

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/26946> holds various files of this Leiden University dissertation

**Author:** Deelen, Joris

**Title:** Genetic and biomarker studies of human longevity

**Issue Date:** 2014-06-25

# Chapter 5

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

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

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

## Abstract

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

## Introduction

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

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

Human longevity is a complex trait that is assumed to be determined by variation

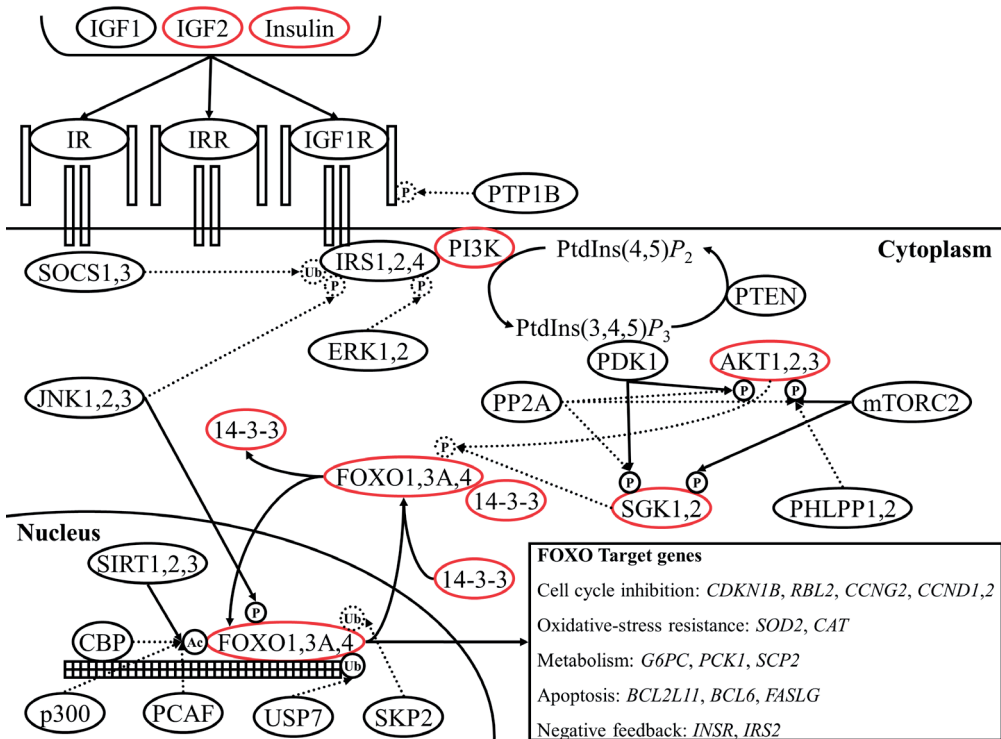
in many genes with small effects. Previous GWAS, in which single SNP analyses were performed [6,7], have identified only one genome-wide significant locus contributing to survival into old age; *APOE*. However, the genetic contribution to human lifespan variation, determined in twin studies, is estimated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

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

Another candidate pathway for studying human longevity is the mechanism of telomere maintenance (TM). Telomeres are structures at the end of chromosomes, consisting of TTAGGG tandem repeats [31], which protect chromosomes from degradation or rearrangement [32]. In normal human cells, telomere length declines with every cell division [33] and when a critical length is reached, the cell will enter replicative senescence [34]. In human epidemiological studies in blood, increased telomere length has been associated with longevity [35], while decreased telomere

length has been associated with increased mortality [36-38], although some studies showed contradictory results [39,40]. Telomere integrity is essentially regulated by two protein networks, telomerase and its associated factors, which regulate telomere length, and the shelterin complex, which covers the telomeres [41,42] (Figure 5.2). Several studies have investigated associations between single SNPs in telomerase and shelterin genes and telomere length. The most promising results came from *TERC* and *TERT* [35,43-46], of which the latter has also been associated with human longevity [35].

