Genomic loci influence patterns of structural covariance in the human brain
Supporting Information

Junhao Wen[1,2*], Ilya M. Nasrallah[2,3], Ahmed Abdulkadir[2], Theodore D. Satterthwaite[2,4], Zhijian Yang[2], Guray Erus[2], Timothy Robert-Fitzgerald[5], Ashish Singh[2], Aristeidis Sotiras[6], Aleix Boquet-Pujadas[7], Elizabeth Mamourian[2], Jimit Doshi[2], Yuhan Cui[2], Dhivya Srinivasan[2], Ioanna Skampardoni[2], Jiong Chen[2], Gyujoon Hwang[2], Mark Bergman[2], Jingxuan Bao[8], Yogasudha Veturi[9], Zhen Zhou[2], Shu Yang[8], Paola Dazzan[10], Rene S. Kahn[11], Hugo G. Schnack[12], Marcus V. Zanetti[13], Eva Meisenzahl[14], Geraldo F. Busatto[13], Benedicto Crespo-Facorro[15], Christos Pantelis[16], Stephen J. Wood[17], Chuanjun Zhuo[18], Russell T. Shinohara[2,5], Ruben C. Gur[4], Raquel E. Gur[4], Nikolaos Koutsouleris[19], Daniel H. Wolf[2,4], Andrew J. Saykin[20], Marylyn D. Ritchie[9], Li Shen[8], Paul M. Thompson[21], Olivier Colliot[22], Katharina Wittfeld[23], Hans J. Grabe[23], Duygu Tosun[24], Murat Bilgel[25], Yang An[25], Daniel S. Marcus[26], Pamela LaMontagne[26], Susan R. Heckbert[27], Thomas R. Austin[27], Lenore J. Launer[28], Mark Espeland[29], Colin L Masters[30], Paul Maruff[30], Jurgen Fripp[31], Sterling C. Johnson[32], John C. Morris[33], Marilyn S. Albert[34], R. Nick Bryan[3], Susan M. Resnick[25], Yong Fan[2], Mohamad Habes[35], David Wolk[2,36], Haochang Shou[2,5], and Christos Davatzikos[2*]

[1]Laboratory of AI and Biomedical Science (LABS), Stevens Neuroimaging and Informatics Institute, Keck School of Medicine of USC, University of Southern California, Los Angeles, California, USA.
[2]Artificial Intelligence in Biomedical Imaging Laboratory (AIBIL), Center for Biomedical Image Computing and Analytics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, USA.
[3]Department of Radiology, University of Pennsylvania, Philadelphia, USA.
[4]Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, USA
[5]Penn Statistics in Imaging and Visualization Center, Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, USA
[6]Department of Radiology and Institute for Informatics, Washington University School of Medicine, St. Louis, USA
[7]Biomedical Imaging Group, EPFL, Lausanne, Switzerland
[8]Department of Biostatistics, Epidemiology and Informatics University of Pennsylvania Perelman School of Medicine, Philadelphia, USA
[9]Department of Genetics and Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[10]Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
[11]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, USA
[12]Department of Psychiatry, University Medical Center Utrecht, Utrecht, Netherlands
[13]Institute of Psychiatry, Faculty of Medicine, University of São Paulo, São Paulo, Brazil
[14] Department of Psychiatry and Psychotherapy, HHU Düsseldorf, Germany
[15]Hospital Universitario Virgen del Rocio, University of Sevilla-IBIS; IDIVAL-CIBERSAM, Sevilla, Spain
[16]Melbourne Neuropsychiatry Centre, Department of Psychiatry, University of Melbourne and Melbourne Health, Carlton South, Australia
[17]Orygen and the Centre for Youth Mental Health, University of Melbourne; and the School of Psychology, University of Birmingham, UK
[18]key Laboratory of Real Tine Tracing of Brain Circuits in Psychiatry and Neurology (RTBCPN-Lab), Nankai University Affiliated Tianjin Fourth Center Hospital; Department of Psychiatry, Tianjin Medical University, Tianjin, China
[19]Department of Psychiatry and Psychotherapy, Ludwig-Maximilian University, Munich, Germany
[20]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
[21]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
[22]Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013, Paris, France

[23]Department of Psychiatry and Psychotherapy, German Center for Neurodegenerative Diseases (DZNE), University Medicine Greifswald, Germany
[24]Department of Radiology and Biomedical Imaging, University of California, San Francisco, CA, USA
[25]Laboratory of Behavioral Neuroscience, National Institute on Aging, NIH, USA
[26]Department of Radiology, Washington University School of Medicine, St. Louis, Missouri, USA
[27]Cardiovascular Health Research Unit and Department of Epidemiology, University of Washington, Seattle, WA, USA
[28]Neuroepidemiology Section, Intramural Research Program, National Institute on Aging, Bethesda, Maryland, USA
[29]Sticht Center for Healthy Aging and Alzheimer's Prevention, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA
[30]Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Parkville, VIC, Australia
[31]CSIRO Health and Biosecurity, Australian e-Health Research Centre CSIRO, Brisbane, Queensland, Australia
[32]Wisconsin Alzheimer's Institute, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin, USA
[33]Knight Alzheimer Disease Research Center, Washington University in St. Louis, St. Louis, MO, USA
[34]Department of Neurology, Johns Hopkins University School of Medicine, USA
[35]Glenn Biggs Institute for Alzheimer's & Neurodegenerative Diseases, University of Texas Health Science Center at San Antonio, San Antonio, USA
[36]Department of Neurology and Penn Memory Center, University of Pennsylvania, Philadelphia, USA

[*]Corresponding authors:
Junhao Wen, Ph.D. – junhaowe@usc.edu
2025 Zonal Ave, Los Angeles, CA 90033, United States
Christos Davatzikos, Ph.D. – Christos.Davatzikos@pennmedicine.upenn.edu
3700 Hamilton Walk, 7th Floor, Philadelphia, PA 19104, United States

eText 1: Sensitivity check analysis for the GWAS using PSC C32_1

We used the GWAS results (233 significant SNPs in 5 genomic loci) of the first PSC in C32 (C32_1) from the UKBB discovery set to demonstrate this.

We replicated all the 233 significant SNPs in 5 genomic loci both at the nominal level ($-\log_{10}$[p-value] > 1.31), and the Bonferroni corrected p-value threshold ($-\log_{10}$[p-value] > 3.67) using the combined discovery and replication sets (*N*=33,541) (**SI eFigure 4b**), the 20,438 participants with all ancestries in the discovery set (**SI eFigure 4c**), and the 16,743 participants in the discovery set with four additional imaging-related covariates (3 parameters for the brain position in the lateral, longitudinal, and transverse directions, and 1 parameter for the head motion from fMRI) (**SI eFigure 4d**). While replicating the results in 2386 participants with non-European ancestries, we only replicated 41 SNPs (17.6%), passing the nominal significant threshold (**SI eFigure 4e**). Finally, only 14 SNPs (6.4%) were replicated when replicating the results using 1481 whole-genome sequencing (WGS) data from ADNI consolidated by the AI4AD consortium[16] (**SI eFigure 4f**). The low replication rates in other ancestries and independent disease-specific populations are expected due to population stratification, disease-specific effects, and reduced sample sizes. This further emphasizes the urge to enrich and diversify genetic research with non-European ancestries and disease-specific populations.

eText 2: Institutional Review Board (IRB) statement

All individual studies were approved by their local corresponding Institutional Review Boards (IRB). The iSTAGING and PHENOM consortia consolidated all individual imaging and clinical data; imputed genotype data were directly downloaded from the UKBB website. Data from the UKBB for this project pertains to application 35148. For iSTAGING, the IRB at the University of Pennsylvania (protocol number: 825722) reviewed the research proposal on August 31st, 2016, and updated it on August 31st, 2022. No human subjects were recruited or scanned. Existing de-identified data will be used in this mega-analysis study pooling data from 17 studies: BLSA, ADNI1, ADNI2, ADNI3, ACCORD-MIND, LookAhead, SPRINT, CARDIA, MESA, SHIP, BIOCARD, WRAP, Penn-ADC, WHIMS-MRI, AIBL, OASIS, UKBB, MESA, HANDLS. For PHENOM, the IRB at the University of Pennsylvania (protocol number: 828077) reviewed the research proposal on August 19th, 2017. No human subjects were recruited or scanned. Existing de-identified data will be used in this meta-analysis study pooling data from 10 studies at Penn, Ludwick-Maximmilian University of Munich, Kings College-London, University of Utrecht, University of Melbourne, University of Cantabria, University of Sao Paolo, Xijing Hospital Shaanxi, Tianjin Anning Hospital, and Institute of Mental Health Peking University.

eText 3: The four datasets and populations defined in this study

We defined four populations or data sets per analysis across the paper: *i*) *discovery set*, *ii*) *replication set*, *iii*) *training population*, and *iv*) *comparison population* (refer to **SI eText 2** for details).

