Chapter 4

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

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

Hum Mol Genet 2014; In press.

Abstract

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

Introduction

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

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

Results

Genome-wide association analysis

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

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

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

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

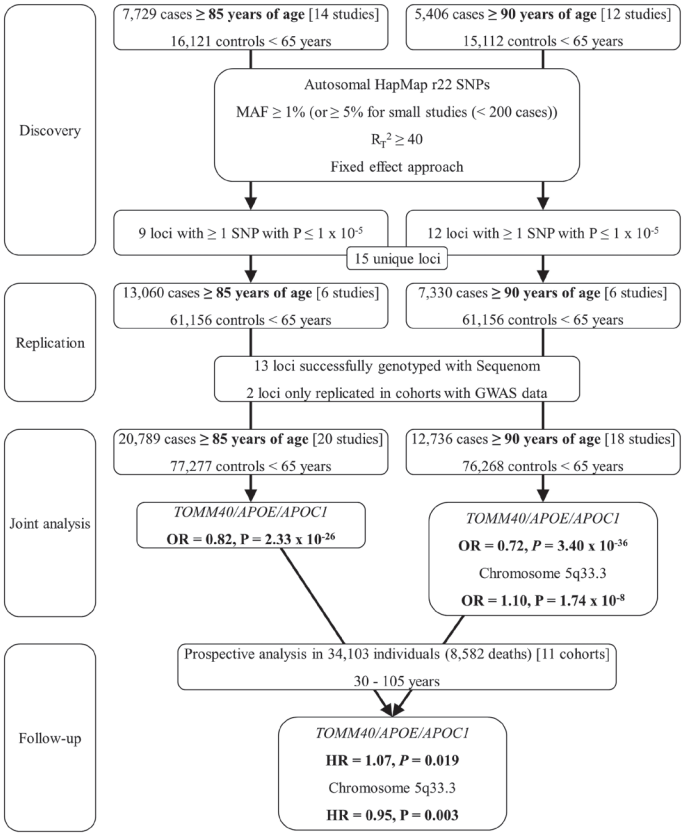
Replication

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

Prospective analysis

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

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



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

Association with cardiovascular disease and blood pressure

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

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

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

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

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

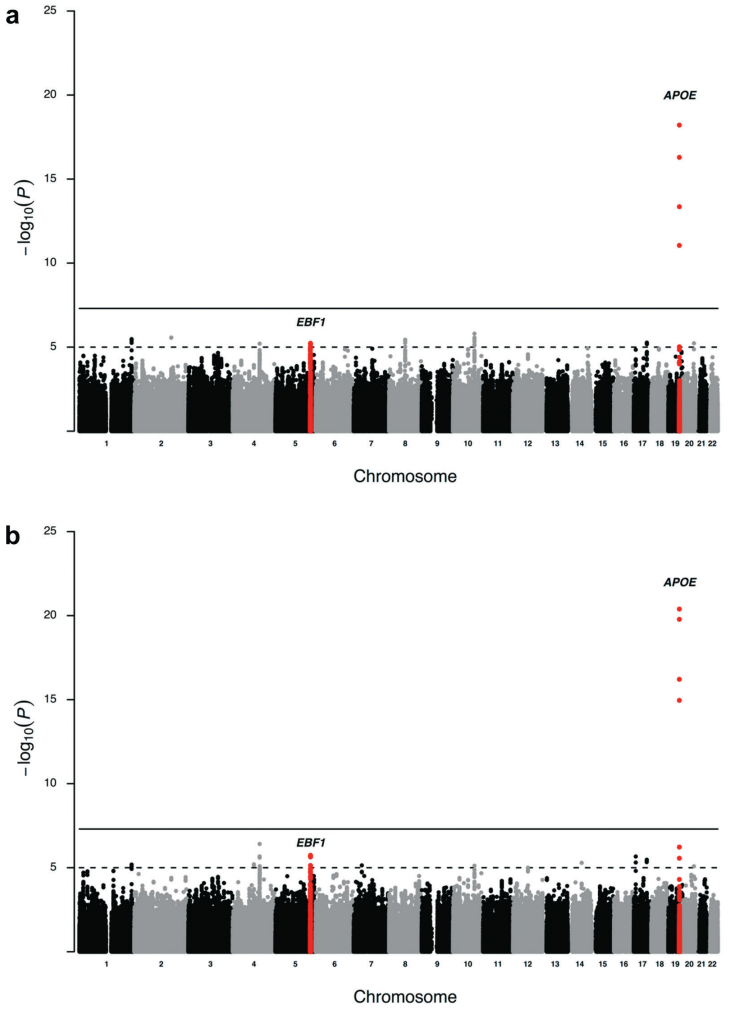
Table 4.2 Results of the discovery phase, replication phase and joint analysis of cases aged ≥ 90 years.

