



Article

Long-Lived Individuals Show a Lower Burden of Variants Predisposing to Age-Related Diseases and a Higher Polygenic Longevity Score

Guillermo G. Torres ^{1,†} , Janina Dose ^{1,†}, Tim P. Hasenbein ^{1,2,3} , Marianne Nygaard ^{4,5} , Ben Krause-Kyora ¹ , Jonas Mengel-From ^{4,5} , Kaare Christensen ^{4,5,6}, Karen Andersen-Ranberg ^{4,7}, Daniel Kolbe ¹, Wolfgang Lieb ⁸, Matthias Laudes ⁹, Siegfried Görg ¹⁰, Stefan Schreiber ¹, Andre Franke ¹, Amke Caliebe ¹¹, Gregor Kuhlenbäumer ² and Almut Nebel ^{1,*}

- ¹ Institute of Clinical Molecular Biology, Kiel University, University Hospital Schleswig-Holstein, Campus Kiel, Rosalind-Franklin-Str. 12, 24105 Kiel, Germany
 - ² Department of Neurology, Kiel University, University Hospital Schleswig-Holstein, Campus Kiel, Arnold-Heller-Str. 3, 24105 Kiel, Germany
 - ³ Institute of Pharmacology and Toxicology, Technical University Munich, Biedersteiner Str. 29, 80802 Munich, Germany
 - ⁴ Department of Public Health, Epidemiology, Biostatistics and Biodemography, University of Southern Denmark, J.B. Winsloews Vej 9B, 5000 Odense, Denmark
 - ⁵ Department of Clinical Genetics, Odense University Hospital, J.B. Winsloews Vej 4, 5000 Odense, Denmark
 - ⁶ Department of Clinical Biochemistry, Odense University Hospital, Klørvænget 47, 5000 Odense, Denmark
 - ⁷ Department of Geriatric Medicine, Odense University Hospital, Klørvænget 23, 5000 Odense, Denmark
 - ⁸ Institute of Epidemiology and Biobank Popgen, Kiel University, University Hospital Schleswig-Holstein, Campus Kiel, Niemannsweg 11, 24105 Kiel, Germany
 - ⁹ Clinic for Internal Medicine I, Division of Endocrinology, Diabetes and Clinical Nutrition, Kiel University, University Hospital Schleswig-Holstein, Campus Kiel, Arnold-Heller-Straße 3, 24105 Kiel, Germany
 - ¹⁰ Institute of Transfusion Medicine, University Hospital Schleswig-Holstein, Campus Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany
 - ¹¹ Institute of Medical Informatics and Statistics, Kiel University, University Hospital Schleswig-Holstein, Campus Kiel, Brunswiker Str. 10, 24105 Kiel, Germany
- * Correspondence: a.nebel@mucosa.de; Tel.: +49-(0)431-500-15155; Fax: +49-(0)431-500-15168
† These authors contributed equally to this work.



Citation: Torres, G.G.; Dose, J.; Hasenbein, T.P.; Nygaard, M.; Krause-Kyora, B.; Mengel-From, J.; Christensen, K.; Andersen-Ranberg, K.; Kolbe, D.; Lieb, W.; et al. Long-Lived Individuals Show a Lower Burden of Variants Predisposing to Age-Related Diseases and a Higher Polygenic Longevity Score. *Int. J. Mol. Sci.* **2022**, *23*, 10949. <https://doi.org/10.3390/ijms231810949>

Academic Editor: Irmgard Tegeder

Received: 9 August 2022

Accepted: 7 September 2022

Published: 19 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Longevity is a complex phenotype influenced by both environmental and genetic factors. The genetic contribution is estimated at about 25%. Despite extensive research efforts, only a few longevity genes have been validated across populations. Long-lived individuals (LLI) reach extreme ages with a relative low prevalence of chronic disability and major age-related diseases (ARDs). We tested whether the protection from ARDs in LLI can partly be attributed to genetic factors by calculating polygenic risk scores (PRSs) for seven common late-life diseases (Alzheimer's disease (AD), atrial fibrillation (AF), coronary artery disease (CAD), colorectal cancer (CRC), ischemic stroke (ISS), Parkinson's disease (PD) and type 2 diabetes (T2D)). The examined sample comprised 1351 German LLI (≥ 94 years, including 643 centenarians) and 4680 German younger controls. For all ARD-PRSs tested, the LLI had significantly lower scores than the younger control individuals (areas under the curve (AUCs): ISS = 0.59, $p = 2.84 \times 10^{-35}$; AD = 0.59, $p = 3.16 \times 10^{-25}$; AF = 0.57, $p = 1.07 \times 10^{-16}$; CAD = 0.56, $p = 1.88 \times 10^{-12}$; CRC = 0.52, $p = 5.85 \times 10^{-3}$; PD = 0.52, $p = 1.91 \times 10^{-3}$; T2D = 0.51, $p = 2.61 \times 10^{-3}$). We combined the individual ARD-PRSs into a meta-PRS (AUC = 0.64, $p = 6.45 \times 10^{-15}$). We also generated two genome-wide polygenic scores for longevity, one with and one without the *TOMM40/APOE/APOC1* gene region (AUC (incl. *TOMM40/APOE/APOC1*) = 0.56, $p = 1.45 \times 10^{-5}$, seven variants; AUC (excl. *TOMM40/APOE/APOC1*) = 0.55, $p = 9.85 \times 10^{-3}$, 10,361 variants). Furthermore, the inclusion of nine markers from the excluded region (not in LD with each other) plus the *APOE* haplotype into the model raised the AUC from 0.55 to 0.61. Thus, our results highlight the importance of *TOMM40/APOE/APOC1* as a longevity hub.

Keywords: longevity; PRS; healthy aging; age-related diseases

1. Introduction

Decades of extensive research on the etiology of human longevity have revealed the complex nature of this phenotype. In addition to healthy behavior, environment and chance, genetic factors have been shown to influence human longevity, with the genetic contribution estimated between 20–30% [1]. Despite an impressive variety of methodological approaches, the output of genetic longevity studies has been limited. Most of the longevity-associated variants exert only low to moderate effects. So far, just a relatively small number of genes (e.g., *APOE*, *FOXO3*, locus chr. 5q33.3, *CDKN2B*) have been confirmed to play a role in the phenotype across populations [2–6].

Long-lived individuals (LLI) reach extreme ages in relatively good health and are, therefore, considered models of healthy aging. Although LLI may suffer from comorbidities [7], they seem to age at a slower pace [4] and to avoid, postpone or survive age-related diseases (ARDs) [8]. It has not yet been conclusively clarified whether a lower genetic risk for ARDs contributes to longevity. Like longevity, most ARDs have a complex genetic architecture, but show a higher heritability than does longevity (e.g., up to 80% for type 2 diabetes (T2D) [9] and Alzheimer’s disease (AD) [10] and 40–60% for coronary artery disease (CAD) [11]). Studying genetic risk factors for major ARDs in LLI might help us identify variants relevant for both longevity and aging [12]. Taken further, a shared genetic architecture between longevity and ARDs might facilitate the prediction of an individual’s longevity potential. Data in the literature generally, although not consistently (e.g., [13]), strengthen a genetic link between longevity and ARDs [14–17]. For example, Fortney et al. successfully applied a disease-informed GWAS approach to detect new longevity genes [14]. They reported a large genetic overlap between longevity and AD and CAD, respectively, which is supported by other studies (e.g., [15,18–20]). For other ARDs, such as T2D, the evidence for a possible genetic link with human longevity is less clear (see, e.g., [14,15,20,21]).

The concept of “genome-wide polygenic score” (GPS) or “polygenic risk score” (PRS) has been successfully applied for several diseases, e.g., CAD, AD, and T2D [22]. GPSs for longevity have only been developed in a few studies so far [4,17,23]. Here, we applied published PRSs for seven common ARDs (AD, atrial fibrillation (AF), CAD, colorectal cancer (CRC), ischemic stroke (ISS), Parkinson’s disease (PD), and T2D) to a German longevity sample comprising 1351 LLI (≥ 94 years) including 643 centenarians, and 4680 younger controls (age range: 18–83 years, mean age: 50.5 years). We calculated the PRSs using genotyping data generated with the Illumina Infinium Global Screening Array-24 (GSAv1; 700,078 single-nucleotide variants (SNVs)). In addition, variants in and near the *APOE* gene have been repeatedly shown to have a strong negative impact on longevity [15] and thus could mask much smaller positive or negative effects of other SNVs. Therefore, we developed two new GPSs for longevity based on published summary statistics from the latest meta-GWAS on longevity [15], one with and one without considering SNVs in the *TOMM40/APOE/APOC1* gene region, and compared them with published longevity GPSs.

2. Results

2.1. Association Analyses Do Not Reveal New Longevity Loci

In the single-variant analyses, we considered either the whole data set or male/female and centenarian-only subsets, respectively (Supplementary Tables S1–S4). Most association signals disappeared after conditioning on variants in the *TOMM40/APOE/APOC1* region on chromosome 19, which is well-known for being negatively associated with human longevity [5,24]. Of note, also a signal in the gene *PVRL2* near *TOMM40/APOE/APOC1* did not survive our conditioning, supporting our previous observation that *PVRL2* most likely does not represent an independent genetic longevity-associated locus, but rather influences the phenotype via epigenetic mechanisms [25]. The remaining longevity-associated variants were either too rare to yield reliable results or they did not show an association in the Danish replication sample (Supplementary Table S5). In the gene-based analysis, only *TOMM40* remained significant after correction for multiple testing ($P_{\text{adj}} = 4.32 \times 10^{-2}$;

Supplementary Table S6). The poor outcome of the longevity association analyses highlights the need for multifactorial analyses and refined methodological strategies to unravel the etiology of human longevity.

