The escape from Alzheimer's disease underlies genetic variants involved into immune response and endocytosis pathways

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 M. van Schoor [5, 6], Philip Scheltens [1], Wiesje M. van der Flier [1, 5], Martijn Huisman [5, 6], Marcel J. T. Reinders [3], and Henne Holstege [1, 2, 3]

[1] *Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands*

[2] *Department of Clinical Genetics, 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, Amsterdam Neuroscience, VU University, Amsterdam, the Netherlands.*

[5] *Department of Epidemiology and Biostatistics, Amsterdam UMC, Amsterdam, The Netherlands*

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

Keywords: centenarians, immune response, endosomal system, Alzheimer's disease, pathways, polygenic risk score, PRS, longevity, resilience, cognitive health, etiology

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

Abstract

The risk to develop and escape Alzheimer's disease (AD) is influenced by a constellation of genetic variants, each associated with specific molecular pathways. Different pathways may differentially contribute to the modification of the AD-risk. We studied the molecular mechanisms that explain the extreme ends of the cognitive spectrum by comparing pathway-specific polygenic risk scores (pathway-PRS) in individuals with AD and those who escaped AD until old age.

We used 29 genetic variants associated with AD to calculate pathway-PRS for five major pathways involved in AD. We developed an integrative framework that allows multiple genes to associate with a variant, and multiple pathways to associate with a gene. We studied pathway-PRS in patients with AD (N=1,909), population controls (N=1,654), and cognitively healthy centenarians who escaped AD (N=293). Last, we estimated the contribution of each pathway to the genetic risk of AD in the general population.

All pathway-PRS significantly associated with the modification of AD-risk ($p<0.05$). The pathway that contributed the most was β-amyloid metabolism (32%), driven mainly by *APOE* variants. After excluding *APOE* variants, all pathway-PRS associated with increased AD-risk ($p<0.05$), while specifically immune response ($p=3.1\times10^{-3}$) and endocytosis ($p=3.8\times10^{-4}$) associated with escaping AD. These pathways were the main contributors to the genetic risk of AD (41.3% and 21.4%, respectively), and their effect on escaping AD was larger or comparable to that of developing AD.

Our work suggests that immune response and endocytosis might be involved in general neuro-protective functions, and highlights the need to study these pathways, next to β-amyloid metabolism.

Introduction

Due to changes in lifestyle and advances in healthcare, life expectancy has greatly increased during the last century.[1] A consequence of an increased fraction of aged individuals in the population is the increased prevalence of age-related diseases. A major contribution to poor health and disability at old age is cognitive decline due to Alzheimer's disease (AD).[2] The incidence of AD increases exponentially with age and reaches ~40% per year at 100 years, making it one of the most prevalent diseases in the elderly.[3] Yet, a small proportion of the population (<0.1%) escapes the disease, reaching at least 100 years while maintaining a high level of cognitive health.[4]

Both the development and the escape from AD are determined by a combination of beneficial and harmful environmental and genetic factors that is unique for each individual.[1, 5, 6] Thus far, large collaborative genome-wide association studies (GWAS) have discovered common genetic variants associated with a small modification of the risk of AD.[7–20] Of these, the alleles that encompass the *APOE* gene explain the largest proportion of the risk to develop or the chance to escape AD. We previously showed that those who escaped cognitive decline until extreme ages (cognitively healthy centenarians) were relatively depleted with genetic variants associated with an increased risk of AD.[21] However, the degree of depletion of these variants in the genomes of cognitively healthy centenarians relative to the middle-aged healthy individuals was not constant, which might point towards a differential impact of associated biological pathways on either escaping or developing AD. This led us to hypothesize that an individuals' chance to develop or escape AD may be determined by pathway-specific risk.

Previous enrichment analyses of GWAS summary statistics indicated that five specific biological pathways associate strongly with AD risk: *immune response*, *ß-amyloid metabolism*, *cholesterol/lipid metabolism*, *endocytosis* and *vascular dysfunction*.[22–27] It is not known the extent to which different pathways contribute to the polygenic risk of AD. However, the degree to which pathways functionality may contribute to the individual risk can be studied with pathway-specific polygenic risk scores (PRS).[28, 29] In a typical polygenic risk score, the effect-sizes of all genetic variants that significantly associate with a trait are combined.[30] In a pathway-specific PRS, however, additional information is necessary, namely: (*i*) the association of genetic variants to genes, and (*ii*) the association of genes to pathways. Previous studies of pathway-PRS in AD approached these challenges using the closest gene for variant mapping. For this, a 1:1 relationship between variants and genes is assumed, however, as AD-associated variants are mostly intronic or intergenic, the closest gene is not necessarily the gene affected by the variant. Additionally, different databases often have different functional annotations of genes, and this uncertainty is usually not taken into account when constructing pathway-PRS.[28, 29]

An accurate mapping of the genetic risk of AD conferred by specific molecular pathways may lead to a greater comprehension of individual AD subtypes and might represent a first important step for the development of targeted intervention strategies and personalized medicine.[31] Here, we propose a novel integrative framework to construct pathway-PRS for the five major pathways suggested to be involved in AD. We then tested whether specific pathways differentially contributed to the risk of AD as well as to the chance of escaping AD until extreme old ages. Finally, we estimated the contribution of each pathway to the polygenic risk of AD in the general (healthy middle-aged) population.

Methods

Populations

Population subjects are denoted by *P*: they consist of a representative Dutch sample of 1,779 individuals aged 55-85 years from the Longitudinal Aging Study Amsterdam (LASA).[32, 33] Patients diagnosed with AD are denoted by *A*. The patients are either clinically diagnosed probable AD patients from the Amsterdam Dementia Cohort (N=1,630) or pathologically confirmed AD patients from the Netherlands Brain Bank (N=436).[34–36] Escapers of AD are denoted by *C*: these are 302 cognitively healthy centenarians from the 100-plus Study cohort. This study includes individuals who can provide official evidence for being aged 100 years or older and self-report to be cognitively healthy, which is confirmed by a proxy.[4] All participants and/or their legal representatives provided written informed consent for participation in clinical and genetic studies. The Medical Ethics Committee of the Amsterdam UMC (METC) approved all studies.

Genotyping and imputation

We selected 29 common genetic variants (minor allele frequency >1%) for which a genome-wide significant association with clinically identified AD cases was found (*Table S1*).[7–18, 37–40] All cohorts were genotyped on the same commercial genotyping array.[21] Genetic variants were determined by standard genotyping or imputation methods and we applied established quality control methods.[21, 41] A total of 1,779 population subjects, 302 centenarians and 2,066 AD cases passed quality control. Prior to analysis, we excluded individuals of non-European ancestry ($N_C = 2$, $N_P = 63$ and $N_A = 94$ based on 1000Genomes clustering)[42] and individuals with a family relation ($N_C = 7$, $N_P = 62$ and $N_A = 63$, identity-by-descent >0.3), leaving 1,654 population subjects, 293 cognitively healthy centenarians and 1,909 AD cases for the analyses.