- Discovery set*:* It consists of a multi-disease and lifespan population that includes participants from all 12 studies (*N*=32,440). Note that this population does not contain the entire UKBB population but only our first download (July 2017, *N*=21,305).
- Replication set: We held 18,259 participants from the UKBB dataset to replicate the GWAS results. We took these data from our second download of the UKBB dataset (November 2021, *N*=18,259).
- Training population: We randomly drew 250 patients (PT), including AD, MCI, SCZ, ASD, MDD, HTN (hypertension), DM (diabetes mellitus), and 250 healthy controls (CN) per decade from the discovery set, ensuring that the PT and CN groups have similar sex, study and age distributions. The resulting set of 4000 imaging data was used to generate the MuSIC atlas with the sopNMF algorithm. The rationale is to maximize variability across a balanced sample of multiple diseases or risk conditions, age, and study protocols rather than overfit the entire data by including all images in training.
- Comparison population: To validate sopNMF compared to the original opNMF algorithm, we randomly subsampled 800 participants from the training population (100 per decade for balanced CN and PT). For this scale of sample size, opNMF can load all images into memory for batch learning.[1]

eMethod 1: Empirical validation of sopNMF.

For the empirical validation of sopNMF, the comparison population (**Method 1** in the main manuscript) was used so that the machine's memory could be sufficient to read the entire data for opNMF. For sopNMF, different choices of batch size (i.e., BS=32, 64, 128, and 256) were tested. We hypothesized that sopNMF could approximate the optima of opNMF during optimization, i.e., resulting in similar parts-based representation, training loss, and sparsity. TensorboardX was embedded into the sopNMF framework to monitor the training process dynamically. All experiments were performed on an Ubuntu machine with a maximum RAM of 32 GB and 8 CPUs. The predefined maximum number of epochs for all experiments is 50,000, and the tolerance of early stopping criteria is 100 epochs based on the training loss.

We qualitatively compared the extracted PSCs and quantitatively for the training loss, the sparsity of the component matrix $\boldsymbol{W}$, and the memory consumption for C=20 (number of PSCs). The 20 PSCs were spatially consistent between opNMF and sopNMF, despite that some regions were decomposed into different PSCs (i.e., the white ellipse in **eFig. 1A**). For the training loss, opNMF obtained the lowest loss (1.103×10^6), and the loss of sopNMF were 1.107×10^6, 1.108×10^6, 1.111×10^6 and 1.210×10^6 for BS =256, 128, 64, and 32, respectively (**eFig. 1D**). For the sparsity of the component matrix, all models obtained comparable results (sparsity ≈ 0.83, **eFig. 1E**). The estimated memory consumptions during the training process were 28.65, 4.02, 3.81, 2.60, 1.47 GB for opNMF and sopNMF (BS =256, 128, 64, and 32), respectively (**Fig. e1F**).

eMethod 2: Reproducibility index.

We proposed a reproducibility index (RI) to test the reproducibility of sopNMF for brain parcellation:

- We used the Hungarian match algorithm[2] to match the pairs of PSCs between two splits under the specific condition that maximizes the similarity (i.e., minimizes the cost of workers/jobs in its original formulation).
- For each pair of PSCs, we calculated the inner product of the vectors (R^d), referred to as RI. This index takes values between [0, 1], with higher values indicating higher reproducibility.
- For each scale *C*, we presented the mean/standard deviation of the RIs for all PSCs.

eMethod 3: Inter-site image harmonization

We used an extensively validated statistical harmonization approach, i.e., ComBat-GAM,[3] to harmonize the extracted multi-scale PSCs. This method estimates the variability in volumetric measures due to differences in site/cohort-specific imaging protocols based on variances observed within and across control groups while preserving normal variances due to age, sex, and intracranial volume (ICV) differences. The model was initially trained on the discovery set and then applied to the replication set.

eMethod 4: Quality check of the image processing pipeline.

Raw T1-weighted MRIs were first quality checked (QC) for motion, image artifacts, or restricted field-of-view. Another QC was performed: First, the images were examined by manually evaluating for pipeline failures (e.g., poor brain extraction, tissue segmentation, and registration errors). Furthermore, a second step automatically flagged images based on outlying values of quantified metrics (i.e., PSC values); those flagged images were re-evaluated.

eMethod 5: Definition of the index, candidate, independent significant, and lead SNP and genomic locus.

Index SNP

They are defined as SNPs with a p-value threshold ≤ 5e-8 (*clump-p1*) from GWAS summary statistics.

Independent significant SNP

They are defined as the index SNPs, which are independent of each other (not in linkage disequilibrium) with $r^2 \leq 0.6$ (*clump-r2*) within 250 kilobases (non-overlapping, *clump-kb*) away from each other.

Lead SNP and *genomic loci*

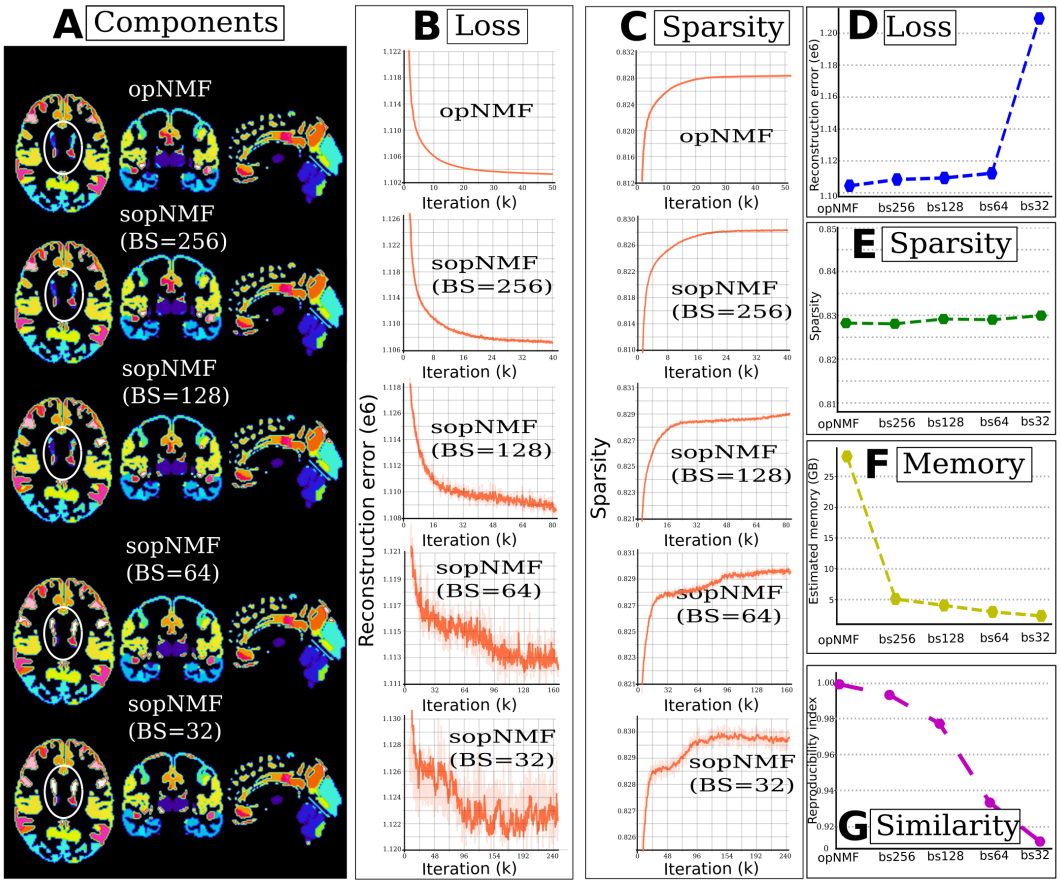
They are defined as the independent significant SNPs, which are independent of each other with a more stringent $r^2 \leq 0.1$ (*clump-r2*) within 250 kilobases (non-overlapping, *clump-kb*) away from each other. Each of these clumps is defined as a *genomic locus*.

