

## Rare Genetic Variants Correlate with Better Processing Speed

Zeyuan Song<sup>1</sup>, Anastasia Gurinovich<sup>2</sup>, Marianne Nygaard<sup>3,8</sup>, Jonas Mengel-From<sup>3,8</sup>, Stacy Andersen<sup>4</sup>, Stephanie Cosentino<sup>5</sup>, Nicole Schupfs<sup>5</sup>, Joseph Lee<sup>5</sup>, Joseph Zmuda<sup>6</sup>, Svetlana Ukraintseva<sup>7</sup>, Konstantin Arbeev<sup>7</sup>, Kaare Christensen<sup>3,8,9</sup>, Thomas Perls<sup>4</sup>, Paola Sebastiani<sup>2</sup>

1: Department of Biostatistics, Boston University School of Public Health, Boston, MA ([zeyuan@bu.edu](mailto:zeyuan@bu.edu), [agurinov@bu.edu](mailto:agurinov@bu.edu))

2: Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA ([agrinovich@tuftsmedicalcenter.org](mailto:agrinovich@tuftsmedicalcenter.org), [psebastiani@tuftsmedicalcenter.org](mailto:psebastiani@tuftsmedicalcenter.org))

3: Epidemiology, Biostatistics and Biodemography, The Danish Aging Research Center, and The Danish Twin Registry, Institute of Public Health, University of Southern Denmark, Odense, Denmark ([KChristensen@health.sdu.dk](mailto:KChristensen@health.sdu.dk), [jmengel-from@health.sdu.dk](mailto:jmengel-from@health.sdu.dk), [mnygaard@health.sdu.dk](mailto:mnygaard@health.sdu.dk))

4: Department of Medicine, Boston University School of Medicine, Boston, MA ([stacy@bu.edu](mailto:stacy@bu.edu), [thperls@bu.edu](mailto:thperls@bu.edu))

5: Departments of Epidemiology and Neurology, Columbia University Medical Center, New York, NY ([sc2460@cumc.columbia.edu](mailto:sc2460@cumc.columbia.edu), [ns24@cumc.columbia.edu](mailto:ns24@cumc.columbia.edu), [JHL2@columbia.edu](mailto:JHL2@columbia.edu))

6: Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA ([zmudaj@edc.pitt.edu](mailto:zmudaj@edc.pitt.edu))

7: Biodemography of Aging Research Unit, Social Science Research Institute, Duke University, Durham, NC ([Svetlana.Ukraintseva@duke.edu](mailto:Svetlana.Ukraintseva@duke.edu), [ka29@duke.edu](mailto:ka29@duke.edu))

8: Department of Clinical Genetics, Odense University Hospital, Odense, Denmark ([KChristensen@health.sdu.dk](mailto:KChristensen@health.sdu.dk), [jmengel-from@health.sdu.dk](mailto:jmengel-from@health.sdu.dk), [mnygaard@health.sdu.dk](mailto:mnygaard@health.sdu.dk))

9: Department of Clinical Biochemistry and Pharmacology, Odense University Hospital, Odense, Denmark ([KChristensen@health.sdu.dk](mailto:KChristensen@health.sdu.dk))

### Abstract

We conducted a genome-wide association study (GWAS) of Digit Symbol Substitution Test (DSST) scores administered in 4207 family members of the Long Life Family Study (LLFS). Genotype data were imputed to the HRC panel of 64,940 haplotypes resulting in ~15M genetic variants with quality score > 0.7. The results were replicated using genetic data imputed to the 1000 Genomes phase 3 reference panel from two Danish twin cohorts: the study of Middle Aged Danish Twins and the Longitudinal Study of Aging Danish Twins. The GWAS in LLFS discovered 20 rare genetic variants (minor allele frequency (MAF) < 1.0%) that reached genome-wide significance ( $p$ -value <  $5 \times 10^{-8}$ ). Among these, 18 variants had large protective effects on the processing speed, including rs7623455, rs9821776, rs9821587, rs78704059 on chromosome 3, which were replicated in the combined Danish twin cohort. These SNPs are located in/near two genes, *THRB* and *RARB*, that belonged to thyroid hormone receptors family that may influence speed of metabolism and cognitive aging. The gene-level tests in LLFS confirmed that these two genes are associated with processing speed.

**Key Words:** processing speed, DSST, genome-wide association study, rare variants

### Introduction

Human cognitive functions change with age, both normally and pathologically. Some cognitive abilities, such as vocabulary, are resilient to brain aging, while other abilities such as conceptual reasoning, memory, and processing speed, decline with age [Harada 2013]. Age-related cognitive impairment is an increasing problem for society. As the global population aged over 65

years continue to increase, and the number of older persons is projected to double to 1.5 billion in 2050 [United Nations 2019], it is expected that the clinical and economic burden of cognitive impairment on society will increase. Understanding cognitive aging and its possible therapeutic targets should thus be an important research focus to help delay or prevent age-related cognitive impairment.

Cognitive functions are usually assessed by administering a battery of neuropsychological tests. For example, verbal episodic memory can be assessed by logical memory tests for immediate and delayed recall. Working memory can be assessed by administering the digit span forward and backward tests, and processing speed and executive control can be assessed by administering the Digit Symbol Substitution Test (DSST). Several of these tests are under strong genetic control with heritability ranging from 16% to 68% [Cirulli 2010], and over the past years several genome-wide association studies (GWAS) have identified common genetic variants associated with cognitive functions [Trampush 2017; Davies 2015]. A recent GWAS of 300,486 individuals identified 148 independent genetic loci that associate with general cognitive function defined as the first principal component of multiple cognitive test scores [Davies 2019].

The heritability of the DSST score is strong, ranging from 0.36-0.68 [Cirulli 2010] but, despite such a high heritability, previous genome-wide association studies failed to identify and replicate common genetic variants associated with DSST score [Luciano 2011; Ibrahim-Verbaas 2015]. Previous studies focused attention on the association between common genetic variants and DSST scores in the general population. Here, we leverage the unique family-based design and the enrichment for healthy agers of the Long Life Family Study (LLFS) to identify rare variants potentially associated with preservation of DSST scores with aging. The LLFS includes very old subjects (90+ years old) who maintained good DSST scores compared to individuals without familial longevity. The study administered the DSST at enrollment and approximately 8 years after and, by studying the longitudinal changes of DSST, we showed that there are individuals whose DSST scores decline slower compared to others [Sebastiani 2020]. In addition, we showed that the heritability of DSST at baseline and its change over time was over 40% [Wojczynski 2019]. These results suggest that those extremely long lived individuals with normal cognitive function may carry protective genetic variants that make them resilient to cognitive decline as they age. We therefore conducted a genome-wide association study of DSST scores measured in subjects enrolled in the LLFS. The genetic findings were further replicated in middle-aged and elderly twins from the Danish Twin Registry.