Polygenic risk score

To calculate the personal polygenic risk score, or the genetic risk of AD that affects a single individual, the effect-sizes of all genetic variants that significantly associate with AD are combined. Formally, a PRS is defined as the sum of trait-associated alleles carried by an individual across a defined set of genetic loci, weighted by effect-sizes estimated from a GWAS.[30] We constructed a polygenic risk score (PRS) using 29 variants that were previously associated with AD. As weights for the PRS, we used the variant effect-sizes (log of odds ratio) as published in large GWAS of AD (*Table S1*). Given a subject *s*, the PRS is defined as:

$$PRS^s = \sum_k^K (dos_k^s * \beta_k) \quad (1)$$

where K is the full set of variants, dos_k^s is the allele dosage from the (imputed) genotype of variant k in subject s and β_k is the variant effect size as derived from literature (*Table S1*).

Mapping variants to pathways

We studied the five pathways implicated in AD: *immune response, ß-amyloid metabolism*, *cholesterol/lipid metabolism*, *endocytosis* and *vascular dysfunction*.[22–27] For these pathways we developed the *variant-pathway mapping* M_p^k, which represents the degree of involvement of a given variant in the pre-selected pathways. To generate this value, we (*i*) associated genetic variants to genes (*variant-gene mapping*), (*ii*) associated genes to pathways (*gene-pathway mapping*) and (*iii*) combined these mappings in the *variant-pathway mapping*.

Variant-gene mapping: the association of a variant with a specific gene is not straight-forward as the closest gene is not necessarily the gene affected by the variant. The two most recent and largest GWAS of AD addressed the relationship between genetic variants and associated genes applying two independent methods.[19, 20] Briefly, one study used (*i*) gene-based annotation, (*ii*) expression-quantitative trait loci (eQTL) analyses, (*iii*) gene cluster/pathway analyses, and (*iv*) differential gene expression analysis between AD cases and healthy controls.[19] The other study integrated (*i*) positional mapping, (*ii*) eQTL gene-mapping, and (*iii*) chromatin interaction as implemented in the tool Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA).[20, 43] The list of genes most likely affected by each variant was obtained from both studies and used to derive a weighted mapping for each genetic variant k to one or more genes g, m_g^k, denoted as the *variant-gene mapping* weight. This weight was calculated by counting the number of times a variant k was associated with gene g across the two studies and dividing this by the total number of genes associated with the variant.

Gene-pathway mapping: each gene from the *variant-gene mapping* was classified into the pre-defined set of pathways integrating four sources of information:

1. Gene-sets from the pathway analysis of *Kunkle et al.,*[19] in which the authors identified 9 significant pathways (coupled with the genes involved in each pathway), which we mapped to 3 of the 5 pathways of interest (*Table S2*);
2. Associated genes from Gene-ontology (GO) terms resembling the 5 pathways of interest within the biological processes tree (including all child-terms) (*Table S3*);[44, 45]
3. Gene-sets derived from the functional clustering within DAVID:[46, 47] the gene-set from the *variant-gene mapping* was used to obtain 15 functional clusters which were then mapped to the 5 pre-selected pathways using a set of keywords (*Table S4* and *Table S5*);
4. Gene-pathway associations from a recent review concerning the genetic landscape of AD (*Table S6*);[22]

By counting the number of times each gene was associated to each pathway according to these sources, and dividing by the total number of associations per gene, we obtained a weighted mapping of each gene *g* to one or more pathways *p*, w_p^g, denoted as the *gene-pathway mapping* weight. In case the *gene-pathway mapping* could not be calculated (*i.e* there was no mapping to any of the pathways of consideration), we excluded the gene from further analyses (*Table S7*).

To associate variants with pathways, we combined the *variant-gene mapping* and the *gene-pathway mapping.* Given a variant *k,* mapping to a set of genes *G*, and a pathway *p*, we define the weight of the variant to the pathway (M_p^k) as:

$$M_p^k = \sum_g^G (m_g^k * w_p^g) \quad (2)$$

where m_g^k is the *variant-gene mapping* weight of variant *k* to gene *g*, and w_p^g is the *gene-pathway mapping* weight of gene *g* to pathway *p*. In this way, for each variant, we calculated a score indicative of the involvement of the variant in each of the five pathways (*variant-pathway mapping*). For some variants no *variant-pathway mapping* was possible. We marked these variants as unmapped.

Pathway-specific polygenic risk score

For the pathway-specific polygenic risk score (pPRS), we extended the definition of the PRS by adding as multiplicative factor the *variant-pathway mapping* weight of each variant. Given a sample *s* and a pathway *p*, we defined the pPRS as:

$$pPRS_p^s = \sum_k^K (dos_k^s * \beta_k * M_k^p) \quad (3)$$

where M_k^p is the *variant-pathway mapping* of variant *k* to pathway *p*.

Association of PRSs in the three cohorts

We calculated the polygenic risk score (PRS) and pathway-PRS (pPRS) for the population subjects, the AD cases and the cognitively healthy centenarians (escapers of AD) (*P,* A and *C*, respectively). In addition, we interrogated the influence of the *APOE* variants on these risk scores by calculating the PRSs and pPRSs with and without *APOE* variants. To investigate the differential contributions of the risk scores to AD development as well as escape from AD, we calculated (*i*) the association of the risk scores (PRS and pPRS) with AD status by comparing AD cases and population subjects (*A vs. P*), and (*ii*) the association of the risk scores with escaping AD until extreme ages by comparing cognitively healthy centenarians and population subjects (*C vs. P* comparison). For the associations, we used logistic regression models with the PRS and pPRS as predictors. The PRS were scaled (mean=0, SD=1) and therefore the effect-sizes (log of odds ratio), can be interpreted as the odds ratio difference per one standard deviation (SD) increase in the PRS, with corresponding estimated 95% confidence intervals (95% CI).

To further investigate the relationship between the effect of each pathway on AD and on escaping AD, we calculated the *change* in effect-size as the ratio between the effect-size of the association with escaping AD (log of odds ratios of *C vs. P* comparison) and the effect-size of the association with AD (log of odds ratios of *A vs. P* comparison). A *change* larger than 1 is indicative of a more pronounced effect on escaping AD, whereas a *change* smaller than 1 indicates a more pronounce AD effect.

Contribution of each pathway to polygenic risk of AD

We estimated the contribution of each pathway to the genetic risk of AD in the general population: this is the ratio between the variance of each pathway-PRS and the variance of the combined PRS as calculated in the individuals in the general population. As such, it is a function of the *variant-pathway mapping*, the effect-size (log of odds ratio) of the variants, and the variant frequencies. Given a variant *k* and the relative *variant-pathway mapping* M_k^p, we define the percentage of the risk explained by each pathway *p* as:

$$P^p = \frac{\sum_k^K (M_k^p * (\beta_k^2 * MAF_k * (1 - MAF_k)))}{\sum_k^K (\beta_k^2 * MAF_k * (1 - MAF_k))} \quad (4)$$

where β_k is the variant effect-size from literature, and $MAF_k * (1 - MAF_k)$ is the variance of a Bernoulli random variable that occurs with probability MAF_k, *i.e* the minor allele frequency of each variant *k* in our cohort of middle-aged healthy population subjects. Here, M_k^p is interpreted as the probability that variant *k* belongs to pathway *p*. When calculating the contributions of each pathway, we also considered variants with missing *variant-pathway mapping*. For these variants, the *variant-pathway mapping* was set to 1 for an *unmapped* pathway. Together, the pathway PRS variances sum to the total PRS variance.

