



Rare genetic coding variants associated with human longevity and protection against age-related diseases

Jhih-Rong Lin^{1,16}, Patrick Sin-Chan^{2,16}, Valerio Napolioni³, Guillermo G. Torres⁴, Joydeep Mitra¹, Quanwei Zhang¹, M. Reza Jabalameli¹, Zhen Wang¹, Nha Nguyen¹, Tina Gao⁵, Regeneron Genetics Center*, Matthias Laudes⁶, Siegfried Görg⁷, Andre Franke⁴, Almut Nebel⁴, Michael D. Greicius⁸, Gil Atzmon^{5,9}, Kenny Ye¹⁰, Vera Gorbunova¹¹, Warren C. Ladiges¹², Alan R. Shuldiner², Laura J. Niedernhofer¹³, Paul D. Robbins¹³, Sofiya Milman^{1,5}, Yousin Suh^{1,14,15}, Jan Vijg¹, Nir Barzilai^{1,5} and Zhengdong D. Zhang¹✉

Extreme longevity in humans has a strong genetic component, but whether this involves genetic variation in the same longevity pathways as found in model organisms is unclear. Using whole-exome sequences of a large cohort of Ashkenazi Jewish centenarians to examine enrichment for rare coding variants, we found most longevity-associated rare coding variants converge upon conserved insulin/insulin-like growth factor 1 signaling and AMP-activating protein kinase signaling pathways. Centenarians have a number of pathogenic rare coding variants similar to control individuals, suggesting that rare variants detected in the conserved longevity pathways are protective against age-related pathology. Indeed, we detected a pro-longevity effect of rare coding variants in the Wnt signaling pathway on individuals harboring the known common risk allele APOE4. The genetic component of extreme human longevity constitutes, at least in part, rare coding variants in pathways that protect against aging, including those that control longevity in model organisms.

Species-specific lifespan is limited by aging, a multifactorial process accompanied by a general decline in tissue function and increased risk for many diseases¹. Rather than a passive, entropic process of deterioration, aging is subject to active modulation by signaling pathways and transcription factors conserved across species^{2,3}. In model organisms, single-gene mutations have been demonstrated to affect lifespan⁴. For example, at the extreme end, the lifespan of nematode worms can be increased up to nearly tenfold by mutations in genes involved in insulin/insulin-like growth factor 1 signaling (IIS)^{5,6}. Even in more complicated organisms, however, such as flies and mice, lifespan can be extended up to 50% by mutations affecting the same pathway^{7–9} or other pathways involved in growth, metabolism and nutrient sensing, such as the mechanistic target of rapamycin (mTOR) and AMP-activating protein kinase (AMPK)¹⁰. On the basis of homology, it is widely hypothesized that these conserved signaling pathways are similarly involved in human aging and longevity.

In humans, lifespan is a complex trait affected by multiple factors that vary considerably within human populations. While nongenetic

factors, including diet, physical activity, health habits and psychosocial factors, are important, lifespan clearly has a genetic component as suggested by human population-based studies^{11,12}. At increasingly older ages, especially beyond 100 years, this genetic component becomes exceedingly strong^{13,14}. As a highly complex trait, the genetic underpinnings of human lifespan probably encompass different types of genetic variants and epistasis across the allele frequency spectrum. Common variants associated with human survival have been extensively searched for in many recent genome-wide association studies (GWAS), using a variety of trait definitions and study designs¹⁵. Together, these studies have identified more than 50 longevity-associated genetic loci of genome-wide significance, among which only a few, especially *APOE*, were replicated by multiple studies¹⁶. On the other hand, several previous studies have detected an association of human longevity with variants in several aging genes—including insulin signaling genes¹⁷ and *FOXO3* (refs. ^{18,19})—using candidate gene approaches. Most of these longevity-associated SNPs have small effect sizes, and currently common variants collectively explain only a very small proportion

¹Department of Genetics, Albert Einstein College of Medicine, New York, NY, USA. ²Regeneron Pharmaceuticals Inc., New York, NY, USA. ³School of Biosciences and Veterinary Medicine, University of Camerino, Camerino, Italy. ⁴Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany. ⁵Department of Medicine, Albert Einstein College of Medicine, New York, NY, USA. ⁶Division of Endocrinology, Diabetes and Clinical Nutrition, Department of Internal Medicine I, Kiel University, Kiel, Germany. ⁷Institute of Transfusion Medicine, University Hospital Schleswig-Holstein, Lübeck, Germany. ⁸Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, CA, USA. ⁹Department of Biology, Faculty of Natural Sciences, University of Haifa, Haifa, Israel. ¹⁰Department of Epidemiology & Population Health, Albert Einstein College of Medicine, New York, NY, USA. ¹¹Department of Biology, University of Rochester, Rochester, NY, USA. ¹²Department of Comparative Medicine, School of Medicine, University of Washington, Seattle, WA, USA. ¹³Institute on the Biology of Aging and Metabolism and Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, Minneapolis, MN, USA. ¹⁴Department of Obstetrics and Gynecology, Columbia University Irving Medical Center, New York, NY, USA. ¹⁵Department of Genetics and Development, Columbia University Irving Medical Center, New York, NY, USA. ¹⁶These authors contributed equally Jhih-Rong Lin, Patrick Sin-Chan. *A full list of members and their affiliations appears in Acknowledgements. ✉e-mail: zhengdong.zhang@einsteinmed.org

of heritability for human longevity. As several recent studies suggest, rare variants probably account for at least some of the ‘missing’ heritability^{20–22}.

Here we examined rare coding variants in a cohort of 515 Ashkenazi Jewish centenarians by whole-exome sequencing (WES) and tested for enrichment using a case-control design. The exceptional longevity of this cohort and their homogeneous genetic background provided us with increased power to detect causal rare variants²³. As controls we used 496 Ashkenazi Jewish individuals, mostly from the same households as the centenarians and between age ~70 and 95 years with no parental history of extreme longevity (neither parent survived beyond 95 years of age) (Table 1 and Supplementary Table 1).

Results

Longevity genes and pathways implicated by rare variants. Using a joint genotyping procedure and stringent quality control metrics, we identified 130,297 rare coding variants, including 126,405 single-nucleotide polymorphism (SNPs) and 3,892 indels with minor allele frequencies <0.01 and missing rates <0.1 in 17,561 genes in centenarians and controls. Of all SNPs, a total of 45,493 were found to be synonymous. The remaining 84,804 nonsynonymous SNPs and all indels include 75,567 missense variants, >3,500 loss-of-function variants (1,755 frameshift, 1,736 stop-gain and 79 stop-loss variants) and other variants with multiple functional annotations. We did not exclude synonymous rare variants from our analysis because not all are functionally silent²⁴. At the whole-exome level we found no significant difference in the number of rare coding variants between centenarians and controls ($P=0.243$, logistic regression including gender and the top ten multidimensional scaling (MDS) components as covariates).

We next examined the association of rare variants with longevity at the variant or gene level. At the variant level, we applied the Firth logistic regression for rare variant association tests to examine the association between the minor allele count of each rare coding variant and longevity status. The variant with the strongest association signal was rs2229426 in fatty acid synthase (*FASN*) ($P=6.23 \times 10^{-5}$) (Supplementary Table 2). At the gene level, we applied two complementary region-based association tests²⁵—the ‘burden test of rare variants’ and the sequence kernel association test (SKAT)—to examine the association between the aggregate effect of rare coding variants in each gene and longevity. The burden test searches for a significant excess of rare alleles in longevity cases or controls, while SKAT implements a variance component test to detect the effects of variants on longevity even if they have opposite directions. Chloride voltage-gated channel6 (*CLCN6*) presented the strongest variant association with longevity (burden test, $P=3.45 \times 10^{-6}$ and 1.03×10^{-5} as the lowest and combined P value, respectively) (Supplementary Table 3). Although these top associations at the variant or gene level did not reach genome-wide significance after multiple-test correction, quantile–quantile (QQ) plots of association signals showed upward deviation in the tails—the lowest P values smaller than expected from uniform distribution (0, 1)—for several groups of rare variants such as functional rare variants (combined annotation-dependent depletion (CADD)²⁶ score ≥ 20) and functional but recessive benign rare variants (CADD ≥ 20 and PrimateAI²⁷ score <0.5) (see the variant masking in Methods for variant groups and their interpretation; Fig. 1a,b, Supplementary Figs 1 and 2 and Supplementary Table 4). These rare variants include several genes known to be related to aging, such as *FASN*²⁸ and the DNA repair gene BLM RecQ-like helicase (*BLM*)²⁹.

The extreme rarity of centenarians in human populations essentially constrains the possibility of performing the large studies necessary to discover rare variants through statistically significant genetic associations with longevity. Instead of a candidate gene approach, we used integrated gene signal processing (IGSP) to prioritize genes based on the longevity association of rare variants in an unbiased

Table 1 | Data on study cohorts

Characteristic	Centenarians	Controls (noncentenarians)
Rare variant association study cohort		
Number of subjects	515	496
Female (%)	72.4	53.0
Age at enrollment, years (mean $\pm$ s.d.)	97.6 $\pm$ 3.5	73.3 $\pm$ 8.4
Disease PRS study cohort		
Number of subjects	479	431
Female (%)	73.1	51.3
Age at enrollment, years (mean $\pm$ s.d.)	97.6 $\pm$ 3.5	73.2 $\pm$ 8.7
Lifespan study cohort		
Number of subjects	356	197
Female (%)	74.2	44.2
Age at enrollment, years (mean $\pm$ s.d.)	97.7 $\pm$ 3.7	77.9 $\pm$ 7.8
Age at death, years (mean $\pm$ s.d.)	100.5 $\pm$ 3.4	84.2 $\pm$ 7.3

manner through data integration³⁰. About 94% of human protein-coding genes were in the functional linkage network used by IGSP, and only half of them also had knockout phenotype data for their mouse homologs. To include most genes in our analysis, we opted for a network integration rather than a full one, which needs both gene network and mouse phenotype data. Individual genes were scored by joint analysis of the longevity association of genes implicated by rare variants in a gene functional linkage network (predicted based on independent genomic high-throughput data)³¹, which implicitly incorporates information on gene–gene functional similarity. Data simulation showed that such integrated scoring greatly increases the prioritization power and effectively uncovers risk genes with marginal association signals³⁰. The negative-control evaluation showed that ~100 top-ranked genes had higher IGSP scores when scored by real data than by randomized data ($P=0.037$) (Supplementary Fig. 3 and Supplementary Table 5), which suggests a clustering of longevity-associated genes implicated by rare variants in the gene network captured by a network integration in gene scoring³⁰. Subsequent pathway enrichment analysis showed that these predicted longevity genes are significantly enriched in insulin signaling (false discovery rate (FDR)=0.00879) and mTOR signaling (FDR=0.0129) (Fig. 1c and Supplementary Fig. 4). Some predicted longevity genes have an indirect connection to insulin signaling as they are in the pathway of signaling by the insulin receptor (for example, *PSMB9*). Interestingly, many of the putative longevity genes carry a burden of rare variants in centenarians, among which potential protective rare variants were also found in previous studies such as those on *ABCA1* (ref. ³²) and *PLCG2* (ref. ³³) (Supplementary Table 5).

To further increase power, the longevity association of rare coding variants can be studied at the pathway level. Since aging is characterized by evolutionarily conserved, parallel and interacting mechanistic hallmarks, we next analyzed rare variants collectively in 20 pathways of all nine aging hallmarks¹ (Supplementary Table 6). Functional but recessively benign rare variants in insulin signaling (SKAT, $P=5.57 \times 10^{-5}$, FDR=0.012) and AMPK signaling (SKAT, $P=1.59 \times 10^{-4}$, FDR=0.017) pathways were found to be significantly associated with extreme longevity (Fig. 1d and Supplementary Tables 7 and 8) after multiple testing correction that took into account the total numbers of pathways, tests and variant masks.

When studying genetic variants in association studies, it is important to validate the results by replicating any observed association signals in unrelated cohorts. Our approach followed the sequence-based replication strategy, which is more powerful than the variant-based strategy that analyzes only rare variants uncovered in the discovery cohort³⁴. Specifically, we examined three replication cohorts for the longevity association of rare coding variants in the insulin and AMPK signaling pathways (Supplementary Table 9): a German longevity cohort of 1,265 centenarians (mean age 99 years) and 4,195 blood donors (mean age 35 years) as controls; a UK Biobank longevity cohort of 104 participants with at least one long-lived parent (lifespan ≥ 100 years) and 23,405 participants with parents of usual survival (lifespan < 95 years); and an Alzheimer's disease (AD) sequencing project (ADSP) longevity cohort of 1,121 non-AD individuals aged ≥ 90 years and 38 non-AD individuals aged < 75 years³⁵.

In the German longevity cohort, we detected a significant longevity association of functional but recessive benign ultra-rare variants (alternative allele frequency (AAF) $< 0.05\%$ among non-Finnish Europeans in gnomAD³⁶) in insulin signaling (SKAT, $P = 4.41 \times 10^{-4}$, FDR = 0.018) after appropriate multiple testing correction (Extended Data Fig. 1a). In the UK Biobank longevity cohort, we identified significant longevity associations of functional rare variants in insulin signaling (SKAT, $P = 9.64 \times 10^{-6}$, FDR = 3.87×10^{-4}) and functional but recessive benign ultra-rare variants in AMPK signaling (SKAT, $P = 2.08 \times 10^{-3}$, FDR = 0.041) pathways (Extended Data Fig. 2a). In the ADSP longevity cohort, we identified significant longevity associations of recessive pathogenic rare variants in insulin signaling (burden test, $P = 8.98 \times 10^{-5}$, FDR = 3.6×10^{-3} ; direction on controls) (Extended Data Fig. 3).

