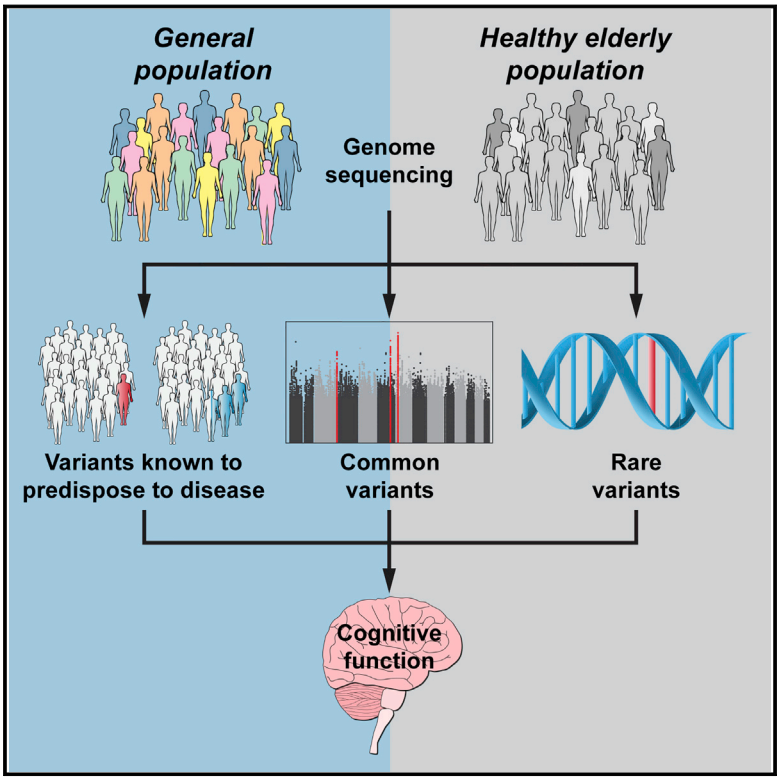
Resource

Cell

Whole-Genome Sequencing of a Healthy Aging Cohort

Graphical Abstract



Highlights

- Healthy aging is a complex polygenic trait related but distinct from longevity
- Healthy aging is associated with decreased genetic risk for select diseases
- Healthy aging is potentially linked to protection against cognitive decline
- Genome data are made available for further analysis

Authors

Galina A. Erikson, Dale L. Bodian, Manuel Rueda, ..., John E. Niederhuber, Eric J. Topol, Ali Torkamani

Correspondence

etopol@scripps.edu (E.J.T.), atorkama@scripps.edu (A.T.)

In Brief

The genomic characterization of humans that age without developing diseases suggests that healthy aging is a distinct phenotype from exceptional longevity and that it may be enriched with disease-protective genetic factors, such as resistance against cognitive decline.

Erikson et al., 2016, Cell *165*, 1–10
May 5, 2016
http://dx.doi.org/10.1016/j.cell.2016.03.022

CellPress

Please cite this article in press as: Erikson et al., Whole-Genome Sequencing of a Healthy Aging Cohort, Cell (2016), http://dx.doi.org/10.1016/j.cell.2016.03.022

Cell

Whole-Genome Sequencing of a Healthy Aging Cohort

Galina A. Erikson,[1,5] Dale L. Bodian,[2,5] Manuel Rueda,[1] Bhuvan Molparia,[1] Erick R. Scott,[1] Ashley A. Scott-Van Zeeland,[4] Sarah E. Topol,[1] Nathan E. Wineinger,[1] John E. Niederhuber,[2,3] Eric J. Topol,[1,6,*] and Ali Torkamani[1,4,6,*]
[1]The Scripps Translational Science Institute, Scripps Health, and The Scripps Research Institute, La Jolla, CA 92037, USA
[2]Inova Translational Medicine Institute, Inova Health System, Falls Church, VA 22042 USA
[3]Department of Oncology, School of Medicine, Johns Hopkins University, Baltimore, MD 21205, USA
[4]Cypher Genomics, Inc., San Diego, CA 92121, USA
[5]Co-first author
[6]Co-senior author
*Correspondence: etopol@scripps.edu (E.J.T.), atorkama@scripps.edu (A.T.)
http://dx.doi.org/10.1016/j.cell.2016.03.022

SUMMARY

Studies of long-lived individuals have revealed few genetic mechanisms for protection against age-associated disease. Therefore, we pursued genome sequencing of a related phenotype—healthy aging—to understand the genetics of disease-free aging without medical intervention. In contrast with studies of exceptional longevity, usually focused on centenarians, healthy aging is not associated with known longevity variants, but is associated with reduced genetic susceptibility to Alzheimer and coronary artery disease. Additionally, healthy aging is not associated with a decreased rate of rare pathogenic variants, potentially indicating the presence of disease-resistance factors. In keeping with this possibility, we identify suggestive common and rare variant genetic associations implying that protection against cognitive decline is a genetic component of healthy aging. These findings, based on a relatively small cohort, require independent replication. Overall, our results suggest healthy aging is an overlapping but distinct phenotype from exceptional longevity that may be enriched with disease-protective genetic factors.

INTRODUCTION

Age-associated diseases account for two-thirds of human deaths globally and 90% of all deaths in industrialized nations. Management of age-related disease is a major source of disability and health care costs, and thus management of age-associated disease is a major driver of medical research and health policy (Goldman et al., 2013). Age-standardized death rates for the leading causes of death have declined substantially over the past half century due to an improved understanding of behavioral risks and more effective medical interventions (Ma et al., 2015). However, as average life expectancy improves, a shift in increased death due to frailty and chronic degenerative disease has been observed (Lozano et al., 2012). Thus, overall longevity is likely due to the avoidance of multiple vulnerabilities, such as death due to infectious disease and severe early onset genetic diseases, mid-life death due to chronic and potentially sporadic diseases, and late-life death due to degenerative diseases and frailty, each of which is differentially influenced by the interplay of genetics and environment. Moreover, technologic progress with pharmacotherapy for cancer and with implantable devices, such as defibrillators and prosthetic joints, can promote longevity as a function of medical interventions per se.

Age at death in adulthood has a moderate genetic component overall, with a heritability of approximately 25% (Murabito et al., 2012). Heritability of longevity increases with age, with a negligible genetic contribution to survival up to approximately 60 years of age, after which an increasing genetic component to survival is observed (Brooks-Wilson, 2013; Christensen et al., 2006). Most genetic studies of aging have focused on long-lived individuals, typically defined as centenarians 100 years or older, who may have had exceptional survival due to medical interventions (Murabito et al., 2012). A number of genetic associations with exceptional longevity have been made (Atzmon et al., 2006; Bojesen and Nordestgaard, 2008; Hurme et al., 2005; Kuningas et al., 2007; Melzer et al., 2007; Pawlikowska et al., 2009; Sanders et al., 2010; Suh et al., 2008; Willcox et al., 2008), with only markers at *APOE* and *FOXO3A* being well replicated (Murabito et al., 2012). Overall, the results of genetic and epidemiological longevity studies suggest aging is a complex trait and that achievement of exceptional longevity may not best capture the genetics of resistance to or delay of age-associated disease (Christensen et al., 2006).