Implementation

We performed quality control of genotype data as well as population stratification analysis and relatedness analysis with PLINK (v2.0). All subsequent analyses were performed with R (v3.5.2), Bash and Python (v2.7.14) scripts.

Results

After quality control of the genetic data, we included 1,654 population subjects (with mean age at inclusion 62.7±6.4, 53.2% females), 1,909 AD cases (with mean age at onset 69.2±9.9, 56.4% females), and 293 cognitively healthy centenarians (with mean age at inclusion 101.4±1.3, 72.6% females) (*P, A* and C respectively, *Table 1*).

Polygenic risk scores associate with AD and escape from AD

To each subject, we assigned a PRS representative of all 29 AD-associated variants, including and excluding *APOE* variants. We found that the PRS, when including *APOE* variants, significantly associated with an increased risk of AD and, in the opposite direction, with increased chance of escaping AD (*A vs. P: OR*=2.61, 95% CI=[2.41-2.82], p=9.7x10^{-112} and *C vs. P: OR*=0.58, 95% CI=[0.50-0.68], p=7.5x10^{-11}) (*Figure 1A* and *Table S7*). When excluding *APOE* variants, the PRS was still significantly associated with an increased risk of AD and, in the opposite direction, with increased risk of escaping AD (*A vs. P: OR*=1.30, 95% CI=[1.23-1.38], p=7.3x10^{-14} and *C vs. P: OR=0.79*, 95% CI=[0.70-0.89], p=3.2x10^{-4}) (*Figure 1B* and *Table S7*).

Pathway-specific PRS associate with AD and escape from AD

Subsequently, we annotated the 29 AD-associated genetic variants to 5 pathways *(Figure 2)*. According to our *variant-gene mapping*, the 29 AD-associated variants mapped to 155 genes (*Table S8*). The number of genes associated with each variant ranged from 1 (*e.g.* for variants in/near *SORL1*, *APOE* or *PLCG2*), to 34 (a variant in the gene-dense region near *ZCWPW1* gene) (*Figure 2* and *Table S9*). We were able to calculate the *gene-pathway mapping* weight for 74 genes (*Table S10*). The remaining 81 genes were not mapped to the 5 pathways (*Table S8*). In total, we calculated the *variant-pathway mapping* for 23 loci to at least one of the pre-selected biological pathways (*Figure 2* and *Table S11*).

We then calculated the pPRS for each pathway in population subjects, AD cases and cognitively healthy centenarians including and excluding *APOE* variants (*Figure 1B* and *1C)*. We found that, when including *APOE* variants, the pPRSs of all pathways significantly associated with increased risk of AD (*A vs. P*, *immune response*: *OR*=1.70, 95% CI=[1.57-1.84], p=2.9x10^{-46}; *β-amyloid metabolism: OR*=2.53, 95% CI=[2.34-2.74], p=2.6x10^{-109}; *endocytosis*: *OR*=2.56, 95% CI=[2.37-2.77], p=3.4x10^{-108}; *cholesterol/lipid metabolism*: *OR*=2.56, 95% CI=[2.37-2.77], p=1.7x10^{-110}; *vascular dysfunction: OR*=1.09, 95%

CI=[1.03-1.16], p=1.2x10^{-2}) (*Figure 1B* and *Table S7*). The association of pPRSs with increased chance of escaping AD was as expected in the opposite direction for all pathways, and the association was significant for all pathways except for *vascular dysfunction* (*C vs. P*, *immune response*: *OR*=0.72, 95% CI=[0.63-0.83], p=1.2x10^{-6}; *β -amyloid metabolism: OR*=0.63, 95% CI=[0.54-0.74], p=3.9x10^{-8}; *endocytosis*: *OR*=0.59, 95% CI=[0.50-0.69], p=1.1x10^{-10}; *cholesterol/lipid metabolism: OR*=0.61, 95% CI=[0.52-0.71], p=9.6x10^{-9}; *vascular dysfunction: OR*=0.90, 95% CI=[0.80-1.01], p=1.1x10^{-1}) (*Figure 1B* and *Table S7*).

When excluding *APOE* variants, the pPRSs of all pathways was still significantly associated with increased risk of AD (*A vs. P*, *immune response*: *OR*=1.16, 95% CI=[1.09-1.23], p=1.1x10^{-5}; *β-amyloid metabolism: OR*=1.20, 95% CI=[1.13-1.27], p=8.8x10^{-8}; *endocytosis*: *OR*=1.27, 95% CI=[1.20-1.35], p=2.0x10^{-12}; *cholesterol/lipid metabolism*: *OR*=1.18, 95% CI=[1.11-1.25], p=9.1x10^{-7}; *vascular dysfunction: OR*=1.08, 95% CI=[1.02-1.15], p=1.2x10^{-2}) (*Figure 1C* and *Table 2*). The association of pPRSs with increased chance of escaping AD was as expected in the opposite direction for all pathways, yet the association was significant only for *immune response* and *endocytosis* (*C vs. P*, *immune response*: *OR*=0.83, 95% CI=[0.74-0.93], p=3.1x10^{-3}; *β-amyloid metabolism: OR*=0.94, 95% CI=[0.84-1.06], p=3.3x10^{-1}; *endocytosis*: *OR*=0.80, 95% CI=[0.71-0.90], p=3.8x10^{-4}; *cholesterol/lipid metabolism: OR*=0.90, 95% CI=[0.80-1.01], p=1.3x10^{-1}; *vascular dysfunction: OR*=0.90, 95% CI=[0.80-1.01], p=1.1x10^{-1}) (*Figure 1C* and *Table 2*).

Comparison of effect on AD and escaping AD

To compare the association of different pPRSs with AD and with escaping AD until extreme old ages more in depth, we calculated for all pPRSs the *change* in effect-size. A *change* in effect-size of 1 indicates a similar effect on causing AD as on escaping AD, whereas a *change* larger than 1 is indicative of larger effect on escaping AD compared to causing AD. When including *APOE* variants, the *change* in effect-size was <1 for all pathways except for the *vascular dysfunction* pathway (*Figure 3*). This is expected as the effect-size of *APOE* variants on causing AD alone is much larger than its effect on escaping AD. When excluding *APOE* variants, the *change* in effect-size was still <1 for *β-amyloid metabolism* and *cholesterol/lipid metabolism* (respectively 0.33 and 0.59), but it approximated 1 for *endocytosis* (0.93) and it was larger than 1 for the *immune response* and *vascular dysfunction* (respectively 1.28 and 1.15) (*Figure 3*).

Contributions of each pathway to the polygenic risk of AD

Finally, we estimated the relative contribution of each pathway to the polygenic risk of AD in the general population. This is indicative of the degree of involvement of each pathway to the total polygenic risk of AD. Including *APOE* variants, the contribution of the pathways to the total polygenic risk of AD was 32% for *β-amyloid metabolism,* 24.2% for

endocytosis, 20.7% for *cholesterol/lipid metabolism*, 19% for *immune response, 0.7%* for *vascular dysfunction* and 3.4% for the *unmapped* variants (*Figure 4A*).

