















A genetically informed brain atlas for enhancing brain imaging genomics

Received: 14 November 2024

Accepted: 24 February 2025

Published online: 14 April 2025

 Check for updates

Jingxuan Bao ^{1,2}, Junhao Wen ^{3,4,5}, Changgee Chang ⁶, Shizhuo Mu¹, Jiong Chen^{1,7}, Manu Shivakumar ^{1,2}, Yuhao Cui⁸, Guray Erus ⁸, Zhijian Yang^{8,9}, Shu Yang ¹, Zixuan Wen^{1,9}, The Alzheimer's Disease Neuroimaging Initiative*, Yize Zhao ¹⁰, Dokyoon Kim ¹, Duy Duong-Tran ^{1,11}, Andrew J. Saykin ¹², Bingxin Zhao ¹³, Christos Davatzikos ¹⁴, Qi Long ¹✉ & Li Shen ¹✉

Brain imaging genomics has manifested considerable potential in illuminating the genetic determinants of human brain structure and function. This has propelled us to develop the GIANT (Genetically Informed brAiN aTlas) that accounts for genetic and neuroanatomical variations simultaneously. Integrating voxel-wise heritability and spatial proximity, GIANT clusters brain voxels into genetically informed regions, while retaining fundamental anatomical knowledge. Compared to conventional (non-genetics) brain atlases, GIANT exhibits smaller intra-region variations and larger inter-region variations in terms of voxel-wise heritability. As a result, GIANT yields increased regional SNP heritability, enhanced polygenicity, and its polygenic risk score explains more brain volumetric variation than traditional neuroanatomical brain atlases. We provide extensive validation to GIANT and demonstrate its neuroanatomical validity, confirming its generalizability across populations with diverse genetic ancestries and various brain conditions. Furthermore, we present a comprehensive genetic architecture of the GIANT regions, covering their functional annotation at the molecular levels, their associations with other complex traits/diseases, and the genetic and phenotypic correlations among GIANT-defined imaging endophenotypes. In summary, GIANT constitutes a brain atlas that captures the complexity of genetic and neuroanatomical heterogeneity, thereby enhancing the discovery power and applicability of imaging genomics investigations in biomedical science.

The advance of large-scale, collaborative brain imaging genomics consortia, such as ENIGMA^{1,2} and UK Biobank^{3,4}, has ushered unprecedented opportunities to gain insights into the human brain - the most intricate organ in the human body. Seizing upon this trend of open science, researchers have discovered many genetic variants associated with brain function and structure⁵⁻¹². The identified genetic variants have facilitated the understanding of disease etiology, biological pathways, and gene-guided drug discovery/repurposing, possibly

paving the road towards personalized medicine¹³⁻¹⁶. An exemplary illustration can be found in a recent genome-wide association study (GWAS) by Zhao et al., where a multitude of single nucleotide polymorphisms (SNP) was found to be associated with imaging-derived phenotypes (IDP) obtained from diffusion magnetic resonance imaging (MRI)¹⁷. These IDPs exhibited a significant enrichment of heritability in glial cells, but not in neurons. Nevertheless, conventional neuroanatomically defined brain atlases such as Desikan atlas¹⁸, were

A full list of affiliations appears at the end of the paper. *A list of authors and their affiliations appears at the end of the paper.

✉ e-mail: qlong@pennmedicine.upenn.edu; li.shen@pennmedicine.upenn.edu

employed in these studies to generate these IDPs. These brain atlases solely account for neuroanatomical variations, but may not necessarily be genetically relevant, thereby impeding the discovery power in subsequent GWAS.

To address this limitation, an urgent need exists to develop a data-driven brain parcellation approach that incorporates information on both genetic and neuroanatomical variations. Pioneering works in this area have already been initiated. For instance, Chen et al. developed a genetically informed atlas that utilized MRI and genetic data from 406 twins, which partitioned the cortical surface area into genetic subdivisions¹⁹. Additionally, a recent study utilized generative adversarial networks to generate disease dimensions, also known as subtypes, that were informative of both imaging and genetic variations²⁰. Within this research trajectory, we proposed GIANT (Genetically Informed brAiN aTlas) to parcellate the human brain via a genetically guided approach. Specifically, we developed a heritability-aware brain parcellation model to cluster the spatially connected voxels of the brain into regions with similar heritability.

Brain morphological development and changes are largely influenced by genetic factors¹². To advance the field of brain imaging genomics and better understand the genetic underpinnings of brain morphology, we hypothesize that the region-level grey- and white-matter densities derived from GIANT, compared to those of neuroanatomy-aware brain atlases such as a multi-atlas parcellation method (MUSE)²¹, can provide higher discovery power and serve as more robust instruments in imaging genomics analyses. In the present study, GIANT divides the human brain into 50 regions of interest (ROIs) that are in alignment with established brain anatomy but are guided by voxel-wise SNP heritability. Our experiments demonstrated that GIANT showed greater discovery power than non-genetic brain atlases in identifying genetic variants associated with brain atrophy. Moreover, we offered a comprehensive landscape of the genetic architecture of GIANT, including its functional annotation, the

associations between GIANT regions and other complex traits, and the genetic/phenotypic correlations among GIANT-defined imaging endophenotypes. Additionally, we map the genes associated with our GIANT imaging-genomics GWAS hits through positional and expression quantitative trait loci (eQTL) mapping. Our results highlight the capability of GIANT to understand the genetic underpinning of brain structures, potentially facilitating gene-guided drug discovery/repurposing and personalized medicine in brain-related disorders.

Results

Atlas Delineation: A framework to define genetically informed brain atlas

We introduced the GIANT atlas along with a framework to define it, aimed at enhancing the discovery power for brain imaging-genomics studies (Fig. 1). Briefly, we first designed a heritability-aware brain parcellation model - a three-dimensional clustering method that integrates heritability and spatial proximity (Method 1)²². Then, we applied our heritability-aware brain parcellation model to the SNP heritability derived from voxel-level gray matter and white matter densities (Supplementary Methods 1–4). As a result, our framework grouped the spatially connected brain voxels with similar heritability, leading to the creation of 50 genetically informed brain ROIs. Specifically, to create and validate GIANT atlas, we downloaded raw T1-weighted MRIs and imputed genotyping data from the UK Biobank (UKBB) and the Alzheimer's disease neuroimaging initiative (ADNI) (Fig. 1A (1)). We extracted voxel-level brain gray matter and white matter densities for each individual using regional analysis of volumes examined in normalized space (RAVENS) (Fig. 1A (2)). After performing initial quality control, we were left with 38,290 subjects (35,181 white British ancestry and 3,109 other ancestries) for the UKBB imaging-genomics cohort. For the ADNI imaging-genomics cohort, we were left with 1,809 subjects. We harmonized the imputed genotyping for UKBB and ADNI, which resulted in 6,965,659 SNP

