

The effect of Alzheimer's disease-associated genetic variants on longevity

Niccolò Tesi^{1,2,3}, Marc Hulsman^{1,2,3}, Sven J. van der Lee^{1,2}, Iris E. Jansen^{1,4}, Najada Stringa⁵, Natasja M. van Schoor⁵, Philip Scheltens¹, Wiesje M. van der Flier^{1,5}, Martijn Huisman⁵, Marcel J. T. Reinders³, and Henne Holstege^{1,2,3}

¹ Section Genomics of Neurodegenerative Diseases and Aging, Department of Clinical Genetics, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands

² Alzheimer Centre, Department of Neurology, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands

³ Delft Bioinformatics Lab, Delft University of Technology, Delft, The Netherlands

⁴ Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, VU, Amsterdam, The Netherlands

⁵ Department of Epidemiology and Data Sciences, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands

Keywords: cognitively healthy, centenarians, aging, Alzheimer's disease, effect on aging, protective variants

Corresponding Author

Dr. Henne Holstege

Alzheimer's Center Amsterdam

Amsterdam UMC

De Boelelaan 1118

1081 HZ, Amsterdam, The Netherlands

Tel: +31 20 4440816

Email: h.holstege@amsterdamumc.nl

27 **Abstract**

28 The genetics underlying human longevity is influenced by the genetic risk to develop -or escape- age-related
29 diseases. As Alzheimer's disease (AD) represents one of the most common conditions at old age, an interplay
30 between genetic factors for AD and longevity is expected.

31 We explored this interplay by studying the prevalence of 38 AD-associated single-nucleotide-polymorphisms
32 (SNPs) identified in AD-GWAS, in self-reported cognitively healthy centenarians, and we replicated findings in the
33 largest GWAS on parental-longevity.

34 We found that 28/38 SNPs identified to associate with increased AD-risk also associated with decreased odds of
35 longevity. For each SNP, we express the imbalance between AD- and longevity-risk as an effect-size distribution.

36 When grouping the SNPs based on these distributions, we found three groups: 17 variants increased AD-risk more
37 than they decreased the risk of longevity (AD-group): these variants were functionally enriched for β -amyloid
38 metabolism and immune signaling, and they were enriched in microglia. 11 variants reported a larger effect on
39 longevity as compared to their AD-effect (Longevity-group): these variants were enriched for endocytosis/immune
40 signaling, and at the cell-type level were enriched in microglia and endothelial cells. Next to AD, these variants
41 were previously associated with other aging-related diseases, including cardiovascular and autoimmune diseases,
42 and cancer. Unexpectedly, 10 variants associated with an increased risk of both AD and longevity (Unexpected-
43 group). The effect of the SNPs in AD- and Longevity-groups replicated in the largest GWAS on parental-longevity,
44 while the effects on longevity of the SNPs in the Unexpected-group could not be replicated, suggesting that these
45 effects may not be robust across different studies.

46 Our study shows that some AD-associated variants negatively affect longevity primarily by their increased risk of
47 AD, while other variants negatively affect longevity through an increased risk of multiple age-related diseases,
48 including AD.

49 Introduction

50 The human lifespan is determined by a beneficial combination of environmental and genetic factors.^{1,2} Long-lived
51 individuals tend to cluster in families, suggesting that the role of the genetic factors is considerable,^{3,4} however,
52 the research of genetic variants that influence human lifespan has yielded contrasting results: only the longevity-
53 association of the *APOE* alleles and few additional variants consistently replicated across studies (*CDKN2B*, *ABO*).^{5,6}
54 While the replication rate in independent studies is low, a large collection of genetic variants has been associated
55 with longevity through genome-wide association studies (GWAS).^{5,6} The majority of these variants was previously
56 identified to associate with other age-related conditions, including cardiovascular disease, autoimmune and
57 neurological disorders, suggesting that the genetics underlying human longevity depends on a low risk for several
58 age-related diseases.^{2,5,6}

59 Of all age-related diseases, late-onset Alzheimer's Disease (AD) is the most common type of dementia and one of
60 the most prevalent causes of death at old age.⁷ The largest risk factor for AD is aging: at 100 years of age, the
61 disease's incidence is about 40% per year.⁸ Genetic factors play a significant role in AD as heritability was estimated
62 to be 60–80%:⁹ the strongest common genetic risk factor for AD is the *APOE-ε4* allele, and large collaborative
63 GWAS have identified ~40 additional common variants associated with a slight modification of the risk of AD.^{10–13}
64 Despite high incidence rates of AD at very old ages, AD is not an inevitable consequence of aging, as demonstrated
65 by individuals who surpass the age of 100 years with high levels of cognitive health.¹⁴

66 As AD-associated variants increase the risk of AD, leading to earlier death, a negative effect on longevity for these
67 variants is to be expected. However, apart from *APOE* alleles, genetic variants that influence the risk of AD were
68 not found to affect the human lifespan in previous GWAS. When assuming that AD-associated variants affect AD
69 only and that these variants' effect is constant during aging, then the effect on longevity for these variants should
70 be proportional to their effect on AD, albeit in a different direction. In other words, if a variant increases the risk
71 of AD 2-fold, then carriers will have twice as much AD-related mortality as non-carriers, and as a consequence
72 (given the assumptions), they will have twice as little chance to age into a cognitively healthy centenarian. For
73 example, variant *rs72824905* (Pro522Arg) in the *PLCG2* gene was recently found to decrease the risk of AD 1.75-fold,
74 1.63-fold frontotemporal dementia, and 1.85-fold dementia with Lewy bodies, while being associated with a 1.49-
75 fold increased likelihood of longevity.¹⁵ For a variant that is protective against multiple conditions, it might be
76 expected that the overall effect on longevity should be larger than the inverse of the effect on AD alone.

77 We have previously shown that cognitively healthy centenarians are depleted with genetic variants that increased
78 the risk of AD compared to a general population; however, the extent of depletion was variant specific, suggesting
79 that a subset of AD-variants may be specifically beneficial to reach extremely old ages in good cognitive health.^{16,17}
80 There is, however, little evidence of an age-dependent effect for AD variants and the extent to which these variants
81 affect other age-related diseases is mostly unknown.¹⁸ Using the assumption of effect-size proportionality, we set
82 out to investigate the relationship between AD- and longevity- risk for genetic variants associated with AD.

83

84 **Methods**

85 **Populations and selection of genetic variants**

86 We included N=358 centenarians from the 100-plus Study cohort, which comprises Dutch-speaking individuals
87 aged 100 years or older who self-report to be cognitively healthy, which is confirmed by a proxy.¹⁴ As population
88 controls, we used population-matched, cognitively healthy individuals from five studies: (i) the Longitudinal Aging
89 Study of Amsterdam (LASA, N=1,779),^{19,20} (ii) the memory clinic of the Alzheimer center Amsterdam and SCIENCE
90 project (N=1,206),^{21,22} (iii) the Netherlands Brain Bank (N=40),²³ (iv) the twin study of Amsterdam (N=201)²⁴ and (v)
91 the 100-plus Study (partners of centenarian's children, N=86).¹⁴ See *Supplementary Methods: Populations* for a
92 detailed description of these cohorts. Throughout the manuscript, we will refer to the union of the individuals from
93 these five studies as population subjects. The Medical Ethics Committee of the Amsterdam UMC (METC) approved
94 all studies. All participants and/or their legal representatives provided written informed consent for participation
95 in clinical and genetic studies.

96 Genetic variants in our populations were determined by standard genotyping and imputation methods. After
97 establishing quality control of the genetic data (see *Supplementary Methods: Quality control*), 2,905 population
98 subjects and 343 cognitively healthy centenarians were left for the analyses (*Table 1*). We then selected 41 variants
99 representing the current genetic landscape of AD (*Table S1*).¹³ We restricted our analysis to high-quality variants
100 with a minor allele frequency >1% in our cohorts, which led to the exclusion of 3/41 variants (rare variants in the
101 *TREM2* gene *rs143332484* and *rs75932628* and *AB13* gene *rs616338*), leaving 38 variants for the analyses.

102

103 **AD and longevity variant effect sizes**

104 We first retrieved the effect-size on AD (E_{AD}^k) for each AD variant, k , from a large genome-wide meta-analysis of
105 AD.¹³ To estimate a confidence interval, we bootstrapped the published effect-sizes (log of odds ratios) and their
106 respective standard errors ($B=10,000$).

107 To calculate the effect-size on longevity (E_{LGV}^k) for the same variants, we used a logistic regression model with
108 cognitive healthy centenarians as cases and population subjects as controls while adjusting for population
109 stratification (PC 1-5). The number of principal components to include as covariates was arbitrarily chosen;
110 however, as all individuals were population-matched, we expected these components to correct all major
111 population effects. The resulting p -values were corrected for multiple testing (False Discovery Rate, FDR). To
112 calculate the confidence interval, we repeated this procedure for bootstraps ($B=10,000$) of the data. For
113 convenience, variant effect-sizes on AD and longevity were calculated with respect to the allele that increases the
114 risk of AD, such that $E_{AD}^k > 0$.

115 Given a variant k , with a relative effect-size on AD (E_{AD}^k) and on longevity (E_{LGV}^k), we defined that the variant has
116 an *expected direction* if the variant increases the risk of AD, *i.e.* $E_{AD}^k > 0$, and at the same time decreases the risk of
117 longevity, *i.e.* $E_{LGV}^k < 0$. Inversely, we define that the longevity effect has an *unexpected direction* if the allele that
118 increased AD risk also increased the risk of longevity, *i.e.* $E_{AD}^k > 0$ and $E_{LGV}^k > 0$. The probability of observing an
119 *expected direction* was considered a Bernoulli variable with $p=0.5$ (*i.e.* equal chance of having an
120 *expected/unexpected direction*), thus the number of variants with an *expected direction* follows a binomial
121 distribution.

122

123 **Imbalance of variant effect direction**

124 We represented each variant as a data point whose coordinates were defined by the variant's effect on AD (E_{AD}^k ,
125 on the y-axis) and its effect on longevity (E_{LGV}^k , on the x-axis). See *Figure S1* for an example. For each variant, we
126 then calculated the normalized angle, α_k , of the vector representing the data point with the x-axis: $\alpha_k =$
127 $\frac{\text{atan2}(E_{AD}^k, E_{LGV}^k)}{\pi/2} + 1$, with $\alpha_k \in [-1; 1]$. This normalized angle relates to the imbalance between the risk of AD and
128 the risk of longevity. That is, for $\alpha_k < 0$ the variant has an *expected direction*, while for $\alpha_k > 0$ the variant has an
129 *unexpected direction*.