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

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

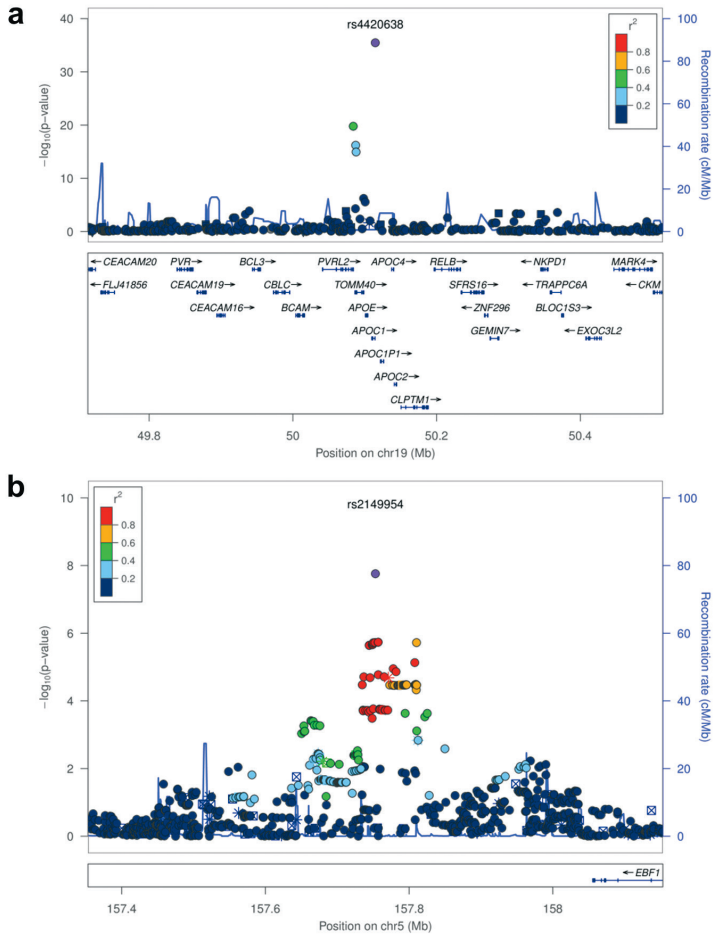
Chr, chromosome according to NCBI build 36; *Position* (bp); position of the lead SNP according to NCBI build 36; *EA*, effect allele; *EAF*, effect allele frequency after pooling the data of all analyzed individuals; *OR*, odds ratio for the effect allele; *95% CI*, 95% confidence interval; I^2, heterogeneity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

