the heritability of human lifespan, and is not able to provide evidence for causative genetic effects, also an epigenetic contribution to this causality should be taken into account. Genome-wide epigenetic studies, i.e. looking at DNA methylation, histone marks, nucleosomal remodeling and noncoding RNAs can provide new insights into the biology of ageing. It would be specifically interesting to investigate changes in epigenetic markers over time in longitudinal cohorts (75). However, the mentioned modifications affecting the genomic transcription machinery are gained and lost at specified rates and they may contribute to disease risk, but might not explain the missing heritability (76).

Gene-gene and gene-environment interactions

Ageing is multifactorial and in this sense may be influenced by many genes interacting with each other, what is called epistasis. SNPs in a gene may modify the effect of SNPs in other genes, and these gene-gene interactions may enhance the risk for a particular phenotype in an additive or multiplicative way, or the negative effect of one SNP may be neutralized by the positive effect of another SNP. Furthermore, there is evidence showing that the environment, for instance nutrition, has an impact on the expression of genes related to lifespan, and the best example is the effect of calorie restriction on induced *SIRT1* expression in model animals (77;78). It emphasizes the importance of gene-environment interactions and the likeliness they occur in humans as well. In humans, next to the suggested impact of dietary intake, smoking, an established factor influencing human lifespan itself, may be an important environmental factor that changes the effects of SNPs

on lifespan. Thus, to better understand ageing, gene-gene and gene-environment interactions should also be studied.

Final conclusions

Findings in this thesis support a role of *SIRT1*, *ADAM33* and *NFE2L2* as candidate genes related to human ageing in the general population. Candidate gene studies, including those described in this thesis, allow better understanding of biological pathways involved in human ageing and survival. Nonetheless, probably due to the complexity of genetics determining human ageing and lifespan, results from ageing/longevity candidate gene studies in humans were hardly replicated in independent cohorts. In this thesis, scanning the whole genome of subjects from the general population, we found an association between *VWC2* and all-cause mortality. This finding needs to be replicated in other cohorts and ideally validated in a functional study on its biological mechanism. In general, GWA studies indeed identified new susceptibility loci for complex diseases, but explained only a very small part of the missing heritability (55). Perhaps some of the missing heritability can be identified by the use of new whole genome sequencing techniques, recently becoming less expensive and promising tools to discover new rare variants with strong effects on complex traits such a human ageing (70).

Phenotypic variance is traditionally a result of genetic and environmental components. The genetic factors can affect phenotypes in additive, dominant or recessive manners and ideally all these models should be considered for testing. On top of that there are interaction effects between genes and between genes and environment, which should be investigated to disentangle the complex associations underlying human ageing and lifespan.

In conclusion, genetic determinants of human ageing consequently affecting human lifespan still remain largely unknown. It is of great interest to know who is at risk for progressive age-related diseases, since this would enable intervention at a very early stage of disease. The advances in high-throughput sequencing technology will probably contribute to the identification of yet unknown or very rare mutations predisposing for development of age-related diseases and/or poor survival. Hence results obtained from sequencing may provide targets for an accurate prediction of age-related changes/ diseases at an individual level and eventually may be used for personalized anti-ageing therapy, all together potentially extending healthy years of life in humans.