Next, we focused on identifying rare variants associated with human age-related disease. A genetic relationship between extreme human longevity and disease is supported by multiple observations in independent studies of a genetic association between extreme human longevity and *APOE*, a locus causally related to both cardiovascular and neurodegenerative disease³⁷. Here we hypothesized that rare genetic variants associated with human longevity can exert their beneficial effects, at least in part, by protecting against chronic disease. Hence, we examined the rare coding variants in the 20 aging hallmark pathways in more refined subgroups of our cohort based on their *APOE* haplotype status, and analyzed separately the longevity subcohorts of *APOE4* carriers and noncarriers (hereinafter *APOE4*⁺ and *APOE4*⁻, respectively) to identify longevity-associated pathways in these two distinct genetic backgrounds. Among *APOE4*⁻, functional but recessively benign rare variants in both insulin and AMPK signaling pathways were again found to be significantly associated with longevity (SKAT, $P = 6.21 \times 10^{-6}$ and 7.9×10^{-5} , FDR = 2.63×10^{-3} and 0.013, respectively) (Extended Data Fig. 4 and Supplementary Tables 10 and 11). Interestingly, among *APOE4*⁺ we detected a significant association between longevity and 152 functional rare variants in Wnt signaling genes after multiple testing correction using both the burden test and SKAT (burden test, $P = 9.16 \times 10^{-5}$,

FDR = 0.013; SKAT, $P = 3.40 \times 10^{-4}$, FDR = 0.036) (Extended Data Fig. 4 and Supplementary Tables 12 and 13). The direction of association suggests that these rare variants are enriched for protective variants among centenarian *APOE4*⁺ in our cohort (Supplementary Table 14). Indeed, only six of them were predicted as highly pathogenic rare variants (PrimateAI score ≥ 0.9) and they are not enriched, either individually or collectively, among *APOE4*⁺ centenarians. The Wnt association was replicated in the UK Biobank longevity cohort with a significant longevity association of functional rare variants in Wnt signaling pathway among *APOE4*⁺ (SKAT, $P = 1.79 \times 10^{-10}$, FDR = 2.14×10^{-8}) (Extended Data Fig. 2b and Supplementary Table 9). We did not detect significant longevity association signals from rare variants in the Wnt signaling pathway in either of the *APOE4* stratified German longevity subcohorts (Extended Data Fig. 1b).

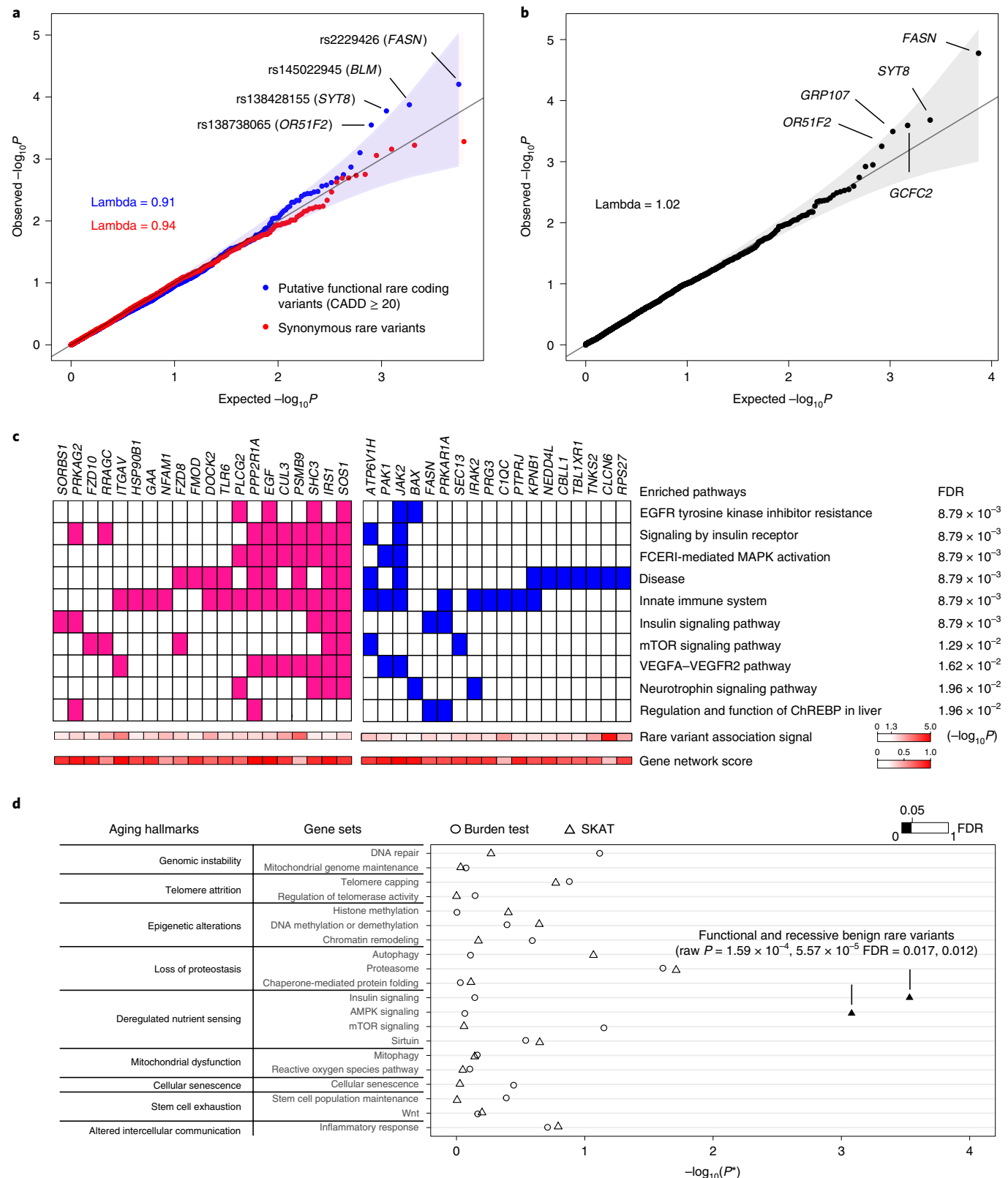
We examined further the protective effect of functional rare variants in Wnt signaling genes on individual human lifespan in our lifespan cohort of 553 individuals with verifiable age at death. Starting with the full linear model of lifespan that included gender, *APOE4* status, the alternative allele count of protective rare variants in Wnt signaling and all two- and three-way interaction terms among them, we identified a statistically significant interaction, the only one, between *APOE4* status and allele count in Wnt signaling ($P = 1.13 \times 10^{-4}$) (Supplementary Table 15). Because *APOE4* status is determined mainly by rs429358, a common variant (minor allele frequency (MAF) = 0.14) associated with aging and age-related diseases, the lifespan analysis result indicates the existence of epistasis between rare variants and aging-associated common variants in the genetic architecture of human aging. We then analyzed the relationship between individual lifespan and the alternative allele count of protective rare variants in Wnt signaling in subcohorts, stratified by the status of both longevity and *APOE4* (Fig. 2a). Among centenarians, allele count in Wnt signaling has no effect on lifespan regardless of *APOE4* status. Among noncentenarians there was a significant positive correlation between the burden of Wnt rare variants and lifespan among *APOE4*⁺ ($r = 0.406$, $P = 8.39 \times 10^{-3}$, FDR = 0.026; Fig. 2a, middle) compared with non-carriers. The relationship between *APOE4* status and allele count in Wnt signaling can also be more readily appreciated by comparing the average lifespan of subcohorts stratified based on both *APOE* and Wnt signaling. Among *APOE4*⁺, the median difference in lifespan was > 9 years ($P = 2.38 \times 10^{-3}$) between individuals with low and high allele counts in Wnt signaling (Fig. 2b). The negative effect of *APOE4* on lifespan became weaker among individuals with a high burden of potentially protective Wnt rare variants (Fig. 2c). Interestingly, the aforementioned 152 rare variants in Wnt signaling genes are associated with the disease status of individuals in the ADSP (SKAT, $P = 4.82 \times 10^{-3}$). Finally, using the same framework, we analyzed lifespans of centenarians and noncentenarians separately and demonstrated a similar protective effect of the 152 rare variants in Wnt signaling genes (Supplementary Table 14) on lifespan among noncentenarian *APOE4*⁺ (Supplementary Table 9 and Extended Data Figs. 5 and 6).

Fig. 1 | Longevity association of rare variants. **a**, QQ plots for single rare variant association tests. *P* values from tests of 2,787 functional rare variants (CADD ≥ 20) and 3,127 synonymous rare variants were used to construct two separate QQ plots. Only rare variants with a minor allele count ≥ 15 in the case-control cohort are included in the plots. **b**, QQ plot for gene-based rare variant association tests. SKAT *P* values from tests of functional but recessive benign rare variants (CADD ≥ 20 and PrimateAI scores < 0.5) in 3,717 genes were used to construct the QQ plot. Only genes with two or more masked rare variants in the case-control study cohort are included in the QQ plot (QQ plots under different rare-variant masks are given in Supplementary Fig. 2). **c**, Pathway enrichment analysis of genes implicated by rare variants aggregated in a gene functional network. The top 100 IGSP-scored genes showing the trend of network aggregation were analyzed. The top 10 nonredundant (covering unique putative longevity genes) enriched pathways are shown. For a gene in a pathway in the heatmap, the color of its cell indicates a weighted burden of rare variants in centenarians (deep pink) or controls (blue) (Supplementary Table 5). Genes in the heatmap are ordered based on their hierarchical clustering. **d**, Gene set rare variant association for aging-related pathways. *P*^{*} denotes *P* value corrected for six categories of rare variants using the minimal-*P* value test from ref. ⁶⁷ (Methods). The text for significant association denotes the lowest nominal *P* value among different categories of rare variants and FDR.

Longevity and common polygenic risk of age-related diseases.

The phenotypic outcome of individuals that carry rare variants of large effects can also be influenced by the background of common polygenic variation. To assess how rare variants may interact with the genetic background of common variants to affect human aging, we specifically examined in our longevity cohort common variants

associated with seven age-related diseases: AD, coronary artery disease (CAD), type 2 diabetes (T2D), stroke, breast cancer, prostate cancer and pancreatic cancer. This analyzed cohort consists of 479 centenarians and 431 controls with both WES and SNP array data available (Table 1 and Supplementary Table 1). We calculated polygenic risk scores (PRS) of individuals for these diseases using summary statistics



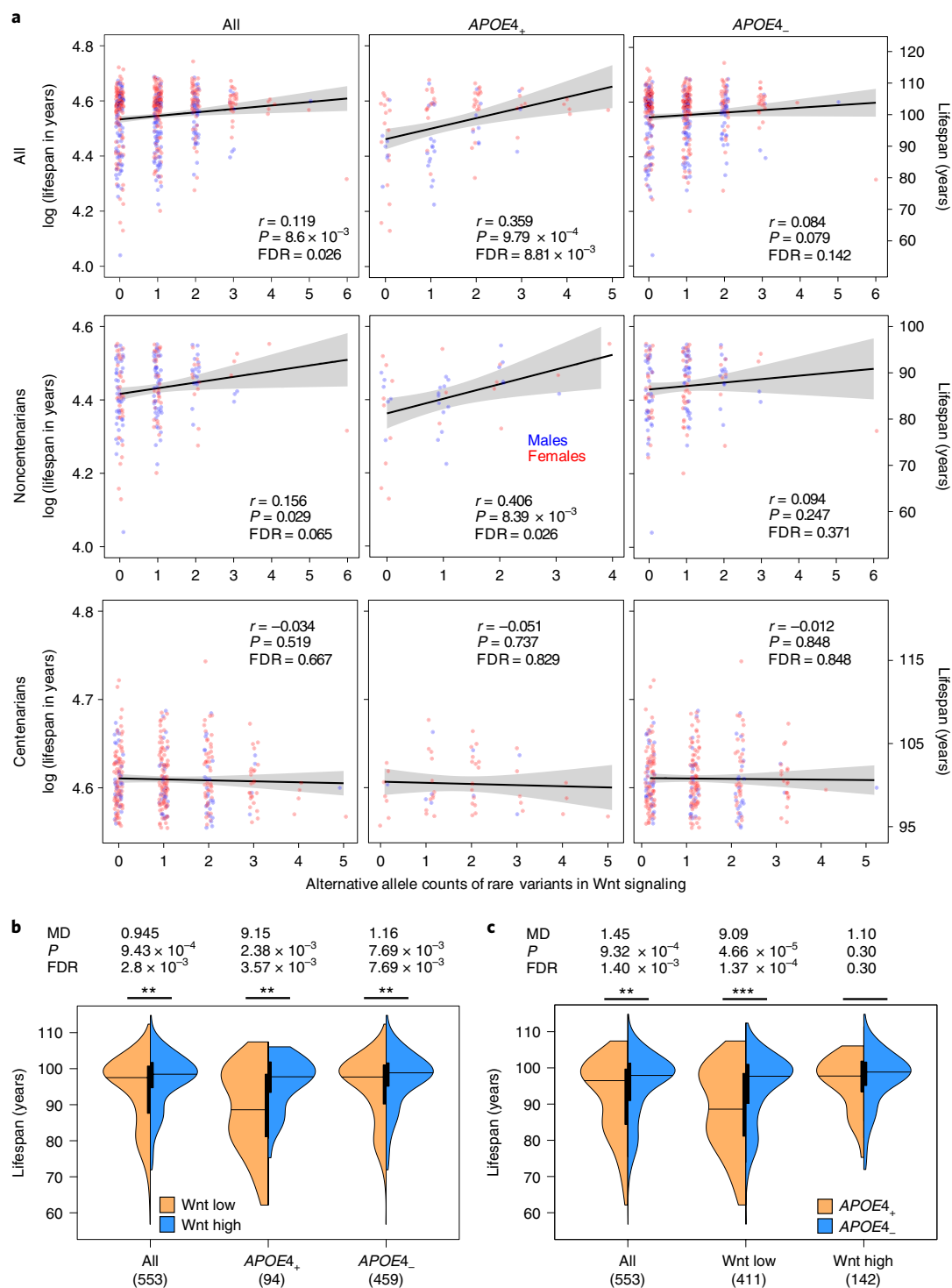


Fig. 2 | Protective rare variants in Wnt signaling genes for $APOE4_+$. **a**, Correlation between lifespan and the alternative allele count of protective rare variants in Wnt signaling genes. **b**, Lifespan difference for individuals carrying a high or low burden of protective rare variants in Wnt signaling genes. The horizontal lines and thick vertical lines in violin plots represent median and interquartile range (IQR), respectively. **c**, Negative effects of $APOE4$ on lifespan compensated by protective rare variants in Wnt signaling genes. P , uncorrected P value derived from linear regression with log-transformed age at death as the outcome and gender as a covariate (Methods). Wnt low and Wnt high represent the alternative allele count of rare variants in Wnt signaling genes ≤ 1 and > 1 (the median), respectively. In parentheses are the numbers of individuals. MD, median difference. Double and triple asterisks denote FDR smaller than 0.01 and 0.001, respectively.