In this light, we pursued a genomic study of an alternate but related aging phenotype—healthy aging—in order to expose its potential to uncover genetic factors for protection against age-associated disease. It is important to differentiate longevity from our healthy aging phenotype, which, as we have defined it for our healthy aging cohort (Wellderly), attempts to understand the genetics of disease-free aging in humans without medical interventions. Toward this end, we performed whole-genome sequencing (WGS) of the Wellderly and compared their genetic characteristics to an ethnicity-matched population control. Our

Table 1. Wellderly Demographic Data

Characteristic	Wellderly Cohort	US Population	ITMI Cohort
Median age	84.2 years (±9.3 years)	70+ years[a] or 80+ years[b]	33.3 years
Gender[a]	male = 39.3% (±1.3%)	male = 36.3%	male = 52.7%
	female = 60.7% (±1.3%)	female = 63.7%	female = 47.3%
Median height[b]	male = 69" (±5.0")	male = 68"	male = N/A
	female = 63" (±5.7")	female = 63"	female = 66"
Median weight[b]	male = 168 lbs (±47 lbs)	male = 190 lbs	male = na
	female = 132 lbs (±45 lbs)	female = 160 lbs	female = 142 lbs
Ever smoker[b]	male = 61% (±2.6%)	male = 54%	male = 31.5%
	female = 42% (±2.6%)	female = 43%	female = 26.2%
Exercise[b]	66.8% (±2.5%)	44.1%	na
Education[b]			
No high school	0.5% (±0.4%)	24.6%	8.0%
High school	16.8% (±2.0%)	56.2%	17.2%
Bachelor's	55.0% (±2.7%)	11.9%	38.7%
Advanced	27.7% (±2.4%)	7.3%	36.1%

The 95% confidence interval is in parentheses. na, not applicable.
[a]US population 80+ years.
[b]US population 70+ years.

findings suggest that healthy aging is associated with a disease-protective genetic profile that overlaps with but differs from that observed in exceptional longevity cohorts. These findings include no enrichment of true longevity variants, a lower genetic risk from common susceptibility alleles for Alzheimer and coronary artery disease, and no decrease in the rate of rare pathogenic variants. We identify suggestive common and rare variant genetic associations that implicate genetic protection against cognitive decline in healthy aging. Our data are made available for the discovery of additional disease protective genetic factors by the research community.

RESULTS

Wellderly Cohort Definition and Demographics

The Wellderly phenotype is defined as individuals who are >80 years old with no chronic diseases and who are not taking chronic medications. Individuals with any of the following phenotypes were excluded from enrollment: autoimmune disease, blood clots, cancer (except basal and squamous cell carcinoma), type I or II diabetes, dementia, myocardial infarction, renal failure, and stroke. The enrollment procedure is described in the Experimental Procedures, and the specific ICD code, CPT code, lab value, and medication status exclusion criteria are provided in the Supplemental Experimental Procedures.

To date, a total of 1,354 Wellderly individuals ranging from 80 to 105 years old have enrolled in our study. Demographic characteristics of these Wellderly individuals were compared with the characteristics of the general US population of 70+- or 80+-year-old individuals (depending upon availability) utilizing data from the US 2010 Census data or the National Health and Nutrition Examination Survey (Table 1). The Wellderly cohort individuals differ from these age-matched controls in the following ways: (1) they are comprised of a slightly but significantly higher proportion of male individuals overall, (2) they contain a small but significantly elevated rate of male smokers, (3) they exercise significantly more frequently, (4) they are leaner on average, but not significantly so, due to a wide weight distribution (Figure 1C), and (5) they have attained a significantly higher level of education relative to the general population. The majority of individuals were enrolled between the age of 80 and 85 with similar overall age distributions for female and male Wellderly individuals (Figures 1A and 1B).

Healthy Aging versus Longevity

It is possible that the Wellderly phenotype simply represents the tail end of a stochastic age-related disease incidence process in which the Wellderly individuals are simply "lucky" and by chance do not develop disease. If this were true, one would expect that the siblings of Wellderly individuals should live no longer than the average individual, and that no shared factor, whether genetic or environmental, should have an influence on Wellderly sibling survival.

To address this possibility, we performed survival analysis comparing the US Social Security 1920s birth cohort life table versus Wellderly siblings—the majority being deceased individuals who were obviously genetically similar to the Wellderly individuals. Survival analysis was initiated at 10 years of age to eliminate a significant excess of childhood death in the 1920s birth cohort life table (Bell and Miller, 2005) relative to that reported by the Wellderly individuals. Survival analysis demonstrated significant increased survival (log-rank test p value = 3.67×10^{-8}) in Wellderly siblings relative to the 1920s birth cohort (Figure 1D). Inspection of the survival curve reveals a strong survival advantage in middle age to middle old age (40–79 years), a limited survival advantage from 80–90 years of age, and no survival advantage after 90+ years of age. 83% of Wellderly siblings survive to age 70 yrs, whereas 75% of the

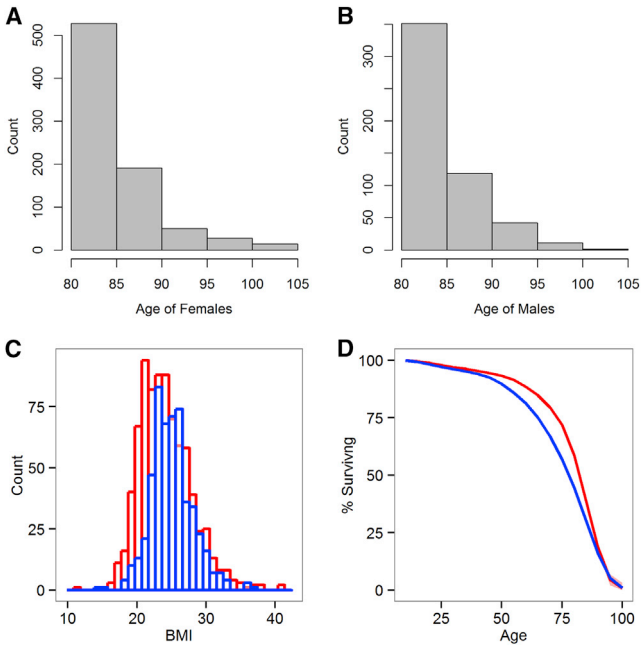


Figure 1. Wellderly Demographic Characteristics

(A) The age distribution of Wellderly female individuals.

(B) The age distribution of male Wellderly individuals.

(C) The BMI distribution of male (blue) and female (red) Wellderly individuals.

(D) Survival curves for the Wellderly siblings (red) and the expected survival of a 1920s birth cohort (blue). A significant difference in survival was observed (log-rank test p value = 3.67×10^{-8}) in middle age (40–79 years), but not overall longevity. The 95% confidence interval is provided for the Wellderly siblings in red shading (only visible at the end of the survival curve).

1920s birth cohort survives to 70 yrs. This suggests Wellderly individuals avoid or delay the onset of middle-age onset diseases through some shared factor. This analysis cannot differentiate between genetic, environmental, or behavioral characteristics shared by Wellderly individuals and their siblings, but indicates the phenotype is not simply due to chance.

Genome Sequencing and Variant Filtration

To determine whether genetic factors underlie the Wellderly phenotype, WGS of 600 Wellderly individuals was performed using the Complete Genomics platform (Complete Genomics) (see the Experimental Procedures) (Figure S1). These genomes were compared to 1,507 adult members of the Inova Translational Medicine Institute (ITMI) pre-term birth cohort also sequenced on the Complete Genomics platform (Bodian et al., 2014). Individuals from these cohorts were filtered to retain only individuals of $\geq$95% European ancestry and a maximum relatedness of 12.5% (see the Experimental Procedures). After cohort filtration, 511 Wellderly individuals and 686 ITMI individuals were carried forward to downstream analyses. Variants were combined within and across cohorts and filtered stringently for downstream statistical analyses (see the Experimental Procedures). Genomic variant filtration resulted in reduction of the initial ~57 million total raw variants to 24,205,551 variants after filtration. Table S1 displays the number of variants removed by each filtration step. The count of variants by type and coding impact is presented in Table S2. Our variant filtration strategy was reinforced by observing no evidence of genomic inflation in the genome-wide association results (see below; $\lambda = 0.98$) (Reich and Goldstein, 2001).

Genetic Longevity and Disease Risk

First, we explored whether previously reported longevity variants may underlie the Wellderly phenotype (Atzmon et al., 2006; Bojesen and Nordestgaard, 2008; Hurme et al., 2005; Kuningas et al., 2007; Melzer et al., 2007; Pawlikowska et al., 2009; Sanders et al., 2010; Suh et al., 2008; Willcox et al., 2008). No deviation in allele frequency of longevity variants was observed between the Wellderly cohort and the ITMI cohort or the 1000 Genomes European individuals (Sudmant et al., 2015) (Table 2). No trend in allele frequency deviations implying even minimal enrichment of longevity variants was observed, suggesting the Wellderly healthy aging phenotype is likely distinct from traditionally defined longevity.