Candidate SNP

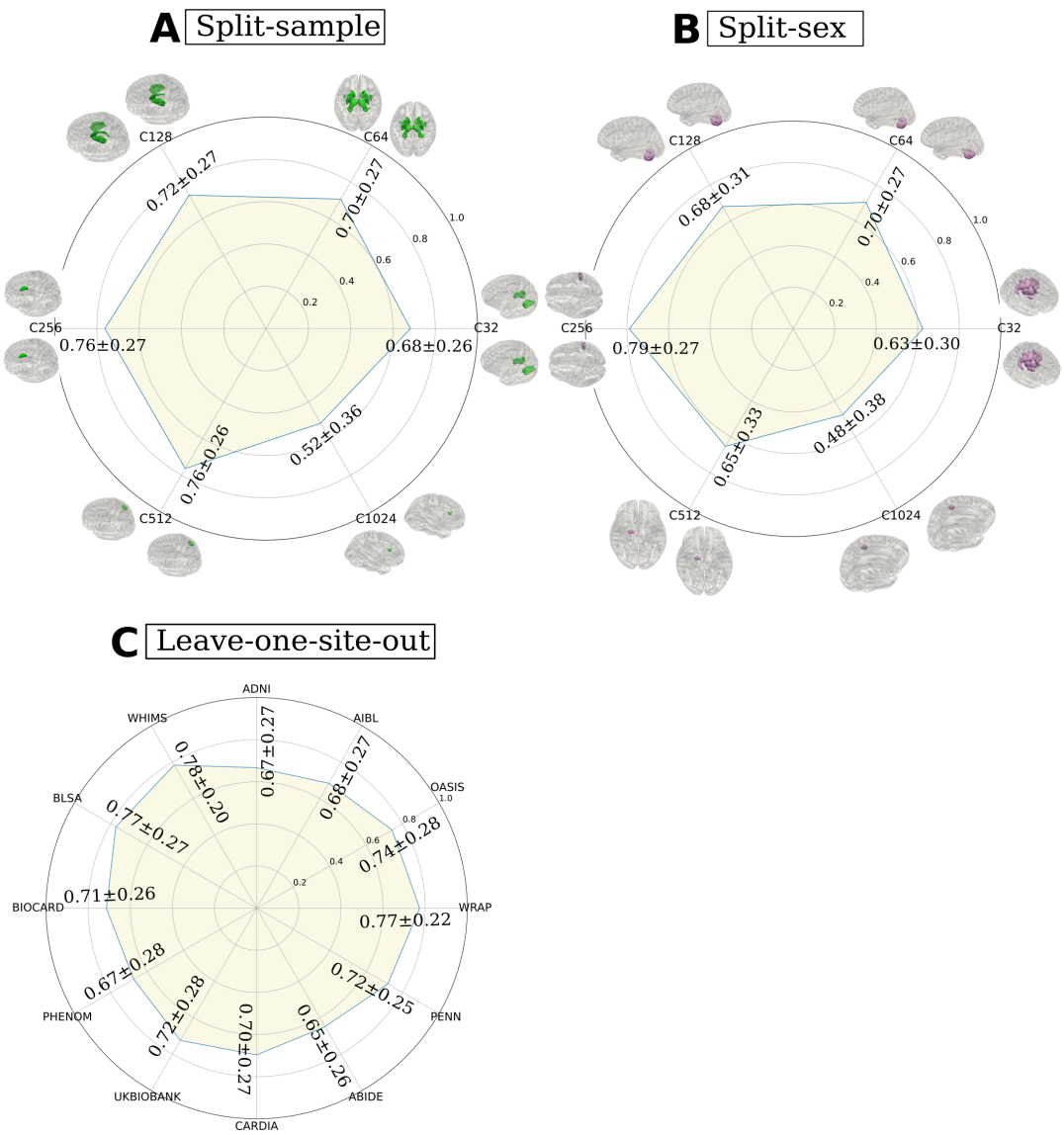
With each genomic locus, candidate SNPs are defined as the SNPs whose association p-values are smaller than 0.05 (*clump-p2*). The definitions followed instructions from FUMA[4] and Plink[5] software.

eMethod 6: Cross-validation procedure for PAML.

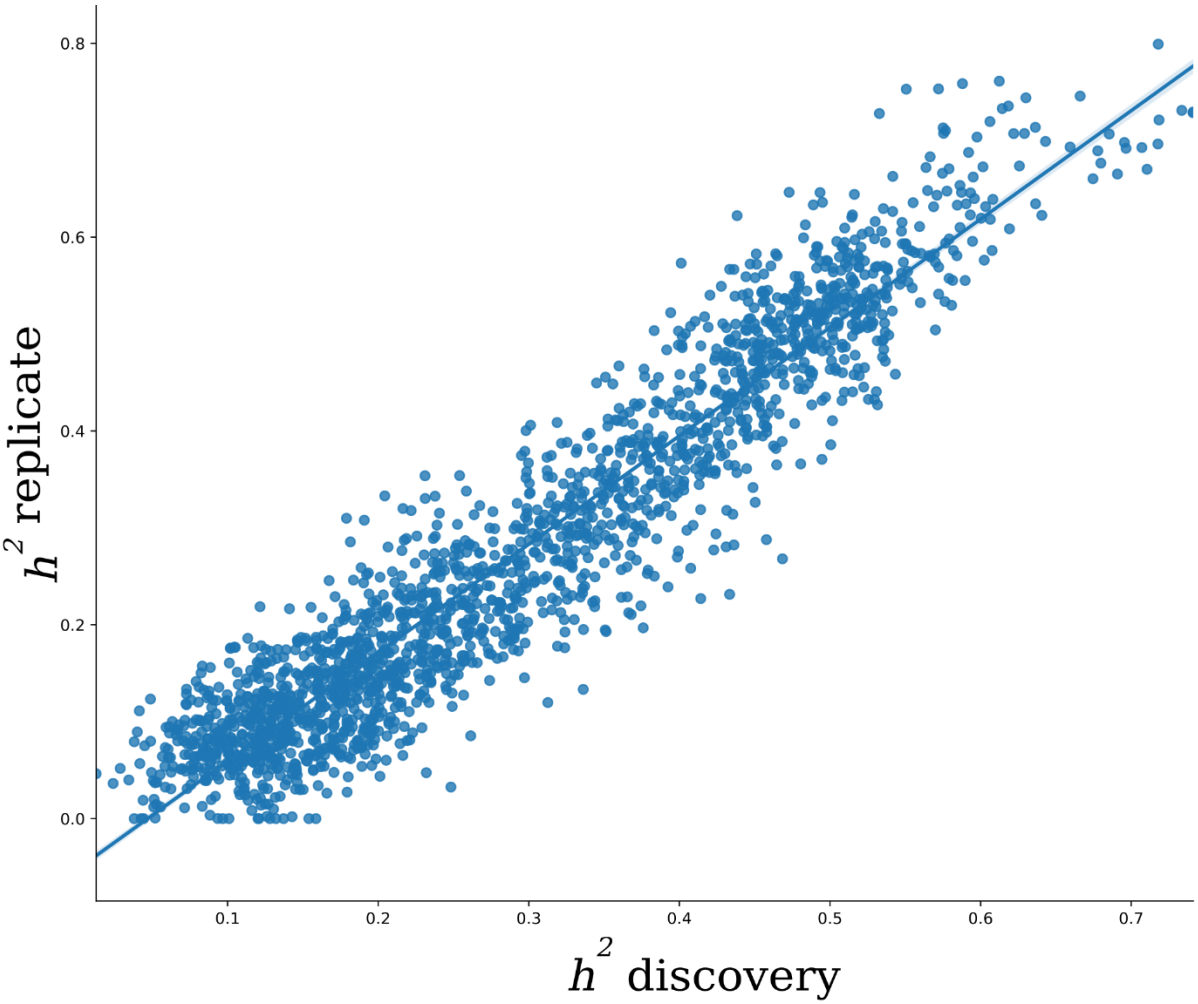
Nested cross-validation was adopted for all tasks following the good-practice guidelines proposed in our previous works[6–8]. In particular, an outer loop was used to evaluate the task performance (250 repetitions of random hold-out splits with 80% of data for training). In contrast, an inner loop focused on tuning the hyperparameters (10-fold splits). We computed the balanced accuracy (BA) to evaluate the classification tasks. We calculated the effect size (Cohen's *d*) and p-value for each SPARE index to quantify its discriminative power.



