Chapter 5
GWAS and Meta-Analysis in Aging/Longevity

Linda Broer and Cornelia M. van Duijn

The Role of Genes in Longevity

In the past century, most Western countries have experienced substantial increases in life expectancy. This has been mostly due to a marked reduction in early life mortality during the first half of the twentieth century, followed by an almost twofold reduction in mortality at ages above 70 years in the past 50 years [1, 2]. Longevity is often defined as reaching extreme age. There is no single accepted age threshold and given the ever increasing life expectancy and the differences in life expectancy across countries, the definition is time and place dependent. At present, the 'oldest-old' in Western societies are often defined as individuals of 85 years and older and this cut-off has been used in genetic studies in the past [3].

However, the percentage of individuals reaching 90 years of age, or even 100 years of age, is growing enormously [4]. The proportions of individuals in a given birth cohort projected to reach 90 or 100 years of age are shown in Fig. 5.1 [4]. The figure illustrates that the proportion of individuals who survive to age 90 has increased dramatically over the past century. When we consider the elderly of today (born between 1919–1929), less than 5% of the women and men reached age 90 years. Also for more recent cohorts (now middle age), reaching 90 years of age is still relatively rare, and reaching 100 years of age even more so. For example, 10% of women from the 1959 birth cohort are projected to reach 90 years of age, and only 0.3% are projected to reach 100 years of age [4]. How difficult it is to reach age 100 can be seen by comparing the likelihood of making it from birth to age 90 with the likelihood of making it from age 90 to age 100. These are similar implying surviving from 90 to 100 years is as difficult as living from 0 to 90 years [4].

C. M. van Duijn (✉) · L. Broer
Department of Epidemiology, Erasmus University Medical Center, Dr. Molewaterplein 50, PO-Box 2040, 3000 CA Rotterdam, Netherlands
e-mail: c.vanduijn@erasmusmc.nl

L. Broer
e-mail: l.broer@erasmusmc.nl

© Springer Science+Business Media New York 2015

G. Atzmon, PhD (ed.), *Longevity Genes,* Advances in Experimental Medicine and Biology 847, DOI 10.1007/978-1-4939-2404-2_5

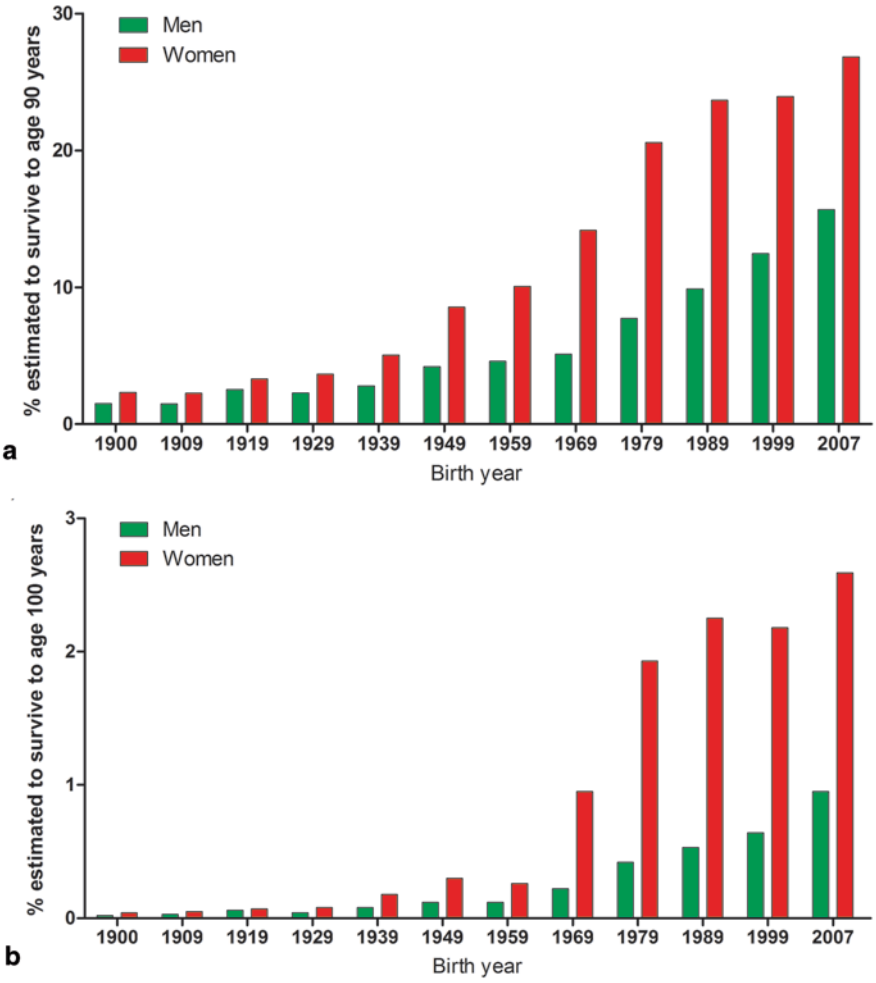

Fig. 5.1 Survivorship to ages 90 years (**a**) and 100 years (**b**) for the 1900–2007 birth cohorts by sex, United States. (Data were obtained from Arias [3])

Without a doubt, control of environmental risk factors (from hygiene to diet) and improved treatment of major diseases (cardiovascular and cancer) underlie the increase in life expectancy. Yet genes play a key role in reaching extreme age as shown by the fact that the heritability of age at death is higher at more exceptional thresholds for longevity [5]. Heritability estimates of age at death, range from 20 to 30% in twin registries [6, 7] and 15–25% in population-based samples [8, 9]. However, studying the birth cohort up that reached old age now, the heritability of surviving past 85 years was found to be 40% [10]. This is very similar to the heritability of other complex genetic traits such as blood pressure, lipids and diabetes [11, 12].