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



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

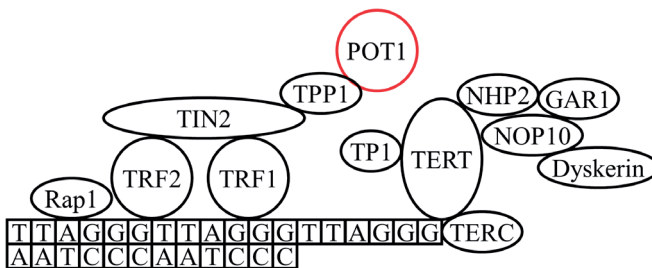
## Results

For the IIS pathway, we selected genes encoding proteins that belong to the well-described core of the pathway, consisting

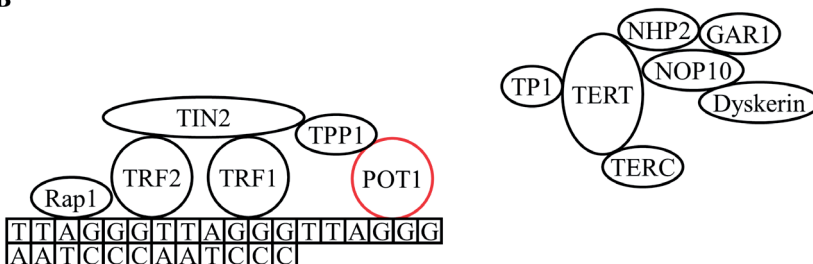
of IGF1R/IR/IRR, IRS, PI3K, AKT/SGK, FOXO, and SIRT, or that had a direct activating or inhibiting effect on these core components [51,52]. In addition, we selected several FOXO target genes that play a role in cell-cycle inhibition, oxidative-stress resistance, metabolism, and apoptosis [52] (Figure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

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

**A**



**B**



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

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

To determine which genes are mainly responsible for the observed association of the pathway SNP sets from the IIS and TM pathways with longevity, we also investigated the association of gene SNP sets from these pathways. Although the power to detect an association using gene SNP set analysis is lower than for pathway SNP set analysis, due to the larger number of tests, it provides a ranking of genes based on the contribution to the observed associations of the pathways.

To analyze the gene SNP sets, we used the PLINK set-based test, Global test, and SNP ratio test. GRASS was not used, since the underlying statistical method of this test is less suitable for analysis of gene SNP sets. Nine of the 68 IIS pathway gene SNP sets (*AKT1*, *AKT3*, *FOXO4*, *IGF2*, *INS*, *PIK3CA*, *SGK1*, *SGK2*, and *YWHAG*) and 1 of the 13 TM pathway gene SNP sets (*POT1*) showed an association ( $P < 0.05$ ) with longevity in at least two tests (Tables 5.5 and 5.6).

