# Genetic, clinical underpinnings of subtle early brain change along Alzheimer's dimensions

Junhao Wen[1*], Zhijian Yang[1], Ilya M. Nasrallah[1], Yuhan Cui[1], Guray Erus[1], Dhivya Srinivasan[1], Ahmed Abdulkadir[1,2], Elizabeth Mamourian[1], Gyujoon Hwang[1], Ashish Singh[1], Mark Bergman[1], Jingxuan Bao[3], Erdem Varol[4], Zhen Zhou[1], Aleix Boquet-Pujadas[5], Jiong Chen[1], Arthur Toga[6], Andrew J. Saykin[7], Timothy J. Hohman[8], Paul M. Thompson[9], Sylvia Villeneuve[10], Randy Gollub[11], Aristeidis Sotiras[12], Katharina Wittfeld[13,14], Hans J. Grabe[13,14], Duygu Tosun[15], Murat Bilgel[16], Yang An[16], Daniel S. Marcus[17], Pamela LaMontagne[17], Susan R. Heckbert[18], Thomas R. Austin[18], Lenore J. Launer[19], Mark Espeland[20], Colin L Masters[21], Paul Maruff[21], Jurgen Fripp[22], Sterling C. Johnson[23], John C. Morris[24], Marilyn S. Albert[25], R. Nick Bryan[26], Susan M. Resnick[16], Luigi Ferrucci[27], Yong Fan[1], Mohamad Habes[28], David Wolk[1, 29], Li Shen[3], Haochang Shou[1,30], Christos Davatzikos[1*], for the iSTAGING, the AI4AD, and the ADSP phenotypic harmonization consortia, the BLSA, the PREVENT-AD, and the ADNI studies

[1]Artificial Intelligence in Biomedical Imaging Laboratory (AIBIL), Center for AI and Data Science for Integrated Diagnostics (AI$^2$D), Perelman School of Medicine, University of Pennsylvania, Philadelphia, USA.
[2]Research Lab in Neuro-imaging of the Department of Clinical Neurosciences at Lausanne University Hospital, Lausanne, Switzerland
[3]Department of Biostatistics, Epidemiology and Informatics University of Pennsylvania Perelman School of Medicine, Philadelphia, USA
[4]Department of Statistics, Center for Theoretical Neuroscience, Zuckerman Institute, Columbia University, New York, New York
[5]Biomedical Imaging Group, EPFL, Lausanne, Switzerland
[6]Laboratory of Neuro Imaging, Stevens Neuroimaging and Informatics Institute, Keck School of Medicine of USC, University of Southern California, Marina del Rey, California
[7]Radiology and Imaging Sciences, Center for Neuroimaging, Department of Radiology and Imaging Sciences, Indiana Alzheimer's Disease Research Center and the Melvin and Bren Simon Cancer Center, Indiana University School of Medicine, Indianapolis
[8]Vanderbilt Memory and Alzheimer's Center, Vanderbilt Genetics Institute, Department of Neurology, Vanderbilt University Medical Center, Nashville, TN USA
[9]Imaging Genetics Center, Mark and Mary Stevens Neuroimaging and Informatics Institute, Keck School of Medicine of USC, University of Southern California, Marina del Rey, California
[10]Douglas Mental Health University Institute, McGill University, Montréal, Québec, Canada
[11]Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Charlestown, Massachusetts
[12]Department of Radiology and Institute for Informatics, Washington University School of Medicine, St. Louis, USA
[13]German Center for Neurodegenerative Diseases (DZNE), Site Rostock/ Greifswald, Greifswald, Germany
[14]Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany
[15]Department of Radiology and Biomedical Imaging, University of California, San Francisco, CA, USA
[16]Laboratory of Behavioral Neuroscience, National Institute on Aging, NIH, USA
[17]Department of Radiology, Washington University School of Medicine, St. Louis, Missouri, USA
[18]Cardiovascular Health Research Unit and Department of Epidemiology, University of Washington, Seattle, WA, USA
[19]Neuroepidemiology Section, Intramural Research Program, National Institute on Aging, Bethesda, Maryland, USA
[20]Sticht Center for Healthy Aging and Alzheimer's Prevention, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA
[21]Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Parkville, VIC, Australia
[22]CSIRO Health and Biosecurity, Australian e-Health Research Centre CSIRO, Brisbane, Queensland, Australia
[23]Wisconsin Alzheimer's Institute, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin, USA
[24]Knight Alzheimer Disease Research Center, Washington University in St. Louis, St. Louis, MO, USA
[25]Department of Neurology, Johns Hopkins University School of Medicine, USA

[26]Department of Radiology, University of Pennsylvania, Philadelphia, USA.
[27]Translational Gerontology Branch, Longitudinal Studies Section, National Institute on Aging, National Institutes of Health, MedStar Harbor Hospital, 3001 S. Hanover Street, Baltimore, MD, 21225, USA
[28]Glenn Biggs Institute for Alzheimer's & Neurodegenerative Diseases, University of Texas Health Science Center at San Antonio, San Antonio, USA
[29]Department of Neurology and Penn Memory Center, University of Pennsylvania, Philadelphia, USA
[30]Penn Statistics in Imaging and Visualization Center, Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, USA

[*]Corresponding authors:
Junhao Wen, Ph.D. – junhao.wen89@gmail.com
Christos Davatzikos, Ph.D. – Christos.Davatzikos@pennmedicine.upenn.edu
3700 Hamilton Walk, 7th Floor, Philadelphia, PA 19104

Word counts: 3547 words

## Abstract

Alzheimer's disease (AD) is associated with heterogeneous atrophy patterns,[1,2] which are increasingly manifested throughout the disease course, driven by underlying neuropathologic processes. Herein, we show that manifestations of these brain changes in early asymptomatic stages can be detected via a novel deep semi-supervised representation learning method.[3] We first identified two dominant dimensions of brain atrophy in symptomatic mild cognitive impairment (MCI) and AD patients[4]: the "diffuse-AD" (R1) dimension shows widespread brain atrophy, and the "MTL-AD" (R2) dimension displays focal medial temporal lobe (MTL) atrophy. Critically, only R2 was associated with known genetic risk factors (e.g., *APOE ε4*) of AD in MCI and AD patients at baseline. We then showed that brain changes along these two dimensions were independently detected in early stages in a cohort representative of the general population[5] and two cognitively unimpaired cohorts of asymptomatic participants.[6,7] In the general population, genome-wide association studies found 77 genes unrelated to *APOE* differentially associated with R1 and R2. Functional analyses revealed that these genes were overrepresented in differentially expressed gene sets in organs beyond the brain (R1 and R2), including the heart (R1) and the pituitary gland, muscle, and kidney (R2). These genes were also enriched in biological pathways implicated in dendritic cells (R2), macrophage functions (R1), and cancer (R1 and R2). The longitudinal progression of R1 and R2 in the cognitively unimpaired populations, as well as in individuals with MCI and AD, showed variable associations with established AD risk factors, including *APOE ε4*, tau, and amyloid. Our findings deepen our understanding of the multifaceted pathogenesis of AD beyond the brain. In early asymptomatic stages, the two dimensions are associated with diverse pathological mechanisms,

including cardiovascular diseases[8], inflammation,[9–11] and hormonal dysfunction[12,13] – driven by genes different from *APOE* – which collectively contribute to the early pathogenesis of AD.

## Main

Alzheimer's disease (AD) is the most common cause of dementia in older adults and remains incurable despite many pharmacotherapeutic clinical trials, including anti-amyloid drugs[14,15] and anti-tau drugs.[16] This is largely due to the complexity and multifaceted nature of the underlying neuropathological processes leading to dementia. The research community has embraced several mechanistic hypotheses to elucidate AD pathogenesis.[17–19] Among these, the amyloid hypothesis has been dominant over the past decades and has proposed a dynamic biomarker chain: extracellular beta-amyloid (Aβ) triggers a cascade that leads to subsequent intracellular neurofibrillary tangles, including hyperphosphorylated tau protein (tau and p-tau), neurodegeneration, including medial temporal lobe atrophy, and cognitive decline.[20,21] However, the amyloid hypothesis has been reexamined and revised due to substantial evidence which questions its current form[21–23]. While amyloid remains critical to AD development, the amyloid cascade model has been continually refined as other biological factors are discovered to influence the pathway from its accumulation to cell death.

Cardiovascular dysfunction has been widely associated with an increased risk for AD.[8] There is also growing evidence that inflammatory[10–12] and neuroendocrine processes[5,13] influence pathways of amyloid accumulation and neuronal death. The inflammation hypothesis claims that microglia and astrocytes release pro-inflammatory cytokines as drivers, byproducts, or beneficial responses associated with AD progression and severity.[11,24,25] The neuroendocrine hypothesis, first introduced in the context of aging[26], has been extended to AD[13], where it proposes that neurohormones secreted by the pituitary and other essential endocrine glands can affect the central nervous system (CNS), which subsequently contribute to developing AD. For example, Xiong and colleagues[12] recently found that blocking the action of follicle-stimulating hormone in

mice abrogates the AD-like phenotype (e.g., cognitive decline) by inhibiting the neuronal C/EBPβ–δ-secretase pathway. These findings emphasize the need to further elucidate early brain and body changes well before they lead to irreversible clinical progression.[27]

Recent advances in artificial intelligence (AI), especially deep learning (DL), applied to magnetic resonance imaging (MRI), showed great promise in biomedical applications[28,29]. DL models discover complex non-linear relationships between phenotypic and genetic features and clinical outcomes, thereby providing informative imaging-derived endophenotypes[30]. In particular, AI has been applied to MRI to disentangle the neuroanatomical heterogeneity of AD with categorical disease subtypes.[1,2,31] The genetic underpinnings[32,33] of this neuroanatomical heterogeneity in AD are also complex and heterogeneous. The most recent large-scale genome-wide association study[32] (GWAS: 111,326 AD vs. 677,633 controls) has identified 75 genomic loci, including *APOE* genes, associated with AD. However, such case-control group comparisons conceal genetic factors that might contribute differentially to different dimensions of AD-related brain change. More importantly, the genetic variants that contribute to the initiation and early progression of brain change in younger and asymptomatic individuals are poorly understood.

The current study models disease neuroanatomical heterogeneity using a newly-developed semi-supervised representation deep learning method, Surreal-GAN.[3] It provides multiple continuous dimensional scores (i.e., neuroanatomical dimensions) instead of categorical subtypes,[1] considering disease heterogeneity spatially and temporally. These multiple dimensional scores reflect the co-expression level of respective brain atrophy dimensions. Refer to **Method 1** and **Supplementary eMethod 1** for more methodological details and strengths of semi-supervised representation learning. We hypothesized that genetic variants, potentially unrelated to *APOE* genes, contribute to early manifestations of multiple dimensions of brain

atrophy in early asymptomatic stages. To test this hypothesis, we first defined the Surreal-GAN dimensions in late symptomatic stages and then examined their expression back to early asymptomatic stages. In our previous study,[3] we derived two neuroanatomical dimensions (R1 and R2) by applying Surreal-GAN to the MCI/AD participants and cognitively unimpaired (CU) participants from the Alzheimer's Disease Neuroimaging Initiative study (ADNI[4]). Herein, we applied the trained model to three asymptomatic populations and one symptomatic population: the *general population* (*N*=39,575; age: 64.12±7.54 years) from the UK Biobank (UKBB[5]) excluding demented individuals; the *cognitively unimpaired population* (*N*=1658; age: 65.75±10.90 years) from ADNI and the Baltimore Longitudinal Study of Aging study (BLSA[7]); the *cognitively unimpaired population with a family risk* (*N*=343; age: 63.63±5.05 years) from the Pre-symptomatic Evaluation of Experimental or Novel Treatments for Alzheimer's Disease (PREVENT-AD[6]); the *MCI/AD population* (*N*=1534; age: 73.45±7.69 years) from ADNI and BLSA. Refer to **Method 2** and **Table 1** for details of the definition of these populations.

