Centenarian Controls Increase Variant Effect-sizes by an average two-fold in an Extreme Case-Extreme Control Analysis of Alzheimer's Disease

Niccolò Tesi,[1,2,3]* Sven J. van der Lee,[1,2]* Marc Hulsman,[1,2,3] Iris E. Jansen,[1,4] Najada Stringa,[5,6] Natasja van Schoor,[5,6] Hanne Meijers-Heijboer,[2] Martijn Huisman, [5,6] Philip Scheltens,[1] Marcel J.T. Reinders,[3] Wiesje M. van der Flier [1,5] and Henne Holstege [#1,2,3]

*authors contributed equally

[1] Alzheimer Centre, Department of Neurology, VU University Medical Centre, Amsterdam, The Netherlands

[2] Department of Clinical Genetics, VU University Medical Centre, 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 Biostatistics, VU University Medical Centre, Amsterdam, The Netherlands

[6] Amsterdam Public Health Research Institute, Amsterdam, The Netherlands

Keywords: cognitively healthy centenarians, extreme phenotypes, Alzheimer's disease, age-related diseases, case-control study, increased power of association

Corresponding Author

Dr. Henne Holstege

VUmc Alzheimer's Center

De Boelelaan 1118

1081 HZ Amsterdam, The Netherlands

Tel: +31 20 4440816

30 Email: h.holstege@vumc.nl

Abstract

The detection of genetic loci associated with Alzheimer's disease (AD) requires large numbers of cases and controls because variant effect-sizes are mostly small. We hypothesized that variant effect-sizes should increase when individuals who represent the extreme ends of a disease spectrum are considered, as their genomes are assumed to be maximally enriched or depleted with disease-associated genetic variants.

We used 1,073 extensively phenotyped AD cases with relatively young age at onset as extreme cases (66.3±7.9 years), 1,664 age-matched controls (66.0±6.5 years) and 255 cognitively healthy centenarians as extreme controls (101.4±1.3 years). We estimated the effect-size of 29 variants that were previously associated with AD in genome-wide association studies.

Comparing extreme AD-cases with centenarian-controls increased the variant effect-size relative to published effect-sizes by on average 1.90-fold (*SE*=0.29, p=9.0x10^{-4}). The effect-size increase was largest for the rare high-impact *TREM2 (R74H)* variant (6.5-fold), and significant for variants in/near *ECHDC3* (4.6-fold), *SLC24A4-RIN3* (4.5-fold), *NME8* (3.8-fold), *PLCG2* (3.3-fold), *APOE-ε2* (2.2-fold) and *APOE-ε4* (2.0-fold). Comparing extreme phenotypes enabled us to replicate the AD association for 10 variants (p<0.05) in relatively small samples. The increase in effect-sizes depended mainly on using centenarians as extreme controls: the average variant effect-size was not increased in a comparison of extreme AD cases and age-matched controls (0.94-fold, p=6.8x10^{-1}), suggesting that on average the tested genetic variants did not explain the extremity of the AD-cases.

Concluding, using centenarians as extreme controls in AD case-controls studies boosts the variant effect-size by on average two-fold, allowing the replication of disease-association in relatively small samples.

Introduction

Alzheimer's disease (AD) is characterized by a slow but progressive loss of cognitive functions, leading to loss of autonomy.[1] AD is rare at the age of 65 years, but its incidence increases exponentially to 40% at the age of 100 years.[2] It is currently the most prevalent cause of death at old age and one of the major health threats of the 21st century.[1] Better understanding of the etiological factors that determine AD is warranted as no treatment is currently available. Heritability plays an important role as genetic factors are estimated to determine 60-80% of the risk of AD.[3] About 30% of the genetic risk is attributable to the *ε4* allele of *APOE* gene, and large collaborative efforts have identified over two dozen additional genetic loci that are associated with a slight modification of the risk of AD.[4–17] The design of these association studies relies on the comparison of very large numbers of cases with age-matched controls, such that detected associations can be attributed specifically to the disease.[18] However, given the prevalence of AD in the aging population, it is likely that a significant fraction of the controls will develop the disease at a later age. Therefore, as the AD risk for future cases likely involves the same genetic variants, using age-matched controls may quench variant association signals. This may, in part, explain the mostly small variant effect-sizes associated with common variants. Also, GWAS studies mostly compare common genetic variants that are widely propagated in the population; as a consequence, these have mostly small effects on AD risk.[19] Rare genetic variants often have larger effect-sizes than common variants, but as there are fewer carriers available in the population, the requirement for large sample sizes stands.[20]

Instead of increasing sample sizes of genetic studies to detect novel disease-associated genetic loci, an alternative strategy is to increase variant effect-sizes by sampling individuals with extreme phenotypes.[20–22] For AD and other age-related diseases, extreme cases may be defined by having a relatively early age at disease-onset, and having the phenotypic features characteristic for the disease, as defined by diagnostic assessment. Extreme

controls are represented by individuals who reach extreme ages without the disease.[21,23,24] Indeed, in a case-control study of type-2 diabetes, the effect-sizes for variants that were previously associated with the disease were increased when using centenarians as extreme controls.[23] The effect of using extreme phenotypes in other age-related diseases has not been studied.

Here, we explored the potential of using extreme phenotypes for genetic studies of Alzheimer's disease (AD) by investigating the change in effect-size of known AD-associated variants. Furthermore, using an age- and population-matched reference group, we investigated the contribution of each extreme phenotype.

Methods

Cohort description

As extreme AD cases group (denoted by *EA*) we used 1,149 AD cases from the Amsterdam Dementia Cohort (ADC). The ADC comprises patients who visit the memory clinic of the VU University Medical Center, The Netherlands.[25,26] This cohort of AD patients is extensively characterized and comprises 503 early-onset cases (denoted by *eEA*) with an age at onset <65 years, and 646 late-onset cases (denoted by *lEA*). At baseline, all subjects underwent a diagnostic assessment including neurological examination, standard laboratory tests of blood and cerebrospinal fluid, electroencephalogram and brain magnetic resonance imaging. Clinical diagnosis is made in consensus-based, multidisciplinary meetings. The diagnosis of probable AD was based on the clinical criteria formulated by the National Institute of Neurological and Communicative Disorders and Stroke - Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) and based on National Institute of Aging - Alzheimer association (NIA-AA).[25–28]

As extreme control group (denoted by *EC*) we used 268 self-reported cognitively healthy centenarians from the 100-plus Study cohort.[29] This study includes Dutch-speaking individuals who (i) can provide official evidence for being aged 100 years or older, (ii) self-report to be cognitively healthy, which is confirmed by a proxy, (iii) consent to donation of a blood sample, (iv) consent to (at least) two home visits from a researcher, and (v) consent to undergo an interview and neuropsychological test battery.

As 'normal controls' (denoted by *NC*) we used 1,717 middle-aged (55-85 year-old) individuals from a representative sample of Dutch individuals from the Longitudinal Aging Study Amsterdam (LASA) cohort.[30,31] LASA is an ongoing longitudinal study of older adults initiated in 1991, with the main objective to determine predictors and consequences of aging.

The Medical Ethics Committee of the VU University Medical Center (METC) approved the ADC cohort, the LASA study and the 100-plus Study. All participants and/or their legal guardians gave written informed consent for participation in clinical and genetic studies.