When we excluded *APOE* variants, the contribution of the pathways to the total polygenic risk of AD was 41.3% for *immune response,* 21.4% for *endocytosis*, 13.3% for *ß-amyloid metabolism*, 6% for *cholesterol/lipid metabolism*, 3% for *vascular dysfunction* and 15% for the *unmapped* variants (*Figure 4B*).

Discussion

In this work, we studied 29 common genetic variants known to associate with AD using polygenic risk scores and pathway-specific polygenic risk scores. As expected, we found that a higher PRS for AD was associated with a higher risk of AD. In addition, using our unique cohort of cognitively healthy centenarians, we showed that a lower PRS for AD associated with escaping AD until extreme old ages. This suggests that the long-term preservation of cognitive health is associated with the selective survival of individuals with the lowest burden of risk-increasing variants or, vice versa, the highest burden of protective variants. Then, using an innovative approach, we studied the five pathways previously found to be involved in AD as well as the contribution of these pathways to the polygenic risk of AD. We showed that all pathways-PRS associate with increased AD risk, both including and excluding *APOE* variants. When we studied the association of pathways-PRS with escaping AD until extreme old ages, we found that, as expected, *APOE* variants played a major factor also in escaping AD. However, when excluding these variants only *immune response* and *endocytosis* significantly associated with an increased chance to escape AD. Interestingly, the effect of these pathways on escaping AD was larger or comparable to that of developing AD, suggesting that these pathways might be involved in general neuro-protective functions. Based on the effect size and frequency of all AD-associated variants, we found that the *ß-amyloid metabolism* (32%) followed by *endocytosis* (24.2%) were the major contributors to the modification of AD-risk. After excluding *APOE* variants, the pathways that contributed the most to the modification of AD-risk were *immune response* (41.3%) and *endocytosis* (21.4%).

Previous studies showed that polygenic risk score of AD not only associated with increased risk of AD, but also with neuropathological hallmarks of AD, lifetime risk and the age at onset in both *APOE ε4* carriers and non-carriers.[28, 29, 48–52] We now add that the PRS for AD also associates with escaping AD at extremely old ages. This adds further importance to the potentiality of using PRS and *APOE* genotype in a clinical setting.[48, 49, 51, 53]

Our approach to map variants to associated genes and to map genes to pathways resulted in a weighted annotation of variants to pathways that allowed for uncertainty in gene as well as pathway assignment, which was not done

previously. We note that considering uncertainty in variant-gene as well as gene-pathway assignments is crucial because most genetic variants are in non-coding regions, which makes the closest gene not necessarily the culprit gene, and because different functional annotation-sources often do not overlap. The pathway-specific PRS that we proposed in this manuscript can be re-used for the identification of subtypes of AD patients compromised in a specific AD-associated pathway. This is of interest for clinical trials, in order to test responsiveness to compounds in specific subsets of patients. For example, monoclonal antibody targeting *TREM2* receptors could work better in AD patients where the immune response pathway is most deficient. Recently, several studies attempted to construct pathway-specific PRS to find heterogeneity in AD patients based on a genetic basis.[28, 29] In line with our findings, *Ahmad et al.* found that genes capturing *endocytosis* pathway significantly associated with AD and with the conversion to AD.[29] Other studies used less variants [28] or less stringent selection for variants, and did not observe a differential involvement of pathways in AD etiology.[54]

The amyloid cascade hypothesis has been dominating AD-related research in the last two decades. However, treatments targeting amyloid have, so far, not been able to slow or stop disease progression. This has led to an increased interest for the other pathways that are important in AD pathogenesis.[22] Part of the current view of the etiology of AD is that the dysregulation of the immune response is a major causal pathway, and that AD is not only a consequence of β-amyloid metabolism.[55, 56] In addition, previous studies showed that healthy immune and metabolic systems are associated with longer and healthier lifespan.[1, 57] Our results indicate that, excluding *APOE* variants, the effect of immune response and endocytosis on escaping AD is stronger or comparable to the effect on causing AD. This suggests that these pathways might be involved in the maintenance of general cognitive health, as the cognitively healthy centenarians represent the escape of all neurodegenerative diseases until extreme ages. We recently found evidence for this hypothesis in the protective low frequency variant in *PLCG2*, which is involved in the regulation of the immune response.[50] This variant is enriched in cognitively healthy centenarians, and protects against AD as well as frontotemporal dementia and dementia with Lewy bodies.[50] We included this variant in the total PRS as well as in the pathway-PRS for the immune response (*variant-pathway mapping* was 60%) and endocytosis (*variant-pathway mapping* was 40%). Regarding endocytosis, this pathway is thought to play a role both in neurons, as part of the β-amyloid metabolism, but also in glia cells, as part of the immune response. Thus, a dysregulation in the interplay between these pathways might lead to an imbalance of immune signaling factors, favouring the engulfment of synapses and AD-associated processes. This, in turn, may contribute to the accumulation of amyloid and tau pathologies.[58–61].

In this work, we assessed the effect of common and low frequency variants on the development and the escape of AD. Therefore, the contributions of rare, causative variants associated with increased AD risk, such as those in *APP, PSEN1, PSEN2*, *TREM2* and *SORL1* were not considered. Despite the large odds ratios to develop AD associated with carrying such variants, the frequency of these variants in the population is very low, and therefore have a minor effect on the total AD risk in the population.[11, 12] However, future versions of the PRS will most likely include the effect of carrying disease-associated rare variants. This will affect individual PRS scores and the necessity to accordingly adapt the results generated with current PRSs. Compared to the sizes of recent GWAS of AD, we included relatively small sample sizes, particularly with respect to the cognitively healthy centenarians, a very rare phenotype in the population (<0.1%).[4] These sample sizes are however sufficient to study PRSs. The cohorts that we used in this study were not used in any GWAS of AD, therefore we provide independent replication of AD PRS in a homogeneous group of (Dutch) individuals.

The goal of this study was not to associate new pathways to AD. For this reason, we limited our analysis to 5 pathways known to be relevant in AD. We acknowledge that, as new variants as well as pathways will be associated with AD, the contributions and the pathway-specific PRSs will need to be recalculated. A limitation, not exclusive to our work, is the highly debated role of *APOE* gene. We mapped the effect of *APOE* to four pathways and we are aware this assignment is relatively arbitrary. For example, *APOE* has vascular properties in cholesterol transport,[62] but in the annotation sources that we consulted these links were not present. The combination of a large effect and unclear pathway assignment makes that pathway-PRS including *APOE* challenging to use. Lastly, we note that the variance contributions might change in different populations, as it depends on variant frequency and population heterogeneity.

Concluding, with the exclusion of *APOE* variants, the aggregate contribution of the immune response and endocytosis represents more than 60% of the currently known polygenic risk of AD. This indicates that an intervention in these systems may have large potential to prevent AD and potentially other related diseases and highlights the critical need to study (neuro)immune response and endocytosis, next to β-amyloid metabolism.

Acknowledgements

Research of the Alzheimer center Amsterdam is part of the neurodegeneration research program of Amsterdam Neuroscience (www.amsterdamresearch.org). The Alzheimer Center Amsterdam is supported by Stichting Alzheimer Nederland (WE09.2014-03) and Stiching VUmc fonds. The clinical database structure was developed with funding

from Stichting Dioraphte (VSM 14 04 14 02). The Dutch case–control study is part of EADB (European Alzheimer DNA biobank) funded by JPcofundNL (ZonMW project number: 733051061). This work was in part carried out on the Dutch national e-infrastructure with the support of SURF Cooperative.

