

## **Title: Compositional structural brain signatures capture Alzheimer's genetic risk on brain structure along the disease *continuum***

Authors: Patricia Genius (1-3), M.Luz Calle (4), Blanca Rodríguez-Fernández (1-3), Carolina Minguillon (1,2,5), Raffaele Cacciaglia (1,2,5), Diego Garrido-Martin (3,6), Manel Esteller (7-10), Arcadi Navarro (1,3,8,11,12), Juan Domingo Gispert (1,2,13,14), Natalia Vilor-Tejedor\* (1,2,3,15) for the Alzheimer's Disease Neuroimaging Initiative<sup>^</sup> and for the ALFA study<sup>†</sup>

### **Affiliations:**

1. Barcelonaβeta Brain Research Center (BBRC), Pasqual Maragall Foundation, Barcelona, Spain.
2. IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain.
3. Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain.
4. Biosciences Department, Faculty of Sciences, Technology and Engineering, University of Vic-Central University of Catalonia, Vic, Spain
5. Centro de Investigación Biomédica en Red de Fragilidad y Envejecimiento Saludable (CIBERFES), Instituto de Salud Carlos III, Madrid, Spain.
6. University of Barcelona. Department of Genetics, Microbiology and Statistics, Barcelona, Spain.
7. Josep Carreras Leukaemia Research Institute (IJC), Badalona, Barcelona, Catalonia, Spain.
8. Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain.
9. Physiological Sciences Department, School of Medicine and Health Sciences, University of Barcelona (UB), Barcelona, Catalonia, Spain
10. Centro de Investigación Biomédica en Red Cancer (CIBERONC), Madrid, Spain
11. Pompeu Fabra University, Barcelona, Spain.
12. Institute of Evolutionary Biology (CSIC-UPF), Department of Experimental and Health Sciences, Universitat Pompeu Fabra, 08003, Barcelona, Spain.
13. Centro de Investigación Biomédica en Red Bioingeniería, Biomateriales y Nanomedicina, Instituto de Salud Carlos III, Madrid, Spain.
14. Centro Nacional de Investigaciones Cardiovasculares (CNIC). Madrid, Spain.
15. Radboud University Nijmegen Medical Center, Department of Genetics, Nijmegen, the Netherlands

<sup>†</sup> *The complete list of collaborators of the ALFA study can be found in the acknowledgments section.*

<sup>^</sup>*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](http://adni.loni.usc.edu)). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at:*

*[http://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf)*

<b>Address</b>	<b>for</b>	<b>Correspondence:</b>
Natalia Vilor-Tejedor, BarcelonaBeta Brain Research Center, Pasqual Maragall Foundation, Barcelona, Spain.	C. Wellington, 30, 08005	Barcelona, Spain.
ORCID:	0000-0003-	4935-6721
E-mail: <a href="mailto:nvilor@barcelonabeta.org">nvilor@barcelonabeta.org</a>		

## Abstract

**INTRODUCTION:** Traditional brain imaging genetics studies have primarily focused on how genetic factors influence the volume of specific brain regions, often neglecting the overall complexity of brain architecture and its genetic underpinnings.

**METHODS:** This study analyzed data from participants across the Alzheimer's disease (AD) *continuum* from the ALFA and ADNI studies. We exploited compositional data analysis to examine relative brain volumetric variations that (i) differentiate cognitively unimpaired (CU) individuals, defined as amyloid-negative (A-) based on CSF profiling, from those at different AD stages, and (ii) associated with increased genetic susceptibility to AD, assessed using polygenic risk scores.

**RESULTS:** Distinct brain signatures differentiated CU A- individuals from amyloid-positive MCI and AD. Moreover, disease stage-specific signatures were associated with higher genetic risk of AD.

**DISCUSSION:** The findings underscore the complex interplay between genetics and disease stages in shaping brain structure, which could inform targeted preventive strategies and interventions in preclinical AD.

**Keywords:** AD Genetic Predisposition; Structural Brain Signature; Multiphenotype Analysis; Compositional Data Analysis; Polygenic Risk Scoring; Brain Imaging Genetics; Neurodegeneration.

## 1. Background

Alzheimer's disease (AD) is a complex multifactorial and heterogeneous neurodegenerative disorder, characterized by progressive changes in brain structure, among other pathophysiological hallmarks<sup>1</sup>. Structural magnetic resonance imaging (MRI) serves as a powerful tool for quantifying cerebral atrophy in the *AD continuum*. Common structural brain changes in AD display a limbic-to-frontal sequence<sup>2</sup> throughout the disease *continuum*. Notably, brain atrophy and hippocampal reduction are closely linked with cognitive decline<sup>3-4-5-6</sup>. Recent studies have shown the high heritability of these structural changes<sup>7-8</sup>, which increases the interest in their role as intermediate phenotypes. Genetics plays a pivotal role in AD, with the *APOLIPOPROTEIN E* gene allele  $\epsilon 4$  (*APOE- $\epsilon 4$* ) significantly increasing disease risk<sup>9</sup>. Nonetheless, beyond *APOE- $\epsilon 4$* , genome-wide association studies (GWAS) have identified numerous genetic variants associated with AD<sup>10-11-12</sup>. In a recent published study<sup>13</sup> we showed that the genetic risk of AD, estimated through polygenic risk scoring, was higher in clinical groups than in cognitively unimpaired (CU) amyloid negative individuals, but not amyloid positive. These findings highlight a similar genetic burden between CU individuals displaying AD-related pathology and individuals at clinical stages of the disease. Therefore, these results reinforce the need to carefully examine how genetics impact on AD structural endophenotypes to examine brain cortical and subcortical changes not only based on the clinical diagnosis but also on the genetic load for AD at early stages of the disease. By combining imaging and genetic markers, there is potential for a more comprehensive exploration and understanding of the individuals' profiles during the preclinical stages of the disease. In this line, brain imaging genetics (IG) studies aim to integrate neuroimaging (NI) and genetic data to uncover new genetic variants associated with AD-related brain features<sup>14-15</sup>. Among the different approaches in IG, univariate analyses that look for the association



between individual brain volumetric structures and AD genetic risk factors<sup>16-18</sup>, do not consider the compositional origin of the brain. The brain can be segmented in regions, and the volumes of the segmented regions can be expressed as a fraction relative to the total intracranial volume. Similarly, at any level of segmentation, a specific brain area and its segmented regions can be defined as a composition and its components, respectively. Thus, given the interconnected and compositional structure of the brain, it is crucial to analyze brain features in a multivariate way to fully understand the complex interactions and influences across different regions<sup>19</sup>. Recent studies have used compositional approaches in the field of brain imaging due to the compositional structure of the data obtained from structural MRI and anatomical brain segmentation<sup>20</sup>. Compositional data analysis (CoDA) methods offer a sophisticated approach to evaluating the brain's complex structure, by viewing it as a cohesive entity. In CoDA, instead of exploring each component separately, the analysis focuses on the relative variation between them<sup>21-22</sup>. This approach moves beyond the limitations of traditional analyses that either focus on single volumetric structures or employ multivariate techniques that fail to capture the compositional origin of structural brain features. This aspect is particularly crucial in AD, where changes in one region can have cascading effects on others, reflecting the multifaceted nature of the disease<sup>23</sup>. This study aimed to identify relative brain volumetric variations in cortical and subcortical areas that (i) distinguished between CU individuals and others at different stages of the disease and (ii) were associated with higher genetic susceptibility to AD within groups along the *continuum*. Compositional methods were employed to explore these relative brain volumetric variations, summarized in structural brain signatures. Results showed that structural brain signatures were accurately distinguishing between CU individuals and others at different stages of the disease. Moreover, disease stage-specific signatures were associated with higher genetic risk of AD. These findings highlight

the modifying effect of disease stage and the heterogeneous profiles within the *continuum* on how AD genetic burden impacts brain structure, demonstrating the complex interplay between genetics and disease stage in shaping structural brain volumetry.