Despite the high heritability, in the previous century there has been little progress in unraveling the genetics of longevity. As has been the case for other diseases, candidate genes studies have limited few genes that have been replicated including apolipoprotein E and FOXO3a nor have family based studies yielded genes with major impact in the population [13–18]. Interestingly known age-related disease-causing genes have not been found to be associated with longevity [19, 20], suggesting there are specific domains in the genome that determine longevity beyond those that determine morbidity in the population. Genome-wide association studies (GWAS) have been able to identify hundreds of genetic loci for traits with similar or even lower heritability. The basic rationale of GWAS is that thousands to millions of genetic variants (single nucleotide polymorphisms (SNPs)) are measured across the genome and then associated to the phenotype of interest. Here we review the GWAS for longevity, distinguishing those using 85+ or 90+ as a cut-off and those studying more extreme phenotypes (100+ or centenarians). Further, we discuss an alternative approach to genetic studies of longevity using time to death as an outcome.

GWAS on Longevity (85+)

The Table 5.1 lists all currently performed GWAS on longevity, with their phenotype definition and sample sizes. The first GWAS investigating the longevity phenotype was by Newman et al [21]. This study consisted of 1836 individuals who achieved longevity, defined as 90 years and over. The comparison group consisted of 1955 individuals who died between 55 and 80 years of age [21]. The youngest age in the comparison group was set to match the minimum age at enrollment in one of the included cohorts. The maximal age at death in the comparison group was set arbitrarily at 80 years of age to include the majority of deaths, while excluding those individuals who survived far beyond average life expectancy for their respective birth cohort and nearly reached longevity [21]. None of the SNP-longevity associations achieved genome-wide significance (p-value $< 5 * 10^{-8}$). 24 independent regions with suggestive association levels (p-value $< 1 * 10^{-4}$) were identified (Table 5.2). 16 SNPs were successfully genotyped in a second stage including two independent cohorts. Only one of the SNPs had a smaller p-value after including the replication cohorts in the meta-analysis. This SNP, rs9664222, is located approximately 25 kb from the *MINPP1* gene and had an OR(odds Ratio) of 0.82 for the minor allele in the final meta-analysis (p-value $= 6.77 * 10^{-7}$) [21]. *MINPP1* encodes multiple inositol polyphosphate phosphatase 1, which is an enzyme that removes 3-phosphate from inositol phosphate substrates. MINPP1-deficient mice have no obvious defects, though targeted deletion *in vitro* is associated with slowed cellular proliferation [22].

Deelen et al published a longevity GWAS consisting of 4149 individuals over 85 years of age and a comparison group of 7582 younger controls [23]. In a first round including only one study (403 longevity cases and 1670 controls) no genome-wide significant SNPs were identified. For 58 out of 62 SNPs reaching a p-value $< 1 * 10^{-4}$

Table 5.1 GWAS studies on longevity phenotype

Author	Year	Phenotype definition	Sample size	Reference
Newman, AB	2010	Longevity: 90+ Comparison: died between 55–80	1836 cases 1955 controls	[21]
Deelen, J	2011	Longevity: 85+ Comparison: middle age	4149 cases 7582 controls	[23]
Nebel, A	2011	Longevity: 90 Comparison: middle age	763 cases 1085 controls	[27]
Malovini, A	2011	Longevity 90 Comparison: 18–45	582 cases 784 controls	[28]
Walter, S	2011	Survival: all-cause mortality Follow-up: 10.6 (5.4) years	25,007 total (8444 deaths)	[36]
Sebastiani, P	2012	Longevity 100+ Comparison: middle age	801 cases 914 controls	[30]

successful genotyping was obtained in the other cohorts (Table 5.3). One SNP on chromosome 19, rs2075650, was associated to longevity at genome-wide significance level (p-value = $3.39 * 10^{-17}$) with an OR of 0.71 [23]. This SNP is located in the *TOMM40* gene, next to the *APOE* gene. *APOE* had previously been identified as a longevity gene in candidate gene studies [24, 25], prompting the authors to test for independence of the signal. Conditional analysis confirmed the observed association was caused by the *APOE* locus. As *APOE* is known to be associated with Alzheimer's disease (AD), the authors investigated all other AD associated SNPs as summarized in AlzGene [26], but found no further significant associations.

In a GWAS including 763 longevity cases (90+) and 1085 control subjects of middle age Nebel et al tackled the longevity phenotype using both allele- and genotype-based case-control comparisons [27]. Their validation sample included 754 cases and 860 controls. 16 SNPs were selected for follow-up with *p*-values ranging from $3.7 * 10^{-10}$ to $9.1 * 10^{-6}$ (Table 5.4). Only rs4420638 was significant in the replication stage at a Bonferroni corrected level of significance (OR = 0.55; *p*-value = $1.9 * 10^{-8}$) [27]. This SNP is located 14 kb downstream of the *APOE* locus. Genome-wide haplotype analysis resulted in 13 significant haplotype pairs, but none were replicated.

Malovini *et al* used 582 longevity cases (90+) and 784 younger controls in their GWAS [28]. Three genetic models, allelic, dominant and recessive, were evaluated. In order to overcome the small sample size, resulting in a low power, a simulation study was performed which suggested that at a *p*-value cut-off for significance of 10^{-4} for at least one of the evaluated genetic models would guarantee a false-positive rate of approximately 2 in 10,000 independent tests. 67 SNPs with *p*-value < $1 * 10^{-4}$ were identified (Table 5.5). The authors claim that many of these SNPs mapped to genes potentially relevant to the aging process. One of the SNPs, rs10491334; *CAMKIV* had previously been associated with high diastolic blood pressure [29]. Replication of this SNP in 116 cases and 160 controls confirmed the finding (joint OR = 0.55; *p*-value = $1.68 * 10^{-6}$; dominant model) [28]. No replication for the other

Table 5.2 Top results of GWAS performed by Newman et al [21]