## **Method**

### **Study populations.**

Long Life Family Study (LLFS). The LLFS is a multicenter, multinational longitudinal two generation family study of longevity and healthy aging [Wojczynski 2019]. A total of 4953 subjects from 539 families were enrolled between 2006 and 2009, and carefully phenotyped in person. Participants were enrolled at three American field centers (Boston, Pittsburgh and New York), and a European field center in Denmark. Potential probands were recruited based on older age, capacity to understand the study and their Family Longevity Selection Score (FLoSS). The FLoSS quantifies the degree of familial longevity using sex and birth-year cohort survival probabilities of the proband and their siblings [Sebastiani 2009]. Eligibility of sibships for the study was based on a FLoSS score >7 and having at least one living sibling and at least one

offspring willing to be enrolled in the study. Socio-demographic, medical history data, current medical conditions and medications, physical and cognitive function data, and blood samples were collected via in-person visits and phone questionnaires for all subjects at the time of enrollment as described elsewhere [Newman 2011; Elo 2013]. Participants have been followed-up annually to track vital and health status. All subjects provided informed consent approved by the field centers IRB.

The Danish Twin Registry. The Danish Twin Registry sample included 1432 individuals who completed the Symbol Digit Substitution Task (SDST) and were collected as part of the study of Middle Aged Danish Twins (MADT, N=1146) and the Longitudinal Study of Aging Danish Twins (LSADT, N=286). MADT was initiated in 1998 and includes 4,314 twins randomly chosen from the birth years 1931-1952. Surviving participants were revisited from 2008 to 2011, where the blood samples used in the present study were collected. The survey data used in the present study was obtained from the intake survey undertaken in 1998 [Pedersen 2019]. LSADT was initiated in 1995 and includes twins aged 70 years and older. Follow-up assessments were conducted every second year until 2005. The individuals included here all participated in the 1997 assessment, where blood samples were donated by same sex twin pairs, and in 1999 participants completed the processing speed test used in the present study [Pedersen 2019]. Written informed consents were obtained from all participants. Collection and use of biological material, survey and registry information were approved by the Regional Scientific Ethical Committees for Southern Denmark (MADT: S-VF-19980072, LSADT: S-VF-20040241), and the study is registered in SDU's internal list (notification no. 11.108) and complies with the rules in the General Data Protection Regulation.

### **Genetic data.**

DNA samples of 4577 LLFS participants were genotyped at the Center for Inherited Disease Research (CIDR) using the Illumina Omni 2.5 platform, and genotype calls were cleaned following a strict quality control process described in [Bae 2013]. The genotype data were imputed with Michigan Imputation Server to the HRC panel (version r1.1 2016) of 64,940 haplotypes with 39,635,008 sites. [Das et al. 2016]. Genome-wide genotype data are available from dbGaP (**dbGaP Study Accession:** phs000397.v1.p1). The Danish Twin Registry samples were genotyped using the Illumina Infinium PsychArray (Illumina San Diego, CA, USA). Pre-imputation quality control included filtering SNPs on genotype call rate <98%, HWE  $P < 10^{-6}$ , and MAF = 0, and individuals on sample call rate <99%, relatedness and gender mismatch. Pre-phasing and imputation to the 1000 Genomes phase 3 reference panel was performed using IMPUTE2 version 2.3.2.

### **Assessment of Processing speed.**

The assessment of processing speed in participants in LLFS was conducted by administering the Digital Symbol Substitution Test (DSST) during the in-person visit at enrollment (2006—2009) and the second in-person visit conducted between 2014 and 2017. The test consists of recoding numbers into symbols using a look-up table, and the test score represents the number of substitutions completed in 90 seconds. The test score ranges from 0 to 100. Table 1 provides a summary table of the participants in LLFS included in the final analysis. The administration of the cognitive speed test in the two Danish cohorts were completed in 1998 (MADT) and 1999 (LSADT). Processing speed in the Danish twins was assessed by using a cognitive speed test

similar to the DSST, in which symbols were to be replaced with numbers (Symbol Digit Substitution Task (SDST)) within 90 seconds, and the test score represents the number of symbol substitutions. Table 1 provides a summary table of the two Danish cohorts included in the replication analysis.

### Genome-wide association study.

We removed duplicate, monomorphic and ambiguous SNPs (A/T, C/G) and selected SNPs with high imputation quality ( $R_{sq} > 0.8$ ), which resulted in 15,346,220 clean SNPs. The associations between baseline DSST scores and dosages of SNPs were tested using a linear mixed effect regression model to account for cryptic relationships. The genetic relations and population stratification were analyzed by the R/3.6.0 packages PC-Relate and PC-Air following the method by Conomos [Conomos 2015 and 2016]. The regression analysis was adjusted by age at enrollment, sex, education, and the first four genome-wide principal components. Significance of the associations was tested using the Wald test and the score test. The association analyses were conducted using the R/3.6.0 package GENESIS [Gogarten 2019] and the results were annotated using ANNOVAR [Wang 2010]. The baseline analysis as described was performed by a recently developed GWAS pipeline [Song 2021]. The associations between DSST scores at both visit 1 and visit 2 were tested using a linear mixed effect model to account for both cryptic relationships and repeated measures as implemented in the R/3.6.0 package GMMAT [Chen 2016]. This analysis was adjusted by age at enrollment, age differences between two visits, sex, education, and the first four genome-wide principal components. The score test was used for each SNP using genotype count first, and the Wald test was used on the top associated SNPs using expected dosages to obtain comparable effect sizes to the baseline GWAS.

In the replication study, the analyses were carried out using a linear regression model adjusted by sex, interview age, education, cohort (LSADT=1, MADT=2), and twin pair number as a random effect to account for the correlation between twins from the same pair. Analyses were only done on the combined sample. The regression was not adjusted by principal components since the Danish population is highly homogeneous. Among the 20 genome-wide significant SNPs discovered by the LLFS GWAS, six SNPs with imputation quality score  $> 0.3$  were tested.

### Gene-based analysis.

Burden tests were performed on the two genes Retinoic Acid Receptor Beta (*RARB*) and Thyroid Hormone Receptor Beta (*THRB*), tagged by 17 protective rare genetic variants on chromosome 3 in LLFS following the method by Madsen [Madsen 2009]. 3029 rare variants within *RARB* and 1356 rare variants within *THRB* with minor allele frequency  $< 0.01$  were aggregated by weighted sum. The weights were assigned considering the effect direction using the following formula:

$$W_i = \frac{\text{sign}(\beta_i)}{\sqrt{n_i f_i (1 - f_i)}}$$

where  $\beta_i$  is the effect of  $SNP_i$ ,  $n_i$  is the total number of subjects, and  $f_i$  is the MAF of  $SNP_i$ . The summary score of one gene for an individual was calculated as:

$$S = \sum_i^N W_i C_i$$

where  $C_i$  is the number of copies of  $SNP_i$  on that gene.