Genotyping and imputation of 29 selected AD-associated genetic variants

We selected 29 single nucleotide variants for which evidence for a genome-wide significant association with Alzheimer's disease was found in previous studies (*Table S1*).[4–17] Genetic variants were determined by standard genotyping or imputation methods. Briefly, we genotyped all individuals using the Illumina Global Screening Array (GSAsharedCUSTOM_20018389_A2) and applied established quality control methods.[32] We used high quality genotyping in all individuals (individual call rate >98%, variant call rate >98%), individuals with sex mismatches were excluded and HWE-departure (d-HWE) was considered significant at $p<1\times10^{-6}$. Genotypes were prepared for imputation using provided scripts (HRC-1000G-check-bim.pl).[33] This script compares variant ID, strand and allele frequencies to the haplotype reference panel (HRC v1.1, April 2016).[33] Finally, all

autosomal variants were submitted to the Michigan imputation server (https://imputationserver.sph.umich.edu).[32] The server uses *SHAPEIT2* (v2.r790) to phase data and imputation to the reference panel (v1.1) was performed with *Minimac3.*[32,34] A total of 1,149 extreme AD cases, 1,717 normal controls and 268 extreme (centenarian) controls passed quality control. Prior to analysis, we excluded individuals of non-European ancestry (N_{EA} = 67, based on 1000Genomes[35] clustering) and individuals with a family relation (N_{EA} = 9, N_{EC} = 13, N_{NC} = 53, identity-by-descent ≥ 0.3),[36] leaving 1,073 extreme AD cases (N_{eEA} = 464 and N_{lEA} = 609), 1,664 normal controls and 255 centenarian controls for the analysis.

Statistical Analysis

For each AD-associated variant, we explored the *change in effect-size* (E) relative to reported effect-sizes when 1) comparing extreme AD cases with extreme (centenarian) controls (*EA vs*. *EC*); 2) comparing extreme AD cases with normal controls (*EA vs*. *NC*); and 3) comparing normal AD cases with extreme (centenarian) controls (*NA vs. EC*). To calculate variant effect-sizes, we used logistic regression models correcting for population stratification (principal components 1 to 6).[37,38] We calculated odds ratios (OR) relative to the HRC alternative allele assuming additive genetic effects, and estimated 95% confidence intervals (CIs).

We estimated the *change in effect-size* relative to reported effect sizes (E) as follows:

$$E_{1-2}^{k} = \frac{\log OR_{1-2}^{k}}{\log OR_{l}^{k}} \quad (1)$$

where E_{1-2}^{k} indicates the effect-size change for variant k in a comparison of cohort 1 and cohort 2, *e.g*, $E_{EA-EC}^{APOE\ \varepsilon 4}$ indicates the effect-size change for the *APOE ε4* variant when extreme AD cases (*EA*) are compared with cohort healthy centenarians (*EC*). The

159 $\log OR^k_{1-2}$ denotes the *effect-size* of variant k when comparing cohort 1 and cohort 2. The
160 effect-size of variant k reported in literature (*Table S1*) is denoted by $\log OR^k_l$.
161
162 We estimated the added value of using extreme (centenarian) controls rather than 'normal
163 age-matched controls' in a case-control analysis. For this, we wanted to compute the change
164 in effect size when comparing non-extreme AD cases with extreme controls (*NA* vs *EC*). As
165 we do not have direct access to 'normal AD cases', we estimated the effect-size for the *NA-*
166 *EC* comparison by summing (1) the effect-size when comparing 'normal AD cases' and
167 'normal controls', as reported in literature ($\log OR^k_l$), and (2) the effect size when comparing
168 normal controls (*NC*) with extreme (centenarian) controls (*NC* vs *EC*), *i.e.* $\log OR^k_{NA-EC} =$
169 $\log OR^k_l + \log OR^k_{NC-EC}$. The added value of using extreme controls in a case control analysis
170 then becomes:

$$E^k_{NA-EC} = \frac{\log OR^k_l + \log OR^k_{NC-EC}}{\log OR^k_l} \quad (2)$$

171
172 To assess whether age at disease onset had an impact on the change in effect-size due to
173 the extreme cases (E_{EA-NC}), we estimated the $\log OR^k_{eEA-NC}$ (early-onset extreme AD cases
174 *vs.* normal controls) and $\log OR^k_{lEA-NC}$ (late-onset extreme AD cases *vs.* normal controls),
175 and their 95% confidence intervals. Then, we computed the probability that the effect size
176 changes E^k_{eEA-NC} and E^k_{lEA-NC} differed using a two-samples z-test (two-tailed *p-value*).
177
178 **Determining significance of change in effect-size**
179 For each variant, we estimated E^k_{1-2} and a 95% confidence intervals (CI) by sampling
180 (*S*=10,000) from the $\log OR^k_{1-2}$ and $\log OR^k_l$ based on their respective standard errors. The
181 probability of divergence between the distributions of the $\log OR^k_{1-2}$ and the $\log OR^k_l$ was
182 determined using a two-sample z-test (two-tailed *p-value*).
183

The probability of observing $E_{1-2}^{k} > 1$, *i.e* an increased effect-size for variant k, is considered to be a Bernoulli variable with *p*=0.5 (equal chance of having an increased/decreased effect). The number of variants that show an increase in effect ($E_{1-2}^{k} > 1$) then follows a binomial distribution.

The average change in effect-size across all K=29 tested variants is calculated as follows:

$$\bar{E}_{1-2} = \frac{1}{K}\sum_{k}^{K} E_{1-2}^{k} \quad (3)$$

Confidence intervals and probability of divergence between $\bar{E}_{1-2}$ and previously reported effect-sizes were estimated by sampling (*S*=10,000, two-tailed *p-value*).

All described statistical analysis were performed with *PLINK* (v1.90b4.6) or *R* (v3.3.2).[39,40]

Results

After quality control of the genetic data, we included 1,073 extreme AD cases (with mean age at onset 66.4±7.8 and 52.7% females), 1,664 normal (age-matched) controls (mean age at inclusion 66.0±6.5, 53.7% females), and 255 cognitive healthy centenarians as extreme controls (mean age at inclusion 101.4±1.3, 74.7% females) (*Table 1*). Within the extreme AD cases group, there were 464 early-onset cases (mean age at onset 59.1±4.1, 54% females), and 609 late-onset cases (mean age at onset 72.1±4.8, 51% females). The age at onset of the extreme AD cases was on average 8.2 years earlier compared to previous GWA studies; the age at disease onset was on average 15.4 years earlier in early-onset cases and 2.5 years earlier in late-onset cases, while the age of study inclusion of our centenarian controls was on average 29.5 years higher than for previously published controls (*Figure 1*).