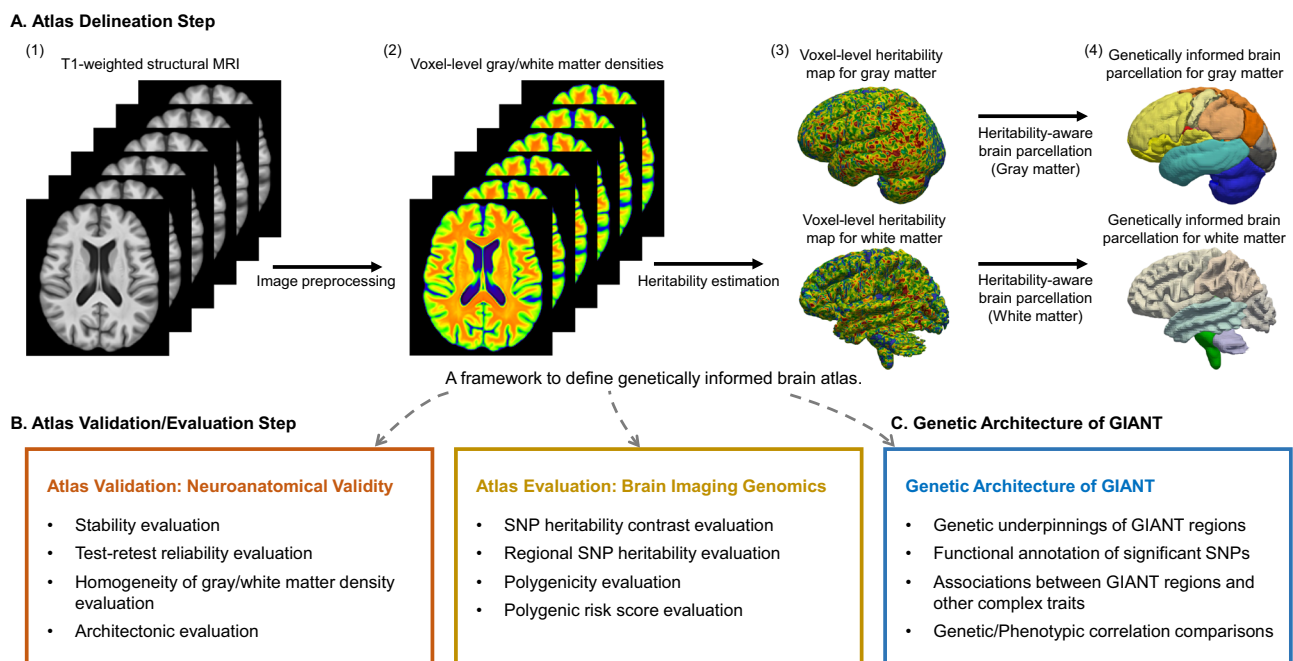


Fig. 1 | A framework to define genetically informed brain atlas. A Atlas Delineation: The framework for defining GIANT begins with preprocessing T1-weighted structural MRI data. The preprocessed data is then used to calculate the voxel-level gray matter and white matter densities. We then estimate the SNP heritability for each brain voxel in gray matter and white matter, respectively. Next, a heritability-aware brain parcellation model is applied to both gray matter and white matter to cluster the brain voxels into regions according to their heritability information and

spatial proximity. The GIANT is defined by combining the gray matter and white matter parcellations. **B** Atlas Validation/Evaluation: Subsequently, we performed a series of validation and evaluation steps, including neuroanatomical validation and brain imaging genomics evaluations, to demonstrate that GIANT functions as a neuroanatomical brain atlas and enhances discovery power in brain imaging genomics studies. **C** Genetic Architecture: Finally, we present the genetic architecture of GIANT. GIANT genetically informed brain atlas, MRI magnetic resonance imaging,

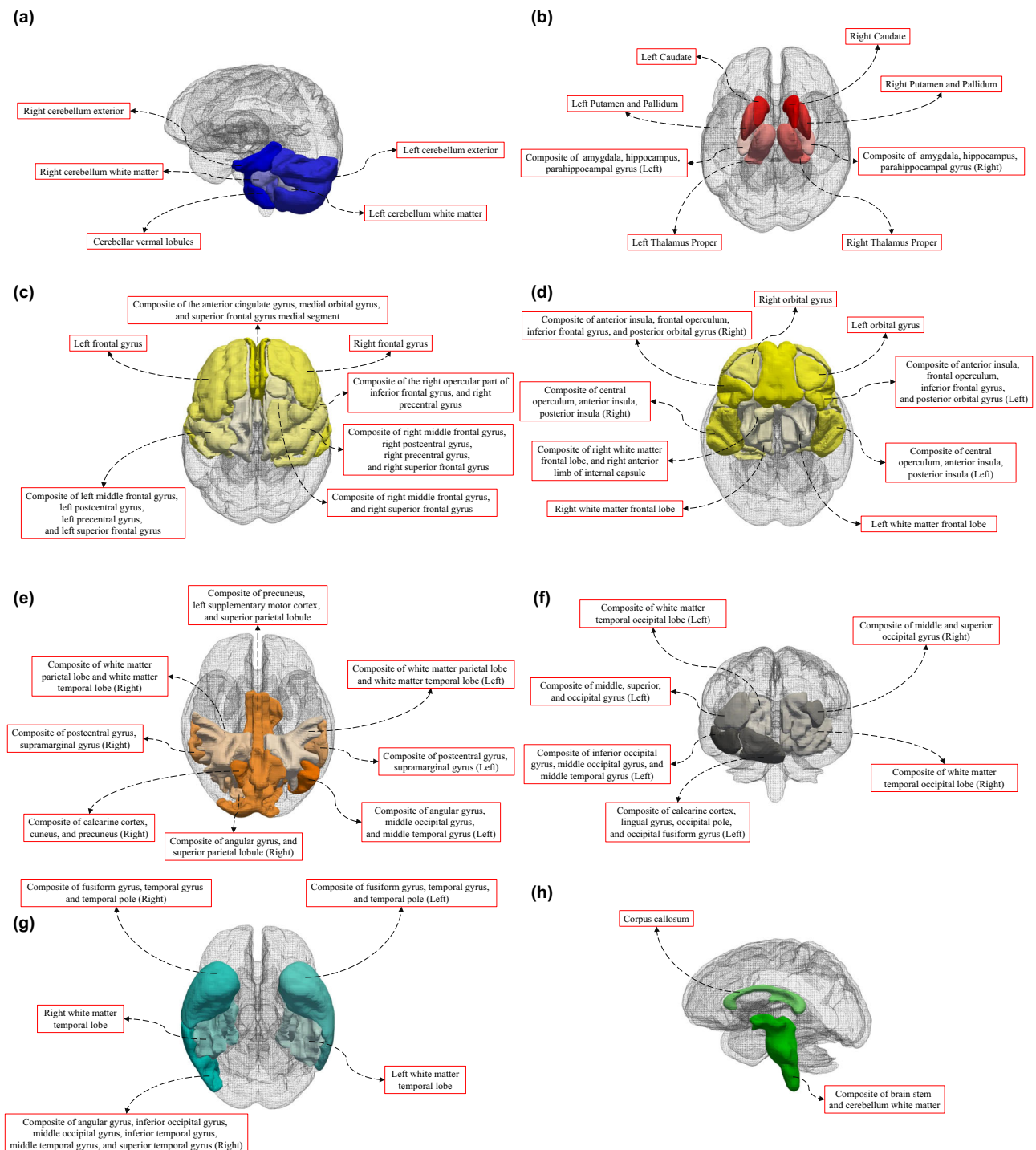


Fig. 2 | GIANt: Genetically informed brain atlas. GIANt integrates the SNP heritability and spatial proximity. We annotate the GIANt into 7 brain substructures, including **a** cerebellum; **b** the deep structure of gray matter and white matter structure; **c** frontal structures (superior); **d** frontal structures (inferior); **e** parietal

structure; **f** occipital structure; **g** temporal structure; and **h** others. We used the MUSE²¹ atlas to annotate the GIANt (Method 6). The region specification for GIANt and MUSE can be found in Supplementary Data 1 and 2. GIANt genetically informed brain atlas, SNP single nucleotide polymorphism.

variants (Method 2 - 4). To define GIANt, we estimated the SNP heritability for both gray matter density (84,090 voxels) and white matter density (67,795 voxels) using 5,000 randomly selected UKBB white British imaging-genomics cohort (Fig. 1A (3)). The SNP heritability was derived using the LD-adjusted kinships (LDAK) software^{23,24}. Finally, we applied our heritability-aware brain parcelation model to the gray matter and white matter separately to segment the brain into genetically informed ROIs (Fig. 1A (4)), which were then combined to define the GIANt atlas.

To evaluate the GIANt atlas, we undertook analyses from three distinct perspectives. First, from an imaging standpoint (Fig. 1B), we conducted a systematic analysis to compare the similarity of the architectonic boundaries of GIANt with those of traditional non-genetic brain atlases. To affirm the neuroanatomical validity of the GIANt atlas, we further validated its stability, test-retest reliability, and gray/white matter homogeneity. Subsequently, from a genomics angle (Fig. 1B), compared to MUSE and a genetically informed brain atlas created using the Watershed algorithm (Watershed-based atlas)

(Supplementary Methods 5), the GIANT atlas exhibited a larger ratio of between-region to within-region SNP heritability dispersion, along with enhanced regional-level SNP heritability and polygenicity. Lastly, we provided the genetic architecture of GIANT-defined IDPs (Fig. 1C).

Atlas validation: neuroanatomical validity of GIANT

We introduced GIANT, a Genetically Informed brAiN aTlas (Fig. 2, Supplementary Data 1), developed through a three-dimensional clustering algorithm (Method 1) applied to the densities of gray and white matter, resulting in tissue-specific brain parcellations. The best-tuned parcellations for each tissue were selected and combined to formulate GIANT (Method 5), and the brain regions were annotated based on existing brain atlases (Method 6). GIANT was subdivided into 7 anatomical sub-structures: cerebellum (Fig. 2a), deep gray matter and white matter structure (Fig. 2b), frontal structure (Fig. 2c, d), parietal structure (Fig. 2e), occipital structure (Fig. 2f), temporal structure (Fig. 2g), and others (Fig. 2h). In the present section, we conducted extensive neuroanatomical assessments (Method 7) of GIANT to confirm its neuroanatomical validity.