from their corresponding GWAS (Methods). Empirical P values provided by PRSice-2 that account for overfitting indicated significant genetic overlap between longevity and each of AD, CAD and T2D (Fig. 3, Extended Data Fig. 7 and Supplementary Table 16), which

was further supported by the significant results of cross-validation (Supplementary Table 17). PRS for AD, CAD and T2D explained 1.93% ($P = 0.0019$), 1.32% ($P = 0.013$) and 1.29% ($P = 0.015$) variance of longevity status, respectively. Measured by PRS, centenarians

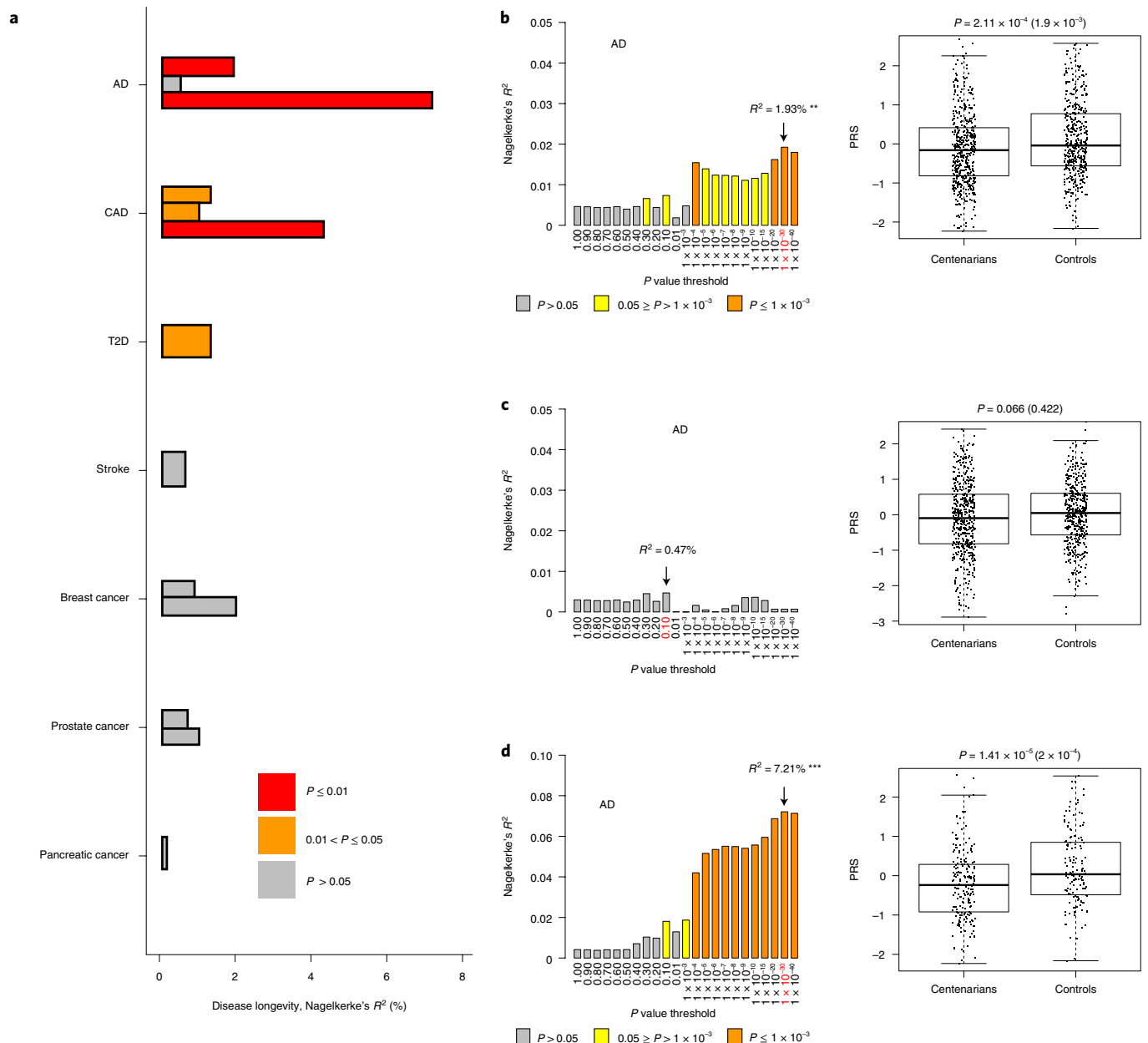


Fig. 3 | Common polygenic risk of age-related diseases. a, Common polygenic risk for seven different age-related diseases of subjects were calculated using PRS of the corresponding disease. Nagelkerke's R^2 is based on correlation between PRS and centenarian status. Bar color denotes the statistical significance of R^2 after adjusting for MDS1-10 and gender (except breast cancer and prostate cancer, which were tested with females and males, respectively) as covariates. Statistical significance is based on the permutation P values of using PRSice-2. For AD and CAD, middle bars show the results of PRS analyses excluding SNPs within 1 megabase pair (Mb) of *APOE* haplotype SNPs rs429358 and rs7412. Bottom bars (for AD, CAD, breast cancer and prostate cancer) show the results of PRS analyses using extreme-longevity phenotypes (cases and controls aged ≥ 100 and < 80 years, respectively) (Supplementary Table 16). **b–d**, Three PRS analyses of AD (**b**), AD excluding SNPs within 1 Mb of rs7412 or rs429358 (**c**), or AD using extreme-longevity phenotypes (**d**). In boxplots, points represent individuals and horizontal lines represent upper fence (maximum in $Q3 + 1.5 \times IQR$), upper quartile (Q3), median, lower quartile (Q1) and lower fence (minimum in $Q1 - 1.5 \times IQR$), sequentially from top to bottom. IQR, 25th to 75th percentile. $n = 910$ biologically independent samples. Above the boxplot on the right are raw and adjusted (in parentheses) P values for the best prediction in Nagelkerke's R^2 plot on the left, which were calculated based on logistic regression and permutation testing in PRSice-2, respectively. Double and triple asterisks denote adjusted P values smaller than 0.01 and 0.001, respectively.

tend to have reduced genetic susceptibility not only to AD and CAD (Bonferroni–Holm $P^* = 0.0067$ and 0.039 , respectively), which were previously found to be associated with healthy aging³⁸, but also T2D ($P^* = 0.039$). The predictive power of PRS for AD was mainly driven by the *APOE* haplotype defined by SNPs rs7412 and rs429358 (Fig. 3b,c), but this is not the case for CAD (Extended Data Fig. 7b)³⁹. To further examine genetic overlap between longevity and diseases,

we applied an ‘extreme-longevity phenotyping’ strategy and found that the variance explained by PRS for AD and CAD increased by almost fourfold to ~ 4 – 7% for people aged ≥ 100 and < 80 years (Fig. 3, Supplementary Fig. 5 and Supplementary Table 16). PRS for T2D showed a stronger association with longevity status among males than females in our cohort; PRS for AD and CAD, however, showed no such gender difference (Supplementary Fig. 6).

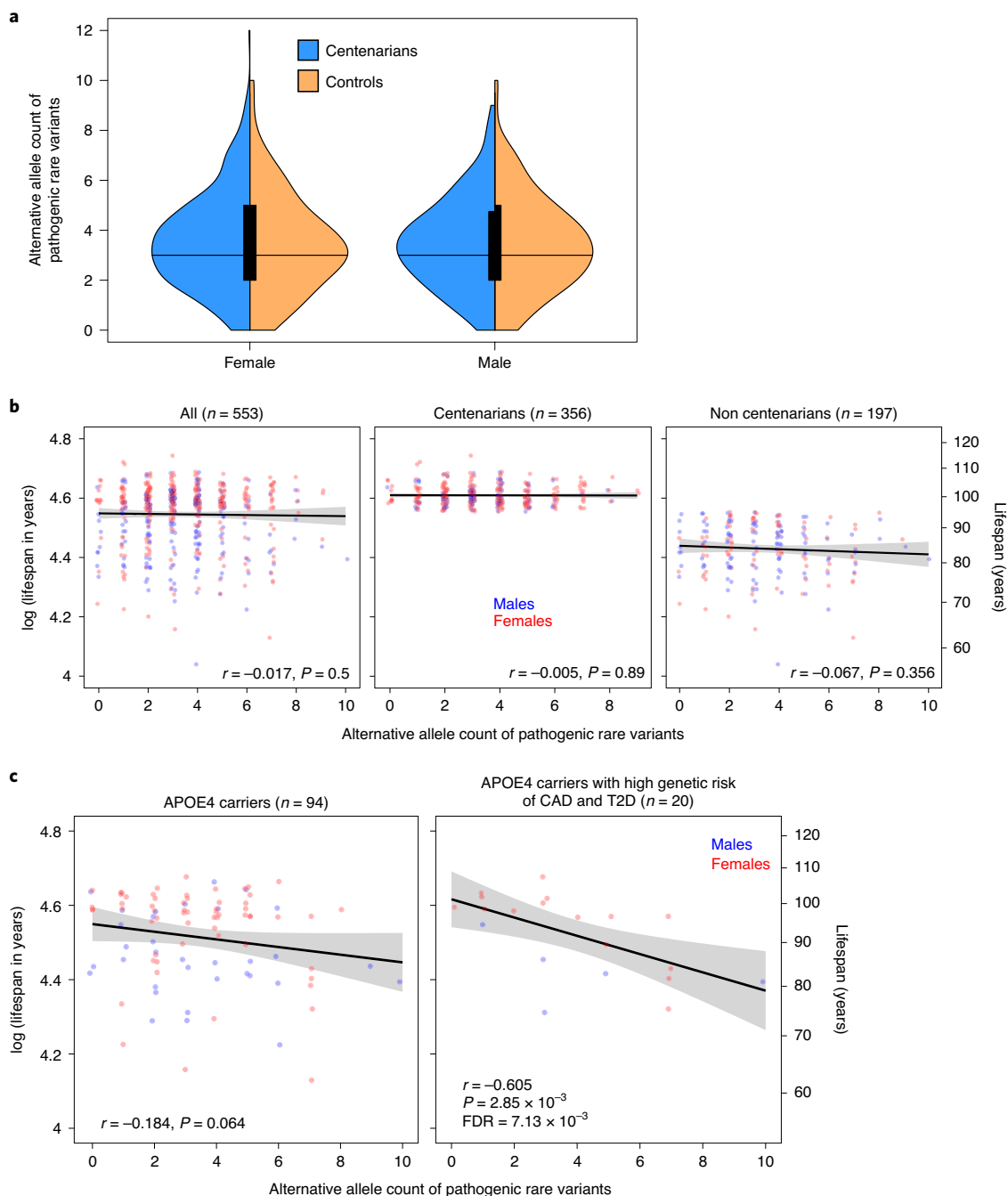


Fig. 4 | Analysis of pathogenic rare variants. **a**, Exome-wide burden of pathogenic rare variants in centenarians and controls. **b**, Correlation between lifespan and the exome-wide burden of pathogenic rare variants. Left: results based on all 553 individuals; middle and right: results based on individuals with lifespan ≥ 95 and < 95 years, respectively. **c**, Correlation between lifespan and the exome-wide burden of pathogenic rare variants among individuals with high genetic risk of age-related diseases. Left: correlation among 94 APOE4⁺; right: correlation among 20 APOE4⁺ with PRS among the top 45% for CAD and T2D (see Supplementary Table 18 for results using other cutoffs).

Pathogenic rare variants and longevity. Since the genetic component of extreme longevity can be explained, at least in part, by a reduced burden of pathogenic variants as compared with that of the general population, we compared the counts of predicted pathogenic rare coding variants (PrimateAI score ≥ 0.9). No significant difference between centenarians and controls was observed ($P = 0.476$, logistic regression including gender and the top ten MDS components as covariates; Fig. 4a). Using our lifespan cohort, we next investigated whether pathogenic rare coding variants affect lifespan, whether the effect depends on the common polygenic disease

background and whether the effect is different between centenarians and controls. Consistent with the general observation, females also had significantly better survivorship than males in our lifespan cohort ($P = 1.71 \times 10^{-7}$; Extended Data Fig. 8), as did APOE4[−] compared with APOE4⁺ ($P = 9.32 \times 10^{-4}$; Extended Data Fig. 9). In our lifespan cohort, 853 pathogenic rare variants were identified. No correlation between the exome-wide burden of pathogenic rare variants and lifespan was observed among centenarians and non-centenarians together (the full lifespan cohort), or between either of them separately (Fig. 4b).

Human extreme longevity could be causally driven by a lack of genetic risk factors for chronic disease, by protective variants or by both. As measured by PRS, centenarians in our cohort tend to have reduced genetic susceptibility to AD, CAD and T2D among the seven age-related diseases that we examined (Fig. 3 and Supplementary Table 16). Using *APOE4* status and PRS of CAD and T2D, we stratified our lifespan cohort according to their common genetic risk of AD, CAD and T2D—the three age-related diseases having significant genetic overlap with longevity in our cohort—and examined how the common polygenic disease risk background and pathogenic rare variants may together affect human lifespan. We first re-examined the effect of pathogenic rare variants on lifespan with an AD risk background based on *APOE4* status, and found a weak negative correlation ($r = -0.184$, $P = 0.064$) among *APOE4*⁺. However, this relationship became significantly stronger ($r = -0.605$, $P = 2.85 \times 10^{-3}$, $FDR = 7.13 \times 10^{-3}$) when substantial genetic risk of both CAD and T2D was included (that is, *APOE4*⁺ with PRS for both diseases is higher than the respective median of the longevity cohort) (Fig. 4c and Supplementary Table 18). These results suggest interaction between pathogenic rare variants and disease-associated common variants. Such genetic interactions may affect the deleterious effect of pathogenic rare variants on human lifespan, a possibility that we formally investigated using a full linear model of lifespan including gender, *APOE4* status, separate PRS of CAD and T2D, pathogenic rare variant counts and all two-way and higher-order interaction terms among them. The subsequent stepwise model selection identified multiple interactions, among which the most significant is a three-way interaction between pathogenic rare variant count and common polygenic disease risk of AD and T2D in our lifespan cohort ($P = 3.12 \times 10^{-4}$; Supplementary Table 15). Our analyses of stratified subcohorts showed that the negative effect of common polygenic disease risk on human lifespan can intensify under a high burden of pathogenic rare variants. For example, the presence of *APOE4* reduced lifespan by ~1.5 years on average in our cohort. However, among individuals with at least seven pathogenic rare variants (median = 3), *APOE4*⁺ lived ~17 years fewer than noncarriers in general ($P = 2.77 \times 10^{-4}$, $FDR = 8.31 \times 10^{-4}$; Supplementary Fig. 7). To replicate this discovery of the relationship between pathogenic rare variants and lifespan, we first constructed a UK Biobank parental lifespan cohort (Methods), which consists of 20,823 unrelated (to first-degree kinship) participants with known parental age at death, and then examined the relationship between the exome-wide burden of pathogenic rare coding variants and parental lifespan in this cohort (Supplementary Fig. 8). We observed a negative correlation among *APOE4*⁺ ($r = -0.024$, $P = 0.044$; Supplementary Fig. 9 and Supplementary Table 9). The stepwise model selection procedure identified a significant interaction related to parental lifespan between *APOE4* status and the exome-wide burden of pathogenic rare coding variants ($P = 5.48 \times 10^{-5}$).