130 As the effect-sizes are sample estimates, we subsequently took their confidence interval into account to create,
131 for each variant, a distribution of the imbalance in the effect direction (*IED*). Hereto, we assumed a Gaussian

132 density for both E_{AD}^k and E_{LGV}^k , centered around \bar{E}_{AD}^k and \bar{E}_{LGV}^k and with a variance equal to the estimated
133 confidence interval for both effect sizes, respectively. We sampled 10,000 times from these distributions and
134 calculated the corresponding imbalance (α_k), to get a (non-Gaussian) distribution of the IED for that variant, IED_k .
135 To group variants with similar patterns of their IED distributions, we ordered the IED by their median value \widetilde{IED}_k ,
136 and defined a group of variants in which the effect sizes were in the expected direction ($\widetilde{IED}_k \leq 0$), which we
137 subsequently split in those that have (i) a larger effect on longevity as compared to the effect on AD ($\widetilde{IED}_k \leq -1/2$,
138 *Longevity-group*), and those that have (ii) a larger effect on AD as compared to the effect on longevity
139 ($-1/2 < \widetilde{IED}_k \leq 0$, *AD-group*). We defined a third group of variants that have an effect in the unexpected direction
140 ($\widetilde{IED}_k > 0$, *Unexpected-group*).

141

142 **Linking variants with functional clusters**

143 To investigate each variant's functional consequences, we calculated the variant-pathway mapping, which
144 indicates the degree of involvement of each genetic variant in AD-associated pathways (*Figure S2*). See
145 *Supplementary Methods: variant-pathway mapping* for a detailed explanation of our approach. Briefly, the *variant-*
146 *pathway mapping* depends on (i) the number of genes each variant was associated with and (ii) the biological
147 pathways each gene was associated with. We calculated the variant-pathway mapping for all 38 AD-associated
148 variants. Finally, we compared the variant-pathway mapping within each group of variants defined based on the
149 $IEDs$ (Longevity-, AD- and Unexpected-groups) using Wilcoxon sum rank tests and correcting p -values using FDR:
150 this was indicative of whether a group of variants was enriched for a specific functional cluster (*Figure S2*).

151

152 **Cell-type annotation at the level of each cluster**

153 To further explore the biological basis of the different groups of variants (Longevity-, AD- and Unexpected-groups),
154 we calculated the degree of enrichment of each group for specific brain cell-types (see *Supplementary Methods:*
155 *cell-type annotation* for a detailed description). This annotation depends on the number of genes each variant was
156 associated with, and the expression of these genes in the different brain cell-types, *i.e.* astrocytes, oligodendrocytes,
157 microglia, endothelial cells, and neurons. We finally compared the cell-specific annotations within each group of
158 variants (Longevity-, AD- and Unexpected-groups) using Wilcoxon sum rank tests and correcting p -values using
159 FDR, which indicated whether a group of variants was enriched for specific brain cell-types (*Figure S2*).

160

161 **Replication of findings in large GWAS cohorts**

162 To find additional evidence for our findings, we inspected the association statistics of the 38 AD-associated
163 variants in the largest GWAS on parental longevity.⁶ Briefly, in this study offspring's genotypes were used to model
164 parental age at death. In this dataset, we looked at the significance of association with longevity for the 38 variants
165 (*p-values* were corrected with FDR) and their direction of effect. Finally, we tested the consistency in the
166 expected/unexpected directions between our study and the GWAS on parental longevity using binomial tests.

167

168 **Implementation**

169 Quality control of genotype data, population stratification analysis and relatedness analysis were performed with
170 PLINK (v2.0 and v1.9). All subsequent analyses were performed with R (v3.6.3), Bash, and Python (v3.6) scripts. All
171 scripts are freely available at https://github.com/TesiNicco/Disentangle_AD_Age. Variant-gene annotation and
172 gene-set enrichment analyses are implemented in a stand-alone package available at
173 <https://github.com/TesiNicco/AnnotateMe>. Annotation and gene-set enrichment analysis of SNP-sets can also be
174 run on our web-server at <https://snpxplorer.eu.ngrok.io>.

175

176 **Results**

177 **AD-associated variants also associate with longevity**

178 We explored the association with longevity of 38 genetic variants previously associated with AD in GWAS (*Table*
179 *S1*). We tested these variants in 343 centenarians who self-reported to be cognitively healthy (mean age at inclusion
180 101.4±1.3, 74.7% females), as opposed to 2,905 population subjects (mean age at inclusion 68.3±11.5, 50.7%
181 females). We found a significant association with longevity for two variants after multiple testing correction
182 (FDR<5%, variants in the *APOE* gene; *rs429358* and *rs7412*, *Table S2*). We compared the direction of effect on
183 longevity with that on AD as found in literature: of the 38 variants, 28 showed an association in the expected
184 direction, *i.e.* alleles that increased AD risk were associated with lower odds of longevity, and this was significantly
185 more than expected by chance ($p=0.005$ including *APOE* variants, $p=0.01$ excluding *APOE* variants, see *Methods*).

186

187 **Distributions of the imbalance in the effect direction (IED)**

188 To study the relationship between the effect on AD and longevity for all 38 AD-associated variants in more detail,
189 we created distributions of the imbalance in the variant effect direction (*IED*): *Figure 1*. The *IED* of a variant

190 indicates whether the effects on AD and longevity are in the expected direction (values<0) or in the unexpected
191 direction (values>0). A variant with an *IED* value approaching 0 has a pure AD effect (AD-end), while a variant
192 with values *IED* value close to -1 has a pure longevity effect (Longevity-end); see *Methods* for a detailed
193 explanation. Based on the median value of each *IED* distributions, \widetilde{IED}_k , we grouped the variants into (i) a
194 Longevity-group (variants with a \widetilde{IED}_k skewed towards the longevity-end of the spectrum), (ii) an AD-group
195 (variants with a \widetilde{IED}_k skewed towards the AD-end of the spectrum), and (iii) an Unexpected-group (variants with
196 a \widetilde{IED}_k in the unexpected direction). The AD-group included 17 variants (*APOE* (1), *APOE* (2), *SCIMP*, *PLCG2* (1),
197 *MS4A6A*, *BIN1*, *PILRA*, *APP*, *PLCG2* (2), *CR1*, *SLC24A4*, *TREML2*, *ACE*, *APH1B*, *FERMT2*, *PICALM*, *CD33*) and the
198 longevity-group included 11 variants (*SHARPIN* (1), *SHARPIN* (2), *HS3ST1*, *EPAH1*, *IQCK*, *PRKD3*, *CD2AP*, *PLCG2* (3),
199 *SPI1*, *HLA*, *EDHDC3*), such that the effect of 28/38 (74%) of all variants was in the expected direction. The effect of
200 10 variants was in the unexpected direction, the Unexpected-group: (*PTK2B*, *CLU*, *KANSL1*, *INPP5D*, *ABCA7*, *CHRNE*,
201 *SORL1*, *IL34*, *ADAM10*, *CASS4*) (*Figure 1*).

202

203 **AD-associated variants in large GWAS of longevity**

204 To find additional evidence for longevity associations, we inspected the AD-associated variants' effect in the
205 largest GWAS on parental longevity.⁶ Of the 38 AD-associated variants, association statistics were available for 34
206 of the variants (missing from longevity-group: *PLCG2* (3), *SPI1*; missing from Unexpected-group: *KANSL1*, *INPP5D*).
207 Overall, 21/26 (81%) of the variants in the expected direction in our study (of which 6/9 variants in Longevity- and
208 15/17 variants in the AD-group), were also in the expected direction in the independent parental longevity dataset.
209 Variants in the expected direction in the first analysis are significantly more likely to be in the expected direction
210 in the replication analysis than in the unexpected direction ($p=0.01$, based on a binomial test, *Figure 3*). Six AD-
211 associated variants reached significance in the parental-longevity GWAS after correcting for multiple comparisons
212 (FDR<5%): variants in the *APOE* gene (*rs429358* and *rs7412*) and variants in/near *PRKD3* (*rs8764613*), *CD2AP*
213 (*rs9381564*), *APH1B* (*rs117618017*) and *BIN1* (*rs6733839*). Of these, variants in/near *PRKD3* and *CD2AP* belonged to the
214 Longevity-group in our analysis.

215 Conversely, only 2/8 (25%) variants that we observed in the unexpected direction in our study were also in the
216 unexpected direction in the parental-longevity GWAS, such that these variants were *not* more likely to be in the
217 unexpected direction ($p=0.29$, based on a binomial test, *Figure 3*).

218

219 **Functional characterization of variants**

220 The 38 AD-associated variants included coding variants (N=10), intronic variants (N=20), and intergenic variants
221 (N=8) (Table S3). 12/28 of the intronic/intergenic variants had eQTL associations. In total, the 38 variants mapped
222 to 68 unique genes, with most variants mapping to one gene (N=21) and fewer mapping to 2 genes (N=10), 3 genes
223 (N=2), 4 genes (N=1), 5 genes (N=2), 6 and 7 genes (N=1, respectively) (Figure S3 and Table S3).

224 We performed gene-set enrichment analysis using a sampling-based approach to explore the biological processes
225 enriched in the 68 genes associated with AD-variants (see Methods and Figure S2). We found 115 significantly
226 enriched biological processes after correction for multiple tests (FDR<5%, Table S4). After clustering these terms
227 based on their semantic similarity, we found four main clusters of biological processes: (i) β -amyloid metabolism,
228 (ii) lipid/cholesterol metabolism, (iii) endocytosis/immune signaling and (iv) synaptic plasticity (Figure 1, Figure S4
229 and Table S5).

230 Next, we calculated the *variant-pathway mapping score* (see Methods and Figure S2), which indicates how well a
231 variant is associated with each of the 4 functional clusters. In total, we calculated the *variant-pathway mapping*
232 for 30 variants; we imputed the annotation of 6 variants (Table S5 and Table S5), while 2 variants could not be
233 annotated (variants *rs7185636* and *rs1582763* in/near *IQCK* and *MS4A6A* genes), because the associated genes were
234 not annotated with any biological process function (Table S5). Finally, we tested whether the Longevity-, AD- and
235 Unexpected-groups were enriched for specific functional clusters by comparing the distribution of variant-
236 pathway mapping within each group (see Methods, Figure 2, and Figure S2). The Longevity-group was significantly
237 (FDR<10%) enriched for the endocytosis/immune signaling functional cluster; the AD-group for the
238 endocytosis/immune signaling, β -amyloid metabolism and to a smaller extent for the synaptic plasticity
239 functional clusters; the Unexpected-group was mainly enriched for the endocytosis and β -amyloid metabolism
240 functional clusters.

241

242 **Expression of AD-associated genes in brain cell-types**

243 We explored whether specific brain cell types, *i.e.* astrocytes, oligodendrocytes, microglia, endothelial cells and
244 neurons, were enriched within each group of variants (see Methods, Table S5, and Table S6). Figure 1 shows the
245 collapsed cell-type expression for all 38 AD-associated variants. We then tested the enrichment for cell-type
246 expression within the Longevity-, AD- and Unexpected-groups. The Longevity-group was significantly enriched for

247 myeloid and endothelial cells, the *AD*-group for myeloid cells, while the Unexpected-group was significantly
248 enriched for endothelial cells (FDR<10%).