Summary scores for *RARB* and *THRB* genes can be found in **Supplement Table 4**. The associations between the scores and DSST were tested using the same model as the single SNP association test at baseline.

## Results.

**Demographic characteristics.** We identified 4207 subjects in LLFS with genotype data, and valid results of the DSST who were included in the GWAS of baseline DSST. Only 2160 of these subjects participated in the second assessment, due to mortality or dropout (88% in the oldest generation and 14% in the youngest generation). **Table 1** displays the characteristics of the 4207 LLFS participants included in the analysis of DSST at baseline, and the 2160 participants at the second visit. The enrollment age of LLFS participants ranged from 25 to 109 years with mean age of 70 years old. This group included 1286 participants from the oldest generation, with 1122 members from long-lived families (mean age = 90), and 164 spousal controls (mean age = 84). The LLFS group also included 2921 participant from the offspring generation, with 2188 members from long-lived families (mean age = 61) and 733 spousal controls (mean age = 62). The MADT included only middle-aged twins with age ranging between 45 and 68 years. The LSADT included older twins with ages 75-96 years old. The average speed score was 49.14 measured by SDST in MADT, 43.94 by DSST in LLFS, and 28.43 by SDST in LSADT, and it declined with increasing average age. The pooled Danish cohort (MADT+LSADT) with average age 60 and speed score 45.00 was comparable to the speed score in the LLFS cohort.

**Genome-wide association Study.** Analyses of the baseline DSST scores in the LLFS identified 20 SNPs in four distinct genomic regions that passed the genome-wide significance threshold with  $p\text{-value} < 5 \times 10^{-8}$  (**Table 2**). The complete list of SNPs with “suggestive” level of significance ( $p\text{-value} < 5 \times 10^{-6}$ ) is in **Supplement Table 1**. All the 20 genome-wide level significant SNPs were rare variants with  $\text{MAF} < 0.01$ . 18 SNPs showed a positive effect on the DSST score, and carriers of the rare alleles had a DSST score greater than 20 points. 17 SNPs were within a region of 1.5 MB on chromosome 3 that includes two genes that belong to a family of thyroid hormone receptors: Retinoic Acid Receptor Beta (*RARB*) and Thyroid Hormone Receptor Beta (*THRB*), as well as miR-4792 which may regulate a number of mitochondria-related genes, among other targets [Liu 2019]. An additional rare protective variant rs58169119 was in an intergenic region on chromosome 8, and the rare allele of this variant was associated with an average increase of 28 points in DSST. The two deleterious rare SNPs also showed very large negative effects, whereby the rare alleles were associated with a decrease in DSST score  $> 50$ , specifically: SNP rs146299120 located in an intergenic region on chromosome 9, and a SNP without rs ID located on chromosome 18 in an intergenic region of the Deleted in Colorectal Cancer (*DCC*) gene, which encodes the netrin 1 receptor, that is involved in nervous system development [Boyer and Gupton 2018; Duman-Scheel M 2009]. Six of the rare variants with imputation score  $R^2 > 0.3$  were analyzed in the combined Danish twin cohort (**Table 2; Supplement Table 3**). The analyses confirmed the association of four SNPs on chromosome 3 with SDST score after Bonferroni correction for 3 independent loci ( $p\text{-value} < 0.017$ ). The SNP rs7623455 had the lowest  $p\text{-value} = 2.7 \times 10^{-18}$  with large positive effect size ( $\beta=10.7$ ) in the combined Danish cohort. All the rare variants that replicated in the Danish cohort also showed positive effects and increase SDST scores by more than 8 points.

The GWAS of the repeated assessment of DSST identified 7 SNPs reaching genome-wide significance level (**Table 2; Supplement Table 2**). Four SNPs were among the 20 SNPs identified in the baseline analysis. The remaining three SNPs with p-values  $< 5 \times 10^{-8}$  in the longitudinal analysis all had p-values  $< 5 \times 10^{-7}$  in the baseline analysis (**Table 2**). SNP rs140387800 was in an intergenic region on chromosome 3 with large protective effect size ( $\beta=40.5$ ). SNP rs6754826 was in an intergenic region on chromosome 2 and the rare allele was associated with a 26.6-point increase in DSST. The SNP rs116193953 was in an intergenic region on chromosome 1 and the rare allele was associated with an average decrease of 28 points in DSST.

### **Gene-based analysis.**

The GWAS results showed strong evidence that two genes, *RARB* and *THRB*, on chromosome 3 were potentially related to processing speed. We therefore performed Burden tests on these two genes in LLFS. The weighted score for *RARB* ranged from -3.29 to 12.98, with mean 0.052 by aggregating 3029 rare alleles as described in the Methods section. The weighted score for *THRB* ranged from -3.26 to 30.50, with mean 0.016 by aggregating 1356 rare alleles. Both gene scores were associated with DSST scores with p-value  $< 0.025$  after Bonferroni correction for two tests. The two genes also showed overall protective effects on processing speed. Each unit increase in *RARB* score increased DSST by 2.13 points (p-value =  $7.7 \times 10^{-17}$ ) and each unit increase in *THRB* scores increased DSST by 0.96 points (p-value =  $1.3 \times 10^{-8}$ ) (**Table 3**).

### **Replication of Published Results.**

We looked at a replication of the recently published GWAS of global cognition [Davies 2019]. Out of the 11,600 genome-wide level significant SNPs in that study, 11,564 SNPs were present in the LLFS, and only 12 SNPs passed the Bonferroni corrected p-value threshold and show concordant effects ( $1.15 \times 10^{-4}$ , 434 independent significant SNPs). These 12 SNPs were within a 10kb intergenic region on chromosome 1 and their associations in LLFS are summarized in **Table 4**. None of our results showed a genome-wide significant association with global cognition.

### **Discussion.**

Our GWAS of the processing speed (based on DSST scores) measured in 4207 LLFS participants, discovered 20 genome-wide significant SNPs (p-value  $< 5 \times 10^{-8}$ ) on chromosomes 3, 8, 9 and 18. Sixteen of these SNPs formed a cluster of rare protective variation on chromosome 3. These rare variants showed very large effect sizes compared to common variants that reached only suggestive significance threshold (p-value  $< 5 \times 10^{-6}$ ) [Supplement Table 1]. All sixteen rare SNPs were confirmed in the longitudinal analysis with comparable effect sizes [Table 2]. We also replicated the association of four of the sixteen variants, rs7623455, rs9821776, rs9821587 and rs78704059, in an independent sample of Danish ancestry (p-values  $< 0.017$ ). Note that SNPs shown in Table 2 were rare and independent in White participants; however, the same SNPs may potentially be common and in LD in different populations/race groups. E.g., rs9821776 and rs9821587 from Table 2 are in complete LD in individuals of African ancestry, according to LDlink, NIH supported portal.