					Discovery			Replication	
SNP	Chr	Gene	EA	EAF	OR	*p*-value		OR	*p*-value
rs4443878	1	RGS7	T	0.04	0.41	$1.30*10^{-6}$		0.83	0.068
rs9825185	3	C3orf21	A	0.87	0.69	$2.50*10^{-6}$		0.91	0.045
rs954551	6	GRIK2	A	0.75	1.30	$5.30*10^{-6}$		NA	NA
rs7624691	3	IL20RB	T	0.57	1.25	$8.80*10^{-6}$		1.05	0.092
rs10888267	1	OR2W3	T	0.55	0.80	$9.70*10^{-6}$		NA	NA
rs9972933	17	ACCN1	T	0.23	0.77	$1.10*10^{-5}$		0.89	0.003
rs2739532	4		C	0.27	1.48	$1.10*10^{-5}$		NA	NA
rs8029244	15	LASS3	A	0.49	0.79	$1.20*10^{-5}$		0.90	0.002
rs16850255	1	PAPPA2	T	0.79	1.33	$1.20*10^{-5}$		1.09	0.041
rs1543505	14	REM2	A	0.72	0.79	$1.30*10^{-5}$		0.89	0.001
rs7321904	13	SPRY2	T	0.07	0.64	$1.30*10^{-5}$		0.92	0.179
rs17401847	1	OTUD3	A	0.85	0.72	$1.40*10^{-5}$		0.89	0.015
rs3124736	10	CASP7	A	0.03	2.30	$1.40*10^{-5}$		NA	NA
rs690232	9	DIRAS2	A	0.30	1.27	$1.60*10^{-5}$		NA	NA
rs9664222	10	MINPP1	A	0.21	0.77	$1.60*10^{-5}$		0.82	$6.8*10^{-7}$
rs11157721	14	LOC196913	T	0.39	0.79	$1.70*10^{-5}$		0.90	0.002
rs4690810	4	SC4MOL	T	0.65	1.27	$1.90*10^{-5}$		1.08	0.044
rs11605096	11	TMPRSS5	A	0.12	0.71	$1.90*10^{-5}$		NA	NA
rs16972414	18	PIK3C3	A	0.70	1.27	$2.00*10^{-5}$		NA	NA
rs12935091	16	ZNF19	A	0.93	1.61	$2.00*10^{-5}$		1.25	0.002
rs210332	14	BMP4	T	0.81	0.75	$2.30*10^{-5}$		NA	NA
rs17369174	8	CRISPLD1	T	0.90	1.45	$2.30*10^{-5}$		1.16	0.014
rs6721003	2	SCN7A	A	0.45	1.23	$2.40*10^{-5}$		1.09	0.006
rs4734457	8	ANKRD46	A	0.10	1.75	$2.50*10^{-5}$		1.10	0.098

EA effective allele, *EAF* effective allele frequency, *OR* odds ratio

suggested associations was attempted. Functional analysis showed that individuals homozygote for the polymorphism had significantly lower protein levels of CAMKIV compared to individuals carrying the wild-type gene. Additionally, *CAMKIV* incudes phosphorylation of a known longevity gene identified in candidate gene studies, *FOXO3* [28].

GWAS on Longevity (Centenarians)

To date, there is only the study of Sebastiani *et al* that included 801 unrelated centenarian cases and 914 population controls [30]. The controls were genetically matched to the cases that were either spouses of centenarian offspring ($n=241$) or

Table 5.3 Top results of GWAS performed by Deelen et al [23]

			Discovery		Replication	
SNP	Chr	Position	OR	p-value	OR	p-value
rs2075650	19	50087459	0.53	$2.65*10^{-6}$	0.71	$3.39*10^{-17}$
rs2003499	7	98852920	1.59	$8.07*10^{-5}$	1.19	$2.19*10^{-4}$
rs4736209	8	140208116	0.64	$9.50*10^{-6}$	0.90	$4.51*10^{-4}$
rs1516507	10	78707638	1.36	$9.67*10^{-5}$	1.09	0.002
rs4110518	10	96640560	1.52	$3.26*10^{-6}$	1.10	0.004
rs6577989	8	140182076	0.68	$3.24*10^{-5}$	0.92	0.006
rs7830605	8	140200576	0.69	$4.07*10^{-5}$	0.92	0.007
rs10401068	18	65414116	1.36	$9.62*10^{-5}$	1.08	0.008
rs1893132	18	2133155	2.07	$3.19*10^{-6}$	1.20	0.013
rs7661225	4	186275651	1.58	$6.59*10^{-5}$	1.12	0.013
rs886550	7	43302768	1.60	$5.75*10^{-5}$	1.12	0.019
rs11129533	3	32810964	1.39	$5.65*10^{-5}$	1.07	0.019
rs7005993	8	22845367	0.64	$5.38*10^{-5}$	0.92	0.021
rs2033563	8	103723247	1.48	$1.76*10^{-5}$	1.08	0.024
rs13248142	8	140182096	0.55	$1.43*10^{-5}$	0.92	0.024
rs625249	11	93149789	0.72	$9.29*10^{-5}$	0.94	0.027
rs1421746	5	127179875	1.38	$4.04*10^{-5}$	1.07	0.028
rs660100	1	4462210	1.48	$1.63*10^{-5}$	1.08	0.029
rs9827142	3	192599618	1.49	$4.88*10^{-6}$	1.07	0.029
rs9868286	3	192564180	1.48	$7.40*10^{-6}$	1.07	0.034
rs4681554	3	150992952	0.64	$9.82*10^{-5}$	0.94	0.050
rs642990	1	54461104	0.73	$9.81*10^{-5}$	0.95	0.052

Table 5.3 (continued)