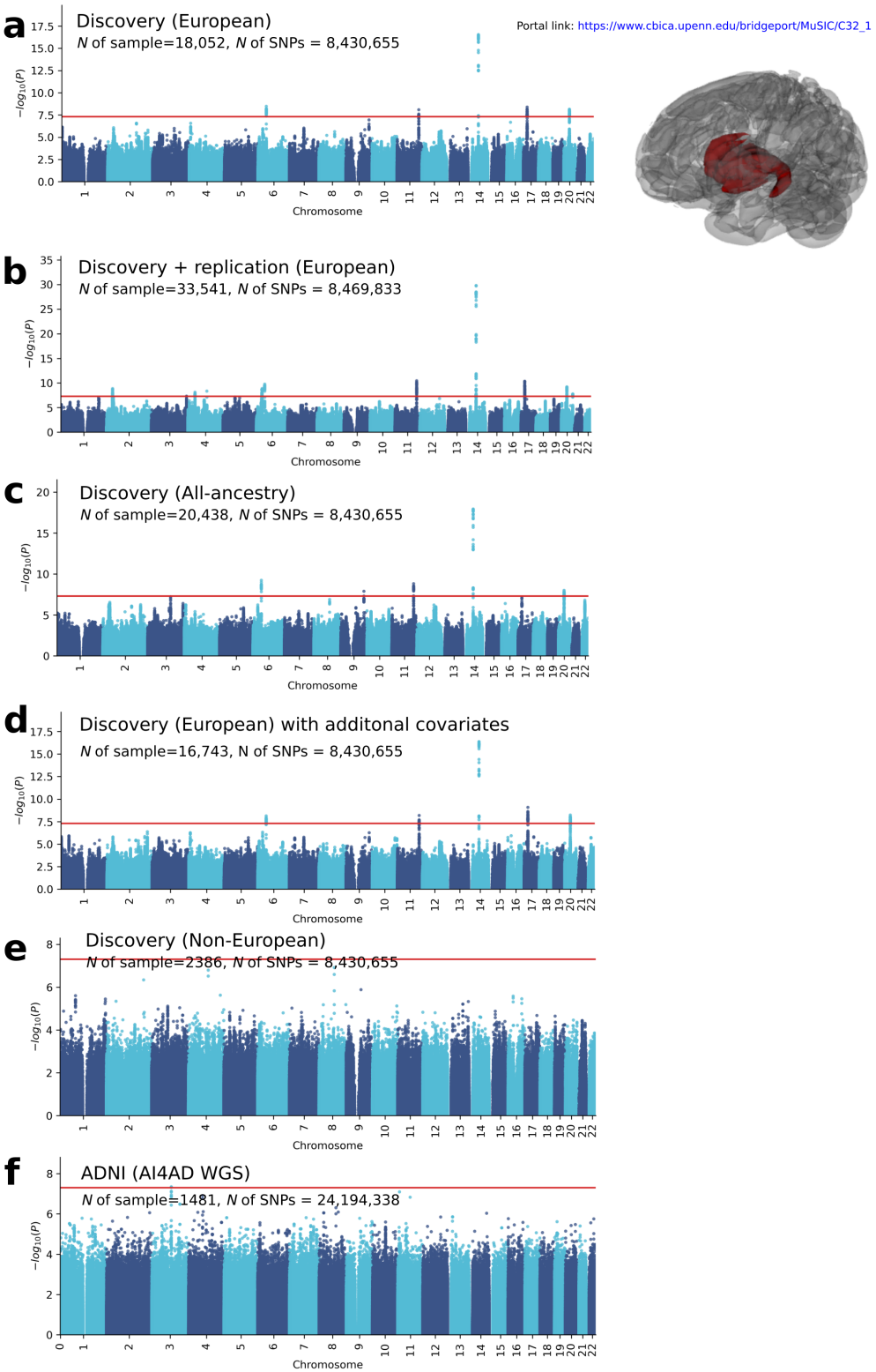
eFigure 1: Comparison between opNMF and sopNMF. (**A**) Qualitative evaluation: The extracted components are shown in the original image space, with each PSC displayed in a distinct color. The white ellipse indicates the region where the models diverge. Quantitative evaluation: training loss (**B**, **D**) and sparsity (**C**, **E**) demonstrated similar patterns between models, except that batch size (BS) = 32 had a larger loss than the other models. Comparing the estimated memory consumption during training across models shows significant advantages for all sopNMF models compared to opNMF.



eFigure 2: Reproducibility of the sopNMF brain parcellation. In general, sopNMF demonstrated high reproducibility under various conditions. For each brain PSC, the reproducibility index (RI) was calculated (**Supplementary eMethod 2**). (**A**) Split-sample analyses, where the training population (*N*=4000) was randomly split into two halves while maintaining similar age, sex, and site distribution between groups. (**B**) Split-sex analyses, where the training population was divided into males and females. Colored PSCs on the brain template illustrate the same PSC independently derived from the two splits. (**C**) Leave-one-site-out analyses for C32 PSCs., where the training populations excluding participants from each site (BIOCARD, ADNI, WARP, AIBL, ABIDE, BLSA, OASIS, CARDIA, PHENOM, PENN, UKBB, and WHIMS) were independently trained with sopNMF. The RI indices were compared to the sopNMF results using the full training sample (N=4000).