REFERENCES

1. Flatt T, Schmidt PS. Integrating evolutionary and molecular genetics of aging. Biochim Biophys Acta 2009 Oct;1790(10):951-62.
2. Tosato M, Zamboni V, Ferrini A, Cesari M. The aging process and potential interventions to extend life expectancy. Clin Interv Aging 2007;2(3):401-12.
3. 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 Mar;97(3):319-23.
4. Accili D, de Cabo R, Sinclair DA. An unSIRTain role in longevity. Nat Med 2011;17(11):1350-1.
5. Kaeberlein M, McVey M, Guarente L. The SIR2/3/4 complex and SIR2 alone promote longevity in Saccharomyces cerevisiae by two different mechanisms. Genes Dev 1999 Oct 1;13(19):2570-80.
6. Rogina B, Helfand SL. Sir2 mediates longevity in the fly through a pathway related to calorie restriction. Proc Natl Acad Sci U S A 2004 Nov 9;101(45):15998-6003.
7. Tissenbaum HA, Guarente L. Increased dosage of a sir-2 gene extends lifespan in Caenorhabditis elegans. Nature 2001 Mar 8;410(6825):227-30.
8. Bordone L, Cohen D, Robinson A, Motta MC, van VE, Czopik A, et al. SIRT1 transgenic mice show phenotypes resembling calorie restriction. Aging Cell 2007 Dec;6(6):759-67.
9. Moynihan KA, Grimm AA, Plueger MM, Bernal-Mizrachi E, Ford E, Cras-Meneur C, et al. Increased dosage of mammalian Sir2 in pancreatic beta cells enhances glucose-stimulated insulin secretion in mice. Cell Metab 2005 Aug;2(2):105-17.
10. Baur JA, Ungvari Z, Minor RK, Le Couteur DG, de CR. Are sirtuins viable targets for improving healthspan and lifespan? Nat Rev Drug Discov 2012;11(6):443-61.
11. Tajbakhsh N, Sokoya EM. Sirtuin 1 is upregulated in young obese Zucker rat cerebral arteries. Eur J Pharmacol 2013 Oct 8.
12. Orimo M, Minamino T, Miyauchi H, Tateno K, Okada S, Moriya J, et al. Protective role of SIRT1 in diabetic vascular dysfunction. Arterioscler Thromb Vasc Biol 2009 Jun;29(6):889-94.
13. Van Eerdewegh P, Little RD, Dupuis J, Del Mastro RG, Falls K, Simon J, et al. Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. Nature 2002 Jul 25;418(6896):426-30.
14. Aierken H, Wushouer Q, Shayhidin E, Jing W, Wufuer D. Polymorphisms of the ADAM33 gene and Chronic Obstructive Pulmonary Disease Risk: a Meta-Analysis. Clin Respir J 2013 Aug 1.

15. Gosman MM, Boezen HM, van Diemen CC, Snoeck-Stroband JB, Lapperre TS, Hiemstra PS, et al. A disintegrin and metalloprotease 33 and chronic obstructive pulmonary disease pathophysiology. Thorax 2007 Mar;62(3):242-7.

16. Holloway JW, Laxton RC, Rose-Zerilli MJ, Holloway JA, Andrews AL, Riaz Z, et al. ADAM33 expression in atherosclerotic lesions and relationship of ADAM33 gene variation with atherosclerosis. Atherosclerosis 2010 Feb 24.

17. Sadeghnejad A, Ohar JA, Zheng SL, Sterling DA, Hawkins GA, Meyers DA, et al. Adam33 polymorphisms are associated with COPD and lung function in long-term tobacco smokers. Respir Res 2009;10:21.

18. van Diemen CC, Postma DS, Vonk JM, Bruinenberg M, Schouten JP, Boezen HM. A disintegrin and metalloprotease 33 polymorphisms and lung function decline in the general population. Am J Respir Crit Care Med 2005 Aug 1;172(3):329-33.

19. Hybertson BM, Gao B, Bose SK, McCord JM. Oxidative stress in health and disease: the therapeutic potential of Nrf2 activation. Mol Aspects Med 2011 Aug;32(4-6):234-46.

20. Barajas B, Che N, Yin F, Rowshanrad A, Orozco LD, Gong KW, et al. NF-E2-related factor 2 promotes atherosclerosis by effects on plasma lipoproteins and cholesterol transport that overshadow antioxidant protection. Arterioscler Thromb Vasc Biol 2011 Jan;31(1):58-66.

21. Freigang S, Ampenberger F, Spohn G, Heer S, Shamshiev AT, Kisielow J, et al. Nrf2 is essential for cholesterol crystal-induced inflammasome activation and exacerbation of atherosclerosis. Eur J Immunol 2011 Jul;41(7):2040-51.

22. Zakkar M, Van der Heiden K, Luong lA, Chaudhury H, Cuhlmann S, Hamdulay SS, et al. Activation of Nrf2 in endothelial cells protects arteries from exhibiting a proinflammatory state. Arterioscler Thromb Vasc Biol 2009 Nov;29(11):1851-7.

23. Sussan TE, Jun J, Thimmulappa R, Bedja D, Antero M, Gabrielson KL, et al. Disruption of Nrf2, a key inducer of antioxidant defenses, attenuates ApoE-mediated atherosclerosis in mice. PLoS One 2008;3(11):e3791.