## Discussion

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

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

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

Protein	Gene	ID	Chr	Start (bp)	End (bp)	Size (kb)	SNPs	Coverage
AKT1	<i>AKT1</i>	207	14	104,333,732	104,333,125	26.39	2	25.00%
AKT2	<i>AKT2</i>	208	19	45,428,064	45,483,142	55.08	6	45.45%
AKT3	<i>AKT3</i>	10000	1	241,718,158	242,073,509	355.35	25	50.00%
BIM	<i>BCL2L1</i>	10018	2	111,594,962	111,642,493	47.53	12	47.06%
BCL-6	<i>BCL6</i>	604	3	188,921,859	188,946,207	24.35	6	23.81%
CAT	<i>CAT</i>	847	11	34,417,048	34,450,183	33.14	18	80.00%
Cyclin D1	<i>CCND1</i>	595	11	69,165,054	69,178,423	13.37	3	27.27%
Cyclin D2	<i>CCND2</i>	894	12	4,253,163	4,284,783	31.62	20	45.00%
Cyclin G2	<i>CCNG2</i>	901	4	78,297,381	78,310,237	12.86	4	33.33%
p27kip	<i>CDKN1B</i>	1027	12	12,761,569	12,766,572	5.00	9	75.00%
CBP	<i>CREBBP</i>	1387	16	3,715,057	3,870,122	155.07	15	50.00%
Deptor (mTORC2)	<i>DEPDC6</i>	64798	8	120,955,081	121,132,338	177.26	47	61.90%
p300	<i>EP300</i>	2033	22	39,818,560	39,906,027	87.47	6	50.00%
Fas ligand	<i>FASLG</i>	356	1	170,894,808	170,902,635	7.83	7	45.45%
FOXO1	<i>FOXO1</i>	2308	13	40,027,801	40,138,734	110.93	19	65.38%
FOXO3a	<i>FOXO3A</i>	2309	6	108,987,719	109,112,664	124.95	21	68.75%
FOXO4	<i>FOXO4</i>	4303	X	70,232,724	70,240,109	7.39	3	NA
G6P	<i>G6PC</i>	2538	17	38,306,340	38,319,976	13.63	5	83.33%
IGF-1	<i>IGF1</i>	3479	12	101,313,775	101,398,508	84.73	20	47.06%
IGF1R	<i>IGF1R</i>	3480	15	97,010,284	97,325,282	315.00	102	56.34%
IGF2	<i>IGF2</i>	3481	11	2,106,923	2,127,409	20.49	7	63.64%
Insulin	<i>INS</i>	3630	11	2,137,585	2,139,015	1.43	4	80.00%
IR	<i>INSR</i>	3643	19	7,063,266	7,245,011	181.75	52	50.00%
IRR	<i>INSRR</i>	3645	1	155,077,289	155,095,336	18.05	6	37.50%
IRS1	<i>IRS1</i>	3667	2	227,304,277	227,371,750	67.47	11	53.33%
IRS2	<i>IRS2</i>	8660	13	109,204,185	109,236,915	32.73	15	59.09%
IRS4	<i>IRS4</i>	8471	X	107,862,383	107,866,263	3.88	2	NA
PCAF	<i>KAT2B</i>	8850	3	20,056,528	20,170,900	114.37	44	62.96%
ERK2	<i>MAPK1</i>	5594	22	20,443,947	20,551,970	108.02	12	58.33%
ERK1	<i>MAPK3</i>	5595	16	30,032,927	30,042,131	9.20	2	33.33%
JNK1	<i>MAPK8</i>	5599	10	49,279,661	49,317,408	37.75	6	35.71%
JNK2	<i>MAPK9</i>	5601	5	179,593,201	179,651,677	58.48	17	40.00%
JNK3	<i>MAPK10</i>	5602	4	87,155,300	87,593,307	438.01	71	55.81%
mSIN1 (mTORC2)	<i>MAPKAP1</i>	79109	9	127,239,494	127,509,334	269.84	26	64.00%
mLST8 (mTORC2)	<i>MLST8</i>	64223	16	2,195,179	2,199,419	3.97	4	80.00%
mTOR (mTORC2)	<i>MTOR</i>	2475	1	11,089,175	11,245,195	156.02	11	50.00%