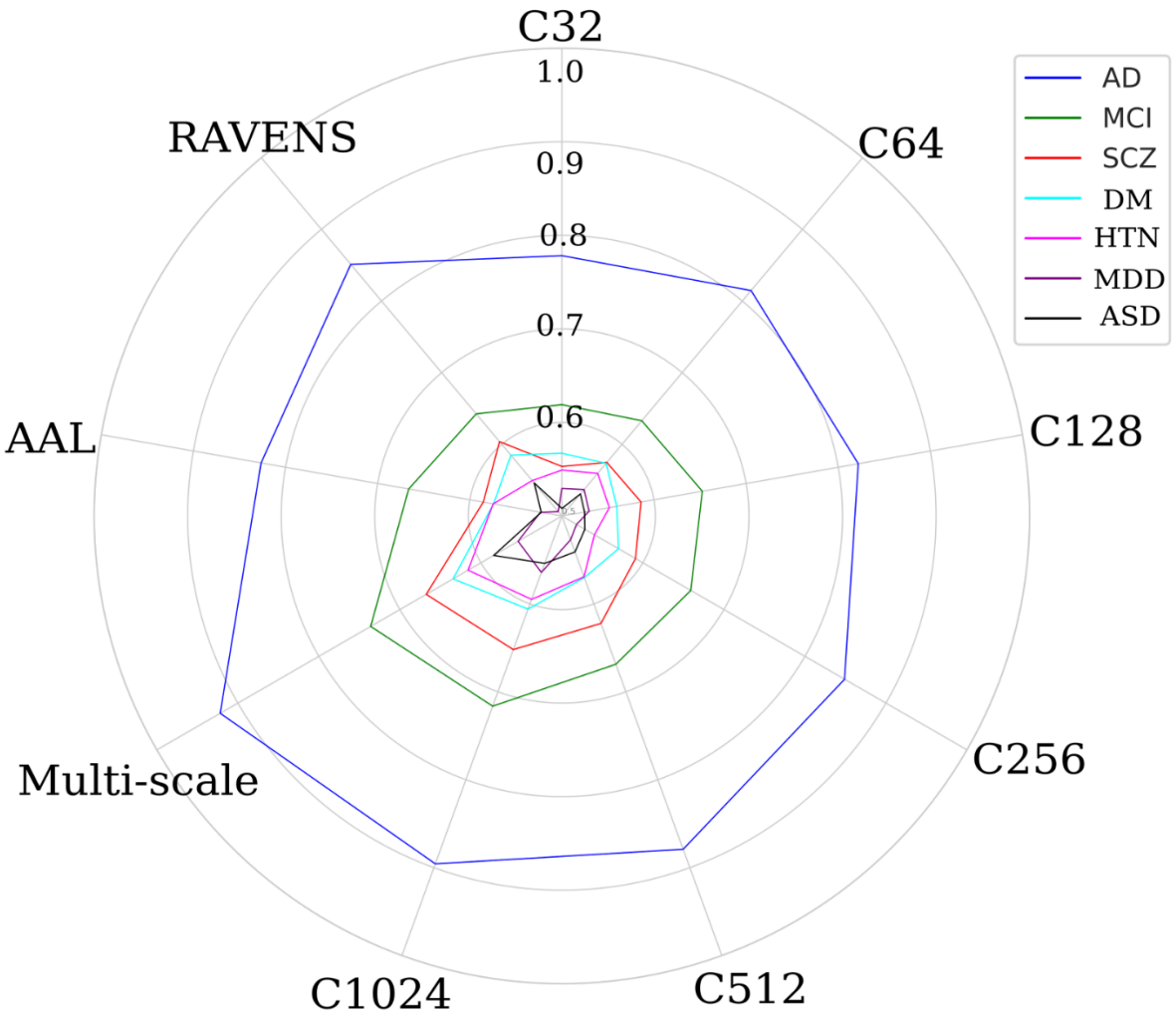


eFigure 3: Scatter plot for the h^2 estimates from the discovery and replication sets. The SNP-based heritability was estimated independently for the discovery set (*N*=18,052) and replication set (*N*=15,243). In particular, the two estimates were highly correlated (r = 0.94, p-value < 10^{-6}), demonstrating a highly similar genetic architecture across different sets of UKBB data.

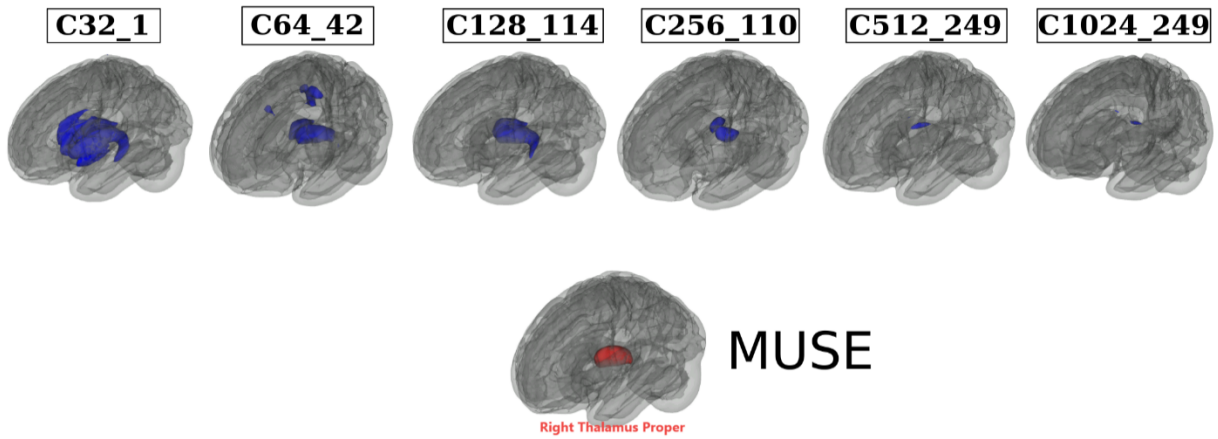


eFigure 4: Sensitivity check for the GWAS results using the discovery set in UKBB. A) The GWAS results for participants with European ancestry in the discovery set. **B**) The GWAS results for participants with European ancestry in the discovery and replication sets. **C**) The GWAS results for participants with all different ancestries in the discovery set. **D**) The GWAS

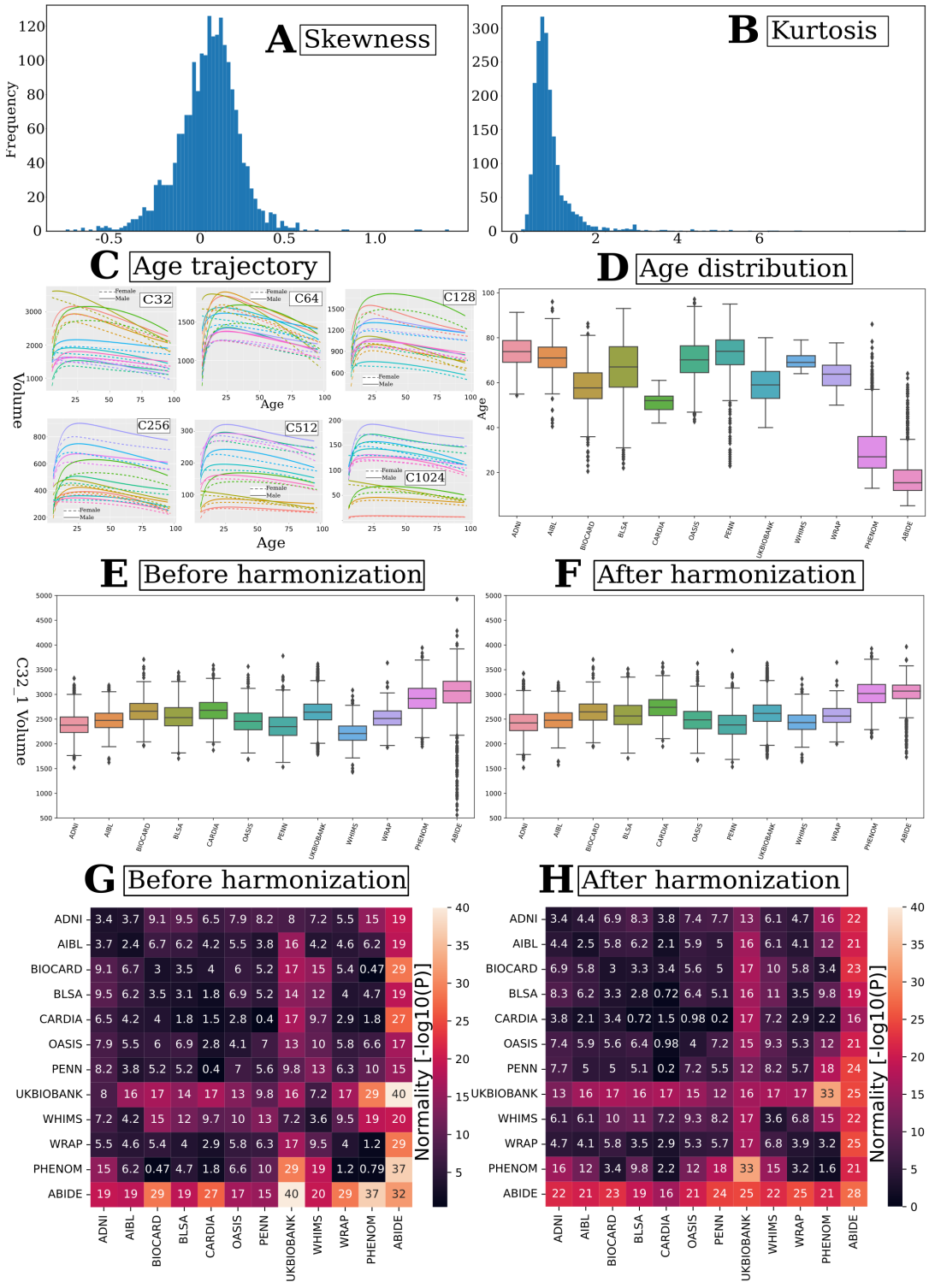
results for participants with European ancestry in the discovery set by adding four additional imaging-related covariates. **E**) The GWAS results for participants with non-European ancestry in the discovery set. **F**) The GWAS results for participants with the independent ADNI WGS data.