Discussion

In summary, in this large-scale genetic study of rare coding variants and human longevity, our network-integrated analysis identified enrichment of longevity-associated rare coding variants in conserved aging pathways, and gene set association tests confirmed the longevity association of rare variants in the insulin and AMPK signaling pathways. These results suggest that rare variants in conserved aging pathways, important for aging of model organisms, also affect human lifespan and constitute a part of the genetic architecture of human longevity. As expected, based on the many species-specific characteristics of aging, this pattern is not completely identical between human and animal longevity. For example, we did not find any association of extreme longevity with variants in the mTOR pathway, which has been associated with longevity in model organisms including the mouse. On the other hand, we did

find other pathways critical to human aging not yet identified in model organisms. For example, we demonstrated protective effects of rare variants in Wnt signaling on human lifespan. Interestingly, in the *klotho*-knockout mouse model of accelerated aging, continuous Wnt exposure triggered accelerated cellular senescence, implicating Wnt signaling in mammalian aging⁴⁰. Finally, our results confirm previous reports that centenarians do not have a lower burden of pathogenic variants. Instead, from our present study, it appears that rare protective variants suppress the adverse effects of pathogenic variants on longevity.

To investigate whether the same conserved pathways are important to aging of both model organisms and humans, we could examine the effects of rare variants on lifespan-related traits. This is particularly challenging, however, due to a strong intrinsic stochasticity in aging processes: among isogenic *Caenorhabditis elegans* in a constant environment, lifespan of long-lived (age 1) mutants overlapped with that of wild-type controls^{41,42}. Thus, the same genetic variants may have highly variable effects on lifespan among different individuals. While this stochasticity complicates the identification of longevity-associated variants in conserved aging pathways, the use of appropriate statistical tests and study cohorts can help overcome the challenge.

In this study we identified rare coding variants in aging pathways that affect human longevity. Future studies of their molecular functions could generate actionable biological insights on aging. In particular, uncovering the downstream pathways that mediate protective effects of rare variants found in Wnt signaling genes is imperative to translate this finding into therapeutic interventions against age-related diseases. Experiments with mouse cells suggest that *APOE4* may inhibit Wnt signaling⁴³. Dysregulation of Wnt signaling contributes to different types of age-related diseases such as cancer⁴⁴, AD⁴⁵ and cardiovascular disease⁴⁶. Thus, protective rare variants in Wnt signaling genes could counteract the adverse effects of *APOE4*-induced Wnt inhibition on the progression of downstream age-related diseases and thus affect lifespan. While coding variants are more likely to reduce than to enhance the function of the protein product, conclusive confirmation and understanding of the functional effects at the molecular, cellular and organismal levels require experimental validation using functional assays and genome editing¹⁶.

Our study suggests that rare variants can have distinct effects on lifespan in different genetic backgrounds of age-related diseases (such as *APOE4* status), underlying the difficulty in detecting and replicating the effects of rare variants on lifespan without considering other genetic factors. On the other hand, while common variants associated with age-related diseases are known to influence lifespan⁴⁷, our finding of potential genetic interactions between common and rare variants in the context of the human lifespan provides insights into the mechanism of disease resilience as a part of the genetics of healthy aging among centenarians. How perturbation of conserved aging pathways contributes to human longevity and healthspan cannot be answered by genetics alone. However, while the molecular mechanisms of many conserved aging pathways have been widely studied in model organisms, our findings on rare variants—especially those from centenarians with high common polygenic risk of age-related diseases—can help translate those established longevity-regulating mechanisms in model organisms to therapeutic targets for healthy aging of humans.

A limitation of WES, as used in our present study, is the absence of rare, noncoding variants that have been implicated in aging of model organisms^{48–50} and that are thus of potential interest regarding human longevity. These include, for example, rare variants in noncoding RNAs or other regulatory elements relevant for tissue specificities, and variants in long tandem repeats connected to brain health and various neurological disorders. To identify the latter, long sequencing reads are required⁵¹.

Methods

Our WES study on the Einstein longevity cohorts complies with all relevant ethical regulations and was approved by the Institutional Review Board at Albert Einstein College of Medicine. Informed consent was obtained from participants, or from a proxy if the participant lacked decisional capacity. The WES studies of all three replication cohorts have informed consent from participants and were approved by the respective ethics committees or institutions: the Ethics Committee at Medical Faculty of Kiel University for the German longevity cohort; the Ethics Advisory Committee and the external ethics committees for the UK Biobank; and the ethics committees of the Broad Institute, Baylor College of Medicine's Human Genome Sequencing Center and Washington University's McDonnell Genome Institute for the ADSP cohort.

Recruitment of Einstein longevity cohorts. The study subjects were Ashkenazi Jewish participants from two longevity cohorts, the Longevity Genes Project (LGP) and the LonGenity study, who have been recruited and characterized at the Albert Einstein College of Medicine since 1998. Cases (centenarians) were defined as individuals of age ≥ 95 years, while individuals of age < 95 years without a parental history of longevity (neither parent survived beyond 95 years of age) were classified as controls. The centenarians' dates of birth were confirmed by birth certificates or government-issued identification. Vital status and date of death, where applicable, were determined as of 3 April 2019, based on documentation of last contact with the study participant, reports from the next of kin and a search of publicly available databases. In the LGP and LonGenity cohorts, 555 and 508 individuals were classified as longevity cases (mean age 101 years) and controls (mean age 83 years), respectively. Mortality status was confirmed for 650 individuals, and these individuals were adopted as subjects for the lifespan analysis.

SNP array genotyping. SNP array genotyping was performed using Illumina Global Screening Array-24 v.1.0 BeadChip with 642,824 markers, 7,201 of which could not be 'lifted over' to human genome assembly GRCh38 and were thus removed; a total of 2,026 samples were genotyped by SNP array. After removal of duplicates and samples not in our longevity studies, 635,623 variants in 1,830 samples were processed and analyzed (1,740 samples also have WES data). Quality control of array-based genotyped data was carried out using PLINK software (v.1.9)⁵². First, we checked the missing rate of SNPs and samples. SNPs and samples that missed $> 20\%$ genotype calls were removed, and this missingness filtering was then repeated with a more stringent threshold of 2%. Individuals whose self-reported gender was different from that predicted based on sex chromosome heterozygosity were removed. SNPs whose genotype frequencies deviated from Hardy–Weinberg equilibrium with a χ^2 -test $P < 1 \times 10^{-6}$ among controls, followed by $P < 1 \times 10^{-10}$ among cases, were removed. Finally, samples whose heterozygosity deviated by more than three standard deviations from the mean were removed.

Exome sequencing and genotyping. Exome sequencing of 2,112 individuals in the LGP and LonGenity cohorts was performed at the Regeneron Genetics Center (RGC). Sample preparation and WES were performed using previously described methods⁵³ (Supplementary Note). Variants in our centenarian cohort were called on human genome assembly GRCh38. For our rare variant analyses using both binary (cases versus controls) longevity and continuous lifespan data, only rare variants with missing rates < 0.1 in the corresponding study cohorts were analyzed; all samples in our study cohorts had a missing rate < 0.01 on rare variants that passed quality control (Supplementary Note).

Aggregation of SNP array and WES data. For PRS-related analyses, we used genotypic data aggregated from WES and SNP array (Extended Data Fig. 10a) for two reasons: (1) genotypes of common variants from the whole genome (not just the exome) need to be imputed (Genotype imputation) for PRS calculation; and (2) genome-wide imputation based on genotypic data from both WES (for better accuracy) and SNP array (for better coverage) is better than that based on WES data alone. After the aggregation process (Supplementary Note), $\sim 1,203,000$ variants were retained in the merged VCF file.

Genotype imputation. We used the Michigan Imputation Server (Minimac3)⁵⁴ for genotype imputation ($n = 1,740$). The Haplotype Reference Consortium (HRC, r1.1 2016)⁵⁵ was used as the reference panel, Eagle v.2.3 for phasing and the European population for quality control. After the postimputation process (Supplementary Note and Supplementary Fig. 10), we obtained $\sim 14,079,000$ polymorphic variants in our cohort. We evaluated the suitability of the HRC reference panel for cross-ethnicity genotype imputation in our study using 196 Ashkenazi Jewish individuals in our cohort for whom the whole-genome sequencing (WGS) data were available. The genotype imputation we performed was highly accurate: in 183 individuals (out of 196), genotypes of $> 99\%$ of 2,020 randomly selected noncoding variants that were not genotyped by either WES or SNP array data were correctly imputed (Supplementary Fig. 11).

Polygenic risk score analysis. We calculated PRS using PRSice-2 (refs. ^{56,57}) to analyze disease risk from common variants in our longevity cohort. We first collected summary statistics from the most recent GWAS of seven complex

diseases of European or predominantly European ancestry: AD⁵⁸, CAD⁵⁹, T2D⁶⁰, stroke⁶¹, prostate cancer⁶², breast cancer⁶³ and pancreatic cancer⁶⁴. From combined genotype data after imputation for 1,740 samples, common SNPs (MAF $> 5\%$) were selected in the cohort and linkage disequilibrium (LD) clumping was carried out if they were within 250 kb and $R^2 > 0.1$. After clumping, we used 19 P values (1.00, 0.90, 0.80, 0.70, 0.60, 0.50, 0.40, 0.30, 0.20, 0.10, 0.01, 1×10^{-3} , 1×10^{-4} , 1×10^{-5} , 1×10^{-6} , 1×10^{-7} , 1×10^{-8} , 1×10^{-9} and 1×10^{-10}) as cutoffs to select SNPs for scoring and, for AD, additional ones to restrict selection to most AD-associated SNPs. After removal of outliers based on MDS analysis (Supplementary Fig. 12), non-Ashkenazi Jewish individuals and kinship, 910 centenarians and controls among 1,740 samples were used to evaluate the association between disease PRS and longevity in the cohort (Extended Data Fig. 10a). To remove population substructure and sex difference, we included the top ten MDS components derived from common SNPs in the combined genotype dataset and gender as covariates in the regression analysis to evaluate PRS association. When analyzing PRS of prostate cancer and breast cancer, only male and female individuals were considered, respectively.

Rare variant association analysis. Among 2,021 Ashkenazi Jewish individuals with WES data, 536 were centenarians and 506 were controls. Pairs of individuals with the proportion of alleles shared identity-by-descent (IBD) > 0.4 were identified as related—that is, monozygotic twins, parents and children and full siblings—and one sample per pair was excluded, with inclusion to achieve more cases, higher ages of cases and lower ages of controls. In our study cohort we identified 31 participants as being related to other participants due to high IBD. After excluding these, we had 515 cases (mean age 101 years) and 496 controls (mean age 83 years) for rare variant association analysis (Table 1 and Extended Data Fig. 10b). In this study, we analyzed rare variants with alternative allele frequencies $< 1\%$ in Ashkenazi Jewish populations, which were calculated based on the average allele frequency in 731 unrelated (to first-degree kinship) Ashkenazi Jewish individuals in our centenarian cohort (2,021) (excluding centenarians and other individuals included in our study (Table 1)) and that in Ashkenazi Jews reported in gnomAD. Longevity association was assessed at the variant, gene and gene set levels. We evaluated the longevity association of each rare coding variant using the Firth logistic regression⁶⁵. For association tests at gene and gene set levels, we performed the burden test and SKAT (implemented in R v.1.3 (ref. ⁶⁶)) to test the longevity association of six different subsets of rare variants within each gene or gene set. The variant-masking scheme⁶⁷ was designed to group similar rare variants of specific properties based on CADD (v.1.4) and PrimateAI (v.0.2) annotation. CADD is widely used as a variant annotation tool to predict the functionality (that is, being functional or neutral) of variants. In contrast, PrimateAI predicts their clinical impact (that is, being pathogenic or benign). We defined different classes of variant based on the recommended thresholds of CADD and PrimateAI scores (Supplementary Table 4): all rare variants (without masking), functional (or non-neutral)²⁶ rare variants (CADD ≥ 20), dominant pathogenic rare variants (PrimateAI score > 0.8), recessive pathogenic rare variants (PrimateAI score > 0.7), functional but dominant benign rare variants (CADD ≥ 20 and PrimateAI score < 0.6) and functional but recessive benign rare variants (CADD ≥ 20 and PrimateAI score < 0.5). The minimum P value test⁶⁷ was used to combine P values of the aforementioned six sets of rare variants at the gene or gene set level. For gene-based association tests, only genes with multiple rare variants after masking were tested for the corresponding rare variant category. In total, 15,935 genes were tested for at least one variant category. For gene set association tests, we compiled 20 gene sets of aging pathways for nine aging hallmarks¹ (Supplementary Table 6) and used the burden test and SKAT to test the longevity association of those six sets of rare variants within each of those 20 gene sets. FDR was used to correct for 130,297 P values at the variant level, 31,870 ($2 \times 15,935$) combined P values at the gene level and 40 (2×20) combined P values at the gene set level. For rare variant association at the gene set level, we conducted an independent analysis using the same framework but in two subcohorts, *APOE4*⁺ and *APOE4*[−]. FDR was used to correct for 80 ($2 \times 2 \times 20$) combined P values in this analysis. Gender and top ten MDS were included as covariates in all rare variant association analyses in the discovery cohort.