## **2. Methods**

### **2.1. Study Population**

This study included 338 CU middle-aged (45-65 years old) participants from the ALFA+ cohort, a nested cohort of the Alzheimer's and FAMILIES study<sup>24</sup>, and 330 CI individuals from the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort, a project funded by the National Institutes of Health (NIH) and launched in 2004 ([adni.loni.usc.edu](http://adni.loni.usc.edu))<sup>25</sup>. Individuals had available information on cortical and subcortical brain regions volume, cerebrospinal fluid (CSF) biomarkers and genetic data. Participants were classified into A/T groups defined by their CSF biomarker profile according to the A/T framework described in<sup>26</sup>. Amyloid-beta pathology positivity (A+) and tau pathology positivity (T+) were defined based on validated cut-offs<sup>27</sup>. Further details on the CSF sampling can be found in [Supplementary Methods](#). The CSF profile of A+T+ individuals in ALFA was determined by the CSF A $\beta$ 42/40 ratio  $< 0.071$  pg/mL and levels of CSF p-tau181  $> 0.24$  pg/mL, while in ADNI A+ was defined by CSF A $\beta$ 42 levels  $\leq 1100$  pg/mL. Participants were classified into four different groups based on both their clinical diagnosis and their CSF amyloid profile. Finally, the sample was covering the whole disease *continuum*: CU A- ALFA (N=220), CU A+ ALFA (N=118), MCI A+ ADNI (N=230), AD A+ ADNI (N=100).

### **2.2. Genetic data acquisition, quality control and imputation**

For the ALFA study, DNA was obtained from blood samples through a salting-out protocol. Genotyping was performed with the Illumina Infinium Neuro Consortium (NeuroChip) Array

(build GRCh37/hg19)<sup>28</sup>. Quality control procedure was performed using PLINK software. Imputation was performed using the Michigan Imputation Server with the haplotype Reference Consortium Panel (HRC r1.1 2016;<sup>29</sup> following default parameters and established guidelines. A full description of the genotyping, quality control and imputation procedures is available elsewhere<sup>13</sup>. ADNI participants underwent genotyping using both the Human 610-Quad BeadChip and the Illumina HumanOmniExpress BeadChip (Illumina, Inc., San Diego, CA). Quality control measures and imputation procedures closely followed the protocols established in the ALFA study. Further details are available elsewhere<sup>30</sup>.

### ***2.3. Genetic predisposition to Alzheimer's disease: polygenic risk score computation***

The polygenic risk score of AD ( $PRS_{AD}$ ) was calculated using the PRSice version 2 tool<sup>31</sup>. Summary statistics from a recent GWAS for AD<sup>11</sup> were obtained to compute the  $PRS_{AD}$ <sup>13</sup> [[Supplementary Table 1](#)]. The algorithm retained the single nucleotide polymorphisms (SNPs) with the smallest p-value in each 250 kb window and removed SNPs that were in linkage disequilibrium ( $r^2 > 0.1$ ). The threshold of SNPs inclusion was defined at  $p\text{-value} < 5 \cdot 10^{-6}$ . The  $PRS_{AD}$  was computed by adding up the alleles carried by participants, weighted by the SNP allele effect size from the GWAS and normalizing by the total number of alleles. The same procedure was performed to estimate the PRS of AD when excluding the *APOE* region ( $PRS_{AD\text{no}APOE}$ ) (chr19:45,409,011-45,412,650; GRCh37/hg19). PRSs were both computed in ALFA and ADNI, respectively<sup>13</sup>. Both PRSs were dichotomized and we created two groups of subjects to differentiate individuals at higher genetic predisposition to AD from the rest of the sample. Categorization was done in each cohort separately. The cut-off point was defined by the quantile 0.8 (High genetic group:  $PRS \geq \text{quantile } 0.8$ ; Low genetic group:  $PRS < \text{quantile } 0.8$ ).

#### **2.4. Brain regions segmentation and quantification**

Volumes for cortical and subcortical brain regions were obtained through high-resolution 3D-T1-weighted MRI scans. Each cohort used customized protocols specific to the scanner. The images were pre-processed and segmented using Freesurfer. In ALFA<sup>32</sup>, protocol parameters were identical for all participants and included high-resolution three-dimensional structural images weighted in T1 with an isotropic voxel of  $0.75 \times 0.75 \times 0.75 \text{ mm}^3$ , acquired with a 3T Philips ingenia CS scanner. The acquisition parameters were TR/TE/TI = 9.9/4.6/900 ms, flip angle =  $8^\circ$  and a matrix size of  $320 \times 320 \times 240$ . In ADNI, MRI acquisition parameters included a repetition time (TR) of 2,400 ms, inversion time (TI) of 1,000 ms, flip angle of 8 degrees, and a field of view (FOV) of 24 cm. The acquisition matrix was set to  $256 \times 256 \times 170$  in the x-, y-, and z-dimensions, resulting in a voxel size of  $1.25 \times 1.26 \times 1.2 \text{ mm}^3$ . The MRI data underwent preprocessing steps, including correction for anterior commissure and posterior commissure, skull stripping, cerebellum removal, intensity inhomogeneity correction. Finally, the data were down-sampled to a matrix size of  $128 \times 128 \times 128$ . Freesurfer version 7 was used for cortical and subcortical quantification in ALFA, while in ADNI, version 5 and version 6 were used in different batches. Cortical surface parcellation was done using the Desikan-Killiany atlas<sup>34</sup>. Subcortical measurements were obtained working with the automatic subcortical segmentation of Freesurfer. Volumes were globally quantified by summing the measurements of both hemispheres. A total of 41 volumes were selected combining 34 measurements from cortical regions and 7 from subcortical ones [[Supplementary Methods](#)]. Harmonization across sites was not assessed because the compositional approach did not work with raw volumes of the individuals. Instead, it worked with all pairs of log-ratios between regions' volumes. Therefore, any differences that could arise between cohorts were removed.

## 2.5. *Statistical analyses*

Differences in demographic characteristics were assessed according to amyloid status (ALFA) and diagnosis group (ADNI), using chi-square tests for categorical variables and parametric (t-test, ANOVA) and non-parametric tests (Wilcoxon Rank-Sum test or Kruskal-Wallis test) for continuous normally and non-normally distributed variables, respectively. P-values for pairwise comparisons were also provided, adjusted for Benjamini-Hochberg false discovery rate. CoDA<sup>35-22</sup> was performed to identify structural brain signatures (i.e. a combination of specific brain regions' volumes) (i) capable of distinguishing between CU A- and other groups along the disease *continuum*, and (ii) associated with higher genetic predisposition to AD within groups [[Supplementary Methods](#)]. In the first scenario, logistic regression models, adjusted for age, sex and the PRS<sub>AD</sub>, were defined to examine the structural brain signature that distinguished CU A- from the rest of the groups in the *continuum* (CU A+, MCI A+, AD A+). In the second step, disease-stage stratified logistic regression models, adjusted for age and sex, were performed to explore the association between the brain signature with higher genetic risk of AD in each group [[Figure 1](#)]. Separate analyses were performed based on the PRS calculation used to determine the genetic groups: 1) The PRS including all the SNPs associated with AD (PRS<sub>AD</sub>) and 2) The PRS excluding the *APOE* region (PRS<sub>ADnoAPOE</sub>). The procedure was exactly the same and analyses were repeated for both cases. Results for the second case can be found in [Supplementary Results](#). All the analyses were performed using the R software version 4.2.2.