eFigure 5: Machine learning performance for disease classification. Balanced accuracy (BA) for each classification task using different features from multi-scale MuSIC, AAL, and RAVENS (higher score better). Details are presented in **eTable 4**.



eFigure 6: Annotation of MUSE ROIs to MuSIC PSCs based on the overlap index. We automatically annotated the 119 MUSE GM PSCs to the MuSIC atlases at all six scales (*C*=32, 64, 128, 256, 512, and 1024). To this end, we calculated an overlap index (OI) to quantify the spatial overlaps between MUSE and MuSIC. For instance, for each MUSE PSC (**eTable 5**) vs. each of the 32 PSCs of MuSIC at *C*=32 scale, the OI equals the proportion of the number of overlap voxels and the total number of voxels in the MUSE PSC. Here we illustrate by mapping the right thalamus of MUSE to all 6 MuSIC atlases. The highest OIs are 0.82, 0.70, 0.86, 0.30, 0.09, 0.05 for C32_1, C64_42, C128_114, C256_110, C512_249, and C1024_249 PSCs. This functionality is available in BRIDGEPORT:
https://www.cbica.upenn.edu/bridgeport/MUSE/Right%20Thalamus%20Proper



290
291 **eFigure 7: Summary statistics of the multi-scale PSCs of MuSIC**. Multi-scale PSCs show
292 considerable normal distributions, i.e., symmetrical distribution (**A**) with a low kurtosis (**B**).
293 Moreover, we fit the Generalized Additive Model for Location, Scale, and Shape (GAMLSS)[9]
294 model (fractional polynomials with 2 degrees) to each PSC to delineate the age trajectory over
295 the lifespan in males (solid lines) and females (dotted lines), respectively (**C**). For visualization

purposes, we selectively display the first 10 PSCs from each scale of the MuSIC atlases. In general, males have larger brain volumes than females. For **D-F**, we selectively showed the distribution of age (**D**) and the distribution of PSC volume before harmonization (**E**) and after harmonization (**F**) for C32_1 within each site in the discovery set. For **G** and **H**, we tested the normality of the PSC volume (C32_1) from each pair of sites using the Shapiro-Wilk test (*scipy.stats.shapiro* function) in the discovery set before (**G**) harmonization and after harmonization (**H**). A higher $-\log_{10}(P)$ indicates the data are less likely to be normally distributed. As a general trend, our statistical harmonization techniques demonstrated a slight improvement in the normality of the data. Additionally, we consistently applied normality transformations to all statistical analyses, including GWAS, to mitigate any non-normality.

eTable 1. Study cohort characteristics.

The current study consists of two main populations/sets: the discovery set (*N*=32,440, including participants from the first download of the UKBB data) and the replication set (*N*=18,259, the second download of the UKBB data). To train the sopNMF model for MuSIC, we selected 250 patients (PT) and 250 healthy controls (CN) for each decade of the discovery set, resulting in 4000 participants in total, referred to as the training population. Age ranges from 5 to 97 years and is shown with mean and standard deviation. Sex is displayed with the number and percentage of female participants. Data was collected from 12 studies, 130 sites, and 12 countries. The number of sites (country) per study is detailed as follows:

- ADNI: 63 sites (USA)
- UKBB: 5 sites (UK)
- AIBL: 2 sites (Australia)
- BIOCARD: 2 sites (USA)
- BLSA: 1 site (USA)
- CARDIA: 3 sites (USA)
- OASIS: 1 site (USA)
- PENN: 1 site (USA)
- WHIMS: 14 sites (USA)
- WRAP 1 site (USA)
- PHENOM: 12 sites (China, Brazil, Australia, Germany, Spain, USA, Netherlands)
- ABIDE: 25 sites (USA, Netherlands, Belgium, Germany, Ireland, Switzerland, France)

Abbreviations: CN: healthy control; AD: Alzheimer's disease; MCI: mild cognitive impairment; SCZ: schizophrenia; ASD: autism spectrum disorder; MDD: major depressive disorder; DM: diabetes; HTN: hypertension.

[a]UKBB data were separately downloaded two times: the first was the *N*=21,305 in the discovery set, and the second was the replication set.

[b]We define CN (healthy controls) as participants that do not have any of the diseases listed here. These CN participants might have diagnoses of other illnesses or comorbidities (e.g., participants from UKBB have a wide range of pathology based on ICD-10).

Study	*N* (50,699)	Age (5-97 year)	Sex (female/%)	CN[b]	AD	MCI	SCZ	ASD	MDD	DM	HTN
Discovery set	32,440	60.04± 14.87	16,868/52	24,980	954	1288	1094	597	1476	1093	958
ADNI	1765	73.66± 7.19	798/45	297	343	875	NA	NA	NA	NA	250
UKBB[a]	21,305	62.58± 7.48	10,101/53	18,735	1	NA	NA	NA	1476	1093	NA
AIBL	830	71.36± 6.78	471/57	625	86	115	NA	NA	NA	NA	4
BIOCARD	288	58.15± 10.54	115/60	283	1	4	NA	NA	NA	NA	NA
BLSA	1114	65.44± 14.11	589/53	729	9	11	NA	NA	NA	NA	365
CARDIA	892	51.21± 3.98	471/53	620	NA	NA	NA	NA	NA	NA	272
OASIS	983	69.92± 9.75	557/57	759	220	NA	NA	NA	NA	NA	4

PENN	807	72.63± 10.65	333/59	173	294	283	NA	NA	NA	NA	57
WHIMS	995	69.61± 3.64	995/100	986	NA	NA	NA	NA	NA	NA	6
WRAP	116	63.36± 6.06	79/68	116	NA	NA	NA	NA	NA	NA	NA
PHENOM	2125	30.21± 10.60	854/40	1031	NA	NA	1094	NA	NA	NA	NA
ABIDE	1220	17.92± 9.01	203/17	623	NA	NA	NA	597	NA	NA	NA
Replication set[a]	18,259	54.70± 7.43	9742/53	NA	NA	NA	NA	NA	NA	NA	NA

337
338

eTable 2: Clinical phenotypes and diagnoses used in machine learning classification.
We harmonized the population of the phenotypes of interest per study definitions:

- We combined AD and MCI patients from ADNI, PENN, and AIBL but excluded OASIS subjects because of the different diagnostic criteria of an AD patient in OASIS.
- For several binary disease phenotypes, we used the ICD-10 diagnosis (https://biobank.ndph.ox.ac.uk/ukb/field.cgi?id=41270). Note that ICD-10 diagnoses are generally collected from the participants' medical inpatient records. We first included diseases from the following categories:
    - Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism (D-XXX, XXX represents the ID of a specific disease);
    - Endocrine, nutritional, and metabolic diseases (E-XXX);
    - Mental and behavioral disorders (F-XXX);
    - Diseases of the nervous system (G-XXX);
    - Diseases of the circulatory system (I-XXX).