2.2. ARD-PRS Distributions Show Significant Differences between LLI and Controls

We investigated the polygenic risk profile of LLI versus younger controls for seven common ARDs, namely CAD, AF, ISS, AD, T2D, CRC and PD, using published PRS models and the imputed and quality-controlled GSA dataset of our German sample. Fractions of 96.4% (AD), 93.2% (ISS), 89.3% (CAD), 88.7% (AF), 96.8% (CRC), 87% (T2D) and 95% (PD) of the published input-SNVs were covered by the imputed GSA data (Table 1). The calculated mean (or median) PRSs were always significantly lower for LLI compared with the younger controls (logistic model p -value < 0.05, Figure 1, Table 1). The potential of differentiating between LLI and controls, based on the single ARD-PRSs, was rather low, with the AUCs ranging between 0.51 and 0.59 (Table 1). The highest discrimination values were achieved by the AD-PRS and ISS-PRS (AUC: AD-PRS = 0.59, ISS-PRS = 0.58, Table 1). Sex and population substructure were found to be significant covariables in all PRS models (Supplementary Table S7), and their inclusion increased the discrimination between LLI and controls by 16% on average (Supplementary Table S8). This means that in our data, LLI and controls could be distinguished to a certain extent by sex and population stratification. Especially in terms of sex, this was to be expected due to the ratio of women to men of 2.7:1 in the sample of LLI (versus 0.7:1 in the controls). To improve the discrimination of LLI and controls based on ARD risk alleles, we used the single ARD-PRSs as influence variables and combined them into a meta-PRS. On the test dataset, this model exhibited an AUC of 0.64 (without accounting for sex and population substructure as covariables; Table 1) and yielded the coefficients shown in Supplementary Table S9. When we included sex and population substructure as covariables, the discriminatory capacity of the model increased by 10% (i.e., the AUC improved to 0.74; Supplementary Table S8). As expected, the inclusion of the covariables also increased the contribution of the single ARD-PRSs (Supplementary Table S10).

Table 1. Accuracy statistics for the calculation of each ARD-PRS in the German study population.

| Age-Related Disease | AUC ¹ | AUC_L95 | AUC_U95 | Beta ³ | OR ⁴ | p -Value ² | PRS-Input-SNVs ⁵ (No.) | Input-SNVs Covered ⁶ (No. (%)) |
|-----------------------|------------------|---------|---------|-------------------|-----------------|-------------------------|-----------------------------------|---|
| PD | 0.52 | 0.50 | 0.53 | −0.142 | 0.87 | 1.91×10^{-3} | 1805 | 1715 (95.01) |
| T2D | 0.51 | 0.50 | 0.53 | −0.056 | 0.95 | 2.61×10^{-3} | 6,917,436 | 6,024,432 (87.09) |
| CRC | 0.52 | 0.50 | 0.54 | −0.097 | 0.91 | 5.85×10^{-3} | 95 | 92 (96.84) |
| CAD | 0.56 | 0.54 | 0.57 | −0.159 | 0.85 | 1.88×10^{-12} | 6,630,150 | 5,920,526 (89.30) |
| AF | 0.57 | 0.55 | 0.58 | −0.170 | 0.84 | 1.07×10^{-16} | 6,730,541 | 5,973,364 (88.75) |
| ISS | 0.59 | 0.57 | 0.61 | −0.283 | 0.75 | 2.84×10^{-35} | 2,759,740 | 2,573,737 (93.26) |
| AD | 0.59 | 0.57 | 0.60 | −0.219 | 0.80 | 3.16×10^{-25} | 167 | 161 (96.41) |
| Meta-PRS ⁷ | 0.64 | 0.63 | 0.67 | −0.403 | 0.67 | 6.45×10^{-15} | 6,087,730 | 6,087,730 (100) |

AD, Alzheimer's disease; AF, atrial fibrillation; ARD, age-related disease; AUC, area under the curve; CAD, coronary artery disease; CRC, colorectal cancer; ISS, ischemic stroke; PD, Parkinson's disease; T2D, type 2 diabetes mellitus. ¹ AUC, area under the curve; calculated for the logistic model using longevity (yes/no) as response variable and PRS only as influence variable; AUC_L95, AUC_U95, lower and upper confidence interval boundaries for AUC. ² p -value of the logistic model for ARD-PRS. ³ Beta values (regression coefficients) for ARD-PRS. ⁴ OR, log(beta). ⁵ Input-SNVs from the original publications; for details see Section 4. ⁶ Covered by SNV genotyping/imputation in the German cohort. ⁷ Diagnostic measurements for single ARD-PRSs were calculated for the whole German dataset; the meta-PRS diagnostics were calculated for the test dataset.

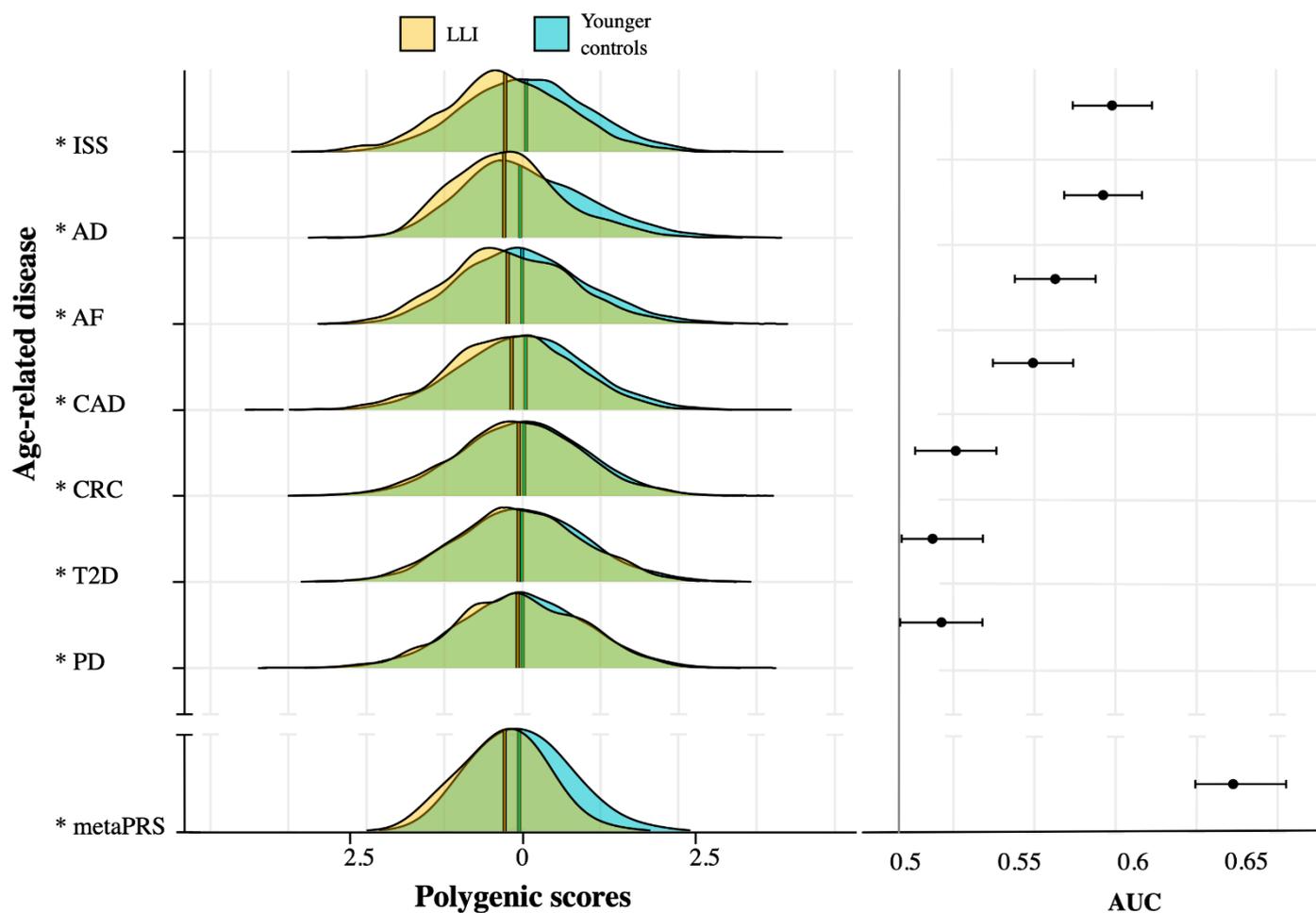


Figure 1. Scaled distribution of the polygenic risk scores of the single age-related diseases (ARD-PRSs) in long-lived individuals (LLI) and younger controls (Ctrl) and areas under the curve (AUCs) with 95% confidence intervals. Depicted values were taken from the model in which only the contributions of the PRSs were taken into account. Asterisks (*) indicate a significant relationship between the PRSs and longevity based on the logistic regression model. ISS, ischemic stroke; AD, Alzheimer’s disease; AF, atrial fibrillation; CAD, coronary artery disease; CRC, colorectal cancer; T2D, type 2 diabetes mellitus; PD, Parkinson’s disease; metaPRS, polygenic score calculated using the single ARD-PRS as influence variable. Within each distribution, the vertical line represents the mean. The black horizontal lines on the right represent the interquartile range and the circles represent the median AUC for each PRS.