249

250 **Discussion**

251 **Summary of the findings**

252 We studied the effect on longevity of 38 genetic variants previously associated with AD through GWAS.¹³ We found
253 that a majority of 74% of the alleles that increase the risk of AD is associated with lower odds of becoming a
254 centenarian (expected direction). Overall, most variants (N=17) had a larger effect on AD than on longevity: these
255 variants were associated with β -amyloid metabolism and endocytosis/immune signaling, and were primarily
256 expressed in microglia. A subset of variants (N=11) had a larger effect on longevity than their effect on AD. These
257 variants were associated mostly with endocytosis and immune signaling, and they were expressed in microglia and
258 endothelial cells. These variant-effects were confirmed for 81% of the alleles in an independent dataset, the largest
259 GWAS on parental longevity. In contrast, 26% of the variants increased both the risk of developing AD *and* the risk
260 of becoming a centenarian (n=10), (unexpected direction). These unexpected effects could only be replicated for 2
261 of the variants in the independent dataset, suggesting that the expected effects were more robust across studies
262 than the unexpected effects. Together, our findings suggest that variants associated with AD-risk may also be
263 linked with the risk of other age-related diseases, and that survival/longevity is affected by these variants.

264

265 **AD-associated variants and their effect on healthy aging**

266 A single study previously explored the extent to which 10 AD-associated variants affect longevity: apart from *APOE*
267 locus, none of the other 10 tested AD-associated variants significantly associated with longevity.²⁵ In addition to
268 *APOE*, four variants showed a negative effect on longevity while increasing AD-risk (in/near *ABCA7*, *EPHA1*, *CD2AP*,
269 and *CLU*). In agreement with these findings, we also found that only the *APOE* variants significantly associated
270 with longevity, and variants in/near *EPHA1* and *CD2AP* belong to the Longevity-group. However, in our study, we
271 found that most alleles associated with an increased risk of AD associated with a decreased chance of longevity.
272 The inability to observe such an inverse relationship between variant effects on AD and longevity in the previous
273 study may be explained by the relatively small sample sizes, combined with a low number of (well-established) AD
274 variants analyzed (N=10). In our study, groups sizes were also relatively small, but the centenarians had a relatively

275 high level of cognitive health, which might have contributed to an increased effect size of AD-associated genetic
276 variants in our comparison.^{16,17}

277

278 **Different trajectories of effect of AD-associated variants on healthy aging**

279 *Variants with a larger effect on AD than longevity:*

280 For most variants with effects in the expected direction, the risk-increasing effect on AD was more extensive than
281 the negative effect on survival/longevity. These variants, which include both *APOE* alleles, might negatively affect
282 lifespan because carriers are removed from the population with increasing age due to AD-associated mortality. For
283 the *APOE* variants specifically, the distribution of the imbalance in the effect directions suggests a nearly similar
284 proportion of the increased risk of AD and decreased risk of longevity for both *APOE* variants ($\widetilde{IED}_k \approx -1/2$). This
285 explains why multiple previous studies have associated *APOE* variants with longevity. In our cohort of centenarians,
286 the frequency of the deleterious $\epsilon 4$ allele is half of that of the population controls (8% vs. 16%, respectively). In
287 comparison, the frequency of the protective $\epsilon 2$ allele is nearly two-fold increased (16% vs. 9%).¹⁶ Note, however,
288 that inclusion criteria of the centenarian cohort required them to self-report to be cognitively healthy, which might
289 have increased the observed longevity effect. Apart from the *APOE* variants, the AD-group included 15 variants, all
290 of which were among the first to be associated with AD through GWAS (*CR1*, *CD33*, *BIN1*, *MS4A6A*, *PICALM*, and
291 *SLC24A4*).^{26,27} These common variants have the strongest associations with AD: largest odds ratios (OR) leading to
292 lowest *p*-values. Functional annotation showed significant enrichment of β -amyloid metabolism, which aligns
293 with the importance of functional APP metabolism in maintaining brain health. We also observed functional
294 enrichment of endocytosis and immune signaling, and a specific cell-type enrichment for microglia. This is in line
295 with the currently growing hypothesis of the involvement of immune dysfunction in the etiology underlying
296 AD.^{28,29}

297

298 *Variants with a larger effect on longevity than AD*

299 The second-largest group of variants constituted a subset of 11 variants with a larger effect on longevity than the
300 effect on AD, which suggests that these variants may be involved in other age-related diseases or general age-
301 related processes. The AD-association of most of these variants is relatively recent, likely due to small effect sizes
302 (ORs) or variants rareness (low minor allele frequency, MAF); both features necessitate a very large number of
303 samples to identify these variants as significantly associating with the disease. The variants within this group were

304 specifically enriched for immune response and endocytosis, which are known hallmarks of longevity.^{1,30,31} In
305 addition to the rare non-synonymous variant in the *PLCG2* gene (*rs72824905*, MAF: 0.6%), which was recently
306 observed to be protective against AD, frontotemporal dementia (FTD) and dementia with Lewy bodies, other
307 variants within this group were previously linked with disease risk factors. One of the two non-synonymous
308 variants in the *SHARPIN* gene, variant *rs34173062* (MAF: 5.7%), has been associated with respiratory system
309 diseases in GWAS.³²⁻³⁴ Variant *rs7185636* (MAF: 17.1%), intronic of the *IQCK* gene, is in complete linkage with a
310 variant (*rs7191155*, $R^2=0.95$), which was previously associated with body-mass index (BMI).³⁵ The variant *rs876461*
311 (MAF: 13.0%) near the *PRKD3* gene is in linkage with variant *rs13420463* ($R^2=0.42$), which has been associated with
312 systolic blood pressure.³⁶ Further, the variant near *CD2AP* gene associates with the development and maintenance
313 of the blood-brain barrier, a specialized vascular structure of the central nervous system which, when disrupted,
314 has been linked with epilepsy, stroke and AD.³⁷ Variant *rs9275152* (MAF: 10.4%) maps to the complex Human-
315 Leukocyte-Antigen (HLA) region, which codes for cell-surface proteins responsible for the regulation of the
316 adaptive immune system. In numerous GWAS, variants in the HLA region were associated with autoimmune
317 diseases, cancer, and longevity.^{6,38} The AD-associated variant in this region (*rs9275152*) is also a risk variant for
318 Parkinson's disease.³⁹ Finally, the genomic region surrounding the *SPI1* gene (in which variant *rs3740688* maps) has
319 been previously associated with cognitive traits (intelligence, depression)^{40(p300)} and, with lower evidence, with
320 kidney disease and cancer.^{41,42} The remaining variants *rs56402156*, *rs7920721*, and *rs4351014* (in/near *EPHA1*, *ECHDC3*,
321 and *HS3ST1*) have not been directly associated with other traits, although their associated genes were implicated
322 in systemic lupus erythematosus (*HS3ST1*) and cancer (*EPHA1*, *ECHDC3*).^{43,44(p1),45} Together, these findings suggest
323 that the counterpart of each risk-increasing allele, the AD-protective alleles, might give a survival advantage that
324 is not only specific to AD. Their functional and cell-type annotations suggest that they contribute to the
325 maintenance of regulatory stimuli in the immune and endosomal systems, which may be essential to maintain
326 brain and overall physical health, necessary to reach extremely old ages in good cognitive health.¹⁷

327

328 **Variants associated with increased risk of AD and increased longevity risk: unexpected group.**

329 Unexpectedly, ten variants increased the risk of AD while at the same time increasing the chance to reach ages
330 over 100 in good cognitive health, which is an *unexpected* balance. We note that the *IED* distributions of these
331 variants were broad, and in some cases even showed a bimodal behavior (in/near *KANSL1*, *IL34*, *CHRNE*): this is
332 attributable to the small effect-sizes (and large standard errors) on longevity for these variants, which caused data

333 points to easily flip between the expected and unexpected direction during the sampling procedure. Replication of
334 the direction of the variant effect in an independent dataset of parental longevity indicated that the unexpected
335 direction was replicated in only the *CLU* and *CHRNE* variants, suggesting that future studies will have to further
336 explore (the robustness of) these unexpected effects.

337 One explanation for such counter-intuitive effects may be a variant interaction with other variants, which was
338 shown for the variant in the *KANSL1* and *CLU* gene with respect to the *APOE* genotype.⁴⁶ Therefore, carrying the
339 risk allele of such variants may specifically affect the risk of AD in *APOE* $\epsilon 4$ allele carriers, which are not prevalent
340 among cognitively healthy centenarians.

341 An alternative explanation may be that these variants have age-dependent effects: for example, high blood
342 pressure at midlife increases the risk of AD, but after the age of 85 a high blood pressure protects against AD.⁴⁷
343 Similarly, a high body-mass-index (BMI) increases the risk of AD at midlife, while being protective at older ages.⁴⁸
344 In line with this hypothesis, the AD variant in/near *IL34* gene codes for a cytokine that is crucial for the
345 differentiation and the maintenance of microglia.⁴⁹ Although further studies are needed, an excessive
346 differentiation in middle-age individuals may increase brain-related inflammation and AD-risk, while it might
347 compensate for the slower differentiation and immune activity at very old ages. Indeed, next to *IL34*, several genes
348 that may be affected by these Unexpected-variants, such as *PTK2B* and *INPP5D*, play a role in aging-associated
349 processes, such as cellular senescence or immunity.^{50,51}

350

351 **Strengths and weaknesses**

352 We acknowledge that our findings are based on relatively small sample sizes, especially for the cognitively healthy
353 centenarian group. This phenotype is rare, and individuals need to be individually approached for study inclusion,¹⁴
354 which is prohibitive for large sample collection. As population subjects in our comparison, we used individuals from
355 five different cohorts: all from the same (Dutch) population, all tested cognitively intact, and did not convert to
356 dementia at the time of analyses. Analyses of variants with low sample sizes in small samples lead to effect sizes
357 large confidence intervals: we took this uncertainty into account by bootstrapping effect sizes, causing the
358 distributions of the imbalance in the variant effect direction of several variants to be widely spread. Although our
359 work represents a first step towards understanding the effect of AD-associated variants on longevity, similar
360 analyses in larger oldest-old or centenarian samples are necessary to support our findings further. Secondly, we
361 had to deal with the problem in GWAS studies, that the driving effect underlying the AD-association of each

362 variant is unclear. Several genes usually map to a specific GWAS locus, represented by one variant. Therefore, to
363 accommodate this uncertainty, we allowed multiple genes mapping to the GWAS locus to be associated with each
364 variant. The subsequent functional annotation of these genes is mainly dependent on the current (limited)
365 knowledge about variant effect and genes function. It is thus likely that our variant-function annotation will
366 change as we gain more understanding about these variant-gene-effects, as well as annotations of gene-functions.
367 When we inspected the parental-longevity GWAS, most of the variants that were in the *expected direction* in our
368 study were also in the same direction in the GWAS; however, this was not true for all variants. The variant that
369 deviated the most between our study and the parental-longevity GWAS was *rs9275152* in the *HLA* region: while we
370 clustered this variant in the Longevity-group, in the parental-longevity GWAS the direction of effect was opposite
371 (*i.e. unexpected*), suggesting that the variant increased the risk of AD and at the same time the chance of a long
372 lifespan.⁶ The genomic region to which *HLA* maps is biologically known to be affected by many recombination
373 events and may be population- and environment-dependent, which may explain this divergence.⁵² In addition to
374 *HLA*-variant, variant *rs34674752* in the *SHARPIN* gene reported the second-largest effect-size in our study (after
375 *APOE-ε4*), while the effect-size of this variant in the GWAS was very small, yet in the *expected direction*. To this
376 end, we note that the individuals used in the parental-longevity GWAS were themselves not extremely old
377 individuals, such that possible pleiotropic effects at very old ages, as described earlier, may not be observable in
378 this GWAS. However, while we observed overall consistency in effect-size direction for variants in the expected
379 direction, 6/8 of the variants in the unexpected direction were in the expected direction in the GWAS, with variants
380 near *SORL1*, *IL34*, and *ADAM10* having the most noticeable differences. We speculate that the relatively young ages
381 of the GWAS samples, together with the small sample size of our centenarian cohort may be the cause of such
382 discrepancy.