## Results

### Two dominant dimensions of brain atrophy found in MCI and AD

In MCI/AD patients, the “diffuse-AD” dimension (R1) showed widespread brain atrophy without an exclusive focus on the medial temporal lobe (**Fig. 1A** and **Supplementary eTable 1** for P-values and effect sizes). In contrast, the “MTL-AD” dimension (R2) displayed more focal medial temporal lobe atrophy, prominent in the bilateral parahippocampal gyrus, hippocampus, and entorhinal cortex (**Fig. 1A**). All results, including P-values and effect sizes (Pearson’s correlation coefficient *r*), are presented in **Supplementary eTable 1**. The atrophy patterns of the two dimensions defined in the symptomatic MCI/AD population (**Fig. 1A**) were present in the

asymptomatic populations, albeit with a smaller magnitude of *r*. (**Supplementary eTable 1**, **4**, and **8**). We visualized the expression of R1 and R2 and the age distribution in the four populations in **Supplementary eFigure 1**. The MCI/AD population exhibited more salient AD signatures than the three asymptomatic populations.

### *APOE* genes are associated with R2 but not with R1 in the MCI/AD population

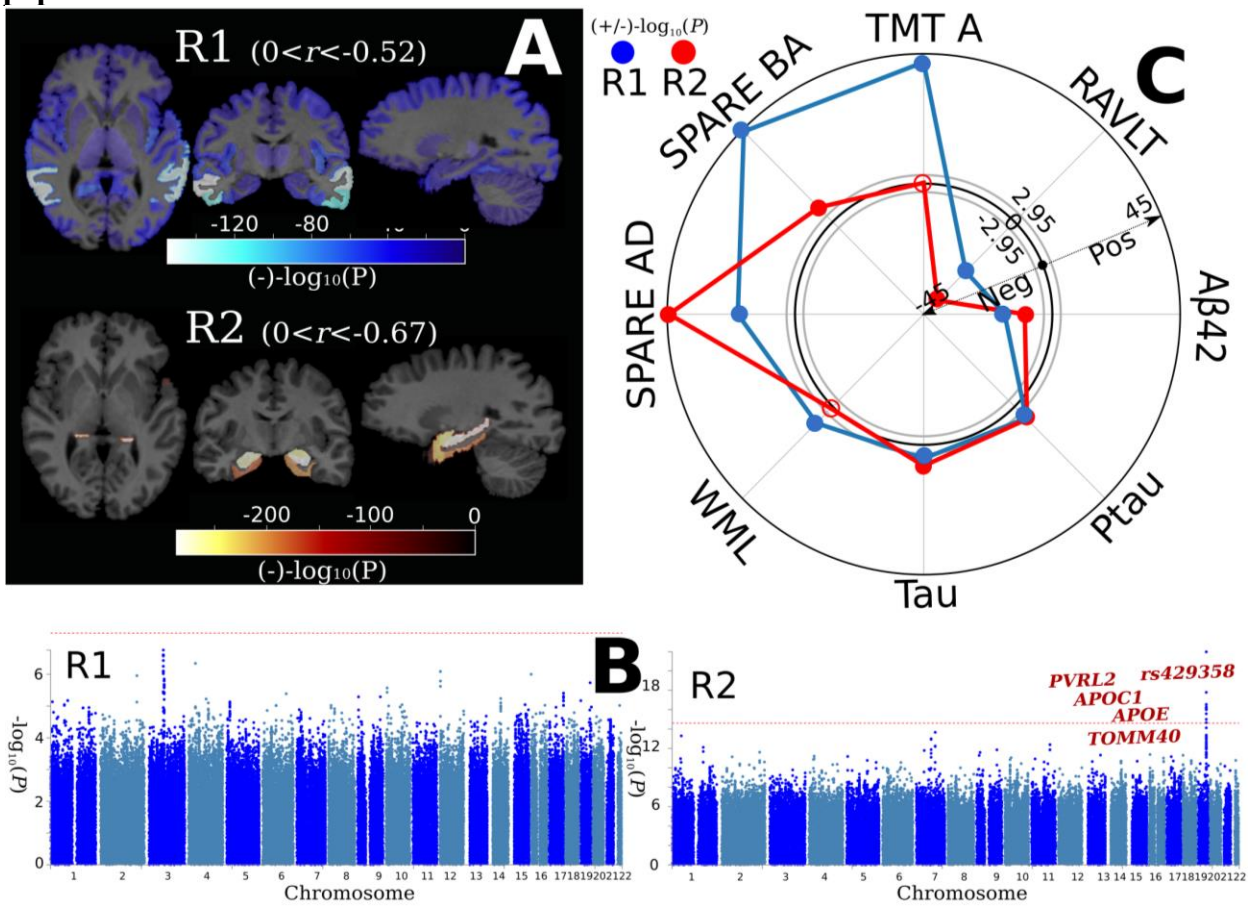
In GWAS (**Method 6 B** & **C**), the R2 dimension, but not R1, was associated with well-established AD genomic loci (*rs429358*, chromosome: 19, 45411941; minor allele: C, P-value: $1.05 \times 10^{-11}$) and genes (*APOE*, *PVRL2*, *TOMM40*, and *APOC1*) (**Fig. 1B**, **Method 6B**). The details of the identified genomic locus and annotated genes are presented in **Supplementary eTable 2**. The polygenic risk scores of AD (**Method 6F**) showed a slightly stronger positive association with the R2 dimension [$r$=0.11, $-\log_{10}$(P-value)=3.14] than with the R1 dimension [$r$=0.09, $-\log_{10}$(P-value)=2.31, **Supplementary eFigure 2**]. The QQ plots of baseline GWAS are presented in **Supplementary eFigure 3**.

### Clinical profiles of the R1 and R2 dimensions in the MCI/AD population

Clinical association studies (**Method 5**) correlated the two dimensions with 45 clinical variables and biomarkers. Compared to the R1 dimension, R2 showed associations, to a larger extent than R1, with SPARE-AD and the Rey Auditory Verbal Learning Test (RAVLT). SPARE-AD quantifies the presence of a typical imaging signature of AD-related brain atrophy, which has been previously shown to predict clinical progression in both CU and MCI individuals.[34] RAVLT measures episodic memory, a reliable neuropsychological phenotype in AD, which is also correlated with medial temporal lobe atrophy.[35,36] The R1 dimension was associated to a

greater extent with 1) SPARE-BA, which captures the individualized expression of advanced brain age from MRI[37]; 2) white matter lesions (WML), which are commonly associated with vascular risk factors and cognitive decline[38], and 3) whole-brain uptake of 18F-fluorodeoxyglucose (FDG) PET, which is a biomarker of brain metabolic function and atrophy. Both dimensions were positively associated with cerebrospinal fluid (CSF) levels of tau and p-tau and negatively associated with the CSF level of Aβ42[39] (**Fig. 1C**), as well as the whole-brain standardized uptake value ratio of 18F-AV-45 PET (**Supplementary eTable 3**). Results for all 45 clinical variables, including cognitive scores, modifiable risk factors, CSF biomarkers, disease/condition labels, demographic variables, and imaging-derived phenotypes, are presented in **Supplementary eTable 3** for P-values and effect sizes (i.e., beta coefficients).

**Figure 1**: **The manifestation of the R1 and R2 dimensions of brain atrophy in the MCI/AD population.**




**A)** Brain association studies (**Method 4**) reveal two dominant brain atrophy dimensions. A linear regression model was fit to the 119 GM ROIs at baseline for the R1 and R2 dimensions. The $-\log_{10}$(P-value) of each significant ROI (Bonferroni correction for the number of 119 ROIs: $-\log_{10}$(P-value) > 3.38) is shown. A negative value denotes brain atrophy with a negative coefficient in the linear regression model. P-value and effect sizes (*r*, Pearson's correlation coefficient) are presented in **Supplementary eTable 1**. The range of *r* for each dimension is also shown. Of note, the sample size (*N*) for R1 and R2 is the same for each ROI. **B)** Genome-wide association studies (**Method 6**) demonstrate that the R2, but not R1, dimension is associated with variants related to *APOE* genes (genome-wide P-value threshold with the red line: $-\log_{10}$(P-value) > 7.30). We associated each common variant with R1 and R2 using the whole-genome sequencing data from ADNI. Gene annotations were performed via positional, expression quantitative trait loci, and chromatin interaction mappings using FUMA.[40] We then manually queried whether they were previously associated with AD-related traits in the GWAS Catalog.[41] Red-colored loci/genes indicate variants associated with AD-related traits in previous literature. **C)** Clinical association studies (**Method 5**) show that the R2 dimension is associated to a larger extent with AD-specific biomarkers, including SPARE-AD[34], an imaging surrogate to AD atrophy patterns, and *APOE ε4*, the well-established risk allele in sporadic AD. The R1 dimension is associated to a larger extent with aging (e.g., SPARE-BA,[37] an imaging surrogate for brain aging) and vascular-related biomarkers (e.g., WML, white matter lesion). The same linear regression model was used to associate the R1 and R2 dimensions with the 45 clinical

variables, including cognitive scores, modifiable risk factors, CSF biomarkers, disease/condition labels, demographic variables, and imaging-derived phenotypes. The radar plot shows representative clinical variables; results for all 45 clinical variables are presented in **Supplementary eTable 3**. The SPARE-AD and SPARE-BA scores are rescaled for visualization purposes. The gray-colored circle lines indicate the P-value threshold in both directions (Bonferroni correction for the 45 variables: $-\log_{10}$(P-value) > 2.95). A positive/negative $-\log_{10}$(P-value) value indicates a positive/negative correlation (beta). The transparent dots represent the associations that do not pass the Bonferroni correction; the blue-colored dots and red-colored dots indicate significant associations for the R1 and R2 dimensions, respectively.