			Discovery		Replication	
SNP	Chr	Position	OR	p-value	OR	p-value
rs12548929	8	140305428	0.55	$3.52*10^{-5}$	0.93	0.058
rs12548622	8	140190418	0.54	$8.78*10^{-6}$	0.93	0.059
rs6774262	3	32814515	1.40	$8.37*10^{-5}$	1.06	0.059
rs4133282	8	140287194	0.54	$2.02*10^{-5}$	0.93	0.066
rs2511703	8	103770272	0.73	$9.96*10^{-5}$	0.95	0.067
rs16861446	1	18221371	0.39	$8.47*10^{-5}$	0.91	0.071
rs970567	20	49281661	0.70	$3.50*10^{-5}$	0.95	0.075
rs10490478	2	207636308	1.63	$5.79*10^{-6}$	1.08	0.080
rs1859416	7	8866426	1.42	$5.04*10^{-5}$	1.06	0.080
rs11047358	12	24343978	1.47	$6.49*10^{-5}$	1.07	0.081
rs12101383	15	65055257	0.69	$9.53*10^{-5}$	0.95	0.112
rs268300	10	43900845	2.13	$4.05*10^{-5}$	1.14	0.113
rs12080088	1	230343981	1.45	$5.19*10^{-5}$	1.06	0.124
rs2290889	9	92679670	2.00	$9.14*10^{-6}$	1.11	0.125
rs2436932	8	103689078	1.59	$1.12*10^{-5}$	1.06	0.132
rs12892152	14	78049436	1.87	$5.78*10^{-5}$	1.11	0.162
rs11122430	1	230334906	1.50	$7.65*10^{-6}$	1.05	0.175
rs11776260	8	128451670	0.61	$4.30*10^{-5}$	0.96	0.250
rs2302951	19	53646233	1.51	$7.59*10^{-5}$	1.05	0.264
rs9662589	1	230344234	1.43	$9.68*10^{-5}$	1.04	0.268
rs7864625	9	92673769	2.00	$1.01*10^{-5}$	1.08	0.282
rs3959143	3	192600773	1.40	$2.61*10^{-5}$	1.03	0.290

Table 5.3 (continued)

			Discovery		Replication	
SNP	Chr	Position	OR	p-value	OR	p-value
rs10191593	2	207567835	1.49	$6.97*10^{-5}$	1.04	0.304
rs11782735	8	128435786	0.61	$4.92*10^{-5}$	0.97	0.369
rs6581191	12	57161965	1.38	$4.06*10^{-5}$	1.02	0.511
rs6852830	4	145726008	1.50	$7.17*10^{-5}$	1.02	0.641
rs7011660	8	30405716	1.42	$8.64*10^{-5}$	0.99	0.761
rs10502005	11	101985631	1.48	$2.42*10^{-5}$	1.01	0.784
rs6941242	6	48479887	1.97	$6.91*10^{-5}$	1.03	0.847
rs857788	1	157051761	1.41	$1.55*10^{-5}$	1.00	0.868
rs10931700	2	196176676	1.41	$6.06*10^{-5}$	1.00	0.869
rs857785	1	157050883	1.43	$2.68*10^{-5}$	0.99	0.877
rs12815289	12	102779349	0.69	$5.40*10^{-5}$	1.00	0.895
rs9473350	6	48475166	1.99	$5.68*10^{-5}$	1.02	0.943
rs10194564	2	142611043	1.55	$8.88*10^{-5}$	1.01	0.962
rs17154903	10	43839414	2.68	$7.31*10^{-5}$	1.03	0.976

OR odds ratio

Table 5.4 Top results of GWAS performed by Nebel et al [27]

				Discovery			Replication	
Chr.	Position	SNP	MAF	OR	*P*-value		OR	*P*-value
19	50114786	rs4420638	0.11	0.53	$3.70*10^{-10}$		0.55	$1.90*10^{-8}$
6	29778631	rs3129046	0.20	0.66	$1.20*10^{-7}$		0.95	0.523
6	29785931	rs1610742	0.20	0.66	$1.60*10^{-7}$		0.94	0.454
2	28515768	rs2338013	0.19	0.67	$7.40*10^{-7}$		1.06	0.496
6	29808162	rs1610601	0.20	0.68	$1.10*10^{-6}$		0.95	0.548
6	29753592	rs3129063	0.18	0.67	$2.20*10^{-6}$		0.92	0.353
6	29834025	rs1633063	0.20	0.68	$2.30*10^{-6}$		0.91	0.289
18	54052953	rs158869	0.48	1.37	$2.30*10^{-6}$		1.11	0.169
5	52022995	rs350450	0.22	0.69	$4.20*10^{-6}$		1.02	0.775
13	47001519	rs1575892	0.03	0.45	$4.50*10^{-6}$		1.13	0.442
6	29731718	rs29228	0.18	0.68	$4.70*10^{-6}$		0.93	0.446
16	73418057	rs16947526	0.06	0.56	$6.30*10^{-6}$		0.88	0.314
9	10715294	rs11790055	0.38	1.38	$7.10*10^{-6}$		1.01	0.942
1	175309154	rs12741354	0.45	0.74	$7.40*10^{-6}$		0.82	0.006
13	46905819	rs9595687	0.04	0.49	$8.00*10^{-6}$		1.13	0.470
9	10726580	rs10959258	0.37	1.36	$9.10*10^{-6}$		1.01	0.897

OR odds ratio

Table 5.5 Top hits of GWAS performed by Malovini et al [28]

SNP	CHR	Gene	Assoc Model	EA	EAF	p-value	OR
rs6504441	17	PRKCA	AM	T	0.30	$1.06 * 10^{-6}$	0.60
rs12413082	10	MSRB2	DM	T	0.32	$1.50 * 10^{-6}$	0.53
rs513154	3	IMPG2	DM	T	0.42	$2.14 * 10^{-6}$	0.52
rs4574762	7	–	RM	G	0.26	$2.29 * 10^{-6}$	0.19
rs1582594	16	–	DM	A	0.32	$3.00 * 10^{-6}$	1.87
rs10514626	1	–	AM	A	0.06	$4.65 * 10^{-6}$	0.38
rs10923806	1	–	RM	G	0.38	$7.24 * 10^{-6}$	2.30
rs2967137	16	–	DM	C	0.37	$8.00 * 10^{-6}$	1.83
rs12088486	1	–	DM	A	0.28	$8.83 * 10^{-6}$	0.56
rs11237644	11	–	AM	A	0.40	$9.24 * 10^{-6}$	1.52
rs2147556	13	–	AM	T	0.29	$9.59 * 10^{-6}$	0.63
rs6592810	11	–	AM	C	0.33	$1.10 * 10^{-5}$	1.54
rs7873259	9	ANKRD19	RM	G	0.27	$1.18 * 10^{-5}$	3.14
rs1563301	6	–	DM	G	0.13	$1.22 * 10^{-5}$	0.50
rs870959	11	–	AM	T	0.40	$1.27 * 10^{-5}$	1.51
rs571391	3	IMPG2	DM	G	0.43	$1.39 * 10^{-5}$	0.55
rs7915479	10	CDH23	DM	T	0.47	$1.47 * 10^{-5}$	1.92
rs10277343	7	–	AM	T	0.28	$1.79 * 10^{-5}$	0.64
rs888808	5	RHOBTB3	AM	G	0.41	$1.94 * 10^{-5}$	0.67
rs6769400	3	–	RM	A	0.48	$1.99 * 10^{-5}$	1.91
rs795602	4	MGST2	AM	T	0.47	$2.11 * 10^{-5}$	0.67
rs4938180	11	IGSF4	RM	T	0.49	$2.16 * 10^{-5}$	1.91
rs6701445	1	TAF5L	DM	T	0.15	$2.24 * 10^{-5}$	1.84