24. Malhotra D, Thimmulappa R, Navas-Acien A, Sandford A, Elliott M, Singh A, et al. Decline in NRF2-regulated antioxidants in chronic obstructive pulmonary disease lungs due to loss of its positive regulator, DJ-1. Am J Respir Crit Care Med 2008 Sep 15;178(6):592-604.

25. Sandford AJ, Malhotra D, Boezen HM, Siedlinski M, Postma DS, Wong V, et al. NFE2L2 pathway polymorphisms and lung function decline in chronic obstructive pulmonary disease. Physiol Genomics 2012 Aug 1;44(15):754-63.

26. Siedlinski M, Postma DS, Boer JM, van der Steege G, Schouten JP, Smit HA, et al. Level and course of FEV1 in relation to polymorphisms in NFE2L2 and KEAP1 in the general population. Respir Res 2009;10:73.

27. Canova C, Dunster C, Kelly FJ, Minelli C, Shah PL, Caneja C, et al. PM10-induced hospital admissions for asthma and chronic obstructive pulmonary disease: the modifying effect of individual characteristics. Epidemiology 2012 Jul;23(4):607-15.

28. Soler Artigas M, Loth DW, Wain LV, Gharib SA, Obeidat M, Tang W, et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. Nat Genet 2011 Nov;43(11):1082-90.

29. Hole DJ, Watt GC, Davey-Smith G, Hart CL, Gillis CR, Hawthorne VM. Impaired lung function and mortality risk in men and women: findings from the Renfrew and Paisley prospective population study. BMJ 1996 Sep 21;313(7059):711-5.

30. Young RP, Hopkins R, Eaton TE. Forced expiratory volume in one second: not just a lung function test but a marker of premature death from all causes. Eur Respir J 2007 Oct;30(4):616-22.

31. Knuiman MW, James AL, Divitini ML, Ryan G, Bartholomew HC, Musk AW. Lung function, respiratory symptoms, and mortality: results from the Busselton Health Study. Ann Epidemiol 1999 Jul;9(5):297-306.

32. Schunemann HJ, Dorn J, Grant BJ, Winkelstein W, Jr., Trevisan M. Pulmonary function is a long-term predictor of mortality in the general population: 29-year follow-up of the Buffalo Health Study. Chest 2000 Sep;118(3):656-64.

33. Chen H, Poon A, Yeung C, Helms C, Pons J, Bowcock AM, et al. A genetic risk score combining ten psoriasis risk loci improves disease prediction. PLoS One 2011;6(4):e19454.

34. Piccolo SR, Abo RP, Allen-Brady K, Camp NJ, Knight S, Anderson JL, et al. Evaluation of genetic risk scores for lipid levels using genome-wide markers in the Framingham Heart Study. BMC Proc 2009;3 Suppl 7:S46.

35. Voll-Aanerud M, Eagan TM, Wentzel-Larsen T, Gulsvik A, Bakke PS. Changes in respiratory symptoms and health-related quality of life. Chest 2007 Jun;131(6):1890-7.

36. Xu X, Rijcken B, Schouten JP, Weiss ST. Airways responsiveness and development and remission of chronic respiratory symptoms in adults. Lancet 1997 Nov 15;350(9089):1431-4.

37. Huijnen B, van der Horst F, van Amelsvoort L, Wesseling G, Lansbergen M, Aarts P, et al. Dyspnea in elderly family practice patients. Occurrence, severity, quality of life and mortality over an 8-year period. Fam Pract 2006 Feb;23(1):34-9.

38. Mooney S. Bioinformatics approaches and resources for single nucleotide polymorphism functional analysis. Brief Bioinform 2005 Mar;6(1):44-56.

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 Sep 16;105(37):13987-92.

40. Frisoni GB, Louhija J, Geroldi C, Trabucchi M. Longevity and the epsilon2 allele of apolipoprotein E: the Finnish Centenarians Study. J Gerontol A Biol Sci Med Sci 2001 Feb;56(2):M75-M78.

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 Feb 24;106(8):2700-5.

42. 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 Apr;12(2):95-104.

43. Zhang WG, Bai XJ, Chen XM. SIRT1 variants are associated with aging in a healthy Han Chinese population. Clin Chim Acta 2010 Nov 11;411(21-22):1679-83.