Effect of comparing extreme cases and centenarian controls

In a genetic comparison of extreme AD cases and centenarian controls (*EA-EC* comparison) the average effect-size over all 29 genetic variants was 1.90-fold increased relative to the effect-sizes reported in published studies ($\bar{E}_{EA-EC}$ = 1.90±0.29; p = 9.0x10^{-4}) (*Figure 3*). For 21 out of 29 variants, we observed an increased effect size (E^{k}_{EA-EC} > 1), which is significantly more than expected by chance (p = 1.2x10^{-2}) (*Figure 2* and *Table 2)*. The increase in effect-size ranged from 1.06 (variant near *CASS4*) to 6.46 (variant in *TREM2 [R47H]*). For variants near or in the genes *TREM2 (R47H), SLC24A4-RIN3* and *ECHDC3,* the increase was more than 4-fold compared to previously reported effect-sizes. For 9 variants the effect-size increase was 2-4-fold (in or near the genes *NME8*, *PLGC, HLA-DRB1, 2CD2AP, ZCWPW1, ABCA7 APOE [ε2], [A>G], HS3ST1* and *ABI3,* in order from high to low effect-size increases*).* For 9 variants the increase was between 1 and 2-fold (in or near genes*, APOE ε4, RPHA1, CELF1, PTK2B, MS4A6A, SORL1, BIN1, PICALM* and *CASS4*) (*Figure 2*). The effect-sizes for 6 genetic variants were not increased in our extreme phenotype analysis compared to previously reported effect-sizes (E^{k}_{EA-EC} between 0 and 1): in or near *TREM2 [R62H], KANSL1 CR1, ABCA7 [G>C], CLU,* and *INPP5D*). Lastly, the effect-sizes of 2 variants were in the opposite direction compared to previously reported effects (E^{k}_{EA-EC} < 0). Specifically, for the variant in *FERMT2* we found an inverted direction of effect-size and a lower magnitude of effect as compared with previous studies (E^{FERMT2}_{EA-EC} between 0 and -1). For the variant near *MEF2C* we observed a larger effect-size as compared to those previously published, but in the opposite direction (E^{MEF2C}_{EA-EC} < -1).

Overall, for 7 variants, the effect size was significantly increased relatively to the previously reported effect-sizes (*Table 2*), in or near genes *APOE ε2* (2.2-fold, p = 1.4x10^{-7}), *APOE ε4* (2.0-fold, p = 1.5x10^{-9})*, SLC24A4-RIN3* (4.5-fold, p = 1.6x10^{-3})*, ECHDC3* (4.6-fold, p = 1.1x10^{-2})*, PLCG2* (3.3-fold p = 1.4x10^{-2})*, NME8* (3.9-fold, p = 1.7x10^{-2}) and *MEF2C* (-1.9-fold, p = 1.8x10^{-2}). Variants with significant effect-size changes, were also more likely to be associated with AD in a comparison of extreme cases and centenarians. The association

with AD reached nominal significance (p<0.05) in 10 out of 21 variants with a changed effect-size (*Table 2*). Next to *APOE ε4* ($\log OR_{EA-EC}^{APOE\ \varepsilon4}$ = 2.1, SE = 0.17, p = 1.3×10^{-33}) and *APOE ε2* ($\log OR_{EA-EC}^{APOE\ \varepsilon2}$ = -1.8, p = 3.2×10^{-21}), variants in or near these genes were significantly associated with AD: *SCL24A4-RIN3, PLCG2, ECHDC3, NME8, BIN1, ZCWPW1, ABCA7 (A>G)* and *HLA-DRB1* (*Table 2).*

Effect of using extreme AD cases

The average effect-size in a comparison of extreme AD cases with normal controls (*EA vs. NC*) did not significantly change relative to the previously reported effect-sizes ($\bar{E}_{EA-NC}$ = 0.94±0.12, p=6.8x10^{-1}) (*Figure 3*). For 14 individual variants, we observed an increased effect size, which was expected by chance (p=0.5), *Figure S1* and *Table S2*). The effect size was significantly increased for *APOE ε4* variant (1.3-fold, p = 1.4x10^{-5}), and nominally significant for *APOE-ε2* (1.4-fold, p = 1.7x10^{-2}).

We then separated AD cases into early-onset extreme AD cases (N_{eEA} = 464) and late-onset extreme AD cases (N_{lEA} = 609), and estimated the change in effect-sizes. When using early onset cases the average effect-size was lower relative to previously published effect sizes ($\bar{E}_{eEA-NC}$ was 0.86±0.16 (p = 7.9x10^{-1}), while for late-onset cases the effect size was similar to published effect sizes ($\bar{E}_{lEA-NC}$ was 1.01±0.14, p = 4.6x10^{-1}) (*Figure S3* and *Table S3*). We found significant differences between the effect-sizes in early-onset and late-onset AD cases ($\log OR_{eEA-NC}^{k}$ and $\log OR_{lEA-NC}^{k}$, respectively) for the variants in or near *APOE ε2* (-1.41 *vs.* -0.89; p=5.0x10^{-2}), *ZCWPW1* (0.01 *vs.* 0.24: p=1.6x10^{-2}) and *MS4A6A* (0.12 *vs.* -0.13; p=7.9x10^{-3}).

Effect of extreme controls

In a comparison of normal AD cases and extreme (centenarian) controls (*NA vs. EC*), the effect-size was on average 1.88-fold higher relative to previously reported effect-sizes

($\bar{E}_{NA-EC}$ = 1.88±0.24, p = 1.0x10^{-4}) (*Figure 3*). This was almost identical to the average increase in effect-size when we compared the extreme cases with centenarian controls ($\bar{E}_{EA-EC}$ = 1.90±0.29; p = 9.0x10^{-4}) (*Figure 3*). At the variant level, the change in effect-sizes was similar in both analyses, with the exception of the rare *TREM2 (R47H)* variant, whose effect-size increase was higher in the comparison with the extreme cases (but with high confidence intervals) (*Figure S4-A*).

In further concordance with the comparison of the extreme phenotypes, we observed an increased effect-size for 24/29 variants relative to published variant effect-sizes (E^{k}_{NA-EC} > 1), which is more than expected by chance (p = 2.7x10^{-4}) (*Figure S2* and *Table S2*). We found a significant increase in effect-size for variants in or near *APOE-ε2* (1.7-fold, p < 5x10^{-5}), *APOE-ε4* (1.7-fold, p < 5x10^{-5}), *NME8* (4.5-fold, p = *3.5x10*$^{-3}$), *SLC24A4-RIN3* (3.9-fold, p = *4.5x10*$^{-3}$) and *PLCG2* (2.9-fold, p = *1.9x10*$^{-2}$). In line with this, for almost all variants individually, the extreme controls contributed more to the effect size change in the extremes-comparison, than the extreme cases (*Figure S4-B*).

Discussion

In this study, we found that the effect sizes of 29 variants previously identified in genetic case-control analyses for Alzheimer's Disease were increased in a case-control analysis of extreme phenotypes. The use of extreme AD-cases and cognitively healthy centenarians as extreme-controls increased effect sizes for association with AD up to 6-fold, relative to previously published effect-sizes. On average, the use of extreme phenotypes almost doubled the variant effect-size. Although changes in effect-size were different per variant, the effect-size increase was driven mainly by the centenarian controls. This profound increase enabled us to replicate the association with AD of 10 variants in relatively small samples.