## 3. Results

### 3.1. *Characteristics of the sample*

The ALFA sample (N=338) included 62.43% of women and 54.14% of *APOE-ε4* carriers, with a median age of 57 years old [IQR 53-61][[Table 1](#)]. The ADNI sample (N=330) was defined by 40% of women and ~64.55% of *APOE-ε4* carriers, with a median age of 74 years [IQR 68.2-78]. In the sample there were 30% of AD dementia patients and 70% of MCI individuals, all of them amyloid positive [[Table 2](#)]. Differences were assessed among disease-stage groups along the *continuum* [[Table 3](#)]. The highest percentage of women was found in ALFA CU A- individuals (63%) while the lowest was observed in ADNI MCI individuals (38%) ( $p < 0.001$ ). Significant differences in *APOE-ε4* carriership were observed among groups, with the lowest percentage found in ALFA CU A- participants (40% carriers) and the highest in ALFA CU A+ (78% carriers) ( $p < 0.001$ ). Notably, the *APOE-ε4* carrier rate in ALFA CU A+ was higher than in clinical groups from ADNI ( $p < 0.001$ ). There were significant differences in the genetic predisposition to AD between CN A- and CU A+, MCI and AD dementia individuals (FDR p-value pairwise comparisons  $< 0.001$ ), but not between CU A+ and the impaired groups. Non-significant differences were found between groups when the *APOE* region was removed from the polygenic risk score [[Supplementary Figure 1](#)]. Significant differences were also found in the median value of cortical and subcortical regions between groups [[Supplementary Table 2](#)].

### ***3.2. Structural brain signatures differentiate amyloid-negative cognitively unimpaired individuals from others at different stages of the disease***

In the ALFA study, the structural brain signature that differentiated CU A- from CU A+ individuals was primarily characterized by the relative volumetric variation between the pallidum and the amygdala, along with volumetric variations in other cortical regions of the parietal, frontal and temporal lobes. The brain signature also involved subcortical regions, specifically the pallidum, putamen and the amygdala, and well as cortical areas in the cingulate cortex [[Figure 2](#)]. The structural brain signature capable of distinguishing between CU A-

and MCI A+ individuals mainly involved the relative volumetric variation between the fusiform and the frontal pole, along with volumetric variations of other temporal and frontal areas, as well as regions in the occipital lobe. This signature also included subcortical regions, specifically the putamen and pallidum, and cortical areas in the cingulate [Figure 2]. Finally, the structural brain signature that distinguished between CU A- and AD A+ individuals was primarily characterized by the relative volumetric variation between the insula and the inferior parietal, along with the volumetric variation of other parietal and frontal regions. The brain signature also involved the putamen and cortical areas in the cingulate cortex [Figure 2]. The structural brain signature showed high prediction accuracy for differentiating CU A- from disease stages, with the discrimination being significantly robust for both MCI A+ (AUC=0.992) and AD A+ (AUC=0.980) [Supplementary Table 3].

### ***3.3. Disease-stage specific structural brain signatures associated with higher genetic risk of Alzheimer's disease***

In CU A- individuals, the structural brain signature associated with higher genetic risk of AD primarily featured structural changes in the hippocampus and the fusiform, along with variations in frontal, temporal and occipital areas together with subcortical regions such as the pallidum [Figure 3]. For CU A+ individuals, the structural brain signature related to higher risk of AD was characterized by volumetric variations in temporal, frontal and occipital regions, with a higher emphasis on the temporal areas, specifically the middle and inferior temporal regions [Figure 3]. When exploring these structural brain signatures in CI individuals, we observed that in MCI participants, the structural brain signature was primarily defined by structural changes in the pars opercularis and the superior parietal, together with variations in other frontal and parietal regions, along with contributions from some temporal and occipital areas as well as subcortical regions such as the amygdala and area accumbens [Figure 3]. In

AD dementia individuals, the structural brain signature associated with higher genetic risk of AD was mainly characterized by structural changes in the pars triangularis and the amygdala, along with variations in other frontal regions, as well as parietal, temporal and occipital areas [Figure 3D]. Although the prediction accuracy of the structural brain signature was low in all the groups [Supplementary Table 4], there were significant differences in the mean score of the structural brain signature between at-high and at-low risk individuals in each group along the disease *continuum* [Supplementary Figure 3].

#### 4. Discussion

In this study, the use of CoDA revealed structural brain signatures capable of discerning between CU A- individuals and subjects within other disease-stage groups along the AD *continuum*. We also found structural brain signatures associated with higher genetic predisposition to AD that were specific to each stage of the disease along the *continuum*. In both cases, the structural brain signatures reflected different structural trajectories among the groups.

On the one hand, regions involved in the brain signatures discriminating between CU A- and the rest of the groups were in line with previous reported patterns of brain networks in CU, MCI and AD<sup>36</sup>. The structural brain signature discerning between CU A- and CU A+ primarily involved subcortical areas such as the amygdala and pallidum, and cortical areas in the insula, cingulate and both frontal and parietal lobes. Despite being inversely related, the amygdala and pallidum emerged as the regions with the most significant contribution to the structural brain signature. Some authors<sup>37</sup> have found that atrophy in the amygdala predicts Mini Mental State Examination scores and hippocampal atrophy in very mild and mild AD subjects. These



findings highlight the relationship between the magnitude of atrophy in this region and cognitive impairment in very early stages of the disease. The structural brain signature that distinguished between CU A- and MCI A+, was mainly characterized by cortical regions in the temporal, frontal and occipital lobe, as well as in the cingulate cortex, together with subcortical areas such as the putamen and pallidum. Despite their inverse relationship, the fusiform and the frontal pole showed the most pronounced contribution to the structural brain signature. Recent studies have revealed different patterns of atrophy in MCI individuals, with observed thinning in both the fusiform and the frontal in distinct incident MCI subtypes<sup>38</sup>. Results highlight the complexity of structural changes in heterogeneous profiles. Finally, CU A- individuals were differentiated from AD A+ through a structural brain signature mainly involving cortical areas in the frontal and parietal lobe, together with the insula and the cingulate cortex, as well as subcortical areas such as the putamen. The insula and the inferior parietal were inversely related but had the highest contribution to the structural brain signature. Previous studies have explored cortical networks in CU, MCI and AD patients<sup>36</sup> and have identified group-specific hub regions, specifically involving the cingulate and the orbital frontal gyrus in AD patients. These findings are in line with the reported regions that we found involved in the structural brain signature associated with higher odds of belonging to the AD group. Moreover, authors<sup>36</sup> also found that compared with CU, the nodal centrality of MCI and AD decreased in the middle temporal and increased in the precuneus, located in the parietal lobe.