Protein	Gene	ID	Chr	Start (bp)	End (bp)	Size (kb)	SNPs	Coverage
PEPCK	<i>PCK1</i>	5105	20	55,569,542	55,574,919	5.38	10	33.33%
PDK1	<i>PDPK1</i>	5170	16	2,527,966	2,593,192	65.23	0	0.00%
PHLPP1	<i>PHLPP1</i>	23239	18	58,533,652	58,798,656	265.00	41	64.29%
PHLPP2	<i>PHLPP2</i>	23035	16	70,236,353	70,306,205	69.85	4	15.38%
PI3K	<i>PIK3CA</i>	5290	3	180,349,005	180,435,191	86.19	10	44.44%
	<i>PIK3CB</i>	5291	3	139,854,230	139,960,891	106.66	8	42.86%
	<i>PIK3CD</i>	5293	1	9,634,377	9,711,759	77.38	8	42.11%
	<i>PIK3RI</i>	5295	5	67,547,340	67,633,405	86.07	31	65.12%
	<i>PIK3R2</i>	5296	19	18,124,988	18,142,343	17.36	5	66.67%
	<i>PIK3R3</i>	8503	1	46,278,399	46,371,295	92.90	10	60.00%
PP2A	<i>PPP2R3B</i>	5526	11	64,448,719	64,458,526	9.81	3	37.50%
Protor-1 (mTORC2)	<i>PRR5</i>	55615	22	43,443,091	43,512,225	69.13	32	50.00%
PTEN	<i>PTEN</i>	5728	10	89,613,175	89,718,512	105.34	8	47.06%
PTPIB	<i>PTPNI</i>	5770	20	48,560,265	48,634,706	74.44	17	44.44%
p130Rb2	<i>RBL2</i>	5934	16	52,025,852	52,083,061	57.21	3	33.33%
RICTOR (mTORC2)	<i>RICTOR</i>	253260	5	38,973,780	39,110,258	136.48	9	30.77%
SCP2	<i>SCP2</i>	6342	1	53,165,489	53,289,877	124.39	18	50.00%
SGK1	<i>SGK1</i>	6446	6	134,532,077	134,680,889	148.81	38	46.75%
SGK2	<i>SGK2</i>	10110	20	41,621,049	41,647,687	26.64	9	34.62%
SIRT1	<i>SIRT1</i>	23411	10	69,314,433	69,348,153	33.72	4	33.33%
SIRT2	<i>SIRT2</i>	22933	19	44,061,035	44,082,342	21.31	7	38.89%
SIRT3	<i>SIRT3</i>	23410	11	205,030	226,362	21.33	17	60.00%
SKP2	<i>SKP2</i>	6502	5	36,187,902	36,219,899	32.00	15	51.72%
SOC31	<i>SOC31</i>	8651	16	11,255,775	11,257,540	1.77	4	50.00%
SOC33	<i>SOC33</i>	9021	17	73,864,453	73,867,755	3.30	4	50.00%
MnSOD	<i>SOD2</i>	6648	6	160,020,139	160,034,343	14.20	4	44.44%
USP7	<i>USP7</i>	7874	16	8,893,452	8,964,842	71.39	12	42.11%
14-3-3	<i>YWHAH</i>	7529	20	42,947,654	42,970,587	22.93	6	50.00%
	<i>YWHAE</i>	7531	17	1,194,584	1,250,306	55.72	16	70.00%
	<i>YWHAG</i>	7532	7	75,794,044	75,826,278	32.23	5	35.71%
	<i>YWHAH</i>	7533	22	30,670,479	30,683,590	13.11	9	43.75%
	<i>YWHAIQ</i>	10971	2	9,641,547	9,688,635	47.08	10	58.33%
	<i>YWHAZ</i>	7534	8	101,999,980	102,034,799	34.82	6	40.00%
<b>Total</b>							<b>1,023</b>	

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

**Table 5.2** Characteristics of the telomere maintenance pathway proteins.

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

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

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

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

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

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

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

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

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

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

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

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

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

Gene	PLINK set-based test*	Global test*	SNP ratio test*
<i>ACD</i>	1	0.491	1
<i>DKC1</i>	1	0.642	1
<i>GAR1</i>	1	0.281	1
<i>NHP2</i>	1	0.759	1
<i>NOPI0</i>	1	0.208	1
<i>POT1</i>	0.007	0.014	0.019
<i>TEP1</i>	1	0.525	1
<i>TERC</i>	1	0.202	1
<i>TERF1</i>	1	0.821	1
<i>TERF2</i>	0.018	0.160	0.164
<i>TERF2IP</i>	1	0.825	1
<i>TERT</i>	1	0.471	1
<i>TINF2</i>	1	0.587	1

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

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

Although SNPs in *FOXO3A* have previously been associated with longevity in several independent studies [25-30], the gene SNP set showed no effect in our study in the PLINK set-based test, Global test, and SNP ratio test ( $P = 0.181$ ,  $P = 0.138$ , and  $P = 0.180$ , respectively, Table 5.5). This might be due to the fact that the effects of *FOXO3A* on longevity are most prominent in centenarians.

As was previously reported by Flachsbart *et al.*, centenarians represent a highly selected phenotype, even among nonagenarians [26]. In addition, the genetic contribution to longevity in general is increased at higher ages [9] and the small effects of longevity-promoting gene variants, relative to other factors, may be larger in centenarians [57] and not detectable in nonagenarians. The cases in our study, which are from long-lived families, have a mean age of 94 years. Yet, we had only 11 individuals > 100 years, which may explain the absence of significance of the *FOXO3A* association in our population.