Table 5.5 (continued)

SNP	CHR	Gene	Assoc Model	EA	EAF	p-value	OR
rs10956502	8	FAM49B	DM	G	0.31	$2.25 * 10^{-5}$	0.57
rs4291539	1	–	DM	C	0.42	$2.37 * 10^{-5}$	1.81
rs1538287	1	KCNH1	DM	A	0.18	$2.64 * 10^{-5}$	0.55
rs10491334	5	CAMKIV	DM	T	0.18	$2.88 * 10^{-5}$	0.55
rs2670104	3	–	AM	C	0.36	$2.98 * 10^{-5}$	0.67
rs1556758	10	SORCS1	RM	C	0.45	$3.47 * 10^{-5}$	0.50
rs130329	22	–	RM	A	0.47	$3.52 * 10^{-5}$	0.50
rs1484583	8	–	AM	T	0.27	$3.74 * 10^{-5}$	1.53
rs285097	13	RNF113B	AM	G	0.08	$3.79 * 10^{-5}$	0.47
rs11738302	5	–	DM	T	0.27	$3.81 * 10^{-5}$	0.57
rs846427	7	–	RM	A	0.45	$3.82 * 10^{-5}$	1.94
rs2277472	14	MAMDC1	AM	T	0.09	$3.88 * 10^{-5}$	1.95
rs2495513	1	TMEM61	AM	G	0.20	$3.98 * 10^{-5}$	0.60
rs9315385	13	DCAMKL1	AM	G	0.14	$3.98 * 10^{-5}$	1.74
rs4740391	9	–	AM	G	0.11	$3.99 * 10^{-5}$	0.52
rs9366292	6	–	AM	G	0.25	$4.26 * 10^{-5}$	0.64
rs4727899	7	–	DM	G	0.48	$4.37 * 10^{-5}$	0.55
rs6540664	1	–	AM	A	0.50	$4.45 * 10^{-5}$	1.46
rs4777170	15	–	RM	C	0.43	$4.52 * 10^{-5}$	2.00
rs10513702	3	–	RM	T	0.48	$4.53 * 10^{-5}$	1.85
rs3134204	8	–	AM	G	0.28	$4.83 * 10^{-5}$	1.51
rs135416	22	–	DM	T	0.44	$4.86 * 10^{-5}$	0.57
rs4594173	14	MADMC1	AM	G	0.08	$4.88 * 10^{-5}$	1.94

Table 5.5 (continued)

SNP	CHR	Gene	Assoc Model	EA	EAF	p-value	OR
rs964403	3	SUMF1	RM	A	0.39	$4.91 * 10^{-5}$	2.07
rs1562688	3	–	RM	C	0.43	$4.93 * 10^{-5}$	0.50
rs712773	3	GRM7	AM	G	0.32	$4.98 * 10^{-5}$	0.66
rs697739	6	ATXN1	DM	A	0.34	$5.05 * 10^{-5}$	0.59
rs2111173	12	PTPRO	DM	C	0.39	$5.22 * 10^{-5}$	1.74
rs4282145	4	–	RM	T	0.33	$5.32 * 10^{-5}$	0.38
rs3864051	3	SUMF1	RM	T	0.39	$5.49 * 10^{-5}$	2.07
rs1428689	5	–	RM	C	0.49	$5.51 * 10^{-5}$	1.85
rs81647	16	–	DM	G	0.40	$5.53 * 10^{-5}$	0.58
rs10134056	14	–	RM	A	0.39	$5.61 * 10^{-5}$	0.46
rs2070325	20	LPLUNC4	RM	G	0.30	$5.98 * 10^{-5}$	2.42
rs2905476	9	–	RM	T	0.28	$6.35 * 10^{-5}$	0.29
rs7583529	2	CFLAR	RM	A	0.21	$6.45 * 10^{-5}$	3.15
rs7842001	8	–	RM	G	0.26	$7.12 * 10^{-5}$	0.32
rs969845	12	–	RM	A	0.18	$7.22 * 10^{-5}$	5.87
rs2354314	4	–	RM	T	0.43	$7.81 * 10^{-5}$	0.51
rs2073586	11	ABCC8	RM	T	0.40	$8.15 * 10^{-5}$	0.48
rs4505466	2	SH3BP4	RM	T	0.40	$8.31 * 10^{-5}$	2.06
rs1584547	14	–	RM	T	0.24	$8.52 * 10^{-5}$	2.96
rs3102484	8	–	RM	G	0.49	$9.20 * 10^{-5}$	1.82
rs731287	13	–	RM	T	0.28	$9.88 * 10^{-5}$	2.51