Network/pathway enrichment of rare variants. In addition to conventional approaches to the study of rare variant association, we investigated whether longevity-associated rare variants aggregate in a gene network and pathways. We first used IGSP³⁰ to score longevity-associated genes by integration of rare-variant association tests at the gene level with gene functional network³¹. To consider information of all rare coding variants in IGSP scoring, we collected gene association signals by applying the weighted burden test (using the R package SKAT) on rare coding variants of each gene weighted by the corresponding CADD scores. We then tested whether the top 100 genes tend to be scored higher than the top 100 genes derived from randomized rare variant association signals using the Wilcoxon rank-sum test. To investigate enriched pathways of those top 100 genes implicated by longevity association of rare variants and the functional gene network in an unbiased manner, we first performed pathway enrichment analysis using ToppGene Suite⁶⁸, in which 1,245 pathways from different pathway databases were analyzed concurrently to summarize the top enriched pathways

across pathway databases. In addition, we compared the top enriched KEGG and Reactome pathways identified by ToppGene Suite and three other widely used tools for pathway enrichment analysis—Enrichr⁶⁹, g:Profiler⁷⁰ and GSEA⁷¹—to derive enriched pathways supported by multiple analysis tools.

Lifespan analysis of rare variants. In our longevity cohort, after removal of kinship relatedness we had the date of death—and thus definitive lifespan information—for 553 Ashkenazi Jewish individuals (202 males and 351 females) (Table 1 and Extended Data Figs. 8 and 10), among which were 550 (~99.5%) individuals with a lifespan ≥ 65 years. Since no censored data were included in our lifespan cohort—that is, all subjects reached the endpoint (death)—for all lifespan analyses of rare variants we tested the association between lifespan and the burden of rare variants in the lifespan cohort using a unified accelerated life linear model⁷² with the log-transformed age at death as the outcome and gender as a covariate. Different from rare variant association analyses that aim to discover longevity-associated rare variants using a longevity case-control design, our lifespan analyses of rare variants investigated how pathogenic and protective rare variants discovered in our case-control study impact human lifespan through quantitative analyses.

Pathogenic rare variants and lifespan. We investigated whether pathogenic rare variants can adversely affect lifespan. We used PrimateAI⁷³, which was specially designed and optimized for prediction of disease-causing variants⁷³, to select highly pathogenic rare coding variants using a stringent score threshold of ≥ 0.9 , and assessed how the total count of their alternative alleles (the exome-wide burden) might affect lifespan. PrimateAI is a machine-learning-based method that expands the dataset for training by inclusion of common variants from nonhuman primates to improve the power for prediction of human pathogenic variants. No direct comparison of variant effects on longevity was made between human and nonhuman primates using PrimateAI.

Protective rare variants and lifespan. Our rare-variant association tests uncovered a burden of rare variants in Wnt signaling genes that may have pro-longevity effects among *APOE4*⁺ (Supplementary Table 14). We investigated their impact on lifespan by examining those protective rare variants in our lifespan cohort through several analyses. We evaluated whether the alternative allele count of those protective rare variants in Wnt signaling genes is correlated with lifespan among *APOE4*⁺ and *APOE4*⁻; from a complementary angle, we investigated whether *APOE4* differentially affects the lifespan of individuals with a high or low burden of those protective rare variants; and, finally, we also examined centenarians and noncentenarians separately in the lifespan analysis to differentiate it from the association study in which longevity status was used.

Replication studies. To maximize the extent of replicating human longevity association of rare variants, we prepared longevity case-control replication studies using cohort-specific criteria of determination of longevity cases and controls. First, longevity cases are individuals older than human life expectancy. Second, longevity controls are individuals substantially (>15 years) younger than cases. We used the WES data from three cohorts—a German longevity cohort, a UK Biobank longevity cohort and a longevity cohort from ADSP—to replicate the longevity association of rare variants discovered in our Ashkenazi Jewish longevity cohort. The German sample comprised 1,265 long-lived individuals (age range 94–110 years, mean age 99 years) as described previously⁷⁴, and 4,195 younger controls (mean age 35 years) recruited as part of the FoCuS cohort⁷⁵ and as blood donors at the University Hospital Schleswig-Holstein in Kiel and Lübeck, Germany. For exome sequencing and data analysis (including alignment and variant calling), the same wet lab processes and bioinformatic pipelines at the RGC were employed as for the Einstein cohort. The UK Biobank longevity cohort was collected from 49,960 individuals whole-exome sequenced in the UK Biobank⁷⁶, consisting of 104 cases and 23,405 controls of UK and white ethnicity with at least one long-lived parent (lifespan ≥ 100 years) (mean longest lifespan of parents, 101 years) and with parents of usual survival (lifespan <95 years) (mean longest lifespan of parents, 80 years), respectively. The ADSP longevity cohort consisted of 1,121 non-AD individuals aged ≥ 90 years (ADSP recorded age is right truncated at 90) as cases and 38 non-AD individuals aged <75 years as controls (mean age, 71 years). Both the German and UK Biobank longevity cohorts were used to replicate our findings in the full and *APOE4* stratified Ashkenazi Jewish longevity cohorts. The ADSP longevity cohort was used only to replicate findings made in the full discovery cohort, due to the limited number of its control samples.

Relatedness to first-degree kinship was removed from all three cohorts. We applied the same framework of rare variant association analysis from our discovery study in the replication analysis. We tested the six masking groups of not only rare variants (AAF $<1\%$) but also of ultra-rare variants (AAF $<0.05\%$), separately, which were not examined specifically in our discovery cohort due to the limited sample size of the allele reference panel in Ashkenazi Jews. The minimum P-value test was used to correct for 12 test P-values for a tested gene set accordingly. Rare variants in the three longevity cohorts were determined based on their AAF frequency in the corresponding WES data (5,460, 42,669 (UK and white ethnicity) and 10,267 individuals in the German, UK Biobank and

ADSP WES data, respectively). Ultra-rare variants were further determined from rare variants based on their AAF reported in the large non-Finnish European reference panel in gnomAD (v2; 56,885 individuals). Rare variants with genotype missing rates of ≥ 0.1 were excluded from our analyses. Gender and the top ten principal components from the principal component analyses accounting for the subpopulation structure were used as covariates in the burden test and SKAT. FDR was used to correct for four (two gene sets: Insulin and AMPK; two tests: SKAT and the burden test) and 12 (three gene sets: Insulin, AMPK and Wnt; two tests: SKAT; two subcohorts: *APOE4*⁺ and *APOE4*⁻) combined P-values for replication tests at the gene set level in the full cohort and *APOE4* stratified cohort, respectively.

We used the UK Biobank WES data of a parental lifespan cohort to replicate the relationship between pathogenic rare variants and lifespan discovered in our Ashkenazi Jewish lifespan cohort, due to the lack of long-lived individuals with WES data (the longest lifespan was ~80 years). After removal of relatedness to first-degree kinship, this cohort consisted of 20,823 individuals of UK and white ethnicity with average parental lifespan ≥ 65 years. We used the same regression framework for replication as for discovery, including the top ten principal components as covariates.

Statistics and reproducibility. No statistical methods were used to predetermine sample size, because all available samples from the WES data were considered. We used various statistical methods to analyze the data; please see the Methods subsections above for details. We used three independent longevity cohorts in which we successfully replicated our findings on the longevity association of rare variants in aging pathways. No data were excluded from the analyses. The experiments were not randomized because this approach was not relevant to the study design. The investigators were not blinded to allocation during experiments and outcome assessment as this was not relevant to the study design.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All summary statistics for the longevity association of rare coding variants in our Ashkenazi Jewish longevity cohort are available at <http://zdslab.einsteinmed.org/1/longevity.html>. Due to privacy concerns for our research participants, individual-level genetic data from the Einstein longevity study are not publicly available; however, anonymized data will be shared by request from a qualified academic investigator, providing the data transfer is approved by the Institutional Review Board and regulated by a material transfer agreement. The German longevity cohort data are part of the PopGen Biobank (Schleswig-Holstein, Germany) and can be accessed through a Material Data Access Form (<http://www.uksh.de/p2n/Information+for+Researchers.html>). Sequence and phenotype data of the UK Biobank and ADSP cohorts are available at <https://bbams.ndph.ox.ac.uk/ams/> and https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000572, respectively. All software used in our analyses was open source and is described in Methods.

Received: 8 December 2020; Accepted: 5 August 2021;
Published online: 13 September 2021

References

- Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–1217 (2013).
- Kenyon, C. J. The genetics of ageing. *Nature* **464**, 504–512 (2010).
- Kirkwood, T. B. Understanding the odd science of aging. *Cell* **120**, 437–447 (2005).
- Campisi, J. et al. From discoveries in ageing research to therapeutics for healthy ageing. *Nature* **571**, 183–192 (2019).
- Kenyon, C., Chang, J., Gensch, E., Rudner, A. & Tabtiang, R. A. C. *elegans* mutant that lives twice as long as wild type. *Nature* **366**, 461–464 (1993).
- Ayyadevara, S., Alla, R., Thaden, J. J. & Shmookler Reis, R. J. Remarkable longevity and stress resistance of nematode PI3K-null mutants. *Aging Cell* **7**, 13–22 (2008).
- Tatar, M. et al. A mutant *Drosophila* insulin receptor homolog that extends life-span and impairs neuroendocrine function. *Science* **292**, 107–110 (2001).
- Clancy, D. J. et al. Extension of life-span by loss of CHICO, a *Drosophila* insulin receptor substrate protein. *Science* **292**, 104–106 (2001).
- Holzenberger, M. et al. IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. *Nature* **421**, 182–187 (2003).
- Johnson, S. C., Rabinovitch, P. S. & Kaeblerlein, M. mTOR is a key modulator of ageing and age-related disease. *Nature* **493**, 338–345 (2013).
- Herskind, A. M. et al. The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870–1900. *Hum. Genet.* **97**, 319–323 (1996).
- Vijg, J. & Suh, Y. Genetics of longevity and aging. *Annu. Rev. Med.* **56**, 193–212 (2005).

13. Christensen, K., Johnson, T. E. & Vaupel, J. W. The quest for genetic determinants of human longevity: challenges and insights. *Nat. Rev. Genet.* **7**, 436–448 (2006).
14. Perls, T. T., Tubrick, E., Wager, C. G., Vijg, J. & Kruglyak, L. Siblings of centenarians live longer. *Lancet* **351**, 1560 (1998).
15. Melzer, D., Pilling, L. C. & Ferrucci, L. The genetics of human ageing. *Nat. Rev. Genet.* **21**, 88–101 (2020).
16. Zhang, Z. D. et al. Genetics of extreme human longevity to guide drug discovery for healthy ageing. *Nat. Metab.* **2**, 663–672 (2020).
17. Deelen, J. et al. Gene set analysis of GWAS data for human longevity highlights the relevance of the insulin/IGF-1 signaling and telomere maintenance pathways. *Age* **35**, 235–249 (2013).
18. Broer, L. et al. GWAS of longevity in CHARGE consortium confirms APOE and FOXO3 candidacy. *J. Gerontol. A* **70**, 110–118 (2015).
19. Deelen, J. et al. A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nat. Commun.* **10**, 3669 (2019).
20. Cash, T. P. et al. Exome sequencing of three cases of familial exceptional longevity. *Ageing Cell* **13**, 1087–1090 (2014).
21. Nygaard, H. B. et al. Whole-exome sequencing of an exceptional longevity cohort. *J. Gerontol. A Biol. Sci. Med. Sci.* **74**, 1386–1390 (2018).
22. Shindyapina, A. V. et al. Germline burden of rare damaging variants negatively affects human healthspan and lifespan. *eLife* **9**, e53449 (2020).
23. Guha, S. et al. Implications for health and disease in the genetic signature of the Ashkenazi Jewish population. *Genome Biol.* **13**, R2 (2012).
24. Hunt, R. C., Simhadri, V. L., Iandoli, M., Sauna, Z. E. & Kimchi-Sarfaty, C. Exposing synonymous mutations. *Trends Genet.* **30**, 308–321 (2014).
25. Lee, S., Abecasis, G. R., Boehnke, M. & Lin, X. Rare-variant association analysis: study designs and statistical tests. *Am. J. Hum. Genet.* **95**, 5–23 (2014).
26. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).
27. Sundaram, L. et al. Predicting the clinical impact of human mutation with deep neural networks. *Nat. Genet.* **50**, 1161–1170 (2018).
28. Fafian-Labora, J. et al. FASN activity is important for the initial stages of the induction of senescence. *Cell Death Dis.* **10**, 318 (2019).
29. Brosh, R. M. Jr. & Bohr, V. A. Human premature aging, DNA repair and RecQ helicases. *Nucleic Acids Res.* **35**, 7527–7544 (2007).
30. Lin, J. R., Zhang, Q., Cai, Y., Morrow, B. E. & Zhang, Z. D. Integrated rare variant-based risk gene prioritization in disease case-control sequencing studies. *PLoS Genet.* **13**, e1007142 (2017).
31. Tasan, M. et al. Selecting causal genes from genome-wide association studies via functionally coherent subnetworks. *Nat. Methods* **12**, 154–159 (2015).
32. Lupton, M. K. et al. The role of ABCA1 gene sequence variants on risk of Alzheimer's disease. *J. Alzheimers Dis.* **38**, 897–906 (2014).
33. Sims, R. et al. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nat. Genet.* **49**, 1373–1384 (2017).
34. Liu, D. J. & Leal, S. M. Replication strategies for rare variant complex trait association studies via next-generation sequencing. *Am. J. Hum. Genet.* **87**, 790–801 (2010).
35. Beecham, G. W. et al. The Alzheimer's Disease Sequencing Project: study design and sample selection. *Neurol. Genet.* **3**, e194 (2017).
36. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
37. Mahley, R. W. Apolipoprotein E: from cardiovascular disease to neurodegenerative disorders. *J. Mol. Med.* **94**, 739–746 (2016).
38. Erikson, G. A. et al. Whole-genome sequencing of a healthy aging cohort. *Cell* **165**, 1002–1011 (2016).
39. Eichner, J. E. et al. Apolipoprotein E polymorphism and cardiovascular disease: a HuGE review. *Am. J. Epidemiol.* **155**, 487–495 (2002).
40. Liu, H. et al. Augmented Wnt signaling in a mammalian model of accelerated aging. *Science* **317**, 803–806 (2007).
41. Kirkwood, T. B. & Finch, C. E. Ageing: the old worm turns more slowly. *Nature* **419**, 794–795 (2002).
42. Kirkwood, T. B. et al. What accounts for the wide variation in life span of genetically identical organisms reared in a constant environment? *Mech. Ageing Dev.* **126**, 439–443 (2005).
43. Caruso, A. et al. Inhibition of the canonical Wnt signaling pathway by apolipoprotein E4 in PC12 cells. *J. Neurochem.* **98**, 364–371 (2006).
44. Klaus, A. & Birchmeier, W. Wnt signalling and its impact on development and cancer. *Nat. Rev. Cancer* **8**, 387–398 (2008).
45. Palomer, E., Buechler, J. & Salinas, P. C. Wnt signaling deregulation in the aging and Alzheimer's brain. *Front. Cell Neurosci.* **13**, 227 (2019).
46. Ng, L. F. et al. WNT signaling in disease. *Cells* **8**, 826 (2019).
47. Timmers, P. R. et al. Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. *eLife* **8**, e39856 (2019).
48. Yang, J. et al. MiR-34 modulates *Caenorhabditis elegans* lifespan via repressing the autophagy gene atg9. *Age* **35**, 11–22 (2013).
49. Piazzesi, A. et al. Replication-independent histone variant H3.3 controls animal lifespan through the regulation of pro-longevity transcriptional programs. *Cell Rep.* **17**, 987–996 (2016).
50. Reichwald, K. et al. High tandem repeat content in the genome of the short-lived annual fish *Nothobranchius furzeri*: a new vertebrate model for aging research. *Genome Biol.* **10**, R16 (2009).
51. Course, M. M. et al. Evolution of a human-specific tandem repeat associated with ALS. *Am. J. Hum. Genet.* **107**, 445–460 (2020).
52. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
53. Dewey, F. E. et al. Inactivating variants in ANGPTL4 and risk of coronary artery disease. *N. Engl. J. Med.* **374**, 1123–1133 (2016).
54. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
55. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
56. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: polygenic risk score software. *Bioinformatics* **31**, 1466–1468 (2015).
57. Choi, S. W. & O'Reilly, P. F. PRSice-2: polygenic risk score software for biobank-scale data. *Gigascience* **8**, giz082 (2019).
58. Jansen, I. E. et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).
59. Nikpay, M. et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
60. Mahajan, A. et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
61. Malik, R. et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537 (2018).
62. Schumacher, F. R. et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat. Genet.* **50**, 928–936 (2018).
63. Michailidou, K. et al. Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92–94 (2017).
64. Wolpin, B. M. et al. Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. *Nat. Genet.* **46**, 994–1000 (2014).
65. Wang, X. Firth logistic regression for rare variant association tests. *Front. Genet.* **5**, 187 (2014).
66. Wu, M. C. et al. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* **89**, 82–93 (2011).
67. Flannick, J. et al. Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature* **570**, 71–76 (2019).
68. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37**, W305–W311 (2009).
69. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
70. Raudvere, U. et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).
71. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
72. Oakes, D. & Peterson, D. R. Survival methods: additional topics. *Circulation* **117**, 2949–2955 (2008).
73. Chen, H. et al. Comprehensive assessment of computational algorithms in predicting cancer driver mutations. *Genome Biol.* **21**, 43 (2020).
74. Flachsbar, F. et al. ImmunoChip analysis identifies association of the RAD50/IL13 region with human longevity. *Ageing Cell* **15**, 585–588 (2016).
75. Muller, N. et al. IL-6 blockade by monoclonal antibodies inhibits apolipoprotein (a) expression and lipoprotein (a) synthesis in humans. *J. Lipid. Res.* **56**, 1034–1042 (2015).
76. Van Hout, C. V. et al. Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* **586**, 749–756 (2020).