**Clinical profiles of the R1 and R2 dimensions in the general population**

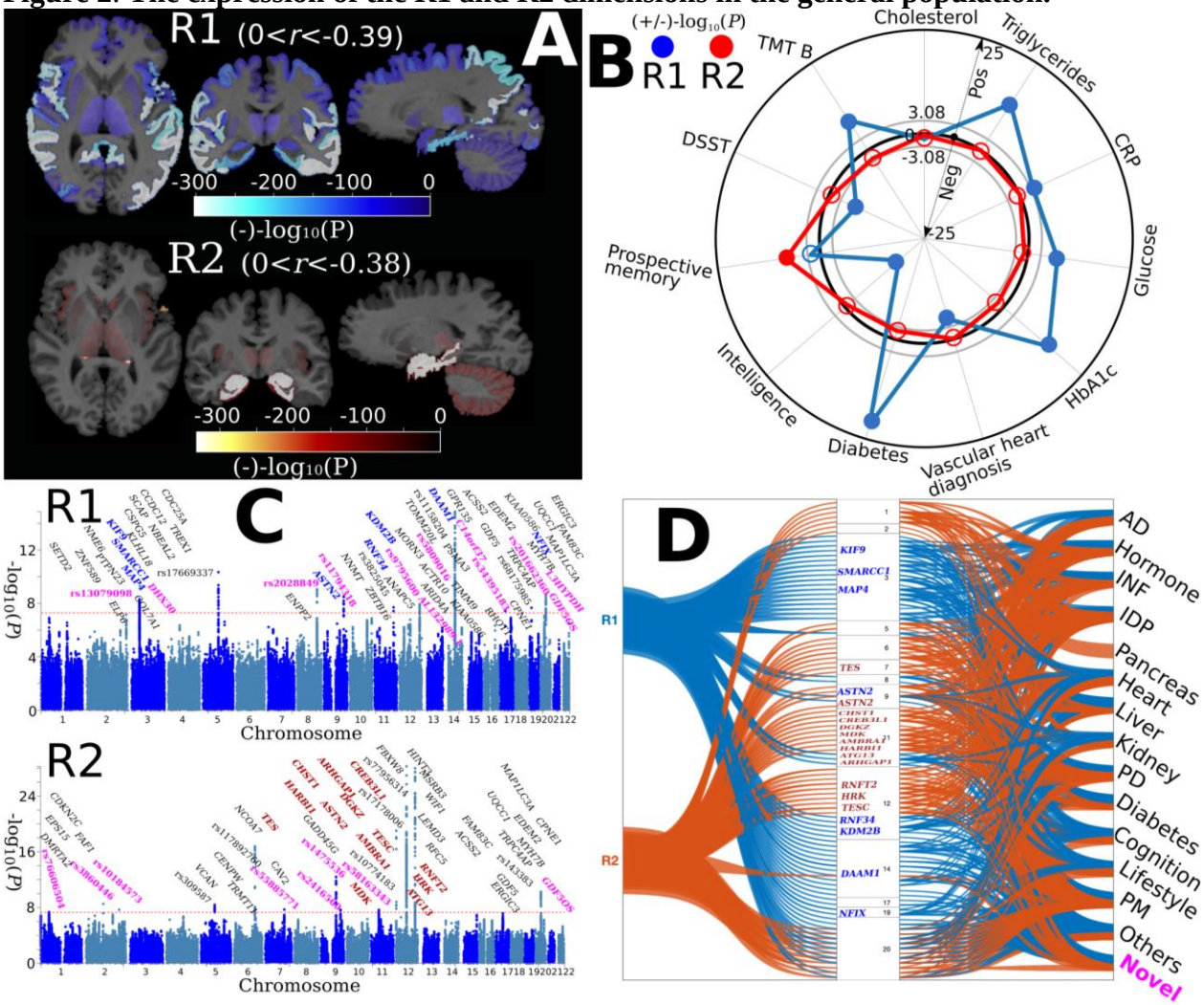
Brain association studies (**Method 4**) confirmed the presence of the two atrophy patterns in the general population (**Fig. 2A** and **Supplementary eTable 4** for P-values and effect sizes). In clinical association studies (**Method 5**), the R1 dimension was significantly associated, to a larger extent than R2, with cardiovascular (e.g., triglycerides) and diabetes factors (e.g., Hba1c and glucose), executive function (TMT-B), intelligence, physical measures (e.g., diastolic blood pressure), SPARE-BA [$-\log_{10}$(P-value) = 236.89 for R1 and -46.35 for R2] and WML [$-\log_{10}$(P-value) = 120.24 for R1 and 2.06 for R2]. In contrast, the R2 dimension was more significantly associated with SPARE-AD [$-\log_{10}$(P-value) = 136.01 for R1 and 250.41 for R2] and prospective memory. (**Fig. 2B**). Results for all 61 clinical variables, including cardiovascular factors, diabetic blood markers, social demographics, lifestyle, physical measures, cognitive scores, and imaging-derived phenotypes, are presented in **Supplementary eTable 5** for P-values and effect sizes.

**Twenty-four genomic loci and seventy-seven genes unrelated to *APOE* are associated with the R1 and R2 dimensions in the general population**

GWAS (**Method 6B** & **C**) identified 24 genomic loci, 14 of which are novel (refer to **Method 6E** for the definition of novel loci/genes), and 77 genes unrelated to *APOE* associated with R1 or

R2. In particular, the R1 dimension was significantly associated with 11 genomic loci and 49 annotated genes. Eight genes (blue-colored genes in **Fig. 2C**) were previously associated with AD-related traits; 12 novel loci/genes (pink-colored in bold genes in **Fig. 2C**) have not been previously associated with any clinical traits (**Method 6C**). The R2 dimension was significantly associated with 13 genomic loci and 40 annotated genes. 13 genes (red-colored genes in **Fig. 2C**) were associated with AD-related traits; 8 loci/genes (pink-colored in bold genes in **Fig. 2C**) were novel (**Fig. 2C**, **Supplementary eTable 6**, **Method 6C**, **D**, **E**). These genomic loci and genes were also associated with many clinical traits in the literature from the GWAS Catalog.[41] These included hormones (e.g., sex hormone-binding globulin measurement vs. *CCKN2C*), inflammatory factors (e.g., macrophage inflammatory protein 1b measurement vs. *CDC25A*), imaging-derived phenotypes (e.g., cerebellar volume measurement from MRIs vs. *DMRTA2*), and psychiatric disorders (e.g., unipolar depression vs. *ASTN2*) (**Fig. 2D**). Details of the GWAS Catalog results are presented in **Supplementary eFile 1**. The QQ plots are presented in **Supplementary eFigure 4**. The two dimensions were significantly heritable in the general population based on the SNP-based heritability estimates (R1: $h^2 = 0.49 \pm 0.02$; R2: $h^2 = 0.55 \pm 0.02$) (**Method 6B**). The polygenic risk scores of AD (**Method 6F**) showed a marginal positive association with the R2 dimension [$-\log_{10}$(P-value)=1.42], but not with the R1 dimension [$-\log_{10}$(P-value)=0.47<1.31] in this population.

**Figure 2**: **The expression of the R1 and R2 dimensions in the general population.**




**A)** Brain association studies (**Method 4**) confirm the presence of the two dimensions in the general population: the R1 dimension shows widespread brain atrophy, whereas the R2 dimension displays focal medial temporal lobe atrophy. P-value and effect sizes (*r*, Pearson's correlation coefficient) are presented in **Supplementary eTable 4**. The range of *r* for each dimension is also shown. Of note, the sample size (*N*) for R1 and R2 is the same for each ROI. **B)** Clinical association studies (**Method 5**) further show that the R2 dimension is associated with prospective memory, and the R1 dimension is associated with several cognitive dysfunctions, cardiovascular risk factors (e.g., triglycerides), and diabetes (e.g., HbA1c). The same linear regression models were used to associate the R1 and R2 dimensions with the 61 clinical variables, including cardiovascular factors, diabetic blood markers, social demographics, lifestyle, physical measures, cognitive scores, and imaging-derived phenotypes. The radar plot shows representative clinical variables; all other results are presented in **Supplementary eTable 5**. The gray circle lines indicate the P-value threshold in both directions (Bonferroni correction for the 61 variables: $-\log_{10}$(P-value) > 3.08). A positive/negative $-\log_{10}$(P-value) value indicates a positive/negative correlation (beta). Transparent dots represent the associations that do not pass the Bonferroni correction; the blue-colored dots and red-colored dots indicate significant associations for the R1 and R2 dimensions, respectively. **C)** Genome-wide association studies

(**Method 6**) demonstrate that the R2 dimension is associated to a larger extent with genomic loci and genes previously associated with AD-related traits in the literature (genome-wide P-value threshold with the red line: $-\log_{10}$(P-value) > 7.30). The R1 dimension identified 8 (blue-colored in bold) out of the 49 mapped genes associated with AD-related traits; 12 novel loci/genes (pink-colored in bold, **Method 6C**) were not associated with any clinical traits. The R2 dimension identified 13 (red-colored in bold) out of 40 mapped genes associated with AD-related traits; 8 loci/genes were novel (pink-colored in bold). Gene annotations were performed via positional, expression quantitative trait loci, and chromatin interaction mappings using FUMA.[40] The genomic loci and mapped genes were manually queried in the GWAS Catalog[41] to determine whether they were previously associated with AD (novel or not). **D)** Besides AD-related traits, the genes and genomic loci in the two dimensions were also associated with other clinical traits, including inflammation, neurohormones, and imaging-derived phenotypes, in the literature from the GWAS Catalog.[41] The flowchart first maps the genomic loci and genes (left) identified in the two dimensions onto the human genome (middle). It then links these variants to any clinical traits identified in previous literature from the GWAS Catalog (right). In the middle of the human genome, we show chromosomes 1 to 22 (above to below); the blue and red-colored genes are AD-related for the R1 and R2 dimensions, respectively. The black-colored genes (Fig. C) are not annotated. INF: inflammation; PD: psychiatric disorder; PM: physical measure; Novel (pink-colored in bold, corresponding to the novel loci/genes in Fig C) indicates that the locus or gene was not associated with any traits in the literature.