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



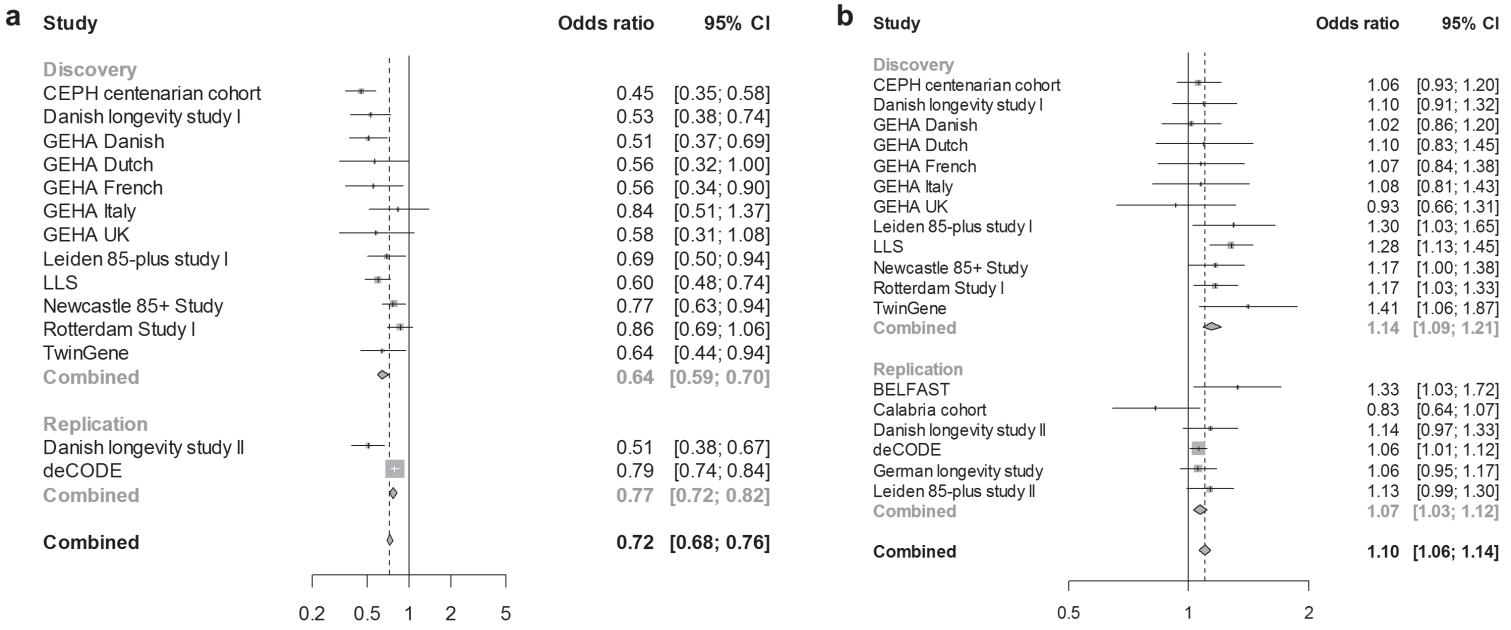
addition, we observed an effect of this SNP on non-CVD mortality (HR = 0.86, *P* = 0.002, Table S4.7). We also examined the Coronary ARtery DIsease Genome-Wide Replication And Meta-Analysis (CARDIoGRAM) GWAS [23], which showed a significant association of rs2149954 with a decreased risk for coronary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

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



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

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



additional to those involved in blood pressure regulation.

Phenotypic characterization and pathway analysis

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

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

Fine mapping and functional characterization

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

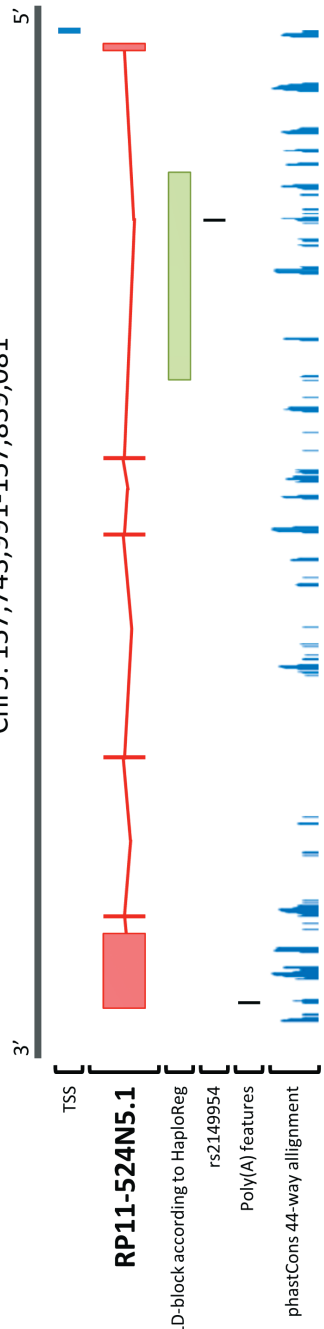


Figure 4.5 Chromosomal region around rs2149954. The region contains a long intronic non-coding RNA (RP11-524N5.1) for which the poly(A) features are supported by PolyA-seq reads from liver, muscle and testis. RP11-524N5.1 is transcribed from the negative strand, and the phastCons 44-way alignment supports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