  We then set a threshold of 75 patients for any ICD-10 diagnosis. We finally randomly selected age and sex-matched healthy controls (excluding all patients in all diagnoses). [a]: For major depressive disorder, we used the inclusion criteria from our previous work.[10]
- For cognitive scores, we included:
    - Tower rearranging (https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=21004)
    - Matrix pattern (https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=6373)
    - TMT-A (https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=6348)
    - TMT-B (https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=6350)
    - DSST (https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=23324)
    - Pairs matching (https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=399)
    - Numerical memory (https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=4282)
    - Prospective memory (https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=4288)
    - Reaction time (https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=20023)
    - Fluid intelligence (https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=20016)

AD: Alzheimer's disease; MCI: mild cognitive impairment; SCZ: schizophrenia; DM: diabetes mellitus; MDD: major depressive disorder; HTN: hypertension; ASD: autism spectrum disorder; CN: healthy control; PT: patient; *N*: number of participants. We decided not to harmonize cognitive scores from different studies.

Trait (ICD-10 code or ID)	Sample size (CN/PT or *N*)	Site	Trait (ICD-10 code or ID)	Sample size (CN/PT or *N*)	Site
AD	1095/723	ADNI, PENN, & AIBL	Carpal tunnel syndrome (G560)	901/901	UKBB
MCI	1273/1095	ADNI, PENN, & AIBL	Lesion of ulnar nerve (G562)	104/104	UKBB
SCZ	1031/1094	PHENOM	Lesion of plantar nerve (G576)	163/163	UKBB
DM	1093/1093	UKBB	Angina pectoris (I20)	1535/1535	UKBB
MDD[a]	1476/1476	UKBB	Acute myocardial infarction (I21)	769/769	UKBB
HTN	934/887	ADNI, BLSA & CARDIA	Chronic ischaemic heart disease (I25)	2217/2217	UKBB

ASD	623/597	ABIDE	Pulmonary embolism (I20)	351/351	UKBB
Iron deficiency anemia (D50)	1012/1012	UKBB	Cardiomyopathy (I42)	116/116	UKBB
Vitamin B12 deficiency anemia (D50)	78/78	UKBB	Paroxysmal tachycardia (I47)	320/320	UKBB
Agranulocytosis (D70)	245/245	UKBB	Heart failure (I50)	436/436	UKBB
Thyrotoxicosis (E05)	205/205	UKBB	Cerebral infarction (I63)	291/291	UKBB
Vitamin D deficiency (E55)	180/180	UKBB	Vitamin B deficiency (E53)	130/130	UKBB
Obesity (E66)	1481/1481	UKBB	Hemiplegia (G81)	111/111	UKBB
Lipoprotein metabolism disorder (E78)	3880/3880	UKBB	Facial nerve disorders (G51)	95/95	UKBB
Mineral metabolism disorder (E83)	291/291	UKBB	Tower rearranging (21004)	8412	UKBB
Volume depletion	240/240	UKBB	Matrix pattern (6373)	8501	UKBB
Delirium	92/92	UKBB	TMT-A (6348)	8599	UKBB
Alcohol abuse	341/341	UKBB	TMT-B (6350)	8599	UKBB
Tobacco abuse	863/863	UKBB	DSST (23324)	8523	UKBB
Bipolar affective disorder	77/77	UKBB	Pairs matching (399)	20945	UKBB
Phobic anxiety disorder	84/84	UKBB	Numerical memory (4282)	9323	UKBB
Multiple sclerosis	109/109	UKBB	Prospective memory (4288)	19681	UKBB
Epilepsy	250/250	UKBB	Reaction time (20023)	21258	UKBB
Migraine	508/508	UKBB	Fluid intelligence (20016)	19184	UKBB
Sleep disorders	590/590	UKBB			

372
373

eTable 3: Comparison of variants identified via MuSIC with other studies. Using the AAL atlas, we found (using the same data in the current study) that 269 independent significant SNPs had 356 pairwise associations with 54 AAL brain regions. 230 out of the 269 SNPs matched with the SNPs in MuSIC. Among the 39 unmatched SNPs, 15 SNPs were in linkage disequilibrium (LD, $r^2 > 0.6$) with MuSIC SNPs (**Supplementary eFile 5**). As a second example, Zhao et al.[11] reported that 251 independent significant SNPs had 346 pairwise associations with 43 GM regions using the Mindboggle atlas on the UKBB (N=19,629).[12] 129 of the 251 SNPs matched with SNPs identified by MuSIC. Among these non-matching SNPs (127), 31 were in LD with MuSIC SNPs (**Supplementary eFile 6**). Similarly, Elliot et al.[13] (N=8428) discovered that 20 independent significant SNPs had 58 pairwise associations with 52 GM regions from atlases in Freesurfer and FSL software. Out of the 20 SNPs, 16 coincided with MuSIC SNPs. Among the four unmatched SNPs, 1 SNP was in LD with MuSIC SNPs (**Supplementary eFile 7**). Note that the definition of independent significant SNPs or genomic loci might slightly differ between studies.

Study/Atlas	Identified genomic loci	Matched loci	Loci in LD	Novel loci	Database	Sample size	Ancestry
MuSIC	915	NA	NA	NA	UKBB	18,052	European
AAL	218	162	13	740	UKBB	18,052	European
Zhao et al.[11]	251	73	14	828	UKBB	19,629	European
Elliot et al.[13]	20	16	1	898	UKBB	8428	European
GWAS Catalog	NA	298	NA	617	NA	NA	NA

eTable 4: Classification balanced accuracy for disease classification and effect size of these imaging signatures.

Disease classification performance is presented using balanced accuracy. The mean and standard deviation are presented. Cohen's *d* was computed to compare the SPARE scores between groups. Multi-scale classification[a]: All 2003 PSCs from multiple scales were fit into the classifier. Multi-scale classification[b]: PSCs from all scales were fit into the classifier with a nested feature selection procedure (SVM-REF). The motivation is that PSCs from different scales are hierarchical and correlated. The nested feature selection can select the features most relevant to the specific task. We avoided any statistical comparison of the performance of machine learning models because available statistical tests are liberal and often lead to false-positive conclusions due to the complexity of the cross-validation procedure.[14]

a): Classification results for all subjects in all sites using a nested CV procedure

PSC	AD	*d*	MCI	*d*	SCZ	*d*	DM	*d*	HTN	*d*	MDD	*d*	ASD	*d*
C32	0.78± 0.02	1.52	0.62± 0.02	0.59	0.55± 0.02	0.30	0.56± 0.02	0.35	0.55± 0.02	0.28	0.52± 0.02	0.16	0.50± 0.02	0.07
C64	0.81± 0.02	1.73	0.63± 0.02	0.66	0.57± 0.02	0.41	0.57± 0.02	0.40	0.56± 0.02	0.31	0.53± 0.02	0.17	0.53± 0.02	0.19
C128	0.82± 0.02	1.82	0.65± 0.02	0.76	0.59± 0.02	0.47	0.56± 0.02	0.33	0.55± 0.02	0.30	0.52± 0.02	0.15	0.52± 0.02	0.15
C256	0.85± 0.02	2.08	0.66± 0.02	0.91	0.59± 0.02	0.50	0.56± 0.02	0.47	0.54± 0.02	0.31	0.51± 0.02	0.13	0.52± 0.02	0.16
C512	0.88± 0.02	2.34	0.67± 0.02	1.06	0.62± 0.02	0.62	0.57± 0.02	0.54	0.56± 0.02	0.42	0.52± 0.02	0.05	0.54± 0.02	0.24
C1024	0.90± 0.02	2.50	0.72± 0.02	1.12	0.65± 0.02	0.75	0.60± 0.02	0.59	0.59± 0.02	0.46	0.56± 0.02	0.13	0.55± 0.02	0.29
Multi-scale[a]	0.91± 0.02	2.54	0.72± 0.02	1.12	0.66± 0.02	0.77	0.61± 0.02	0.64	0.59± 0.02	0.47	0.55± 0.02	0.23	0.56± 0.02	0.30
Multi-scale[b]	0.92± 0.02	2.61	0.73± 0.02	1.13	0.67± 0.02	0.78	0.64± 0.02	0.67	0.61± 0.02	0.49	0.55± 0.02	0.26	0.58± 0.02	0.32
AAL	0.82± 0.02	1.81	0.66 ±0.02	0.75	0.59± 0.02	0.46	0.57± 0.02	0.32	0.57± 0.02	0.35	0.52± 0.02	0.08	0.52± 0.02	0.14
RAVENS	0.85± 0.02	2.04	0.64 ±0.02	0.74	0.60± 0.02	0.45	0.58± 0.02	0.33	0.55± 0.02	0.34	0.50± 0.02	0.05	0.54± 0.02	0.15

b): The classification results of the balanced accuracy (BA) from the test data in the nested CV and the independently left-out site for the task of AD vs. CN were assessed using all available multi-scale PSCs[a]. Three sites, namely ADNI, AIBL, and PENN, were considered for this analysis. However, UKBB, BIOCARD, and BLSA data were excluded due to limited AD cases (**eTable 1**). Similarly, data from OASIS were excluded due to discrepancies in the diagnosis criteria for AD, as previously stated in our previous work[7].