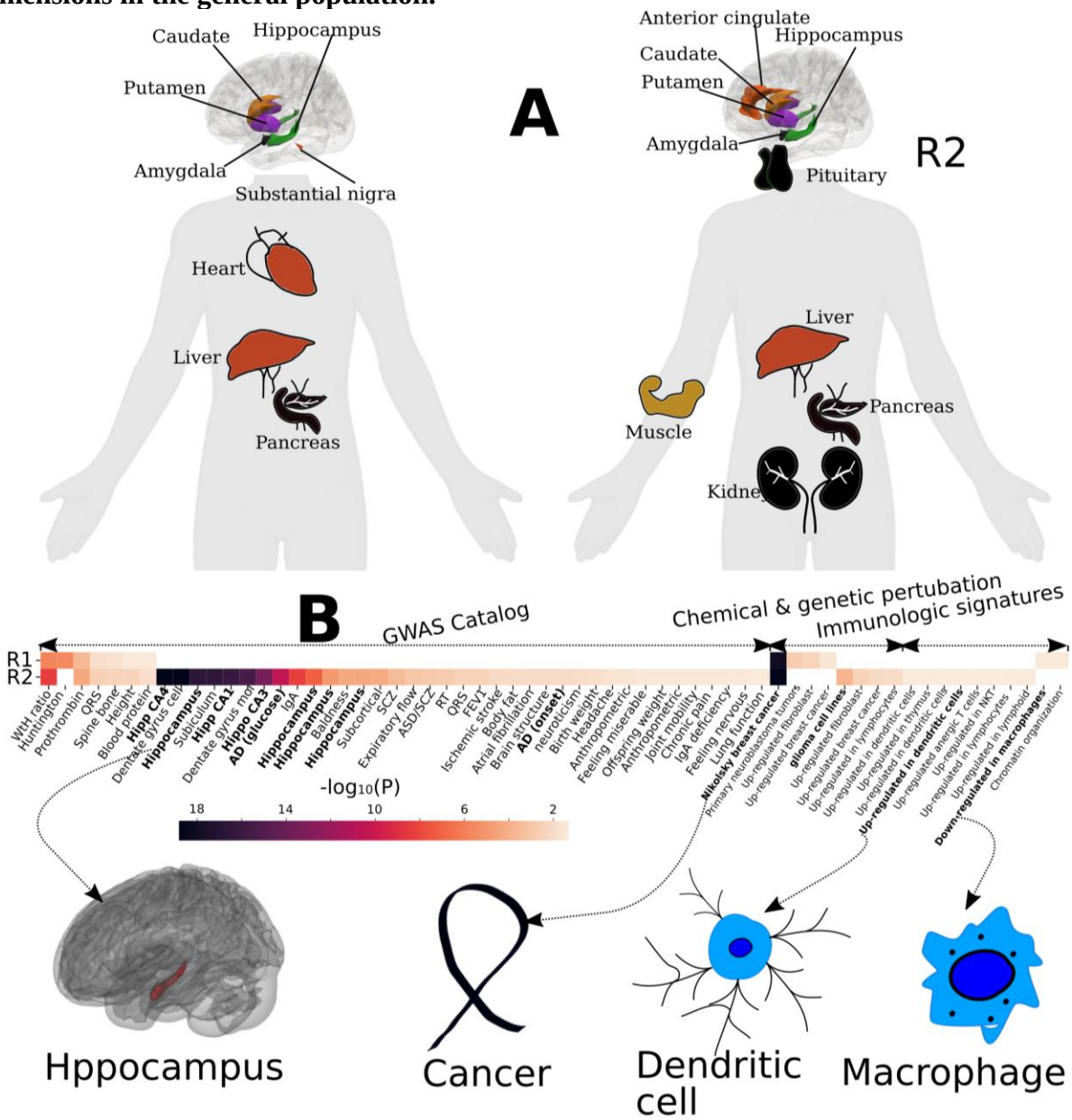
**Genes associated with the R1 and R2 dimensions are overrepresented in organs beyond the brain in the general population**

Tissue specificity analyses test whether the input genes (**Fig. 2C**) are overrepresented in differentially expressed gene sets (DEG) in one organ/tissue compared to all other organs/tissues using different gene expression data[40] (**Method 6D**). The genes associated with the R1 dimension were overrepresented in caudate, hippocampus, putamen, amygdala, substantia nigra, liver, heart, and pancreas; the genes associated with the R2 dimension were overrepresented in caudate, hippocampus, putamen, amygdala, anterior cingulate, pituitary, liver, muscle, kidney, and pancreas (**Fig. 3A** and **Supplementary eFigure 5**). Genes in DEG over-expressed in the heart were only associated with R1, while those in DEG over-expressed in the pituitary gland, muscle, and kidney were unique in R2. The expression values of every single gene for all tissues are presented in **Supplementary eFigure 6**.

**Genes associated with the R1 and R2 dimensions are enriched in key biological pathways in the general population**

Genes associated with the two dimensions were enriched in different biological pathways (**Method 6D**). Genes associated with the two dimensions were implicated in several types of cancer, including up-regulation of fibroblast, breast cancer, and neuroblastoma tumors (**Fig. 3B**), which indicate a certain extent of genetic overlaps and shared pathways that may explain the intriguing inverse relationship between AD and cancer.[42] Genes associated with the R1 dimension were implicated in pathways involved in the down-regulation of macrophages (**Fig. 3B**), which are involved in the initiation and progression of various inflammatory processes, including neuroinflammation and AD.[24] Inflammation is also known to be associated with vascular compromise and dysfunction. This further concurs with the stronger cardiovascular profile of R1, especially with increased WML and predominant SPARE-BA increases. Genes associated with the R2 dimensions were enriched in pathways involved in AD onset, hippocampus-related brain volumes, and dendritic cells (**Fig. 3B**). In particular, dendritic cells may regulate amyloid-β-specific T-cell entry into the brain,[43] as well as the inflammatory status of the brain.[44] The gene set enrichment analysis results are presented in **Supplementary eTable 7**.

**Figure 3**: **Tissue specificity and biological pathway enrichment analysis of the R1 and R2 dimensions in the general population.**




**A)** Tissue specificity analyses (**Method 6D**) show that genes associated with the two dimensions of neurodegeneration are overrepresented in organs/tissues beyond the human brain (R1 and R2). The unique overrepresentation of genes in differentially expressed gene sets (DEG) in the heart (R1) and the pituitary gland, muscle, and kidney (R2) may imply the involvement of inflammation[9–11] and neurohormone dysfunction[12,13,26], respectively. The *GENE2FUNC*[40] pipeline from FUMA was performed to examine the overrepresentation of prioritized genes (Fig. 2C) in pre-defined DEGs (up-regulated, down-regulated, and both-side DEGs) from different gene expression data. The input genes (**Fig. 2C**) were tested against each DEG using the hypergeometric test. We present only the organs/tissues that passed the Bonferroni correction for multiple comparisons. **B)** Gene set enrichment analysis (**Method 6D**) shows that genes associated with the two dimensions are enriched in different biological pathways. For example,

genes associated with the R1 dimension are implicated in down-regulated macrophage functions, which have been shown to be associated with inflammation.[24] In contrast, the R2 dimension is enriched in AD hallmarks (e.g., hippocampus), AD-related gene sets, and the pathway involved in dendritic cells, which may regulate amyloid-β-specific T-cell entry into the brain.[43] Both dimensions are enriched in gene sets involved in cancer, which may indicate overlapped genetic underpinnings between AD and cancer.[42] The *GENE2FUNC*[40] pipeline from FUMA was performed to examine the enrichment of prioritized genes (**Fig. 2C**) in pre-defined gene sets. Hypergeometric tests were performed to test whether the input genes were overrepresented in any pre-defined gene sets. Gene sets were obtained from different sources, including MsigDB[45] and GWAS Catalog.[41] We show the significant results from gene sets defined in the GWAS Catalog, curated gene sets, and immunologic signature gene sets. All results are shown in **Supplementary eTable 7.**

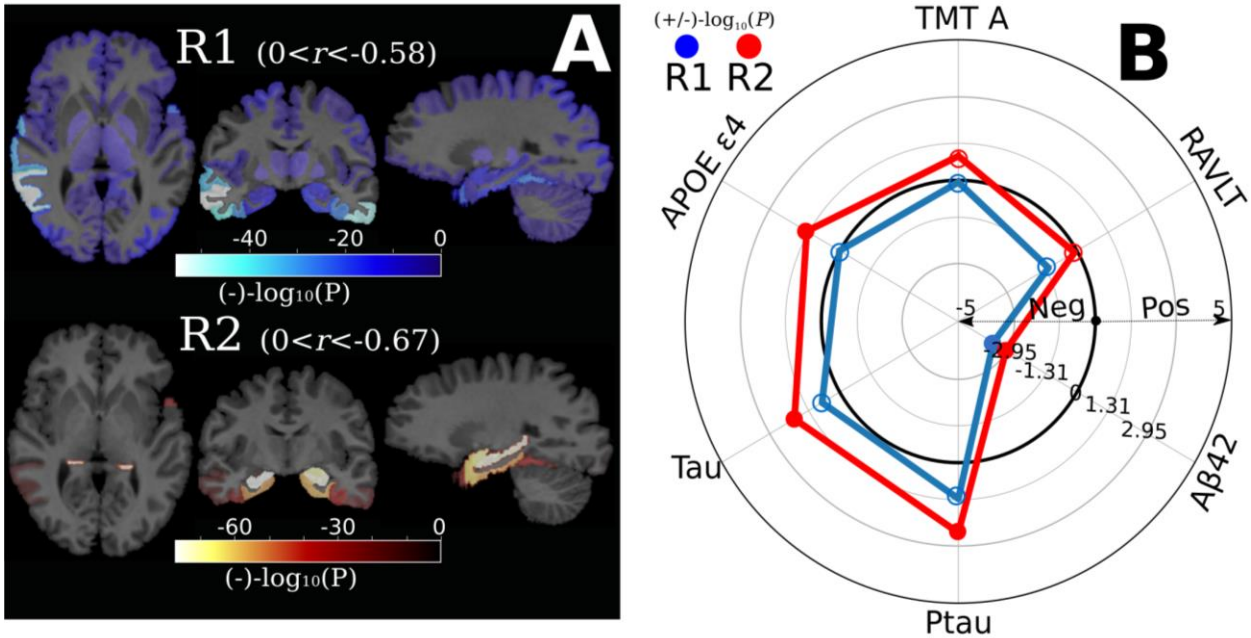
**The longitudinal rate of change in the R2 dimension, but not R1, is marginally associated with the *APOE ε4* allele, tau in cognitively unimpaired individuals**

Using cognitively unimpaired participants from ADNI and BLSA, longitudinal brain association studies (**Method 4B**) showed that the rate of change in the R1 dimension was associated with the change of brain volume in widespread brain regions. In contrast, the rate of change in the R2 dimension was associated with the change of brain volume in the focal medial temporal lobe (**Fig. 4A** and **Supplementary eTable 8** for P-values and effect sizes). This further indicates that the two dominant patterns discovered cross-sectionally also progress in consistent directions longitudinally. The two dimensions were not associated with CSF biomarkers (Aβ42, tau, and p-tau) and the *APOE ε4* allele (*rs429358*) at baseline [$-\log_{10}$(P-value) < 1.31)]. The rate of change of the R2 dimension, but not R1, was marginally [nominal threshold: $-\log_{10}$(P-value) > 1.31] associated with the *APOE ε4* allele, the CSF level of tau, and p-tau (**Fig. 4B** and **Supplementary eTable 9** for P-values and effect sizes), but they did not survive the Bonferroni correction [$-\log_{10}$(P-value) = 2.95]. The longitudinal rate of change of both dimensions was negatively associated [$-\log_{10}$(P-value) > 2.95] with the total CSF level of Aβ42.

We tested these associations using cognitively unimpaired individuals with a high risk of AD based on their family history (refer to **Method 2** for details) from the PREVENT-AD cohort. Similarly, at baseline, the two dimensions were not associated with CSF biomarkers or the *APOE ε4* allele (*rs429358*). The longitudinal rate of change in the R2 dimension, but not R1, was marginally [nominal threshold: $-\log_{10}$(P-value) > 1.31] associated with the *APOE ε4* allele [$-\log_{10}$(P-value) = 1.92], the CSF level of tau [$-\log_{10}$(P-value) = 1.65], and p-tau [$-\log_{10}$(P-value) = 1.66].