383

384 **Conclusions**

385 Each AD-associated variant has a different effect on longevity. Variants that have a larger effect on longevity than
386 on AD were previously associated as disease risk-factors, and associated genes are selectively enriched for
387 endocytosis and immune signaling functions.

388

389 **Acknowledgments:** The following studies and consortia have contributed to this manuscript. Amsterdam Dementia Cohort
390 (ADC): Research at the Alzheimer center Amsterdam is part of the neurodegeneration research program of Amsterdam

391 Neuroscience. 100-plus Study: we are grateful for the collaborative efforts of all participating centenarians and their family
392 members and/or relatives. Wiesje van der Flier holds the Pasma chair. Longitudinal Aging Study of Amsterdam (LASA): the
393 authors are grateful to all LASA participants, the fieldwork team and all researchers for their ongoing commitment to the study.
394 **Funding:** The Alzheimer center Amsterdam is supported by Stichting Alzheimer Nederland and Stichting VUmc fonds. The
395 clinical database structure was developed with funding from Stichting Dioraphte. The SCIENCE project is supported by a
396 research grant from Gieskes-Strijbis fonds and stichting Dioraphte. Genotyping of the Dutch case-control samples was
397 performed in the context of EADB (European Alzheimer DNA biobank), funded by the JPCo-fuND FP-829-029 (ZonMW
398 projectnumber 733051061). The 100-plus Study was supported by Stichting Alzheimer Nederland (WE09.2014-03), Stichting
399 Dioraphte, horstingstuit foundation, Memorabel (ZonMW projectnumber 733050814), and Stichting VUmc fonds. Genotyping
400 of the 100-plus study was performed in the context of EADB (European Alzheimer DNA biobank) funded by the JPCo-fuND FP-
401 829-029 (ZonMW projectnumber 733051061). Longitudinal Aging Study Amsterdam (LASA) is largely supported by a grant from
402 the Netherlands Ministry of Health, Welfare and Sports, Directorate of Long-Term Care.

403 **Author contributions:** Conceptualization, N.T., M.Hul., S.L., M.R. and H.H.; Data curation, S.L., M.Hul., I.J., N.S., N.Sc., P.S., W.F.,
404 M.Hui. and H.H.; Formal analysis, N.T.; Funding acquisition, P.S., W.F., M.Hui., M.R. and H.H.; Methodology, N.T., M.Hul., S.L.,
405 M.R. and H.H.; Project administration, P.S., W.F., M.Hui., M.R. and H.H.; Software, N.T., S.L., M.Hul., M.R.; Supervision, N.T., S.L.,
406 M.Hul., M.R. and H.H.; Visualization, N.T.; Writing – original draft, N.T., S.L., M.Hul., M.R. and H.H.; Writing – review & editing,
407 N.T., S.L., M.Hul., I.J., N.S., N.Sc., P.S., W.F., M.Hui., M.R. and H.H.

408 **Conflicts of Interest:** All the authors in the study declared no conflict of interest. The funders had no role in the study's design
409 at any stage.

410 **Data availability:** data that support the findings of this study are available on request, if reasonable, from the corresponding
411 author. The data are not publicly available due to privacy or ethical restrictions.
412

413 References

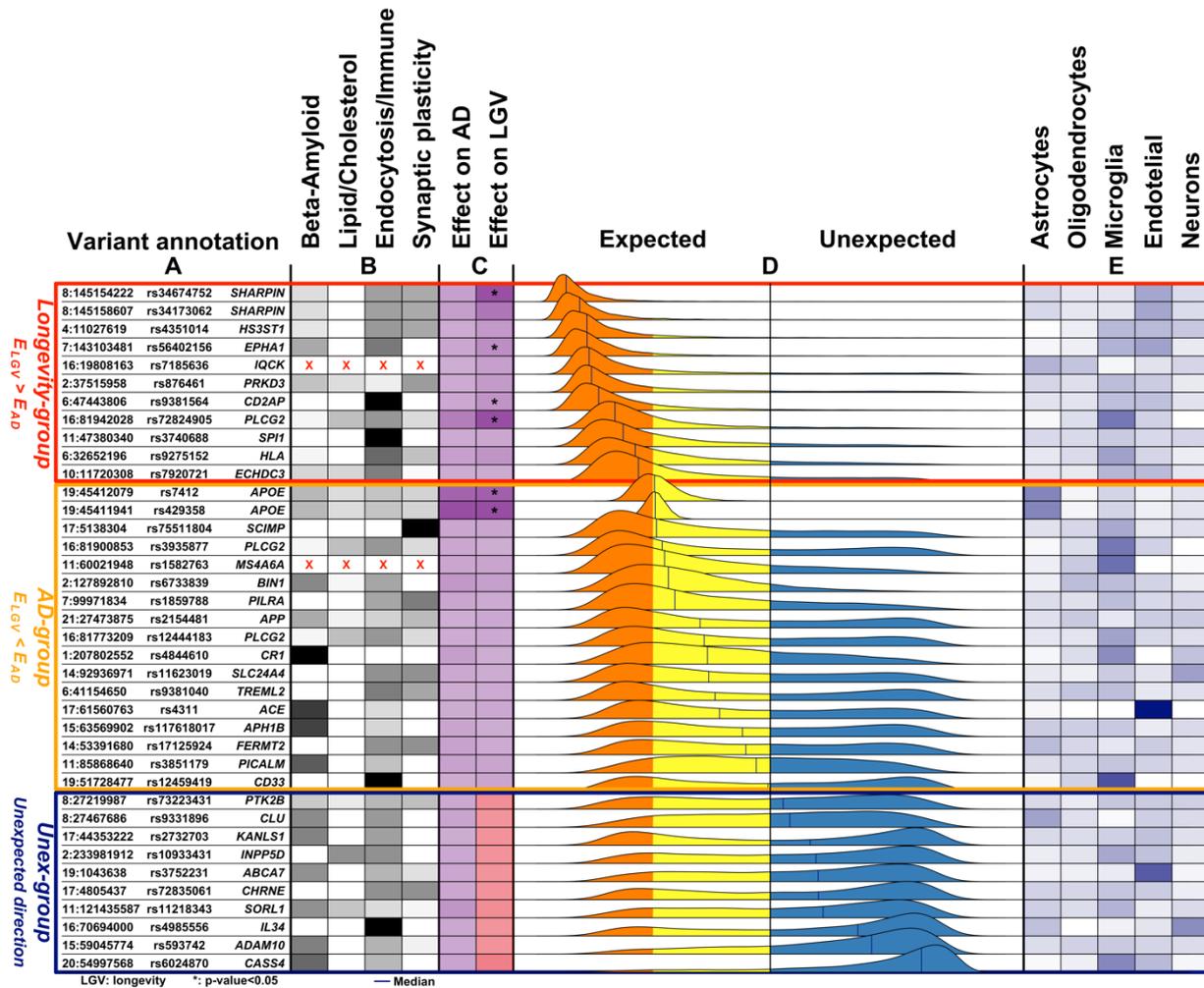
- 414 1. Partridge L, Deelen J, Slagboom PE. Facing up to the global challenges of ageing. *Nature*. 2018;561(7721):45-56. doi:10.1038/s41586-
415 018-0457-8
- 416 2. Melzer D, Pilling LC, Ferrucci L. The genetics of human ageing. *Nat Rev Genet*. Published online November 2019. doi:10.1038/s41576-
417 019-0183-6
- 418 3. Perls TT, Wilmoth J, Levenson R, et al. Life-long sustained mortality advantage of siblings of centenarians. *Proc Natl Acad Sci U S A*.
419 2002;99(12):8442-8447. doi:10.1073/pnas.122587599
- 420 4. Caselli G, Pozzi L, Vaupel JW, et al. Family clustering in Sardinian longevity: a genealogical approach. *Exp Gerontol*. 2006;41(8):727-
421 736. doi:10.1016/j.exger.2006.05.009
- 422 5. Deelen J, Evans DS, Arking DE, et al. A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nat*
423 *Commun*. 2019;10(1). doi:10.1038/s41467-019-11558-2
- 424 6. Timmers PR, Mounier N, Lall K, et al. Genomics of 1 million parent lifespans implicates novel pathways and common diseases and
425 distinguishes survival chances. *eLife*. 2019;8. doi:10.7554/eLife.39856
- 426 7. 2012 Alzheimer's disease facts and figures. *Alzheimers Dement*. 2012;8(2):131-168. doi:10.1016/j.jalz.2012.02.001
- 427 8. Corrada MM, Brookmeyer R, Paganini-Hill A, Berlau D, Kawas CH. Dementia incidence continues to increase with age in the oldest
428 old: The 90+ study. *Ann Neurol*. 2010;67(1):114-121. doi:10.1002/ana.21915
- 429 9. Gatz M, Reynolds CA, Fratiglioni L, et al. Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry*.
430 2006;63(2):168-174. doi:10.1001/archpsyc.63.2.168
- 431 10. Alzheimer Disease Genetics Consortium (ADGC), The European Alzheimer's Disease Initiative (EADI), Cohorts for Heart and Aging
432 Research in Genomic Epidemiology Consortium (CHARGE), et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk
433 loci and implicates A β , tau, immunity and lipid processing. *Nat Genet*. 2019;51(3):414-430. doi:10.1038/s41588-019-0358-2
- 434 11. Jansen IE, Savage JE, Watanabe K, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing
435 Alzheimer's disease risk. *Nat Genet*. 2019;51(3):404-413. doi:10.1038/s41588-018-0311-9
- 436 12. Sims R, van der Lee SJ, Naj AC, et al. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate
437 immunity in Alzheimer's disease. *Nat Genet*. 2017;49(9):1373-1384. doi:10.1038/ng.3916
- 438 13. de Rojas I, Moreno-Grau S, Tesi N, et al. *Common Variants in Alzheimer's Disease: Novel Association of Six Genetic Variants with AD*
439 *and Risk Stratification by Polygenic Risk Scores*. Genetic and Genomic Medicine; 2019. doi:10.1101/19012021
- 440 14. Holstege H, Beker N, Dijkstra T, et al. The 100-plus Study of cognitively healthy centenarians: rationale, design and cohort
441 description. *Eur J Epidemiol*. Published online October 25, 2018. doi:10.1007/s10654-018-0451-3
- 442 15. DESGESCO (Dementia Genetics Spanish Consortium), EADB (Alzheimer Disease European DNA biobank), EADB (Alzheimer Disease
443 European DNA biobank), IFGC (International FTD-Genomics Consortium), IPDGC (The International Parkinson Disease Genomics
444 Consortium), et al. A nonsynonymous mutation in PLCG2 reduces the risk of Alzheimer's disease, dementia with Lewy bodies and
445 frontotemporal dementia, and increases the likelihood of longevity. *Acta Neuropathol (Berl)*. Published online May 27, 2019.
446 doi:10.1007/s00401-019-02026-8
- 447 16. Tesi N, van der Lee SJ, Hulsman M, et al. Centenarian controls increase variant effect sizes by an average twofold in an extreme
448 case-extreme control analysis of Alzheimer's disease. *Eur J Hum Genet*. Published online September 26, 2018. doi:10.1038/s41431-018-0273-5
- 449 17. Tesi N, van der Lee SJ, Hulsman M, et al. Immune response and endocytosis pathways are associated with the resilience against
450 Alzheimer's disease. *Transl Psychiatry*. 2020;10(1):332. doi:10.1038/s41398-020-01018-7
- 451 18. Ponomareva N, Andreeva T, Protasova M, et al. Age-dependent effect of Alzheimer's risk variant of CLU on EEG alpha rhythm in
452 non-demented adults. *Front Aging Neurosci*. 2013;5. doi:10.3389/fnagi.2013.00086
- 453 19. Hoogendijk EO, Deeg DJH, Poppelaars J, et al. The Longitudinal Aging Study Amsterdam: cohort update 2016 and major findings. *Eur*
454 *J Epidemiol*. 2016;31(9):927-945. doi:10.1007/s10654-016-0192-0
- 455 20. Huisman M, Poppelaars J, van der Horst M, et al. Cohort Profile: The Longitudinal Aging Study Amsterdam. *Int J Epidemiol*.
456 2011;40(4):868-876. doi:10.1093/ije/dyq219
- 457 21. Slot RER, Verfaillie SCJ, Overbeek JM, et al. Subjective Cognitive Impairment Cohort (SCIENCe): study design and first results.
458 *Alzheimers Res Ther*. 2018;10(1). doi:10.1186/s13195-018-0390-y
- 459 22. van der Flier WM, Scheltens P. Amsterdam Dementia Cohort: Performing Research to Optimize Care. Perry G, Avila J, Zhu X, eds. *J*
460 *Alzheimers Dis*. 2018;62(3):1091-1111. doi:10.3233/JAD-170850
- 461 23. Rademaker MC, de Lange GM, Palmen SJMC. The Netherlands Brain Bank for Psychiatry. In: *Handbook of Clinical Neurology*. Vol 150.
462 Elsevier; 2018:3-16. doi:10.1016/B978-0-444-63639-3.00001-3
- 463 24. Willemsen G, de Geus EJC, Bartels M, et al. The Netherlands Twin Register Biobank: A Resource for Genetic Epidemiological Studies.
464 *Twin Res Hum Genet*. 2010;13(3):231-245. doi:10.1375/twin.13.3.231
- 465 25. Shi H, Belbin O, Medway C, et al. Genetic variants influencing human aging from late-onset Alzheimer's disease (LOAD) genome-
466 wide association studies (GWAS). *Neurobiol Aging*. 2012;33(8):1849.e5-1849.e18. doi:10.1016/j.neurobiolaging.2012.02.014
- 467 26. Harold D, Abraham R, Hollingworth P, et al. Genome-wide association study identifies variants at CLU and PICALM associated with
468 Alzheimer's disease. *Nat Genet*. 2009;41(10):1088-1093. doi:10.1038/ng.440
- 469 27. Lambert J-C, Heath S, Even G, et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's
470 disease. *Nat Genet*. 2009;41(10):1094-1099. doi:10.1038/ng.439
- 471 28. Efthymiou AG, Goate AM. Late onset Alzheimer's disease genetics implicates microglial pathways in disease risk. *Mol*
472 *Neurodegener*. 2017;12(1). doi:10.1186/s13024-017-0184-x
- 473 29. Hansen DV, Hanson JE, Sheng M. Microglia in Alzheimer's disease. *J Cell Biol*. 2018;217(2):459-472. doi:10.1083/jcb.201709069
- 474 30. Sadighi Akha AA. Aging and the immune system: An overview. *J Immunol Methods*. 2018;463:21-26. doi:10.1016/j.jim.2018.08.005