On the other hand, regions involved in the structural brain signatures associated with higher genetic predisposition of AD within disease-stage groups, reflected the structural atrophy trajectory observed as AD progresses, from medial temporal, to temporo-parietal and frontal regions<sup>39</sup>. The structural brain signature linked to higher risk of AD in CU individuals primarily

involved the relative change of temporal areas. Furthermore, independently of amyloid status, the structural brain signature of both CU A- and A+ individuals, shared some commonalities, such as the involvement of the pallidum and the middle temporal cortex, albeit with varying degrees of prominence. Despite these common similarities, the impact of AD genetic load on the modulation of nearby brain regions was group-specific. These findings suggest that the brain's response to AD genetic predisposition differs even within the preclinical stage, which has been already observed in univariate brain modelization<sup>17-18</sup>. In the MCI group, the structural brain signature discerning individuals at higher genetic risk of AD from those at lower risk, was characterized by relative volumetric changes in temporal and parietal regions. Specifically, MCI at higher genetic risk of AD exhibited a brain signature representing an intermediate position between CU and AD. Notably, the amygdala and accumbens played significant roles in shaping this signature. The amygdala, in particular, is well-established in the processing of emotional and memory-related information, implying a potential role in the early stages of cognitive dysfunction (Al-Ani et al., 2023). In the AD group, the structural brain signature was primarily characterized by frontal regions. The amygdala remained a crucial contributor to this signature. These findings emphasize the genetic susceptibility of the amygdala and its role in advanced stages of AD. Recent studies<sup>40</sup> have reported the association between the *APOE-ε4* genotype and reduced volumes in the amygdala, specifically in the AD dementia and MCI groups compared to young cognitively healthy individuals. However, they did not find significant associations between the volume of the amygdala and an overall polygenic score of AD. In contrast, other studies<sup>41</sup> have found significant associations between higher genetic risk of AD and lower amygdala volumes in individuals from the ADNI cohort.

Altogether, results suggested that when accounting for genetic information, group-specific structural brain signatures better capture the pattern of structural changes observed as the

disease progresses. The identification of these structural brain signatures underscores the complex interplay between genetics and brain morphology in both prodromal and clinical stages of the disease. Moreover, these findings provide a novel perspective for brain IG studies showing that the contribution of genetic factors to brain structure may evolve as the disease progresses<sup>42</sup>. Results also highlighted the impact of polygenicity on the brain morphology. While the *APOE-ε4* allele is widely recognized as the most significant risk factor for AD, our study exposes the importance of a broader genetic landscape in influencing brain structure alterations in AD<sup>43</sup>. This insight is particularly crucial for the understanding of the disease pathogenesis, suggesting that changes in brain structure are a result of a complex genetic interplay rather than being solely dependent on single genetic variants<sup>44</sup>. Besides the *APOE-ε4* allele, other polymorphisms in the *APOE* region (*i.e.* rs405509; T/T as risk allele), have been independently associated with the thickness of specific brain regions (e.g. left parahippocampal gyrus) in old cognitively healthy individuals<sup>45</sup>. Moreover, genetic variants included in the computation of the polygenic risk score of AD were annotated to multiple genes<sup>11</sup> that have been associated with MRI outcomes, such as *TOMM40*<sup>46</sup>, that is involved in a haplotype together with *APOE* (*TOMM40*'523-*APOE* ε4) that has been associated with AD cortical morphological traits<sup>47</sup>.

Despite these findings, our study also showed some limitations. Notably, we acknowledge the need for longitudinal data analysis to assess the joint volumetric trajectories discerning individuals at higher genetic risk of AD from those at lower risk over time. Another aspect to consider is that each structural brain signature captures the relative variation of brain regions' volumes that are most closely associated with the outcome of interest. Thus, the structural brain signature does not reflect actual changes on the brain volumetric composition, but relative volumetric variations. This interpretation might not be straightforward, but the methodology

overcomes biases and limitations of current IG multivariate approaches when working with heterogeneous MRI data from different cohorts. For instance, CoDA surpasses traditional methodologies in its ability to integrate volumetric data from various cohorts without the stringent requirement for harmonization<sup>48</sup>. This flexibility facilitates the analysis of diverse datasets and strengthens the reliability of cross-study comparisons. Nonetheless, another limitation lies in the involvement of other AD risk factors that may interact with genetics affecting the brain volumetric modulation. It is well known that genetic factors significantly contribute to neurodegenerative disorders, but often they are functional under specific exposures<sup>49</sup>. In this context, one could state that individuals from the ALFA cohort are exposed to sufficiently homogeneous conditions to assume an equal random impact of factors external to genetics. However, it should be noted that an extension of these results by evaluating the impact of external factors and in a more diverse population group, would improve the interpretation and translation of the results.

This study also had several strengths. The use of CoDA represents a strong method for overcoming methodological aspects of the traditional analyses used in neuroimaging studies and offers a more accurate and comprehensive understanding of brain structural alterations linked to genetics than common univariate and multivariate methods<sup>50</sup>. Unlike conventional univariate analyses, which often lack the power to detect significant associations due to rigorous multiple testing corrections at brain-wide or genome-wide level, this approach adeptly captures the complex interdependencies among different brain regions. Moreover, CoDA offers distinct advantages over commonly used multivariate methods<sup>19</sup>. While multivariate approaches reduce data dimensionality and identify correlated feature sets<sup>51</sup>, they typically overlook the compositional nature of brain data.

In summary, CoDA emerges as a relevant method in neuroimaging studies, holding significant implications for both research and clinical practice. By providing a more accurate analysis of brain imaging data and addressing its compositional nature, CoDA can influence the development of targeted approaches, opening new avenues for enhancing brain health.

## 5. References

1. Avelar-Pereira B, Belloy ME, O'Hara R, Hosseini SMH. Decoding the heterogeneity of Alzheimer's disease diagnosis and progression using multilayer networks. *Mol Psychiatry*. 2022;28(6):2423-2432.
2. Thompson PM, Hayashi KM, Dutton RA, et al. Tracking Alzheimer's disease. *Ann N Y Acad Sci*. 2007;1097:183-214.
3. Laakso MP, Frisoni GB, Könönen M, et al. Hippocampus and entorhinal cortex in frontotemporal dementia and Alzheimer's disease: a morphometric MRI study. *Biol Psychiatry*. 2000;47(12):1056-1063.
4. Teipel SJ, Grothe M, Lista S, Toschi N, Garaci FG, Hampel H. Relevance of magnetic resonance imaging for early detection and diagnosis of Alzheimer disease. *Med Clin North Am*. 2013;97(3):399-424.
5. Dawe RJ, Yu L, Arfanakis K, Schneider JA, Bennett DA, Boyle PA. Late-life cognitive decline is associated with hippocampal volume, above and beyond its associations with traditional neuropathologic indices. *Alzheimers Dement*. 2020;16(1):209-218.
6. Dong H, Guo L, Yang H, et al. Association between gray matter atrophy, cerebral hypoperfusion, and cognitive impairment in Alzheimer's disease. *Front Aging Neurosci*. 2023;15:1129051.
7. Zhao B, Ibrahim JG, Li Y, et al. Heritability of Regional Brain Volumes in Large-Scale Neuroimaging and Genetic Studies. *Cereb Cortex*. 2019;29(7):2904-2914.
8. Vilor-Tejedor N, Evans TE, Adams HH, et al. Genetic Influences on Hippocampal Subfields: An Emerging Area of Neuroscience Research. *Neurol Genet*. 2021;7(3):e591.
9. Liu CC, Liu CC, Kanekiyo T, Xu H, Bu G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat Rev Neurol*. 2013;9(2):106-118.
10. Jansen IE, Savage JE, Watanabe K, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet*.