Stability evaluation. To evaluate GIANT's neuroanatomical validity from a stability perspective, we compared it with an independently generated genetically informed brain atlas derived from a separate sample of 5,000 non-overlapping UKBB white British individuals. The high concordance between the two atlases, indicated by an adjusted Rand index of 0.91 and an adjusted mutual information score of 0.93, demonstrates that GIANT maintains its structure across bootstrapped UKBB data. This stability is significantly superior to the Watershed-based atlas, which showed an adjusted Rand index of 0.17 and an adjusted mutual information score of 0.55. These results further solidify GIANT's validity as a neuroanatomical brain atlas.

Test-retest reliability evaluation. The test-retest reliability of GIANT was evaluated using intra-class correlation (ICC) coefficients derived from longitudinal data in the UKBB and ADNI cohorts, involving a total of 3273 subjects (1917 subjects in ADNI and 1356 subjects in UKBB). The evaluations were based on the initial and final visits of the same individuals. For each brain region, six different ICC coefficients were calculated^{25,26}, and the mean values across all regions were used to assess overall reliability. GIANT exhibited excellent reliability²⁷, with all correlation coefficients exceeding 0.9 (Supplementary Data 3), outperforming both MUSE and the Watershed-based atlas. These findings reinforce GIANT's validity as a neuroanatomical atlas.

Homogeneity evaluation. We assessed the homogeneity of gray and white matter densities within GIANT across three population cohorts: the UKBB white British discovery cohort, the UKBB non-white-British replication cohort, and the ADNI replication cohort. Using the approach adapted from Schaefer et al.²⁸, we measured the homogeneity by calculating the weighted standard deviation of regional densities, with lower standard deviations indicating greater homogeneity within each brain region. GIANT demonstrated consistently lower weighted average standard deviations across all three cohorts - 93.92 in the UKBB white British discovery cohort, 96.72 in the UKBB non-white-British replication cohort, and 95.47 in the ADNI replication cohort - compared to MUSE (106.05, 108.34, and 106.71, respectively) and the Watershed-based atlas (95.48, 98.21, and 97.37, respectively). These results suggest that GIANT defines brain regions with greater homogeneity across different populations, further validating GIANT as a neuroanatomical brain atlas.

Architectonic comparisons. To assess GIANT's ability to capture known architectonic boundaries, we compared it with a range of established brain atlases that delineate regions based on anatomical landmarks and other neuroimaging modalities²⁹. These atlases

included Automated Anatomical Labeling (AAL) atlas³⁰, the atlas of Intrinsic Connectivity of Homotopic Areas (AICHA)³¹, the whole-brain fMRI atlas generated via spatially constrained spectral clustering (CPAC200)³², and several others, such as the Desikan¹⁸, Hammersmith³³, MUSE²¹, Schaefer²⁸, Talairach³⁴, and Yeo³⁵ atlases.

The evaluation focused on both cortical (Fig. 3a) and gray matter (Fig. 3b) regions. We used the adjusted mutual information (AMI) score to quantify the alignment of GIANT's architectonic boundaries with those of the reference atlases. GIANT exhibited moderate agreement with most of these atlases (AMI scores between 0.4 and 0.8), reflecting its ability to capture key architectonic in both cortical regions and gray matter tissue. As a sanity check, the AMI scores for regions within the Schaefer and Yeo atlas sets were consistently greater than 0.8, which was expected since these atlases were created using the same methodologies.

The moderate agreement of the GIANT atlas with other brain atlases is consistent with our expectations, given that GIANT incorporates SNP heritability information to enhance its discovery power in brain imaging-genomics. These results align with our hypothesis that the GIANT atlas would adjust the architectonic boundaries of established neuroanatomical brain atlases to improve discovery power in brain imaging-genomics while maintaining core anatomical knowledge.

Atlas Evaluation: GIANT for enhancing the brain imaging genomics

GIANT unveils enhanced SNP heritability contrast and increased regional SNP heritability. GIANT reveals an enhanced contrast in SNP heritability, exhibiting a larger ratio of between-region to within-region SNP heritability dispersion. Specifically, we used our three-dimensional clustering algorithm to segment the brain's gray matter and white matter and integrate the optimally tuned brain parcellations for both tissues. We annotated the resulting regions using our brain region annotation strategy (Method 6). We compared the within-region voxel-level SNP heritability dispersion and the between-region voxel-level SNP heritability dispersion of GIANT with two other brain atlases - the MUSE atlas and the Watershed-based atlas. We assessed the relationship between within-region heritability dispersion and between-region heritability dispersion using the Calinski-Harabasz (CH) score³⁶. GIANT consistently exhibits the highest CH score across cohort comparisons, indicating its superiority in grouping voxels with similar heritability estimates into regions while maximizing the heritability differences between regions.

To access the regional SNP heritability for imaging-derived endophenotypes induced by GIANT atlas in brain gray matter and white matter densities, we compared the region-level SNP heritability estimates among GIANT, MUSE, DKT, HarvardOxford, and the Watershed-based atlas (Supplementary Data 4). We identified a significant difference in the distribution of region-level heritability, with the heritability estimates from GIANT demonstrating significantly higher SNP heritability compared to those from the MUSE atlas (one-sided Wilcoxon rank sum test $p\text{-value} = 3.35 \times 10^{-5}$), DKT atlas (one-sided Wilcoxon rank sum test $p\text{-value} = 2.18 \times 10^{-7}$), HarvardOxford atlas (one-sided Wilcoxon rank sum test $p\text{-value} = 6.62 \times 10^{-7}$), and the Watershed-based atlas (one-sided Wilcoxon rank sum test $p\text{-value} = 1.16 \times 10^{-3}$). Our results suggest that genetics may account for a greater portion of the phenotypic variations in imaging-derived endophenotypes for the GIANT atlas than the traditional neuroanatomically defined atlases or those formulated based solely on brain neuroimaging modalities.

GIANT yields enhanced polygenicity. To access the discovery power of GIANT, we conducted region-level GWAS using GIANT, MUSE, and the Watershed-based atlas. In our discovery cohort, we included all individuals of white British ancestry from the UKBB imaging-genomics

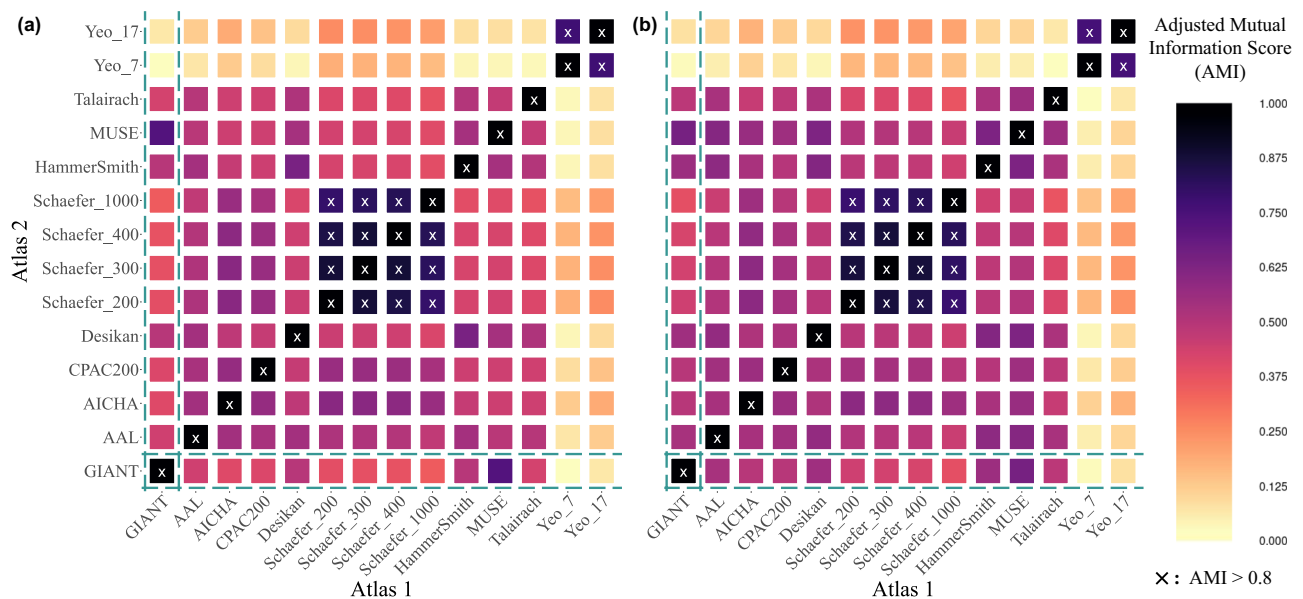


Fig. 3 | The architectonic similarity of the cortical and gray matter regions of selected brain atlases. We evaluate the similarity between brain atlases using adjusted mutual information (AMI) score. We plot the pairwise AMI scores for GIANT atlas and some other selected brain atlas within **a** cortical regions and **b** gray matter tissue. Darker color represents higher concordance between two atlases. We marked the AMI > 0.8 with “x” representing the “perfect alignments”. The GIANT atlas is highlighted using blue dashed lines. CPAC200: a whole brain fMRI atlas generated via spatially constrained spectral clustering; Desikan: an automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral-

based regions of interest; Hammersmith: an automatic segmentation of young children’s brains; MUSE: an ensemble multi-atlas parcellation; Schaefer: a local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI; Talairach: automated Talairach atlas labels for functional brain mapping; Yeo: the organization of the human cerebral cortex estimated by intrinsic functional connectivity. AMI adjusted mutual information, AAL automated anatomical labeling, AICHA an atlas of intrinsic connectivity of homotopic areas, MRI magnetic resonance imaging.