- 475 31. Solé-Domènech S, Cruz DL, Capetillo-Zarate E, Maxfield FR. The endocytic pathway in microglia during health, aging and
476 Alzheimer's disease. *Ageing Res Rev.* 2016;32:89-103. doi:10.1016/j.arr.2016.07.002
- 477 32. Astle WJ, Elding H, Jiang T, et al. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease.
478 *Cell.* 2016;167(5):1415-1429.e19. doi:10.1016/j.cell.2016.10.042
- 479 33. Olafsdottir TA, Theodors F, Bjarnadottir K, et al. Eighty-eight variants highlight the role of T cell regulation and airway remodeling
480 in asthma pathogenesis. *Nat Commun.* 2020;11(1):393. doi:10.1038/s41467-019-14144-8
- 481 34. Kichaev G, Bhatia G, Loh P-R, et al. Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am J Hum Genet.*
482 2019;104(1):65-75. doi:10.1016/j.ajhg.2018.11.008
- 483 35. Hoffmann TJ, Choquet H, Yin J, et al. A Large Multiethnic Genome-Wide Association Study of Adult Body Mass Index Identifies
484 Novel Loci. *Genetics.* 2018;210(2):499-515. doi:10.1534/genetics.118.301479
- 485 36. Warren HR, Evangelou E, Cabrera CP, et al. Genome-wide association analysis identifies novel blood pressure loci and offers
486 biological insights into cardiovascular risk. *Nat Genet.* 2017;49(3):403-415. doi:10.1038/ng.3768
- 487 37. Cochran JN, Rush T, Buckingham SC, Roberson ED. The Alzheimer's disease risk factor CD2AP maintains blood-brain barrier
488 integrity. *Hum Mol Genet.* 2015;24(23):6667-6674. doi:10.1093/hmg/ddv371
- 489 38. Bodis G, Toth V, Schwarting A. Role of Human Leukocyte Antigens (HLA) in Autoimmune Diseases. *Rheumatol Ther.* 2018;5(1):5-20.
490 doi:10.1007/s40744-018-0100-z
- 491 39. Bandres-Ciga S, Ahmed S, Sabir MS, et al. The Genetic Architecture of Parkinson Disease in Spain: Characterizing Population-
492 Specific Risk, Differential Haplotype Structures, and Providing Etiologic Insight. *Mov Disord Off J Mov Disord Soc.* 2019;34(12):1851-1863.
493 doi:10.1002/mds.27864
- 494 40. Davies G, Lam M, Harris SE, et al. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive
495 function. *Nat Commun.* 2018;9(1). doi:10.1038/s41467-018-04362-x
- 496 41. Pattaro C, Teumer A, Gorski M, et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney
497 function. *Nat Commun.* 2016;7:10023. doi:10.1038/ncomms10023
- 498 42. Michailidou K, Lindström S, Dennis J, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature.* 2017;551(7678):92-
499 94. doi:10.1038/nature24284
- 500 43. Langefeld CD, Ainsworth HC, Cunningham Graham DS, et al. Transancestral mapping and genetic load in systemic lupus
501 erythematosus. *Nat Commun.* 2017;8:16021. doi:10.1038/ncomms16021
- 502 44. Herath NI, Doecke J, Spanevello MD, Leggett BA, Boyd AW. Epigenetic silencing of EphA1 expression in colorectal cancer is
503 correlated with poor survival. *Br J Cancer.* 2009;100(7):1095-1102. doi:10.1038/sj.bjc.6604970
- 504 45. Rafiq S, Khan S, Tapper W, et al. A genome wide meta-analysis study for identification of common variation associated with breast
505 cancer prognosis. *PLoS One.* 2014;9(12):e101488. doi:10.1371/journal.pone.0101488
- 506 46. Jun G, Ibrahim-Verbaas CA, Vronskaya M, et al. A novel Alzheimer disease locus located near the gene encoding tau protein. *Mol*
507 *Psychiatry.* 2016;21(1):108-117. doi:10.1038/mp.2015.23
- 508 47. McGrath ER, Beiser AS, DeCarli C, et al. Blood pressure from mid- to late life and risk of incident dementia. *Neurology.*
509 2017;89(24):2447-2454. doi:10.1212/WNL.0000000000004741
- 510 48. Xu WL, Atti AR, Gatz M, Pedersen NL, Johansson B, Fratiglioni L. Midlife overweight and obesity increase late-life dementia risk: a
511 population-based twin study. *Neurology.* 2011;76(18):1568-1574. doi:10.1212/WNL.0b013e3182190d09
- 512 49. Wang Y, Colonna M. Interleukin-34, a cytokine crucial for the differentiation and maintenance of tissue resident macrophages and
513 Langerhans cells: Highlights. *Eur J Immunol.* 2014;44(6):1575-1581. doi:10.1002/eji.201344365
- 514 50. Ryu SJ, Cho KA, Oh YS, Park SC. Role of Src-specific phosphorylation site on focal adhesion kinase for senescence-associated
515 apoptosis resistance. *Apoptosis.* 2006;11(3):303-313. doi:10.1007/s10495-006-3978-9
- 516 51. Pauls SD, Marshall AJ. Regulation of immune cell signaling by SHIP1: A phosphatase, scaffold protein, and potential therapeutic
517 target. *Eur J Immunol.* 2017;47(6):932-945. doi:10.1002/eji.201646795
- 518 52. Jinam TA. Human Leukocyte Antigen (HLA) Region in Human Population Studies. In: Saitou N, ed. *Evolution of the Human Genome I.*
519 *Evolutionary Studies.* Springer Japan; 2017:173-179. doi:10.1007/978-4-431-56603-8_9
- 520 53. McCarthy S, Das S, Kretschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet.*
521 2016;48(10):1279-1283. doi:10.1038/ng.3643
- 522 54. Durbin R. Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). *Bioinforma Oxf Engl.*
523 2014;30(9):1266-1272. doi:10.1093/bioinformatics/btu014
- 524 55. Loh P-R, Danecsek P, Palamara PF, et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet.*
525 2016;48(11):1443-1448. doi:10.1038/ng.3679
- 526 56. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human
527 genome. *Nucleic Acids Res.* 2019;47(D1):D886-D894. doi:10.1093/nar/gky1016
- 528 57. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and
529 functional annotation. *Nucleic Acids Res.* 2016;44(D1):D733-745. doi:10.1093/nar/gkv1189
- 530 58. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580-585. doi:10.1038/ng.2653
- 531 59. McInnes BT, Pedersen T. Evaluating measures of semantic similarity and relatedness to disambiguate terms in biomedical text. *J*
532 *Biomed Inform.* 2013;46(6):1116-1124. doi:10.1016/j.jbi.2013.08.008
- 533

534 **Table 1: Population characteristics**

	Population controls	Cognitively healthy centenarians
Number of individuals	2,905	343
Females (%)	1400 (48.2)	246 (71.7)
Age (SD)^a	68.3 (11.5)	101.4 (1.8)
ApoE ε4 (%)	1,012 (17.38)	48 (7.15)
ApoE ε2 (%)	523 (9.00)	91 (13.26)