2019;51(3):404-413.

11. Wightman DP, Jansen IE, Savage JE, et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat Genet.* 2021;53(9):1276-1282.
12. Bellenguez C, Küçükali F, Jansen IE, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet.* 2022;54(4):412-436.
13. Vilor-Tejedor N, Genius P, Rodríguez-Fernández B, et al. Genetic characterization of the ALFA study: Uncovering genetic profiles in the Alzheimer's continuum. *Alzheimers Dement.* Published online December 13, 2023. doi:10.1002/alz.13537
14. Hashimoto R, Ohi K, Yamamori H, et al. Imaging genetics and psychiatric disorders. *Curr Mol Med.* 2015;15(2):168-175.
15. Zhao B, Luo T, Li T, et al. Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. *Nat Genet.* 2019;51(11):1637-1644.
16. Kannappan B, Gunasekaran TI, Te Nijenhuis J, et al. Polygenic score for Alzheimer's disease identifies differential atrophy in hippocampal subfield volumes. *PLoS One.* 2022;17(7):e0270795.
17. Foo H, Thalamuthu A, Jiang J, et al. Associations between Alzheimer's disease polygenic risk scores and hippocampal subfield volumes in 17,161 UK Biobank participants. *Neurobiol Aging.* 2021;98:108-115.
18. Murray AN, Chandler HL, Lancaster TM. Multimodal hippocampal and amygdala subfield volumetry in polygenic risk for Alzheimer's disease. *Neurobiol Aging.* 2021;98:33-41.
19. Vilor-Tejedor N, Garrido-Martín D, Rodríguez-Fernández B, Lamballais S, Guigó R, Gispert JD. Multivariate Analysis and Modelling of multiple Brain endOphenotypes: Let's MAMBO! *Comput Struct Biotechnol J.* 2021;19:5800-5810.
20. Wang B, Caffo BS, Luo X, et al. Regularized regression on compositional trees with application to MRI analysis. *J R Stat Soc Ser C Appl Stat.* 2022;71(3):541-561.
21. Rivera-Pinto J, Egozcue JJ, Pawlowsky-Glahn V, Paredes R, Noguera-Julian M, Calle ML. Balances: A new perspective for microbiome analysis. *mSystems.* 2018;3(4). doi:10.1128/mSystems.00053-18
22. Calle ML, Pujolassos M, Susin A. coda4microbiome: compositional data analysis for microbiome cross-sectional and longitudinal studies. *BMC Bioinformatics.* 2023;24(1):82.
23. Seeley WW, Crawford RK, Zhou J, Miller BL, Greicius MD. Neurodegenerative diseases target large-scale human brain networks. *Neuron.* 2009;62(1):42-52.

24. Molinuevo JL, Gramunt N, Gispert JD, et al. The ALFA project: A research platform to identify early pathophysiological features of Alzheimer's disease. *Alzheimers Dement*. 2016;2(2):82-92.
25. Petersen RC, Aisen PS, Beckett LA, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology*. 2010;74(3):201-209.
26. Jack CR Jr, Bennett DA, Blennow K, et al. A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology*. 2016;87(5):539-547.
27. Milà-Alomà M, Salvadó G, Gispert JD, et al. Amyloid beta, tau, synaptic, neurodegeneration, and glial biomarkers in the preclinical stage of the Alzheimer's continuum. *Alzheimers Dement*. 2020;16(10):1358-1371.
28. Blauwendraat C, Faghri F, Pihlstrom L, et al. NeuroChip, an updated version of the NeuroX genotyping platform to rapidly screen for variants associated with neurological diseases. *Neurobiol Aging*. 2017;57:247.e9-e247.e13.
29. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48(10):1284-1287.
30. Weber CJ, Carrillo MC, Jagust W, et al. The Worldwide Alzheimer's Disease Neuroimaging Initiative: ADNI-3 updates and global perspectives. *Alzheimers Dement*. 2021;7(1):e12226.
31. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience*. 2019;8(7). doi:10.1093/gigascience/giz082
32. Huguet J, Falcon C, Fusté D, et al. Management and Quality Control of Large Neuroimaging Datasets: Developments From the Barcelonaβeta Brain Research Center. *Front Neurosci*. 2021;15:633438.
33. Jack CR Jr, Bernstein MA, Fox NC, et al. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J Magn Reson Imaging*. 2008;27(4):685-691.
34. Desikan RS, Ségonne F, Fischl B, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*. 2006;31(3):968-980.
35. Susin A, Wang Y, Lê Cao KA, Calle ML. Variable selection in microbiome compositional data analysis. *NAR Genom Bioinform*. 2020;2(2):lqaa029.
36. Yao Z, Zhang Y, Lin L, et al. Abnormal cortical networks in mild cognitive impairment and Alzheimer's disease. *PLoS Comput Biol*. 2010;6(11):e1001006.
37. Poulin SP, Dautoff R, Morris JC, Barrett LF, Dickerson BC, Alzheimer's Disease Neuroimaging Initiative. Amygdala atrophy is prominent in early Alzheimer's disease and relates to symptom severity. *Psychiatry Res*. 2011;194(1):7-13.



38. Machulda MM, Lundt ES, Albertson SM, et al. Cortical atrophy patterns of incident MCI subtypes in the Mayo Clinic Study of Aging. *Alzheimers Dement*. 2020;16(7):1013-1022.
39. Whitwell JL. Progression of atrophy in Alzheimer's disease and related disorders. *Neurotox Res*. 2010;18(3-4):339-346.
40. Lupton MK, Strike L, Hansell NK, et al. The effect of increased genetic risk for Alzheimer's disease on hippocampal and amygdala volume. *Neurobiol Aging*. 2016;40:68-77.
41. Biffi A, Anderson CD, Desikan RS. Genetic variation and neuroimaging measures in Alzheimer disease. *Archives of*. Published online 2010. <https://jamanetwork.com/journals/jamaneurology/article-abstract/800419>
42. Veitch DP, Weiner MW, Aisen PS, et al. Understanding disease progression and improving Alzheimer's disease clinical trials: Recent highlights from the Alzheimer's Disease Neuroimaging Initiative. *Alzheimers Dement*. 2019;15(1):106-152.
43. Chauhan G, Adams HHH, Bis JC, et al. Association of Alzheimer's disease GWAS loci with MRI markers of brain aging. *Neurobiol Aging*. 2015;36(4):1765.e7-e1765.e16.
44. Harrison JR, Foley SF, Baker E, et al. Pathway-specific polygenic scores for Alzheimer's disease are associated with changes in brain structure in younger and older adults. *Brain Commun*. 2023;5(5):fcad229.
45. Chen Y, Li P, Gu B, et al. The effects of an APOE promoter polymorphism on human cortical morphology during nondemented aging. *J Neurosci*. 2015;35(4):1423-1431.
46. Roses AD, Lutz MW, Amrine-Madsen H, et al. A TOMM40 variable-length polymorphism predicts the age of late-onset Alzheimer's disease. *Pharmacogenomics J*. 2010;10(5):375-384.
47. Honea RA, Hunt S, Lepping RJ, et al. Alzheimer's disease cortical morphological phenotypes are associated with TOMM40'523-APOE haplotypes. *Neurobiol Aging*. 2023;132:131-144.
48. Richter S, Winzeck S, Correia MM, et al. Validation of cross-sectional and longitudinal ComBat harmonization methods for magnetic resonance imaging data on a travelling subject cohort. *Neuroimage Rep*. 2022;2(4):None.
49. Wang C, Sun J, Guillaume B, et al. A Set-Based Mixed Effect Model for Gene-Environment Interaction and Its Application to Neuroimaging Phenotypes. *Front Neurosci*. 2017;11:191.
50. Xin Y, Sheng J, Miao M, Wang L, Yang Z. A review of imaging genetics in Alzheimer's disease. *Journal of Clinical*. Published online 2022. [https://www.sciencedirect.com/science/article/pii/S0967586822001758?casa\\_token=skJ4HXq0yTYAAAAA:5m3tFExzT2mYkKcSv6P7alKO3XHz7Aw3p-zfrzTGukiEzW8e5Y4bsNeXE1OWX6QaVGILHjJqdU](https://www.sciencedirect.com/science/article/pii/S0967586822001758?casa_token=skJ4HXq0yTYAAAAA:5m3tFExzT2mYkKcSv6P7alKO3XHz7Aw3p-zfrzTGukiEzW8e5Y4bsNeXE1OWX6QaVGILHjJqdU)