We also investigated the association of 11,564 SNPs that reached a genome-wide significant association with global cognition in a recent GWAS with 300,486 individuals [Davies 2019]. Although our analysis focused on the specific cognitive domain of processing speed and attention

assessed by the DSST score, we replicated the association of 12 common SNPs in an intergenic region on chromosome 1 with concordant effect directions [Table 4]. None of the rare variants we found associated with DSST were reported in the analysis by [Davies 2019], but their study dropped SNPs with minor allele count < 25.

Another GWAS of 30,576 individuals of European ancestry and 5758 individuals of African ancestry investigated the associations of rare variants with processing speed but did not identify any genome-wide significant associations [Bressler 2021]. Their gene-based analysis adjusted for age and gender discovered one genome-wide significant result: ring finger protein 19A (*RNF19A*). We attempted replication of this association in our data using our method and adjusting for age, gender, education and first 4 PCs. The association reached a p-value of  $7.6 \times 10^{-4}$  for this gene, consistent with their result. A possible reason that our study of only 4207 individuals in LLFS was able to identify genome-wide significant rare variants could be that the family based design produced a dataset enriched for protective rare variants, thus increasing our discovery power. The allele counts per 10,000 alleles of the top rare variants were doubled or tripled comparing to the general European population reported on dbSNP.

Although the most significant variants were very rare, they were notably concentrated in/near two functionally related genes on chromosome 3, *RARB* and *THRB*. The *RARB* and *THRB* belong to the family of thyroid hormone receptors, which could potentially contribute to their associations with the processing speed. For example, *THRB* may mediate the biological activities of thyroid hormone and through this influence the speed of metabolism, which in turn may influence the processing speed. Some studies also linked reduced thyroid function during aging with a decline in cognition [Bégin 2008; Jasim 2017]. Additional biological effects of these genes could play a role in their connection to the processing speed as well. E.g., the *RARB* encodes a receptor binding the retinoic acid, the biologically active form of vitamin A. The deficiency of vitamin A was associated with the cognitive decline in aging [Pallet 2015; Wołoszynowska-Fraser 2020]. The *RARB* rare variant burden was associated with severe cognitive deficits in schizophrenia [Reay 2020]. The GTEX portal shows that there are several expression quantitative trait loci in the brain frontal cortex for the *RARB* gene, warranting further exploration of the observed associations. The gene-based analysis also confirmed that *RARB* and *THRB* genes are both associated with DSST scores and have an overall beneficial effect on the processing speed.

In addition, the three top SNPs on chromosome 3 shown in Table 2 are close to miR-4792. Targets of this microRNA include a number of mitochondria-related genes [Liu 2019]. This points to a possible role of miR-4792 in energy metabolism and ATP production, which in turn may also contribute to the processing speed. This connection is, however, underexplored in the literature, and is purely hypothetical.

A potential limitation of our results is that we used imputed data for rare variants. We were extremely conservative and only SNPs with a high quality score were retained in the analysis. The LLFS has recently completed Whole Genome Sequencing (WGS) data in approximately 3,500 participants. We compared the top imputed SNPs to the WGS data in the common set of individuals with both data types, and observed a perfect concordance. Therefore, these rare variants imputed to the HRC panel with high quality appear to be trustworthy.

In summary, our study provided new evidence that rare genetic variants can play an important role in cognitive aging. We discovered 20 rare protective SNPs located in/near two genes on chromosome 3, *RARB* and *THRB*, belonging to the family of thyroid hormone receptors that may help aging individuals maintain younger processing speed, potentially through preserving the speed of metabolism. This explanation, however, needs to be thoroughly explored before these genes and relevant biological pathways could become new targets for therapies helping older adults to maintain the processing speed without undesirable trade-offs, such as depletion of stem cell reserves, damage accumulation, etc. [reviewed in Ukraintseva et al. 2021]. In future research, the gene-based analysis scanning through the whole genome might detect more genomic regions associated with the processing speed with increased power.

**Acknowledgments:** Supported by the National Institute on Aging (K01AG057798 to S.L.A., 5U19AG063893 5U01AG023749 to S.C., 5U01AG023755 to T.T.P., U19AG063893 R01AG062623 to S.U. and K.A., 5U01AG023712, 5U01AG023744, 5U01AG023746); Genotyping of the Danish twins 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).

#### **Reference:**

Harada CN, Natelson Love MC, Triebel KL. Normal cognitive aging. *Clin Geriatr Med.* 2013 Nov;29(4):737-52. doi: 10.1016/j.cger.2013.07.002. PMID: 24094294; PMCID: PMC4015335.

United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Ageing 2019: Highlights (ST/ESA/SER.A/430).

Trampush JW, Yang ML, Yu J, Knowles E, Davies G, Liewald DC, Starr JM, Djurovic S, Melle I, Sundet K, Christoforou A, Reinvang I, DeRosse P, Lundervold AJ, Steen VM, Espeseth T, Räikkönen K, Widen E, Palotie A, Eriksson JG, Giegling I, Konte B, Roussos P, Giakoumaki S, Burdick KE, Payton A, Ollier W, Horan M, Chiba-Falek O, Attix DK, Need AC, Cirulli ET, Voineskos AN, Stefanis NC, Avramopoulos D, Hatzimanolis A, Arking DE, Smyrnis N, Bilder RM, Freimer NA, Cannon TD, London E, Poldrack RA, Sabb FW, Congdon E, Conley ED, Scult MA, Dickinson D, Straub RE, Donohoe G, Morris D, Corvin A, Gill M, Hariri AR, Weinberger DR, Pendleton N, Bitsios P, Rujescu D, Lahti J, Le Hellard S, Keller MC, Andreassen OA, Deary IJ, Glahn DC, Malhotra AK, Lencz T. GWAS meta-analysis reveals novel loci and genetic correlates for general cognitive function: a report from the COGENT consortium. *Mol Psychiatry.* 2017 Mar;22(3):336-345. doi: 10.1038/mp.2016.244. Epub 2017 Jan 17. Erratum in: *Mol Psychiatry.* 2017 Nov;22(11):1651-1652. PMID: 28093568; PMCID: PMC5322272.