In a comparison of AD cases (either normal or extreme) with centenarian controls, we observed significant effect-size increases for variants in or near *PLCG2, NME8, ECHDC3, SLC24A4-RIN3, APOE-ε2* and *APOE-ε4*. This suggests that the tested variants or loci might contribute to the long-term preservation of cognitive health and/or to longevity in general. *PLCG2* and *NME8* are implicated in immunological processes,[8,43] while *SLC24A4*, *ECHDC3* and *APOE* are involved in lipid and cholesterol metabolism.[17,44,45] Both these processes were previously associated with longevity,[46,47] such that an overlapping etiology of maintained cognitive health and maintained overall health may contribute to the observed increase in effect-size. However, with the exception of the *APOE* locus, these loci were thus far not associated with longevity in GWAS studies.[48–51] We speculate that the association might be dependent on the cognitive health in the centenarians of the 100-plus Study cohort.[29] Alternatively, longevity studies may have been underpowered to detect the association of these loci with extreme survival. Future studies will have to establish the mechanism behind the association of these genes with preserved cognitive health. Next to *APOE*, the *HLA-DRB1* locus has been associated with both AD[13] and longevity.[48] However, its most informative variants, *rs9271192* for AD and *rs34831921* for longevity, are not in linkage disequilibrium,[52] suggesting that these are independent signals.

Using extreme cases did not increase the variant effect-sizes relative to published effect-sizes, even though most of the extreme cases were biomarker confirmed and their mean age at onset was 8.2 years younger than the mean age at onset in other studies.[7,8,13] This suggests that based on the tested genetic variants, the “phenotypically extreme” cases presented in this study were not genetically more extreme than cases presented in other studies.

Counter intuitively, in a comparison with normal controls, the variant effect-sizes of early-onset AD cases were on average *lower* than the variant effect-size of late-onset AD cases.

One explanation for this observation may be that the early age at onset may have been driven by rare, high-impact variants,[19] while the disease onset at later ages may depend to a greater extent on more common risk variants, which are tested here. However, at the variant level, we found significant differences between the effect-sizes in early-onset and late-onset cases for variants in/near *ZCWPW1* and *APOE ε2,* and also in —opposite directions— for the variant in *MS4A6A*. These results are a first indication that these variants may differentially influence age of disease onset, however, future experiments will have to confirm this finding.

Here, we find that the majority of the observed increase in effect-size in a genetic case-control study of extreme phenotypes is attributable to the extreme controls. We note that the centenarians used in this study were selected for their preserved cognitive health, which might have further enlarged the effect-size increase for genetic variants that were previously identified for their AD-association. We acknowledge that using centenarians as controls in genetic studies of AD could result in the detection of variants associated with extreme longevity, such that newly detected AD-associations need to be verified in an age-matched AD case-control setting. Nevertheless, the effect-sizes for all but two variants are in the same direction as previously reported, which suggests that the tested AD variants do not have significant pleiotropic activities that counteract their AD-related survival effects. Notably, the two variants with an opposite effect in the comparison of the extremes relative to published effect sizes, in or near *MEF2C* and *FERMT2,* also did not associate with AD in our age-matched case-control analysis. This suggests that the AD-association of the *MEF2C* and *FERMT2* variants might be false positive findings in previous studies. This is in line with results from unpublished GWASs of AD in which AD-associations of variants near the *MEF2C* and *FERMT2* genes were not replicated[41,42] (p = 5.4x10^{-3},[41] p = 3.0x10^{-4} for *MEF2C*[42] and p = 1.6x10^{-5} for *FERMT2*[42] variant, with 5.0x10^{-8} being the genome-wide significance threshold). An additional strength of our study is that our cohorts of AD patients and controls, were not previously used in the discovery of any of the known AD associated

346 variants;[4–17] we thus provide independent replication in a genetically homogeneous group of
347 individuals, as they all came from one specific population (Dutch).
348
349 Concluding, in our comparison of cases and controls with extreme phenotypes we found that
350 on average, the effect of AD-related variants in genetic association studies almost doubled,
351 while at the variant level effect-sizes increased up to six-fold. The observed increment in
352 effect-size was largely driven by the centenarians as extreme controls, identifying
353 centenarians as a valuable resource for genetic studies, with possible applications for other
354 age-related diseases.
355

356 **Acknowledgements**

357 Research of the VUmc Alzheimer center is part of the neurodegeneration research program
358 of Amsterdam Neuroscience (www.amsterdamresearch.org). The VUmc Alzheimer Center is
359 supported by Stichting Alzheimer Nederland (WE09.2014-03) and Stiching VUmc fonds. The
360 clinical database structure was developed with funding from Stichting Dioraphte (VSM 14 04
361 14 02). The Dutch case-control study is part of EADB (European Alzheimer DNA biobank)
362 funded by JPcofundNL (ZonMW project number: 733051061). The authors declare no
363 conflict of interests.

References

1. (2012). 2012 Alzheimer's disease facts and figures. Alzheimers Dement. *8*, 131–168.

2. Corrada, M.M., Brookmeyer, R., Paganini-Hill, A., Berlau, D., and Kawas, C.H. (2010). Dementia incidence continues to increase with age in the oldest old: The 90+ study. Ann. Neurol. *67*, 114–121.

3. Gatz, M., Reynolds, C.A., Fratiglioni, L., Johansson, B., Mortimer, J.A., Berg, S., Fiske, A., and Pedersen, N.L. (2006). Role of genes and environments for explaining Alzheimer disease. Arch. Gen. Psychiatry *63*, 168–174.

4. Lambert, J.-C., Heath, S., Even, G., Campion, D., Sleegers, K., Hiltunen, M., Combarros, O., Zelenika, D., Bullido, M.J., Tavernier, B., et al. (2009). Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat. Genet. *41*, 1094–1099.

5. Harold, D., Abraham, R., Hollingworth, P., Sims, R., Gerrish, A., Hamshere, M.L., Pahwa, J.S., Moskvina, V., Dowzell, K., Williams, A., et al. (2009). Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat. Genet. *41*, 1088–1093.

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

7. Desikan, R.S., Schork, A.J., Wang, Y., Thompson, W.K., Dehghan, A., Ridker, P.M., Chasman, D.I., McEvoy, L.K., Holland, D., Chen, C.-H., et al. (2015). Polygenic Overlap Between C-Reactive Protein, Plasma Lipids, and Alzheimer Disease. Circulation *131*, 2061–2069.

8. Sims, R., van der Lee, S.J., Naj, A.C., Bellenguez, C., Badarinarayan, N., Jakobsdottir, J., Kunkle, B.W., Boland, A., Raybould, R., Bis, J.C., et al. (2017). Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. Nat. Genet. *49*, 1373–1384.