51. Cruciani F, Altmann A, Lorenzi M, Menegaz G, Galazzo IB. What PLS can still do for Imaging Genetics in Alzheimer's disease. In: *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE; 2022:1-4.

## Acknowledgements

This publication is part of the ALFA study. The authors express their most sincere gratitude to the ALFA project participants and relatives without whom this research would have not been possible. Collaborators of the ALFA study are: Müge Akinci, Federica Anastasi, Annabella Beteta, Raffaele Cacciaglia, Lidia Canals, Alba Cañas, Carme Deulofeu, Maria Emilio, Irene Cumplido-Mayoral, Marta del Campo, Carme Deulofeu, Ruth Dominguez, Maria Emilio, Karine Fauria, Sherezade Fuentes, Marina García, Laura Hernández, Gema Huesa, Jordi Huguet, Laura Iglesias, Esther Jiménez, David López-Martos, Paula Marne, Tania Menchón, Paula Ortiz-Romero, Eleni Palpatzis, Wiesje Pelkmans, Albina Polo, Sandra Pradas, Mahnaz Shekari, Lluís Solsona, Anna Soteras, Núria Tort-Colet, and Marc Vilanova.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson

Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health ([www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

## **Funding**

The research leading to these results has received funding from “la Caixa” Foundation (ID 100010434), under agreement LCF/PR/GN17/50300004, the Health Department of the Catalan Government (Health Research and Innovation Strategic Plan (PERIS) 2016-2020 grant# SLT002/16/00201) and the Alzheimer’s Association and an international anonymous charity foundation through the TriBEKa Imaging Platform project (TriBEKa-17-519007). Additional support has been received from the Universities and Research Secretariat, Ministry of Business and Knowledge of the Catalan Government under the grant no. 2021 SGR 00913. All CRG authors acknowledge the support of the Spanish Ministry of Science, Innovation, and Universities to the EMBL partnership, the Centro de Excelencia Severo Ochoa, and the CERCA Programme/Generalitat de Catalunya. NV-T was supported by the Spanish Ministry of Science and Innovation - State Research Agency (IJC2020-043216-1/MCIN/AEI/10.13039/501100011033) and the European Union «NextGenerationEU»/PRTR

and currently receives funding from the Spanish Research Agency MICIU/AEI/10.13039/501100011033 (grant RYC2022-038136-I cofunded by the European Union FSE+ and grant PID2022-143106OA-I00 cofunded by the European Union FEDER).

### **Conflicts of interest**

JDG has served as a consultant for Roche Diagnostics and Prothena Biosciences; he has given lectures at symposiums sponsored by General Electric, Philips, Esteve, Life-MI and Biogen; and he received research support from GE Healthcare, Roche Diagnostics, and Hoffmann-La Roche. The remaining co-authors have no conflicts to disclose.

### **Consent Statement**

The ALFA study was conducted in accordance with the directives of the Spanish Law 14/2007, of July 3 , on Biomedical Research (Ley 14/2007 de Investigación Biomédica). The ALFA study protocol was approved by the Independent Ethics Committee Parc de Salut Mar Barcelona and registered at Clinicaltrials.gov (Identifier: NCT01835717). All participants accepted the study procedures by signing the study's informed consent form that had also been approved by the same Institutional Review Board.

## Tables

	<b>ALFA N=338</b>	<b>N</b>
<b>Sex (n,%)</b>		338
<b>Men</b>	127 (37.574%)	
<b>Women</b>	211 (62.426%)	
<b>Age (median, IQR)</b>	57.000 [53.000;61.000]	338
<b>Education years (median, IQR)</b>	12.000 [11.000;17.000]	338
<b>Amyloid-beta status (n,%)</b>		338
<b>AB-</b>	220 (65.089%)	
<b>AB+</b>	118 (34.911%)	
<b>APOE ε4 carriers (n,%)</b>		338
<b>Non-carrier</b>	155 (45.858%)	
<b>Carrier</b>	183 (54.142%)	
<b>APOE ε4 load (n,%)</b>		338
<b>Non-carriers</b>	155 (45.858%)	
<b>One ε4 allele</b>	152 (44.970%)	
<b>Two ε4 alleles</b>	31 (9.172%)	
<b>Minimental State Examination (median, IQR)</b>	29.000 [28.000;30.000]	338

**Table 1.** Demographic characteristics and APOE genotypes distribution in the ALFA sample

	<b>ADNI N=330</b>	<b>N</b>
<b>Sex (n,%)</b>		330
<b>Women</b>	132 (40.000%)	
<b>Men</b>	198 (60.000%)	
<b>Age (median, IQR)</b>	74.050 [68.200;78.000]	330
<b>Education years (median, IQR)</b>	16.000 [14.000;18.000]	330
<b>Diagnosis (n,%)</b>		330
<b>AD</b>	100 (30.303%)	
<b>MCI</b>	230 (69.697%)	
<b>APOE ε4 carriers (n,%)</b>		330
<b>Non-carrier</b>	117 (35.455%)	
<b>Carrier</b>	213 (64.545%)	
<b>APOE ε4 load (n,%)</b>		330
<b>Non-carriers</b>	117 (35.455%)	
<b>One ε4 allele</b>	153 (46.364%)	
<b>Two ε4 alleles</b>	60 (18.182%)	
<b>Minimental State Examination (median, IQR)</b>	27.000 [25.000;29.000]	330