Davies G, Armstrong N, Bis JC, Bressler J, Chouraki V, Giddaluru S, Hofer E, Ibrahim-Verbaas CA, Kirin M, Lahti J, van der Lee SJ, Le Hellard S, Liu T, Marioni RE, Oldmeadow C, Postmus I, Smith AV, Smith JA, Thalamuthu A, Thomson R, Vitart V, Wang J, Yu L, Zgaga L, Zhao W, Boxall R, Harris SE, Hill WD, Liewald DC, Luciano M, Adams H, Ames D, Amin N, Amouyel P, Assareh AA, Au R, Becker JT, Beiser A, Berr C, Bertram L, Boerwinkle E, Buckley BM,



Campbell H, Corley J, De Jager PL, Dufouil C, Eriksson JG, Espeseth T, Faul JD, Ford I; Generation Scotland, Gottesman RF, Griswold ME, Gudnason V, Harris TB, Heiss G, Hofman A, Holliday EG, Huffman J, Kardina SL, Kochan N, Knopman DS, Kwok JB, Lambert JC, Lee T, Li G, Li SC, Loitfelder M, Lopez OL, Lundervold AJ, Lundqvist A, Mather KA, Mirza SS, Nyberg L, Oostra BA, Palotie A, Papenberg G, Pattie A, Petrovic K, Polasek O, Psaty BM, Redmond P, Reppermund S, Rotter JI, Schmidt H, Schuur M, Schofield PW, Scott RJ, Steen VM, Stott DJ, van Swieten JC, Taylor KD, Trollor J, Trompet S, Uitterlinden AG, Weinstein G, Widen E, Windham BG, Jukema JW, Wright AF, Wright MJ, Yang Q, Amieva H, Attia JR, Bennett DA, Brodaty H, de Craen AJ, Hayward C, Ikram MA, Lindenberger U, Nilsson LG, Porteous DJ, Rääkkönen K, Reinvang I, Rudan I, Sachdev PS, Schmidt R, Schofield PR, Srikanth V, Starr JM, Turner ST, Weir DR, Wilson JF, van Duijn C, Launer L, Fitzpatrick AL, Seshadri S, Mosley TH Jr, Deary IJ. Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53949). *Mol Psychiatry*. 2015 Feb;20(2):183-92. doi: 10.1038/mp.2014.188. Epub 2015 Feb 3. PMID: 25644384; PMCID: PMC4356746.

Cirulli ET, Kasperaviciūte D, Attix DK, Need AC, Ge D, Gibson G, Goldstein DB. Common genetic variation and performance on standardized cognitive tests. *Eur J Hum Genet*. 2010 Jul;18(7):815-20. doi: 10.1038/ejhg.2010.2. Epub 2010 Feb 3. Erratum in: *Eur J Hum Genet*. 2010 Jul;18(7):820. PMID: 20125193; PMCID: PMC2987367.

Liu P, Pu J, Zhang J, Chen Z, Wei K, Shi L. Bioinformatic analysis of miR-4792 regulates Radix Tetrastigma hemsleyani flavone to inhibit proliferation, invasion, and induce apoptosis of A549 cells. *Onco Targets Ther*. 2019 Feb 20;12:1401-1412. doi: 10.2147/OTT.S182525. PMID: 30863107; PMCID: PMC6389004.

Luciano M, Hansell NK, Lahti J, et al. Whole genome association scan for genetic polymorphisms influencing information processing speed. *Biol Psychol*. 2011;86(3):193–202. doi:10.1016/j.biopsycho.2010.11.008

Ibrahim-Verbaas CA, Bressler J, DeBette S, et al. GWAS for executive function and processing speed suggests involvement of the CADM2 gene. *Mol Psychiatry*. 2016;21(2):189–197. doi:10.1038/mp.2015.37

Davies G, Lam M, Harris SE, Trampush JW, Luciano M, Hill WD, Hagenaars SP, Ritchie SJ, Marioni RE, Fawns-Ritchie C, Liewald DCM, Okely JA, Ahola-Olli AV, Barnes CLK, Bertram L, Bis JC, Burdick KE, Christoforou A, DeRosse P, Djurovic S, Espeseth T, Giakoumaki S, Giddaluru S, Gustavson DE, Hayward C, Hofer E, Ikram MA, Karlsson R, Knowles E, Lahti J, Leber M, Li S, Mather KA, Melle I, Morris D, Oldmeadow C, Palviainen T, Payton A, Pazoki R, Petrovic K, Reynolds CA, Sargurupremraj M, Scholz M, Smith JA, Smith AV, Terzikhan N, Thalamuthu A, Trompet S, van der Lee SJ, Ware EB, Windham BG, Wright MJ, Yang J, Yu J, Ames D, Amin N, Amouyel P, Andreassen OA, Armstrong NJ, Assareh AA, Attia JR, Attix D, Avramopoulos D, Bennett DA, Böhmer AC, Boyle PA, Brodaty H, Campbell H, Cannon TD, Cirulli ET, Congdon E, Conley ED, Corley J, Cox SR, Dale AM, Dehghan A, Dick D, Dickinson D, Eriksson JG, Evangelou E, Faul JD, Ford I, Freimer NA, Gao H, Giegling I, Gillespie NA, Gordon SD, Gottesman RF, Griswold ME, Gudnason V, Harris TB, Hartmann AM,

Hatzimanolis A, Heiss G, Holliday EG, Joshi PK, Kähönen M, Kardia SLR, Karlsson I, Klei L, Knopman DS, Kochan NA, Konte B, Kwok JB, Le Hellard S, Lee T, Lehtimäki T, Li SC, Lill CM, Liu T, Koini M, London E, Longstreth WT Jr, Lopez OL, Loukola A, Luck T, Lundervold AJ, Lundquist A, Lyytikäinen LP, Martin NG, Montgomery GW, Murray AD, Need AC, Noordam R, Nyberg L, Ollier W, Papenberg G, Pattie A, Polasek O, Poldrack RA, Psaty BM, Reppermund S, Riedel-Heller SG, Rose RJ, Rotter JI, Roussos P, Rovio SP, Saba Y, Sabb FW, Sachdev PS, Satizabal CL, Schmid M, Scott RJ, Scult MA, Simino J, Slagboom PE, Smyrnis N, Soumaré A, Stefanis NC, Stott DJ, Straub RE, Sundet K, Taylor AM, Taylor KD, Tzoulaki I, Tzourio C, Uitterlinden A, Vitart V, Voineskos AN, Kaprio J, Wagner M, Wagner H, Weinhold L, Wen KH, Widen E, Yang Q, Zhao W, Adams HHH, Arking DE, Bilder RM, Bitsios P, Boerwinkle E, Chiba-Falek O, Corvin A, De Jager PL, DeBette S, Donohoe G, Elliott P, Fitzpatrick AL, Gill M, Glahn DC, Hägg S, Hansell NK, Hariri AR, Ikram MK, Jukema JW, Vuoksimaa E, Keller MC, Kremen WS, Launer L, Lindenberg U, Palotie A, Pedersen NL, Pendleton N, Porteous DJ, Raitakari OT, Ramirez A, Reinvang I, Rudan I, Dan Rujescu, Schmidt R, Schmidt H, Schofield PW, Schofield PR, Starr JM, Steen VM, Trollor JN, Turner ST, Van Duijn CM, Villringer A, Weinberger DR, Weir DR, Wilson JF, Malhotra A, McIntosh AM, Gale CR, Seshadri S, Mosley TH Jr, Bressler J, Lencz T, Deary IJ. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat Commun.* 2018 May 29;9(1):2098. doi: 10.1038/s41467-018-04362-x. Erratum in: *Nat Commun.* 2019 May 1;10(1):2068. PMID: 29844566; PMCID: PMC5974083.