2.3. Longevity GPS Discriminates LLI and Controls with an AUC of 0.56

We calculated a GPS for longevity (GPSlong), i.e., “genome-wide”, for the German longevity sample using published summary statistics from a recent large longevity meta-GWAS [15]. GPSlong was based on 3,298,544 out of 8,597,396 published input-SNVs. The remaining 5,298,852 markers were not considered, as they were not present in our imputed and quality-controlled GSA dataset or because they had been removed due to ambiguity. We divided our study population into a training dataset of 3913 individuals (900 LLI; 3013 younger controls) and a test dataset of 1678 individuals (395 LLI; 1283 younger controls) (see Section 4). The GPSs had AUCs ranging from 0.56 to 0.58 (Figure 2a, Supplementary Table S11). The best polygenic score, termed GPSlong, in the discovery and test datasets exhibited an AUC of 0.58 and 0.56, respectively (p -value threshold for SNV selection = 5×10^{-7} , McFadden $R^2 = 0.016$ and = 0.098 in discovery and test datasets, respectively; Figure 2b, Supplementary Table S11). GPSlong used seven SNVs, four located in the *TOMM40/APOE/APOC1* region, two more in the vicinity of the genes *CEP89*

(rs62127361, chromosome 19) and *GPR78* (rs7676745, chromosome 4), and one on chromosome 2 (rs116362179). LLI exhibited a significantly higher GPSlong than the younger controls (logistic model p -value = 1.45×10^{-5} in the test dataset).

We also calculated a second GPS (GPSlong II) after the removal of the *TOMM40/APOE/APOC1* region (126 variants were excluded from the genotyping data). The best performing score, termed GPSlong II, showed a significantly higher score for LLI than for younger controls, displaying an R^2 of 0.09 and an AUC of 0.55 in the test dataset, i.e., it was equivalent to the AUC of GPSlong but was achieved with the inclusion of 10,361 SNVs (p -value threshold for SNV selection = 0.01; GPSlong II p -value = 9.85×10^{-3} (logistic model); Figure 2c,d, Supplementary Tables S12 and S13). Next, we added all the SNVs from the *TOMM40/APOE/APOC1* region and the *APOE* haplotype to GPSlong II and applied stepwise backward regression. This last model (GPSlong II+) included GPS long II, 9 SNVs from the *TOMM40/APOE/APOC1* region (Supplementary Table S14) and the *APOE* haplotype. GPSlong II+ achieved an AUC of 0.61 (± 0.032) in the test dataset.

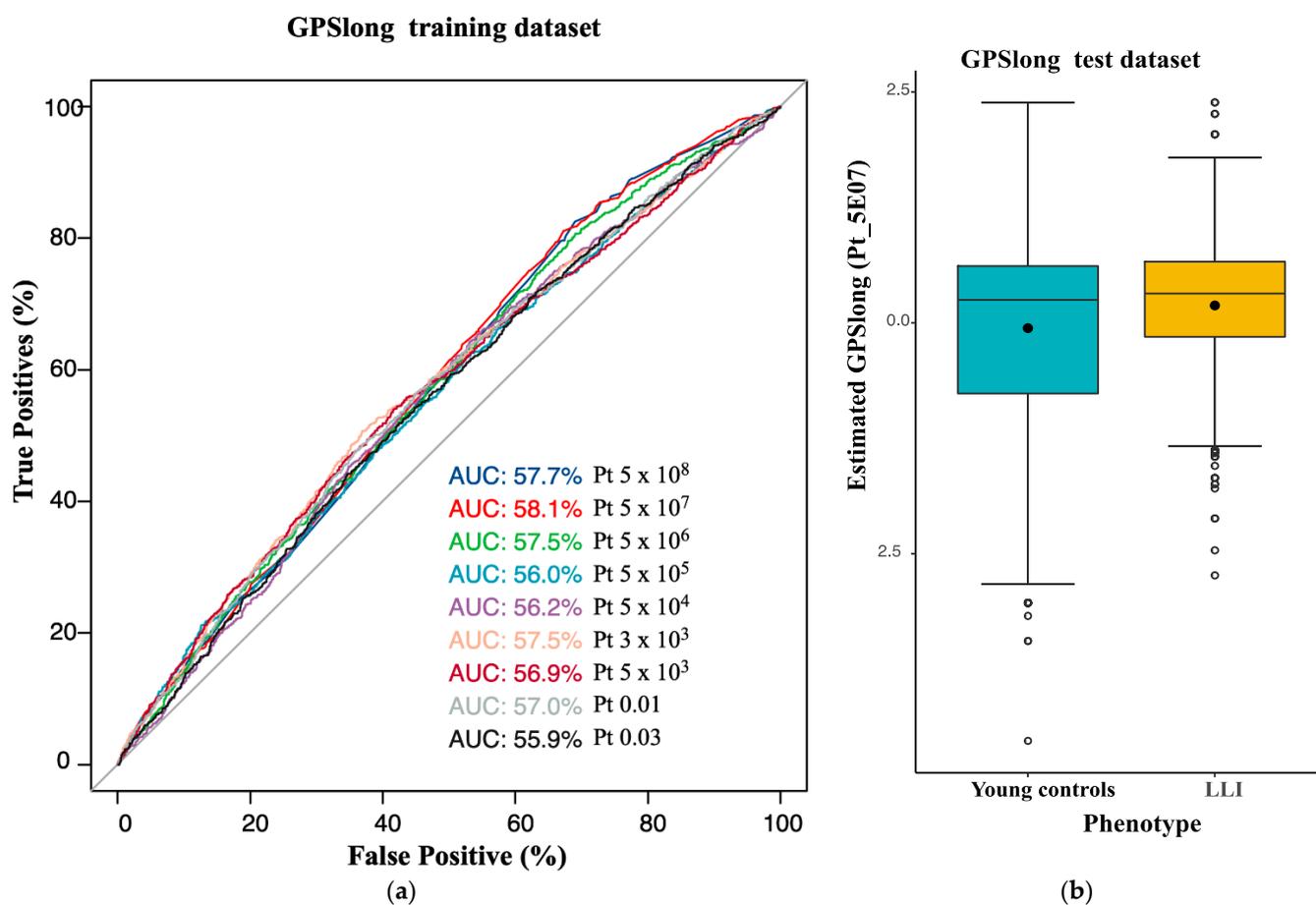


Figure 2. Cont.

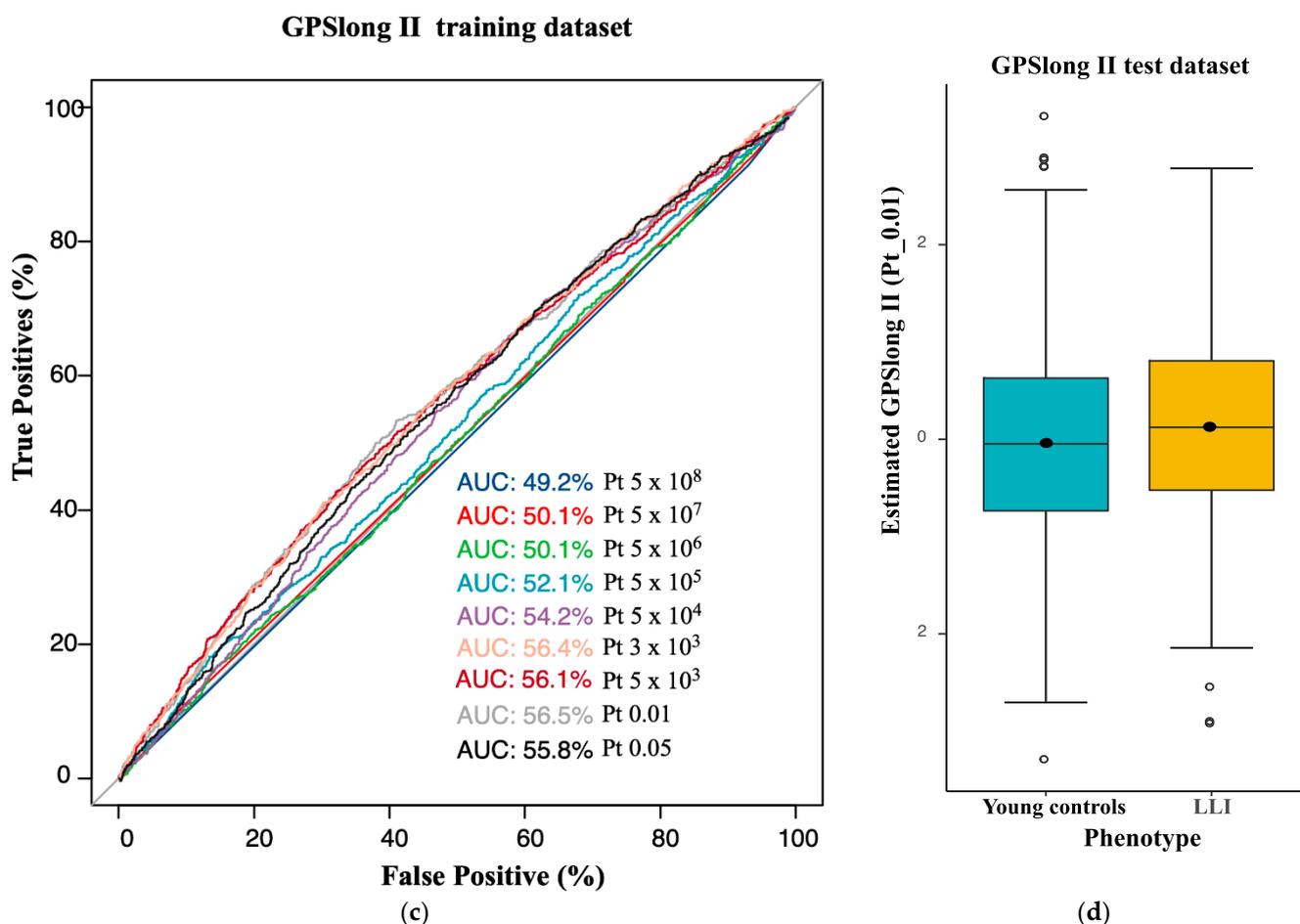


Figure 2. Longevity genome-wide polygenic scores (GPSlong and GPSlong II). (a,c) Discrimination potential of GPSlong (a) and GPSlong II (c) measured by the area under the curve (AUC) with different cut-offs. (b,d) Distribution of GPSlong (b) and GPSlong II (d) among the individuals. The values for LLI and younger controls in the test dataset are represented by boxplots. Pt denotes the different p -value significance cut-offs for the SNVs for GPS calculation.