References

1. Partridge L, Deelen J, Slagboom PE. Facing up to the global challenges of ageing. Nature. 2018;561:45–56.
2. 2012 Alzheimer's disease facts and figures. Alzheimers Dement. 2012;8:131–168.
3. Corrada MM, Brookmeyer R, Paganini-Hill A, Berlau D, Kawas CH. Dementia incidence continues to increase with age in the oldest old: The 90+ study. Ann Neurol. 2010;67:114–121.
4. Holstege H, Beker N, Dijkstra T, Pieterse K, Wemmenhove E, Schouten K, et al. The 100-plus Study of cognitively healthy centenarians: rationale, design and cohort description. Eur J Epidemiol. 2018. 25 October 2018. https://doi.org/10.1007/s10654-018-0451-3.
5. Niccoli T, Partridge L. Ageing as a Risk Factor for Disease. Curr Biol. 2012;22:R741–R752.
6. Brooks-Wilson AR. Genetics of healthy aging and longevity. Hum Genet. 2013;132:1323–1338.
7. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. 2013;45:1452–1458.
8. Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet. 2009;41:1088–1093.
9. Seshadri S, Fitzpatrick AL, Ikram MA, DeStefano AL, Gudnason V, Boada M, et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. JAMA. 2010;303:1832–1840.
10. Desikan RS, Schork AJ, Wang Y, Thompson WK, Dehghan A, Ridker PM, et al. Polygenic Overlap Between C-Reactive Protein, Plasma Lipids, and Alzheimer Disease. Circulation. 2015;131:2061–2069.
11. Sims R, van der Lee SJ, Naj AC, Bellenguez C, Badarinarayan N, Jakobsdottir J, et al. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. Nat Genet. 2017;49:1373–1384.
12. Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, Majounie E, et al. TREM2 variants in Alzheimer's disease. N Engl J Med. 2013;368:117–127.
13. Jonsson T, Stefansson H, Steinberg S, Jonsdottir I, Jonsson PV, Snaedal J, et al. Variant of TREM2 associated with the risk of Alzheimer's disease. N Engl J Med. 2013;368:107–116.
14. Hollingworth P, Harold D, Sims R, Gerrish A, Lambert J-C, Carrasquillo MM, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet. 2011;43:429–435.
15. Naj AC, Jun G, Beecham GW, Wang L-S, Vardarajan BN, Buros J, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet. 2011;43:436–441.
16. Jun G, Ibrahim-Verbaas CA, Vronskaya M, Lambert J-C, Chung J, Naj AC, et al. A novel Alzheimer disease locus located near the gene encoding tau protein. Mol Psychiatry. 2016;21:108–117.
17. Steinberg S, Stefansson H, Jonsson T, Johannsdottir H, Ingason A, Helgason H, et al. Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. Nat Genet. 2015;47:445–447.
18. Marioni R, Harris SE, McRae AF, Zhang Q, Hagenaars SP, Hill WD, et al. GWAS on family history of Alzheimer's disease. 2018.
19. Alzheimer Disease Genetics Consortium (ADGC), The European Alzheimer's Disease Initiative (EADI), Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (CHARGE), Genetic and Environmental Risk in AD/Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease Consortium (GERAD/PERADES), Kunkle BW, Grenier-Boley B, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nat Genet. 2019;51:414–430.
20. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat Genet. 2019;51:404–413.
21. Tesi N, van der Lee SJ, Hulsman M, Jansen IE, Stringa N, van Schoor N, et al. Centenarian controls increase variant effect sizes by an average twofold in an extreme case–extreme control analysis of Alzheimer's disease. Eur J Hum Genet. 2018. 26 September 2018. https://doi.org/10.1038/s41431-018-0273-5.
22. Van Cauwenberghe C, Van Broeckhoven C, Sleegers K. The genetic landscape of Alzheimer disease: clinical implications and perspectives. Genet Med. 2016;18:421–430.

23. Hardy J, Bogdanovic N, Winblad B, Portelius E, Andreasen N, Cedazo-Minguez A, et al. Pathways to Alzheimer's disease. J Intern Med. 2014;275:296–303.
24. Naj AC, Schellenberg GD, for the Alzheimer's Disease Genetics Consortium (ADGC). Genomic variants, genes, and pathways of Alzheimer's disease: An overview. Am J Med Genet B Neuropsychiatr Genet. 2017;174:5–26.
25. Verheijen J, Sleegers K. Understanding Alzheimer Disease at the Interface between Genetics and Transcriptomics. Trends Genet. 2018;34:434–447.
26. Jefferies WA, Price KA, Biron KE, Fenninger F, Pfeifer CG, Dickstein DL. Adjusting the compass: new insights into the role of angiogenesis in Alzheimer's disease. Alzheimers Res Ther. 2013;5:64.
27. Vagnucci AH, Li WW. Alzheimer's disease and angiogenesis. The Lancet. 2003;361:605–608.
28. Darst BF, Koscik RL, Racine AM, Oh JM, Krause RA, Carlsson CM, et al. Pathway-Specific Polygenic Risk Scores as Predictors of Amyloid-β Deposition and Cognitive Function in a Sample at Increased Risk for Alzheimer's Disease. J Alzheimers Dis. 2016;55:473–484.
29. Ahmad S, Bannister C, van der Lee SJ, Vojinovic D, Adams HHH, Ramirez A, et al. Disentangling the biological pathways involved in early features of Alzheimer's disease in the Rotterdam Study. Alzheimers Dement. 2018;14:848–857.
30. Dudbridge F. Power and Predictive Accuracy of Polygenic Risk Scores. PLoS Genet. 2013;9:e1003348.
31. Freudenberg-Hua Y, Li W, Davies P. The Role of Genetics in Advancing Precision Medicine for Alzheimer's Disease-A Narrative Review. Front Med. 2018;5:108.
32. Huisman M, Poppelaars J, van der Horst M, Beekman AT, Brug J, van Tilburg TG, et al. Cohort Profile: The Longitudinal Aging Study Amsterdam. Int J Epidemiol. 2011;40:868–876.
33. Hoogendijk EO, Deeg DJH, Poppelaars J, van der Horst M, Broese van Groenou MI, Comijs HC, et al. The Longitudinal Aging Study Amsterdam: cohort update 2016 and major findings. Eur J Epidemiol. 2016;31:927–945.
34. van der Flier WM, Scheltens P. Amsterdam Dementia Cohort: Performing Research to Optimize Care. J Alzheimers Dis. 2018;62:1091–1111.
35. van der Flier WM, Pijnenburg YAL, Prins N, Lemstra AW, Bouwman FH, Teunissen CE, et al. Optimizing patient care and research: the Amsterdam Dementia Cohort. J Alzheimers Dis JAD. 2014;41:313–327.
36. Rademaker MC, de Lange GM, Palmen SJMC. The Netherlands Brain Bank for Psychiatry. Handb. Clin. Neurol., vol. 150, Elsevier; 2018. p. 3–16.
37. Lambert J-C, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet. 2009;41:1094–1099.
38. Strittmatter WJ, Saunders AM, Schmechel D, Pericak-Vance M, Enghild J, Salvesen GS, et al. Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proc Natl Acad Sci U S A. 1993;90:1977–1981.
39. Corder EH, Saunders AM, Risch NJ, Strittmatter WJ, Schmechel DE, Gaskell PC, et al. Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. Nat Genet. 1994;7:180–184.
40. Jansen I, Savage J, Watanabe K, Bryois J, Williams D, Steinberg S, et al. Genetic meta-analysis identifies 9 novel loci and functional pathways for Alzheimers disease risk. 2018.
41. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016;48:1284–1287.
42. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526:68–74.
43. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8.
44. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–29.
45. The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Res. 2017;45:D331–D338.

46. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4:44–57.
47. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37:1–13.
48. Escott-Price V, Myers AJ, Huentelman M, Hardy J. Polygenic risk score analysis of pathologically confirmed Alzheimer disease: PRS Analysis of AD. Ann Neurol. 2017;82:311–314.
49. Escott-Price V, Shoai M, Pither R, Williams J, Hardy J. Polygenic score prediction captures nearly all common genetic risk for Alzheimer's disease. Neurobiol Aging. 2017;49:214.e7-214.e11.
50. DESGESCO (Dementia Genetics Spanish Consortium), EADB (Alzheimer Disease European DNA biobank), EADB (Alzheimer Disease European DNA biobank), IFGC (International FTD-Genomics Consortium), IPDGC (The International Parkinson Disease Genomics Consortium), IPDGC (The International Parkinson Disease Genomics Consortium), RiMod-FTD (Risk and Modifying factors in Fronto-Temporal Dementia), Netherlands Brain Bank (NBB), et al. A nonsynonymous mutation in PLCG2 reduces the risk of Alzheimer's disease, dementia with Lewy bodies and frontotemporal dementia, and increases the likelihood of longevity. Acta Neuropathol (Berl). 2019. 27 May 2019. https://doi.org/10.1007/s00401-019-02026-8.
51. for the International Genomics of Alzheimer's Project, Chouraki V, Reitz C, Maury F, Bis JC, Bellenguez C, et al. Evaluation of a Genetic Risk Score to Improve Risk Prediction for Alzheimer's Disease. J Alzheimers Dis. 2016;53:921–932.
52. Chaudhury S, Brookes KJ, Patel T, Fallows A, Guetta-Baranes T, Turton JC, et al. Alzheimer's disease polygenic risk score as a predictor of conversion from mild-cognitive impairment. Transl Psychiatry. 2019;9.
53. van der Lee SJ, Wolters FJ, Ikram MK, Hofman A, Ikram MA, Amin N, et al. The effect of APOE and other common genetic variants on the onset of Alzheimer's disease and dementia: a community-based cohort study. Lancet Neurol. 2018;17:434–444.
54. Leonenko G, Shoai M, Bellou E, Sims R, Williams J, Hardy J, et al. Genetic risk for Alzheimer's disease is distinct from genetic risk for amyloid deposition. Ann Neurol. 2019. 14 June 2019. https://doi.org/10.1002/ana.25530.
55. Ransohoff RM. How neuroinflammation contributes to neurodegeneration. Science. 2016;353:777–783.
56. Heneka MT, Carson MJ, Khoury JE, Landreth GE, Brosseron F, Feinstein DL, et al. Neuroinflammation in Alzheimer's disease. Lancet Neurol. 2015;14:388–405.
57. Joshi PK, Pirastu N, Kentistou KA, Fischer K, Hofer E, Schraut KE, et al. Genome-wide meta-analysis associates HLA-DQA1/DRB1 and LPA and lifestyle factors with human longevity. Nat Commun. 2017;8.
58. Srinivasan K, Friedman BA, Larson JL, Lauffer BE, Goldstein LD, Appling LL, et al. Untangling the brain's neuroinflammatory and neurodegenerative transcriptional responses. Nat Commun. 2016;7.
59. Wang Y, Ulland TK, Ulrich JD, Song W, Tzaferis JA, Hole JT, et al. TREM2-mediated early microglial response limits diffusion and toxicity of amyloid plaques. J Exp Med. 2016;213:667–675.
60. Orre M, Kamphuis W, Osborn LM, Jansen AHP, Kooijman L, Bossers K, et al. Isolation of glia from Alzheimer's mice reveals inflammation and dysfunction. Neurobiol Aging. 2014;35:2746–2760.
61. Hansen DV, Hanson JE, Sheng M. Microglia in Alzheimer's disease. J Cell Biol. 2018;217:459–472.
62. Hauser PS, Narayanaswami V, Ryan RO. Apolipoprotein E: from lipid transport to neurobiology. Prog Lipid Res. 2011;50:62–74.

Tables and Figures

Table 1: Population characteristics

	AD cases	Population subjects	Healthy centenarians
Number of individuals	1,909	1,654	292
Females (%)	1,069 (56.4%)	880 (53.2%)	212 (72.6%)
Age (SD)*	69.23 (9.9)	62.73 (6.4)	101.4 (1.3)

* Age at onset for AD cases, age at study inclusion for population subjects and cognitively healthy centenarians.

Table 2: Results of association with PRS and pPRS excluding *APOE* variants

Pathways (pPRS)	Excluding *APOE*					
	AD cases vs. Pop. Subjects (A vs P) (N = 1,909 vs. 1,654)			*AD escapers vs. Pop. Subjects (C vs P)* (N = 293 *vs.* 1,654)		
	OR	95% CI	*P*	OR	95% CI	*P*
Immune response	1.16	1.09-1.23	1.1×10^{-5}	0.83	0.74-0.93	3.1×10^{-3}
β-amyloid metabolism	1.20	1.13-1.27	8.8×10^{-8}	0.94	0.84-1.06	3.3×10^{-1}
Endocytosis	1.27	1.20-1.35	2.0×10^{-12}	0.80	0.71-0.90	3.8×10^{-4}
Cholesterol/lipid metabolism	1.18	1.11-1.25	9.1×10^{-7}	0.90	0.80-1.01	1.3×10^{-1}
Vascular dysfunction	1.08	1.02-1.15	1.2×10^{-2}	0.90	0.80-1.01	1.1×10^{-1}
Combined (PRS)	1.30	1.23-1.38	7.3×10^{-14}	0.79	0.70-0.89	3.2×10^{-4}

Abbreviations: Centenarians, cognitively healthy centenarians; *OR*, odds ratio; *SE,* standard error with respect to the log(OR); *P*, p-value of association; *Combined,* association of PRS, *i.e* including all variants; *pPRS,* pathway-specific PRS; *PRS*, total PRS including all 29 variants, excluded *APOE* variants.