Next, we considered whether the Wellderly have decreased genetic risk of common disease by calculating genetic risk scores for the Wellderly and ITMI individuals for the top five leading causes of death with a genetic component—heart disease, cancer, stroke, Alzheimer disease, and diabetes (CDC Vital Statistics)—using genetic loci identified from the latest large genome-wide association studies and/or meta-analyses (Al-Tassan et al., 2015; Deloukas et al., 2013; Kilarski et al., 2014; Lambert et al., 2013; Michailidou et al., 2015; Mahajan et al., 2014; Seshadri et al., 2010; Timofeeva et al., 2012; Wolpin et al., 2014; Zheng et al., 2008) (Table 3). The five most common and deadly cancer types were considered—breast, colon, lung, pancreatic, and prostate—comprising over 75% of cancer deaths in the United States. We observed no difference in cancer, stroke, or type 2 diabetes genetic risk. On the other hand, the Wellderly have a significantly lower genetic risk for Alzheimer disease (p value = 9.84×10^{-4}) and coronary artery disease (p value = 2.54×10^{-3}). Individual marker level differences are presented in Table S3.

The difference in Alzheimer disease genetic risk is strongly influenced by the *APOE*-ε4 marker (rs2075650; p value = 7.02×10^{-4}), though a trend for marginal differences in allele frequency, in the direction consistent with decreased Alzheimer risk, can be observed for the majority of the 18 Alzheimer disease risk variants (Table S3). The next most significant single marker (rs11218343; p value = 2.96×10^{-2}) was also the next strongest genetic risk marker for Alzheimer disease in the published GWAS meta-analysis—at the *SORL1* locus—suggesting the magnitude of the deviation in allele frequency between the Wellderly and ITMI cohort is related to the strength of the genetic risk locus in mediating disease (correlation between –log10 single marker p value and genetic risk weight = 0.89, p value < 0.0001).

Similarly, 83 of the 141 previously identified markers for coronary artery disease variants displayed allele frequency differences consistent with decreased genetic risk for coronary artery disease in the Wellderly (p value = 0.043). The correlation between deviation in allele frequency between the Wellderly and ITMI cohort and the strength of the genetic risk locus was again significant, though not as strong as that observed for Alzheimer disease (r = 0.15, p value = 0.04). The strongest differences were observed at *APOE* (rs2075650; p value = 7.02×10^{-4}), *PECAM1*

Table 2. Allele Frequencies for Known Longevity Variants

Variant	Gene	Longevity Allele	Wellderly AF	ITMI AF	p Value	1000G CEU AF
rs2802292	FOXO3A	G	0.41	0.38	0.059	0.43
rs1935949	FOXO3A	A	0.31	0.30	0.431	0.35
rs3758391	SIRT1	T	0.32	0.31	0.586	0.33
rs5882	CETP	G	0.33	0.33	0.726	0.33
rs1042522	TP53	C	0.74	0.74	0.677	0.72
rs1800795	IL6	C	0.40	0.40	0.761	0.42
rs2811712	CDKN2A	G	0.11	0.10	0.727	0.10
rs34516635	IGF1R	A	0.01	0.01	0.394	0.00
rs2542052	APOC3	C	0.61	0.59	0.433	0.61
rs3803304	AKT1	C	0.73	0.74	0.511	0.73

Comparison of allele frequency for known longevity variants. The frequency of an allele associated with a longer lifespan is displayed.

(rs2070783; p value = 1.26×10^{-2}), *SMG6* (rs2281727; p value = 1.35×10^{-2}), and *SCARB1* (rs11057841; p value = 1.69×10^{-2}).

Correspondingly, the single strongest risk markers for Alzheimer disease (*APOE* above) and coronary artery disease (9p21 rs1333049; p value = 3.57×10^{-2}) (Wellcome Trust Case Control, 2007) demonstrated deviations in allele frequency between the Wellderly and ITMI cohorts, whereas no significant difference was observed for the strongest risk markers for diabetes or cancer: *FTO* (rs9939609; p value = 0.809) (Wellcome Trust Case Control, 2007), *TCF7L2* (rs7903146; p value = 0.728) (Wellcome Trust Case Control, 2007), and 8q24 (rs6983267; p value = 0.421) (Wokolorczyk et al., 2008).

Common Variant Genome-wide Association Study

Next, we globally considered whether any common variants (minor allele frequency >5%) were associated with the Wellderly phenotype. We performed a WGS-based genome-wide association study (GWAS) using a logistic model with principal component correction (Figure 2A) to account for any remaining population stratification after restriction to individuals with $\geq$95% European ancestry, though inspection of the principal component plots demonstrates the cohorts are well balanced (Figure 2A). The quantile-quantile (QQ) plot of association results demonstrates no genomic inflation ($\lambda = 0.98$) (Figure 2B). No associations were observed at the genome-wide significance level (p value < 5×10^{-8}, Figure 2C), though a number of interesting marginally significant peaks were observed. The full list of top hits is provided in Table S4.

The top region is a large linkage block at MHC locus 6p22.1 containing lead SNPs rs13217620 (chr6:27,653,120) (p value = 6.1×10^{-7}) (Figure 3A) and rs41266839 (chr6:26,409,890) (p value = 4.1×10^{-6}) (Figure 3B). This region contains a handful of SNPs that have been replicated and confidently associated with cognitive traits, including rs1056667, associated with cognitive performance (Rietveld et al., 2013, 2014), rs13194053, and associated with schizophrenia and bipolar disorder (Purcell et al., 2009; Shi et al., 2009). The two lead SNPs (rs13217620 and rs41266839) from our study are in perfect disequilibrium (D' = 1.0) with the aforementioned cognitive trait associated SNPs (rs1056667 and rs13194053, respectively) (Figures 3A and 3B).

The second sub-significant region included lead SNP rs156033 (p value = 1.7×10^{-6}), in the 5q31.1 region, which contains *SLC22A4* and multiple variants (rs7727544, rs419291, rs11950562, rs273914, rs272889, rs272869, and rs274567) strongly associated with carnitine and carnitine-related metabolite levels (Shin et al., 2014) (Figure 3C). Again, our lead SNP displays strong disequilibrium with all of these previously associated SNPs, especially rs11950562 (D' = 1.0) (Figure 3C), which was specifically associated with isovalerylcarnitine levels (p value = 2×10^{-41}) (Shin et al., 2014).

Finally, our third sub-significant region included lead SNP rs10209741 (p value = 7.0×10^{-6}) in the 2q36.1 region containing *KCNE4* and nearby (11.4 kb) rs895767, which was previously associated with cognitive decline (Zhang and Pierce, 2014) (Figure 3C). Again, rs10209741 and rs895767 are in perfect disequilibrium (D' = 1.0).

While no individual locus is genome-wide significant, the observation that at least two of three of our top GWAS hits are associated with cognitive function is unlikely to be due to chance. With the cautious assumption that ~100 loci have been previously associated with cognitive function via GWAS, the probability that two of our three top GWAS hits land among these ~100 cognitive function loci by chance is p value = 3.0×10^{-8}. This observation is also consistent with the finding that the Wellderly have a lower overall genetic risk for Alzheimer disease and cognitive decline (above), a lower APOE-ε4 rate (rs429358 p value = 7.6×10^{-5}), and a greatly reduced frequency of the Alzheimer disease-associated rare coding variant rs145999145 in *PLD3* as previously reported (Cruchaga et al., 2014).