391 9. Guerreiro, R., Wojtas, A., Bras, J., Carrasquillo, M., Rogaeva, E., Majounie, E., Cruchaga,
392 C., Sassi, C., Kauwe, J.S.K., Younkin, S., et al. (2013). TREM2 variants in Alzheimer's
393 disease. N. Engl. J. Med. *368*, 117–127.
394 10. Jonsson, T., Stefansson, H., Steinberg, S., Jonsdottir, I., Jonsson, P.V., Snaedal, J.,
395 Bjornsson, S., Huttenlocher, J., Levey, A.I., Lah, J.J., et al. (2013). Variant of TREM2
396 associated with the risk of Alzheimer's disease. N. Engl. J. Med. *368*, 107–116.
397 11. Hollingworth, P., Harold, D., Sims, R., Gerrish, A., Lambert, J.-C., Carrasquillo, M.M.,
398 Abraham, R., Hamshere, M.L., Pahwa, J.S., Moskvina, V., et al. (2011). Common variants at
399 ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's
400 disease. Nat. Genet. *43*, 429–435.
401 12. Naj, A.C., Jun, G., Beecham, G.W., Wang, L.-S., Vardarajan, B.N., Buros, J., Gallins,
402 P.J., Buxbaum, J.D., Jarvik, G.P., Crane, P.K., et al. (2011). Common variants at
403 MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's
404 disease. Nat. Genet. *43*, 436–441.
405 13. Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C.,
406 DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., et al. (2013). Meta-analysis of
407 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat. Genet.
408 *45*, 1452–1458.
409 14. Jun, G., Ibrahim-Verbaas, C.A., Vronskaya, M., Lambert, J.-C., Chung, J., Naj, A.C.,
410 Kunkle, B.W., Wang, L.-S., Bis, J.C., Bellenguez, C., et al. (2016). A novel Alzheimer
411 disease locus located near the gene encoding tau protein. Mol. Psychiatry *21*, 108–117.
412 15. Steinberg, S., Stefansson, H., Jonsson, T., Johannsdottir, H., Ingason, A., Helgason, H.,
413 Sulem, P., Magnusson, O.T., Gudjonsson, S.A., Unnsteinsdottir, U., et al. (2015). Loss-of-
414 function variants in ABCA7 confer risk of Alzheimer's disease. Nat. Genet. *47*, 445–447.
415 16. Strittmatter, W.J., Saunders, A.M., Schmechel, D., Pericak-Vance, M., Enghild, J.,
416 Salvesen, G.S., and Roses, A.D. (1993). Apolipoprotein E: high-avidity binding to beta-
417 amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease.
418 Proc. Natl. Acad. Sci. U. S. A. *90*, 1977–1981.

419 17. Corder, E.H., Saunders, A.M., Risch, N.J., Strittmatter, W.J., Schmechel, D.E., Gaskell,
420 P.C., Rimmler, J.B., Locke, P.A., Conneally, P.M., and Schmader, K.E. (1994). Protective
421 effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. Nat. Genet. *7*, 180–
422 184.
423 18. Rose, S., and van der Laan, M.J. (2009). Why Match? Investigating Matched Case-
424 Control Study Designs with Causal Effect Estimation. Int. J. Biostat. *5*,.
425 19. Lord, J., Lu, A.J., and Cruchaga, C. (2014). Identification of rare variants in
426 Alzheimerâ€™s disease. Front. Genet. *5*,.
427 20. Lee, S., Abecasis, G.R., Boehnke, M., and Lin, X. (2014). Rare-variant association
428 analysis: study designs and statistical tests. Am. J. Hum. Genet. *95*, 5–23.
429 21. Li, D., Lewinger, J.P., Gauderman, W.J., Murcray, C.E., and Conti, D. (2011). Using
430 extreme phenotype sampling to identify the rare causal variants of quantitative traits in
431 association studies. Genet. Epidemiol. *35*, 790–799.
432 22. Barnett, I.J., Lee, S., and Lin, X. (2013). Detecting rare variant effects using extreme
433 phenotype sampling in sequencing association studies. Genet. Epidemiol. *37*, 142–151.
434 23. Garagnani, P., Giuliani, C., Pirazzini, C., Olivieri, F., Bacalini, M.G., Ostan, R., Mari, D.,
435 Passarino, G., Monti, D., Bonfigli, A.R., et al. (2013). Centenarians as super-controls to
436 assess the biological relevance of genetic risk factors for common age-related diseases: a
437 proof of principle on type 2 diabetes. Aging *5*, 373–385.
438 24. Peloso, G.M., Rader, D.J., Gabriel, S., Kathiresan, S., Daly, M.J., and Neale, B.M.
439 (2016). Phenotypic extremes in rare variant study designs. Eur. J. Hum. Genet. *24*, 924–
440 930.
441 25. van der Flier, W.M., Pijnenburg, Y.A.L., Prins, N., Lemstra, A.W., Bouwman, F.H.,
442 Teunissen, C.E., van Berckel, B.N.M., Stam, C.J., Barkhof, F., Visser, P.J., et al. (2014).
443 Optimizing patient care and research: the Amsterdam Dementia Cohort. J. Alzheimers Dis.
444 JAD *41*, 313–327.
445 26. van der Flier, W.M., and Scheltens, P. (2018). Amsterdam Dementia Cohort: Performing
446 Research to Optimize Care. J. Alzheimers Dis. *62*, 1091–1111.

447 27. Varma, A.R., Snowden, J.S., Lloyd, J.J., Talbot, P.R., Mann, D.M., and Neary, D. (1999).
448 Evaluation of the NINCDS-ADRDA criteria in the differentiation of Alzheimer's disease and
449 frontotemporal dementia. J. Neurol. Neurosurg. Psychiatry *66*, 184–188.
450 28. Blacker, D., Albert, M.S., Bassett, S.S., Go, R.C., Harrell, L.E., and Folstein, M.F. (1994).
451 Reliability and validity of NINCDS-ADRDA criteria for Alzheimer's disease. The National
452 Institute of Mental Health Genetics Initiative. Arch. Neurol. *51*, 1198–1204.
453 29. Holstege, H., Beker, N., Dijkstra, T., Pieterse, K., Wemmenhove, E., Schouten, K.,
454 Thiessens, L., Horsten, D., Rechtuijt, S., Sikkes, S., et al. (2018). The 100-plus Study of
455 Dutch cognitively healthy centenarians: rationale, design and cohort description.
456 30. Huisman, M., Poppelaars, J., van der Horst, M., Beekman, A.T., Brug, J., van Tilburg,
457 T.G., and Deeg, D.J. (2011). Cohort Profile: The Longitudinal Aging Study Amsterdam. Int. J.
458 Epidemiol. *40*, 868–876.
459 31. Hoogendijk, E.O., Deeg, D.J.H., Poppelaars, J., van der Horst, M., Broese van Groenou,
460 M.I., Comijs, H.C., Pasman, H.R.W., van Schoor, N.M., Suanet, B., Thomése, F., et al.
461 (2016). The Longitudinal Aging Study Amsterdam: cohort update 2016 and major findings.
462 Eur. J. Epidemiol. *31*, 927–945.
463 32. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew,
464 E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype imputation service and
465 methods. Nat. Genet. *48*, 1284–1287.
466 33. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang,
467 H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al. (2016). A reference panel of 64,976
468 haplotypes for genotype imputation. Nat. Genet. *48*, 1279–1283.
469 34. O'Connell, J., Gurdasani, D., Delaneau, O., Pirastu, N., Ulivi, S., Cocca, M., Traglia, M.,
470 Huang, J., Huffman, J.E., Rudan, I., et al. (2014). A general approach for haplotype phasing
471 across the full spectrum of relatedness. PLoS Genet. *10*, e1004234.
472 35. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison,
473 E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A
474 global reference for human genetic variation. Nature *526*, 68–74.