**Table 2.** Demographic characteristics and APOE genotypes distribution in the ADNI sample

	CU A- ALFA N=220	CU A+ ALFA N=118	MCI A+ ADNI N=230	AD A+ ADNI N=100	p-value	p-value CU A- vs CU A+	p-value CU A- vs MCI	p-value AD vs CU A-	p-value CU A+ vs MCI	p-value AD vs CU A+	p-value AD vs MCI
<b>Sex (n,%)</b>					<b>&lt;0.001</b>	0.784		<b>0.002</b>		<b>0.018</b>	0.649
<b>Women</b>	139 (63.182%)	72 (61.017%)	89 (38.696%)	43 (43.000%)			<b>&lt;0.001</b>		<b>&lt;0.001</b>		
<b>Men</b>	81 (36.818%)	46 (38.983%)	141 (61.304%)	57 (57.000%)							
<b>Age (median, IQR)</b>	56.000 [52.750;60.000]	58.000 [54.000;62.000]	72.850 [68.125;77.400]	75.600 [69.550;80.300]	<b>&lt;0.001</b>	<b>0.004</b>		<b>&lt;0.001</b>			0.025
<b>Education years (median, IQR)</b>	12.500 [11.000;17.000]	12.000 [10.250;17.000]	16.000 [14.000;18.000]	16.000 [14.000;18.000]	<b>&lt;0.001</b>	0.495		<b>&lt;0.001</b>		<b>&lt;0.001</b>	0.105
<b>APOE ε4 carriers (n,%)</b>					<b>&lt;0.001</b>	<b>&lt;0.001</b>		<b>&lt;0.001</b>		<b>&lt;0.001</b>	0.317
<b>Non-carrier</b>	130 (59.091%)	25 (21.186%)	86 (37.391%)	31 (31.000%)			<b>&lt;0.001</b>		<b>&lt;0.001</b>		
<b>Carrier</b>	90 (40.909%)	93 (78.814%)	144 (62.609%)	69 (69.000%)							
<b>APOE ε4 load (n,%)</b>					<b>&lt;0.001</b>	<b>&lt;0.001</b>		<b>&lt;0.001</b>		<b>0.007</b>	0.170
<b>Non-carriers</b>	130 (59.091%)	25 (21.186%)	86 (37.391%)	31 (31.000%)			<b>&lt;0.001</b>				
<b>One ε4 allele</b>	79 (35.909%)	73 (61.864%)	104 (45.217%)	49 (49.000%)							
<b>Two ε4 alleles</b>	11 (5.000%)	20 (16.949%)	40 (17.391%)	20 (20.000%)							
<b>PRS-AD (median, IQR)</b>	-0.038 [- 0.155;0.090]	0.056 [-0.086;0.220]	0.033 [-0.112;0.204]	0.070 [-0.054;0.234]	<b>&lt;0.001</b>	<b>&lt;0.001</b>	<b>&lt;0.001</b>	<b>&lt;0.001</b>	0.559	0.459	0.150
<b>PRS-ADnoAPOE (median, IQR)</b>	-0.11 [-0.18;-0.02]	-0.12 [-0.22;0.03]	-0.10 [-0.19;-0.01]	-0.08 [-0.19;0.02]	0.605	0.903	0.708	0.708	0.921	0.708	0.708
<b>Minimental State Examination (median, IQR)</b>	29.000 [28.000;30.000]	29.000 [29.000;30.000]	28.000 [27.000;29.000]	23.000 [21.750;25.000]	<b>&lt;0.001</b>	0.303	<b>&lt;0.001</b>	<b>&lt;0.001</b>	<b>&lt;0.001</b>	<b>&lt;0.001</b>	<b>&lt;0.001</b>

**Table 3.** Demographic characteristics and APOE genotypes distribution in the ALFA and ADNI sample.



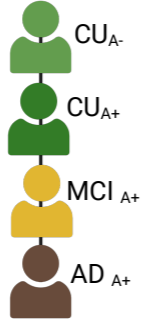
## 1. Sample of the study and genetic characterization



**ALFA sample**  
N=338 CU individuals  
N CU<sub>A-</sub> = 220  
N CU<sub>A+</sub> = 118



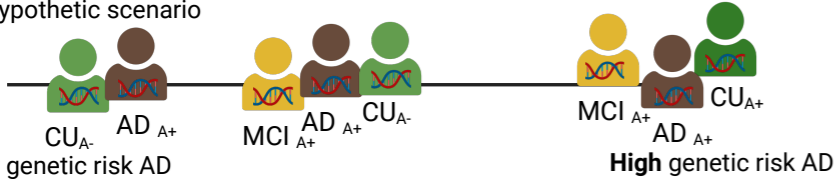
**ADNI sample**  
N=330 CI A+ individuals  
N CI<sub>MCI A+</sub> = 230  
N CI<sub>AD A+</sub> = 100



**AD continuum** (clinical diagnosis and amyloid profile)

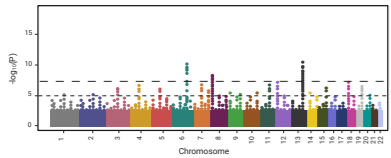
### What about the genetic profile?

Hypothetic scenario



### PRS<sub>AD</sub> computation

#### 1. Data acquisition - GWAS summary statistics (Wightman et al., 2021)



Effect sizes in the population of reference

#### 2. PRS computation



LD  $R_2=0.1$   
window 250kb

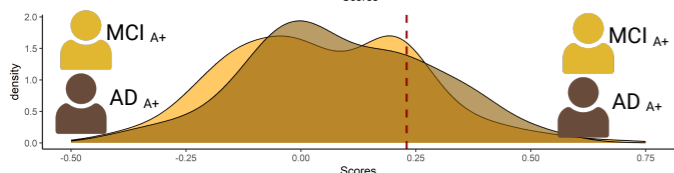
SNPs inclusion:  $p\text{-value} < 5 \cdot 10^{-6}$

$$PRS_j = \sum_i \frac{S_i \times G_{ij}}{M_j}$$

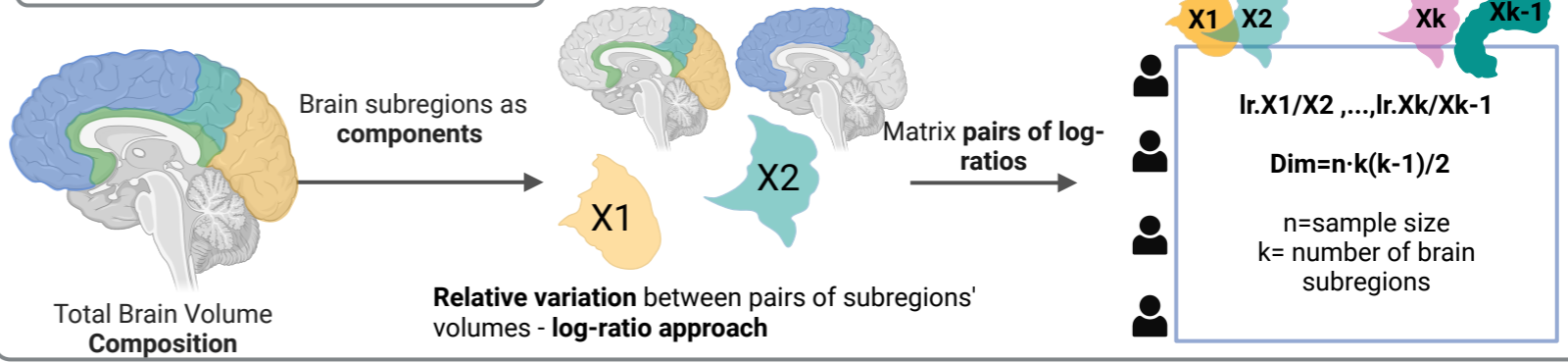
Non-high risk

**Cohort-specific PRS<sub>AD</sub>**

High risk



## 2. Compositional dataset

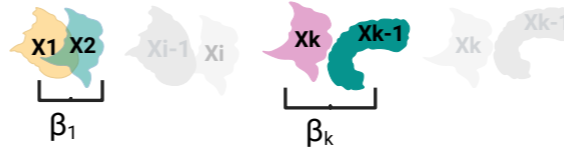


## 3. Modeling and variable selection

Aim: to identify **structural brain signatures** (combination of specific brain subregion's volumes)