Furthermore, we investigated the performance of recently published GPSs for longevity [23] and [17] in our cohort (Table 2). Specifically, we tested the best scores reported by Tesi et al. [23], “PRS-5” (best score excluding *APOE* variants) and “PRS-6” (best score including *APOE* variants) and the score published by Liu et al. [17], in the following designated as “PRS_Liu”. With regard to PRS-5 and PRS-6, our genotyping dataset covered 94.8% (91 SNVs) and 97% (324 SNVs) of the SNVs originally used in PRS-5 and PRS-6, respectively. Both genomic scores reached statistical significance in our logistic model and exhibited similar statistics and density distributions as reported in Tesi et al. [23] (Table 2). With respect to PRS_Liu, our dataset covered 86.1% (3414 SNVs) of the original input-SNVs. Although the score discriminated LLI and controls with $p = 2.22 \times 10^{-3}$, the AUC was only 0.53 in our data, and with that, considerably lower than the AUC of 0.77 reported by the authors (Table 2).

Table 2. Statistics of GPSlong and GPSlong II in the German cohort (test dataset) as well as the results of the replication of the previously published scores compared to the reference data [17,23].

| GPS | AUC ¹ | AUC_L95 | AUC_U95 | Beta ² | OR ³ | p-Value ⁴ | Number of SNVs ⁵ |
|---------------------|------------------|---------|---------|-------------------|-----------------|------------------------|-----------------------------|
| GPSlong (Germans) | 0.56 | 0.53 | 0.58 | 0.281 | 1.32 | 1.45×10^{-5} | 7 |
| GPSlongII (Germans) | 0.55 | 0.52 | 0.55 | 0.158 | 1.17 | 9.85×10^{-3} | 10,361 |
| PRS-5 (Germans) | 0.56 | 0.54 | 0.57 | 0.223 | 1.25 | 5.33×10^{-11} | 324 |
| PRS-5 [23] | NA | NA | NA | 0.149 | 1.41 | 3.50×10^{-9} | 334 |
| PRS-6 (Germans) | 0.58 | 0.56 | 0.59 | 0.329 | 1.39 | 1.37×10^{-20} | 91 |
| PRS-6 [23] | NA | 0.57 | 0.61 | 0.158 | 1.44 | 7.30×10^{-10} | 96 |
| PRS_Liu (Germans) | 0.53 | 0.51 | 0.54 | 0.102 | 1.11 | 2.22×10^{-3} | 3414 |
| PRS_Liu [17] | 0.76 | | | | | 1.90×10^{-5} | 3966 |

AUC, area under the curve; GPS, genome-wide polygenic score; OR, odds ratio; PRS, polygenic risk score; SNV, single-nucleotide variant. ¹ Calculated for the logistic model using longevity (yes/no) as response variable and GPS/PRS, sex and five principal components (population substructure) as influence variables; AUC_L95, AUC_U95, lower and upper confidence interval boundaries for AUC. ² Beta values (regression coefficients) for GPS. ³ OR, log(beta). ⁴ p-value of the logistic model for GPS/PRS. ⁵ The number of variants that contributed to each GPS/PRS.

3. Discussion

Based on ARD-PRSs, the LLI in our study had a significantly lower genetic risk of developing ARDs than the individuals from the control sample. This was true for all seven ARDs analyzed, with the largest PRS effects for ISS and AD. We also showed that PRSs are more informative than single genetic variants, because PRSs capture most of the genetic variance due to common variants in a single number. Therefore, our results with respect to ARDs can end the long controversial debate on the contribution of genetic risk factors for these diseases to longevity.

The coverage of the input-SNVs from the initial ARD-PRSs in our data was good for all seven ARDs. The literature supports a link between disease-associated variants and human longevity, within and beyond the *TOMM40/APOE/APOC1* gene region [14,19,26,27]. Our results substantiated the relatively well-documented shared genetic component of longevity and predisposition to cardiovascular disease (CVD; significantly lower risk scores for CAD, AF and ISS in LLI compared to controls) as well as the genetic link with AD [14,15,18–20,28]. Additionally, we strengthened the evidence of a shared genetic architecture between longevity and T2D, as already indicated in, e.g., [15,20]. Moreover, the German LLI exhibited a lower PRS for CRC than the younger controls, supporting that the lower CRC prevalence, incidence and mortality in centenarians [29] has, at least in part, a genetic basis [30]. Remarkably, despite the relatively low incidence and heritability of PD [31,32] and probably even lower prevalence in centenarians [33], we detected minor differences in the joint impact of PD variants between LLI and younger controls.

Efforts in generating a proper GPS for longevity have been hampered mainly by (a) the generally small sample size of longevity-GWAS and the resulting inaccuracy of selected variants and effect sizes, (b) the non-standardized definition of cases and controls, (c) phenotype dilution (when parental longevity is used as phenotype) and (d) the lack of large datasets to validate and test the developed GPS. To date, only a handful of studies have developed GPSs for longevity to discriminate between LLI and controls [4,17,23]. In total, we constructed two new and replicated three published GPSs for longevity. Our first score, GPSlong, was based on the summary statistics from the last meta-GWAS on longevity [15]. The score included only seven input-SNVs, four of which were located in *TOMM40/APOE/APOC1*, and significantly differentiated LLI from younger controls with an AUC of 0.56. We achieved a similarly high AUC (AUC = 0.55) when we excluded the *TOMM40/APOE/APOC1* region (GPSlong II). Strikingly, in a model (GPSlong II+)

including nine non-LD SNVs from the removed region, the *APOE* haplotype and the scores from GPSlong II, the AUC increased to 0.61. This finding confirms the very strong influence of *TOMM40/APOE/APOC1* in explaining longevity and its masking effect on other variants. The exclusion of this region allowed us to observe the cumulative effect of 10,361 SNVs with moderate to low effect size in GPSlong II. Interestingly, GPSlong II reached almost the same discriminative power as GPSlong. These analyses, in addition to showing significant differences in the distributions of LLI and controls, demonstrated a substantial genetic contribution to human longevity within, but also beyond, the well-described mortality-associated *TOMM40/APOE/APOC1* locus.

The third and fourth scores were both validations of scores published by Tesi et al. [23]. The variant coverage by our genotyping data was almost complete, and the scores provided statistical measures similar to those given in the original report. Notably, we validated these longevity scores despite different phenotype definitions (sporadic longevity versus parental longevity). The AUCs achieved in our cohort were 0.56 and 0.58, respectively. Unfortunately, no AUCs were reported by Tesi et al. [23]. The fifth score was a replication of a longevity-GPS recently published by Liu et al. [17] in a Chinese cohort. The authors had reported an extraordinarily high AUC of 0.77; however, although the score reached significance in our data, the AUC was as low as 0.53 despite relatively good input-SNV coverage (86.1%). This discrepancy might be partly due to population differences between European and Chinese individuals, even though Liu et al. [17] had used a European GWAS for PRS construction.

The development of the PRSs/GPSs carries both promises and limitations. They provide a quantitative measure of the predisposition of a phenotype based on a set of genetic variants. In particular, for longevity, a GPS could help researchers to improve the stratification of individuals into groups with significantly different odds. Although the currently existing, purely genetic scores do not individually predict whether a person will be long-lived or not (and this is also not to be expected given the relatively low heritability of human longevity), longevity scores could potentially be combined with, for instance, environmental, lifestyle or epigenetic factors to serve as a genetically informed phenotyping tool that may enhance the discrimination between LLI and controls in the future. Nonetheless, this type of predictive model is sensitive to the cryptic substructure of the population used (i.e., related to geography or participation bias [34]), and to confounding effects [35]. Moreover, the lack of standardized methods on how to integrate ancestry information [36], gene–environment interactions [37], and high-impact variants [38] and on how to find the best thresholds for SNV inclusion [35] and reliable metrics for selection of the best PRS [39] are factors influencing model performance and reproducibility.

4. Materials and Methods

Materials and methods are summarized in Figure 3.

4.1. Study Populations

The German longevity sample comprised 1351 unrelated LLI (mean age: 99 years, age range: 94–110 years), including 643 centenarians (≥ 100 years). The male:female ratio in the sample was 1:2.7. The participants were all of German ancestry and showed no overt signs of cognitive impairment. The recruitment of the German longevity sample was partly organized by the PopGen biobank and has been described in detail elsewhere [40]. Written informed consent to participate in the study was obtained from all participants, and the project was approved by the Ethics Committee of the Medical Faculty of Kiel University. The 4680 unrelated younger controls (age range: 18–83 years, mean age: 50.5 years) were recruited as part of the FoCUS cohort [41] and as blood donors at the University Hospital Schleswig-Holstein in Kiel and Lübeck, Germany.