High Penetrance Pathogenic Variants

Next, we explored the contribution of rare monogenic disease variants to the Wellderly phenotype. We considered monogenic disorders that are the most prevalent and relevant to the Wellderly phenotype: hereditary dementia, cancer, and common monogenic diseases as defined by the American College of Medical Genetics secondary findings (Green et al., 2013). No difference in the rate of rare pathogenic variants in the Wellderly versus ITMI cohorts was observed for any gene set (chi-square p values 1.0, 0.39, and 0.80, respectively) (Figure 4A) or at the

Table 3. Genetic Risk for Leading Causes of Death with a Genetic Component

	Wellderly Genetic Risk	ITMI Genetic Risk	Wellderly Odds Ratio	p Value
Alzheimer's disease	0.0075 (±0.0418)	0.1060 (±0.0409)	0.906 (±0.051)	0.009
Breast cancer	0.596 (±0.045)	0.600 (±0.045)	0.996 (±0.064)	1.0
Colorectal cancer	1.11 (±0.030)	1.12 (±0.03)	0.992 (±0.032)	1.0
Coronary artery disease	8.27 (±0.035)	8.35 (±0.040)	0.922 (±0.047)	0.023
Lung cancer	0.130 (±0.027)	0.142 (±0.022)	0.988 (±0.034)	1.0
Pancreatic cancer	0.364 (±0.033)	0.387 (±0.026)	0.977 (±0.040)	1.0
Prostate cancer	4.80 (±0.100)	4.75 (±0.090)	1.050 (±0.131)	1.0
Stroke	2.04 (±0.030)	2.09 (±0.030)	0.958 (±0.043)	0.66
Type 2 diabetes	6.40 (±0.050)	6.39 (±0.04)	1.004 (±0.0580)	1.0

The mean genetic risk score is presented with the 95% confidence interval in parentheses. The Wellderly odds ratio is the overall odds ratio of Wellderly individuals for the development of each disease relative to the general population, and the 95% confidence interval is also in parentheses. p values corrected for multiple testing are reported.
See also Table S3.

individual gene level (Table S5). The full list of genes per disease category is presented in the Supplemental Experimental Procedures.

Rare Variant Genome-wide Association Study

Finally, we considered whether rare coding variants in any gene (minor allele frequency <1%) were associated with the Wellderly phenotype using the SKAT-O method (see the Experimental Procedures). The QQ plot of gene-level association results demonstrates no genomic inflation (genomic inflation factor = 0.99) (Figure 4B). No genome-wide significant associations (p value < 4.8×10^{-6} for Bonferroni correction of 10,447 individual gene tests) were observed. However, the top hit, *COL25A1* (p value = 1.56×10^{-5}), is of major interest. The full list of top hits is provided in Table S6.

COL25A1 contained nine ultra-rare coding variants carried by ten distinct Wellderly individuals, where eight variants were observed as singletons and one variant was observed in two distinct individuals (Figure 4C). No *COL25A1* variants were observed in any ITMI individual (variants provided in Table S7; Sanger validation traces displayed in Figure S2). Five of these nine variants have been previously observed in the Exome Aggregation Consortium Database. COL25A1 is a brain-specific, secreted collagenous protein that associates with amyloid plaques (Söderberg et al., 2005). The distribution of the observed coding mutations relative to the various functional domains and elements of COL25A1 is displayed in Figure 4C. No clustering in any particular region of COL25A1 was observed, though many of the mutations result in highly non-conservative amino-acid substitutions (p.S48F, p.G225C, p.R402C, and p.P520L), or are nearby critical regions for plaque association (p.S116L being nearby the cleavage site at amino acid 113 and p.N171S being in the NC2 domain which interacts with Aβ).

DISCUSSION

We identified no major singular contributor to healthy aging. Instead, healthy aging appears to demonstrate characteristics similar to other complex polygenic phenotypes. Further, our results suggest that healthy aging is a genetically overlapping but divergent phenotype from exceptional longevity and that the healthy aging phenotype is potentially enriched for heritable components of both reduced risk of age-associated disease and resistance to age-associated disease.

First, we observe minimal and non-significant enrichment of the previously identified *FOXO3A* longevity allele in the Wellderly (Willcox et al., 2008). In contrast, we observe a significant depletion of *APOE*-ε4 alleles, which is associated with longevity but considered to be a "frailty" allele for late-onset disease rather than a true longevity allele (Christensen et al., 2006; Gerdes et al., 2000).

Second, and again in contrast to exceptional longevity cohorts, where no overall reduction in common disease genetic risk is generally observed (Beekman et al., 2010; Fortney et al., 2015; Sebastiani et al., 2012), we observe a significant decrease in genetic risk for Alzheimer disease and coronary artery disease genetic risk in the Wellderly. Importantly, these results reflect an overall decrease in genetic risk scores, a result not observed in exceptional longevity cohorts (Beekman et al., 2010; Sebastiani et al., 2012), possibly due to the indirect and/or a mixed relationship between individual genetic disease risk loci and exceptional longevity (as discussed by Fortney et al., 2015) versus the potentially more direct relationship between aging in the absence of disease and overall genetic disease risk.

On the other hand, no difference in genetic risk is observed for type 2 diabetes genetic risk and cancer. Some of these findings (type 2 diabetes, colon, and lung cancer) can be explained by the fact that the risk for these diseases is strongly modulated by behavioral and environmental risk factors. However, we also observed no difference in genetic risk for more heritable cancer types (breast and prostate especially), neither via common variant risk nor rare pathogenic variant burden. The lack of a difference in rare pathogenic variant burden, for cancer as well as hereditary dementia and other common monogenic disorders, could be indicative of disease resistance via behavioral or other genetic characteristics.

In fact, while no individually genome-wide significant associations were observed, both our rare and common variant genetic associations implicate protection against cognitive decline as a potential molecular mechanism governing healthy aging and

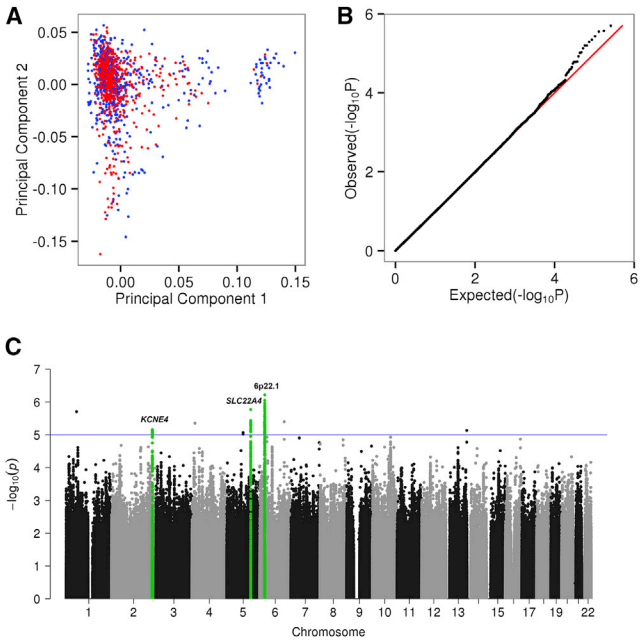


Figure 2. Common Variant Association Results

(A) Principal component plot of the Wellderly (red) and ITMI (blue) cohort based on filtered and LD pruned ($R^2 > 0.5$) common variants (allele frequency >5%).

(B) QQ plot of expected $-\log_{10}$(p values) (x axis) versus observed $-\log_{10}$(p values) (y axis) (one black point per variant). Expected versus expected $-\log_{10}$(p values) (red line) is along the diagonal. The λ value for this QQ plot is 0.98; no genomic inflation is observed.

(C) Manhattan plot of the unbiased GWAS results. Each point represents a single SNP p value determined by a logistic model adjusted for the first ten principal components. Points are organized by chromosome and chromosomal coordinate (x axis) and $-\log_{10}$(p values) (y axis). The blue line indicates p values $<10^{-5}$. Green highlighted SNPs correspond to the top most significant regions, plotted in Figure 3.

See also Table S4.

longevity. This is an area in which genetic associations have been notoriously difficult to identify, despite reasonably high heritability estimates. Perhaps most interestingly, mutation of *COL25A1* may disrupt Alzheimer disease pathogenesis due to the critical role soluble COL25A1 plays in regulating the buildup of Aβ fibrils (Kakuyama et al., 2005; Osada et al., 2005). This finding provides potential genetic support for the targeting of COL25A1 for the prevention or treatment of Alzheimer disease.