cohort, excluding the 5,000 subjects randomly selected for atlas creation, totaling 30,181 individuals. For replication, we divided it into two parts: a UKBB replication cohort comprising individuals from non-white-British ancestries with 3,109 individuals, and the ADNI replication cohort with 1,809 subjects. This design aimed to assess the generalizability of our GIANT atlas: for the UKBB replication cohort, we seek to evaluate its ability to maintain superior discovery power across different ancestries; and for the ADNI replication cohort, we attempt to evaluate its ability to maintain superior discovery power in cohort with significant brain atrophy patterns.

For the GWAS, in our results, GIANT identified an average of 61.72 significant independent SNPs per ROI (genome-wide significance threshold of 5×10^{-8}) in the UKBB white British discovery cohort. Of these, an average of 4.98 were replicated in the UKBB non-white British and the ADNI disease cohort. The significant threshold for the replication is set to be 0.05 due to the small replication sample size. In contrast, using the MUSE atlas, we identified an average of 36.59 significant independent SNPs per ROI (genome-wide significance threshold of 5×10^{-8}) in the UKBB white British discovery cohort, of which an average of 2.96 were replicated in the UKBB non-white-British and the ADNI disease cohort. We performed the meta-analysis (Method 8) to integrate the GWAS results derived from discovery and replication cohorts and prioritized the lead SNPs (Fig. 4). Moreover, we identified an average of 27.01, 22.89, and 46.17 significant independent SNPs per ROI for DKT, HarvardOxford, and the Watershed-based atlas in the UKBB white British discovery cohort where an average of 2.29, 1.93, and 3.93 significant independent SNPs were replicated using UKBB non-white British and ADNI replication cohorts.

Our results demonstrate that the GIANT has significantly more independent GWAS signals than the MUSE brain atlas in the UKBB discovery cohort (one-sided Wilcoxon rank sum test p -value = 6.03×10^{-5}), and exhibits a higher number of independent GWAS signals that could be replicated by both the UKBB and

ADNI replication cohorts (one-sided Wilcoxon rank sum test p -value = 1.28×10^{-4}). Moreover, GIANT yielded significantly more independent GWAS signals across the discovery and replication cohorts compared to the DKT brain atlas³⁷ (one-sided Wilcoxon rank sum test p -value = 9.18×10^{-7} for the UKBB discovery cohort; and one-sided Wilcoxon rank sum test p -value = 1.09×10^{-4} for the significant GWAS results replicated by both UKBB replication cohort and ADNI replication cohort), HarvardOxford brain atlas³⁸ (one-sided Wilcoxon rank sum test p -value = 9.12×10^{-7} for the UKBB discovery cohort; and one-sided Wilcoxon rank sum test p -value = 2.03×10^{-5} for the significant GWAS results replicated by both UKBB replication cohort and ADNI replication cohort), and Watershed-based atlas (one-sided Wilcoxon rank sum test p -value = 4.22×10^{-3} for the UKBB discovery cohort; and one-sided Wilcoxon rank sum test p -value = 4.88×10^{-3} for the significant GWAS results replicated by both UKBB replication cohort and ADNI replication cohort) (Supplementary Fig. 1). These results highlight GIANT’s enhanced discovery power in brain imaging genomics, affirming its generalizability across diverse ancestries and in cohorts with severe brain disorders and brain atrophy patterns.

To assess the robustness of our GWAS findings, we performed a sensitivity analysis by comparing the β coefficients of significant GWAS signals across the discovery cohort and two replication cohorts (Method 8). The results showed strong robustness, with a weighted mean Pearson correlation of 0.82 for the β coefficients across 50 GIANT brain regions between the white British discovery cohort and the non-white British replication cohort. Similarly, a weighted mean Pearson correlation of 0.77 was observed between the white British discovery cohort and the ADNI replication cohort. Additionally, 93% of the β coefficient signs were in agreement between the UKBB white-British discovery cohort and the non-white-British replication cohort, weighted by the number of significant GWAS signals. A similar agreement of 90% was found between the UKBB white-British discovery cohort and the ADNI replication

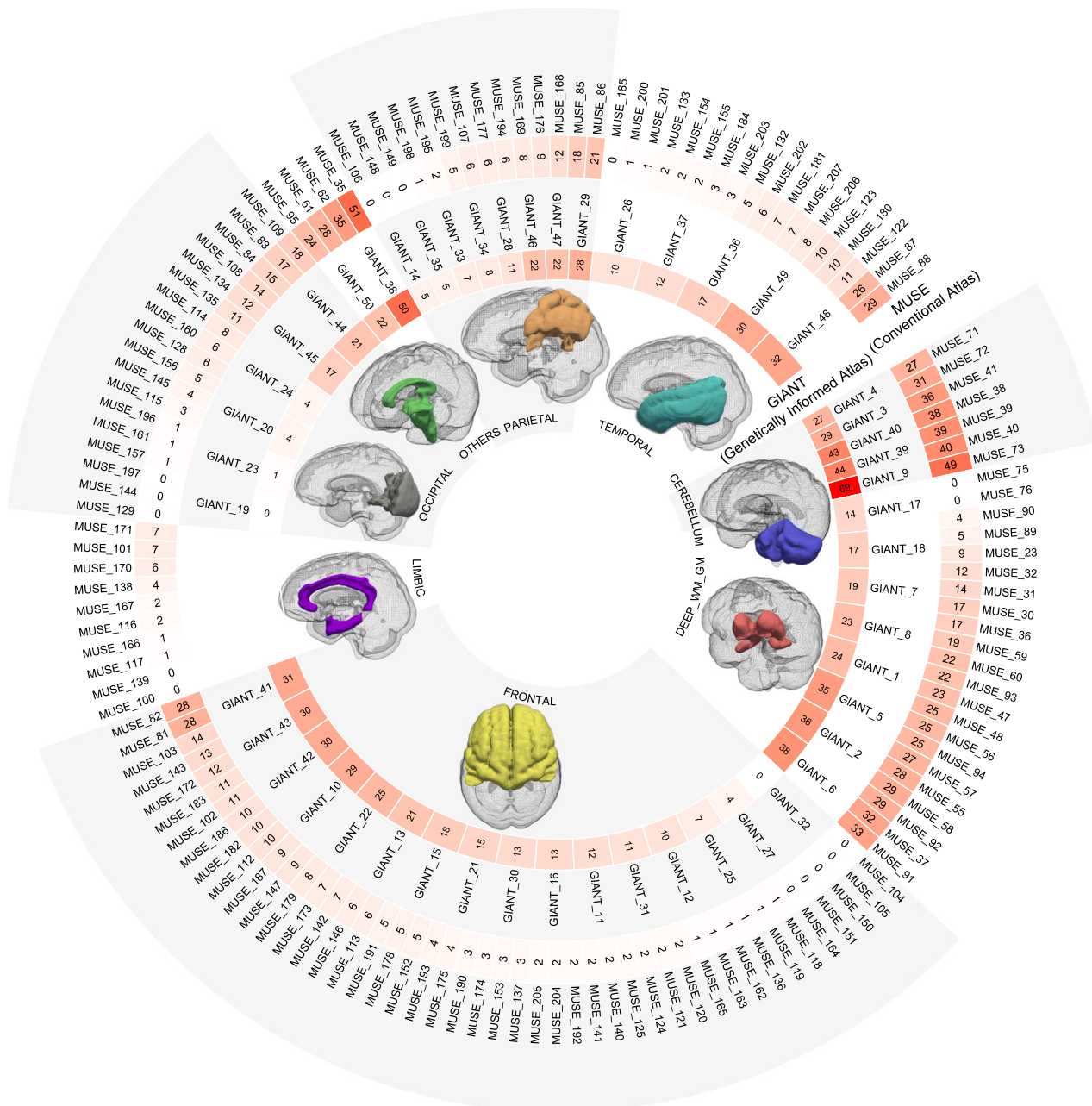


Fig. 4 | Comparison of ROI-level significant lead SNPs between MUSE and GIANT. GIANT identified more significant lead SNPs than MUSE. The number of significant lead SNPs for ROIs of MUSE is plotted in the outer circle, whereas the inner circle illustrates the same for ROIs of GIANT. UKBB UK biobank, GIANT

genetically informed brain atlas, GWAS genome-wide association study, GM gray matter, WM white matter, SNP single nucleotide polymorphism, DEEP_WM_GM deep structure of white matter and gray matter, NONE others.

cohort. These findings highlight a high degree of concordance across the three sets of β values, supporting the robustness of our GWAS results.