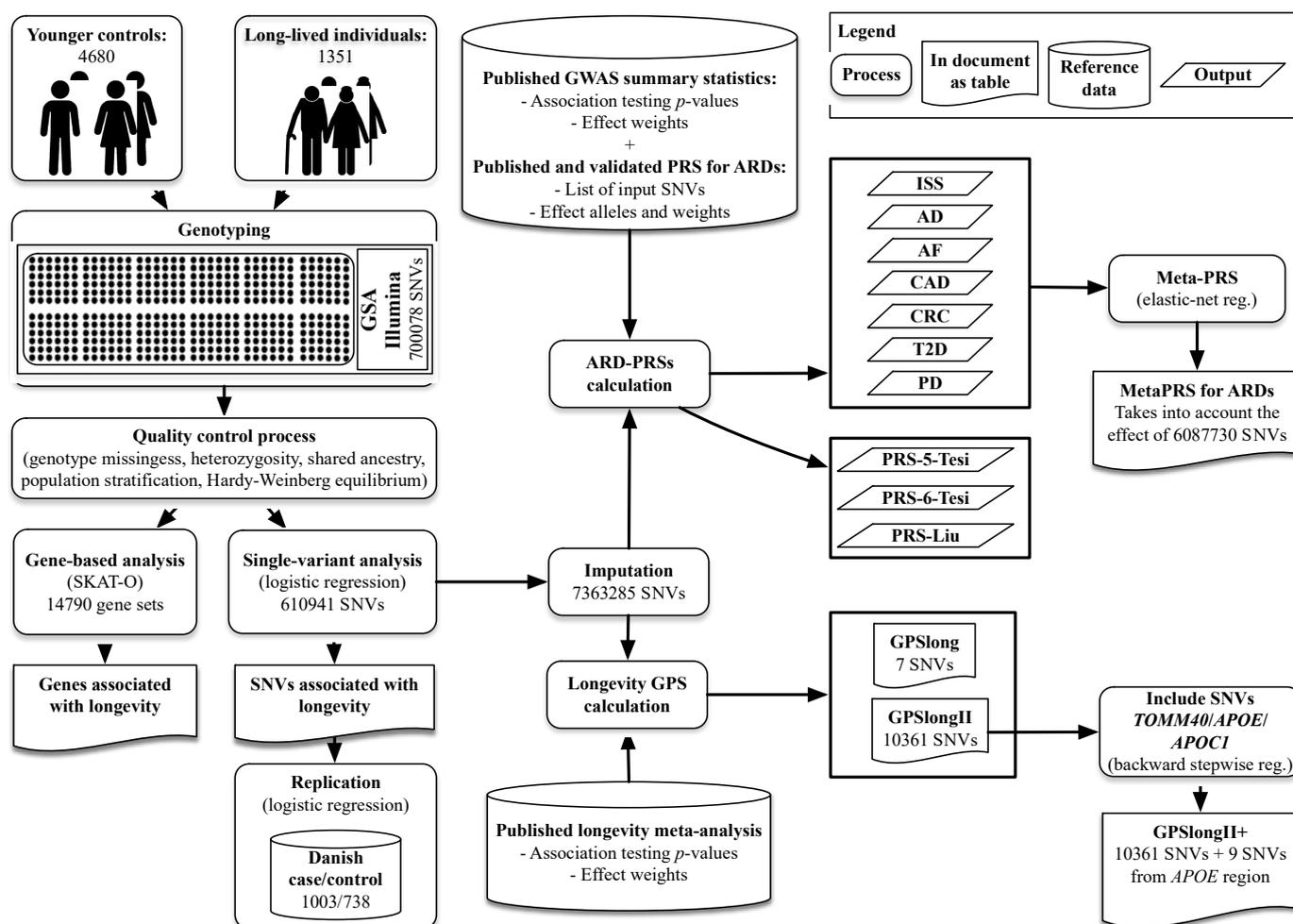


Figure 3. Study design and workflow of the longevity case–control GWAS, calculation of ARD-PRSs, meta-PRS for the diseases and the GPS for longevity. AD, Alzheimer’s disease; AF, atrial fibrillation; ARD, age-related disease; CAD, coronary artery disease; CRC, colorectal cancer; ISS, ischemic stroke; GPS, genome-wide polygenic score; GWAS, genome-wide association study; PD, Parkinson’s disease; PRS, polygenic risk score; SNV, single nucleotide variant; T2D, type 2 diabetes mellitus.

The 1003 Danish cases (mean age: 97.4 years, age range: 90.0–102.5 years, 75.7% women) were included in the present study for the validation of longevity association findings. The sample consisted of participants drawn from seven nation-wide surveys collected at the University of Southern Denmark: the Study of Danish Old Sibs (DOS), the 1905 Birth Cohort Study, the 1910 Birth Cohort Study, the 1911–12 Birth Cohort Study, the 1915 Birth Cohort Study, the Longitudinal Study of Danish Centenarians (LSDC), and the Longitudinal Study of Ageing Danish Twins (LSADT). Briefly, DOS was initiated in 2004 and included families in which at least two siblings were ≥ 90 years of age at intake. The 1905 Birth Cohort Study, 1910 Birth Cohort Study, 1915 Birth Cohort Study, and LSDC are prospective follow-up studies initiated in 1995, 1998, and 2010, when participants were 92–93, 95, and 100 years of age, respectively [42]. The 1911–1912 cohort study consisted of individuals who reached the age of 100 years in the period from May 2011 to July 2012 [43], and LSADT was initiated in 1995 and included Danish twins ≥ 70 years of age [44]. From DOS and LSADT, one individual from each sib-ship or twin pair was randomly selected among participants that had reached an age of at least 91 years for DOS, and 90 years for LSADT. From the 1905 and 1915 Birth Cohort Studies, participants were selected among individuals that had reached a minimum age of 96 years. The 738 controls (mean age: 66.3 years, age range: 55.9–79.9 years, 49.1% women) consisted of individuals recruited by the Danish

Twin Registry (DTR) as part of the study of Middle-Aged Danish Twins (MADT). MADT was initiated in 1998 and included 4314 twins randomly chosen from each of the birth years 1931–1952 [44]. Surviving participants were revisited from 2008 to 2011, where the blood samples used for DNA extraction were collected. To ensure a control sample of unrelated individuals, only one twin from each twin pair was included. Written informed consents were obtained from all participants. Collection and use of biological material, and survey and registry information were approved by the Regional Committees on Health Research Ethics for Southern Denmark. The study was registered in SDU's internal list (notification no. 11.163) and complies with the rules in the General Data Protection Regulation.

4.2. Variant Calling, Quality Control and Imputing for the German and Danish Study Populations

The German samples were genotyped on the Illumina Infinium Global Screening Array-24 (700,078 SNVs) (GSAv1, Illumina[®] Inc., San Diego, CA, USA). Plink 1.9 [45] was used for per-individual and per-marker quality control (QC). In total, 431 individuals failed one or more of the following inclusion criteria: concordant sex information, missing genotype <8%, heterozygosity rate greater or lower than ± 4 standard deviations from the mean, and no individual relatedness. Identity-by-descent (IBD) metric was used to estimate relatedness. In case of relatedness (IBD > 0.1875; halfway between the expected IBD for third- and second-degree relatives), only one individual was included in the analysis. Variants were excluded if the missing rate was higher than 5% and if they deviated from the Hardy–Weinberg equilibrium in control samples (HWE, $p < 1 \times 10^{-5}$). Following this QC procedure, 633,642 variants and 5600 individuals (1295 LLI and 4305 younger controls) remained for the analyses. Population stratification was evaluated with principal component analysis using a common set of independent markers (HapMap3 ancestry set from four ethnic populations). The principal components (PCs) were calculated with Plink 1.9 [45]. For further analyses, the first five PCs were used according to the scree plot. Outliers of the population substructure were identified based on these first five PCs and the local outlier factor metric (LOF > 2.1) [46]. Prior to imputation, a pre-imputation QC was implemented using the HRC-1000G-check-bim script v4.2.7 (<https://www.well.ox.ac.uk/~wrayner/tools/#Checking> accessed on 1 November 2020) to ensure good genotype estimations. Genotype imputation was performed using the secure cloud-based MIS [47] and selecting the Haplotype Reference Consortium HRC r1.1 2016 GRCh37/hg19 as a reference panel [48]. Phasing was performed by applying the Eagle 4.0 engine [48]. In a post-imputation QC, SNVs were excluded if they had an $R^2 < 0.75$, deviated from the HWE ($p < 1 \times 10^{-9}$) in the control sample, had a genotype call rate <95% and/or showed an extremely low minor allele frequency (MAF < 1%). This resulted in 6,010,362 remaining autosomal variants.

The Danish cases were genotyped using the Illumina Human OmniExpress Array (Illumina[®] Inc., San Diego, CA, USA) and imputed to the 1000 Genomes phase1 v3 reference panel using IMPUTE2 [49]. Pre-imputation quality control included filtering of SNPs on genotype call rate <95%, HWE $p < 1 \times 10^{-4}$, and MAF < 1%, and individuals on sample call rate <95%, relatedness and gender mismatch. Controls were genotyped using the Illumina Infinium PsychArray (Illumina[®] Inc., San Diego, CA, USA) and imputed to the 1000 Genomes phase3 reference panel using IMPUTE2 [49]. Pre-imputation quality control included filtering SNPs on genotype call rate <98%, HWE $p < 1 \times 10^{-6}$, and MAF = 0, and individuals on sample call rate <99%, relatedness, and sex mismatch. After imputation, genotype probabilities were converted to hard-called genotypes in Plink using a cut-off of 90%, meaning that only genotypes with a probability of more than 90% were called. Variants with no genotype probabilities above 90% were set to missing.