Assoc Model: *AM* allelic model [1 df]), *DM* dominant model [1, df]; *RM* recessive model [1 df]), *OR* odds ratio

came from the Illumina control database ($n = 673$). For replication two additional sets were used of 253 and 60 centenarians and 341 and 2863 population controls [30]. Four different genetic models (general/genotypic, allelic/additive, recessive and dominant) were investigated. A single SNP, rs2075650, in *APOE/TOMM40* reached genome-wide significance [30]. Table 5.6 contains the top 17 SNPs with a p-value $< 10^{-4}$ in the additive model. They further explored the hypothesis that different sets of SNPs that are associated with exceptional longevity, although with moderate effects, may jointly characterize the genetic predisposition to exceptional longevity [31, 32]. The authors included 281 predictive SNPs in the genetic risk profiles reaching 89% sensitivity and specificity for predicting centenarian status in the discovery sample [30]. In the replication samples the sensitivity was 60% and specificity was 58%. However, this set was slightly younger. In the older subjects sensitivity increased to 85%. The 281 predictive SNPs are located in 130 genes. Some of these genes are known for progeroid (premature aging-like) syndromes, like *LMNA* (Huthcinson-Gilford syndrome) and *WRN* (Werner's Syndrome) [33, 34]. 38 of the 130 genes were linked to AD in literature, 42 to dementia and 38 to tauopathies. The fact that so many genes play a role in dementia is consistent with the epidemiologic finding that dementia is absent or markedly delayed amongst centenarians [35]. Cluster analysis identified 26 groups of 8 to 94 centenarians (90% of the discovery set) with similar genetic risk profiles [30]. The genetic risk profiles associated with each cluster represent different genetic signatures of exceptional longevity. Some of the genetic signatures were significantly associated with different life spans, while others were associated with varying prevalences and ages of onset of various age-related diseases [30].

Alternative Approach

A different approach to study longevity was employed by Walter *et al* [36]. They employed a prospective follow-up design to investigate time to death as a continuous outcome (all-cause mortality) using a Cox proportional hazard model. The GWAS study included 25,007 participants including 8444 deaths. Mean follow-up time was 10.6 years. Mean age at death was 81.1 years of age. 14 SNPs were associated with time to death at a suggestive threshold of p-value $< 1 * 10^{-5}$ (Table 5.7). The strongest association was for rs4936894 (chromosome 11, near *VWA5A*) with a p-value of $3.4 * 10^{-7}$ [36]. Replication for the top 5 SNPs was sought in 4 independent samples ($n = 10{,}411$, deaths $= 1295$). None of the SNPs were consistently replicated. In the combined meta-analysis only rs1425609 near *OTOL1* showed a stronger association compared to discovery (p-value $= 1.61 * 10^{-6}$) [36]. Pathway analysis was applied to investigate the SNPs with p-values $< 1 * 10^{-3}$ in more detail. Relevant biological processes overrepresented in the results were developmental processes, neuronal activities, signal transduction, neurogenesis, ectoderm development and cell adhesion.

Table 5.6 Top hits of GWAS performed by Sebastiani et al [30]

SNP	Gene	EA	EAF	PVAL.GA	PVAL.AA	PVAL.DA	PVAL.RA
rs2075650	TOMM40/APOE	G	0.14	$2.89*10^{-9}$	$2.36*10^{-10}$	$2.50*10^{-4}$	$1.03*10^{-8}$
rs12629971	EIF4E3	G	0.82	$1.95*10^{-5}$	$1.90*10^{-6}$	$7.44*10^{-6}$	0.025
rs4977756	NA	G	0.37	$3.87*10^{-5}$	$7.97*10^{-6}$	$1.44*10^{-4}$	$5.88*10^{-4}$
rs6801173	EIF4E3	G	0.80	$6.81*10^{-5}$	$8.16*10^{-6}$	$2.20*10^{-5}$	0.041
rs1456669	NA	C	0.89	$1.94*10^{-5}$	$8.60*10^{-6}$	$4.05*10^{-6}$	0.374
rs4802234	CEACAM16	C	0.52	$3.06*10^{-5}$	$9.22*10^{-6}$	$3.25*10^{-5}$	0.003
rs1063192	CDKN2A	G	0.41	$9.70*10^{-5}$	$1.66*10^{-5}$	$6.33*10^{-4}$	$4.53*10^{-4}$
rs915179	LMNA	G	0.39	$1.03*10^{-4}$	$2.37*10^{-5}$	0.001	$3.18*10^{-4}$
rs2758331	SOD2	C	0.55	$1.31*10^{-4}$	$2.93*10^{-5}$	$4.39*10^{-4}$	0.001
rs4073968	NA	G	0.76	$5.38*10^{-5}$	$3.64*10^{-5}$	$1.24*10^{-5}$	0.129
rs1412832	NA	G	0.28	$2.18*10^{-4}$	$4.55*10^{-5}$	0.002	$6.89*10^{-4}$
rs9557276	CLYBL	C	0.50	$2.62*10^{-4}$	$4.65*10^{-5}$	$3.21*10^{-4}$	0.004
rs4722094	NA	G	0.18	$8.66*10^{-5}$	$5.00*10^{-5}$	0.239	$1.93*10^{-5}$
rs10483493	TTC6	C	0.25	$3.32*10^{-4}$	$5.96*10^{-5}$	0.004	$6.17*10^{-4}$
rs6997589	SH2D4A	G	0.22	$3.70*10^{-4}$	$6.25*10^{-5}$	0.017	$2.31*10^{-4}$
rs3763305	BTNL2	G	0.96	$1.62*10^{-4}$	$6.27*10^{-5}$	$4.51*10^{-5}$	0.810
rs277432	NA	G	0.62	$3.86*10^{-4}$	$9.12*10^{-5}$	0.001	0.002

PVAL.GA *p*-value for genotype association, *PVAL.AA* *p*-value for allelic association, *PVAL.DA* *p*-value for dominant association, *PVAL.RA* *p*-value for recessive association

Table 5.7 Top hits of GWAS performed by Walter et al [36]