This work represents one of the first combined rare and common variant genome-wide association studies utilizing whole-genome sequencing data and highlights some of the opportunities and challenges associated with WGS-based case-control analysis. For complex polygenic phenotypes, achieving genome-wide significance is improbable without much larger sample sizes, which take years to accrue for rare phenotypes, such as healthy aging. It should be re-emphasized that the Wellderly phenotype is a narrowly defined phenotype with likely fewer eligible individuals than those who would be eligible for an exceptional longevity cohort. Thus, we acknowledge that, due to the relatively small size of our cohort, independent replication is needed to confirm our preliminary findings. In the meantime, it is important to discern patterns that hint at the underlying molecular and physiological pathways that may inform follow-up studies. While the patterns in our unbiased analysis highlight the role of genetic resistance to cognitive decline in healthy aging, due to the global absence of disease in the cohort, we suggest that the Wellderly are likely to be enriched with other genetic factors for disease resistance. These factors will likely be gleaned only via more detailed analyses of the Wellderly dataset via interaction with individual genetic risk factors or as a result of more detailed analyses of genes of interest. Moreover, the Wellderly dataset can act as independent replication and validation of candidate disease protective genetic factors identified in other cohorts. The identification of genetic factors protective against disease is of paramount importance to understanding disease biology and a proven means for drug target identification. Thus, we make the Wellderly genomic data broadly available for further mining by the research community.

EXPERIMENTAL PROCEDURES

The Healthy Elderly Active Longevity Cohort Study (IRB-13-6142) was approved by the Scripps Institutional Review Board in July 2007.

Data Availability

Aggregate unfiltered annotated Wellderly variants and their allele and genotype frequencies are available via Scripps Translational Science Institute Variant Browser (https://genomics.scripps.edu/browser). Variant presence is also queryable via a Global Alliance for Genomics and Health Beacon (https://genomics.scripps.edu/browser/ga4gh). Individual-level variant data are available from Complete Genomics under terms determined by Complete Genomics, from Scripps Genomic Medicine for scientific collaboration with not-for profit entities, and will be deposited in dbGAP under similar data use restrictions.

Wellderly Cohort Recruitment

The Wellderly cohort (1,354 individuals) was collected over the course of eight years. Individuals with no chronic diseases and not taking chronic medications were screened for enrollment. Specific inclusion/exclusion criteria are provided in the Supplemental Experimental Procedures. The study and inclusion criteria were advertised to the public, for example, via the Scripps newsletters, radio announcements, newspaper articles in various cities around the country, and in senior community newsletters. Participants volunteered and were self-referred. Research staff did not recruit individuals who had not already contacted the study staff expressing interest in participating. Potential study participants were then interviewed in order to assess health history according to the inclusion criteria.

ITMI Cohort

Individuals were enrolled in the ITMI research study titled "Molecular Study of Pre-term Birth," as described in (Bodian et al., 2014). The sub-cohort for this study consisted of adults who consented to research use of their genomic and clinical data. Demographic data were obtained from self-reported questionnaires. Of the individuals with European ancestry, 68.1% (467) were parents of a neonate born full-term, and 31.9% (219) had a pre-term infant. They range in age from 20–44 years. Collection of peripheral blood and sample processing for WGS were previously described (Bodian et al., 2014).

Survival Analysis

The number of living and deceased (with age of death) Wellderly siblings was determined by self-report. The number of at-risk for death Wellderly siblings, starting at age 10 (3,024 individuals total), and the number of deaths per five-year interval was calculated based on these self-report data. Survival analysis was conducted by determining the expected number of deaths in a cohort of 3,024 individuals born in the 1920s, under the assumption of equal male and female individuals at 10 years of age, and calculating the expected

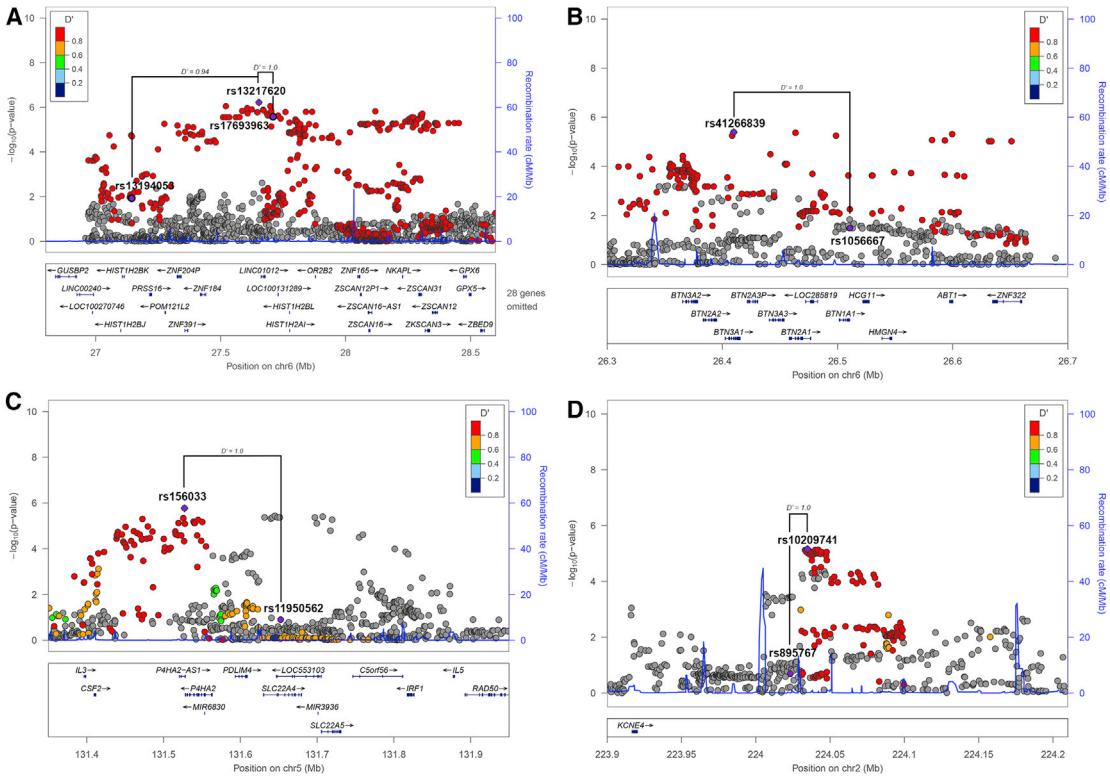


Figure 3. Regional Association Plots

(A–D) Regional plots for the top three most significant regions for GWAS comparing the Wellderly and ITMI cohorts. For all plots, each point represents a SNP, where the x axis represents the position of the SNP and the y axis (right) the –log10 p value of the genome-wide association results. Each point is color-coded with the D' value as calculated within the Wellderly and ITMI cohorts. The D' value within the 1000 Genomes European individuals is indicated by connecting lines. The recombination rate is also plotted at each genomic position with the rate indicated on the y axis (left). (A) Lead SNP rs13217620 (chr6:27,653,120) (p value = 6.1×10^{-7}) is highlighted along with SNPs previously associated with schizophrenia (rs13194053 and rs17693963). (B) Lead SNP rs41266839 (chr6:26,409,890) (p value = 4.1×10^{-6}) is highlighted along with rs1056667, which as previously associated with cognitive performance. (C) Lead SNP rs156033 (p value = 1.7×10^{-6}) is highlighted along with rs11950562, which was previously associated with isovalerylcarnitine levels. (D) Lead SNP rs10209741 (p value = 7.0×10^{-6}) is highlighted along with rs895767, which was previously marginally associated with cognitive decline.

number of deaths per five-year interval based on the conditional probability of death (q_x) as supplied in the Social Security Actuarial Study No. 120 (Bell and Miller, 2005) for the 1920s birth cohort. The significance of the difference in survival between Wellderly siblings and an equal-sized population born in the 1920s was then calculated via the log-rank test for survival analysis.

Genome Sequencing and Cohort Filtration

WGS of both the Wellderly cohort and ITMI cohort was performed by Complete Genomics via their standard WGS service. Variant calling was performed via cgatools (v.2.0.1–v.2.0.4). A mean whole-genome coverage of 56× per Wellderly individual and 55× per ITMI individual was achieved (Figures S1A and S1B). An average of 98.8% and 98.9% of the protein coding portion of the genome was covered by >10 reads in the Wellderly and ITMI cohorts, respectively (Figures S1C and S1D).