44. Walter S, Atzmon G, Demerath EW, Garcia ME, Kaplan RC, Kumari M, et al. A genome-wide association study of aging. Neurobiol Aging 2011 Nov;32(11):2109-28.

45. Bartke A, Brown-Borg H. Life extension in the dwarf mouse. Curr Top Dev Biol 2004;63:189-225.

46. Friedman DB, Johnson TE. A mutation in the age-1 gene in Caenorhabditis elegans lengthens life and reduces hermaphrodite fertility. Genetics 1988 Jan;118(1):75-86.

47. 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 Jan 9;421(6919):182-7.

48. Kaeberlein M, McVey M, Guarente L. The SIR2/3/4 complex and SIR2 alone promote longevity in Saccharomyces cerevisiae by two different mechanisms. Genes Dev 1999 Oct 1;13(19):2570-80.

49. Kenyon C, Chang J, Gensch E, Rudner A, Tabtiang R. A C. elegans mutant that lives twice as long as wild type. Nature 1993 Dec 2;366(6454):461-4.

50. 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 Nov 14;278(5341):1319-22.

51. Schriner SE, Linford NJ, Martin GM, Treuting P, Ogburn CE, Emond M, et al. Extension of murine life span by overexpression of catalase targeted to mitochondria. Science 2005 Jun 24;308(5730):1909-11.

52. Tissenbaum HA, Guarente L. Increased dosage of a sir-2 gene extends lifespan in Caenorhabditis elegans. Nature 2001 Mar 8;410(6825):227-30.

53. Christensen K, Johnson TE, Vaupel JW. The quest for genetic determinants of human longevity: challenges and insights. Nat Rev Genet 2006 Jun;7(6):436-48.

54. Kuningas M, Mooijaart SP, van HD, Zwaan BJ, Slagboom PE, Westendorp RG. Genes encoding longevity: from model organisms to humans. Aging Cell 2008 Mar;7(2):270-80.

55. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature 2009 Oct 8;461(7265):747-53.

56. 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 Aug;10(4):686-98.

57. 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 Jun;132(6-7):324-30.

58. 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(1):e29848.

59. 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 May;65(5):478-87.

60. 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 Oct 19;107(42):18046-9.

61. Tan Q, Kruse TA, Christensen K. Design and analysis in genetic studies of human ageing and longevity. Ageing Res Rev 2006 Nov;5(4):371-87.

62. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology 1990 Jan;1(1):43-6.

63. Streiner DL, Norman GR. Correction for multiple testing: is there a resolution? Chest 2011 Jul;140(1):16-8.

64. Ioannidis JP, Boffetta P, Little J, O'Brien TR, Uitterlinden AG, Vineis P, et al. Assessment of cumulative evidence on genetic associations: interim guidelines. Int J Epidemiol 2008 Feb;37(1):120-32.

65. Evangelou E, Ioannidis JP. Meta-analysis methods for genome-wide association studies and beyond. Nat Rev Genet 2013 Jun;14(6):379-89.

66. Kavvoura FK, Ioannidis JP. Methods for meta-analysis in genetic association studies: a review of their potential and pitfalls. Hum Genet 2008 Feb;123(1):1-14.

67. 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 Jun;14(3):283-91.

68. Brown SD, Moore MW. Towards an encyclopaedia of mammalian gene function: the International Mouse Phenotyping Consortium. Dis Model Mech 2012 May;5(3):289-92.

69. Gibson G. Rare and common variants: twenty arguments. Nat Rev Genet 2011 Feb;13(2):135-45.

70. Matullo G, Di GC, Guarrera S. Next generation sequencing and rare genetic variants: from human population studies to medical genetics. Environ Mol Mutagen 2013 Aug;54(7):518-32.

71. Terry DF, Wilcox M, McCormick MA, Lawler E, Perls TT. Cardiovascular advantages among the offspring of centenarians. J Gerontol A Biol Sci Med Sci 2003 May;58(5):M425-M431.

72. Westendorp RG, van HD, 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 Sep;57(9):1634-7.

73. Perls TT, Wilmoth J, Levenson R, Drinkwater M, Cohen M, Bogan H, et al. Life-long sustained mortality advantage of siblings of centenarians. Proc Natl Acad Sci U S A 2002 Jun 11;99(12):8442-7.