475 36. Anderson, C.A., Pettersson, F.H., Clarke, G.M., Cardon, L.R., Morris, A.P., and
476 Zondervan, K.T. (2010). Data quality control in genetic case-control association studies. Nat.
477 Protoc. *5*, 1564–1573.
478 37. Price, A.L., Zaitlen, N.A., Reich, D., and Patterson, N. (2010). New approaches to
479 population stratification in genome-wide association studies. Nat. Rev. Genet. *11*, 459–463.
480 38. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D.
481 (2006). Principal components analysis corrects for stratification in genome-wide association
482 studies. Nat. Genet. *38*, 904–909.
483 39. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller,
484 J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome
485 association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
486 40. R Core Team R: A language and environment for statistical computing. *R Foundation for*
487 *Statistical Computing*, *Vienna*, *Austria.*,.
488 41. Marioni, R., Harris, S.E., McRae, A.F., Zhang, Q., Hagenaars, S.P., Hill, W.D., Davies,
489 G., Ritchie, C.W., Gale, C., Starr, J.M., et al. (2018). GWAS on family history of Alzheimer's
490 disease.
491 42. Jansen, I., Savage, J., Watanabe, K., Bryois, J., Williams, D., Steinberg, S., Sealock, J.,
492 Karlsson, I., Hagg, S., Athanasiu, L., et al. (2018). Genetic meta-analysis identifies 9 novel
493 loci and functional pathways for Alzheimers disease risk.
494 43. Van Cauwenberghe, C., Van Broeckhoven, C., and Sleegers, K. (2016). The genetic
495 landscape of Alzheimer disease: clinical implications and perspectives. Genet. Med. *18*,
496 421–430.
497 44. Saunders, A.M., Strittmatter, W.J., Schmechel, D., George-Hyslop, P.H., Pericak-Vance,
498 M.A., Joo, S.H., Rosi, B.L., Gusella, J.F., Crapper-MacLachlan, D.R., and Alberts, M.J.
499 (1993). Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic
500 Alzheimer's disease. Neurology *43*, 1467–1472.
501 45. Kraja, A.T., Borecki, I.B., Tsai, M.Y., Ordovas, J.M., Hopkins, P.N., Lai, C.-Q., Frazier-
502 Wood, A.C., Straka, R.J., Hixson, J.E., Province, M.A., et al. (2013). Genetic analysis of 16

503 NMR-lipoprotein fractions in humans, the GOLDN study. Lipids *48*, 155–165.

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

505 1323–1338.

506 47. vB Hjelmborg, J., Iachine, I., Skytthe, A., Vaupel, J.W., McGue, M., Koskenvuo, M.,

507 Kaprio, J., Pedersen, N.L., and Christensen, K. (2006). Genetic influence on human lifespan

508 and longevity. Hum. Genet. *119*, 312–321.

509 48. Joshi, P.K., Pirastu, N., Kentistou, K.A., Fischer, K., Hofer, E., Schraut, K.E., Clark,

510 D.W., Nutile, T., Barnes, C.L.K., Timmers, P.R.H.J., et al. (2017). Genome-wide meta-

511 analysis associates HLA-DQA1/DRB1 and LPA and lifestyle factors with human longevity.

512 Nat. Commun. *8*,.

513 49. Ryu, S., Atzmon, G., Barzilai, N., Raghavachari, N., and Suh, Y. (2016). Genetic

514 landscape of APOE in human longevity revealed by high-throughput sequencing. Mech.

515 Ageing Dev. *155*, 7–9.

516 50. Broer, L., Buchman, A.S., Deelen, J., Evans, D.S., Faul, J.D., Lunetta, K.L., Sebastiani,

517 P., Smith, J.A., Smith, A.V., Tanaka, T., et al. (2015). GWAS of longevity in CHARGE

518 consortium confirms APOE and FOXO3 candidacy. J. Gerontol. A. Biol. Sci. Med. Sci. *70*,

519 110–118.

520 51. Sebastiani, P., Gurinovich, A., Bae, H., Andersen, S., Malovini, A., Atzmon, G., Villa, F.,

521 Kraja, A.T., Ben-Avraham, D., Barzilai, N., et al. (2017). Four Genome-Wide Association

522 Studies Identify New Extreme Longevity Variants. J. Gerontol. Ser. A *72*, 1453–1464.

523 52. Ward, L.D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states,

524 conservation, and regulatory motif alterations within sets of genetically linked variants.

525 Nucleic Acids Res. *40*, D930–D934.

526

Tables and Figures

Table 1: Population characteristics

	Extreme AD Cases (*EA*)	**Centenarian controls (*EC*)**	**Normal controls (*NC*)**
Number of individuals	1,073	255	1,664
Females (%)	564 (52.6)	191 (74.9)	893 (53.7)
Age (SD)*	66.4 (7.8)	101.4 (1.3)	66.0 (6.5)
ApoE ε4 (%)	981 (42.7)	44 (8.6)	533 (16.0)
ApoE ε2 (%)	76 (3.5)	78 (15.3)	304 (9.1)

*Age at onset for extreme Alzheimer's disease cases, age at study inclusion for extreme controls and normal controls; Abbreviations: SD, standard deviation; *ApoE*, Apolipoprotein E allele count for *ε4* and *ε2*, respectively; Reference to the cohorts reported in this table are: [25,26,29,30]