Sebastiani P, Andersen SL, Sweigart B, Du M, Cosentino S, Thyagarajan B, Christensen K, Schupf N, Perls TT. Patterns of multi-domain cognitive aging in participants of the Long Life Family Study. *Geroscience.* 2020 Oct;42(5):1335-1350. doi: 10.1007/s11357-020-00202-3. Epub 2020 Jun 8. PMID: 32514870; PMCID: PMC7525612.

Wojczynski M.K., Province M.A. (2019) The Long Life Family Study (LLFS). In: Gu D., Dupre M. (eds) *Encyclopedia of Gerontology and Population Aging*. Springer, Cham. [https://doi.org/10.1007/978-3-319-69892-2\\_1101-1](https://doi.org/10.1007/978-3-319-69892-2_1101-1)

Stevenson M, Bae H, Schupf N, et al. Burden of disease variants in participants of the Long Life Family Study [published correction appears in *Aging (Albany NY)*. 2015 Jun;7(6):450]. *Aging (Albany NY)*. 2015;7(2):123-132. doi:10.18632/aging.100724

Sebastiani P, Hadley EC, Province M, Christensen K, Rossi W, Perls TT, et al. A family longevity selection score: ranking sibships by their longevity, size, and availability for study. *Am J Epidemiol.* 2009;170(12):1555-62. PMCID: 2800272

Newman AB, Glynn NW, Taylor CA, Sebastiani P, Perls TT, Mayeux R, et al. Health and function of participants in the Long Life Family Study: A comparison with other cohorts. *Aging (Albany NY)*. 2011;3(1):63-76. PMCID: 3047140.

Elo IT, Mykyta L, Sebastiani P, Christensen K, Glynn NW, Perls T. Age validation in the long life family study through a linkage to early-life census records. *J Gerontol B Psychol Sci Soc Sci.* 2013;68(4):580-5. PMCID: 3674734.

Pedersen DA, Larsen LA, Nygaard M, et al. The Danish Twin Registry: An Updated Overview. *Twin Res Hum Genet.* 2019;22(6):499-507. doi:10.1017/thg.2019.72

Bae HT, Sebastiani P, Sun JX, Andersen SL, Daw EW, Terracciano A, Ferrucci L, Perls TT. Genome-wide association study of personality traits in the long life family study. *Front Genet.* 2013;4:65.

Conomos MP, Miller MB, Thornton TA. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. *Genet Epidemiol.* 2015;39(4):276-293. doi:10.1002/gepi.21896

Conomos MP, Reiner AP, Weir BS, Thornton TA. Model-free Estimation of Recent Genetic Relatedness. *Am J Hum Genet.* 2016;98(1):127-148. doi:10.1016/j.ajhg.2015.11.022

Gogarten SM, Sofer T, Chen H, et al. Genetic association testing using the GENESIS R/Bioconductor package. *Bioinformatics.* 2019;35(24):5346-5348. doi:10.1093/bioinformatics/btz567

Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from next-generation sequencing data *Nucleic Acids Research*, 38:e164, 2010

Song et al., (2021). nf-gwas-pipeline: A Nextflow Genome-Wide Association Study Pipeline. *Journal of Open Source Software*, 6(59), 2957, <https://doi.org/10.21105/joss.02957>

Han Chen, Matthew P. Conomos. GMMAT: Generalized linear Mixed Model Association Tests. 2016. [https://content.sph.harvard.edu/xlin/dat/GMMAT\\_user\\_manual\\_v0.7.pdf](https://content.sph.harvard.edu/xlin/dat/GMMAT_user_manual_v0.7.pdf)

Madsen BE, Browning SR. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet.* 2009 Feb;5(2):e1000384. doi: 10.1371/journal.pgen.1000384. Epub 2009 Feb 13. PMID: 19214210; PMCID: PMC2633048.

Liu P, Pu J, Zhang J, Chen Z, Wei K, Shi L. Bioinformatic analysis of miR-4792 regulates *Radix Tetragyna hemsleyana* flavone to inhibit proliferation, invasion, and induce apoptosis of A549 cells. *Onco Targets Ther.* 2019;12:1401-1412. Published 2019 Feb 20. doi:10.2147/OTT.S182525

Boyer NP, Gupton SL. Revisiting Netrin-1: One Who Guides (Axons). *Front Cell Neurosci.* 2018;12:221. Published 2018 Jul 31. doi:10.3389/fncel.2018.00221

Duman-Scheel M. Netrin and DCC: axon guidance regulators at the intersection of nervous system development and cancer. *Curr Drug Targets.* 2009;10(7):602-610. doi:10.2174/138945009788680428

Bressler J, Davies G, Smith AV, et al. Association of low-frequency and rare coding variants with information processing speed. *Transl Psychiatry.* 2021;11(1):613. Published 2021 Dec 4. doi:10.1038/s41398-021-01736-6

Jasim S, Gharib H. THYROID AND AGING. *Endocr Pract.* 2018 Apr;24(4):369-374. doi: 10.4158/EP171796.RA. Epub 2017 Aug 17. PMID: 28816538.

Bégin ME, Langlois MF, Lorrain D, Cunnane SC. Thyroid Function and Cognition during Aging. *Curr Gerontol Geriatr Res.* 2008;2008:474868. doi: 10.1155/2008/474868. Epub 2008 Sep 1. PMID: 19415145; PMCID: PMC2671998.

Ukrainitseva S, Arbeev K, Duan M, Akushevich I, Kulminski A, Stallard E, Yashin A. Decline in biological resilience as key manifestation of aging: Potential mechanisms and role in health and longevity. *Mech Ageing Dev.* 2021 Mar;194:111418. doi: 10.1016/j.mad.2020.111418. Epub 2020 Dec 16. PMID: 33340523; PMCID: PMC7882032.

Wołoszynowska-Fraser MU, Kouchmeshky A, McCaffery P. Vitamin A and Retinoic Acid in Cognition and Cognitive Disease. *Annu Rev Nutr.* 2020 Sep 23;40:247-272. doi: 10.1146/annurev-nutr-122319-034227. PMID: 32966186.

Pallet, V. and Touyarot, K. Vitamin A and Cognitive Processes. *Nutrition and Aging*, vol. 3, no. 1, pp. 21-31, 2015. doi: 10.3233/NUA-150048

Reay, W.R., Atkins, J.R., Quidé, Y. *et al.* Polygenic disruption of retinoid signaling in schizophrenia and a severe cognitive deficit subtype. *Mol Psychiatry* **25**, 719–731 (2020). <https://doi.org/10.1038/s41380-018-0305-0>

## Rare Genetic Variants Correlate with Better Processing Speed

Table1. Demographic characteristics of the LLFS cohort at baseline measurement and at the second visit, and the MADT and the LSADT cohorts.