74. 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 Jan;14(1):79-84.

75. 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 Jan 12;366(1561):35-42.

76. Slatkin M. Epigenetic inheritance and the missing heritability problem. Genetics 2009 Jul;182(3):845-50.

77. Shinmura K, Tamaki K, Bolli R. Impact of 6-mo caloric restriction on myocardial ischemic tolerance: possible involvement of nitric oxide-dependent increase in nuclear Sirt1. Am J Physiol Heart Circ Physiol 2008 Dec;295(6):H2348-H2355.

78. Cohen HY, Miller C, Bitterman KJ, Wall NR, Hekking B, Kessler B, et al. Calorie restriction promotes mammalian cell survival by inducing the SIRT1 deacetylase. Science 2004 Jul 16;305(5682):390-2.

Samenvatting

Veroudering gaat samen met een afname van de fysiologische functies. Deze afname van fysiologische functies kan leiden tot de ontwikkeling van leeftijdsgebonden ziekten, zoals cardiovasculaire ziekte (CVD) of chronische obstructieve longziekte (COPD), en uiteindelijk tot de dood. Er zijn grote verschillen in het tempo van veroudering tussen mensen; mensen worden ziek en sterven op verschillende leeftijden. Genetica lijkt een belangrijke rol te spelen in het tempo van veroudering. Uit eerder onderzoek is gebleken dat de variatie in de menselijke levensduur voor ongeveer 30% erfelijk bepaald is. Dit proefschrift richt zich op de genetische factoren die geassocieerd zijn met sterfte in zijn algemeenheid en met sterfte door specifieke oorzaken (cardiovasculaire en COPD-gerelateerde sterfte) in de algemene bevolking.

In **hoofdstuk 2** wordt de rol van het *Sirtuin 1* (*SIRT1*) gen in relatie tot de menselijke levensduur beschreven. Mutaties die de activiteit van *Sir2* (silent information regulator 2) verhogen zijn geassocieerd met een langere levensduur van gisten, fruitvliegen en wormen. De menselijke homoloog van *Sir2*, *SIRT1*, controleert talrijke fysiologische processen waaronder het glucosemetabolisme. Daarom zou *SIRT1* een goede voorspeller kunnen zijn voor de variatie in de menselijke levensduur. We vonden dat een bepaalde mutatie in *SIRT1* geassocieerd is met een verlaagd risico op sterfte (en dus met een langere levensduur) en met een betere glucosetolerantie.

In **hoofdstuk 3** wordt onderzoek naar genetische variatie in het A Desintegrin and Metalloproteinase 33 (*ADAM33*) gen in relatie tot algemene sterfte en oorzaakspecifieke sterfte (COPD en cardiovasculaire sterfte) in de algemene bevolking beschreven. Het *ADAM33* gen werd aanvankelijk geïdentificeerd als een astma-gevoeligheidsgen. Echter, omdat uit eerder onderzoek is gebleken

dat het *ADAM33* gen ook geassocieerd is met chronische obstructieve longziekte (COPD) en atherosclerose, is het een plausibel kandidaatgen voor ons onderzoek omdat we geïnteresseerd zijn in genen die geassocieerd zijn met meerdere leeftijdsgerelateerde aandoeningen. We vonden dat genetische varianten in *ADAM33* geassocieerd zijn met algemene, cardiovasculaire en COPD-gerelateerde sterfte, onafhankelijk van hun associatie met longfunctie.

In **hoofdstuk 4** onderzochten we de relatie tussen genetische variatie in het Nucleaire Factor (Erythroid-afgeleid 2)-Like 2 (*NFE2L2* of *NRF2*) gen en algemene sterfte, en cardiovasculaire en COPD-gerelateerde sterfte. *NFE2L2* reguleert de transcriptie van genen die betrokken zijn bij het handhaven van de oxidatie-antioxidatie balans, bij ontstekingsreacties of bij weefsel remodellering. De veelomvattende rol van *NFE2L2* in leeftijdsgerelateerde aandoeningen blijkt door de eerder gevonden associaties tussen genetische variatie in *NFE2L2* en de ontwikkeling van atherosclerose en COPD. Daarnaast reguleert *NFE2L2* ook de transcriptie van enzymen die betrokken zijn bij het handhaven van de vetbalans (lipiden homeostase), en eerdere studies in muizen hebben aangetoond dat de expressie van *NFE2L2* de expressie van lipogene genen beïnvloedt en een effect heeft op lipide ophoping en afzetting in aorta laesies. In ons onderzoek vonden we dat genetische variatie in *NFE2L2* geassocieerd is met algemene, cardiovasculaire en COPD-gerelateerde sterfte en met triglyceride niveaus in het bloed.