535 ^a Age at study inclusion; SD, standard deviation; *ApoE*, Apolipoprotein E allele count for ε4 and ε2, and relative allele frequency
536 in population controls and cognitively healthy centenarians. Reference to the cohorts reported in this table are: ^{14,20-24}



537

538

539

540

541

542

543

544

545

546

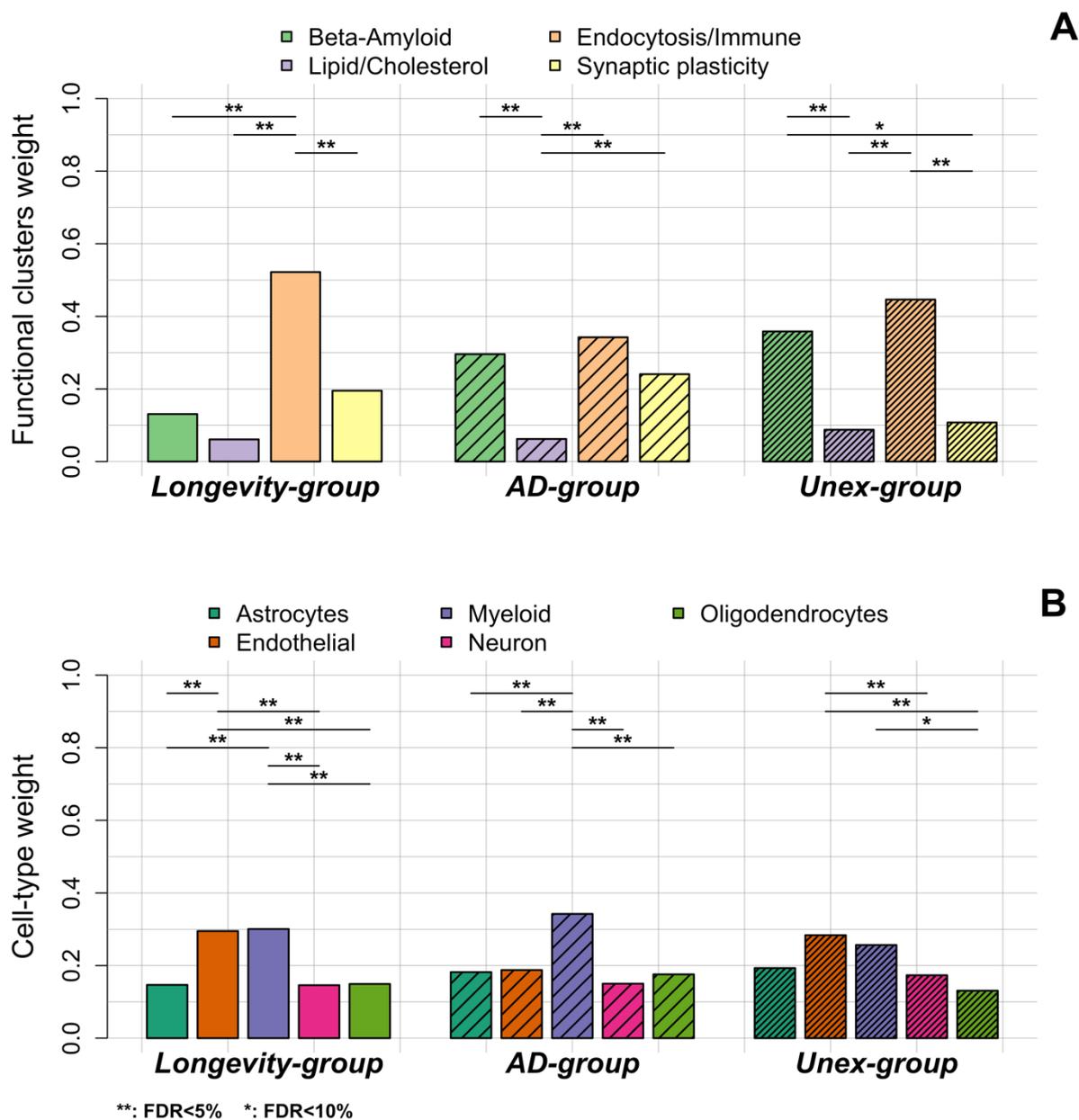
547

548

549

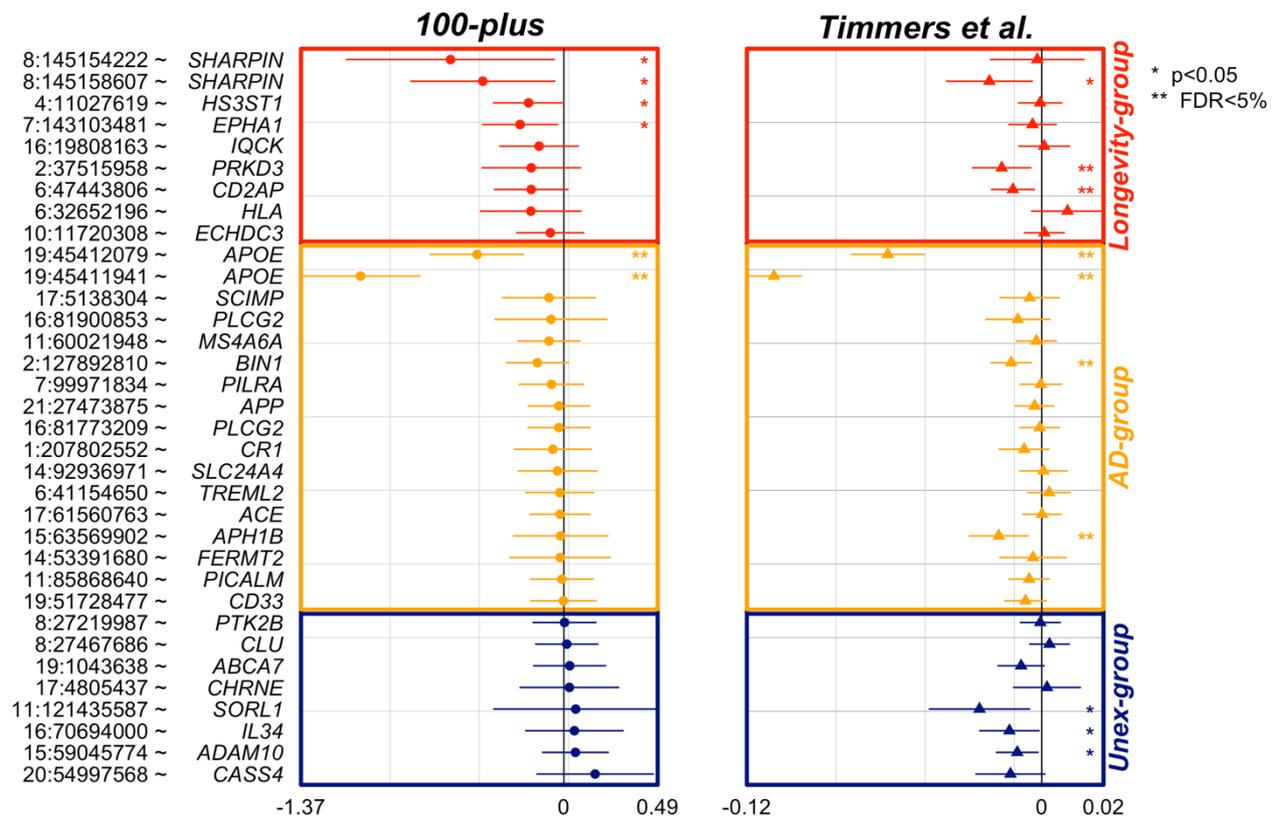
550

Figure 1: Overview of the 38 genetic variants associated with Alzheimer's disease. **A.** The genomic position of the variants (chromosome: position), variant identifier, and closest gene. Genomic positions are with respect to GRCh37 (hg19). **B.** The variant-pathway mapping score of association with the four functional clusters (darker colors representing stronger associations). Variants annotated with red crosses could not be annotated to any one of the functional clusters as no biological processes are associated with the related genes. **C.** The effect size on AD (from literature) and the observed effect size on longevity (LGV) for each variant (darker color indicating stronger effect). The same color indicates *expected direction* (i.e. increased risk of AD and decreased chance of longevity), while different colors, visible in the *Unex-group* of variants, indicates *unexpected direction*. For the longevity effects, we also annotate variants for which we observed a significant association (unadjusted p -value < 0.05). **D.** The distribution of the imbalance direction of variant effect (IED) in AD-risk as compared to cognitive health aging (see *Methods* for details). The *longevity*-, *AD*- and *Unex*-groups were derived based on the median value of the IED. The median value is reported for each IED as a blue vertical line. **E.** Average gene expression of the genes associated with the variant in five different brain cell-types (the darker, the higher the expression).



551

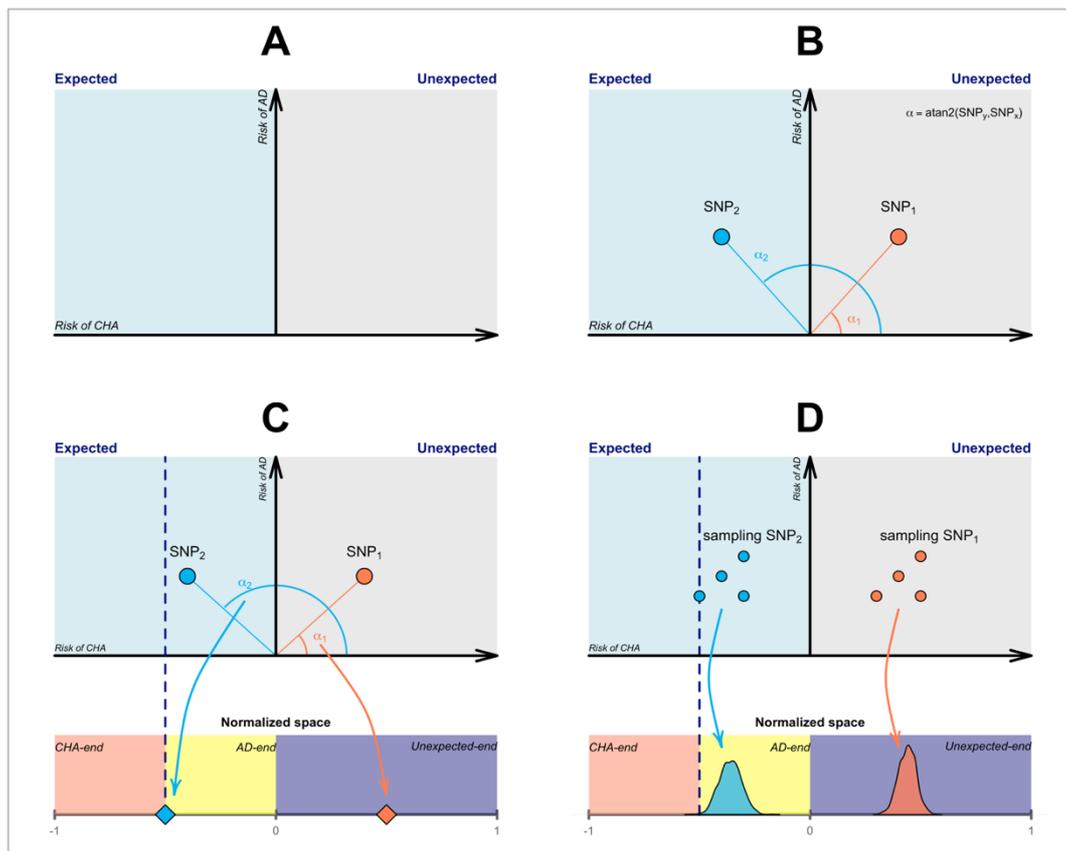
552 **Figure 2: Comparison of functional annotation and cell-type annotation within the Longevity-, AD- and Unex-**
 553 **groups. A.** The weights of the 4 functional clusters within the *longevity-*, *AD-* and *Unex-*
 554 *groups*. **B.** The weights of the different cell-types in the brain, per group. Differences in functional weights and cell-type weights within each
 555 group were calculated using Wilcoxon sum rank tests. The resulting p-values were FDR-corrected.