	LLFS baseline (N=4207)	LLFS offspring baseline (N=2188)	LLFS offspring spouse baseline (N=733)	LLFS proband baseline (N=1122)	LLFS proband spouse baseline (N=164)
DSST score, mean $\pm$ sd [min, max]	43.94 $\pm$ 15.94 [0, 93]	51.00 $\pm$ 12.36 [0, 93]	48.69 $\pm$ 12.97 [1, 85]	28.42 $\pm$ 12.36 [0, 73]	34.65 $\pm$ 13.43 [0, 67]
Age, mean $\pm$ sd [min, max]	69.86 $\pm$ 15.20 [25, 109]	61.09 $\pm$ 8.20 [32, 88]	61.71 $\pm$ 8.54 [25, 88]	90.29 $\pm$ 6.39 [49, 109]	83.68 $\pm$ 7.24 [55, 98]
Female, n (%)	2291 (54.46)	1268 (57.95)	337 (45.98)	560 (49.91)	126 (76.83)
Education years, mean $\pm$ sd [min, max]	11.75 $\pm$ 3.48 [1, 17]	12.57 $\pm$ 3.01 [2, 17]	12.01 $\pm$ 3.35 [3, 17]	10.15 $\pm$ 3.78 [1, 17]	10.49 $\pm$ 3.64 [2, 16]

	LLFS 2nd visit (N=2160)	LLFS offspring 2nd visit (N=1417)	LLFS offspring spouse 2nd visit (N=441)	LLFS proband 2nd visit (N=253)	LLFS proband spouse 2nd visit (N=49)
DSST score, mean $\pm$ sd [min, max]	44.88 $\pm$ 14.13 [0, 90]	48.43 $\pm$ 12.23 [0, 90]	45.16 $\pm$ 12.22 [4, 78]	26.71 $\pm$ 12.45 [0, 68]	33.73 $\pm$ 12.14 [0, 59]
Age, mean $\pm$ sd [min, max]	72.30 $\pm$ 11.18 [43, 106]	68.76 $\pm$ 7.77 [43, 94]	69.90 $\pm$ 7.81 [47, 90]	93.32 $\pm$ 6.52 [57, 106]	88.02 $\pm$ 6.90 [71, 99]
Female, n (%)	1196 (55.37)	810 (57.16)	206 (46.71)	137 (54.15)	43 (87.76)
Education years, mean $\pm$ sd [min, max]	12.25 $\pm$ 3.23 [2, 17]	12.68 $\pm$ 2.99 [2, 17]	12.09 $\pm$ 3.40 [3, 17]	10.51 $\pm$ 3.40 [3, 17]	10.35 $\pm$ 3.62 [2, 15]

	MADT (N=1146)	LSADT (N=286)	Danish combined (MADT+LSADT) (N=1432)
SDST score, mean $\pm$ sd [min, max]	49.14 $\pm$ 10.53 [8, 81]	28.43 $\pm$ 12.67 [0, 61]	45.00 $\pm$ 13.76 [0, 81]
Age, mean $\pm$ sd [min, max]	55.68 $\pm$ 5.97 [45, 68]	79.98 $\pm$ 3.72 [75, 96]	60.54 $\pm$ 11.21 [45, 96]
Female, n (%)	550 (47.99)	190 (66.43)	740 (51.68)
Education, n (%)			
<7 years - pro training	0 (0)	22 (7.89)	22 (1.57)
7+ years -pro training	234 (20.86)	113 (40.50)	347 (24.77)
8-9 years + pro training	299 (26.65)	65 (23.30)	364 (25.98)
9+ years + pro training	262 (23.35)	21 (7.53)	283 (20.20)
short-medium college/UNI	249 (22.19)	47 (16.85)	296 (21.13)
long college/UNI	78 (6.95)	11 (3.94)	89 (6.35)

Table2. SNPs with p-values  $< 5 \times 10^{-8}$  from the baseline or longitudinal GWAS in LLFS and replication in the Danish combined cohort.

SNP Info						DSST LLFS Baseline GWAS				DSST LLFS Longitudinal GWAS			SDST Danish Replication			
SNP	Chr	Position	Gene	Ref	Alt	EAF	Effect	SE	p	Effect	SE	p	EAF	Effect	SE	p
rs7623455	3	24713169	MIR4792(dist=150243) RARB(dist=157645)	G	A	0.0006	29.93	5.15	6.39E-09	26.54	4.81	3.50E-08	0.0003	10.73	1.20	2.72E-18
rs9821776	3	24712332	MIR4792(dist=149406) RARB(dist=158482)	G	C	0.0006	29.93	5.15	6.41E-09	26.54	4.81	3.51E-08	0.0004	9.63	2.23	1.80E-05
rs9821587	3	24712189	MIR4792(dist=149263) RARB(dist=158625)	G	T	0.0006	29.93	5.16	6.41E-09	26.55	4.81	3.50E-08	0.0004	9.61	2.20	1.40E-05
rs75963215	3	24423893	THRB	A	C	0.0010	20.96	3.80	3.44E-08	19.30	3.58	7.06E-08	0.0007	6.83	3.51	0.052
rs59914825	3	24424415	THRB	T	C	0.0010	20.96	3.80	3.44E-08	19.30	3.58	7.07E-08	0.0007	6.94	3.52	0.049
rs4266131	3	24425760	THRB	A	G	0.9990	-20.88	3.79	3.61E-08	-19.25	3.57	7.16E-08				
rs58169119	8	82350918	FABP5(dist=153906) PMP2(dist=1645)	G	A	0.0005	-28.22	5.13	3.74E-08	-27.88	4.86	9.59E-09				
rs148022846	3	25100109	RARB	A	T	0.0005	30.56	5.55	3.77E-08	28.19	5.22	6.70E-08				
rs146299120	9	81654385	PSAT1(dist=709376) LOC101927450(dist=95953)	C	T	0.0001	-53.09	9.67	4.07E-08							
	18	50217262	DCC	G	A	0.0002	-51.99	9.48	4.14E-08							
rs189337466	3	25228798	RARB	T	A	0.0005	30.38	5.56	4.63E-08	28.00	5.22	8.40E-08				
rs78704059	3	25170804	RARB	A	G	0.0005	30.34	5.55	4.64E-08	27.97	5.22	8.39E-08	0.0012	12.67	3.55	4.00E-04
rs59296535	3	25206678	RARB	C	T	0.0005	30.33	5.55	4.67E-08	27.95	5.22	8.54E-08				
rs74467766	3	25206522	RARB	A	C	0.0005	30.33	5.55	4.69E-08	27.95	5.22	8.55E-08				
rs80011850	3	25194665	RARB	A	G	0.0005	30.33	5.55	4.70E-08	27.94	5.22	8.62E-08				
rs77306558	3	25223736	RARB	A	T	0.0005	30.34	5.55	4.70E-08	27.96	5.22	8.53E-08				
rs79046847	3	25217506	RARB	A	G	0.0005	30.33	5.55	4.71E-08	27.95	5.22	8.57E-08				
rs77025115	3	25192336	RARB	T	G	0.0005	30.33	5.55	4.71E-08	27.95	5.22	8.59E-08				
rs80164536	3	25212666	RARB	C	G	0.0005	30.33	5.55	4.71E-08	27.95	5.22	8.58E-08				
rs80223246	3	25206289	RARB	C	T	0.0005	30.32	5.55	4.75E-08	27.94	5.22	8.64E-08				
rs140387800	3	26853308	LRR3B(dist=101043) NEK10(dist=299086)	G	C	0.0003	40.46	7.76	1.86E-07	40.42	7.28	2.78E-08				
rs6754826	2	174325120	CDCA7(dist=91402) SP3(dist=446067)	T	C	0.9995	-26.56	5.20	3.20E-07	-26.29	4.72	2.59E-08				
rs116193953	1	163629475	LOC100422212(dist=236494) PBX1(dist=899122)	A	C	0.0006	-26.87	5.28	3.66E-07	-28.09	5.11	3.83E-08				