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

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

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

were part of this pathway, which indicates that the pathway definition used in this study could have had a large influence on the results of the analysis.

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

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

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

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

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

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

## Materials and methods

### Study Populations

#### *Leiden Longevity Study*

For the LLS, long-lived siblings of European descent were recruited together with their offspring and the partners of the offspring. Families were included if at least two long-lived siblings were alive and fulfilled the age criterion of 89 years or older for men and 91 years or older for women, representing less than 0.5% of the Dutch population in 2001 [65]. In total, 944 long-lived proband siblings were included with a mean age of 94 years (range, 89-104), 1,671 offspring (61 years, 39-81), and 744 partners (60 years, 36-79). DNA from the LLS was extracted from samples at baseline using conventional methods [66]. For the GWAS, 403 unrelated LLS siblings (one sibling from each sibling pair) were included (LLS GWAS cases) [6].

#### *Rotterdam Study*

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

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

#### *Population substructure*

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

#### **Genotyping and SNP selection**

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

We analyzed SNPs within a 10-kb window around genes encoding proteins that belonged to the IIS (Figure 5.1) and TM pathway (Figure 5.2). A gene was defined as an NCBI Entrez Gene (mRNA or RNA)

cluster, corresponding to a set of transcripts (RefSeq) for which the alignments can be obtained from the UCSC genome browser (<http://genome.ucsc.edu/>), in which all transcripts within a cluster agree on strand and overlap. Due to an overlap of the 10-kb windows around *IGF2* and *INS*, two SNPs, rs4320932 and rs7924316, were assigned to both genes.

#### **Statistical analysis**

##### *PLINK set-based test*

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

##### *GRASS*

GRASS (<http://linchen.fhcr.org/grass.html>) [48] calculates “eigenSNPs” for each gene in the pathway SNP set by summarizing the

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

#### *Global test*

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

#### *SNP ratio test*

The SNP ratio test (<http://sourceforge.net/projects/snpratiotest/>) [50] performs a single

SNP analysis (in our case, a trend test) of the original pathway or gene SNP set and of similar-sized SNP sets in which the phenotype status of the individuals is permuted. An empirical  $P$ -value of the SNP set is computed by calculating the ratio between the proportion of SNPs that shows an association below a certain  $P$ -value threshold ( $p$ ) in the original GWAS dataset and in the simulated GWAS datasets. The number of significant SNPs in the simulated GWAS datasets is defined as the top  $n$  SNPs with the lowest  $P$ -values, where  $n$  is the number of SNPs with an association below  $p$  in the original GWAS dataset. For the analysis in this study, we made use of the scripts described in “SRT\_documentation\_090310.pdf” (<http://sourceforge.net/projects/snpratiotest/>). For the analysis in this study,  $p$  was set to 0.05 and the number of simulated datasets used was 10,000.

#### *Statistical significance*

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

## **Acknowledgments**

We thank all participants of the Leiden Longevity Study and Rotterdam Study. The research leading to these results has received funding from the European Union’s Seventh Framework Programme (FP7/2007-2011) under grant agreement n° 259679. This study was supported by a grant from the Innovation-Oriented Research Program on Genomics (SenterNovem IGE05007), the