Longitudinal brain association studies also confirmed the longitudinal progression of the two dimensions in the MCI/AD population (**Supplementary eFigure 7A**). The rates of change in the two dimensions were both associated with *APOE ε4* [$-\log_{10}$(P-value) = 12.54 for R1 and 9.05 for R2] in GWAS (**Supplementary eFigure 7B**), and related to CSF levels of tau [$-\log_{10}$(P-value) = 16.47 for R1 and 9.73 for R2], p-tau [$-\log_{10}$(P-value) = 19.13 for R1 and 10.81 for R2], and Aβ42 [$-\log_{10}$(P-value) = 13.64 for R1 and 13.55 for R2] (**Supplementary eFigure 7C).**

**Figure 4**: **The longitudinal rate of change in R1 and R2 in the cognitively unimpaired population.**

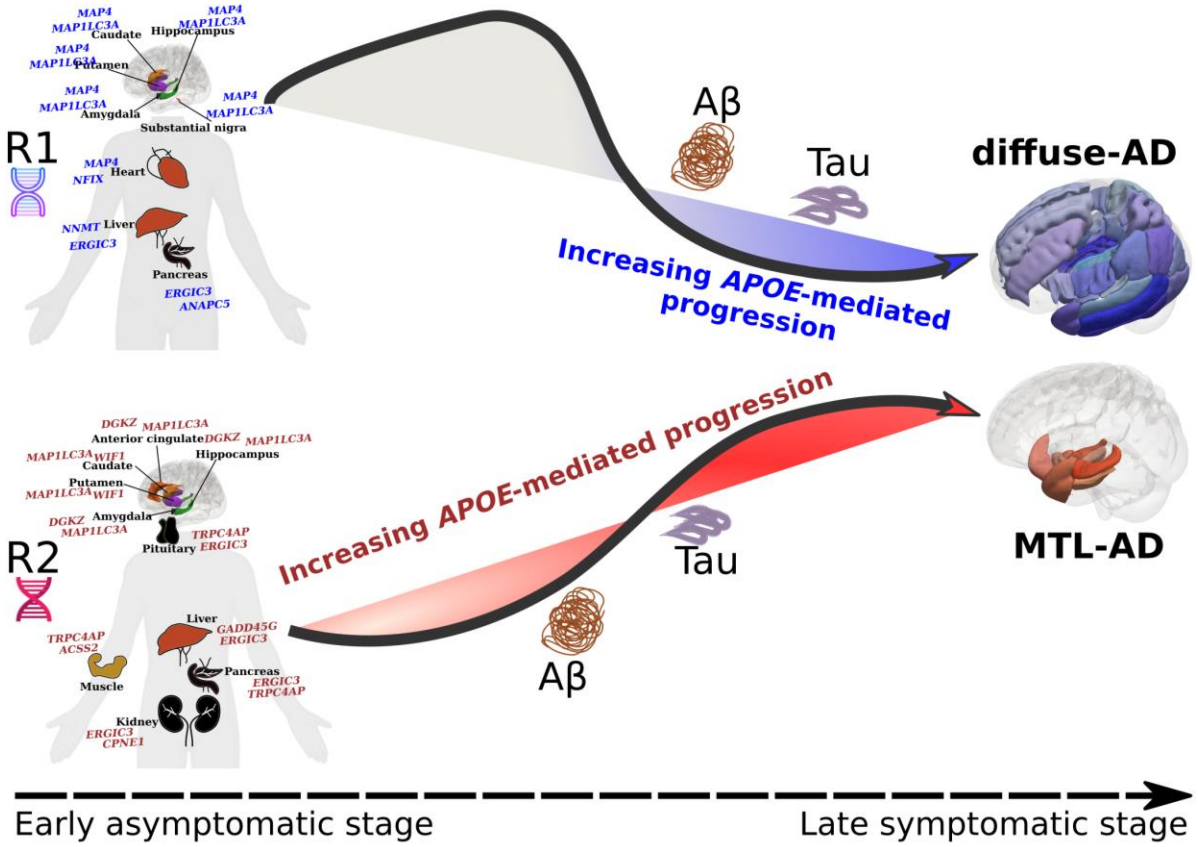



**A)** Longitudinal brain association studies (**Method 4B**) show that the R1 dimension exhibits longitudinal brain volume decrease in widespread brain regions, whereas the R2 dimension displays longitudinal brain volume decrease in the focal medial temporal lobe. We first derived the rate of change of the 119 GM ROIs and the R1 and R2 dimensions using a linear mixed effect model; a linear regression model was then fit to the rate of change of the ROIs, R1, and R2 to derive the beta coefficient value of each ROI. A negative value denotes longitudinal brain changes with a negative coefficient of the rate of change in the linear regression model. P-value and effect sizes (*r*, Pearson's correlation coefficient) are presented in **Supplementary eTable 8**. The range of *r* for each dimension is also shown. Of note, the sample size (*N*) for R1 and R2 is the same for each ROI. **B)** The rate of change, not the baseline measurement, in the two dimensions is negatively associated with the CSF level of Aβ42 (Bonferroni correction for the 45 variables: $-\log_{10}$(P-value) > 2.95). The rate of change in the R2 dimension, not the R1 dimension, was marginally ($-\log_{10}$(P-value) > 1.31) associated with the CSF level of tau and p-tau, and *APOE ε4*. All other clinical associations are presented in **Supplementary eFile 9**. The gray-colored circle lines indicate different P-value thresholds in both directions (Bonferroni correction for the 45 variables: $-\log_{10}$(P-value) > 2.95 and the nominal P-value threshold: $-\log_{10}$(P-value) > 1.31). A positive/negative $-\log_{10}$(P-value) value indicates a positive/negative correlation (beta). Transparent dots represent the associations that do not pass the nominal P-value threshold [$\log_{10}$(P-value) = 1.31]; the blue-colored dots and red-colored dots indicate significant associations [$\log_{10}$(P-value) > 1.31] for the R1 and R2 dimensions, respectively.

## Discussion

The current study leveraged a deep semi-supervised representation learning method to establish two predominant dimensions in the symptomatic MCI/AD population, which were independently found to be expressed, to a lesser degree, in three asymptomatic populations. In particular, the R1 dimension represented a "diffuse-AD" atrophy pattern: varying degrees of brain atrophy throughout the entire brain. In contrast, the R2 dimension showed an "MTL-AD" atrophy pattern: brain atrophy predominantly concentrated in the medial temporal lobe (**Fig. 1A**). Importantly, only R2 was found to be significantly associated with genetic variants of the *APOE* genes in MCI/AD patients. Furthermore, our study examined early manifestations of the R1 and R2 dimensions in asymptomatic populations with varying levels of AD risks and their associations with genetics, amyloid plaques and tau tangles, biological pathways, and body organs. We identified that 24 genomic loci, 14 of which are novel (**Method 6E**), and 77 annotated genes contribute to early manifestations of the two dimensions. Functional analyses showed that genes unrelated to *APOE* were overrepresented in DEG sets in organs beyond the brain (R1 and R2), including the heart (R1) and the pituitary gland (R2), and enriched in several biological pathways involved in dendritic cells (R2), macrophage functions (R1), and cancer (R1 and R2). Longitudinal findings in the cognitively unimpaired populations showed that the rate of change of the R2 dimension, but not R1, was marginally associated with the *APOE ε4* allele, the CSF level of tau, and Aβ42 (R1 and R2). Our findings suggested that diverse pathologic processes, including cardiovascular risk factors, neurohormone dysfunction, and inflammation, might occur in the early asymptomatic stages, supporting and expanding the current amyloid cascade.[20,21]

**Figure 5**: **Genes unrelated to *APOE* influence early manifestations of R1 and R2.**




Genes unrelated to *APOE* and overrepresented in organs beyond the human brain are associated with early manifestations of the R1 (diffuse-AD) and R2 (MTL-AD) dimensions, which capture the heterogeneity of AD-related brain atrophy. For visualization purposes, we display the two genes with the highest expression values in the tissue specificity analyses for each organ/tissue (**Supplementary eFigure 6**). Along these two dimensions, the black arrow line emulates the longitudinal progression trajectory of AD. The positions of beta-amyloid, tau and the text "increasing *APOE*-mediated progression" indicate the time point when they are associations with the two dimensions. The blue/red gradient-color background indicates a higher influence of *APOE*-related genes (left to right). The brain atrophy patterns are presented in the 3D view. In early asymptomatic stages, the R1-related genes are implicated in cardiovascular diseases and inflammation; the R2-related genes are involved in hormone-related dysfunction. Critically, longitudinal progression of the dimension demonstrates an impact of the *APOE* genes in early asymptomatic stages in R2, but this longitudinal effect occurs only in late symptomatic stages in R1. These results suggest that comorbidities (e.g., cardiovascular conditions) or normal aging in R1 may alter or delay the trajectory of neurodegeneration in early asymptomatic stages; *APOE*-related genes may play a pronounced role in the acceleration and progression in late symptomatic stages for both dimensions. Of note, the underlying pathological processes that initiate and drive the progression of the two dimensions are not mutually exclusive. Hence, both R1 and R2 can be co-expressed in the same individual. In addition, the two dimensions can also be affected by other AD hypotheses, such as the mitochondrial hypothesis[46] and the metabolic hypothesis.[47] MTL: medial temporal lobe.

AD has been regarded as a CNS disorder. However, increasing evidence has indicated that the origins or facilitators of the pathogenesis of AD might involve processes outside the brain.[19] For example, recent findings revealed that gut microbiota disturbances might influence the brain through the immune and endocrine system and the bacteria-derived metabolites.[48,49] Our findings support the view that multiple pathological processes might contribute to early AD pathogenesis and identify non-*APOE* genes in the two dimensions overrepresented in tissues beyond the brain (e.g., the heart, pituitary gland, muscle, and kidney). Pathological processes may be involved in different cells, molecular functions, and biological pathways, exaggerating amyloid plaque and tau tangle accumulation and leading to the downstream manifestation of neurodegeneration and cognitive decline.

The genetic and clinical underpinnings of the R1 dimension support inflammation, as well as cardiovascular diseases, as a core pathology contributing to AD.[9–11] Genes associated with the R1 dimension were previously associated with various inflammation-related clinical traits (**Fig. 2D**), and enriched in biological pathways involved in immunological response (e.g., up-regulation in macrophages[50], **Fig. 3B**). In addition, genes in this dimension were overrepresented in DEG sets in the heart (**Fig. 3A**). Previous literature indicated that inflammation is likely an early step that initiates the amyloidogenic pathway – the expression of inflammatory cytokines leads to the production of β-amyloid plaques.[24] Several markers of inflammation are also present in serum and CSF before any indications of Aβ or tau tangles.[51] For example, clusterin, a glycoprotein involved in many processes and conditions (e.g., inflammation, proliferation, and AD) induced by tumor necrosis factor (TNF), was present ten years earlier than Aβ deposition.[52] In addition, the R1 dimension was also strongly associated with cardiovascular and diabetes biomarkers (**Fig. 2B**). Inflammatory processes have been

critical, well-established risk factors for compromised cardiovascular function,[53] such as coronary artery disease and the breakdown of the blood-brain barrier. Our results corroborated the close relationships between AD, cardiovascular diseases, and inflammation.

The genetic and clinical underpinnings of the R2 dimension support that neuroendocrine dysfunction might be an early event contributing to the pathogenesis of AD.[12,13] Genes in the R2 dimension were previously associated with different hormone and pancreas-related traits from GWAS Catalog (**Fig. 2D**); they were also overrepresented in DEG in the pituitary and pancreas glands, muscle and kidney (**Fig. 3A**), which are master glands or key organs in the endocrine system.[54] Previous literature suggested that neuroendocrine dysfunction might contribute to AD development by secreting neurohormonal analogs and affecting CNS function.[13] For example, luteinizing hormone-releasing hormone and follicle-stimulating hormone in serum or neurons were associated with the accumulation of Aβ plaques in the brain.[12,55].[56] However, early experimental studies on antagonists of Luteinizing hormone-releasing hormone and growth hormone-releasing hormone in animal models of AD have shown promising but not entirely convincing evidence.[13] Taken together, neurodegeneration in the R2 dimension represents an AD-specific phenotype that might be driven by hormonal dysfunction, leading to rapid accumulation of amyloid plaques, and was potentially accelerated by the *APOE ε4* allele – the rate of change in R2, but not R1, was associated with the *APOE ε4* allele in cognitively unimpaired individuals (**Fig. 4B**).