Acknowledgements

This work was supported by NIH grant nos. R01 HG008153 (Z.D.Z.), R01 AG061155 (S.M.), R01 AG060747 (M.D.G.), R01 AG057909 (N.B. and Z.D.Z.), P01 AG017242 (J.V.) and U19 AG056278 (J.V., P.D.R., L.J.N., Y.S. and W.C.L.) and a Career Scientist Award from the Irma T. Hirsch Trust to Z.D.Z. We thank the the Popgen Biobank and the Popgen 2.0 Network at Kiel University for help with recruitment of some of the long-lived individuals. G.G.T. was supported by the Deutsche Forschungsgemeinschaft (German Research Foundation) through project no. 390870439 (EXC 2150 – ROOTS). We thank T. Wang (Albert Einstein College of Medicine) for comments and suggestions. We thank the Management and Leadership Team at RGC for contributing to securing funding, study design and oversight and reviewing the manuscript (G. Abecasis, A. Baras,

M. Cantor, G. Coppola, A. Deubler, A. Economides, L. A. Lotta, J. D. Overton, J. G. Reid and A. Shuldiner). We thank Sequencing and Lab Operations at RGC for performing and being responsible for sample sequencing (J. Marcovici, E. Weiheing, A. Lopez and J. D. Overton); for performing and being responsible for exome sequencing (A. DeVito, J. LaRosa, L. Widom, C. Beechert, C. Forsythe, E. D. Fuller, M. Lattari, M. Sotiropoulos Padilla, S. E. Wolf, A. Lopez and J. D. Overton); for conceiving and being responsible for laboratory automation (T. D. Schleicher, Z. Gu, A. Lopez and J. D. Overton); and for being responsible for sample tracking and the library information management system (M. Pradhan, K. Manoochehri, R. H. Ulloa and J. D. Overton). We thank Genome Informatics at RGC for performing, and being responsible for, the analysis needed to produce exome and genotype data (X. Bai, A. Hawes, W. Salerno and J. G. Reid); for providing computing infrastructure development and operational support (G. Eom and J. G. Reid); for providing variant and gene annotations and their functional interpretation of variants (S. Balasubramanian and J. G. Reid); and for conceiving and being responsible for creating, developing and deploying analysis platforms and computational methods for analysis of genomic data (E. K. Maxwell, J. C. Staples, L. Habegger and J. G. Reid). We thank Research Program Management at RGC for contributing to the management and coordination of all research activities, planning, execution and reviewing of the manuscript (M. B. Jones and L. J. Mitnaul).

Author contributions

J.-R.L. and Z.D.Z. conceived the formal analysis. J.-R.L. executed the formal analysis. P.S.-C. and A.S. obtained the study resources. Z.D.Z., P.S.-C., J.M., Q.Z. and T.G. performed data curation. Z.W. performed variant imputation. V.N., G.G.T., M.D.G.,

A.F., A.N. and S.G. participated in replication analysis. Z.D.Z. and N.B. conceived of the research goals and acquired funding. J.-R.L. and Z.D.Z. wrote the original draft. J.V., Y.S., S.M., P.D.R., L.J.N., W.C.L., V.G., K.Y., G.A., M.L., M.R.J. and N.N. participated in review and editing. The R.G.C. performed WES and SNP array genotyping.

Competing interests

J.V. is a founder of Singulomics Corp. P.D.R. and L.J.N. are cofounders of NRTK Biosciences. All other authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s43587-021-00108-5>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43587-021-00108-5>.

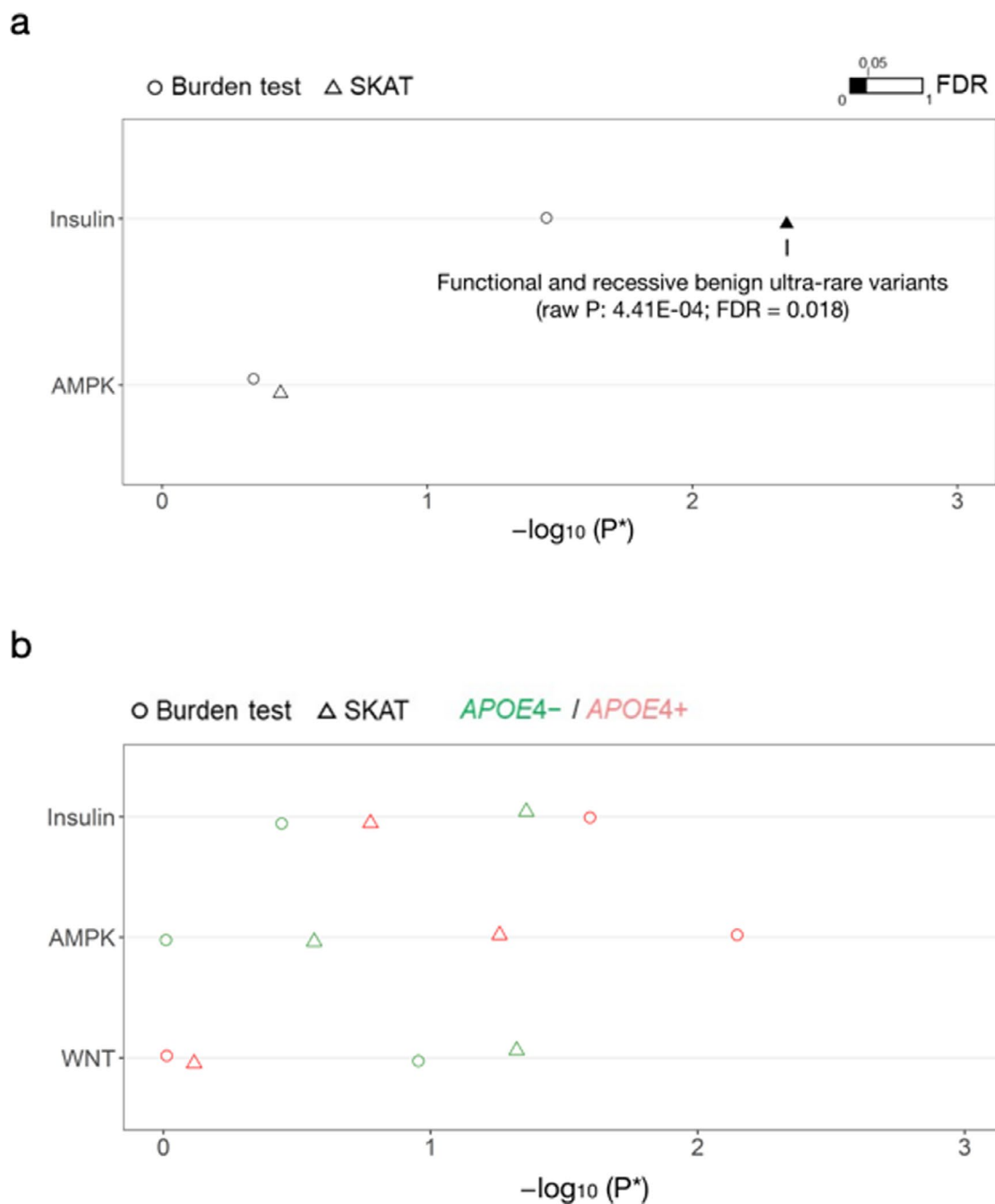
Correspondence and requests for materials should be addressed to Zhengdong D. Zhang.

Peer review information *Nature Aging* thanks George Martin and the other, anonymous reviewer(s) for their contribution to the peer review of this work.

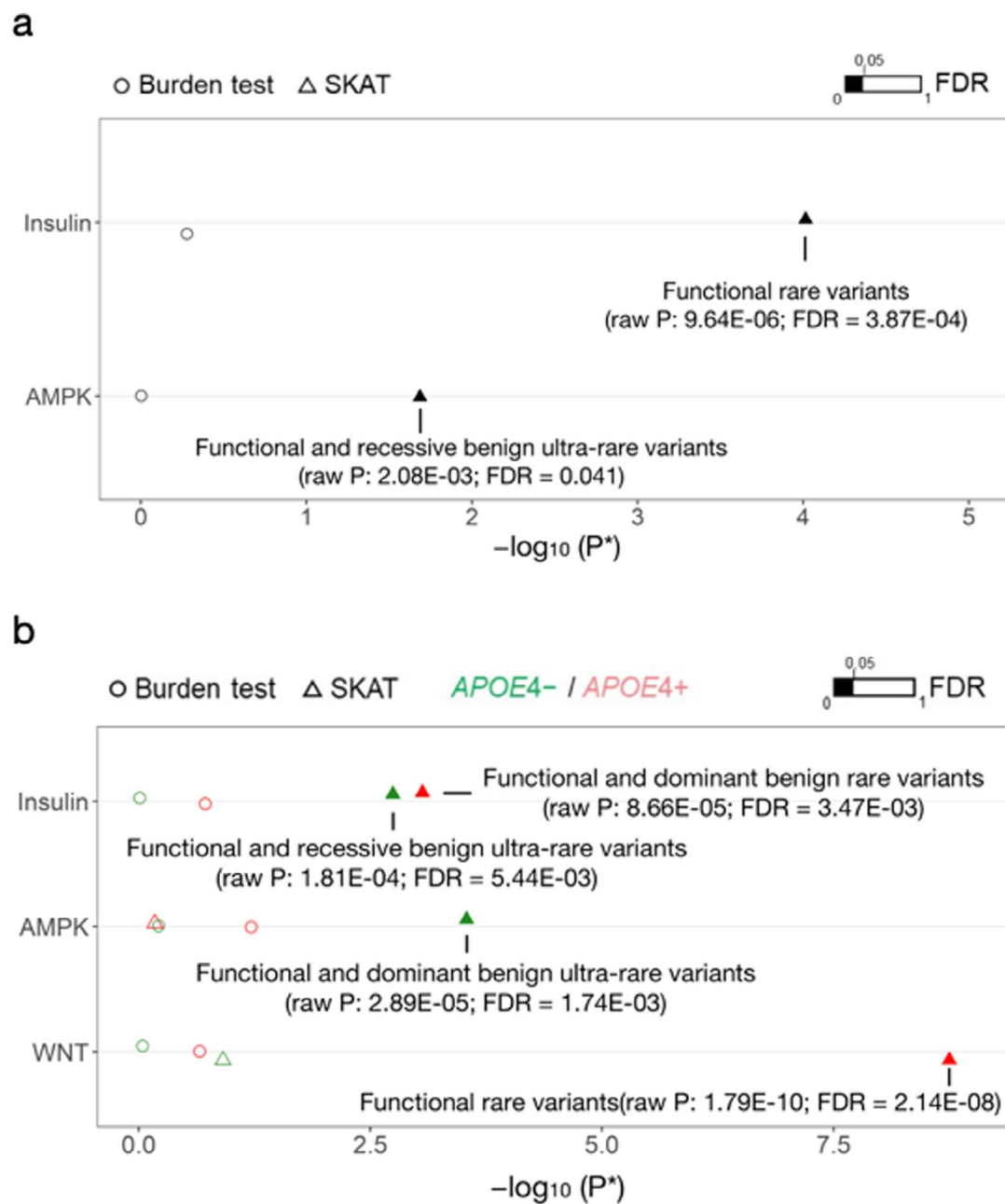
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