In **hoofdstuk 5** hebben we een genoom-brede associatie (GWA) studie in relatie tot algemene en cardiovasculaire sterfte gedaan. We onderzochten 1546 blanke individuen in de Vlagtwedde-Vlaardingen studie, van wie bloedmonsters zijn verzameld in 1989/90. Gedurende 18 jaar follow-up, zijn er 356 van de 1546 geïncludeerde

personen overleden, van wie 154 als gevolg van cardiovasculaire aandoeningen. SNP rs10237193 in het gen *VWC2* was geassocieerd met algemene sterfte met een p-waarde kleiner dan de genoom-brede significantie ($p<2.06x10^{-7}$). Twintig SNPs waren geassocieerd met algemene sterfte ($p< 5x10^{-5}$), van welke 10 SNPs in of rondom genen liggen: *RASAL2* (3 SNPs), *SAMD4A, MSI2, MBTPS1, CSF1R, NAP5, GAP43, TNKS1BP1*. In relatie tot cardiovasculaire sterfte hebben we geen genoom-breed significant geassocieerde SNPs gevonden. Wel vonden we 10 SNPs geassocieerd met p-waarden $<5x10^{-5}$ en 4 daarvan lagen in genen (*AMPH, LRMP, RASAL2, NELL1*).

Hoofdstuk 6 beschrijft het gebruik van een genetische risicoscore (GRS) voor het voorspellen van een verlaagd longfunctie niveau in de algemene bevolking. De geforceerde uitademing in een seconde (FEV_1), gemeten met behulp van spirometrie, weerspiegelt de fysiologische toestand van de luchtwegen en de longen. Daarnaast is een verlaagde FEV_1 een sterke voorspeller van toekomstige ziekte en sterfte. FEV_1 is zeer erfelijk, maar SNPs in kandidaat genen voorspellen elk apart slechts een klein deel van het longfunctie niveau. Uit onze studie blijkt dat een gewogen genetische risicoscore (wGRS, weging op basis van de SNP-specifieke associatie met de longfunctie), bestaande uit 11 SNPs, een beter instrument is om te discrimineren tussen personen met een normaal en een verminderd longfunctie niveau dan een ongewogen GRS (uGRS). Echter het gebruik van de wGRS geeft niet een betere voorspelling dan het opnemen van de 11 afzonderlijke SNPs in het model.

In **hoofdstuk 7** onderzochten we de associatie tussen kortademigheidsklachten en algemene en oorzaakspecifieke sterfte. Het Vlagtwedde-Vlaardingen cohort is gestart in 1965 en deelnemers aan dit onderzoek werden elke drie jaar opnieuw

onderzocht, tot en met de laatste meting in 1989/1990. Dit bood ons de unieke mogelijkheid om de aanwezigheid en de ernst van kortademigheidsklachten en veranderingen hierin tijdens het leven te onderzoeken in relatie tot algemene en oorzaakspecifieke sterfte. Kortademigheid blijkt een belangrijke voorspeller voor algemene en cardiovasculaire sterfte, maar alleen ernstige kortademigheid is geassocieerd met COPD sterfte. Aanhoudende kortademigheid of het ontwikkelen van kortademigheid in de loop der tijd verhoogt het risico op algemene sterfte. Echter, wanneer de kortademigheidsklachten in de loop der tijd verdwijnen, normaliseert ook het sterfterisico weer.

Acknowledgments

Acknowledgements

I would like to acknowledge all the people who contributed to this thesis and who helped me during the PhD period.