The hypothesized implications above of the R1 and R2 dimensions on inflammation, cardiovascular functions, and neuroendocrine dysfunctions are not mutually exclusive and may collectively contribute to AD pathogenesis. It has been shown that dysregulation of the hypothalamic-pituitary-gonadal axis is associated with dyotic signaling, modulating the

expression of TNF and related cytokines in systemic inflammation, and the induction of downstream neurodegenerative cascades within the brain.[57,58] These studies hypothesized that the neuroendocrine dysfunction and the inflammation mechanism might be the upstream and downstream neuropathological processes along the disease course of AD.[13] That is, the loss of sex steroids and the elevation of gonadotropins might lead to a higher level of inflammatory factors in the brain. Finally, other competing hypotheses may also play a role in developing AD in early asymptomatic stages, including the mitochondrial hypothesis,[46] the metabolic hypothesis,[47] and the tau hypothesis.[16]

The NIA-AA framework[59] claims that AD is a continuum in which AD pathogenesis is initiated in early asymptomatic cognitively unimpaired stages and progresses to amyloid-positive and tau-positive (A+T+) in late symptomatic stages.[59] Our findings are consistent with this framework and elucidate the cross-sectional and longitudinal associations of the two dimensions with genetic and clinical markers from early asymptomatic to late symptomatic stages. In early asymptomatic stages, the rates of change in the two dimensions are both associated with amyloid. However, only the R2 dimension, not R1, is marginally associated with the *APOE ε4* allele and the CSF level of tau (**Fig. 4B**). In contrast, in late symptomatic stages, the rates of change in the two dimensions are both associated with the *APOE ε4* allele, CSF levels of tau, p-tau, and amyloid (**Supplementary eFigure 7**). Our findings suggest that comorbidities or normal aging in R1 may alter the rate or trajectory of neurodegeneration at early asymptomatic stages, but *APOE*-related genes might play a more pronounced role in the acceleration and progression during late symptomatic stages for both dimensions (**Fig. 5**).

In conclusion, the current study used a novel deep semi-supervised representation learning method to establish two AD dimensions. Our findings support that those diverse

pathological mechanisms, including cardiovascular diseases, inflammation, hormonal dysfunction, and involving multiple organs,[60] collectively affect AD pathogenesis in asymptomatic stages. Disentangling diverse pathological mechanisms into homogeneous dimensions may guide future therapies in the early asymptomatic stages of AD to target multi-organ dysfunctions beyond the brain.

## Methods

### Method 1: Surreal-GAN deciphers the neuroanatomical heterogeneity of Alzheimer's disease

Surreal-GAN[3] dissects underlying disease-related heterogeneity via a deep representation learning approach under the principle of semi-supervised clustering. Semi-supervised clustering[1,31] seeks the "*1-to-k*" mapping between the reference healthy control group and the patient group, thereby teasing out clusters or subtypes driven by different pathological trajectories instead of global similarity/dissimilarity in data. Refer to **Supplementary eMethod 1**'s schematic figure for the intuition of deep semi-supervised learning.

The methodological advance of Surreal-GAN is to model neuroanatomical heterogeneity by considering both spatial and temporal (i.e., disease severity) variation using only cross-sectional MRI data. Its precursor, the Smile-GAN model[1], disentangled this heterogeneity in MCI/AD patients into four subtypes: *i*) P1, preserved cognitively unimpaired-like brain, exhibits intact brain neuroanatomy compared to cognitively unimpaired participants; *ii*) P2, mild diffuse atrophy without pronounced medial temporal lobe atrophy; *iii*) P3, focal medial temporal lobe atrophy; *iv*) P4, advanced diffuse atrophy over the whole brain, and two longitudinal pathways: *i*) P1→P2→P4; *ii*) P1→P3→P4. However, Smile-GAN sought to derive categorical disease subtypes, ignoring that disease heterogeneity spatially and temporally expands along a continuum (i.e., disease severity or stages). To address this, Surreal-GAN models this heterogeneity as a continuous dimensional representation, enforces monophasic disease severity in each neuroanatomical dimension, and allows non-exclusive manifestations of multiple dimensions. As a result, Surreal-GAN dissects the heterogeneity of AD in two different dimensions: *i*) R1: diffuse brain atrophy without an exclusive focus on the medial temporal lobe;

*ii*) R2: focalized atrophy in medial temporal lobes. Each participant can be presented with different levels of expression in both two dimensions.

By modeling, Surreal-GAN's R1 and R2 dimensions are more appropriate for dimensional analyses as continuous variables. In contrast, the four neuroanatomical subtypes of Smile-GAN are better instruments for case-control analyses as categorical variables. We used ADNI data to show that the R1 and R2 dimensional scores were approximately normally distributed, and the P1, P2, P3, and P4 probability scores followed bimodal distributions (**Supplementary eFigure 8**). Refer to **Supplementary eMethod 1** for mathematical details.

**Method 2: Study populations**

The current study consists of four main populations (**Table 1**), which were jointly consolidated by the iSTAGING and the AI4AD consortia:

- the *MCI/AD population*: MCI/AD patients from ADNI[4] and BLSA[7].
- the *general population*: all participants excluding demented from UKBB[5] (refer to **Table 1** for inclusion criteria).
- the *cognitively unimpaired population*: cognitively unimpaired participants from ADNI and BLSA.
- the *cognitively unimpaired population with a family risk*: cognitively unimpaired participants from PREVENT-AD[6] (refer to **Table 1** for inclusion criteria).

The Surreal-GAN model was trained in the MCI/AD patients and healthy controls from ADNI and then applied to all populations. The iSTAGING consortium[61] consolidated all imaging and clinical data; imputed genotyping data were originally downloaded from UKBB; the

AI4AD consortium (http://ai4ad.org/) consolidated the whole-genome sequencing (WGS) data for the ADNI study.

For ADNI, cognitively unimpaired (CU), MCI, and AD participants were recruited from multiple centers in the United States. The primary goal of ADNI was to derive the two MCI/AD AI- and imaging-derived dimensions by applying the Surreal-GAN model to T1w MRIs. We included all baseline and longitudinal T1w MRI scans and cognitive data available from ADNI in iSTAGING. In addition, we also included whole-genome sequencing (WGS) data for ADNI in AI4AD. BLSA is a longitudinal study that aims to study aging and related diseases, such as AD. At baseline recruitment, participants were mostly cognitively normal and had multiple time points of longitudinal follow-ups for MRIs and cognition. Participants enrolled in PREVENT-AD were cognitively normal older adults with a family history of AD (at least one parent or multiple siblings)[a]. The inclusion criteria are generally similar but more stringent than the proxy-AD diagnosis[62,63] used in UK Biobank (see below), including *i*) being cognitively normal, *ii*) having a family history of AD, *iii*) aging within 15 years from the age of disease onset of their youngest relative, and *iv*) no history of neurological or psychiatric diseases. For UKBB, we defined asymptomatic participants[b] as those that did not have a diagnosis of all sources of dementia (G30 in ICD-10 diagnoses, see below) in our data consolidation. However, these asymptomatic participants might have diagnoses of other illnesses or comorbidities based on ICD-10: https://biobank.ctsu.ox.ac.uk/crystal/coding.cgi?id=19&nl=1. Furthermore, we defined proxy-AD[c] in UKBB as long as the participant satisfied one of the following criteria: *i*) illnesses_of_father_f20107 and *ii*) illnesses_of_mother_f20110.

**Table 1. Study characteristics.**
We present the age with the mean, min, and max in each population. The definition of cognitively unimpaired (CU)[a] in PREVENT-AD, asymptomatic participants[b] in UKBB, and proxy-AD[c] in UKBB are detailed as below:

a) Participants (proxy-AD and CU with a family risk) from the PREVENT-AD study were recruited with the following criteria: *i*) being cognitively normal, *ii*) having a family history of AD, *iii*) aging within 15 years from the age of disease onset of their youngest relative, and *iv*) no history of neurological or psychiatric diseases.
b) The UKBB participants (the general population) represent a general population with healthy aging and diseases (not AD, specifically). We excluded those diagnosed with all sources of dementia (G30 in ICD-10 diagnoses, see below). However, these asymptomatic participants might have diagnoses of other illnesses or comorbidities based on ICD-10: https://biobank.ctsu.ox.ac.uk/crystal/coding.cgi?id=19&nl=1.
c) Participants with proxy-AD in UKBB are defined by a family history of AD with the following criteria: *i*) illnesses_of_father_f20107 and *ii*) illnesses_of_mother_f20110.

| Population | Study | Participant (*N*) | Scan (*N*) | Age (year) | Sex /female | CU | AD | MCI | proxy-AD |
|---|---|---|---|---|---|---|---|---|---|
| MCI/AD | ADNI & BLSA | 1534 | 7019 | 73.45(54.27, 93.00) | 888/58% | 0 | 424 | 1110 | NA |
| General | UKBB | 39,575 | 40,981 | 64.12(44.56,82.27) | 18,625/47% | 39,574[b] | 1 | NA | 10,189[c] |
| CU | ADNI & BLSA | 1658 | 6143 | 65.75(22.00,80.00) | 939/57% | 1658 | 0 | 0 | NA |
| CU with a family risk | PREVENT-AD | 343 | 1215 | 63.63(55.13,84.22) | 243/71% | 343[a] | NA | NA | 343[a] |

**Method 3: Image processing and statistical harmonization**

**(A): Image processing**. The imaging quality check is detailed in **Supplementary eMethod 2**. All images were first corrected for magnetic field intensity inhomogeneity.[64] A deep learning-based skull stripping algorithm was applied for the removal of extra-cranial material. In total, 145 anatomical regions of interest (ROIs) were generated in gray matter (GM, 119 ROIs), white matter (WM, 20 ROIs), and ventricles (6 ROIs) using a multi-atlas label fusion method.[65] The 119 GM ROIs were used for heterogeneity analyses via Surreal-GAN and for brain association studies to examine the brain atrophy patterns of the two dimensions of Surreal-GAN. Of note, this is not to validate the two dimensions[66] since the 119 GM ROIs were used in defining the two dimensions. Instead, we showed the *post hoc* neuroanatomical patterns to elucidate these brain features that drove the two dimensions.

**(B): Statistical harmonization of MUSE ROI**. The 119 ROIs were statistically harmonized by an extensively validated approach, i.e., ComBat-GAM[67], using the entire imaging data of iSTAGING. Site-specific mean and variance were estimated based on variability observed within and across control groups while preserving normal variance due to age, sex, and intracranial volume (ICV).

**Method 4: Brain association studies**

**(A): Baseline brain association studies**: We performed brain association studies for the 119 GM ROIs. Linear regression models were fitted with R1 and R2 dimensions as dependent variables, with each ROI as the exposure of interest, controlling for age, gender, ICV, and/or diagnosis as confounders. A P-value statistic map was generated for each neuroanatomical

dimension: a positive/negative -$\log_{10}$(P-value) value indicates a positive/negative correlation (beta values) between the neuroanatomical dimension and the ROI.