SNP	Chr	Gene	EA	EAF	HR	p-value
rs4936894	11	VWA5A	A	0.23	1.11	$3.38*10^{-7}$
rs1425609	3	OTOL1	A	0.38	0.92	$1.46*10^{-6}$
rs766903	12	BIN2	A	0.83	0.90	$1.61*10^{-6}$
rs12042640	1	ATG4C	T	0.28	1.09	$1.71*10^{-6}$
rs17149227	7	HIP1	T	0.96	0.79	$3.56*10^{-6}$
rs3128591	9	COL5A1	A	0.75	0.92	$3.64*10^{-6}$
rs11582903	1	LMO4	A	0.15	1.12	$3.94*10^{-6}$
rs4850695	2	HECW2	A	0.77	1.09	$4.62*10^{-6}$
rs10259086	7	ORC5L	T	0.69	1.08	$5.16*10^{-6}$
rs2769255	1	KCNQ4	T	0.37	1.08	$5.17*10^{-6}$
rs17291546	6	LOC340156	A	0.96	0.82	$7.65*10^{-6}$
rs12606100	18	NETO1	T	0.20	1.11	$8.72*10^{-6}$
rs1274214	11	GRAMD1B	T	0.50	0.93	$8.87*10^{-6}$
rs10811679	9	SMARCA2	T	0.33	1.08	$9.53*10^{-6}$

HR hazard ratio

The Future of GWAS on Longevity

Published GWAS on longevity have so far failed to identify any new robust associations with longevity that have replicated over studies. The only loci robustly associated stem from candidate genes *APOE* and *FOXO3a* [13, 14, 16, 18, 24]. Though GWAS has proven to be a powerful approach to unravel the genetics of many complex traits, the longevity phenotype remains resistant to the efforts to uncover new genetic associations.

A reason for not finding any replicated associations for longevity could be the sheer complexity of the phenotype. Even centenarians fall into different groups in terms of age of onset of age-related diseases: survivors (onset of aging disease < 80 years), delayers (onset of aging disease between 80 and 100 years) and escapers (onset of aging disease > 100 years) [37]. Taking a younger age-cutoff for longevity cases (85+ or 90+), the number of cases will increase, which is very relevant for prospective cohort studies. However, along with an increase in power, the heterogeneity is expected to increase. The key to success in GWAS of other traits has been to increase samples size, ignoring issues of heterogeneity, which also occur in other complex outcomes such a blood pressure and cardiovascular disease. Without a doubt progress may be achieved by pooling the present studies in a joint analysis and adding as much as possible new studies available to increase the statistical power. Despite the robustness of GWAS to heterogeneity, there is a definite need to harmonize the longevity phenotype across studies. As seen in Table 5.1, almost every study investigating longevity uses different criteria for either longevity cases or the comparison group. This makes comparing the results between studies very difficult.

Why have we not identified new genes for longevity by GWAS? It has been argued that it may require a great number of 'protective' genes all with very small effects to have a genetic advantage to achieve longevity [30]. This model is also referred to as the infinitesimal model [38]. We have recently tested the infinitesimal model in the Rotterdam study and found that 81.3 % of the heritability in longevity defined as survival to age 90+ years is explained by common variants. Such a mechanism has been proposed for other complex traits including height. Though for highly heritable traits like height these genes are uncovered [39], in a trait like longevity this may require extremely large samples size to achieve sufficient statistical power, which have not been achieved yet. Using biomarkers of aging might be a more fruitful pursuit for finding associations with longevity. Unfortunately, no good biomarkers of aging currently exist, though many have been proposed [40]. Telomere length, a marker of cellular senescence, is one of the previously proposed biomarkers of aging [41] and has already proven successful in identifying genes for this trait [42]. As of yet, these genes have been associated with cardiovascular disease [42] but have not been found to associate with longevity [43]. These findings are not final as only a very small percentage of the telomere length variance can be explained by the currently uncovered genes [42].

An alternative approach to solving the heterogeneity issue in longevity is addressing healthy aging, as captured in a healthy aging index (HAI) [44]. The HAIs may include markers of 5 various organ systems that are known to predict mortality and disability. The systems included are vascular (carotid wall thickness), brain (white matter grade on MRI), kidney (cystatin-c), lung (forced vital capacity) and metabolic (fasting glucose levels) [44]. The HAI is able to distinguish a wide risk gradient, but is most remarkable for its advantage in identifying low risk individuals. As a single factor, the HAI prediction of mortality is similar in magnitude to age itself. When entered together, age remained partly independent, but the HAI explained 40 % of the effect of age [44].

Another potential reason for not finding any solid associations with the longevity phenotype stems back to the old debate of the role of common versus rare variants [45]. The common disease, common variant (CDCV) hypothesis states that common traits are caused by common variants with small effect sizes [46]. This theory is essentially targeted in GWAS. However, assuming a role of common variants may be an oversimplification of the true genetic architecture behind complex traits as longevity [47]. An alternate hypothesis is that rare phenotypes such as extreme longevity may be explained by rare variants with large effects that explain the high heritability and the clustering of nonagenarians and centenarians in families [48]. GWAS is not suited to identify rare variants as they are often not properly tagged by the variants present on genotyping arrays. Exome sequencing, or even genome sequencing, might help in uncovering such rare variants [49].

Although findings of GWAS to date have been disappointing, as discussed in this chapter there is ample opportunity to improve the statistical power of studies to find common variants with small effects that appear to explain over 80 % of the heritability in the Rotterdam study. Collaboration between various consortia is most likely the fastest way forward to success and may likely require some a priori titration on the definition of longevity cases and controls with the view to maximize the statistical power.