Genetic ancestry estimation was defined via the ADMIXTURE algorithm (Alexander and Lange, 2011) using ~16,000 ancestry informative markers and a reference panel of 83 populations around the world (Libiger and Schork, 2013). Individuals of ≥95% European ancestry were retained for downstream association analyses. Related individuals were identified via pairwise identity by descent estimation in PLINK (Purcell et al., 2007).

Variant Filtrations

Variants were combined across the Wellderly and ITMI cohorts via GenomeComb and filtered using metrics based on optimized features selected based on concordance between monozygotic twins (Reumers et al., 2011) and refined based on observed GWAS genomic inflation. These filters excluded variants that were (1) labeled as VQLOW in all individuals, (2) clustered in >10% of individuals from either cohort, (3) variants missing in >10% of either cohort, 4) with median coverage <10 or >100 in either cohort, 5) in simple repeats, homopolymer repeats ≥6 bp, segmental duplications, microsatellite repeats, or low-complexity repeats, (5) out of the Hardy-Weinberg Equilibrium (p value < 1×10^{-5}), and (6) in non-unique 36-mers (Derrien et al., 2012). VQLOW genotypes were set to missing. For common variant analyses, an additional filter for variants <5% minor allele frequency in either cohort was applied. For rare variant analyses, an additional filter for variants >1% minor allele frequency in either cohort was applied, as well as a minimum allele depth filter, where if 20% of individuals with a rare variant alternate allele call had a minimum alternate allele depth of <25% of total reads or fewer than three supporting reads, the variant was considered a false positive and removed.

Genetic Risk Scores

Weighted genetic risk scores based on known susceptibility markers for the top five causes of death with a genetic component were calculated in both cohorts. Markers and corresponding weights used to construct the genetic risk scores included (when possible) markers identified in the latest large-scale meta-analyses of Alzheimer disease (replicated markers in Table 2 of Lambert et al. [2013] and *APOE* risk markers [Seshadri et al., 2010]), breast cancer (markers in Tables 1 and S3a of Michailidou et al. [2015]), colorectal cancer

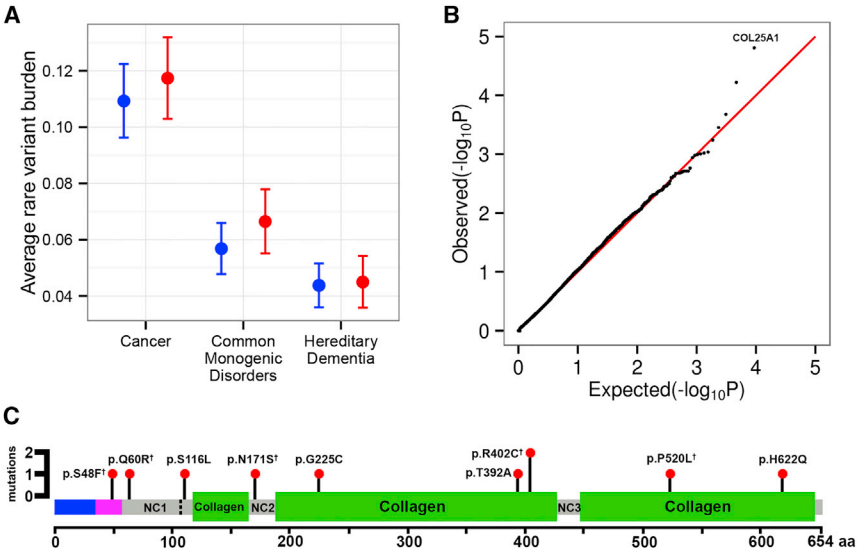


Figure 4. Rare Variant Burden

(A) The rate of rare pathogenic cancer, hereditary dementia, and common monogenic disease (secondary finding) variants is displayed for the Wellderly (red) and ITMI (blue) cohorts with SE bars. No difference is observed between the cohorts. See also Table S5.

(B) QQ plot of expected $-\log_{10}$(p values) (x axis) versus observed $-\log_{10}$(p values) (y axis) (one black point per gene). Expected versus expected $-\log_{10}$(p values) (red line) is along the diagonal. The λ value for this QQ plot is 0.99; no genomic inflation is observed. The point corresponding to the top gene, *COL25A1*, is labeled. See also Table S6.

(C) The location, impact, and count of COL25A1 coding variants are indicated along the COL25A1 protein. The COL25A1 protein and domains are displayed horizontally from the N-terminal to C-terminal amino acid (x axis). The blue region indicates the cytoplasmic region of COL25A1; the maroon region indicates the transmembrane domain; green regions are collagen domains; and gray regions are intervening non-collagen stretches. The cleavage site for secretion of COL25A1 is indicated by the vertical dashed line. COL25A1 interacts with the amyloid beta protein at NC2. Each observed mutation position is indicated by a black pin with a red head. The height of each pin corresponds to the number of alleles observed in the Wellderly cohort (all variants are observed in a single individual, except for p.R402C). The coding impact of each variant is indicated next to each pin. Superscript daggers indicate variants that have been previously observed in the Exome Aggregation Consortium database.

See also Figure S2 and Tables S5, S6, and S7.

(markers in Table S3 of Al-Tassan et al. [2015]), coronary artery disease (markers in Table S9 of Deloukas et al. [2013]), lung cancer (markers in Table S2 of Timofeeva et al. [2012] that were presented in Figures 2A–2G, using the fixed effect model for weights), pancreatic cancer (markers in Table S4 of Wolpin et al. [2014], using stage 1 odds ratios for weights), prostate cancer (markers in Table 3 of Zheng et al. [2008]), stroke (markers in Table e-2 of Kilarski et al. [2014], with the All_IS phenotype), and type 2 diabetes (markers in Table S3 and Table S6 of Mahajan et al. [2014], using the European weight estimates). Markers were coded additively, and the logarithms of the reported odds ratios were used as weights. All markers were pruned by pairwise linkage disequilibrium ($R^2 > 0.8$) prior to constructing the genetic risk score. The full list of these markers, corresponding weights, allele frequencies in each of our cohorts, and our single-marker association statistics are available in Table S3.

Variant Annotations and Pathogenicity

Variant annotation was performed by Cypher Genomics as described by the SG-ADVISER annotation tool (Pham et al., 2015). High confidence known pathogenic variants are defined as any variant whose allele frequency is no greater than 0.5% in all 1000 Genomes (Sudmant et al., 2015), NHLBI Exome Sequencing Project (Fu et al., 2013), and Exome Aggregation Consortium (http://exac.broadinstitute.org) populations and cataloged as pathogenic or likely pathogenic in ClinVar (Landrum et al., 2014) without any benign or likely benign assertions. Secondary finding variants were determined as defined by Green et al. (2013). Analyses of rare pathogenic variants considered variant counts on the cohort level only; individual pathogenic variants and actionability at the individual level were not examined.

Genome-wide Association Studies

Common variants, filtered as described above, were subject to a standard GWAS analysis via PLINK (Purcell et al., 2007) using a logistic model and corrected for the first ten principal components. The QQ plot was generated by pruning association results at a linkage disequilibrium threshold of 0.50. Chromosome positions are provided in GrCh37/hg19 coordinates. Regional plots were generated with LocusZoom with modifications (Pruim et al., 2010). Pairwise linkage disequilibrium from 1000 Genomes (Sudmant et al., 2015) European individuals was calculated with SNAP (Johnson et al., 2008). For rare variants, burden testing was performed with SKAT-O on all rare coding exonic variants (Lee et al., 2012), adjusted for the first two principal components. A gene must have had five or more rare variants (as recommended by the SKAT authors) to be included in testing; 9,343 genes satisfied this requirement. The QQ plot was generated as one point per gene.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, two figures, and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2016.03.022.

A video abstract is available at http://dx.doi.org/10.1016/j.cell.2016.03.022#mmc6.