Left-out site	Test BA in CV	Test BA in the left-out site
ADNI	0.90±0.02	0.88±0.02
AIBL	0.88±0.02	0.95±0.02
PENN	0.90±0.02	0.95±0.02

eTable 5: 119 MUSE gray matter regions of interest.
L: Left hemisphere; R: Right hemisphere; ROI: region of interest.

MUSE ROI	MUSE ROI	MUSE ROI
Precentral gyrus (R)	Occipital fusiform gyrus (R)	Anterior insula (L)
Precentral gyrus (L)	Planum temporale (R)	Anterior orbital gyrus (R)
Accumbens area (R)	Cerebellar vermal lobules I-V	Anterior orbital gyrus (L)
Accumbens area (L)	Cerebellar vermal lobules VI-VII	Angular gyrus (R)
Amygdala (R)	Cerebellar vermal lobules VIII-X	Angular gyrus (L)
Amygdala (L)	Basal forebrain (R)	Calcarine cortex (R)
Occipital pole (L)	Basal forebrain (L)	Calcarine cortex (L)
Caudate (R)	Middle temporal gyrus (L)	Central operculum (R)
Caudate (L)	Occipital pole (R)	Central operculum (L)
Cerebellum exterior (R)	Planum temporale (L)	Cuneus (R)
Cerebellum exterior (L)	Parietal operculum (L)	Cuneus (L)
Planum polare (L)	Postcentral gyrus (R)	Entorhinal area (R)
Middle temporal gyrus (R)	Postcentral gyrus (L)	Entorhinal area (L)
Hippocampus (R)	Posterior orbital gyrus (R)	Frontal operculum (R)
Hippocampus (L)	Temporal pole (R)	Frontal operculum (L)
Precentral gyrus medial segment (R)	Temporal pole (L)	Frontal pole (R)
Precentral gyrus medial segment (L)	Triangular part of the inferior frontal gyrus (R)	Frontal pole (L)
Superior frontal gyrus medial segment (R)	Triangular part of the inferior frontal gyrus (L)	Fusiform gyrus (R)
Superior frontal gyrus medial segment (L)	Transverse temporal gyrus (R)	Fusiform gyrus (L)
Pallidum (R)	Superior frontal gyrus medial segment (L)	Gyrus rectus (R)
Pallidum (L)	Planum polare (R)	Gyrus rectus (L)
Putamen (R)	Transverse temporal gyrus (L)	Inferior occipital gyrus (R)
Putamen (L)	Anterior cingulate gyrus (R)	Inferior occipital gyrus (L)
Thalamus proper (R)	Anterior cingulate gyrus (L)	Inferior temporal gyrus (R)
Thalamus proper (L)	Anterior insula (R)	Inferior temporal gyrus (L)
Lingual gyrus (R)	Occipital fusiform gyrus (L)	Subcallosal area (R)
Lingual gyrus (L)	Opercular part of inferior frontal gyrus (R)	Subcallosal area (L)
Lateral orbital gyrus (R)	Opercular part of inferior frontal gyrus (L)	Superior frontal gyrus (R)
Lateral orbital gyrus (L)	Orbital part of inferior frontal gyrus (R)	Superior frontal gyrus (L)
Middle cingulate gyrus (R)	Orbital part of inferior frontal gyrus (L)	Supplementary motor cortex (R)
Middle cingulate gyrus (L)	Posterior cingulate gyrus (R)	Supplementary motor cortex (L)
Medial frontal cortex (R)	Posterior cingulate gyrus (L)	Supramarginal gyrus (R)
Medial frontal cortex (L)	Precuneus (R)	Supramarginal gyrus (L)
Middle frontal gyrus (R)	Precuneus (L)	Superior occipital gyrus (R)
Middle frontal gyrus (L)	Parahippocampal gyrus (R)	Superior occipital gyrus (L)
Middle occipital gyrus (R)	Parahippocampal gyrus (L)	Superior parietal lobule (R)
Middle occipital gyrus (L)	Posterior insula (R)	Superior parietal lobule (L)
Medial orbital gyrus (R)	Posterior insula (L)	Superior temporal gyrus (R)
Medial orbital gyrus (L)	Parietal operculum (R)	Superior temporal gyrus (L)
Superior frontal gyrus medial segment (R)	Posterior orbital gyrus (L)	

eAlgorithm 1: Algorithm for sopNMF.

The source code of the Python implementation of sopNMF is available here: https://github.com/anbai106/SOPNMF

Algorithm 1: sopNMF

- **Input:** maximum number of epochs e, number of component C or r, batch size b, early stopping criteria θ (i.e., the loss without decreasing for a certain epochs) ;
- **Output:** $\boldsymbol{W} \in \mathbb{R}^{d \times r}, \boldsymbol{H} \in \mathbb{R}^{r \times n}$;
- **Initialization:** $\boldsymbol{W}$;

if *not* θ *or epoch* $\neq e$ **then**
  for $p \leftarrow 0$ **to** e **do**
    for $i \leftarrow 0$ **to** t **do**
      Read mini-batch $\boldsymbol{X}_{bi}$
      Update $\boldsymbol{W}_{i+1}$ *via Eq. 2*
    end
    $loss = \sum_{i=1}^{\lceil \frac{n}{b} \rceil} \|\boldsymbol{X}_{bi} - \boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X}_{bi}\|_F^2 (Eq.3)$
    if *loss in* θ **then**
      Stop
    else
      Shuffle $\boldsymbol{X}$
      Continue
    end
  end
else
  Stop
end

References

1. Shalev-Shwartz, S., Singer, Y. & Ng, A. Y. Online and batch learning of pseudo-metrics. in *Proceedings of the twenty-first international conference on Machine learning* 94 (Association for Computing Machinery, 2004). doi:10.1145/1015330.1015376.
2. Kuhn, H. W. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly* **2**, 83–97 (1955).
3. Pomponio, R. *et al.* Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. *Neuroimage* **208**, 116450 (2020).
4. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).
5. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* **81**, 559–575 (2007).
6. Samper-González, J. *et al.* Reproducible evaluation of classification methods in Alzheimer's disease: Framework and application to MRI and PET data. *NeuroImage* **183**, 504–521 (2018).
7. Wen, J. *et al.* Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation. *Medical Image Analysis* **63**, 101694 (2020).
8. Wen, J. *et al.* Reproducible Evaluation of Diffusion MRI Features for Automatic Classification of Patients with Alzheimer's Disease. *Neuroinformatics* **19**, 57–78 (2021).
9. Stasinopoulos, D. M. & Rigby, R. A. Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. *Journal of Statistical Software* **23**, 1–46 (2008).

10. 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.
11. 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).
12. Klein, A. & Tourville, J. 101 Labeled Brain Images and a Consistent Human Cortical Labeling Protocol. *Frontiers in Neuroscience* **6**, 171 (2012).
13. Elliott, L. T. *et al.* Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210–216 (2018).
14. Nadeau, C. & Bengio, Y. Inference for the Generalization Error. in *Advances in Neural Information Processing Systems 12* (eds. Solla, S. A., Leen, T. K. & Müller, K.) 307–313 (MIT Press, 2000).