533 **Table 2:** Association statistics of the 29 tested AD-associated variants

Chr	*Position*	*Rs ID*	*Gene*	*A1*	$\log OR_l^k$ *(SE)*	$\log OR_{EA-EC}^k$ *(SE)*	*P-value*$_{EA\text{-}EC}$	E_{EA-EC}^k *(95% CI, p)*	*AF*$_{EA}$	*AF*$_{NC}$	*AF*$_{EC}$
6	41,129,252	rs75932628	*TREM2 (R47H)*	T	0.89 (0.09)	5.75 (5.13)	$2.63x10^{-1}$	6.46 (-4.83 - 18.46, $3.5x10^{-1}$)	0.003	0.001	0.000
10	**11,720,308**	**rs7920721**	***ECHDC3***	**G**	**0.07 (0.01)**	**0.31 (0.10)**	**$2.93x10^{-3}$ ***	**4.56 (1.55 - 8.93, $1.8x10^{-2}$)**	**0.430**	**0.389**	**0.357**
14	**92,926,952**	**rs10498633**	***SLC24A4-RIN3***	**T**	**-0.09 (0.01)**	**-0.42 (0.11)**	**$1.30x10^{-4}$ ***	**4.50 (2.08 - 7.93, $2.8x10^{-3}$)**	**0.206**	**0.236**	**0.292**
7	**37,841,534**	**rs2718058**	***NME8***	**G**	**-0.08 (0.01)**	**-0.29 (0.10)**	**$3.72x10^{-3}$ ***	**3.80 (1.17 - 7.28, $3.3x10^{-2}$)**	**0.360**	**0.367**	**0.433**
16	**81,942,028**	**rs72824905**	***PLCG2***	**G**	**-0.39 (0.06)**	**-1.27 (0.40)**	**$1.38x10^{-3}$ ***	**3.28 (1.26 - 5.98, $2.8x10^{-2}$)**	**0.008**	**0.012**	**0.025**
6	32,578,530	rs9271192	*HLA-DRB1*	A	-0.11 (0.01)	-0.35 (0.16)	$3.06x10^{-2}$ *	3.20 (0.35 - 6.65, $1.3x10^{-1}$)	0.712	0.727	0.780
7	100,004,446	rs1476679	*ZCWPW1*	T	0.09 (0.01)	0.26 (0.11)	$1.34x10^{-2}$ *	2.97 (0.60 - 6.10, $1.0x10^{-1}$)	0.703	0.674	0.649
19	1,063,443	rs4147929	*ABCA7 (A>G)*	G	-0.14 (0.02)	-0.32 (0.14)	$2.11x10^{-2}$ *	2.26 (0.30 - 4.42, $2.2x10^{-1}$)	0.809	0.834	0.855
19	**45,412,079**	**rs7412**	***APOE (ε2)***	**T**	**-0.79 (0.03)**	**-1.76 (0.18)**	**$3.16x10^{-21}$ ***	**2.24 (1.75 - 2.77, $1.4x10^{-7}$)**	**0.033**	**0.091**	**0.149**
4	11,711,232	rs13113697	*HS3ST1*	G	-0.07 (0.01)	-0.14 (0.12)	$2.41x10^{-1}$	2.06 (-1.49 - 6.13, $5.4x10^{-1}$)	0.265	0.268	0.247
17	47,297,297	rs616338	*ABI3*	C	-0.36 (0.05)	-0.74 (0.57)	$1.93x10^{-1}$	2.06 (-0.99 - 5.59, $5.2x10^{-1}$)	0.017	0.009	0.006
6	47,487,762	rs10948363	*CD2AP*	G	0.10 (0.01)	0.19 (0.11)	$8.84x10^{-2}$	2.00 (-0.34 - 4.60, $4.1x10^{-1}$)	0.284	0.272	0.245
19	**45,411,941**	**rs429358**	***APOE (ε4)***	**C**	**1.05 (0.03)**	**2.08 (0.17)**	$1.31x10^{-33}$ *	**1.99 (1.65 - 2.33, $1.5x10^{-9}$)**	**0.429**	**0.166**	**0.082**
7	143,110,762	rs11771145	*EPHA1*	A	-0.10 (0.01)	-0.20 (0.10)	$5.96x10^{-2}$	1.94 (-0.09 - 4.29, $3.7x10^{-1}$)	0.325	0.345	0.371
11	47,557,871	rs10838725	*CELF1*	C	0.08 (0.01)	0.14 (0.11)	$2.05x10^{-1}$	1.78 (-0.95 - 5.11, $5.8x10^{-1}$)	0.328	0.314	0.302
8	27,195,121	rs28834970	*PTK2B*	C	0.10 (0.01)	0.18 (0.10)	$8.96x10^{-2}$	1.76 (-0.23 - 4.09, $4.7x10^{-1}$)	0.395	0.376	0.353
11	59,923,508	rs983392	*MS4A6A*	G	-0.11 (0.01)	-0.17 (0.10)	$9.39x10^{-2}$	1.56 (-0.20 - 3.61, $5.4x10^{-1}$)	0.397	0.403	0.439
11	121,435,587	rs11218343	*SORL1*	C	-0.26 (0.03)	-0.39 (0.25)	$1.21x10^{-1}$	1.48 (-0.39 - 3.51, $6.2x10^{-1}$)	0.033	0.040	0.047
2	127,892,810	rs6733839	*BIN1*	T	0.20 (0.01)	0.25 (0.10)	$1.12x10^{-2}$ *	1.28 (0.31 - 2.29, $5.8x10^{-1}$)	0.456	0.413	0.390
11	85,867,875	rs10792832	*PICALM*	G	0.14 (0.01)	0.15 (0.10)	$1.26x10^{-1}$	1.09 (-0.30 - 2.56, $9.1x10^{-1}$)	0.653	0.614	0.612
20	55,018,260	rs7274581	*CASS4*	C	-0.13 (0.02)	-0.14 (0.18)	$4.41x10^{-1}$	1.06 (-1.83 - 4.07, $9.7x10^{-1}$)	0.075	0.088	0.084
6	41,129,207	rs143332484	*TREM2 (R62H)*	T	0.50 (0.07)	0.48 (0.48)	$3.21x10^{-1}$	0.97 (-0.96 - 3.09, $9.8x10^{-1}$)	0.017	0.015	0.009
17	44,353,222	rs118172952	*KANSL1*	G	-0.14 (0.03)	-0.13 (0.14)	$3.44x10^{-1}$	0.97 (-1.08 - 3.64, $9.6x10^{-1}$)	0.191	0.202	0.221
1	207,692,049	rs6656401	*CR1*	G	-0.17 (0.01)	-0.12 (0.12)	$3.11x10^{-1}$	0.75 (-0.75 - 2.21, $7.4x10^{-1}$)	0.781	0.803	0.806
19	1,061,892	rs200538373	*ABCA7 (G>C)*	C	-0.65 (0.14)	-0.44 (0.80)	$5.81x10^{-1}$	0.68 (-1.83 - 3.54, $7.9x10^{-1}$)	0.004	0.004	0.006
8	27,467,686	rs9331896	*CLU*	T	0.15 (0.01)	0.09 (0.10)	$3.99x10^{-1}$	0.60 (-0.78 - 2.06, $5.8x10^{-1}$)	0.361	0.400	0.378
2	234,068,476	rs35349669	*INPP5D*	T	0.08 (0.01)	0.03 (0.10)	$7.83x10^{-1}$	0.36 (-2.33 - 3.16, $6.2x10^{-1}$)	0.474	0.496	0.486
14	53,400,629	rs17125944	*FERMT2*	C	0.13 (0.02)	-0.11 (0.16)	$4.99x10^{-1}$	-0.82 (-3.46 - 1.60, $1.3x10^{-1}$)	0.104	0.105	0.114
5	**88,223,420**	**rs190982**	***MEF2C***	**A**	**0.08 (0.01)**	**-0.14 (0.10)**	**$1.70x10^{-1}$**	**-1.86 (-5.01 - 0.77, $3.3x10^{-2}$)**	**0.408**	**0.406**	**0.372**
		AVERAGE						1.90 ± 0.29, $p = 9.0x10^{-4}$			

Abbreviations: *Chr,* chromosome: *Position,* chromosomal position; *Rs ID*, variant ID; *Gene*, gene associated with the variant according to paper in which the variant was found; *A1*, allele tested; $\log OR_l^k$*(SE)*, log(odds ratio) and relative standard error for variant *k* reported by study with largest sample size; $\log OR_{EA-EC}^k$ (*SE*), log(odds ratio) and relative standard error in extreme-control association; *P-value*, *p*-value of AD-association of extreme AD cases *vs.* centenarian controls; E_{EA-EC}^k *(95% CI, p)*, change in effect-size, 95% confidence intervals and p-value of difference when using extreme phenotypes relative to published effect-sizes; AF_{EA}, tested allele frequency in AD extreme cases; AF_{NC}, tested allele frequency in normal controls; AF_{EC}, tested allele frequency in centenarian controls. Bold: variants for which the E_{EA-EC}^k was significantly different from published effect-size; Stars*: significant at $p<0.05$.