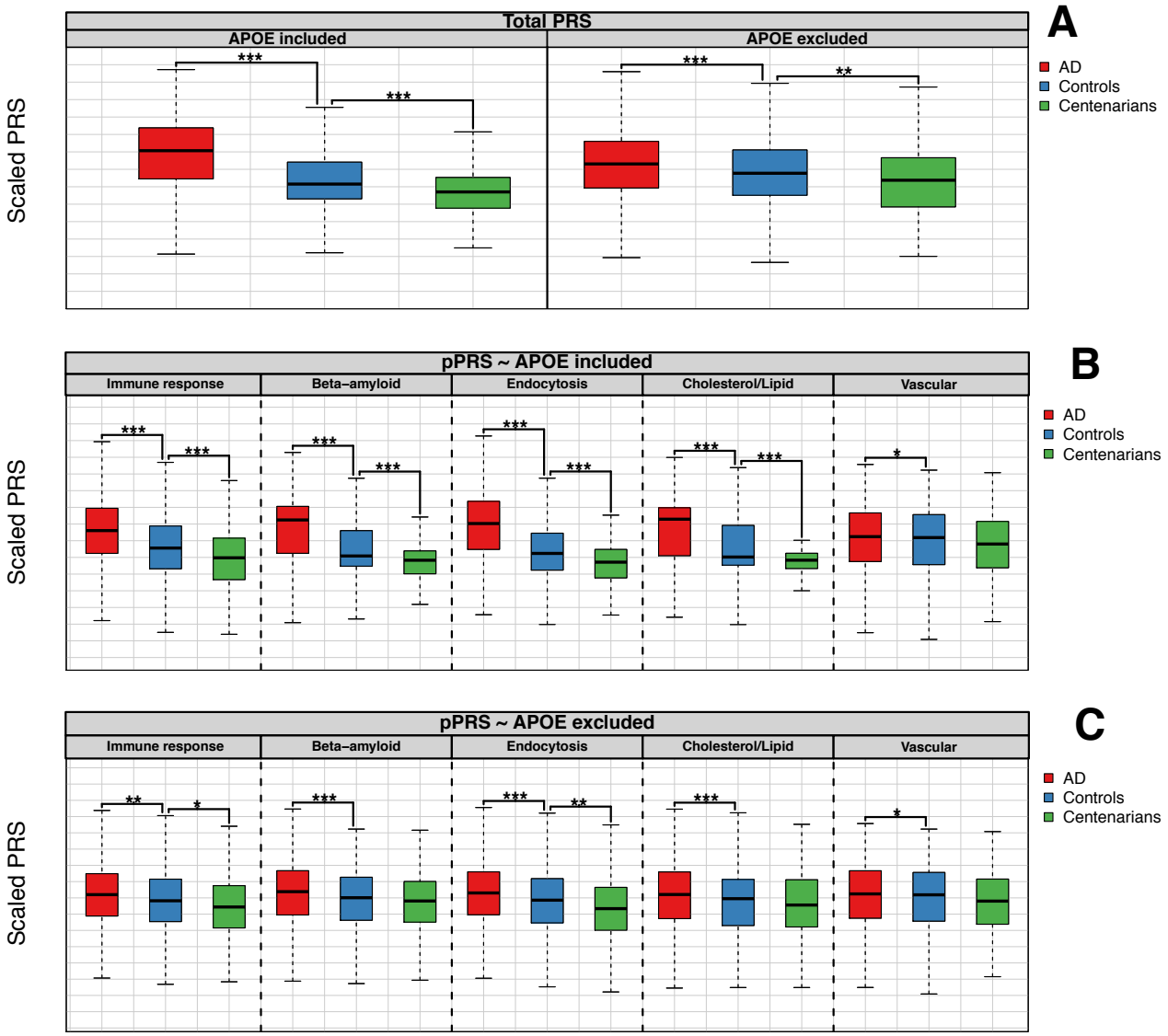

Figure 1: Boxplots of PRS and pPRS in the different settings. Figure A (above) shows the PRS including all the 29 known AD-associated variants, with and without *APOE* variants. As weight for the PRS, we used published variant effect-sizes (*Table S1*). Figure B (central) and Figure C (bottom) show the pPRS for each of the selected molecular pathways, including and excluding *APOE* variants, respectively. For all plots, risk scores were calculated for AD cases, population subjects and cognitively healthy centenarians. Then, risk scores were compared between (*i*) AD cases and population subjects (*A vs. P* comparison) and (*ii*) cognitively healthy centenarians and population subjects (*C vs. P* comparison). For representation, we scaled all PRS and pathway-PRS to be mean=0 and SD=1. For the comparison, we used logistic regression models with risk scores as predictors. Annotation: ***, *p-value* of association < $5x10^{-6}$; **, *p-value* of association < $5x10^{-4}$; *, *p-value* of association < $5x10^{-2}$.