References

1. Oeppen J, Vaupel JW (2002) Demography. Broken limits to life expectancy. Science 296(5570):1029–1031
2. Vaupel JW et al (1998) Biodemographic trajectories of longevity. Science 280(5365):855–860
3. Suzman R, Riley MW (1985) Introducing the “oldest old”. Milbank Mem Fund Q Health Soc 63(2):177–186
4. Arias E (2011) United States life tables, 2007. Natl Vital Stat Rep 59(9):1–60
5. vB Hjelmborg J et al (2006) Genetic influence on human lifespan and longevity. Hum Genet 119(3):312–321
6. Herskind AM et al (1996) The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870–1900. Hum Genet 97(3):319–323
7. McGue M et al (1993) Longevity is moderately heritable in a sample of Danish twins born 1870–1880. J Gerontol 48(6):B237–244
8. Kerber RA et al (2001) Familial excess longevity in Utah genealogies. J Gerontol A Biol Sci Med Sci 56(3):B130–139

9. Mitchell BD et al (2001) Heritability of life span in the Old Order Amish. Am J Med Genet 102(4):346–352
10. Murabito JM, Yuan R, Lunetta KL (2012) The search for longevity and healthy aging genes: insights from epidemiological studies and samples of long-lived individuals. J Gerontol A Biol Sci Med Sci 67(5):470–479
11. McIlhany ML, Shaffer JW, Hines EA Jr (1975) The heritability of blood pressure: an investigation of 200 pairs of twins using the cold pressor test. Johns Hopkins Med J 136(2):57–64
12. Pilia G et al (2006) Heritability of cardiovascular and personality traits in 6148 Sardinians. PLoS Genet 2(8):e132
13. Anselmi CV et al (2009) Association of the FOXO3A locus with extreme longevity in a southern Italian centenarian study. Rejuvenation Res 12(2):95–104
14. Bathum L et al (2006) Apolipoprotein e genotypes: relationship to cognitive functioning, cognitive decline, and survival in nonagenarians. J Am Geriatr Soc 54(4):654–658
15. Beekman M et al (2013) Genome-wide linkage analysis for human longevity: genetics of healthy aging study. Aging Cell 12(2):184–193
16. Flachsbart F et al (2009) Association of FOXO3A variation with human longevity confirmed in German centenarians. Proc Natl Acad Sci U S A 106(8):2700–2705
17. Gerdes LU et al (2000) Estimation of apolipoprotein E genotype-specific relative mortality risks from the distribution of genotypes in centenarians and middle-aged men: apolipoprotein E gene is a frailty gene, not a longevity gene. Genet Epidemiol 19(3):202–210
18. Willcox BJ et al (2008) FOXO3A genotype is strongly associated with human longevity. Proc Natl Acad Sci U S A 105(37):13987–13992
19. Beekman M et al (2010) Genome-wide association study (GWAS)-identified disease risk alleles do not compromise human longevity. Proc Natl Acad Sci U S A 107(42):18046–18049
20. Ganna A et al (2013) Genetic determinants of mortality. Can findings from genome-wide association studies explain variation in human mortality? Hum Genet 132(5):553–561
21. Newman AB et al (2010) 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 65(5):478–487
22. Chi H et al (2000) Targeted deletion of Minpp1 provides new insight into the activity of multiple inositol polyphosphate phosphatase in vivo. Mol Cell Biol 20(17):6496–6507
23. Deelen J et al (2011) Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. Aging Cell 10(4):686–698
24. Schachter F et al (1994) Genetic associations with human longevity at the APOE and ACE loci. Nat Genet 6(1):29–32
25. Christensen K, Johnson TE, Vaupel JW (2006) The quest for genetic determinants of human longevity: challenges and insights. Nat Rev Genet 7(6):436–448
26. Bertram L et al (2007) Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. Nat Genet 39(1):17–23
27. Nebel A et al (2011) A genome-wide association study confirms APOE as the major gene influencing survival in long-lived individuals. Mech Ageing Dev 132(6–7):324–330
28. Malovini A et al (2011) Association study on long-living individuals from Southern Italy identifies rs10491334 in the CAMKIV gene that regulates survival proteins. Rejuvenation Res 14(3):283–291
29. Levy D et al (2007) Framingham heart study 100 k project: genome-wide associations for blood pressure and arterial stiffness. BMC Med Genet 8(Suppl 1):S3
30. Sebastiani P et al (2012) Genetic signatures of exceptional longevity in humans. PLoS One 7(1):e29848
31. Hekimi S (2006) How genetic analysis tests theories of animal aging. Nat Genet 38(9):985–991
32. Terry DF et al (2008) Disentangling the roles of disability and morbidity in survival to exceptional old age. Arch Intern Med 168(3):277–283
33. Gray MD et al (1997) The Werner syndrome protein is a DNA helicase. Nat Genet 17(1):100–103

34. Eriksson M et al (2003) Recurrent de novo point mutations in lamin A cause Hutchinson–Gilford progeria syndrome. Nature 423(6937):293–298
35. Hitt R et al (1999) Centenarians: the older you get, the healthier you have been. Lancet 354(9179):652
36. Walter S et al (2011) A genome-wide association study of aging. Neurobiol Aging 32(11):2109 e15–28
37. Evert J et al (2003) Morbidity profiles of centenarians: survivors, delayers, and escapers. J Gerontol A Biol Sci Med Sci 58(3):232–237
38. Gibson G (2011) Rare and common variants: twenty arguments. Nat Rev Genet 13(2):135–145
39. Lango AH et al (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467(7317):832–838
40. Johnson TE (2006) Recent results: biomarkers of aging. Exp Gerontol 41(12):1243–1246
41. von Zglinicki T, Martin-Ruiz CM (2005) Telomeres as biomarkers for ageing and age-related diseases. Curr Mol Med 5(2):197–203
42. Codd V et al (2013) Identification of seven loci affecting mean telomere length and their association with disease. Nat Genet 45(4):422–427, 427e1–2
43. Deelen J et al (2014) Leukocyte telomere length associates with prospective mortality independent of immune-related parameters and known genetic markers. Int J Epidemiol 43(3):878–886
44. Newman AB et al (2008) A physiologic index of comorbidity: relationship to mortality and disability. J Gerontol A Biol Sci Med Sci 63(6):603–609
45. Schork NJ et al (2009) Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev 19(3):212–219
46. Reich DE, Lander ES (2001) On the allelic spectrum of human disease. Trends Genet 17(9):502–510
47. Maher B (2008) Personal genomes: the case of the missing heritability. Nature 456(7218):18–21
48. Pritchard JK (2001) Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet 69(1):124–137
49. Cirulli ET, Goldstein DB (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet 11(6):415–425