AUTHOR CONTRIBUTIONS

G.A.E., D.L.B., and M.R. performed data curation, developed the software, and performed formal analysis and visualization. B.M. developed the software for data curation. E.R.S. conceived of the Wellderly sibling survival analysis. A.A.S.-V.Z. and S.E.T. obtained the study resources. N.E.W conceived of the genetic risk score and provided formal statistical analysis and supervision. E.J.T. and J.E.N. conceived of the research goals, reviewed and edited the manuscript, and acquired funding. A.T. conceived of the research goals, wrote the original draft and edits, executed formal analysis and visualization, provided project supervision and administration, and acquired funding.

ACKNOWLEDGMENTS

Complete Genomics produced and provided all Wellderly WGS data. We would like to thank Wayne Pfeiffer (San Diego Supercomputer Center) for his helpful comments on the manuscript and Chunlei Wu, Jiwen Xin, Jackie Atkins, and Quang Tran for IT support. This work was supported by Scripps Genomic Medicine, an NIH-NCATS Clinical and Translational Science Award (CTSA; 5 UL1 RR025774) (to STSI), the Inova Health System, and a generous gift from the Odeen family to ITMI. Further support was from the Gary and Mary West Foundation, Lavin Family Foundation, and the NIH (NIH-U01 HG006476 and U54GM114833) (to A.T.).

Received: December 17, 2015
Revised: February 5, 2016
Accepted: March 14, 2016
Published: April 21, 2016

REFERENCES

Al-Tassan, N.A., Whiffin, N., Hosking, F.J., Palles, C., Farrington, S.M., Dobbins, S.E., Harris, R., Gorman, M., Tenesa, A., Meyer, B.F., et al. (2015). A new GWAS and meta-analysis with 1000Genomes imputation identifies novel risk variants for colorectal cancer. Sci. Rep. *5*, 10442.

Alexander, D.H., and Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics *12*, 246.

Atzmon, G., Rincon, M., Schechter, C.B., Shuldiner, A.R., Lipton, R.B., Bergman, A., and Barzilai, N. (2006). Lipoprotein genotype and conserved pathway for exceptional longevity in humans. PLoS Biol. *4*, e113.

Beekman, M., Nederstigt, C., Suchiman, H.E., Kremer, D., van der Breggen, R., Lakenberg, N., Alemayehu, W.G., de Craen, A.J., Westendorp, R.G., Boomsma, D.I., et al. (2010). Genome-wide association study (GWAS)-identified disease risk alleles do not compromise human longevity. Proc. Natl. Acad. Sci. USA *107*, 18046–18049.

Bell, F.C., and Miller, M.L. (2005). Life tables for the United States Social Security area 1900-2100. Social Security Actuarial Study No 120. https://www.ssa.gov/oact/NOTES/pdf_studies/study120.pdf.

Bodian, D.L., McCutcheon, J.N., Kothiyal, P., Huddleston, K.C., Iyer, R.K., Vockley, J.G., and Niederhuber, J.E. (2014). Germline variation in cancer-susceptibility genes in a healthy, ancestrally diverse cohort: implications for individual genome sequencing. PLoS ONE *9*, e94554.

Bojesen, S.E., and Nordestgaard, B.G. (2008). The common germline Arg72-Pro polymorphism of p53 and increased longevity in humans. Cell Cycle *7*, 158–163.

Brooks-Wilson, A.R. (2013). Genetics of healthy aging and longevity. Hum. Genet. *132*, 1323–1338.

Christensen, K., Johnson, T.E., and Vaupel, J.W. (2006). The quest for genetic determinants of human longevity: challenges and insights. Nat. Rev. Genet. *7*, 436–448.

Cruchaga, C., Karch, C.M., Jin, S.C., Benitez, B.A., Cai, Y., Guerreiro, R., Harari, O., Norton, J., Budde, J., Bertelsen, S., et al.; UK Brain Expression Consortium; Alzheimer's Research UK Consortium (2014). Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. Nature *505*, 550–554.

Deloukas, P., Kanoni, S., Willenborg, C., Farrall, M., Assimes, T.L., Thompson, J.R., Ingelsson, E., Saleheen, D., Erdmann, J., Goldstein, B.A., et al.; CARDIoGRAMplusC4D Consortium; DIAGRAM Consortium; CARDIOGENICS Consortium; MuTHER Consortium; Wellcome Trust Case Control Consortium (2013). Large-scale association analysis identifies new risk loci for coronary artery disease. Nat. Genet. *45*, 25–33.

Derrien, T., Estellé, J., Marco Sola, S., Knowles, D.G., Raineri, E., Guigó, R., and Ribeca, P. (2012). Fast computation and applications of genome mappability. PLoS ONE *7*, e30377.

Fortney, K., Dobriban, E., Garagnani, P., Pirazzini, C., Monti, D., Mari, D., Atzmon, G., Barzilai, N., Franceschi, C., Owen, A.B., and Kim, S.K. (2015). Genome-wide scan informed by age-related disease identifies loci for exceptional human longevity. PLoS Genet. *11*, e1005728.

Fu, W., O'Connor, T.D., Jun, G., Kang, H.M., Abecasis, G., Leal, S.M., Gabriel, S., Rieder, M.J., Altshuler, D., Shendure, J., et al.; NHLBI Exome Sequencing Project (2013). Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature *493*, 216–220.

Gerdes, L.U., Jeune, B., Ranberg, K.A., Nybo, H., and Vaupel, J.W. (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*, 202–210.

Goldman, D.P., Cutler, D., Rowe, J.W., Michaud, P.C., Sullivan, J., Peneva, D., and Olshansky, S.J. (2013). Substantial health and economic returns from delayed aging may warrant a new focus for medical research. Health Aff. (Millwood) *32*, 1698–1705.

Green, R.C., et al. (2013). ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet. Med. *15*, 565–574.

Hurme, M., Lehtimäki, T., Jylhä, M., Karhunen, P.J., and Hervonen, A. (2005). Interleukin-6 -174G/C polymorphism and longevity: a follow-up study. Mech. Ageing Dev. *126*, 417–418.

Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., O'Donnell, C.J., and de Bakker, P.I. (2008). SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics *24*, 2938–2939.

Kakuyama, H., Söderberg, L., Horigome, K., Winblad, B., Dahlqvist, C., Näslund, J., and Tjernberg, L.O. (2005). CLAC binds to aggregated Abeta and Abeta fragments, and attenuates fibril elongation. Biochemistry *44*, 15602–15609.

Kilarski, L.L., Achterberg, S., Devan, W.J., Traylor, M., Malik, R., Lindgren, A., Pare, G., Sharma, P., Slowik, A., Thijs, V., et al.; GARNET Collaborative Research Group, Wellcome Trust Case Control Consortium 2, Australian Stroke Genetic Collaborative, the METASTROKE Consortium, and the International Stroke Genetics Consortium (2014). Meta-analysis in more than 17,900 cases of ischemic stroke reveals a novel association at 12q24.12. Neurology *83*, 678–685.

Kuningas, M., Putters, M., Westendorp, R.G., Slagboom, P.E., and van Heemst, D. (2007). SIRT1 gene, age-related diseases, and mortality: the Leiden 85-plus study. J. Gerontol. A Biol. Sci. Med. Sci. *62*, 960–965.

Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., et al.; European Alzheimer's Disease Initiative (EADI); Genetic and Environmental Risk in Alzheimer's Disease; Alzheimer's Disease Genetic Consortium; Cohorts for Heart and Aging Research in Genomic Epidemiology (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat. Genet. *45*, 1452–1458.

Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M., and Maglott, D.R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. *42*, D980–D985.

Lee, S., Emond, M.J., Bamshad, M.J., Barnes, K.C., Rieder, M.J., Nickerson, D.A., Christiani, D.C., Wurfel, M.M., and Lin, X.; NHLBI GO Exome Sequencing Project—ESP Lung Project Team (2012). Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am. J. Hum. Genet. *91*, 224–237.

Libiger, O., and Schork, N.J. (2013). A method for inferring an individual's genetic ancestry and degree of admixture associated with six major continental populations. Front. Genet. *3*, 322.

Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., Abraham, J., Adair, T., Aggarwal, R., Ahn, S.Y., et al. (2012). Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet *380*, 2095–2128.