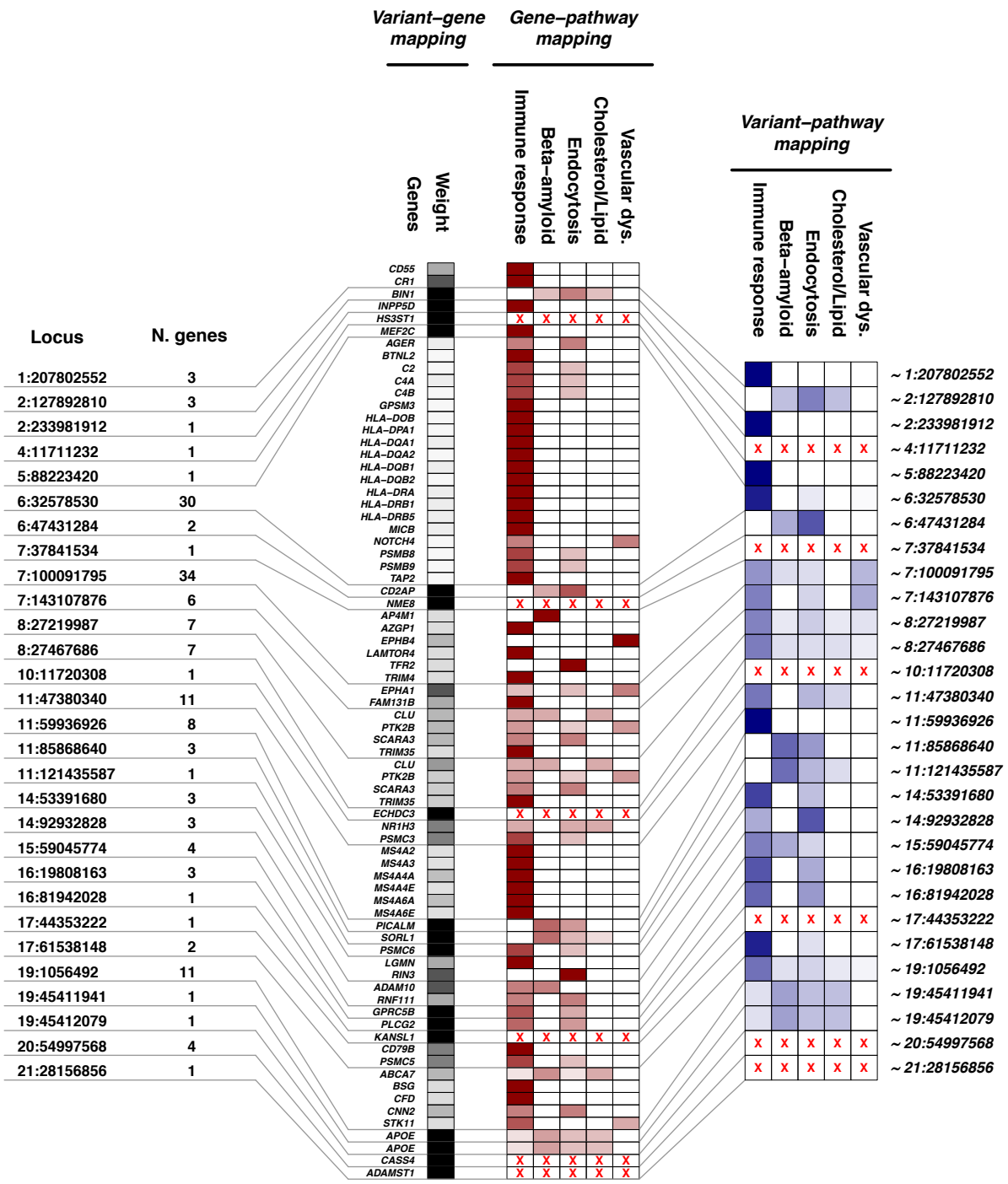

Figure 2: Variant-pathways mapping: *Locus*: chromosome and position of the AD-associated genetic variants (coordinates are with respect to GRCh37). *N. genes*: total number of genes associated with each variant according to *variant-gene mapping* (equivalent to locational mapping). *Variant-gene mapping: Genes*: all genes with at least one annotation to the 5 selected molecular pathways associated with AD. *Weight*: the weight of the *variant-gene mapping*. *Gene-pathway mapping: Immune response, Beta-amyloid, Endocytosis, Cholesterol/lipid, Vascular dysfunction*: the weight of each molecular pathway at the gene level. *Variant-pathway mapping*: summarization of each variant's effect after combining *variant-gene* and *gene-pathway* mappings. Red crosses indicate unmapped genes.

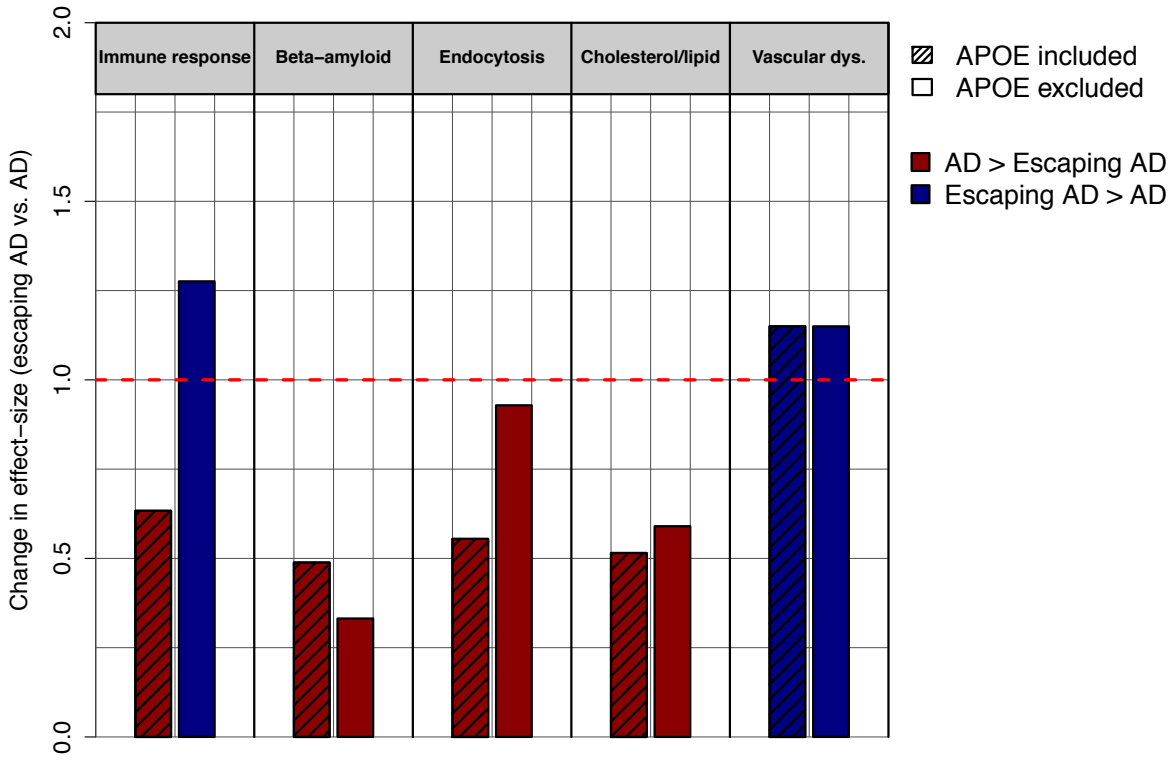

Figure 3: *Change* in effect-size between association with escaping AD and causing AD for the five pPRSs: bars are grouped by pathway. Each bar represents the ratio between the effect-size of the association with escaping AD (*C vs. P* comparison) and with causing AD (*A vs. P* comparison), respectively with and without *APOE* variants. Red dotted line at 1 is indicative of the same effect on both developing and escaping AD. Ratios larger than 1 are then indicative of larger effect-size on escaping AD. Blue bars are those with a *change* in effect-size larger than 1, while red bars refer to a *change* in effect-size smaller than 1.

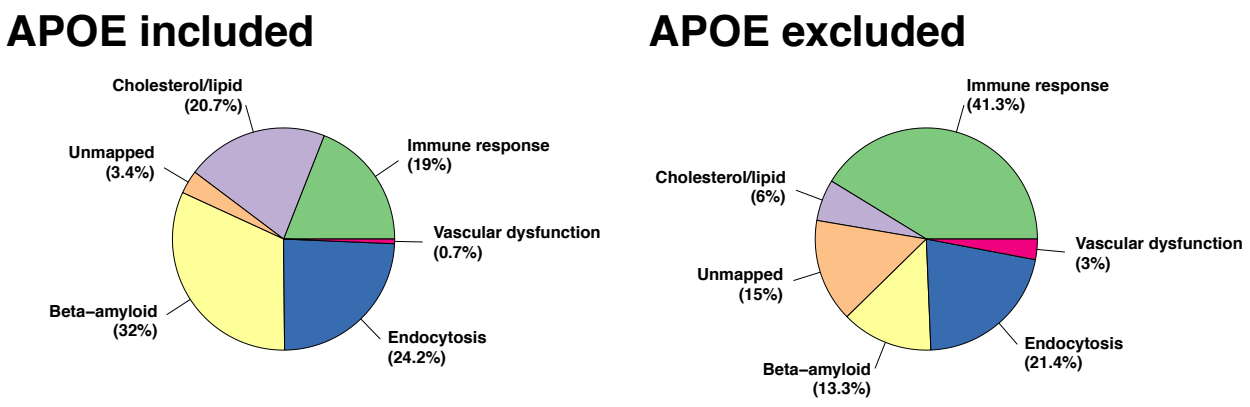

Figure 4: Explained variance of each pathway-specific PRS to polygenic risk of AD: the pie charts represents the explained variance of each pathway-specific PRS to the polygenic risk of AD, including and excluding *APOE* variants. The contributions are calculated according to (*i*) our *variant-pathway mapping*, (*ii*) the effect size (log of odds ratio) of each variant from literature (*Table S1*), and (*iii*) variant's frequency in our cohort of middle-aged healthy population subjects. We also considered variants with missing *variant-pathway mapping* (*unmapped* pathway).