Genetic and biomarker studies of human longevity

Deelen, J.

Chapter 7

General discussion

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

Main findings

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

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

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

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

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

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

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

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

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

Functional characterization of longevity loci

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

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

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

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

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

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

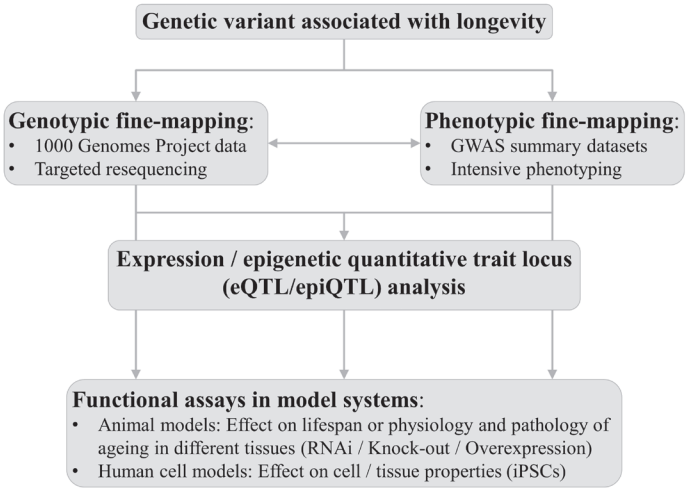
The second step is phenotypic fine-mapping, i.e., to identify other (combinations of) metabolic phenotypes, clinical endpoints,

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

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

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

Figure 7.1 Functional characterization of longevity loci.



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

Reducing heterogeneity in the healthy aging phenotype

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

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

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

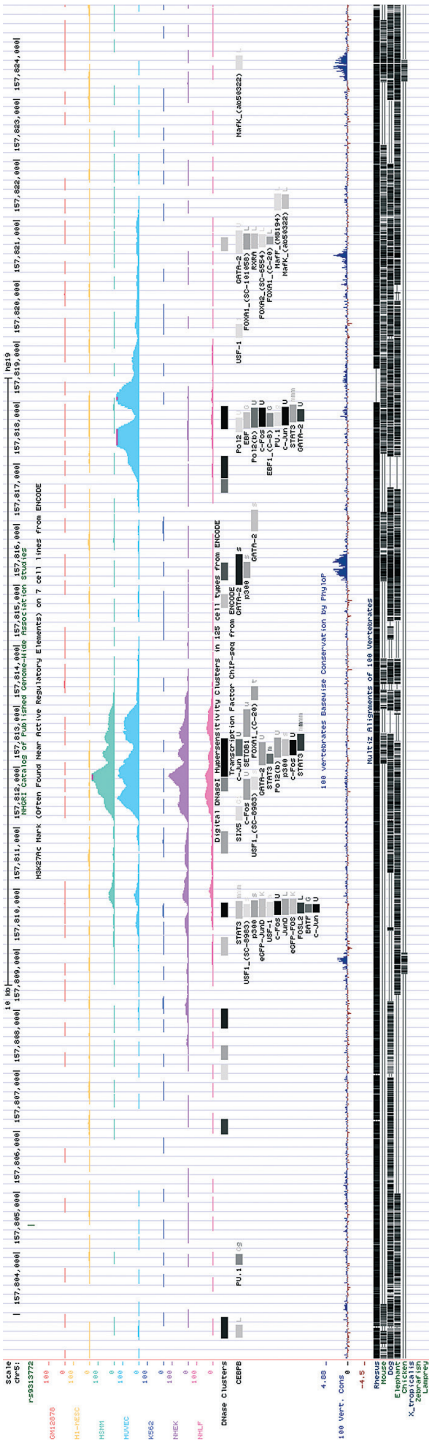


Figure 7.2 UCSC plot of the ~22.3 kb intergenic region on chromosome 5q33.3.