**(B): Longitudinal brain association studies**: We performed two-step brain association studies for the 119 GM ROIs using longitudinal data in the MCI/AD and the cognitively unimpaired populations (≥ 4-time points). First, we estimated the RC of age using a linear mixed-effects model. Specifically, we included age as the main variable of interest while adjusting for study and ICV as fixed effects. We fit linear mixed-effects models with a participant-specific random slope for age and random intercept. The slope of age for each participant is estimated as the summation of two parts: *i*) the fixed coefficient of age and *ii*) the random coefficient of age for each participant. For the second step, the same linear regression model, as in baseline brain association studies, was fitted with the age change rate in the R1 and R2 dimensions. Bonferroni correction of 119 GM ROIs was performed to adjust for the multiple comparisons.

**Method 5: Clinical association studies**

We performed clinical association studies for all clinical biomarkers and neuropsychological testing available in the iSTAGING consolidation for each population. Linear regression models were fitted with the R1 and R2 dimensions as dependent variables, with each clinical variable as the variable of interest, controlling for age, gender, ICV, and/or diagnosis as confounders. A P-value statistic was generated for each clinical variable: a positive/negative -$\log_{10}$(P-value) value indicates a positive/negative correlation (beta values). Bonferroni correction was performed to adjust for the multiple comparisons.

**Method 6: Genetic analyses**

Genetic analyses were performed for the whole-genome sequencing (WGS) data from ADNI and the imputed genotype data from UKBB. Our quality check protocol (**Method 5A**) for ADNI resulted in 1487 participants and 24,194,338 SNPs. For UKBB, we limited our analysis to European ancestry participants, resulting in 33,541 participants and 8,469,833 SNPs.

**(A): Genetic data quality check protocol**. For ADNI WGS data, we first convert the VCF files into plink binary format. We excluded related individuals (up to 2$^{nd}$-degree) using the KING software for family relationship inference.[68] Further QC steps are: excluding criteria were: i) individuals with more than 2% of missing genotypes; ii) variants with minor allele frequency (MAF) of less than 0.1%; iii) variants with larger than 5% missing genotyping rate; iv) variants that failed the Hardy-Weinberg test at $1x10^{-5}$. We then removed duplicated variants from all 22 autosomal chromosomes. We also excluded individuals for whom either imaging or genetic data were not available. To adjust for population stratification,[69] we derived the first 40 genetic principal components (PC) using the SmartPCA software[70]. For UKBB, the genetic pipeline was previously described elsewhere.[71,72] All QC steps were documented in our BRIDGEPORT web portal: https://www.cbica.upenn.edu/bridgeport/data/pdf/BIGS_genetic_protocol.pdf.

**(B): Heritability estimates and genome-wide association analysis**. Using UKBB data, we first estimated the SNP-based heritability using GCTA-GREML,[73] controlling for confounders of age (at imaging), age-squared, sex, age-sex interaction, age-squared-sex interaction, ICV, and the first 40 genetic principal components, following a previous pioneer study.[74] In GWAS, we performed a linear regression for each neuroanatomical dimension and included the same

covariates as in the heritability estimates. We adopted the genome-wide P-value threshold (5 x $10^{-8}$) in all GWAS.

**(C): Annotation of genomic loci and gene mappings**. The annotation of genomic loci and gene mappings was performed on the online platform of FUMA (*SNP2GENE*, https://fuma.ctglab.nl/, version: v1.3.8). For annotation of genomic loci, default parameters were set in FUMA. First, lead SNPs (correlation $r^2 \leq 0.1$, distance < 250 kilobases) are assigned to a genomic locus (non-overlapping). The SNP with the lowest P-value represents each genomic locus. For gene mappings, three different strategies were used to map the SNPs to genes. First, positional mapping maps SNPs to genes if the SNPs are physically located inside a gene (a 10 kb window by default). Second, expression quantitative trait loci (eQTL) mapping maps SNPs to genes showing a significant eQTL association. Lastly, chromatin interaction mapping maps SNPs to genes when there is a significant chromatin interaction between the disease-associated regions and nearby or distant genes.[40]

**(D): Prioritized gene set enrichment and tissue specificity analysis**. FUMA provides the functionality *GENE2FUNC* to study the expression of prioritized genes and test for enrichment of the set of genes in pre-defined pathways. We used the mapped genes as prioritized gene inputs. The background genes were specified as all genes in FUMA, and default values were defined for all other parameters. *GENE2FUNC* outputs a single gene-level expression heat map that quantifies the expression values (average expression per label or average of normalized expression per label) in different tissues, including the GTEx v8[75] 54 tissue types and 30 general tissue types. For tissue specificity analysis, differentially expressed gene sets (DEG) were pre-

calculated by performing a two-sided t-test for any one label of tissue against all others. Input genes were tested against the pre-defined DEG sets using the hypergeometric test. The tissue specificity plot highlights significant enrichment at the Bonferroni corrected P-value < 0.05.

**(E): Annotation of novel genomic loci and genes related to AD**. A two-step procedure was performed to determine if a genomic locus or gene was associated with any AD-related clinical traits. First, we manually queried the identified genomic loci, mapped genes, checked if any AD-related traits were previously reported in GWAS Catalog, and downloaded these associations. In addition, we checked if any input genes overlap with the gene set pathways (defined in GWAS catalog reported genes) related to AD in (**D**). We defined a genomic locus or a gene as a novel association if the variant was not associated with any clinical traits in GWAS Catalog. For these clinical traits reported in GWAS Catalog, we mapped them into several different categories (**Fig. 2D**, **Supplementary eFile 5**).

**(F): Polygenic risk score calculation**. We calculated the PRS[76] using both ADNI and UKBB genetic data. The weights of the PRS were defined based on independent base data,[33] ensuring that the base population does not overlap with the target population in ADNI and UKBB (European ancestry). The QC steps for the base data are as follows: *i*) SNP-based heritability estimate ($h^2 > 0.05$) using LDSC[77] to avoid spurious SNP data; *ii*) the genome reference consortium human build of the base data is on GRCh37[78]; *iii*) removal of duplicated and ambiguous SNPs. The QC steps for the target data are as follows: *i*) using LiftOver[79] to convert the ADNI WGS from GRCh38 to GRCh37; *ii*) standard GWAS QC (low minor allele frequency, low genotyping rate, etc.); *iii*) pruning to remove highly correlated SNPs; *iv*) removal of high

heterozygosity samples; *v*) removal of duplicated, mismatching and ambiguous SNPs. After rigorous QC, we used PLINK to generate PRS for ADNI and UKBB by adopting the classic C+T method (clumping + thresholding: C+T). To approximate the "best-fit" PRS, we performed a logistic regression using the PRS calculated at different P-value thresholds, controlling for age, sex, and the first five genetic PCs. We chose the PRS that explains the highest phenotypic variance (AD vs. CU in ADNI) (**Supplementary eFigure 9**).

## Data Availability

The GWAS summary statistics corresponding to this study are publicly available on the BRIDGEPORT web portal (https://www.cbica.upenn.edu/bridgeport/), the FUMA online platform (https://fuma.ctglab.nl/), and the GWAS Catalog platform (https://www.ebi.ac.uk/gwas/home).

## Code Availability

The software and resources used in this study are all publicly available:

- Surreal-GAN: https://pypi.org/project/SurrealGAN/, generation of R1 and R2
- Smile-GAN: https://pypi.org/project/SmileGAN/, generation of P1, P2, P3, and P4
- BIGS: https://www.cbica.upenn.edu/bridgeport/data/pdf/BIGS_genetic_protocol.pdf, genetic processing protocol
- BRIDGEPORT: https://www.cbica.upenn.edu/bridgeport/, web portal for dissemination
- MUSE: https://www.med.upenn.edu/sbia/muse.html, image preprocessing
- PLINK: https://www.cog-genomics.org/plink/, GWAS
- FUMA: https://fuma.ctglab.nl/, genetic analysis
- GCTA: https://yanglab.westlake.edu.cn/software/gcta/#Overview, heritability estimates
- MAGMA: https://ctg.cncr.nl/software/magma, gene analysis
- LDSC: https://github.com/bulik/ldsc, PRS

## Competing Interests

DAW served as Site PI for studies by Biogen, Merck, and Eli Lilly/Avid. He has received consulting fees from GE Healthcare and Neuronix. He is on the DSMB for a trial sponsored by Functional Neuromodulation. AJS receives support from multiple NIH grants (P30 AG010133, P30 AG072976, R01 AG019771, R01 AG057739, U01 AG024904, R01 LM013463, R01 AG068193, T32 AG071444, and U01 AG068057 and U01 AG072177). He has also received support from Avid Radiopharmaceuticals, a subsidiary of Eli Lilly (in-kind contribution of PET tracer precursor); Bayer Oncology (Scientific Advisory Board); Eisai (Scientific Advisory Board); Siemens Medical Solutions USA, Inc. (Dementia Advisory Board); Springer-Nature Publishing (Editorial Office Support as Editor-in-Chief, Brain Imaging, and Behavior). ME receives support from multiple NIH grants, the Alzheimer's Association, and the Alzheimer's Therapeutic Research Institute. HJG has received travel grants and speaker honoraria from Fresenius Medical Care, Neuraxpharm, Servier, and Janssen Cilag as well as research funding from Fresenius Medical Care. HJG had personal contracts approved by the university administration for speaker honoraria and one IIT with Fresenius Medical Care. AA was supported by grants 191026 and 206795 from the Swiss National Science Foundation.

## Authors' contributions

Dr. Wen has full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis..

*Study concept and design*: Wen, Davatzikos

*Acquisition, analysis, or interpretation of data*: Wen, Davatzikos

*Drafting of the manuscript*: Wen, Davatzikos

*Critical revision of the manuscript for important intellectual content*: all authors

*Statistical and genetic analysis*: Wen

## References


1. Yang, Z. *et al.* A deep learning framework identifies dimensional representations of Alzheimer's Disease from brain structure. *Nat Commun* **12**, 7065 (2021).
2. Young, A. L. *et al.* Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. *Nat Commun* **9**, 4273 (2018).
3. Yang, Z., Wen, J. & Davatzikos, C. Surreal-GAN:Semi-Supervised Representation Learning via GAN for uncovering heterogeneous disease-related imaging patterns. *ICLR* (2021).
4. Petersen, R. C. *et al.* Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology* **74**, 201–209 (2010).
5. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
6. Breitner, J. C. S., Poirier, J., Etienne, P. E. & Leoutsakos, J. M. Rationale and Structure for a New Center for Studies on Prevention of Alzheimer's Disease (StoP-AD). *J Prev Alzheimers Dis* **3**, 236–242 (2016).
7. Resnick, S. M. *et al.* One-year age changes in MRI brain volumes in older adults. *Cereb Cortex* **10**, 464–472 (2000).
8. Barisano, G. *et al.* Blood–brain barrier link to human cognitive impairment and Alzheimer's disease. *Nat Cardiovasc Res* **1**, 108–115 (2022).
9. Heppner, F. L., Ransohoff, R. M. & Becher, B. Immune attack: the role of inflammation in Alzheimer disease. *Nat Rev Neurosci* **16**, 358–372 (2015).
10. Kinney, J. W. *et al.* Inflammation as a central mechanism in Alzheimer's disease. *Alzheimers Dement (N Y)* **4**, 575–590 (2018).