Dissect the enhanced polygenicity for GIANT. To the source of increased polygenicity detected in the GIANT atlas, we aligned regions delineated by the GIANT atlas with those defined by the MUSE atlas, based on the most overlap voxels between regions specified by the two atlases. We conducted GWAS for both sets of regions using three cohorts: the UKBB white British discovery cohort, the UKBB replication cohort, and the ADNI replication cohort. Then, we integrated the GWAS summary statistics from these cohorts through meta-analysis. As a result, we observed enhanced discovery power for GIANT in most paired regions. We

now consider the left central operculum region as an example (Fig. 5). In contrast to MUSE, the GIANT atlas integrated some voxels from the left anterior and posterior insula into the creation of this brain region, leveraging both heritability information and spatial proximity. This redefinition by GIANT led to several GWAS loci, which had not achieved genome-wide significance threshold in GWAS using the MUSE atlas, reaching significance (highlighted by dashed red circles). Moreover, loci identified as significant by MUSE exhibited even more significant p-values after reclassification by GIANT. Thus, by consolidating spatially proximate brain voxels with similar heritability, GIANT not only achieved more significant p-values than MUSE but also enhances polygenicity. This demonstrates GIANT's capacity to significantly enhance discovery power for brain imaging genomics.

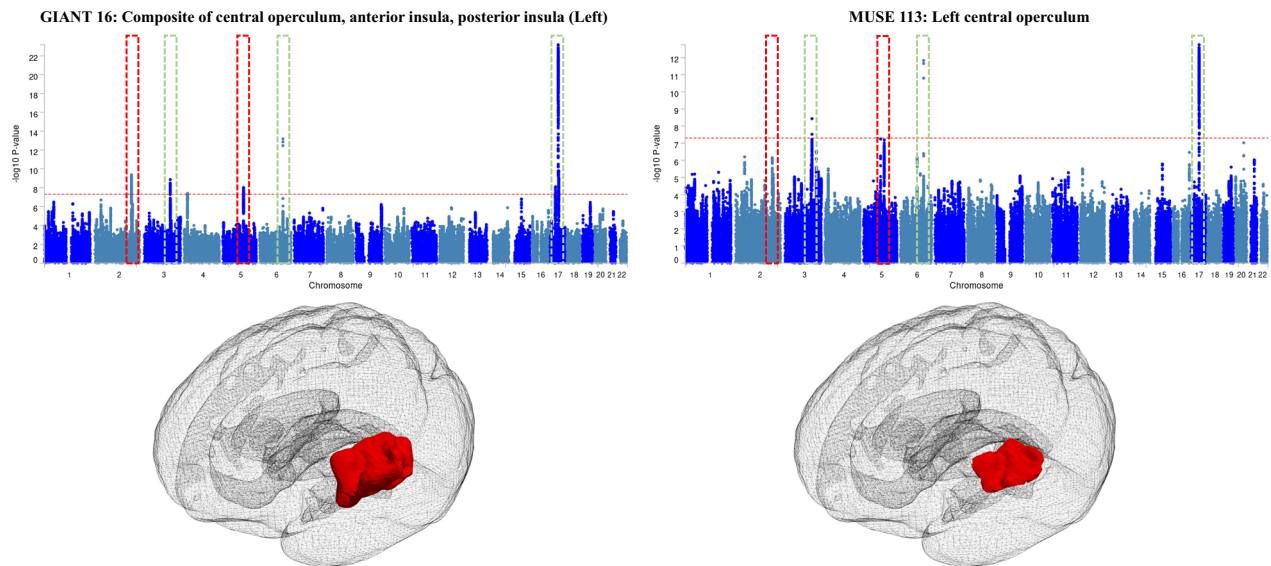


Fig. 5 | Dissect the enhanced polygenicity for GIANT. GIANT defined brain regions yield enhanced discovery power compared to MUSE. In this example, we present the GWAS results for left central operculum (MUSE 113) and left composite of central operculum, anterior insula, and posterior insula regions (GIANT 16). We observed that some GWAS loci were detected to be significant in GIANT 16 that did

not reach the significant threshold in MUSE 113 (circled by dashed red rectangles) and some of the GWAS loci yield more significant p values in GIANT 16 than in MUSE 113 (circled by dashed green rectangles). GIANT Genetically informed brain atlas, MUSE ensemble multi-atlas, GWAS genome-wide association study.

GIANT atlas polygenic risk score explains more brain volumetric variation than traditional neuroanatomical brain atlases. We conducted a systematic evaluation of the region-level polygenic risk score (PRS) on the UKBB and ADNI imaging-genomics cohorts. Our analyses of region-level PRS were based on the GWAS summary statistics estimated using the UKBB white British discovery cohort. We estimated the PRS for regional gray matter and white matter densities across the GIANT atlas, MUSE atlas, and the Watershed-based atlas (Supplementary Data 5). Consequently, for the UKBB non-white-British imaging genetics cohort, the average coefficient of determination (R^2) for PRS derived from the GIANT atlas defined regional brain gray matter and white matter densities was 3.74%, compared to 2.64% for MUSE and 3.13% for the Watershed-based atlas. Our PRS results suggest that a significantly larger proportion of brain volumetric variation is explained by the regional PRS for the GIANT atlas than by the MUSE-defined ROIs (one-sided Wilcoxon rank sum test p -value = 1.5×10^{-4}). Similarly, a significantly larger proportion of brain volumetric variation is accounted for by the regional PRS for the GIANT atlas than for the Watershed-defined ROIs (one-sided Wilcoxon rank sum test p -value = 0.028). This pattern was also observed in the ADNI cohort, where the average R^2 for PRS derived from GIANT atlas was 4.01%, compared to 2.73% for MUSE, and 3.30% for Watershed. In ADNI, our PRS results indicate that a significantly larger proportion of brain volumetric variation is explained by the regional PRS for the GIANT atlas than for the MUSE-defined ROIs (one-sided Wilcoxon rank sum test p -value = 1.7×10^{-5}) and the Watershed-defined ROIs (one-sided Wilcoxon rank sum test p -value = 2.6×10^{-3}). Additionally, we compared the R^2 derived by GIANT in UKBB non-white-British cohort with the one for regional brain volume imaging derived endophenotypes derived by Yang et al.³⁹, which indicated an average R^2 of 1.13%. Our GIANT-derived regional brain volume imaging-derived endophenotypes presented a much higher R^2 , with a one-sided Wilcoxon rank sum test p -value < 2.2×10^{-16} . In summary, our results suggest that the PRS for GIANT atlas captures a larger proportion of the variance in brain volumetric measures than the traditional neuroanatomical ROIs such as MUSE.

Genetic architecture of GIANT

We conducted a comprehensive assessment of the genetic architecture of GIANT through a multi-faceted approach: (1) investigating the genetic underpinnings of each GIANT region, (2) annotating the function of significant SNP variants, (3) examining the associations between GIANT regions and other phenotypic traits, (4) comparing pairwise regional genetic and phenotypic correlations, and (5) interpreting genetic determinants of GIANT regions.