4.3. Longevity Association Analyses in the German and Danish Study Populations

Single-variant association analysis was performed using the logistic regression test in Plink 1.9 [45] assuming an additive genetic model and adjusting for sex and the first five PCs for the German dataset. To test for independency of the candidate SNVs from the

known longevity-associated locus *TOMM40/APOE/APOC1* [5,50], an additional logistic regression was employed, adjusting for the effects of the SNVs rs769449, rs157582 and rs150966173 in this region. Variants that reached a (borderline) significant p -value ($p \leq 0.05$) after conditioning for the SNVs in the *APOE* gene were selected for replication in the Danish longevity sample. Gene-based association analysis was performed with both burden and non-burden approaches using the SKATO algorithm from the R-package SKAT [51]. False discovery rate [52] was used for multiple testing corrections.

4.4. Application of Published PRSs for Common ARDs in the German Longevity Study Population

PRSs describe the cumulative impact of many common variants on a specific disease. Variants are selected based on the association with the disease, taking into account the LD structure. The weights assigned to each genetic variant are usually derived from the effect sizes of the summary statistics from large GWAS. We investigated differences in the PRS distributions between LLI and younger controls for seven common ARDs, specifically CAD, AF, ISS, T2D, CRC, AD and PD. The diseases were selected because of their association with age [53] and the availability of summary statistics. In more detail, PRS summary statistics for CAD, AF, and T2D were acquired from [22], for ISS from [54], for AD from [55], for PD from [56], and for CRC from [57]. PRSs for ARDs were computed by first multiplying for each selected variant the genotype dose of the risk allele (coded by 0, 1 or 2) by its respective effect size (the log odds ratio (OR) from the GWAS summary statistics). Afterwards, the resulting values of all variants in the score were summed up using Plink 1.9 [45]. The areas under the curve (AUCs) of the ARD-PRSs, as a measure of the performance of the model, were calculated with pROC [58], and the significance of influence variables was assessed by logistic regression using the glmnet R package [59], once only for ARD-PRS as influence variable and once accounting additionally for sex and the first five ancestry PCs as covariables.

4.5. Computation of a Longevity-GPS

First, we defined a discovery dataset, sampling randomly 70% of our cohort (900 LLI and 3013 younger controls). The remaining 30% comprised the test dataset and were used for performance evaluation. For longevity, we computed a genome-wide polygenic score, GPSlong, based on the summary statistics of the latest meta-GWAS on longevity [15]. GPSlong was developed on the discovery dataset using PRSice-2 [60], including a linkage disequilibrium pruning approach by the clumping option (without reference panel, R2 and physical distance thresholds of 0.1 and 250 kb, respectively). In total, 11 different values for GPSlong were calculated using different p -value significance cut-offs for the SNVs, specifically $p = 5 \times 10^{-8}$, $p = 5 \times 10^{-7}$, $p = 5 \times 10^{-6}$, $p = 5 \times 10^{-5}$, $p = 5 \times 10^{-4}$, $p = 3 \times 10^{-3}$, $p = 5 \times 10^{-3}$, $p = 1 \times 10^{-2}$, $p = 3 \times 10^{-2}$, $p = 5 \times 10^{-1}$ and $p = 1$. The score with the best discriminative capacity (i.e., best at separating LLI and controls) was determined based on the maximal area under receiving-operating characteristic curve (AUC). The regression model was calculated using the glmnet R package [59], considering “being long-lived” as outcome and GPSlong as an influence variable with and without additional adjustment for sex and the first five ancestry PCs as predictors. AUC confidence intervals and McFadden’s pseudo- R^2 were calculated with the pROC [58] and rcompanion [61] R packages, respectively. Due to the strong association of the *TOMM40/APOE/APOC1* region with longevity, an additional GPS was calculated after removal of the region (chr. 19: 45351000–45500000) (GPSlong II). To further assess the influence of this region, we employed a backward stepwise regression using GPSlong II scores, 30 SNVs (not in LD with any marker in the deleted region and MAF > 0.01) and *APOE* haplotype ($\epsilon 2$, $\epsilon 3$, and $\epsilon 4$; determined by the alleles at rs429358 and rs7412) as influential variables. The resulting model was denoted GPSlong II+. The performances of the models (i.e., GPSlong, GPSlong II and GPSlong II+) were evaluated in the test dataset using Plink 1.9. The comparison of the mean scores between LLI and younger controls and the replication of the longevity

scores published by Tesi et al. [23] and Liu et al. [17] in our data were performed in the same way as described above for the ARD-PRSs.

4.6. MetaPRS Estimation

We applied an elastic-net logistic regression [62] to model the associations between longevity and the standardized ARD-PRSs into a metaPRS on the discovery dataset. This approach used the single ARD-PRSs as influence variables adjusting for sex and the first five ancestry PCs. The model was computed with the caret R package [63], model penalties were evaluated using 10-fold cross-validation, and class imbalance effects were controlled employing a smoothed bootstrap re-sampling approach [64]. The metaPRS was evaluated using the test dataset similar to GPS long and GPS long II.

5. Conclusions

Our PRSs analyses in the German sample showed an inverse correlation between longevity and the risk for the ARDs tested. The strong influence of the *TOMM40/APOE/APOC1* region on longevity was corroborated, supporting the view that this region is a longevity hub [25]. The PRS approach also helped us identify a considerable number of variants with mostly small effect sizes that influence longevity; however, future studies will be needed to refine PRS methodologies to better integrate genetic, environmental and disease-risk factors. Furthermore, exome-only GPSs could be very informative on the functional level. In general, an optimization of the existing models would require the implementation of additional validation cohorts.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms231810949/s1>.

Author Contributions: Conceptualization, G.G.T., J.D., A.C., G.K. and A.N.; Data curation, G.G.T. and T.P.H.; Formal analysis, G.G.T. and T.P.H.; Investigation, G.G.T., J.D. and T.P.H.; Methodology, G.G.T.; Resources, J.D., M.N., B.K.-K., J.M.-F., K.C., K.A.-R., W.L., M.L., S.G., S.S., A.F. and A.N.; Validation, M.N. and A.C.; Supervision, A.C., G.K. and A.N.; Writing—original draft, G.G.T. and J.D.; Writing—review and editing, G.G.T., J.D., M.N., B.K.-K., D.K., J.M.-F., K.C., A.F., A.C., G.K. and A.N. All authors have read and agreed to the published version of the manuscript.

Funding: GGT was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through project number 390870439 (EXC 2150-ROOTS). AC was supported by the DFG through project number 287074911 (FOR2488). DK was supported by the DFG through project number 400993799 (GRK 2501-Translational Evolutionary Research). The Popgen Biobank and the Popgen 2.0 Network were supported by the Bundesministerium für Bildung und Forschung (BMBF, Federal Ministry of Education and Research; grant no. 01EY1103). The Danish replication sample was supported by The National Program for Research Infrastructure 2007 (grant no. 09-063256) from the Danish Agency for Science and Innovation, the Velux Foundation, and the US National Institute of Health (P01 AG08761). Genotyping of the Danish controls was conducted by the SNP&SEQ Technology Platform, Science for Life Laboratory, Uppsala, Sweden (<http://snpseq.medsci.uu.se/genotyping/snp-services/>) and supported by NIH R01 AG037985 (Pedersen).

Institutional Review Board Statement: The study was conducted in accordance with the approval from the Ethics Committee of the Medical Faculty of Kiel University and the Regional Committees on Health Research Ethics for Southern Denmark, registered in SDU's internal list (notification no. 11.163) and complies with the rules in the General Data Protection Regulation.

Informed Consent Statement: All participants gave written informed consent to participate in the study.

Data Availability Statement: All German samples and information on their corresponding phenotypes were obtained from the PopGen Biobank (Schleswig-Holstein, Germany) and can be accessed through a Material Data Access Form. Information about the Material Data Access Form and how to apply can be found at <https://www.uksh.de/p2n/Information+for+Researchers.html>.

Acknowledgments: We thank Mike A. Nalls, Center for Alzheimer's and Related Dementias, a collaborative initiative of the US National Institute on Aging and the US National Institute of