Effect size on longevity of AD risk-increasing alleles

556

557 **Figure 3: Forest plot of association statistics of AD-variants in our study and the largest GWAS of parental**
 558 **longevity.** The plot shows the association of AD-variants in our study and the largest by-proxy GWAS on parental
 559 longevity.⁶ The association statistics of 34/38 variants were available from publicly available summary statistics of
 560 Timmers et al. study. Plotted effect-sizes are with respect to the AD-risk increasing allele. Thus, an expected
 561 direction of effect is shown for variants with a negative estimate. Nominally significant associations with AD
 562 ($p < 0.05$) are annotated with an asterisk (*), and significant associations after FDR correction are annotated with
 563 two asterisks (**).



564

565

566

567

568

569

570

571

572

573

574

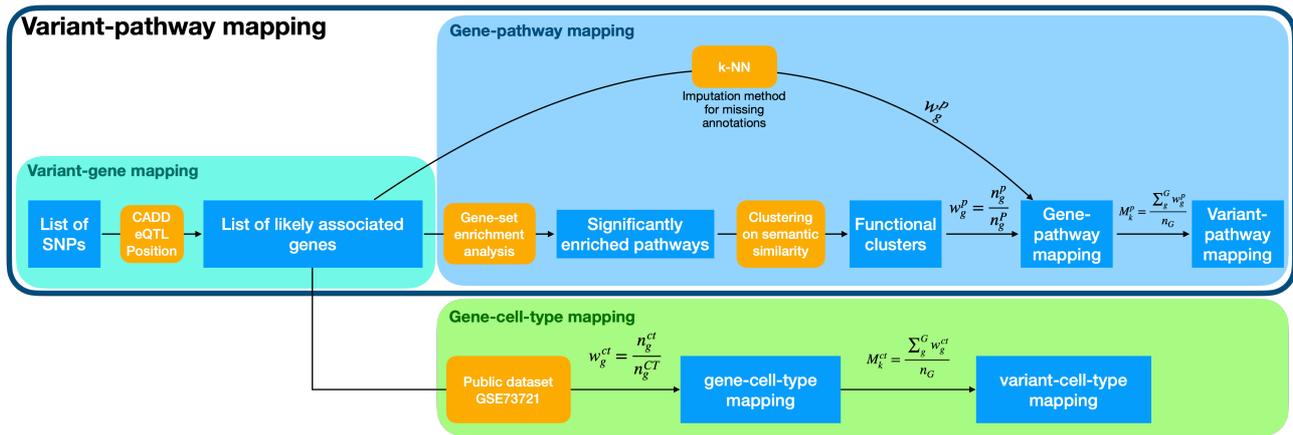
575

576

577

578

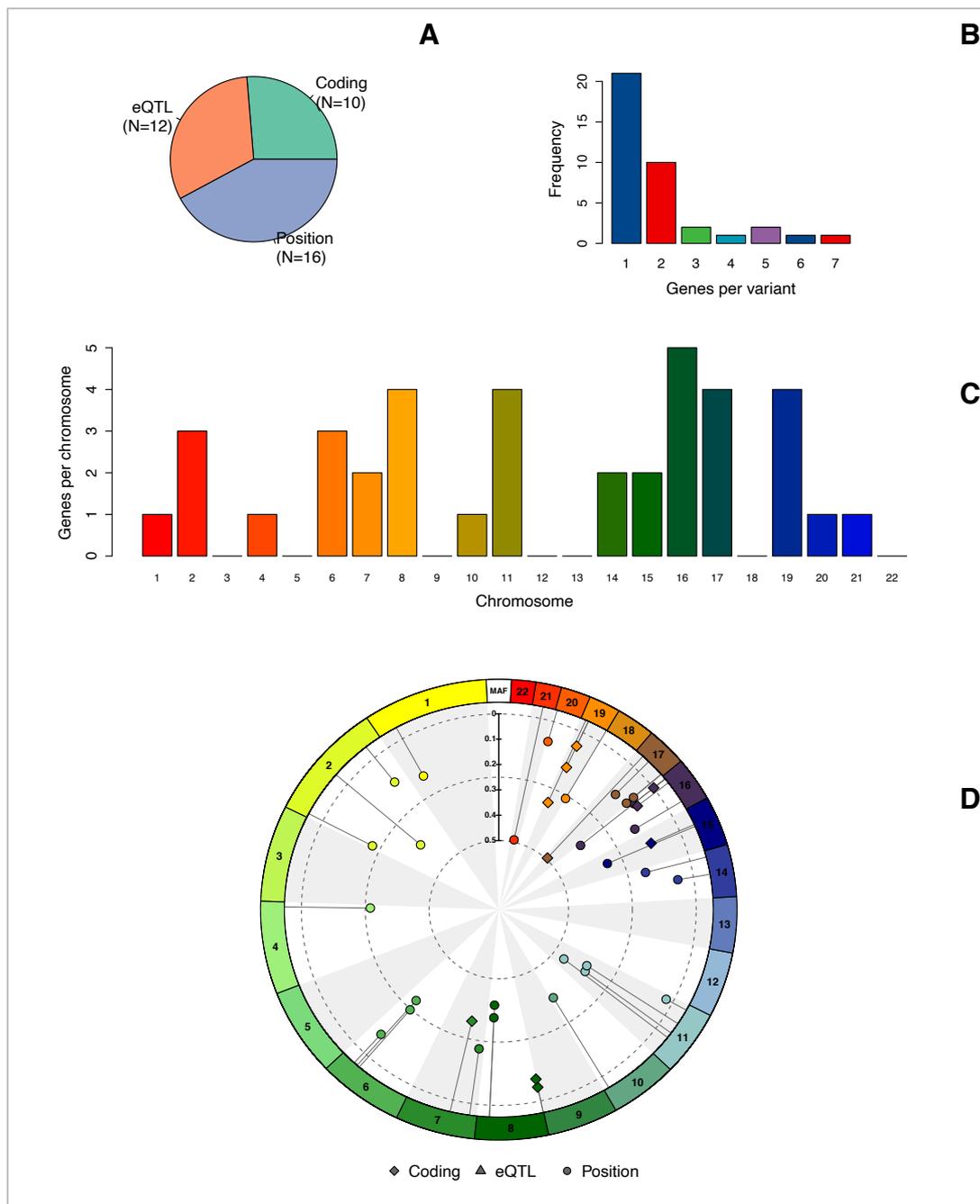
Figure S1: Explanation of the distribution of imbalance variant effect direction (IED). The figure shows the sequential steps for constructing the distribution of the expected direction of variant effect for AD-risk compared to longevity for two toy variants (SNP_1 and SNP_2). **A.** Axes definition, with the y-axis being the effect-size for AD-risk (log of odds ratio) of a variant, derived from literature and set positive by definition. The x-axis identifies the effect-size of a variant on longevity. This can be either positive or negative depending on the variant's association in cognitively healthy centenarians as opposed to population subjects. The blue area represents that the two effects are in the expected direction with respect to each other, *i.e.* a variant increases the risk of AD and at the same time decreases the chance of longevity. Oppositely, the grey area refers to the unexpected direction of effect. **B.** Two toy variants (SNP_1 and SNP_2) are shown as data points. α_{1-2} represents the angle of the data point vector with the x-axis. **C.** Normalization of the α_{1-2} value into an arbitrary space. Here, we used $[-1; 1]$. **D.** Repeating this procedure for each bootstrap iteration of each variant, we obtained the distribution of imbalance effect direction for each variant (*IED*). Values smaller than 0 indicate the *expected direction* of effect, whereas values larger than 0 refer to the *unexpected direction* of effects. Additionally, values close to 0 indicate a larger AD effect than longevity effect, and values close to -1 suggest that the variant's longevity effect is larger than the AD effect.