Ma, J., Ward, E.M., Siegel, R.L., and Jemal, A. (2015). Temporal trends in mortality in the United States, 1969-2013. JAMA *314*, 1731–1739.

Mahajan, A., Go, M.J., Zhang, W., Below, J.E., Gaulton, K.J., Ferreira, T., Horikoshi, M., Johnson, A.D., Ng, M.C., Prokopenko, I., et al.; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium; Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium; South Asian Type 2 Diabetes (SAT2D) Consortium; Mexican American Type 2 Diabetes (MAT2D) Consortium; Type 2 Diabetes Genetic Exploration by Nex-generation sequencing in muylti-Ethnic Samples (T2D-GENES) Consortium (2014). Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat. Genet. *46*, 234–244.

Melzer, D., Frayling, T.M., Murray, A., Hurst, A.J., Harries, L.W., Song, H., Khaw, K., Luben, R., Surtees, P.G., Bandinelli, S.S., et al. (2007). A common

variant of the p16(INK4a) genetic region is associated with physical function in older people. Mech. Ageing Dev. *128*, 370–377.

Michailidou, K., Beesley, J., Lindstrom, S., Canisius, S., Dennis, J., Lush, M.J., Maranian, M.J., Bolla, M.K., Wang, Q., Shah, M., et al.; BOCS; kConFab Investigators; AOCS Group; NBCS; GENICA Network (2015). Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. Nat. Genet. *47*, 373–380.

Murabito, J.M., Yuan, R., and Lunetta, K.L. (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*, 470–479.

Osada, Y., Hashimoto, T., Nishimura, A., Matsuo, Y., Wakabayashi, T., and Iwatsubo, T. (2005). CLAC binds to amyloid beta peptides through the positively charged amino acid cluster within the collagenous domain 1 and inhibits formation of amyloid fibrils. J. Biol. Chem. *280*, 8596–8605.

Pawlikowska, L., Hu, D., Huntsman, S., Sung, A., Chu, C., Chen, J., Joyner, A.H., Schork, N.J., Hsueh, W.C., Reiner, A.P., et al.; Study of Osteoporotic Fractures (2009). Association of common genetic variation in the insulin/IGF1 signaling pathway with human longevity. Aging Cell *8*, 460–472.

Pham, P.H., Shipman, W.J., Erikson, G.A., Schork, N.J., and Torkamani, A. (2015). Scripps Genome ADVISER: Annotation and Distributed Variant Interpretation SERver. PLoS ONE *10*, e0116815.

Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics *26*, 2336–2337.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., and Sklar, P.; International Schizophrenia Consortium (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature *460*, 748–752.

Reich, D.E., and Goldstein, D.B. (2001). Detecting association in a case-control study while correcting for population stratification. Genet. Epidemiol. *20*, 4–16.

Reumers, J., De Rijk, P., Zhao, H., Liekens, A., Smeets, D., Cleary, J., Van Loo, P., Van Den Bossche, M., Catthoor, K., Sabbe, B., et al. (2011). Optimized filtering reduces the error rate in detecting genomic variants by short-read sequencing. Nat. Biotechnol. *30*, 61–68.

Rietveld, C.A., Medland, S.E., Derringer, J., Yang, J., Esko, T., Martin, N.W., Westra, H.J., Shakhbazov, K., Abdellaoui, A., Agrawal, A., et al.; LifeLines Cohort Study (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science *340*, 1467–1471.

Rietveld, C.A., Esko, T., Davies, G., Pers, T.H., Turley, P., Benyamin, B., Chabris, C.F., Emilsson, V., Johnson, A.D., Lee, J.J., et al. (2014). Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. Proc. Natl. Acad. Sci. USA *111*, 13790–13794.

Sanders, A.E., Wang, C., Katz, M., Derby, C.A., Barzilai, N., Ozelius, L., and Lipton, R.B. (2010). Association of a functional polymorphism in the cholesteryl ester transfer protein (CETP) gene with memory decline and incidence of dementia. JAMA *303*, 150–158.

Sebastiani, P., Solovieff, N., Dewan, A.T., Walsh, K.M., Puca, A., Hartley, S.W., Melista, E., Andersen, S., Dworkis, D.A., Wilk, J.B., et al. (2012). Genetic signatures of exceptional longevity in humans. PLoS ONE *7*, e29848.

Seshadri, S., Fitzpatrick, A.L., Ikram, M.A., DeStefano, A.L., Gudnason, V., Boada, M., Bis, J.C., Smith, A.V., Carassquillo, M.M., Lambert, J.C., et al.; CHARGE Consortium; GERAD1 Consortium; EADI1 Consortium (2010). Genome-wide analysis of genetic loci associated with Alzheimer disease. JAMA *303*, 1832–1840.

Shi, J., Levinson, D.F., Duan, J., Sanders, A.R., Zheng, Y., Pe'er, I., Dudbridge, F., Holmans, P.A., Whittemore, A.S., Mowry, B.J., et al. (2009). Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature *460*, 753–757.

Shin, S.Y., Fauman, E.B., Petersen, A.K., Krumsiek, J., Santos, R., Huang, J., Arnold, M., Erte, I., Forgetta, V., Yang, T.P., et al.; Multiple Tissue Human Expression Resource (MuTHER) Consortium (2014). An atlas of genetic influences on human blood metabolites. Nat. Genet. *46*, 543–550.

Söderberg, L., Kakuyama, H., Möller, A., Ito, A., Winblad, B., Tjernberg, L.O., and Näslund, J. (2005). Characterization of the Alzheimer's disease-associated CLAC protein and identification of an amyloid beta-peptide-binding site. J. Biol. Chem. *280*, 1007–1015.

Sudmant, P.H., Rausch, T., Gardner, E.J., Handsaker, R.E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Hsi-Yang Fritz, M., et al.; 1000 Genomes Project Consortium (2015). An integrated map of structural variation in 2,504 human genomes. Nature *526*, 75–81.

Suh, Y., Atzmon, G., Cho, M.O., Hwang, D., Liu, B., Leahy, D.J., Barzilai, N., and Cohen, P. (2008). Functionally significant insulin-like growth factor I receptor mutations in centenarians. Proc. Natl. Acad. Sci. USA *105*, 3438–3442.

Timofeeva, M.N., Hung, R.J., Rafnar, T., Christiani, D.C., Field, J.K., Bickeböller, H., Risch, A., McKay, J.D., Wang, Y., Dai, J., et al.; Transdisciplinary Research in Cancer of the Lung (TRICL) Research Team (2012). Influence of common genetic variation on lung cancer risk: meta-analysis of 14 900 cases and 29 485 controls. Hum. Mol. Genet. *21*, 4980–4995.

Wellcome Trust Case Control, C.; Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature *447*, 661–678.

Willcox, B.J., Donlon, T.A., He, Q., Chen, R., Grove, J.S., Yano, K., Masaki, K.H., Willcox, D.C., Rodriguez, B., and Curb, J.D. (2008). FOXO3A genotype is strongly associated with human longevity. Proc. Natl. Acad. Sci. USA *105*, 13987–13992.

Wokolorczyk, D., Gliniewicz, B., Sikorski, A., Zlowocka, E., Masojc, B., Debniak, T., Matyjasik, J., Mierzejewski, M., Medrek, K., Oszutowska, D., et al. (2008). A range of cancers is associated with the rs6983267 marker on chromosome 8. Cancer Res. *68*, 9982–9986.

Wolpin, B.M., Rizzato, C., Kraft, P., Kooperberg, C., Petersen, G.M., Wang, Z., Arslan, A.A., Beane-Freeman, L., Bracci, P.M., Buring, J., et al. (2014). Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. Nat. Genet. *46*, 994–1000.

Zhang, C., and Pierce, B.L. (2014). Genetic susceptibility to accelerated cognitive decline in the US Health and Retirement Study. Neurobiol. Aging *35*, 1512.e11-8.

Zheng, S.L., Sun, J., Wiklund, F., Smith, S., Stattin, P., Li, G., Adami, H.O., Hsu, F.C., Zhu, Y., Bälter, K., et al. (2008). Cumulative association of five genetic variants with prostate cancer. N. Engl. J. Med. *358*, 910–919.