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

Combining study designs for biomarker research

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

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

Novel methods and technologies plea for data integration

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

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

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

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

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

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

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

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

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

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

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

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

Conclusions

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

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

References

1. Perls T, Shea-Drinkwater M, Bowen-Flynn J, Ridge SB, Kang S, Joyce E, *et al*. Exceptional familial clustering for extreme longevity in humans. *J Am Geriatr Soc* 2000; **48**: 1483-5.
2. Schoenmaker M, de Craen AJ, de Meijer PH, Beekman M, Blauw GJ, Slagboom PE, *et al*. Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur J Hum Genet* 2006; **14**: 79-84.
3. Atzmon G, Schechter C, Greiner W, Davidson D, Rennert G, Barzilai N. Clinical phenotype of families with longevity. *J Am Geriatr Soc* 2004; **52**: 274-7.
4. Newman AB, Glynn NW, Taylor CA, Sebastiani P, Perls TT, Mayeux R, *et al*. Health and function of participants in the Long Life Family Study: A comparison with other cohorts. *Aging (Albany NY)* 2011; **3**: 63-76.
5. Bos SD, Beekman M, Maier AB, Karsdal MA, Kwok WY, Bay-Jensen AC, *et al*. Metabolic health in families enriched for longevity is associated with low prevalence of hand osteoarthritis and influences OA biomarker profiles. *Ann Rheum Dis* 2013; **72**: 1669-74
6. Terry DF, Wilcox MA, McCormick MA, Pennington JY, Schoenhofen EA, Andersen SL, *et al*. Lower all-cause, cardiovascular, and cancer mortality in centenarians' offspring. *J Am Geriatr Soc* 2004; **52**: 2074-6.
7. Westendorp RG, van Heemst D, Rozing MP, Frolich M, Mooijaart SP, Blauw GJ, *et al*. Nonagenarian siblings and their offspring display lower risk of mortality and morbidity than sporadic nonagenarians: The Leiden Longevity Study. *J Am Geriatr Soc* 2009; **57**: 1634-7.
8. Beekman M, Nederstigt C, Suchiman HE, Kremer D, van der Breggen R, Lakenberg N, *et al*. Genome-wide association study (GWAS)-identified disease risk alleles do not compromise human longevity. *Proc Natl Acad Sci U S A* 2010; **107**: 18046-9.
9. Christensen K, Johnson TE, Vaupel JW. The quest for genetic determinants of human longevity: challenges and insights. *Nat Rev Genet* 2006; **7**: 436-48.
10. Schachter F, Faure-Delanef L, Guenot F, Rouger H, Froguel P, Lesueur-Ginot L, *et al*. Genetic associations with human longevity at the APOE and ACE loci. *Nat Genet* 1994; **6**: 29-32.
11. Malovini A, Illario M, Iaccarino G, Villa F, Ferrario A, Roncarati R, *et al*. Association study on long-living individuals from Southern Italy identifies rs10491334 in the CAMKIV gene that regulates survival proteins. *Rejuvenation Res* 2011; **14**: 283-91.
12. Nebel A, Kleindorp R, Caliebe A, Nothnagel M, Blanche H, Junge O, *et al*. A genome-wide association study confirms APOE as the major gene influencing survival in long-lived individuals. *Mech Ageing Dev* 2011; **132**: 324-30.
13. Newman AB, Walter S, Lunetta KL, Garcia ME, Slagboom PE, Christensen K, *et al*. A meta-analysis of four genome-wide association studies of survival to age 90 years or older: the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium. *J Gerontol A Biol Sci Med Sci* 2010; **65**: 478-87.
14. Sebastiani P, Solovieff N, Dewan AT, Walsh KM, Puca A, Hartley SW, *et al*. Genetic signatures of exceptional longevity in humans. *PLoS One* 2012; **7**: e29848.
15. Walter S, Atzmon G, Demerath EW, Garcia ME, Kaplan RC, Kumari M, *et al*. A genome-wide association study of aging. *Neurobiol Aging* 2011; **32**: 2109-28.
16. Finch CE, Tanzi RE. Genetics of aging. *Science* 1997; **278**: 407-11.
17. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, *et al*. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 2010; **467**: 832-8.
18. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, *et al*. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* 2012; **44**: 981-90.
19. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M,

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

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