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

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

## References

1. Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. *J Clin Invest* 2008; **118**: 1590-605.
2. Fridley BL, Biernacka JM. Gene set analysis of SNP data: benefits, challenges, and future directions. *Eur J Hum Genet* 2011; **19**: 837-43.
3. Holmans P. Statistical methods for pathway analysis of genome-wide data for association with complex genetic traits. *Adv Genet* 2010; **72**: 141-79.
4. Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. *Nat Rev Genet* 2010; **11**: 843-54.
5. Goeman JJ, Buhlmann P. Analyzing gene expression data in terms of gene sets: methodological issues. *Bioinformatics* 2007; **23**: 980-7.
6. Deelen J, Beekman M, Uh HW, Helmer Q, Kuningas M, Christiansen L, *et al.* Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. *Aging Cell* 2011; **10**: 686-98.
7. Newman AB, Walter S, Lunetta KL, Garcia ME, Slagboom PE, Christensen K, *et al.* A meta-analysis of four genome-wide association studies of survival to age 90 years or older: the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium. *J Gerontol A Biol Sci Med Sci* 2010; **65**: 478-87.
8. Gudmundsson H, Gudbjartsson DF, Frigge M, Gulcher JR, Stefansson K. Inheritance of human longevity in Iceland. *Eur J Hum Genet* 2000; **8**: 743-9.
9. Hjelmborg JV, Iachine I, Skytthe A, Vaupel JW, McGue M, Koskenvuo M, *et al.* Genetic influence on human lifespan and longevity. *Hum Genet* 2006; **119**: 312-21.
10. Skytthe A, Pedersen NL, Kaprio J, Stazi MA, Hjelmborg JV, Iachine I, *et al.* Longevity studies in GenomEUtwin. *Twin Res* 2003; **6**: 448-54.
11. Tatar M, Bartke A, Antebi A. The endocrine regulation of aging by insulin-like signals. *Science* 2003; **299**: 1346-51.
12. Blüher M, Kahn BB, Kahn CR. Extended longevity in mice lacking the insulin receptor in adipose tissue. *Science* 2003; **299**: 572-4.
13. Clancy DJ, Gems D, Harshman LG, Oldham S, Stocker H, Hafén E, *et al.* Extension of life-span by loss of CHICO, a Drosophila insulin receptor substrate protein. *Science* 2001; **292**: 104-6.
14. Friedman DB, Johnson TE. A mutation in the age-1 gene in *Caenorhabditis elegans* lengthens life and reduces hermaphrodite fertility. *Genetics* 1988; **118**: 75-86.
15. Giannakou ME, Goss M, Junger MA, Hafén E, Leevers SJ, Partridge L. Long-lived *Drosophila* with overexpressed dFOXO in adult fat body. *Science* 2004; **305**: 361.
16. Holzenberger M, Dupont J, Ducos B, Leneuve P, Geloen A, Even PC, *et al.* IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. *Nature* 2003; **421**: 182-7.
17. Hwangbo DS, Gershman B, Tu MP, Palmer M, Tatar M. *Drosophila* dFOXO controls lifespan and regulates insulin signalling in brain and fat body. *Nature* 2004; **429**: 562-6.
18. Kenyon C, Chang J, Gensch E, Rudner A, Tabtiang R. A *C. elegans* mutant that lives twice as long as wild type. *Nature* 1993; **366**: 461-4.
19. Kimura KD, Tissenbaum HA, Liu Y, Ruvkun G. *daf-2*, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. *Science* 1997; **277**: 942-6.
20. Lin K, Dorman JB, Rodan A, Kenyon C. *daf-16*: An HNF-3/forkhead family member that can function to double the life-span of *Caenorhabditis elegans*. *Science* 1997; **278**: 1319-22.
21. Morris JZ, Tissenbaum HA, Ruvkun G. A phosphatidylinositol-3-OH kinase family member regulating longevity and diapause in *Caenorhabditis elegans*. *Nature* 1996; **382**: 536-9.
22. Ogg S, Paradis S, Gottlieb S, Patterson GI, Lee L, Tissenbaum HA, *et al.* The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C. elegans*. *Nature* 1997; **389**: 994-9.