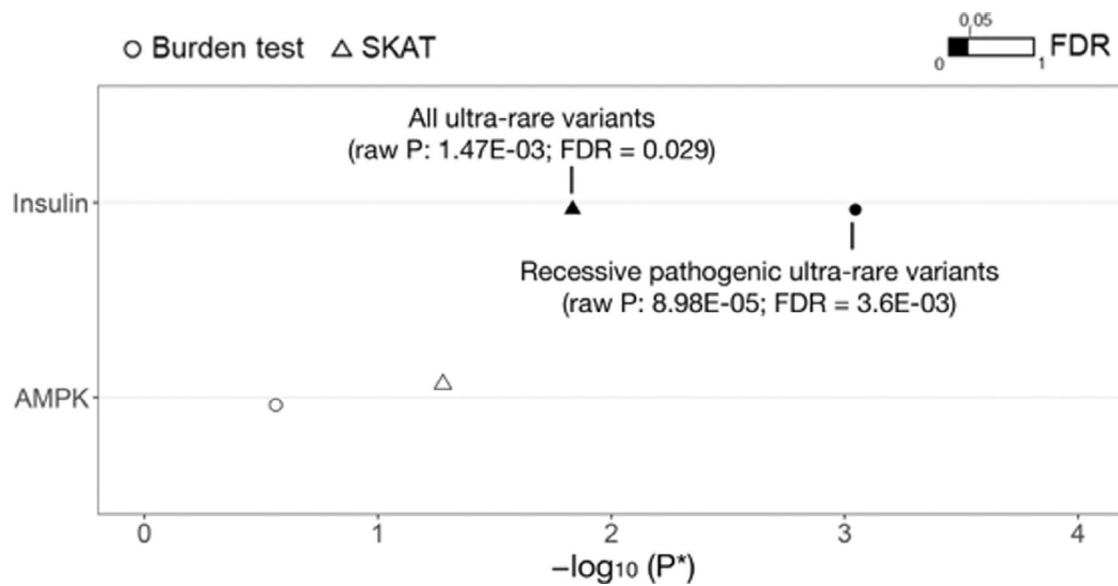
© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021



Extended Data Fig. 1 | The replication study of gene-set longevity association using the WES data of the German longevity cohort. The longevity case-control study consists of 1,265 longevity cases and 4,195 longevity controls. P^* denotes P -value corrected for 12 categories of rare variants using the minimal- P value test from Flannick et al.⁶⁷ (Methods). The text for the significant association denotes the lowest raw P -value among different groups of tested rare variants and FDR. **(A)** Full longevity cohort. **(B)** *APOE4* stratified cohorts.



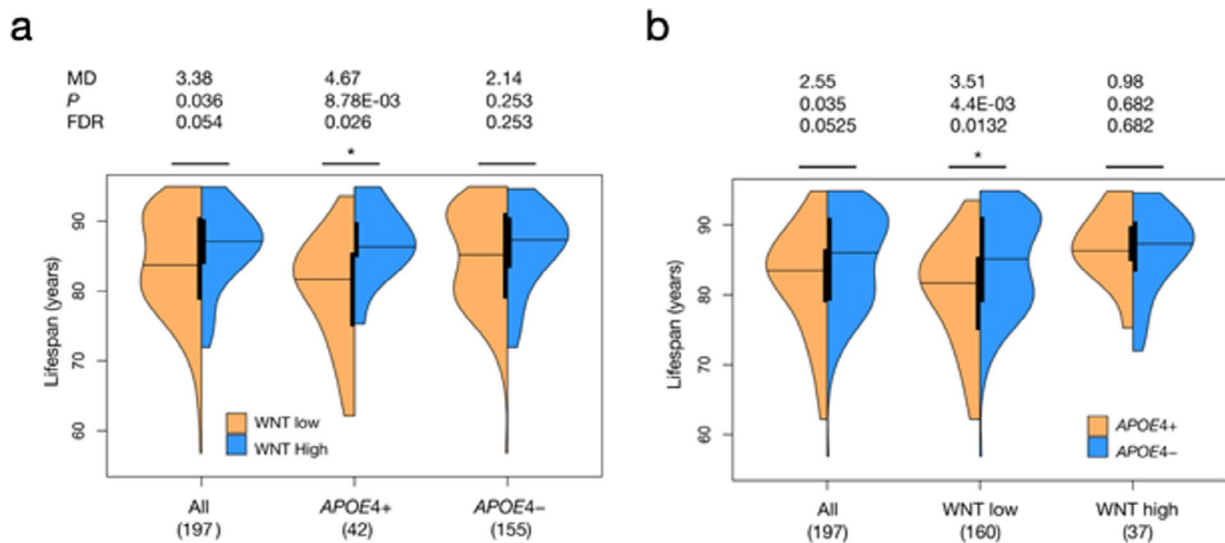
Extended Data Fig. 2 | The replication study of gene-set longevity association using the UK Biobank WES data. The longevity case-control study consists of 104 cases with at least one parent age at death ≥ 100 years and 23,405 controls with both parent age at death < 95 years. P^* denotes P -value corrected for 12 categories of rare variants using the minimal- P value test from Flannick et al.⁶⁷ (Methods). The text for the significant association denotes the lowest raw P -value among different groups of tested rare variants and FDR. **(a)** Full longevity cohort. **(b)** *APOE4* stratified cohorts.



Extended Data Fig. 3 | The replication study of gene-set longevity association using the ADSP WES data. The longevity case-control study consists of 1,121 non-AD cases with age ≥ 90 years and 38 non-AD controls with age < 75 years. P^* denotes P -value corrected for 12 categories of rare variants using the minimal- P value test from Flannick et al.⁶⁷ (Methods). The text for the significant association denotes the lowest raw P -value among different groups of tested rare variants and FDR.

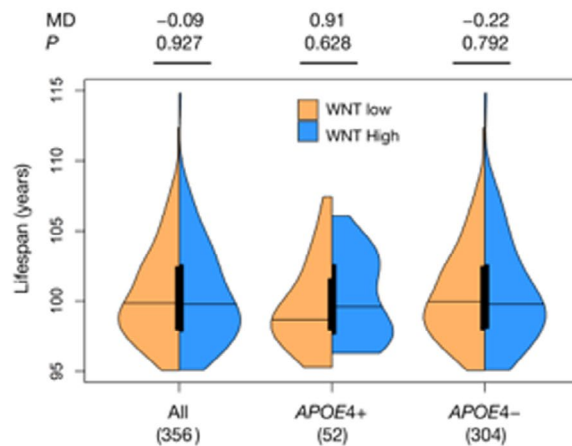


Extended Data Fig. 4 | Gene-set rare variant association in the *APOE4*-stratified cohorts of the discovery (Ashkenazi Jewish) longevity cohort. P* denotes *P*-value corrected for 6 categories of tested variants using the minimal-*P* value test from Flannick et al.⁶⁷ (Methods). The text for the significant association denotes the lowest raw *P*-value among different groups of tested rare variants and FDR.

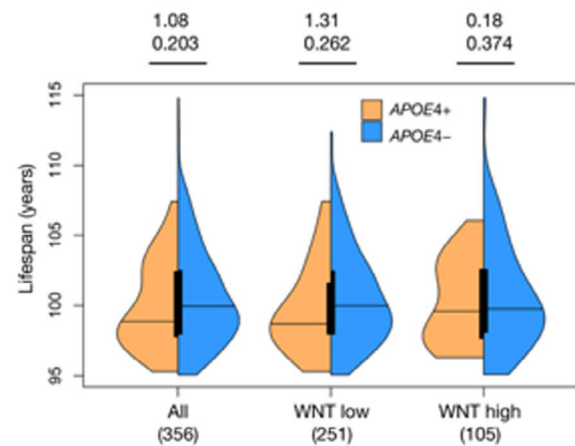


Extended Data Fig. 5 | Lifespan analysis of protective variants in WNT signaling genes for noncentenarians. *P* denotes uncorrected *P*-value derived from linear regression with the log-transformed age at death as the outcome and the gender as a covariate (See Methods). ‘WNT low’ and ‘WNT high’ represent the alternative allele count of rare variants in WNT signaling genes ≤ 1 and > 1 (the median), respectively. In parentheses are the numbers of individuals. MD stands for ‘median difference’. The asterisk denotes $\text{FDR} < 0.05$. **(a)** The lifespan difference of individuals carrying a high and low burden of protective rare variants in WNT signaling genes. **(b)** Negative effects of *APOE4* on lifespan with high and low burden of protective rare variants in WNT signaling for noncentenarians.