I would like to express my deep gratitude to my promotor Prof. Marike Boezen and co-promotor Dr. Judith Vonk from the Department of Epidemiology, UMCG, for their continuous support, scientific advices and all the comments that let me improve our manuscripts. Dear Marike and Judith, I appreciate your enthusiasm for my ideas and the freedom in the research I had from the beginning. During these four years of my PhD studies under your supervision I became a better and independent researcher. Thank you both for the excellent atmosphere, patience and understanding. Moreover, I am grateful that you gave me the great opportunity to work at the Mayo Clinic during the last six months of my project. It was a real pleasure to work with you. Hartstikke bedankt!

I thank Prof. James Kirkland from the Mayo Clinic Robert and Arlene Kogod Center on Aging for the possibility to work in his lab for six months. I thank the whole Kirkland group that I could learn from their experience in a friendly atmosphere. My thanks also go to Ms. Armstrong for her enormous hospitality and the farewell party for me. I was so happy I met people like you and also very sad that it was already time to leave. I enjoyed my stay in Rochester very much.

I would like to thank also the Dean of the Faculty of Medical Sciences and the Supervisory Board of the UMCG, Prof. Folkert Kuipers, for his help to arrange my stay at the Mayo Clinic.

I would like to thank the members of the reading committee (Prof. J. Kirkland, Prof. E. Slagboom and Prof. F. Kuipers) for their willingness to read and review this thesis.

Dear Simona, I learned from you how to deal with common issues during the PhD project, like stress before a presentation or a rejection of the paper. Now it is funny when I think how many times you calmed me down when I was panicking. Thank you for your patience.

Dear Despo, it was a great pleasure to share the office with you. Thank you for all the talks we had, at work and in the center by night. It was fun sharing the office with you. I have to say it was tough when you left to Boston, then I realized how much your optimistic attitude helped me.

I would like to thank my colleague Niloofar for discussing issues regarding the Vlagtwedde/Vlaardingen cohort, for the time we spent in the serum archive preparing blood samples and the work in the lab.

I thank my colleague Kim for discussing different scientific issues and for all help with Dutch, whenever I needed it. Dank je Kim.

I also appreciate all the short talks I had on the corridor or next to the coffee machine with my colleagues from the 4th floor.

I am grateful to all people from Department of Epidemiology and from the Groningen Research Institute for Asthma and COPD (GRIAC) for stimulating research meetings.

I had fantastic time and many parties with my friends I met in Groningen. Asia i Marcin, cieszę się, że Was poznałam. Dzięki Wam

mój czas spędzony w Gro był ciekawszy. Wiele imprez jak „Pierogi Party" albo obiad z zupą cebulową na pewno będę długo pamiętać. Również dziękuję innym uczestnikom imprez na Jensemaheerd. Asia, poza czasem rozrywkowym dziękuję również za pomoc w moich pytaniach dotyczących GWAS i innych spraw genetycznych.

Gosia, dziękuję za spędzony czas, zarówno w deszczowym Groningen jak i gorącej Jordanii. Ogromna ilość niesamowitych przygód nas spotkała. Wiele razy miałyśmy wielki ubaw, często z powodow zrozumiałych tylko dla nas. Dziękuję również za cząstkę polskości w pysznym bigosie, kiedy naprawde brakowalo mi polskości. Dzięki za długie rozmowy, za wieczorne wyjścia i noce przetańczone w De Kaar i Pakhuisie.

Ula, dziękuję za ułatwienie początków mojego życia w Groningen, czyli pomoc w znalezieniu mieszkania, ponieważ szukanie na odległość z Polski było lekko stresujące.

Justyn, dziękuję za wpadanie na kawę, za krótkie rozmowy w ramach przerwy w pracy, za naukę reggaetonu.

Dziękuje Dagmarze i Marcinowi, moim przyjaciołom ze studiów w Łodzi, za długie rozmowy na Skype dotyczące min. rozterek doktoratowych, za pielęgnowanie przyjaźni mimo odległości i za miły czas kiedy byłam w Polsce.

I thank Lovebird for the beautiful cover and printing of the thesis.

Last but not least, I would like to thank my family: my brother Bartuś, called "Bubak", for his quite often visits and our own way of spending Easter in Groningen, for long talks and lots of fun. To wielkie szczęście mieć takiego brata jak Ty.

I thank my parents for supporting me with all decisions I made, for their visits in Groningen as well as good time we had every time I visited you in Poland. Serdecznie dziękuję.