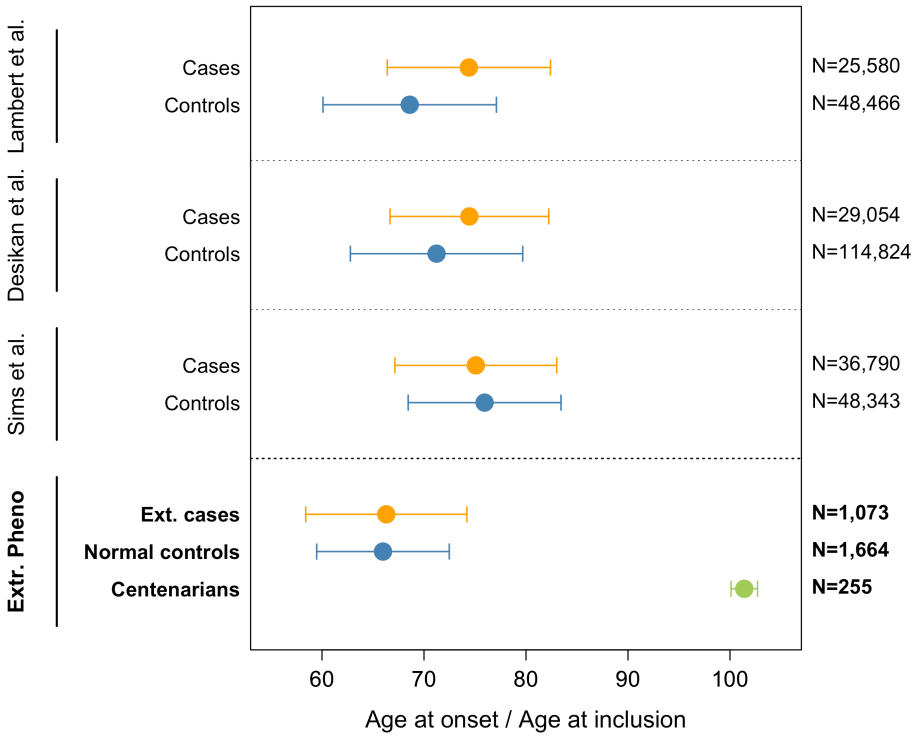

***Figure 1*: Comparison of age at disease-onset and age at inclusion for cases and controls in previously reported case-control comparisons, and in our extreme phenotypes comparison.** Weighted mean and (combined) standard deviation of the age at onset for AD cases and age at inclusion for controls. As weights, we used the sample sizes of each GWA study. Note that previous case-control studies of AD included samples from multiple cohorts, sometimes overlapping across studies. References to the cohorts reported in this figure are: [7,8,13,25,26,30]

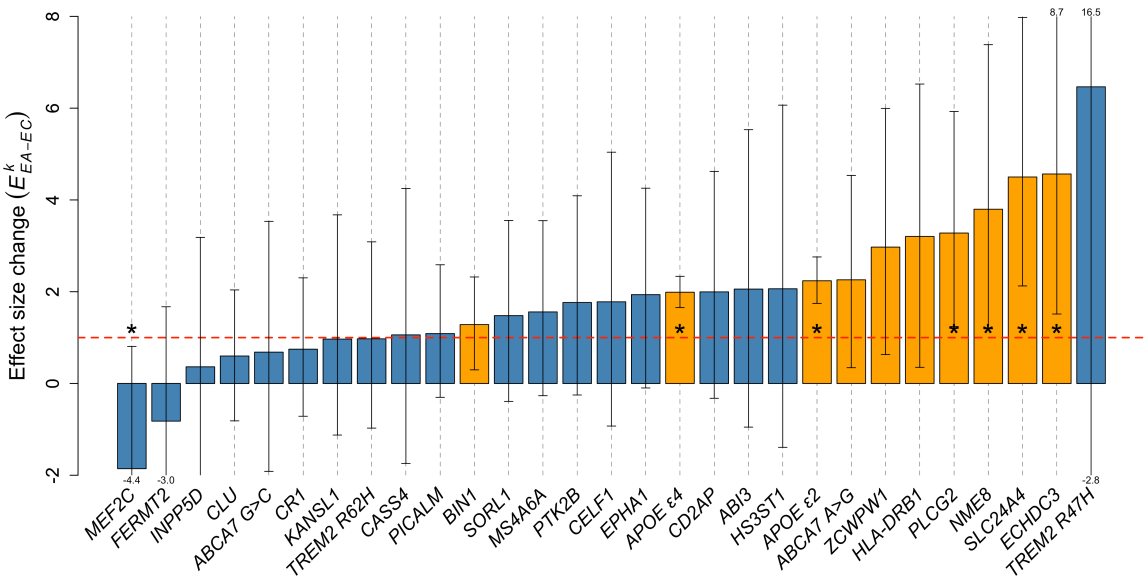

***Figure 2*: Change in variant effect-size using extreme cases and centenarian controls relative to published effect-sizes, for 29 AD associated genetic variants.** Dashed red line at E^k_{EA-EC}=1 indicates same effect-size as reported in literature. Orange bars indicate nominal statistical significance for the association with AD (p<0.05). Stars indicate significant changes of effect-size relative to previously reported effect-sizes (p<0.05, two-sample z-test).

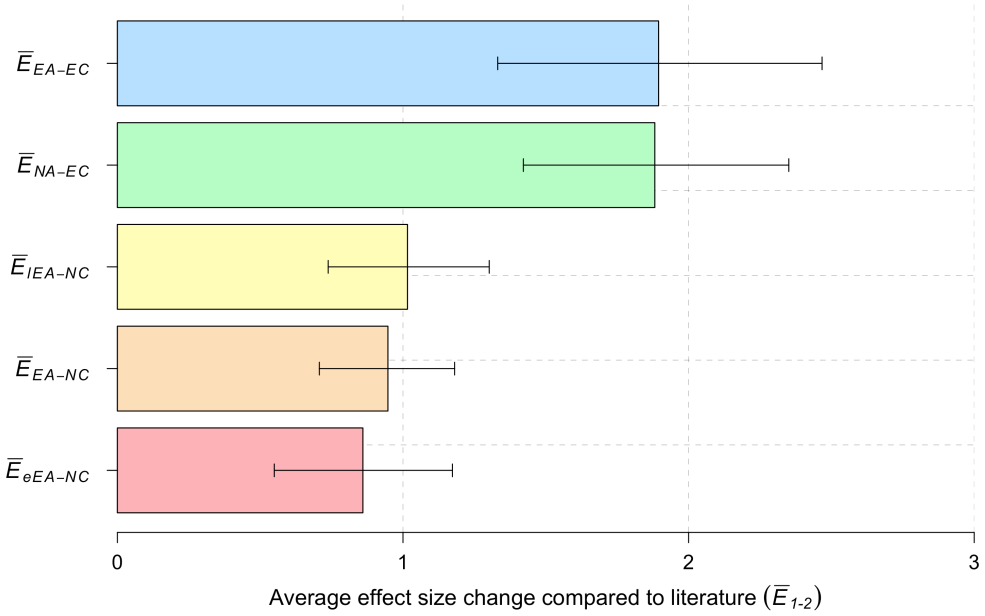

***Figure 3:* Average increase in effect-size for the different comparisons.** Average increase in effect sizes for: Extreme AD cases (N_{EA}=1,073), of which early onset cases (N_{eEA}=464) late onset cases (N_{lEA}=609); centenarian controls (N_{EC}=255); normal controls (N_{NC}=1,664). 95% confidence intervals were estimated by random sampling (S=10,000).