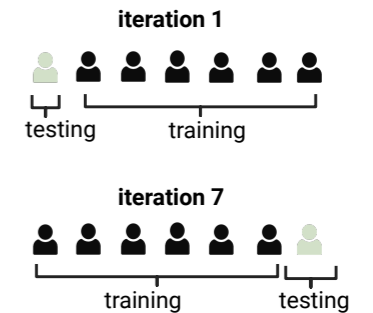
- (i) capable of **distinguishing** between **CU A-** and the rest of the groups in the disease **continuum**.
- (ii) associated with **higher genetic risk of AD** in individuals at different stages of the disease

### 1. Variable selection through elastic net penalized regression

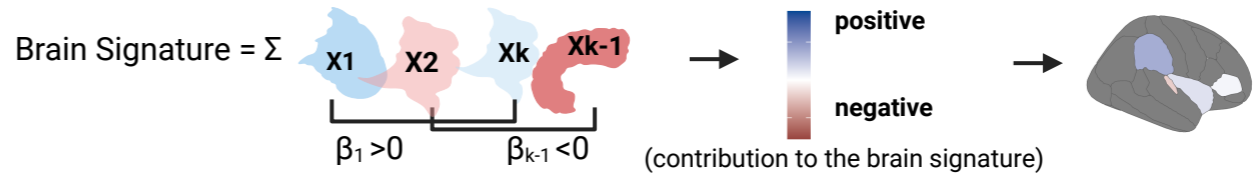


Cross-validation

- $\alpha=0.9$
- $\lambda=\text{lambda}1.\text{se}$
- N-fold=7



### 2. The joint effect of the selected brain regions is summarized in a brain structural signature (log-contrast function)



3. The joint change of the structural brain signature and the binary outcomes of interest, after adjusting for covariates, is given by logistic regression models.

$$(i) \text{logit}(CU_{A-} \text{ vs } Group_i) = \beta_0 + \sum \beta_i \cdot X_i + \gamma_1 \cdot Age + \gamma_2 \cdot Sex + \gamma_3 \cdot PRS_{AD}$$

$$(ii) \text{logit}(Non\text{-}high \text{ vs } High \text{ risk } AD) = \beta_0 + \sum \beta_i \cdot X_i + \gamma_1 \cdot Age + \gamma_2 \cdot Sex, \text{ stratified by disease-stage}$$

where  $\sum \beta_i \cdot X_i = \text{Brain Signature}$

**Figure 1.** Workflow *coda4microbiome* algorithm: implementation in our Imaging Genetics study.

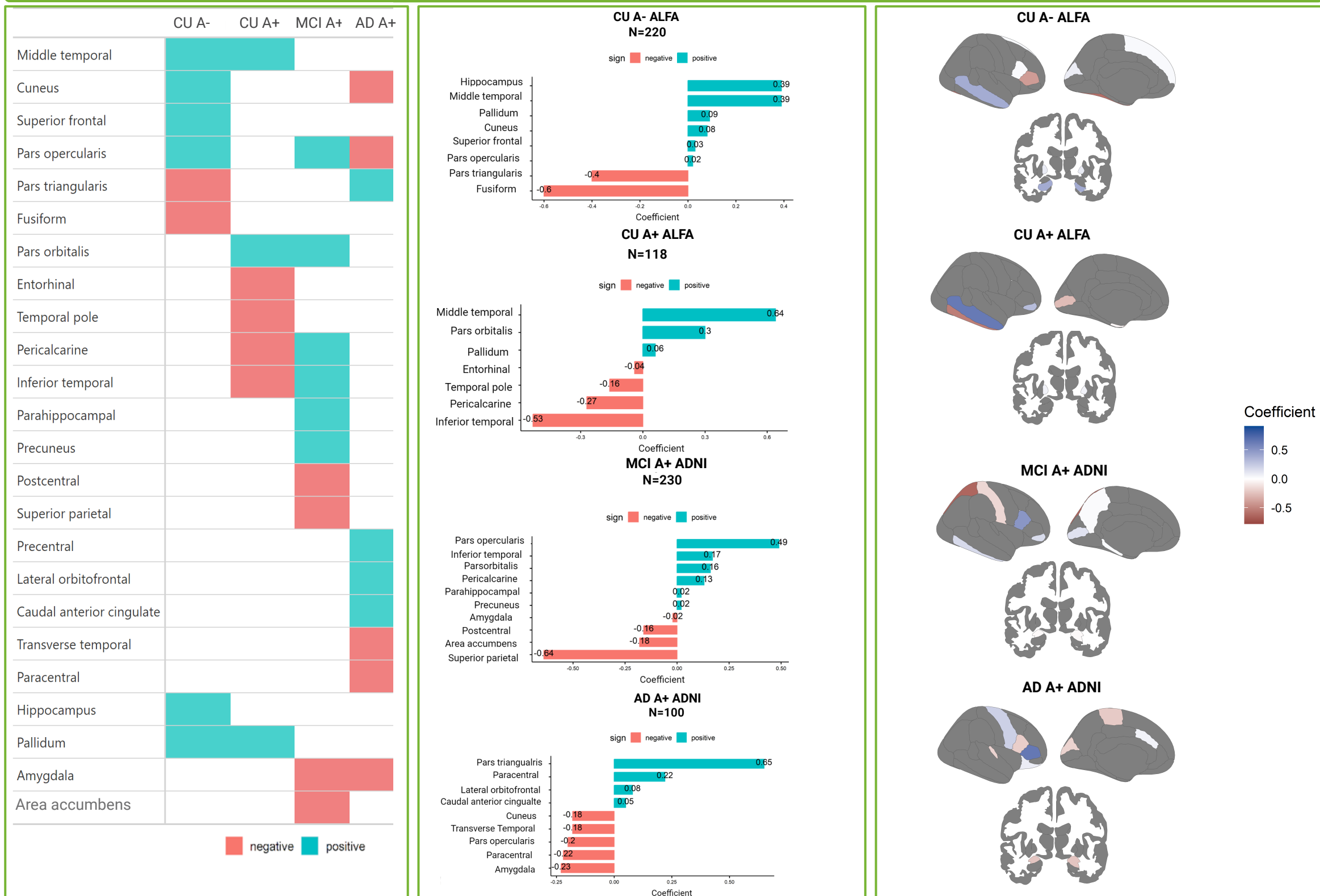
# Structural brain signatures that discriminate between CU A- and other disease-stage groups in the AD continuum



**Figure 2.** Brain structural signature distinguishing between CU A- individuals and others at different stages of the disease. In blue, brain regions that positively contribute to the structural brain signature. In red, brain regions that negatively contribute to the structural brain signature. Legend: CU A- (cognitively unimpaired amyloid-beta negative individuals), CU A+ (cognitively unimpaired amyloid-beta positive individuals), MCI A+ (mild cognitive impaired amyloid-beta positive individuals), AD A+ (Alzheimer's disease amyloid-beta positive individuals).



# Structural brain signatures associated with higher genetic risk of AD (PRS<sub>AD</sub>) along the AD continuum



**Figure 3.** Brain structural signature associated with higher genetic predisposition to AD (PRS<sub>AD</sub>). In blue, brain regions that positively contribute to the structural brain signature. In red, brain regions that negatively contribute to the structural brain signature. *Legend: CU A-* (cognitively unimpaired amyloid-beta negative individuals), *CU A+* (cognitively unimpaired amyloid-beta positive individuals), *MCI A+* (mild cognitive impaired amyloid-beta positive individuals), *AD A+* (Alzheimer's disease amyloid-beta positive individuals).