Neurological Disorders and Stroke, for providing us with the list of the 1805 SNVs (together with reference alleles and effect sizes β) included in their published PRS.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Hjelmborg, J.V.; Iachine, I.; Skytthe, A.; Vaupel, J.W.; McGue, M.; Koskenvuo, M.; Kaprio, J.; Pedersen, N.L.; Christensen, K. Genetic influence on human lifespan and longevity. *Qual. Life Res.* **2006**, *119*, 312–321. [\[CrossRef\]](#)
- Flachsbart, F.; Caliebe, A.; Kleindorp, R.; Blanché, H.; von Eller-Eberstein, H.; Nikolaus, S.; Schreiber, S.; Nebel, A. Association of *FOXO3A* variation with human longevity confirmed in German centenarians. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 2700–2705. [\[CrossRef\]](#) [\[PubMed\]](#)
- Nebel, A.; Kleindorp, R.; Caliebe, A.; Nothnagel, M.; Blanché, H.; Junge, O.; Wittig, M.; Ellinghaus, D.; Flachsbart, F.; Wichmann, H.-E.; et al. A genome-wide association study confirms *APOE* as the major gene influencing survival in long-lived individuals. *Mech. Ageing Dev.* **2011**, *132*, 324–330. [\[CrossRef\]](#) [\[PubMed\]](#)
- Sebastiani, P.; Solovieff, N.; DeWan, A.T.; Walsh, K.M.; Puca, A.; Hartley, S.W.; Melista, E.; Andersen, S.; Dworkis, D.A.; Wilk, J.B.; et al. Genetic Signatures of Exceptional Longevity in Humans. *PLoS ONE* **2012**, *7*, e29848. [\[CrossRef\]](#) [\[PubMed\]](#)
- Deelen, J.; Beekman, M.; Uh, H.-W.; Broer, L.; Ayers, K.L.; Tan, Q.; Kamatani, Y.; Bennet, A.M.; Tamm, R.; Trompet, S.; et al. Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. *Hum. Mol. Genet.* **2014**, *23*, 4420–4432. [\[CrossRef\]](#)
- Torres, G.G.; Nygaard, M.; Caliebe, A.; Blanché, H.; Chantalat, S.; Galan, P.; Lieb, W.; Christiansen, L.; Deleuze, J.-F.; Christensen, K.; et al. Exome-Wide Association Study Identifies *FN3KRP* and *PGP* as New Candidate Longevity Genes. *J. Gerontol. Ser. A* **2021**, *76*, 786–795. [\[CrossRef\]](#)
- Andersen-Ranberg, K.; Schroll, M.; Jeune, B. Healthy centenarians do not exist, but autonomous centenarians do: A population-based study of morbidity among Danish centenarians. *J. Am. Geriatr. Soc.* **2001**, *49*, 900–908. [\[CrossRef\]](#)
- Evert, J.; Lawler, E.; Bogan, H.; Perls, T. Morbidity Profiles of Centenarians: Survivors, Delayers, and Escapers. *J. Gerontol. Ser. A* **2003**, *58*, M232–M237. [\[CrossRef\]](#)
- Ali, O. Genetics of type 2 diabetes. *World J. Diabetes* **2013**, *4*, 114. [\[CrossRef\]](#)
- Barber, R.C. The Genetics of Alzheimer's Disease. *Scientifica* **2012**, *2012*, 1–14. [\[CrossRef\]](#)
- McPherson, R.; Tybjaerg-Hansen, A. Genetics of Coronary Artery Disease. *Circ. Res.* **2016**, *118*, 564–578. [\[CrossRef\]](#) [\[PubMed\]](#)
- Soerensen, M.; Nygaard, M.; Debrabant, B.; Mengel-From, J.; Dato, S.; Thinggaard, M.; Christensen, K.; Christiansen, L. No Association between Variation in Longevity Candidate Genes and Aging-related Phenotypes in Oldest-old Danes. *Exp. Gerontol.* **2016**, *78*, 57–61. [\[CrossRef\]](#) [\[PubMed\]](#)
- Stevenson, M.; Bae, H.; Schupf, N.; Andersen, S.; Zhang, Q.; Perls, T.; Sebastiani, P. Burden of disease variants in participants of the long life family Study. *Ageing* **2015**, *7*, 123–132. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fortney, K.; Dobriban, E.; Garagnani, P.; Pirazzini, C.; Monti, D.; Mari, D.; Atzmon, G.; Barzilai, N.; Franceschi, C.; Owen, A.B.; et al. Genome-Wide Scan Informed by Age-Related Disease Identifies Loci for Exceptional Human Longevity. *PLoS Genet.* **2015**, *11*, e1005728. [\[CrossRef\]](#) [\[PubMed\]](#)
- Deelen, J.; Evans, D.S.; Arking, D.E.; Tesi, N.; Nygaard, M.; Liu, X.; Wojczynski, M.K.; Biggs, M.L.; van der Spek, A.; Atzmon, G.; et al. A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nat. Commun.* **2019**, *10*, 3669. [\[CrossRef\]](#)
- Melzer, D.; Pilling, L.C.; Ferrucci, L. The genetics of human ageing. *Nat. Rev. Genet.* **2019**, *21*, 88–101. [\[CrossRef\]](#)
- Liu, X.; Song, Z.; Li, Y.; Yao, Y.; Fang, M.; Bai, C.; An, P.; Chen, H.; Chen, Z.; Tang, B.; et al. Integrated genetic analyses revealed novel human longevity loci and reduced risks of multiple diseases in a cohort study of 15,651 Chinese individuals. *Ageing Cell* **2021**, *20*, e13323. [\[CrossRef\]](#)
- Tesi, N.; Hulsman, M.; van der Lee, S.J.; Jansen, I.E.; Stringa, N.; van Schoor, N.M.; Scheltens, P.; van der Flier, W.M.; Huisman, M.; Reinders, M.J.T.; et al. The Effect of Alzheimer's Disease-Associated Genetic Variants on Longevity. *Front. Genet.* **2021**, *12*, 748781. [\[CrossRef\]](#)
- Pilling, L.C.; Kuo, C.-L.; Sicinski, K.; Tamosauskaite, J.; Kuchel, G.A.; Harries, L.W.; Herd, P.; Wallace, R.; Ferrucci, L.; Melzer, D. Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. *Ageing* **2017**, *9*, 2504–2520. [\[CrossRef\]](#)
- Lin, J.-R.; Sin-Chan, P.; Napolioni, V.; Torres, G.G.; Mitra, J.; Zhang, Q.; Jabalameli, M.R.; Wang, Z.; Nguyen, N.; Gao, T.; et al. Rare genetic coding variants associated with human longevity and protection against age-related diseases. *Nat. Ageing* **2021**, *1*, 783–794. [\[CrossRef\]](#)
- Mooijaart, S.P.; Van Heemst, D.; Noordam, R.; Rozing, M.P.; Wijman, C.A.; De Craen, A.J.; Westendorp, R.G.; Beekman, M.; Slagboom, E.P. Polymorphisms associated with type 2 diabetes in familial longevity: The Leiden Longevity Study. *Ageing* **2011**, *3*, 55–62. [\[CrossRef\]](#)
- Khera, A.V.; Chaffin, M.; Aragam, K.G.; Haas, M.E.; Roselli, C.; Choi, S.H.; Natarajan, P.; Lander, E.S.; Lubitz, S.A.; Ellinor, P.T.; et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **2018**, *50*, 1219–1224. [\[CrossRef\]](#) [\[PubMed\]](#)