579

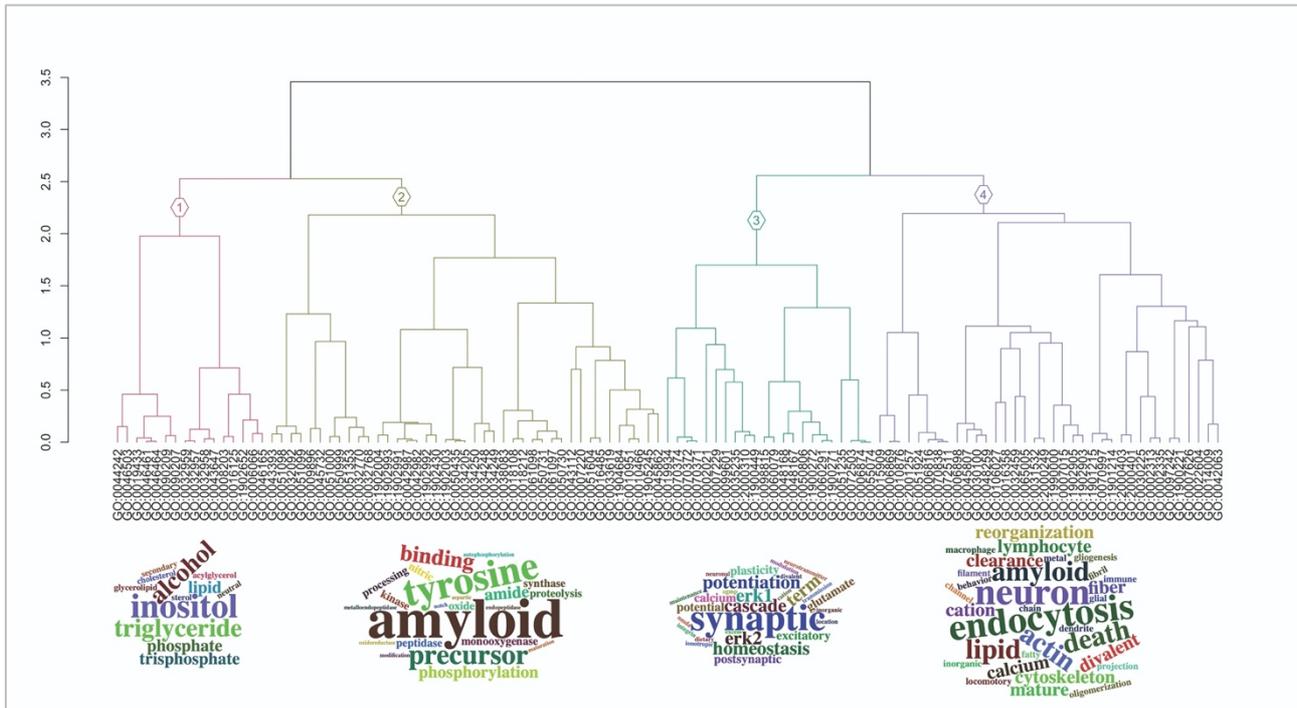
580 **Figure S2: Schematic representation of the variant-pathway and variant-cell-type mapping.** The figure shows
 581 a schematic representation of the annotation framework used to functionally annotate AD-associated variants
 582 and perform cell-type enrichment. Outputs are represented as blue squares, while methods are represented in
 583 orange. In the variant-gene mapping, showed in the grey box, we start from a list of variants and, through the
 584 integration of predicted variant consequences (CADD), eQTL and position, we obtain a list of genes. Note that here
 585 multiple genes may be associated with each variant. The yellow box shows the gene-pathway mapping: briefly, we
 586 perform gene-set enrichment analysis followed by clustering of the significantly enriched pathways to obtain
 587 functional clusters. We then calculate the gene-pathway mapping by looking at the (enriched) pathways
 588 associated with each gene and their associated functional clusters to get a weight for each gene-functional cluster
 589 association. Finally, we average the gene-pathway mapping of each gene associated with the same variant.
 590 Imputation methods (k-NN) are implemented for genes with missing annotation to obtain the gene-pathway
 591 mapping. Together, the grey box and the yellow box form the variant-pathway mapping. At the bottom, the green
 592 box shows the gene-cell-type enrichment using the public dataset GSE73721 of gene expression in different brain
 593 cell-types. Similar to the gene-pathway mapping, we calculate a weight of association of each gene to each cell-
 594 type, and we average these weights in case multiple genes mapped to the same variant (variant-cell-type mapping).
 595

596



597

598 **Figure S3: Variant-gene mapping for the 38 AD-associated variants.** **A.** The sources used to annotate each
 599 variant to the likely affected genes. Coding: variants located in the coding region of a gene (*e.g.* synonymous or
 600 non-synonymous variants). eQTL: variants associated with RNA expression changes in blood from the GTEx
 601 consortium. *Position*: variants intronic or intergenic without evidence of eQTL associations that were annotated
 602 based on neighboring genes. **B.** Barplot of the number of genes associated with each variant. **C.** Distribution of
 603 genes across the chromosomes. **D.** Distribution of the previously identified variants along the genome together
 604 with each variant's minor allele frequency and annotation.



605

606 **Figure S4: Hierarchical clustering of the semantic similarity matrix and the 4 functional clusters' definition.**

607 Dendrogram of the hierarchical clustering analysis and the 4 functional clusters, along with word-clouds of the
608 most frequent terms per cluster. Hierarchical clustering was performed on the semantic similarity distance matrix
609 (using *Lin* as semantic similarity metric). We used the dynamic tree-cut method to define the number of functional
610 clusters, specifying 15 as the minimum number of terms per cluster. We then used word-cloud visualization as well
611 as manual interpretation of the biological processes underlying each functional cluster to label each cluster to
612 *Lipid/Cholesterol metabolism* (cluster 1), *β -Amyloid metabolism* (cluster 2), *Synaptic plasticity* (cluster 3) and
613 *Endocytosis/Immune signaling* (cluster 4).

614 **Supplementary Methods**

615 **Populations**

616 The 100-plus Study focuses on the biomolecular aspect of preserved cognitive health until extremely old ages. This
617 study includes (1) Dutch-speaking centenarians who can (2) provide official evidence for being aged 100 years or
618 older, (3) self-report to be cognitively healthy, which is confirmed by an informant (*i.e.* a child or close relation), (4)
619 consent to donation of a blood sample and (5) consent to (at least) two home-visits from a researcher, which
620 includes an interview and neuropsychological testing.¹⁴ This study also includes (1) siblings or children from
621 centenarians who participate in the 100-plus Study, or partners thereof who (2) agree to donate a blood sample,
622 (3) agree to fill in a family history, lifestyle history, and disease history questionnaire. The Longitudinal Aging Study
623 of Amsterdam (LASA) is an ongoing longitudinal study of older adults initiated in 1991, with the main objective to
624 determine predictors and consequences of aging.^{19,20} The SCIENCe is a prospective cohort study of subjective
625 cognitive decline (SCD) patients.^{21,22} Participants undergo extensive assessment, including cerebrospinal fluid
626 collection (CSF) and optional amyloid positron emission tomography scan (PET), with annual follow-up. The
627 primary outcome measure is clinical progression. All individuals were labeled cognitively intact. The Netherlands
628 Brain Bank (NBB) cohort is a prospective donor program for psychiatric diseases. All subjects were labeled
629 cognitively intact after neuropathological examination.²³ The Netherland Twin Registry study (NTR) was
630 established in 2004 to collect biological and environmental data in twin families to create a resource for genetic
631 studies on health, lifestyle, and personality.²⁴

632

633 **Genotyping and imputation**

634 Genetic variants in our populations were determined by standard genotyping and imputation methods, and we
635 applied established quality control methods: we genotyped all individuals with the Illumina Global Screening Array
636 (GSAsharedCUSTOM_20018389_A2) and excluded individuals with low-quality genotypes (individual call rate
637 <98%, variant call rate <98%), individuals with sex mismatches and variants deviating from Hardy-Weinberg
638 equilibrium ($p < 1 \times 10^{-6}$). Genotypes were prepared for imputation comparing variants identifiers, strand and allele
639 frequencies to the Haplotype Reference Panel (HRC v1.1, April 2016), and all remaining variants were submitted to
640 the Sanger imputation server (<https://imputation.sanger.ac.uk>).⁵³ The server uses EAGLE2 (v2.0.5) to phase the
641 data, and imputation to the reference panel was performed with PBWT.^{54,55} Before analysis, we excluded

642 individuals of non-European ancestry and individuals with a family relation, leaving 2,905 population subjects and
643 343 cognitively healthy centenarians for the analysis.

644

645 **Variant-gene mapping**

646 We annotated each variant to the likely affected gene(s), so-called *variant-gene mapping*, combining annotation
647 from *Combined Annotation Dependent Depletion* (CADD, v1.3), *expression-quantitative-trait-loci* in the blood (eQTL
648 from GTEx v8), and positional mapping (from RefSeq build 98).⁵⁶⁻⁵⁸ In the case of coding variants, we confidently
649 associated the variant with the corresponding gene. Alternatively, we first considered possible eQTL associations.
650 When these were not available, we included all genes at increasing distance d from the variant (starting with $d \leq$
651 $50kb$, up to $d \leq 500kb$, increasing by $50kb$ until at least 1 gene was found). Our procedure allows the association
652 of each variant with one or multiple genes (*Figure S2*).

653

654 **Gene-pathway mapping**

655 The resulting list of genes was used to find the molecular pathways enriched in the AD variants. See *Figure S2* for a
656 schematic representation of our annotation framework. We realized that allowing multiple genes to associate with
657 each variant could result in an enrichment bias, as neighboring genes are often functionally related. To control this,
658 we implemented a sampling technique: at each iteration, we (i) sampled one gene from the pool of genes
659 associated with each variant, and (ii) performed a gene-set enrichment analysis with the resulting list of genes. The
660 gene-set enrichment analysis was performed considering biological processes (BP) and implemented with the
661 *enrichGO* function of the R package *clusterProfiler*, with all genes as background and correcting p-values controlling
662 the False Discovery Rate (FDR). Finally, we averaged p-values for each enriched term over the iterations (N=1,000).
663 To facilitate interpretation, we merged significantly enriched biological processes. First, we calculated the
664 semantic similarity between all significant biological processes (*i.e.* FDR<5%) using *Lin* as a distance measure.⁵⁹ We
665 then applied hierarchical clustering on the resulting distance matrix and selected the number of functional clusters
666 using the dynamic tree-cut method as implemented in *cutreeDynamic* function from the R package *WGCNA*,
667 specifying 15 as the minimum number of terms per cluster (using the default value of 20 resulted in 2 functional
668 clusters only). To provide an interpretation of each functional cluster, we selected the most frequent words
669 describing the biological processes underlying each cluster, and show this as word-clouds as implemented in R
670 package *wordcloud2*. Finally, by counting how often a functional cluster was associated with a gene, we could

671 calculate a weighted annotation of each gene to the 4 functional clusters, so-called gene-pathway mapping (*Figure*
672 *S2*).

673 Due to the initial selection of significantly enriched BP, not every gene in the list of variant-associated genes is
674 annotated with (at least one of) these terms. Consequently, these genes could not be related to the final functional
675 clusters. To overcome this, we connect these genes to the functional clusters using a k-nearest neighbor (k-NN)
676 imputation. The k-NN model was initially trained using the functional clusters as classes and the semantic
677 similarity matrix between the enriched biological processes as features (*feature terms*). Then, for each gene with
678 missing annotation, we (i) extracted all the biological processes the gene is involved in (*input biological processes*),
679 and (ii) calculated the semantic similarity matrix between these terms and the *feature terms*, which defines the
680 similarity between the *input biological processes* and the *feature terms*. Finally, we (iii) predicted the probability of
681 classification of the similarity matrix to the classes (functional clusters), and used this as weight for the gene-
682 pathway mapping (*Figure S2*).

683

684 **Variant-pathway mapping**

685 The variant-pathway mapping represents the combined annotation of each variant to the different functional
686 clusters. As such, it depends on the variant-gene mapping and the gene-pathways mapping. Briefly, given a variant
687 k , we (i) retrieved all the genes that were associated with the variant in the variant-gene mapping, G_k , and (ii)
688 retrieved all the biological processes (gene ontology term identifiers) that were associated with these genes, GO_G .
689 Because we clustered biological processes into functional clusters, by looking at which functional clusters the GO_G
690 belonged to, we could assign a weight of association for variant k to each of the functional clusters.

691

692 **Variant-cell-type mapping**

693 To study brain-specific cell-types and their relationship with AD-associated variants, we used the publicly available
694 gene expression dataset GSE73721: this dataset includes gene expression values of 6 fetal astrocyte samples, 12
695 adult astrocyte samples, 8 sclerotic hippocampal samples, 4 whole human cortex samples, 4 adult mouse
696 astrocyte samples, and 11 human samples of other purified central-nervous-system (CNS) cell-types. We restricted
697 to the gene expression of 12 astrocyte samples and 11 samples of purified CNS cell-types from the cortex of adult
698 humans (total $N=23$, mean age of 41.5 ± 19.6 years). To calculate the variant-cell-type mapping, we averaged the
699 gene expression of the genes mapping to the same variant.