Genetic underpinnings of GIANT. We first presented a thorough genetic analysis of GIANT using the UKBB white British discovery cohort, the UKBB non-white-British replication cohort, and the ADNI replication cohort. We applied the random effect model of METAL (version released on 2020-05-05) software⁴⁰ to the GWAS summary statistics we derived using the above three cohorts. We identified 773 significant region-lead-SNP associations (p -value < 5×10^{-8}) (Fig. 6). Specifically, we found 472 unique lead SNPs located within 386 genome loci significantly associated with 50 GIANT regions. From the imaging perspective, the cerebellum structure has the most lead SNP association signal density (31.20 associations per region) whereas the occipital structure has the least lead SNP association signal density (6.00 associations per region). The GIANT region 9 (a composite of cerebellum exterior and cerebellar vermal lobules) has the most significant association signals and the GIANT region 32 (a composite of right middle frontal gyrus and right superior frontal gyrus) shows no significant associations. From the genomics perspective, most of our lead SNPs are located in the noncoding regions mapped by the FUMA GWAS. In detail, among the 472 lead SNPs, 235 SNPs are located in the intronic regions, and 134 SNPs are located in intergenic regions. The lead SNP rs1935952:G > C (chr6:108998905:G > C in hg19) has the most associations (with 17 different GIANT regions). We further plot the functional annotation distribution of significant brain-region-lead-SNP associations (Fig. 7 b). The intronic and intergenic regions of the genome are associated with the most GIANT regions ($N = 48$ for intronic regions and $N = 45$ for intergenic regions). The exonic non-coding RNA ($N = 3$) and splicing ($N = 1$) have the least significantly associated GIANT regions. Our findings deepen the understanding of

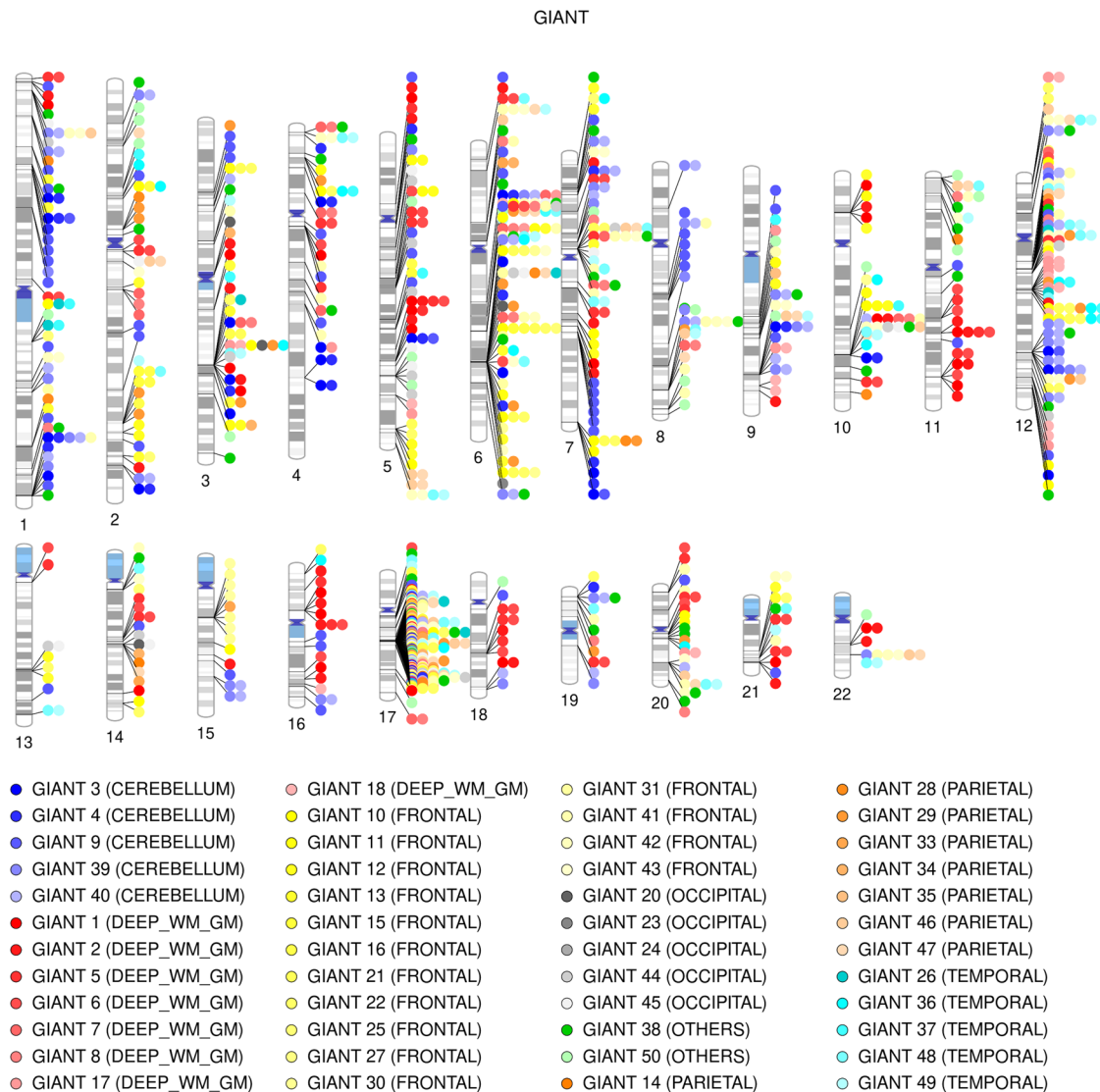


Fig. 6 | Genetic underpinnings of GIANT. GIANT can improve the discovery power of imaging-genomics studies. We provide a landscape of the genetic determinants of GIANT. We found 472 lead SNPs within 386 genome loci significantly associated with 50 GIANT regions. Our findings can deepen the understanding of the genetic

architecture of GIANT and may shed light on the potential mechanisms underlying GIANT. GIANT genetically informed brain atlas, DEEP_WM_GM deep structure of white matter and gray matter, NONE others.

the genetic architecture of GIANT and highlight the importance of non-coding SNP variants in brain structure and function. The imaging-genomics GWAS for both neuroanatomical brain atlas MUSE and genetically informed brain atlas GIANT indicated that chromosome 17 had the largest chromosome length-weighted independent SNP associations.

Functional assessment of genetic variants in GIANT. To better understand the genetic underpinnings of GIANT, we assessed the functions of significant GWAS lead SNPs. By integrating 63 functional annotations^{41–43}, we identified 56 lead SNPs with combined annotation-dependent depletion scores (CADD)⁴¹ >12.37, which suggests they are likely deleterious. Notably, the GIANT regions 18 (composite of left amygdala, left hippocampus, and left parahippocampal gyrus), region 2 (left caudate), and region 9 (cerebellar vermal lobules) had the most deleterious lead SNPs, with a total of 6 deleterious lead SNP GWAS signals. Using RegulomeDB⁴⁴, we assessed the regulatory functions of the lead SNPs by integrating eQTLs and chromatin marks^{42,43}. Our analysis identified 384 lead SNPs with regulatory annotations. Among these, SNP rs12928404:T>C and rs11022131:C>G had the most

regulatory annotations. Our findings suggest that both variants have significant associations with gene expression levels, chromatin accessibility, and direct effects on transcription factor (TF) binding. Moreover, we found direct evidence of binding of the variant through ChIP-seq and DNase with either a matched positional weight matrix (PWM)^{45–51} or a DNase footprint^{52,53}. The variant rs12928404:T>C and rs11022131:C>G was associated with right caudate, calcarine cortex, cuneus, and precuneus, in brain subcortical, occipital, and parietal regions. Overall, our functional assessment of genetic variants sheds light on the potential mechanisms underlying GIANT and may help identify new therapeutic targets for neurological disorders.

Associations of GIANT regions with other phenotypic traits. To explore the relationships between GIANT regions and other phenotypic traits, we conducted enrichment analyses using the NHGRI-EBI GWAS Catalog database v1.0.3.1⁵⁴ through FUMA (Fig. 7a). Specifically, we assessed the traits in the NHGRI-EBI GWAS Catalog with significant GWAS signals overlapped with our imaging-genomics findings (Method 9). Through our analyses, we found that brain measurement, handedness, BMI-adjusted waist-hip ratio, total cortical area