23. Tesi, N.; van der Lee, S.J.; Hulsman, M.; Jansen, I.E.; Stringa, N.; van Schoor, N.M.; Scheltens, P.; van der Flier, W.M.; Huisman, M.; Reinders, M.J.T.; et al. Polygenic Risk Score of Longevity Predicts Longer Survival Across an Age Continuum. *J. Gerontol. Ser. A* **2021**, *76*, 750–759. [[CrossRef](#)] [[PubMed](#)]
24. Schächter, F.; Faure-Delaneuf, L.; Guénot, F.; Rouger, H.; Froguel, P.; Lesueur-Ginot, L.; Cohen, D. Genetic associations with human longevity at the APOE and ACE loci. *Nat. Genet.* **1994**, *6*, 29–32. [[CrossRef](#)] [[PubMed](#)]
25. Szymczak, S.; Dose, J.; Torres, G.G.; Heinsen, F.-A.; Venkatesh, G.; Datlinger, P.; Nygaard, M.; Mengel-From, J.; Flachsbar, F.; Klapper, W.; et al. DNA methylation QTL analysis identifies new regulators of human longevity. *Hum. Mol. Genet.* **2020**, *29*, 1154–1167. [[CrossRef](#)] [[PubMed](#)]
26. McDaid, A.F.; Joshi, P.K.; Porcu, E.; Komljenovic, A.; Li, H.; Sorrentino, V.; Litovchenko, M.; Bevers, R.P.J.; Rüeger, S.; Reymond, A.; et al. Bayesian association scan reveals loci associated with human lifespan and linked biomarkers. *Nat. Commun.* **2017**, *8*, 15842. [[CrossRef](#)]
27. Timmers, P.R.; Mounier, N.; Lall, K.; Fischer, K.; Ning, Z.; Feng, X.; Bretherick, A.D.; Clark, D.W.; Agbessi, M.; Ahsan, H.; et al. Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. *eLife* **2019**, *8*, e39856. [[CrossRef](#)]
28. Maruszak, A.; Pepłońska, B.; Safranow, K.; Chodakowska-Żebrowska, M.; Barcikowska, M.; Żekanowski, C. TOMM40 rs10524523 Polymorphism's Role in Late-Onset Alzheimer's Disease and in Longevity. *J. Alzheimer's Dis.* **2012**, *28*, 309–322. [[CrossRef](#)]
29. Pavlidis, N.; Stanta, G.; Audisio, R.A. Cancer prevalence and mortality in centenarians: A systematic review. *Crit. Rev. Oncol.* **2012**, *83*, 145–152. [[CrossRef](#)]
30. Nolen, S.C.; Evans, M.A.; Fischer, A.; Corrada, M.M.; Kawas, C.H.; Bota, D.A. Cancer—Incidence, prevalence and mortality in the oldest-old. A comprehensive review. *Mech. Ageing Dev.* **2017**, *164*, 113–126. [[CrossRef](#)]
31. von Campenhausen, S.; Bornschein, B.; Wick, R.; Bötzel, K.; Sampaio, C.; Poewe, W.; Oertel, W.; Siebert, U.; Berger, K.; Dodel, R. Prevalence and incidence of Parkinson's disease in Europe. *Eur. Neuropsychopharmacol.* **2005**, *15*, 473–490. [[CrossRef](#)] [[PubMed](#)]
32. Lin, M.T.; Simon, D.K. No evidence for heritability of Parkinson disease in Swedish twins. *Neurology* **2005**, *64*, 932. [[CrossRef](#)] [[PubMed](#)]
33. Marcon, G.; Manganotti, P.; Tettamanti, M. Is Parkinson's Disease a Very Rare Pathology in Centenarians? A Clinical Study in a Cohort of Subjects. *J. Alzheimer's Dis.* **2020**, *73*, 73–76. [[CrossRef](#)]
34. Kerminen, S.; Martin, A.R.; Koskela, J.; Ruotsalainen, S.E.; Havulinna, A.S.; Surakka, I.; Palotie, A.; Perola, M.; Salomaa, V.; Daly, M.J.; et al. Geographic Variation and Bias in the Polygenic Scores of Complex Diseases and Traits in Finland. *Am. J. Hum. Genet.* **2019**, *104*, 1169–1181. [[CrossRef](#)]
35. Janssens, A.C.J.W. Validity of polygenic risk scores: Are we measuring what we think we are? *Hum. Mol. Genet.* **2019**, *28*, R143–R150. [[CrossRef](#)]
36. Marnetto, D.; Pärna, K.; Läll, K.; Molinaro, L.; Montinaro, F.; Haller, T.; Metspalu, M.; Mägi, R.; Fischer, K.; Pagani, L. Ancestry deconvolution and partial polygenic score can improve susceptibility predictions in recently admixed individuals. *Nat. Commun.* **2020**, *11*, 1628. [[CrossRef](#)]
37. Rudolph, A.; Song, M.; Brook, M.; Milne, R.L.; Mavaddat, N.; Michailidou, K.; Bolla, M.K.; Wang, Q.; Dennis, J.; Wilcox, A.; et al. Joint associations of a polygenic risk score and environmental risk factors for breast cancer in the Breast Cancer Association Consortium. *Int. J. Epidemiol.* **2018**, *47*, 526–536. [[CrossRef](#)] [[PubMed](#)]
38. Barnes, D.R.; Rookus, M.A.; McGuffog, L.; Leslie, G.; Mooij, T.M.; Dennis, J.; Mavaddat, N.; Adlard, J.; Ahmed, M.; Aittomäki, K.; et al. Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of BRCA1 and BRCA2 pathogenic variants. *Genet. Med.* **2020**, *22*, 1653–1666. [[CrossRef](#)] [[PubMed](#)]
39. Polygenic Risk Score Task Force of the International Common Disease Alliance; Adeyemo, A.; Balaconis, M.K.; Darnes, D.R.; Fatumo, S.; Moreno, P.G.; Hodonsky, C.J.; Inouye, M.; Kanai, M.; Kato, K.; et al. Responsible use of polygenic risk scores in the clinic: Potential benefits, risks and gaps. *Nat. Med.* **2021**, *27*, 1876–1884. [[CrossRef](#)]
40. Nebel, A.; Croucher, P.J.P.; Stiegeler, R.; Nikolaus, S.; Krawczak, M.; Schreiber, S. No association between microsomal triglyceride transfer protein (MTP) haplotype and longevity in humans. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 7906–7909. [[CrossRef](#)]
41. Müller, N.; Schulte, D.M.; Türk, K.; Freitag-Wolf, S.; Hampe, J.; Zeuner, R.; Schröder, J.O.; Gouni-Berthold, I.; Berthold, H.K.; Krone, W.; et al. IL-6 blockade by monoclonal antibodies inhibits apolipoprotein (a) expression and lipoprotein (a) synthesis in humans. *J. Lipid Res.* **2015**, *56*, 1034–1042. [[CrossRef](#)] [[PubMed](#)]
42. Rasmussen, S.H.; Andersen-Ranberg, K.; Thinggaard, M.; Jeune, B.; Skytthe, A.; Christiansen, L.; Vaupel, J.W.; McGue, M.; Christensen, K. Cohort Profile: The 1895, 1905, 1910 and 1915 Danish Birth Cohort Studies—secular trends in the health and functioning of the very old. *Int. J. Epidemiol.* **2017**, *46*, 1746–1746j. [[CrossRef](#)] [[PubMed](#)]
43. Robine, J.-M.; Cheung, S.L.K.; Saito, Y.; Jeune, B.; Parker, M.G.; Herrmann, F.R. Centenarians Today: New Insights on Selection from the 5-COOP Study. *Curr. Gerontol. Geriatr. Res.* **2010**, *2010*, 120354. [[CrossRef](#)] [[PubMed](#)]
44. Pedersen, D.A.; Larsen, L.A.; Nygaard, M.; Mengel-From, J.; McGue, M.; Dalgård, C.; Hvidberg, L.; Hjelmberg, J.; Skytthe, A.; Holm, N.V.; et al. The Danish Twin Registry: An Updated Overview. *Twin Res. Hum. Genet.* **2019**, *22*, 499–507. [[CrossRef](#)]
45. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [[CrossRef](#)] [[PubMed](#)]

46. Breunig, M.M.; Kriegel, H.-P.; Ng, R.T.; Sander, J. LOF: Identifying density-based local outliers. In Proceedings of the 2000 ACM SIGMOD international conference on Management of data, Dallas, TX, USA, 16–18 May 2000; pp. 93–104.
47. Das, S.; Forer, L.; Schönherr, S.; Sidore, C.; Locke, A.E.; Kwong, A.; Vrieze, S.I.; Chew, E.Y.; Levy, S.; McGue, M.; et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **2016**, *48*, 1284–1287. [[CrossRef](#)]
48. Loh, P.-R.; Danecek, P.; Palamara, P.F.; Fuchsberger, C.; Reshef, Y.A.; Finucane, H.K.; Schoenherr, S.; Forer, L.; McCarthy, S.; Abecasis, C.F.G.R.; et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **2016**, *48*, 1443–1448. [[CrossRef](#)]
49. Howie, B.N.; Donnelly, P.; Marchini, J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genet.* **2009**, *5*, e1000529. [[CrossRef](#)]
50. Newman, A.B.; Walter, S.; Lunetta, K.; Garcia, M.E.; Slagboom, P.; Christensen, K.; Arnold, A.M.; Aspelund, T.; Aulchenko, Y.; Benjamin, E.; et al. A Meta-analysis of Four Genome-Wide Association Studies of Survival to Age 90 Years or Older: The Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium. *J. Gerontol. Ser. A* **2010**, *65*, 478–487. [[CrossRef](#)]
51. Ionita-Laza, I.; Lee, S.; Makarov, V.; Buxbaum, J.D.; Lin, X. Sequence Kernel Association Tests for the Combined Effect of Rare and Common Variants. *Am. J. Hum. Genet.* **2013**, *92*, 841–853. [[CrossRef](#)]
52. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **1995**, *57*, 289–300. [[CrossRef](#)]
53. Irizar, P.A.; Schäuble, S.; Esser, D.; Groth, M.; Frahm, C.; Priebe, S.; Baumgart, M.; Hartmann, N.; Marthandan, S.; Menzel, U.; et al. Transcriptomic alterations during ageing reflect the shift from cancer to degenerative diseases in the elderly. *Nat. Commun.* **2018**, *9*, 327. [[CrossRef](#)] [[PubMed](#)]
54. Abraham, G.; Malik, R.; Yonova-Doing, E.; Salim, A.; Wang, T.; Danesh, J.; Butterworth, A.S.; Howson, J.M.M.; Inouye, M.; Dichgans, M. Genomic risk score offers predictive performance comparable to clinical risk factors for ischaemic stroke. *Nat. Commun.* **2019**, *10*, 5819. [[CrossRef](#)] [[PubMed](#)]
55. Chaudhury, S.; Brookes, K.J.; Patel, T.; Fallows, A.; Guetta-Baranes, T.; Turton, J.C.; Guerreiro, R.; Bras, J.; Hardy, J.; Francis, P.T.; et al. Alzheimer’s disease polygenic risk score as a predictor of conversion from mild-cognitive impairment. *Transl. Psychiatry* **2019**, *9*, 154. [[CrossRef](#)]
56. Nalls, M.A.; Blauwendraat, C.; Vallerga, C.L.; Heilbron, K.; Bandres-Ciga, S.; Chang, D.; Tan, M.; Kia, D.A.; Noyce, A.J.; Xue, A.; et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson’s disease: A meta-analysis of genome-wide association studies. *Lancet Neurol.* **2019**, *18*, 1091–1102. [[CrossRef](#)]
57. Jia, G.; Lu, Y.; Wen, W.; Long, J.; Liu, Y.; Tao, R.; Li, B.; Denny, J.C.; Shu, X.-O.; Zheng, W. Evaluating the Utility of Polygenic Risk Scores in Identifying High-Risk Individuals for Eight Common Cancers. *JNCI Cancer Spectr.* **2020**, *4*, pkaa021. [[CrossRef](#)]
58. Robin, X.; Turck, N.; Hainard, A.; Tiberti, N.; Lisacek, F.; Sanchez, J.-C.; Müller, M. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **2011**, *12*, 77. [[CrossRef](#)]
59. Friedman, J.H.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1–22. [[CrossRef](#)]
60. Choi, S.W.; O’Reilly, P.F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience* **2019**, *8*, giz082. [[CrossRef](#)]
61. Mangiafico, S. rcompanion: Functions to support extension education program evaluation. *Cran Repos* **2017**, *20*, 1–71.
62. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Stat. Methodol. Ser. B* **2005**, *67*, 301–320. [[CrossRef](#)]
63. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **2008**, *28*, 1–26. [[CrossRef](#)]
64. Menardi, G.; Torelli, N. Training and assessing classification rules with imbalanced data. *Data Min. Knowl. Discov.* **2014**, *28*, 92–122. [[CrossRef](#)]