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

Discussion

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

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

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

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

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

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

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

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

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

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

Materials and methods

Study populations

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

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

Genotyping, imputation and genome-wide association analysis

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

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

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

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

Meta-analyses

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

Conditional analysis

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

Sex-stratified analysis

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

Prospective analysis

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

Pathway analysis

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

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

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

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

eQTL analysis

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

Acknowledgments

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

Funding

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

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

References

1. Oeppen J, Vaupel JW. Demography. Broken limits to life expectancy. *Science* 2002; **296**: 1029-31.
2. Jagger C, Gillies C, Moscone F, Cambois E, van Oyen H, Nusselder W, *et al*. Inequalities in healthy life years in the 25 countries of the European Union in 2005: a cross-national meta-regression analysis. *Lancet* 2008; **372**: 2124-31.
3. Barzilai N, Atzmon G, Schechter C, Schaefer EJ, Cupples AL, Lipton R, *et al*. Unique lipoprotein phenotype and genotype associated with exceptional longevity. *JAMA* 2003; **290**: 2030-40.
4. Derhovanessian E, Maier AB, Beck R, Jahn G, Hahnel K, Slagboom PE, *et al*. Hallmark features of immunosenescence are absent in familial longevity. *J Immunol* 2010; **185**: 4618-24.
5. Newman AB, Glynn NW, Taylor CA, Sebastiani P, Perls TT, Mayeux R, *et al*. Health and function of participants in the Long Life Family Study: A comparison with other cohorts. *Aging (Albany NY)* 2011; **3**: 63-76.
6. Slagboom PE, Beekman M, Passtoors WM, Deelen J, Vaarhorst AA, Boer JM, *et al*. Genomics of human longevity. *Philos Trans R Soc Lond B Biol Sci* 2011; **366**: 35-42.
7. Wijsman CA, Rozing MP, Streefland TC, Le Cessie, Mooijaart SP, Slagboom PE, *et al*. Familial longevity is marked by enhanced insulin sensitivity. *Aging Cell* 2011; **10**: 114-21.
8. Atzmon G, Schechter C, Greiner W, Davidson D, Rennert G, Barzilai N. Clinical phenotype of families with longevity. *J Am Geriatr Soc* 2004; **52**: 274-7.
9. Westendorp RG, van Heemst D, Rozing MP, Frolich M, Mooijaart SP, Blauw GJ, *et al*. Nonagenarian siblings and their offspring display lower risk of mortality and morbidity than sporadic nonagenarians: The Leiden Longevity Study. *J Am Geriatr Soc* 2009; **57**: 1634-7.
10. Herskind AM, McGue M, Holm NV, Sorensen TI, Harvald B, Vaupel JW. The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870-1900. *Hum Genet* 1996; **97**: 319-23.
11. Hjelmborg JV, Iachine I, Skytthe A, Vaupel JW, McGue M, Koskenvuo M, *et al*. Genetic influence on human lifespan and longevity. *Hum Genet* 2006; **119**: 312-21.
12. Finch CE, Tanzi RE. Genetics of aging. *Science* 1997; **278**: 407-11.
13. Kirkwood TB, Cordell HJ, Finch CE. Speed-bumps ahead for the genetics of later-life diseases. *Trends Genet* 2011; **27**: 387-8.
14. Sacher GA. 1976 Robert W. Kleemeier Award Lecture: Longevity, aging, and death: an evolutionary perspective. *Gerontologist* 1978; **18**: 112-9.
15. Ganna A, Rivadeneira F, Hofman A, Uitterlinden AG, Magnusson PK, Pedersen NL, *et al*. Genetic determinants of mortality. Can findings from genome-wide association studies explain variation in human mortality? *Hum Genet* 2013; **132**: 553-61.
16. Beekman M, Nederstigt C, Suchiman HE, Kremer D, van der Breggen R, Lakenberg N, *et al*. Genome-wide association study (GWAS)-identified disease risk alleles do not compromise human longevity. *Proc Natl Acad Sci U S A* 2010; **107**: 18046-9.
17. Deelen J, Beekman M, Uh HW, Helmer Q, Kuningas M, Christiansen L, *et al*. Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. *Aging Cell* 2011; **10**: 686-98.
18. Newman AB, Walter S, Lunetta KL, Garcia ME, Slagboom PE, Christensen K, *et al*. A meta-analysis of four genome-wide association studies of survival to age 90 years or older: the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium. *J Gerontol A Biol Sci Med Sci* 2010; **65**: 478-87.
19. Deelen J, Beekman M, Capri M, Franceschi C, Slagboom PE. Identifying the genomic determinants of aging and longevity in human population studies: progress and challenges. *Bioessays* 2013; **35**: 386-96.
20. Gavrilova NS, Gavrilov LA, Evdokushkina GN, Semyonova VG, Gavrilova AL, Evdokushkina NN, *et al*. Evolution, mutations, and human longevity: European

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

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

Supplementary Information

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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