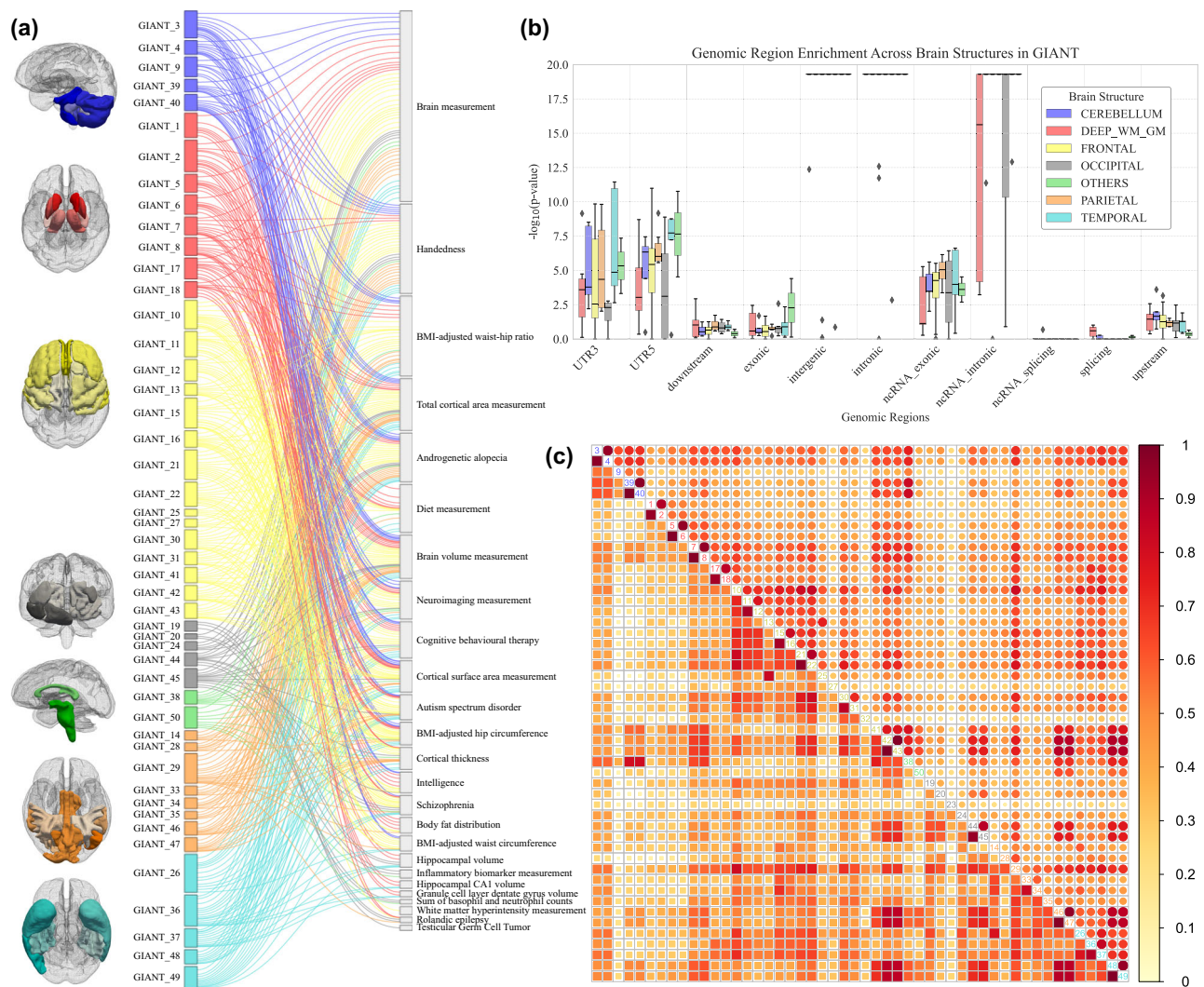


Fig. 7 | Genetic architecture of GIANT. We conducted a comprehensive assessment of the genetic architecture of GIANT through a multifaceted approach. **a** We investigated the relationships between GIANT regions and other phenotypic traits. We reported only those traits in the NHGRI-EBI GWAS Catalog that have significant overlapping GWAS signals. The GIANT regions are color-coded based on their sub-structures, and each region and trait are labeled by the proportion of their shared significant GWAS hits out of the total number of significant GWAS signals. **b** We annotated the functional significance of SNP variants. We counted the number of regions with significant GWAS signals that were functionally annotated in various

genome regions. The GIANT regions were color-coded based on their sub-structures. **c** We conducted a comparison of pairwise regional genetic and phenotypic correlations, with the genetic correlations presented in the lower-left triangular regions and the phenotypic correlations located in the upper-right regions. The GIANT regions were grouped and color-coded based on their sub-structures. GIANT genetically informed brain atlas, DEEP_WM_GM deep structure of white matter and gray matter, NONE others, GWAS genome-wide association study, UTR5 5' untranslated region, UTR3 3' untranslated region, upstream upstream regulatory region, downstream downstream regulatory region, ncRNA non-coding RNA.

measurement, androgenetic alopecia, diet measurement, and brain volume measurement had the highest proportion ($\geq 5\%$) of overlapping significant GWAS signals with our imaging-genomics associations across the entire brain. In addition, we observed a moderate amount ($\geq 3\%$) of associated SNPs for neuroimaging measurement, cognitive behavioral therapy, cortical surface area measurement, and autism spectrum disorder. These findings suggested potential links between GIANT regions and various phenotypic traits.

Pairwise genetic and phenotypic correlations among GIANT Regions. We analyzed the pairwise genetic correlations and pairwise phenotypic correlations among the GIANT regions (Fig. 7c). Using the meta-analysis GWAS summary statistics derived from UKBB white British discovery cohort, UKBB nonwhite British cohort, and ADNI cohort, we estimated the genetic correlations using LDAC³⁵ (Method 10 and Supplementary Data 6). We plot the genetic correlation in the lower triangular area of Fig. 7c. For comparison purpose, we estimated

the phenotypic correlations using Pearson correlations and plotted them in the upper triangular region of Fig. 7c. Our results show that the genetic correlations among GIANT regions are significantly lower than their phenotypic correlations (Wilcoxon rank sum test $p\text{-value} = 2.05 \times 10^{-14}$), indicating that GIANT-defined brain regions are genetically distinct from each other. This result matches our expectations of GIANT.

Mapping SNPs to genes in GIANT regions. To identify potential target genes for the significant SNPs in GIANT regions, we employed three mapping approaches: positional mapping, eQTL mapping, and chromatin interaction mapping using FUMA GWAS^{42,43}.

Using positional mapping, we linked SNPs to genes within a 10-kilobase distance. The GIANT region 38 (composite of brain stem and cerebellum white matter) had the most associations with a total of 258 mapped genes. Gene *ARL17B*, *KANSL1-AS1*, *LRRC37A*, *MAPT*, *NSF*, *RNU7*, *RP11*, *RPS7P11*, and *STH* had the highest number of associations. The

GIANT region 24, the composite region of left middle, superior, and occipital gyri, at the occipital structure, has the most averaged number of SNPs mapped genes (166.2 SNPs per gene). We also assessed the intolerance of the mapped genes to loss-of-function mutations using the probability of loss-of-function intolerance (pLI) score from ExAC (Exome Aggregation Consortium)⁵⁶ and gnomAD (Genome Aggregation Database)⁵⁷. GIANT region 23, the composite region of the right middle and superior occipital gyrus at the occipital structure, was most resistant to such mutations, while the GIANT region 39 (right cerebellum white matter) has the most loss-of-function mutation intolerant gene.

For eQTL mapping, we mapped those significant SNPs to genes using eQTL summary statistics from 13 brain tissues extracted from GTEx project v8⁵⁸, cis- and trans-eQTLs from the CommonMind Consortium⁵⁹, 11 brain tissues from Brainiac of the UK Brain Expression Consortium⁶⁰, and eQTL data from PsychENCODE⁶¹. The GIANT region 38 (a composite of brain stem and cerebellum white matter) has the highest number of significant eQTL-mapped genes across all aforementioned tissue types. On the other hand, gene *CRHR1-IT1* has the highest number of significant SNP-eQTL associations - it is significantly associated with multiple SNPs that are discovered to be the GWAS hits of 35 different GIANT regions.

Using chromatin interaction mapping, we identified genes using Hi-C data of the dorsolateral prefrontal cortex and hippocampus tissues in the GSE87112 dataset⁶² of the Gene Expression Omnibus database⁶³. We further annotated the enhancer and promoter regions using 12 brain tissues using Roadmap 111 epigenomes⁶⁴. There are 152 genes mapped by significant GWAS signals using 3D chromatin interactions. GIANT region 44, comprising the right white matter temporal and occipital lobe, had the most genes mapped through 3D chromatin interactions.

Discussion

In this study, we introduced an biologically interpretable three-dimensional clustering model tailored for brain parcellation, named the heritability-aware brain parcellation model. This framework simultaneously integrates SNP heritability information with spatial information from brain voxels. It can process the brain voxel-level data efficiently without necessitating extensive denoising imaging pre-processing steps, as a smoothing process is achieved by the incorporation of Ising prior. Our method achieves fast convergence (Supplementary Fig. 3). Furthermore, although originally developed to create a genetically informed brain atlas based on SNP heritability of brain volume, this framework can be applied to other applications involving clustering of three-dimensional objects while considering specific voxel attributes. For example, it is applicable to cluster the cell types using three-dimensional spatial transcriptomics data, where the transcript reads are the attribute of interest⁶⁵. Moreover, our framework is designed to accommodate various distance metrics for incorporating spatial information. An alternative to Euclidean distance, for example, could be the use of voxel-level brain functional connectomes⁶⁶ as similarity matrices in constructing the genetically informed brain atlas. The creation of such a multi-modal, genetically informed brain atlas can significantly boost the discovery power in brain imaging genomics studies.