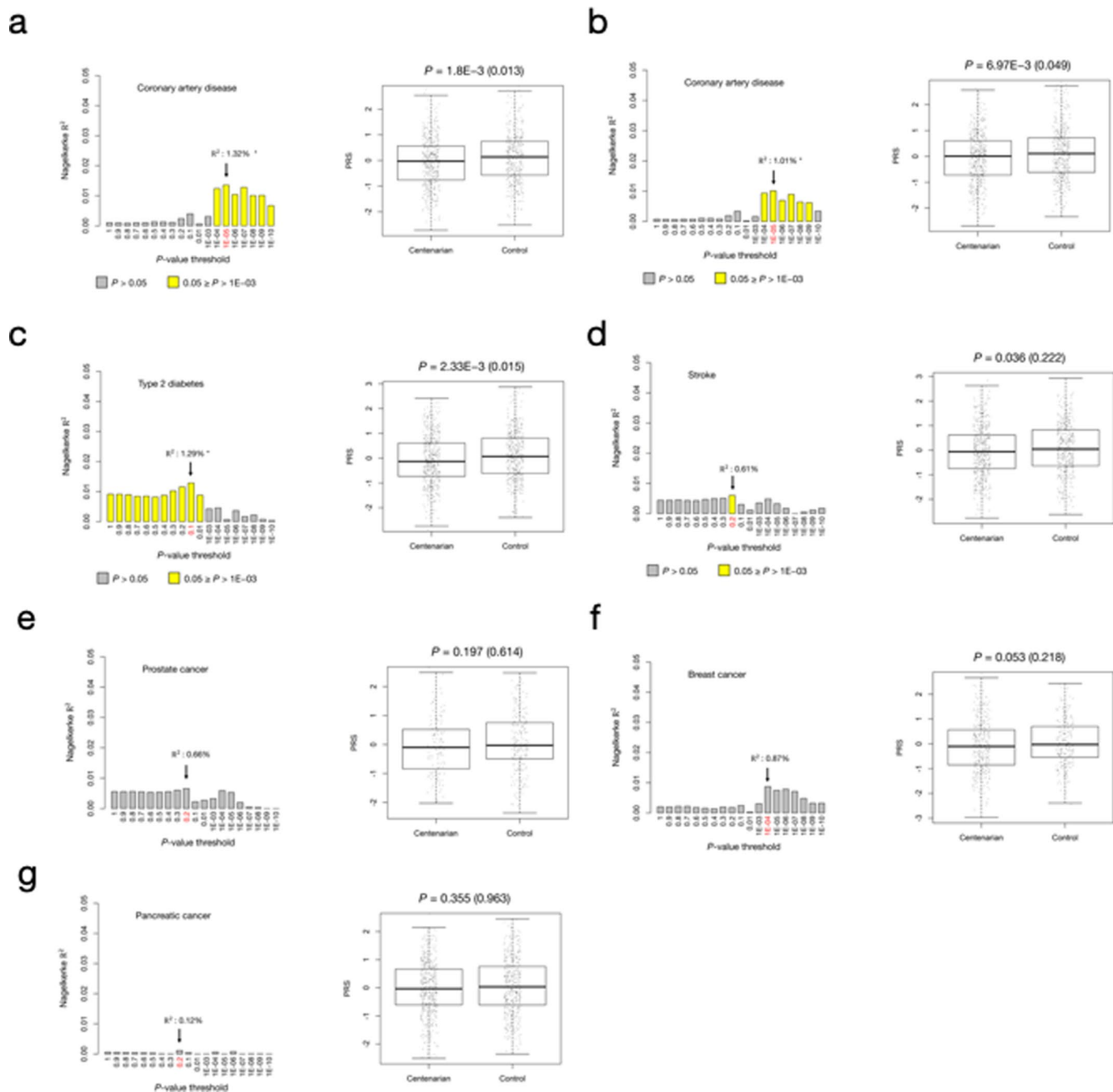
a



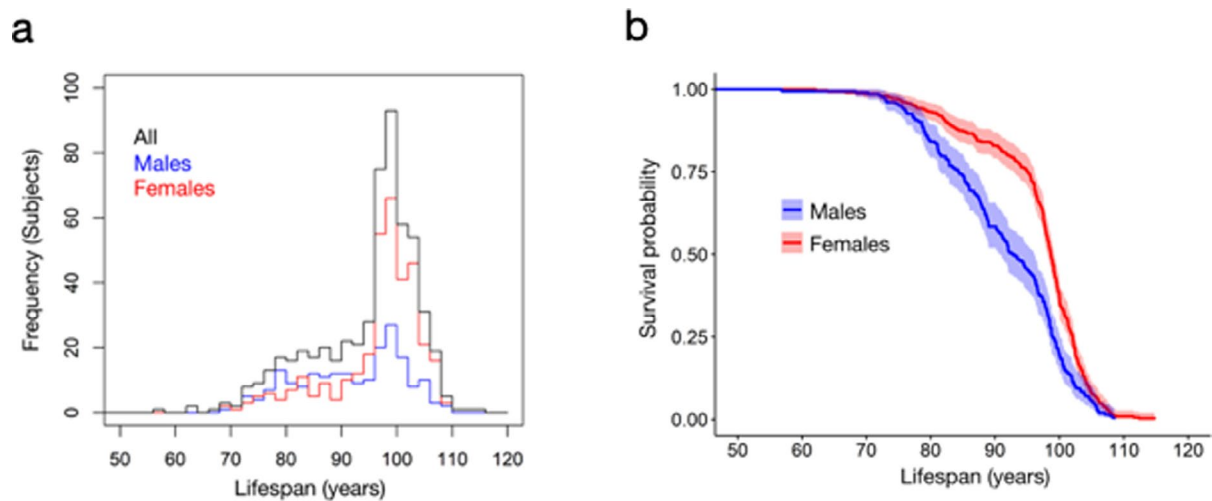
b



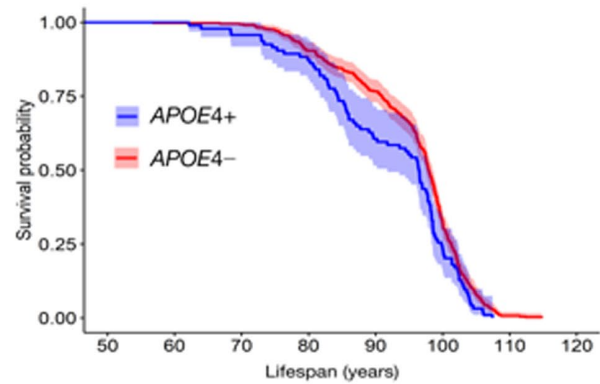
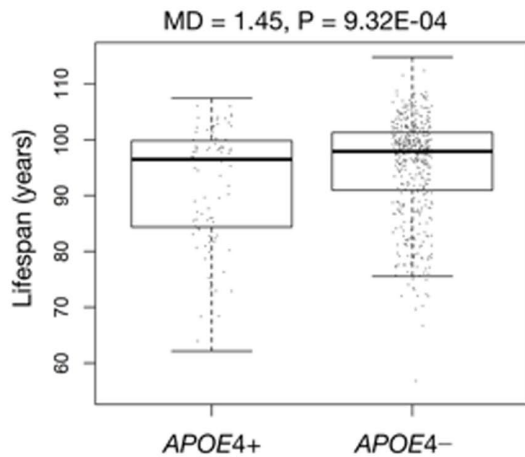
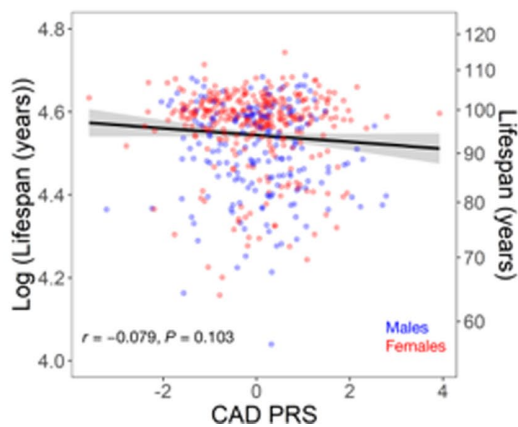
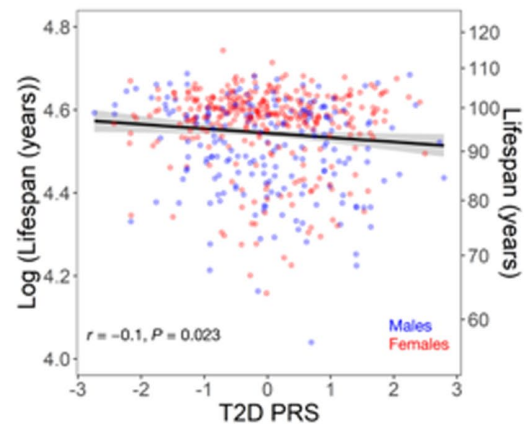
Extended Data Fig. 6 | Lifespan analysis of protective variants in WNT signaling genes for centenarians. *P* denotes uncorrected *P*-value derived from linear regression with the log-transformed age at death as the outcome and the gender as a covariate (See Methods). ‘WNT low’ and ‘WNT high’ represent the alternative allele count of rare variants in WNT signaling genes ≤ 1 and > 1 (the median), respectively. In parentheses are the numbers of individuals. MD stands for ‘median difference’. **(a)** The lifespan difference of individuals carrying a high and low burden of protective rare variants in WNT signaling genes. **(b)** Negative effects of APOE4 on lifespan with high and low burden of protective rare variants in WNT signaling for centenarians.



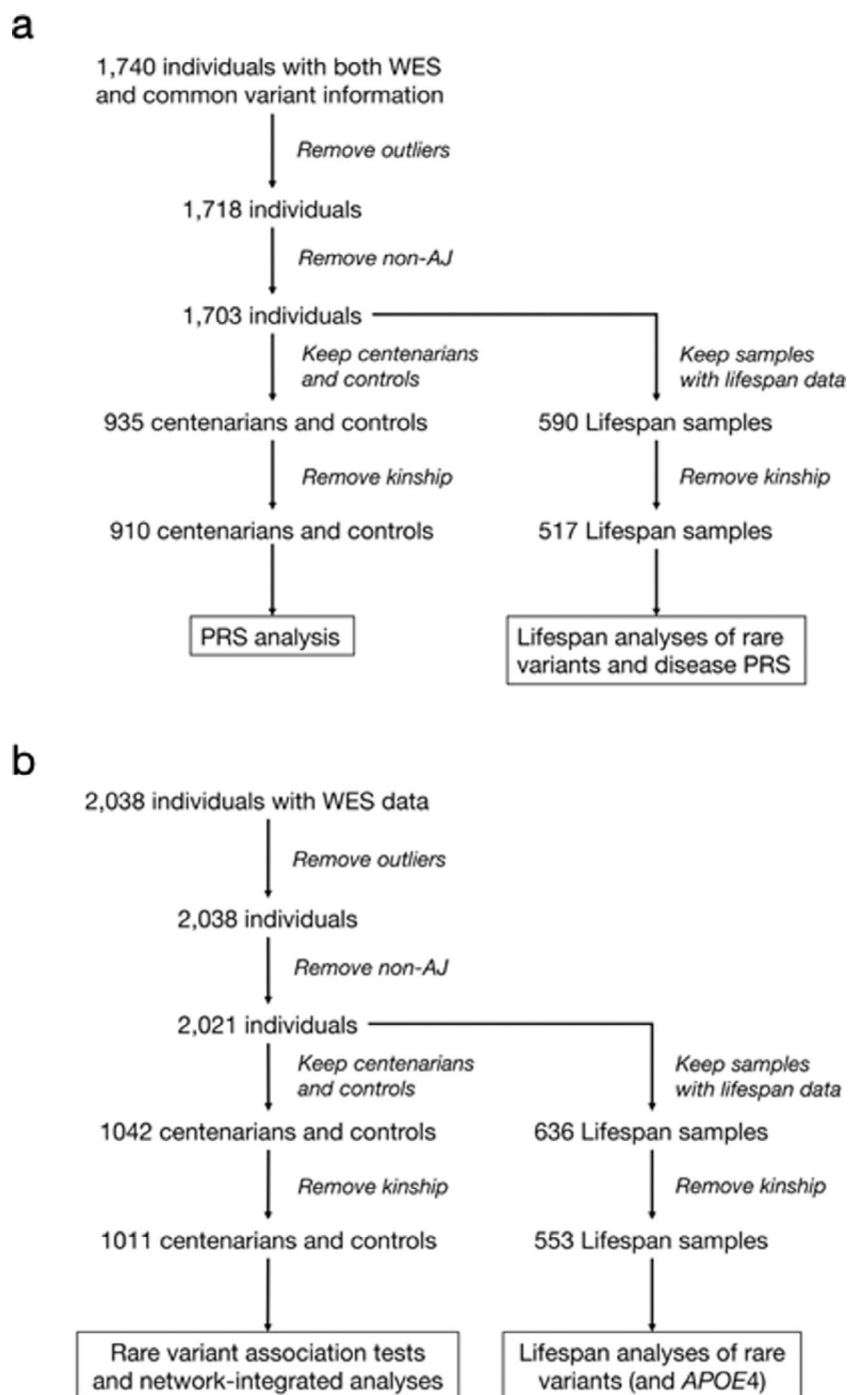
Extended Data Fig. 7 | Disease-PRS analyses for centenarian and control. This shows the results of PRS analyses for age-related diseases in the centenarian cohort. In the boxplots, points represent individuals, and horizontal lines represent upper fence (maximum in $Q3 + 1.5 \times IQR$), upper quartile ($Q3$), median, lower quartile ($Q1$), lower fence (minimum in $Q1 - 1.5 \times IQR$), sequentially from top to bottom; IQR : interquartile range (25th to the 75th percentile). $n = 910$ biologically independent samples in the boxplots on the right panels for coronary artery disease, type 2 diabetes, stroke, and pancreatic cancer. $n = 339$ and 571 biologically independent samples in the boxplots on the right panels for prostate cancer and breast cancer, respectively. Above the boxplot on the right are raw and adjusted (in parentheses) P -values for the best prediction in the Nagelkerke's R^2 plot on the left, which were calculated based on logistic regression and the permutation test in PRSice-2, respectively. For stroke, breast cancer, prostate cancer, and pancreatic cancer, no robust association was observed between their PRS and the longevity status as originally defined in our cohort. **(a)** Coronary artery disease. **(b)** Coronary artery disease without considering SNPs within 1Mb of rs7412 or rs429358 (SNPs for the *APOE* haplotype). **(c)** Type 2 diabetes. **(d)** Stroke. **(e)** Prostate cancer. Only males are considered. **(f)** Breast cancer. Only females are considered. **(g)** Pancreatic cancer.



Extended Data Fig. 8 | Basic statistics of the lifespan cohort. (a) Lifespan distribution of 553 individuals. (b) Survival curves of 202 males and 351 females composing the analyzed cohort. Females have a significant survival rate than males based on cox regression model ($P=1.71\text{E-}07$; coxph in R package).

a**b****c**

Extended Data Fig. 9 | Correlation between lifespan and common-variant genetic risk of age-related diseases. *P*-values were based on the result of linear regression (regress log lifespan on genetic disease risk) corrected for gender. **(a)** Alzheimer's disease. The plots on the left and right show the boxplot and survival curves of $APOE4+$ and $APOE4-$, respectively. MD stands for 'Median Difference'. In the boxplots, points represent individuals, and horizontal lines represent upper fence (maximum in $Q3 + 1.5 \times IQR$), upper quartile ($Q3$), median, lower quartile ($Q1$), lower fence (minimum in $Q1 - 1.5 \times IQR$), sequentially from top to bottom; IQR : interquartile range (25th to the 75th percentile). $n = 553$ biologically independent samples. **(b)** Coronary artery disease. r represents 'correlation coefficient'. **(c)** Type 2 diabetes.



Extended Data Fig. 10 | Flowcharts of sample collection for different analyses. (a) Flowchart of sample collection for PRS analyses and lifespan analyses of rare variants and disease PRS. Refer 'Rare variant association analysis' subsection for the strategy of removing kinship for PRS analysis that involves longevity status. The strategy of removing kinship in lifespan analyses is to randomly exclude one in pairs of individuals with the proportion of alleles shared identity-by-descent (IBD) > 0.4. (b) Flowchart of sample collection for rare variant association tests, network-integrated analyses, and lifespan analyses of rare variants (and APOE4).

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|---|
| n/a | Confirmed |
| ☐ | ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☐ | ☒ A description of all covariates tested |
| ☐ | ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☒ | ☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☐ | ☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection No specific software was used to collect the data.

Data analysis All used softwares in our analyses were open source and were described in the Methods section of the manuscript. Plink 1.9, PRSice-2 2.1, SKAT (R package 1.3), CADD 1.4, PrimateAI 0.2, Michigan Imputation Server (<https://imputationserver.sph.umich.edu/index.html#!>), IGSP (<https://zenodo.org/record/1034362#.X-JWQNgzY2w>), ToppGene Suite (<https://toppgene.cchmc.org/>), Enrichr (<https://maayanlab.cloud/Enrichr/>), g:Profiler (<https://biit.cs.ut.ee/gprofiler/gost>), GSEA (<http://www.gsea-msigdb.org/gsea/msigdb/annotate.jsp>).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All summary statistics for the longevity association of rare coding variants in our Ashkenazi Jewish longevity cohort are available at <http://zdzlab.einsteinmed.org/1/longevity.html>. Due to privacy concerns for our research participants, individual-level genetic data from the Einstein longevity study are not publicly available; however, anonymized data will be shared by request from a qualified academic investigator as long as the data transfer is approved by the Institutional Review Board and regulated by a material transfer agreement. The German longevity cohort data are part of the PopGen Biobank (Schleswig-Holstein, Germany) and can be accessed through a Material Data Access Form (<http://www.uksh.de/p2n/Information+for+Researchers.html>). Sequence and phenotype data of the UK Biobank and

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size as all the available samples from the WES data were considered.
Data exclusions	No data were excluded with pre-established criteria in favor of results.
Replication	We used three independent longevity cohorts in which we successfully replicated our finding on longevity association of rare variant in aging pathways.
Randomization	The experiments were not randomized as this approach was not relevant to the study design.
Blinding	The investigators were not blinded to allocation during experiments and outcome assessment as this was not relevant to the study design.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a	Involved in the study	n/a	Involved in the study
☒	☐ Antibodies	☒	☐ ChIP-seq
☒	☐ Eukaryotic cell lines	☒	☐ Flow cytometry
☒	☐ Palaeontology and archaeology	☒	☐ MRI-based neuroimaging
☒	☐ Animals and other organisms		
☐	☒ Human research participants		
☒	☐ Clinical data		
☒	☐ Dual use research of concern		

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	The discovery cohort is Einstein Longevity cohort: Longevity cases were defined as individuals age ≥ 95 years, and individuals age < 95 without a parental history of longevity (neither parent survived beyond 95 years of age) as controls. To maximize the extent of replicating human longevity association of rare variants, we prepared three longevity case-control replication studies using cohort-specific criteria of determining longevity cases and controls. German longevity cohort: Longevity cases are long-lived individuals (age: 94-101 years), and controls are blood donors (mean age: 35 years); UK Biobank longevity cohort: Longevity cases are individuals with at least one long-lived parent (lifespan ≥ 100 years), and controls are individuals with parents of usual survival (lifespan < 95 years); ADSP longevity cohort: Longevity cases are individuals with age ≥ 90 years, and control are individuals with age < 75 years. In all longevity case-control studies, gender and top 10 principal components from PCA or MDS derived from common variants were considered as population characteristics and included as covariates in our analyses.
Recruitment	The discovery study subjects were Ashkenazi Jewish participants from two longevity cohorts, the Longevity Genes Project (LGP) and the LonGenity study, who were recruited and characterized at the Albert Einstein College of Medicine since 1998. The German longevity subjects consist of long lived individuals previously described (PMID: 27004735) and blood donors at the University Hospital Schleswig-Holstein in Kiel and Lübeck, German. The UK Biobank longevity subjects were recruited from general population as previously described (PMID: 33087929). The ADSP longevity subjects were recruited from a study

cohort for Alzheimer's disease.

Ethics oversight

The Einstein WES data: The Institutional Review Board at Albert Einstein College of Medicine

The German WES data: The Ethics Committee at the Medical Faculty of Kiel University

The UKB WES data: The Ethics Advisory Committee and the external ethics committees

The ADSP WES data: Broad Institute, the Baylor College of Medicine's Human Genome Sequencing Center, and Washington University's McDonnell Genome Institute

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval, sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com