- Chr, Chromosome; Position, position of SNP on GRCh37 reference panel; Ref and Alt, reference allele and alternative allele; EAF, effect allele frequency

Table3. Association results of the two gene summary scores in LLFS.

Gene	Chr	start	end	Number of rare variants (MAF < 0.01)	Effect	SE	p
RARB	3	24829321	25597932	3029	2.13	0.25	7.68x10 <sup>-17</sup>
THRB	3	24117153	24495708	1356	0.96	0.17	1.28x10 <sup>-8</sup>

Table4. SNPs replicated in LLFS baseline GWAS from published results.

SNP Info						GCF as Discovery			DSST LLFS GWAS as Replication			
SNP	Chr	Position	Gene	Ref	Alt	EAF	Z_Score	P	EAF_LLFS	Effect	SE	p
rs6659327	1	183405853	NMNAT2(dist=18219),SMG7-AS1(dist=24158)	G	A	0.4858	5.571	2.53E-08	0.5112	0.985	0.235	2.71E-05
rs6684015	1	183405794	NMNAT2(dist=18160),SMG7-AS1(dist=24217)	G	A	0.5141	-5.55	2.86E-08	0.4889	-0.984	0.235	2.77E-05
rs6424896	1	183405648	NMNAT2(dist=18014),SMG7-AS1(dist=24363)	C	T	0.4858	5.562	2.67E-08	0.5110	0.983	0.235	2.84E-05
rs4047801	1	183415318	NMNAT2(dist=27684),SMG7-AS1(dist=14693)	G	A	0.4859	5.539	3.04E-08	0.5112	0.969	0.235	3.71E-05
rs7550777	1	183425047	NMNAT2(dist=37413),SMG7-AS1(dist=4964)	C	T	0.4452	-5.968	2.41E-09	0.4196	-0.962	0.237	4.83E-05
rs4652802	1	183426758	NMNAT2(dist=39124),SMG7-AS1(dist=3253)	G	A	0.555	5.981	2.22E-09	0.5806	0.961	0.237	4.89E-05
rs3965889	1	183407953	NMNAT2(dist=20319),SMG7-AS1(dist=22058)	G	A	0.5535	5.928	3.06E-09	0.5795	0.938	0.236	7.18E-05
rs10911334	1	183405382	NMNAT2(dist=17748),SMG7-AS1(dist=24629)	G	C	0.4465	-5.923	3.16E-09	0.4205	-0.937	0.237	7.63E-05
rs3856126	1	183416374	NMNAT2(dist=28740),SMG7-AS1(dist=13637)	G	T	0.4473	-5.978	2.26E-09	0.4202	-0.935	0.237	7.69E-05
rs12039578	1	183413886	NMNAT2(dist=26252),SMG7-AS1(dist=16125)	G	A	0.4466	-5.978	2.26E-09	0.4206	-0.930	0.237	8.39E-05
rs11808408	1	183415589	NMNAT2(dist=27955),SMG7-AS1(dist=14422)	G	A	0.4464	-5.98	2.23E-09	0.4206	-0.929	0.237	8.60E-05
rs17558314	1	183415188	NMNAT2(dist=27554),SMG7-AS1(dist=14823)	C	T	0.4463	-5.961	2.51E-09	0.4206	-0.929	0.237	8.64E-05

- Chr, Chromosome; Position, position of SNP on GRCh37 reference panel; Ref and Alt, reference allele and alternative allele; EAF, effect allele frequency

Figure1. QQ-plot (left) and Manhattan plot (right) of the genome-wide association with baseline DSST scores in LLFS. The x-axis of QQ-plot reports the expected  $-\log_{10}(p\text{-value})$  and the y-axis reports the observed  $-\log_{10}(p\text{-value})$ . The x-axis of Manhattan plot reports chromosomes and coordinates within chromosomes. The y-axis reports the  $-\log_{10}(p\text{-value})$ . The genomic control parameter is 1.04, indicating population stratification addressed properly.

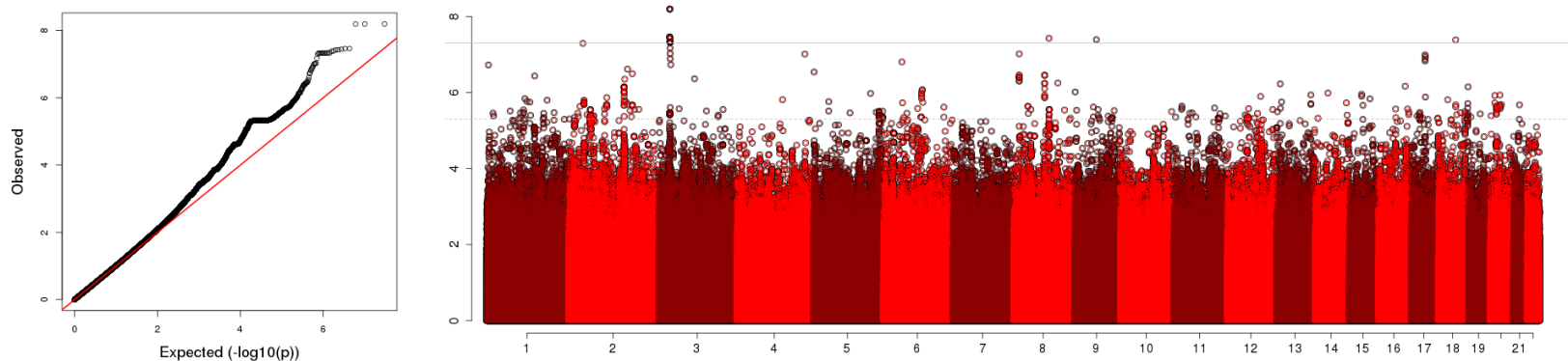


Figure2. Examples of boxplots of baseline DSST scores by SNP status. The x-axis reports the number of SNPs carried by individuals in this group. The y-axis reports the baseline DSST scores. Carriers of all three SNPs show better average performance in the DSST assessment.