11. Wyss-Coray, T. Inflammation in Alzheimer disease: driving force, bystander or beneficial response? *Nat Med* **12**, 1005–1015 (2006).

12. Xiong, J. *et al.* FSH blockade improves cognition in mice with Alzheimer's disease. *Nature* **603**, 470–476 (2022).

13. Schally, A. V. Endocrine approaches to treatment of Alzheimer's disease and other neurological conditions: Part I: Some recollections of my association with Dr. Abba Kastin: A tale of successful collaboration. *Peptides* **72**, 154–163 (2015).

14. Guthrie, H. *et al.* Safety, Tolerability, and Pharmacokinetics of Crenezumab in Patients with Mild-to-Moderate Alzheimer's Disease Treated with Escalating Doses for up to 133 Weeks. *J Alzheimers Dis* **76**, 967–979 (2020).

15. Sevigny, J. *et al.* The antibody aducanumab reduces Aβ plaques in Alzheimer's disease. *Nature* **537**, 50–56 (2016).

16. Congdon, E. E. & Sigurdsson, E. M. Tau-targeting therapies for Alzheimer disease. *Nat Rev Neurol* **14**, 399–415 (2018).

17. Hardy, J. & Selkoe, D. J. The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics. *Science* **297**, 353–356 (2002).

18. Brinkmalm, G. & Zetterberg, H. The phosphorylation cascade hypothesis of Alzheimer's disease. *Nat Aging* **1**, 498–499 (2021).

19. Du, X., Wang, X. & Geng, M. Alzheimer's disease hypothesis and related therapies. *Translational Neurodegeneration* **7**, 2 (2018).

20. Jack, C. R. *et al.* Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol* **12**, 207–216 (2013).

21. Frisoni, G. B. *et al.* The probabilistic model of Alzheimer disease: the amyloid hypothesis revised. *Nat Rev Neurosci* **23**, 53–66 (2022).

22. Makin, S. The amyloid hypothesis on trial. *Nature* **559**, S4–S7 (2018).

23. Herrup, K. The case for rejecting the amyloid cascade hypothesis. *Nat Neurosci* **18**, 794–799 (2015).

24. Heneka, M. T. *et al.* Neuroinflammation in Alzheimer's Disease. *Lancet Neurol* **14**, 388–405 (2015).

25. Leng, F. & Edison, P. Neuroinflammation and microglial activation in Alzheimer disease: where do we go from here? *Nat Rev Neurol* **17**, 157–172 (2021).

26. Dean, D. W. Neuroendocrine Theory of Aging: Chapter 1. (2012).

27. Dubois, B. *et al.* Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimers Dement* **12**, 292–323 (2016).

28. Rajpurkar, P., Chen, E., Banerjee, O. & Topol, E. J. AI in health and medicine. *Nat Med* **28**, 31–38 (2022).

29. Wen, J. *et al.* Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation. *Medical Image Analysis* **63**, 101694 (2020).

30. Kendler, K. & Neale, M. Endophenotype: a conceptual analysis. *Mol Psychiatry* **15**, 789–797 (2010).

31. Wen, J. *et al.* Multi-scale semi-supervised clustering of brain images: Deriving disease subtypes. *Med Image Anal* **75**, 102304 (2021).

32. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet* **54**, 412–436 (2022).

33. Lambert, J.-C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **45**, 1452–1458 (2013).

34. Davatzikos, C., Xu, F., An, Y., Fan, Y. & Resnick, S. M. Longitudinal progression of Alzheimer's-like patterns of atrophy in normal older adults: the SPARE-AD index. *Brain* **132**, 2026–2035 (2009).

35. Moradi, E., Hallikainen, I., Hänninen, T. & Tohka, J. Rey's Auditory Verbal Learning Test scores can be predicted from whole brain MRI in Alzheimer's disease. *NeuroImage: Clinical* **13**, 415–427 (2017).

36. Squire, L. R., Stark, C. E. L. & Clark, R. E. The medial temporal lobe. *Annu Rev Neurosci* **27**, 279–306 (2004).

37. Bashyam, V. M. *et al.* MRI signatures of brain age and disease over the lifespan based on a deep brain network and 14 468 individuals worldwide. *Brain* **143**, 2312–2324 (2020).

38. Prins, N. D. & Scheltens, P. White matter hyperintensities, cognitive impairment and dementia: an update. *Nat Rev Neurol* **11**, 157–165 (2015).

39. Andreasen, N. *et al.* Evaluation of CSF-tau and CSF-Aβ42 as Diagnostic Markers for Alzheimer Disease in Clinical Practice. *Archives of Neurology* **58**, 373–379 (2001).

40. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).

41. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005–D1012 (2019).

42. Lanni, C., Masi, M., Racchi, M. & Govoni, S. Cancer and Alzheimer's disease inverse relationship: an age-associated diverging derailment of shared pathways. *Mol Psychiatry* **26**, 280–295 (2021).

43. Fisher, Y., Nemirovsky, A., Baron, R. & Monsonego, A. Dendritic cells regulate amyloid-β-specific T-cell entry into the brain: the role of perivascular amyloid-β. *J Alzheimers Dis* **27**, 99–111 (2011).

44. Brezovakova, V., Valachova, B., Hanes, J., Novak, M. & Jadhav, S. Dendritic Cells as an Alternate Approach for Treatment of Neurodegenerative Disorders. *Cell Mol Neurobiol* **38**, 1207–1214 (2018).

45. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).

46. Swerdlow, R. H., Burns, J. M. & Khan, S. M. The Alzheimer's disease mitochondrial cascade hypothesis: progress and perspectives. *Biochim Biophys Acta* **1842**, 1219–1231 (2014).

47. Demetrius, L. A. & Driver, J. Alzheimer's as a metabolic disease. *Biogerontology* **14**, 641–649 (2013).

48. Jiang, C., Li, G., Huang, P., Liu, Z. & Zhao, B. The Gut Microbiota and Alzheimer's Disease. *J Alzheimers Dis* **58**, 1–15 (2017).

49. Seo, D., Boros, B. D. & Holtzman, D. M. The microbiome: A target for Alzheimer disease? *Cell Res* **29**, 779–780 (2019).

50. Fujiwara, N. & Kobayashi, K. Macrophages in inflammation. *Curr Drug Targets Inflamm Allergy* **4**, 281–286 (2005).

51. Laurin, D., David Curb, J., Masaki, K. H., White, L. R. & Launer, L. J. Midlife C-reactive protein and risk of cognitive decline: a 31-year follow-up. *Neurobiol Aging* **30**, 1724–1727 (2009).

52. Thambisetty, M. *et al.* Association of Plasma Clusterin Concentration With Severity, Pathology, and Progression in Alzheimer Disease. *Archives of General Psychiatry* **67**, 739–748 (2010).

53. Ruparelia, N., Chai, J. T., Fisher, E. A. & Choudhury, R. P. Inflammatory processes in cardiovascular disease: a route to targeted therapies. *Nat Rev Cardiol* **14**, 133–144 (2017).

54. Vitale, G., Salvioli, S. & Franceschi, C. Oxidative stress and the ageing endocrine system. *Nat Rev Endocrinol* **9**, 228–240 (2013).

55. Strittmatter, W. J. Old Drug, New Hope for Alzheimer's Disease. *Science* **335**, 1447–1448 (2012).

56. Strittmatter, W. J. Medicine. Old drug, new hope for Alzheimer's disease. *Science* **335**, 1447–1448 (2012).

57. Clark, I. A. & Atwood, C. S. Is TNF a Link between Aging-Related Reproductive Endocrine Dyscrasia and Alzheimer's Disease? *JAD* **27**, 691–699 (2011).

58. Clark, I. A., Alleva, L. M. & Vissel, B. The roles of TNF in brain dysfunction and disease. *Pharmacol Ther* **128**, 519–548 (2010).

59. Jack, C. R. *et al.* NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers Dement* **14**, 535–562 (2018).

60. Tian, Y. E. *et al.* Biological aging of human body and brain systems. 2022.09.03.22279337 Preprint at https://doi.org/10.1101/2022.09.03.22279337 (2022).

61. Habes, M. *et al.* The Brain Chart of Aging: Machine-learning analytics reveals links between brain aging, white matter disease, amyloid burden, and cognition in the iSTAGING consortium of 10,216 harmonized MR scans. *Alzheimer's & Dementia* **17**, 89–102 (2021).

62. Liu, J. Z., Erlich, Y. & Pickrell, J. K. Case-control association mapping by proxy using family history of disease. *Nat Genet* **49**, 325–331 (2017).

63. Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet* **51**, 404–413 (2019).

64. Tustison, N. J. *et al.* N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* **29**, 1310–1320 (2010).

65. Doshi, J. *et al.* MUSE: MUlti-atlas region Segmentation utilizing Ensembles of registration algorithms and parameters, and locally optimal atlas selection. *Neuroimage* **127**, 186–195 (2016).

66. Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F. & Baker, C. I. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat. Neurosci.* **12**, 535–540 (2009).

67. Pomponio, R. *et al.* Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. *Neuroimage* **208**, 116450 (2020).

68. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).

69. Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nat Rev Genet* **11**, 459–463 (2010).

70. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904–909 (2006).

71. Wen, J. *et al.* Characterizing Heterogeneity in Neuroimaging, Cognition, Clinical Symptoms, and Genetics Among Patients With Late-Life Depression. *JAMA Psychiatry* (2022) doi:10.1001/jamapsychiatry.2022.0020.

72. Wen, J. *et al.* Novel genomic loci and pathways influence patterns of structural covariance in the human brain. 2022.07.20.22277727 Preprint at https://doi.org/10.1101/2022.07.20.22277727 (2022).

73. Yang, J., Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. GCTA-GREML accounts for linkage disequilibrium when estimating genetic variance from genome-wide SNPs. *PNAS* **113**, E4579–E4580 (2016).

74. Zhao, B. *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* **51**, 1637–1644 (2019).

75. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580–585 (2013).

76. Choi, S. W., Mak, T. S.-H. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc* **15**, 2759–2772 (2020).

77. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291–295 (2015).

78. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).

79. Karolchik, D. *et al.* The UCSC Genome Browser Database. *Nucleic Acids Res* **31**, 51–54 (2003).

## Acknowledgments

The iSTAGING consortium is a multi-institutional effort funded by NIA by RF1 AG054409. The Baltimore Longitudinal Study of Aging neuroimaging study is funded by the Intramural Research Program, National Institute on Aging, National Institutes of Health and by HHSN271201600059C. Other supporting grants include 5U01AG068057-02, 1U24AG074855-01. S.R.H is Supported by multiple grants and contracts from NIH. This research has been conducted using the UK Biobank Resource under Application Number 35148. Data used in the preparation of this article were in part obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (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 the analysis or writing of this report. A complete listing of ADNI investigators can be found

at: http://adni.loni.usc.edu/wpcontent/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. 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). 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. Dr. Wen had full access to all the data in the study. He took responsibility for the integrity of the data and the accuracy of the data analysis.