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

44. Levy D, Neuhausen SL, Hunt SC, Kimura M, Hwang SJ, Chen W, *et al.* Genome-wide association identifies OBFC1 as a locus involved in human leukocyte telomere biology. *Proc Natl Acad Sci U S A* 2010; **107**: 9293-8.
45. Mirabello L, Yu K, Kraft P, De Vivo I, Hunter DJ, Prescott J, *et al.* The association of telomere length and genetic variation in telomere biology genes. *Hum Mutat* 2010; **31**: 1050-8.
46. Rafnar T, Sulem P, Stacey SN, Geller F, Gudmundsson J, Sigurdsson A, *et al.* Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nat Genet* 2009; **41**: 221-7.
47. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559-75.
48. Chen LS, Hutter CM, Potter JD, Liu Y, Prentice RL, Peters U, *et al.* Insights into colon cancer etiology via a regularized approach to gene set analysis of GWAS data. *Am J Hum Genet* 2010; **86**: 860-71.
49. Goeman JJ, van de Geer SA, de Kort F, van Houwelingen HC. A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics* 2004; **20**: 93-9.
50. O'Dushlaine C, Kenny E, Heron EA, Segurado R, Gill M, Morris DW, *et al.* The SNP ratio test: pathway analysis of genome-wide association datasets. *Bioinformatics* 2009; **25**: 2762-3.
51. Taniguchi CM, Emanuelli B, Kahn CR. Critical nodes in signalling pathways: insights into insulin action. *Nat Rev Mol Cell Biol* 2006; **7**: 85-96.
52. van der Horst A, Burgering BM. Stressing the role of FoxO proteins in lifespan and disease. *Nat Rev Mol Cell Biol* 2007; **8**: 440-50.
53. Harrington L, McPhail T, Mar V, Zhou W, Oulton R, Bass MB, *et al.* A mammalian telomerase-associated protein. *Science* 1997; **275**: 973-7.
54. Vulliamy T, Beswick R, Kirwan M, Marrone A, Digweed M, Walne A, *et al.* Mutations in the telomerase component NHP2 cause the premature ageing syndrome dyskeratosis congenita. *Proc Natl Acad Sci U S A* 2008; **105**: 8073-8.
55. Eriksson N, Macpherson JM, Tung JY, Hon LS, Naughton B, Saxonov S, *et al.* Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet* 2010; **6**: e1000993.
56. Blake JA, Bult CJ, Kadin JA, Richardson JE, Eppig JT. The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics. *Nucleic Acids Res* 2011; **39**: D842-D848.
57. Perls T, Kunkel LM, Puca AA. The genetics of exceptional human longevity. *J Mol Neurosci* 2002; **19**: 233-8.
58. Yang Q, Zheng YL, Harris CC. POT1 and TRF2 cooperate to maintain telomeric integrity. *Mol Cell Biol* 2005; **25**: 1070-80.
59. He H, Wang Y, Guo X, Ramchandani S, Ma J, Shen MF, *et al.* Pot1b deletion and telomerase haploinsufficiency in mice initiate an ATR-dependent DNA damage response and elicit phenotypes resembling dyskeratosis congenita. *Mol Cell Biol* 2009; **29**: 229-40.
60. Hockemeyer D, Palm W, Wang RC, Couto SS, de Lange T. Engineered telomere degradation models dyskeratosis congenita. *Genes Dev* 2008; **22**: 1773-85.
61. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**: 25-9.
62. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000; **28**: 27-30.
63. Torkamani A, Topol EJ, Schork NJ. Pathway analysis of seven common diseases assessed by genome-wide association. *Genomics* 2008; **92**: 265-72.
64. Luo L, Peng G, Zhu Y, Dong H, Amos CI, Xiong M. Genome-wide gene and pathway analysis. *Eur J Hum Genet* 2010; **18**: 1045-53.
65. Schoenmaker M, de Craen AJ, de Meijer PH, Beekman M, Blauw GJ, Slagboom PE, *et al.* Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur J Hum Genet* 2006; **14**: 79-84.

66. Beekman M, Blauw GJ, Houwing-Duistermaat JJ, Brandt BW, Westendorp RG, Slagboom PE. Chromosome 4q25, microsomal transfer protein gene, and human longevity: novel data and a meta-analysis of association studies. *J Gerontol A Biol Sci Med Sci* 2006; **61**: 355-62.
67. Teichert M, Eijgelsheim M, Rivadeneira F, Uitterlinden AG, van Schaik RH, Hofman A, *et al.* A genome-wide association study of acenocoumarol maintenance dosage. *Hum Mol Genet* 2009; **18**: 3758-68.
68. Chapman J, Whittaker J. Analysis of multiple SNPs in a candidate gene or region. *Genet Epidemiol* 2008; **32**: 560-6.
69. Pan W. Asymptotic tests of association with multiple SNPs in linkage disequilibrium. *Genet Epidemiol* 2009; **33**: 497-507.

## Supplementary Information

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

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

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

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

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

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