In our study, we introduced GIANT, a genetically informed brain atlas for brain imaging-genomics studies. GIANT is generated by integrating the SNP heritability of brain volumetric endophenotype and spatial proximity, making it suitable for brain imaging-genomics studies. We established GIANT using a subset of randomly selected 5,000 subjects from the UKBB white British imaging genomics cohort, totaling 35,181 individuals; and we validated the atlas through a comprehensive multi-perspective approach. We assessed GIANT's neuroanatomical validity in three distinct ways. Specifically, we examined the concordance between GIANT and a genetically

informed brain atlas generated from a separate, non-overlapping subset of 5,000 UKBB white British individuals. Such neuroanatomical validation experiment suggests that GIANT, though defined by a subset of the population, is representative and can be generalized across the entire white British imaging genomics cohort. Furthermore, we evaluated the stability of imaging-derived endophenotypes within GIANT, particularly focusing on regional brain gray and white matter densities. The high test-retest reliability and greater homogeneity affirm GIANT's stability as a neuroanatomical brain atlas, demonstrating GIANT's generalizability to various cohorts with different brain conditions.

We assessed the capability of GIANT to capture known architectonic boundaries. We compared the alignment of architectonic boundaries of GIANT and other widely used brain atlases, including AAL³⁰, AICHA³¹, CPAC200³², Desikan⁴⁸, Schaefer²⁸, HammerSmith³³, Talairach³⁴, and Yeo³⁵ atlases. GIANT atlas exhibited moderate alignment with the AAL, Desikan, HammerSmith, and Talairach atlases. These atlases, which are delineated based on major sulci and gyri across diverse age groups²⁹, are pivotal for understanding brain neuroanatomical structures. The observed moderate alignment with these atlases demonstrates GIANT's capability in capturing the architectonic boundaries that define brain neuroanatomical structures. In addition, GIANT showed moderate alignment with the AICHA, CPAC200, and Schaefer atlases, which are outlined based on brain resting-state networks²⁹, suggesting GIANT's capacity to reflect the architectonic boundaries of brain functional structures to a considerable extent. In contrast, GIANT demonstrated lower concordance with the Yeo atlases, indicating limitations in capturing networks of functionally coupled regions across the cerebral cortex³⁵. Nevertheless, GIANT's moderate alignment with many neuroanatomical brain atlases, without achieving very high concordance, illustrates it can retain fundamental anatomical and functional brain knowledge, even as it aims to advance brain imaging genomics studies.

Pioneering work in the development of a genetically informed brain atlas was initiated by Dr. Chi-hua Chen and their colleagues¹⁹. In their study, they delineated the human brain's cortical area into 12 regions of interest using a hierarchical clustering strategy based on genetic correlations, derived from 406 twins. This genetically informed brain cortical atlas is able to identify more significant genetic loci¹¹. To serve as a complement to the atlas created by Chen et al., our GIANT offers distinct perspectives: it is based on gray matter and white matter densities, rather than cortical surface area. Additionally, to capture the genetic heterogeneity at the finest resolution, GIANT is defined at the brain voxel level, incorporating SNP heritability information. This is complementary to the atlas generated by Chen et al. where pre-defined anatomical brain regions were grouped based on genetic correlations. Moreover, our GIANT is derived from a cohort of 5,000 White British individuals in comparison of 406 twins in Chen et al. Through extensive validation, GIANT has been assessed as neuroanatomically valid and demonstrates broad generalizability across populations with diverse genetic ancestries and various brain conditions.

When compared to the MUSE atlas, the regional brain volumetric measures defined by GIANT exhibit significantly enhanced voxel-level SNP heritability contrasts, increased estimates of regional SNP heritability, improved polygenicity, and a larger variation of phenotype explained by PRS. Specifically, to avoid the potential circularity concerns in our genetics analysis, we excluded the 5000 individuals from the UKBB white-British cohort who were randomly selected for the generation of the atlas. To evaluate the generalizability of the GIANT atlas, we conducted two independent replication studies using the UKBB non-white-British cohort and the ADNI replication cohort. The genetics analysis results from these cohorts confirmed the GIANT atlas's enhanced discovery power in brain imaging genomics studies, demonstrating its generalizability across diverse population ancestries and various brain conditions. In summary, GIANT increases the power

to dissect the genetic underpinnings of brain neuroimaging studies in cohorts with different genetic ancestries and brain conditions.

We present an in-depth evaluation of the genetic architecture of GIANT through a comprehensive, multi-angle approach. Our study identified 773 significant region-genome-locus associations that shed light on the genetic underpinnings of GIANT. We dissect the genetic determinants of GIANT regions, functionally annotating their underlying genetic variants from multiple resources. We fully explore the genetic relationships between GIANT regions and various phenotypic traits, revealing that our GIANT regions share a multitude of genetic determinants with several brain-related traits and disorders. This suggests potential genetic associations between GIANT regions and a range of phenotypic traits. Through the comparison of pairwise genetic and phenotypic correlations, GIANT reveals significantly lower genetic correlations than phenotypic correlations, indicating the ability of our algorithm to group the genetically homogeneous brain voxels into regions. These findings not only support the anatomical validity of GIANT but also align the intuitions behind the formation of GIANT. Additionally, we identify potential target genes for the significant SNPs in the GWAS of GIANT regions by employing positional, eQTL, and chromatin interaction mapping approaches, followed by their regulatory annotations. Our findings deepen the understanding of the genetic architecture of GIANT and may shed light on the potential mechanisms underlying GIANT, providing new candidate therapeutic targets for brain disorders.

Our investigation acknowledges several limitations. First, the performance of GIANT might be undermined by potential imaging artifacts, as changes in MRI hardware and software can introduce unwanted variability into the downstream genetic analyses, particularly when integrating data from multiple sites and phases of neuroimaging studies. Second, errors in imaging segmentation may lead to imprecise voxel-level heritability estimations, especially at the boundary of gray matter and white matter and within the cortical areas of the brain. Third, in addition to the gray matter and white matter densities, there are different types of IDPs worthy of investigation, such as brain cortical surface area and cortical thickness, where previous studies have shown their distinct genetic influences⁶⁷. Last, as in many other genetic studies, GIANT's development relies on data predominantly from individuals of European ancestry. As more genetically diverse datasets become available in the future, there exists the opportunity to retrain our model. This advancement will allow GIANT to encompass populations with genetic ancestries that are presently underrepresented, thereby enhancing its applicability.

Methods

Method 1: Genetically informed brain parcellation via three-dimensional Gaussian mixture model

A Bayesian model for heritability-aware brain parcellation. We developed a flexible Bayesian model for learning the heritability-aware brain parcellation which models the voxel-level heritability using a Gaussian mixture model with Ising prior to incorporate the spatial information. Similar modeling approaches have been widely applied to different research areas, including microarray image analyses, imaging processing, and spatial transcriptomics^{68–70}. In our heritability-aware brain parcellation framework, we extend the two-dimensional model to three dimensions and apply the framework to coordinate-based brain neuroimaging data. Our Bayesian model encourages grouping spatially connected brain voxels into regions to achieve enhanced discovery power for brain imaging genomics studies.

Our data consists of a three-dimensional matrix that describes the estimated heritability for volumetric changes of brain voxels and a binary brain mask that indicates brain structures. We model the voxel-level volumetric heritability as a three-dimensional matrix $Y = \{y_{i,j,l}\}_{1 \leq i \leq N_l, 1 \leq j \leq N_j, 1 \leq l \leq N_l}$ where N_l, N_j, N_l represent the number

of voxels in each dimension. To focus only on the voxels of interest, our model applies the mask and considers only the voxels in a certain user-defined region of interest. For example, these regions can be a certain type of tissue such as gray matter or white matter. Or it can be certain brain structures such as brain cortical or subcortical regions. Since our model only considers part of the 3D matrix by masking, from now on, we will use a linear index $i \in \{1, \dots, N\}$ to replace the coordinate-based index system, where N denotes the total number of voxels within the masked regions. Each voxel-level heritability $y_i \in [0, 1]$ in the region of interests is modeled by the Gaussian distribution

$$y_i | z_i = k \sim N(\mu_k, \sigma^2) \quad (1)$$

where $z_i \in \{1, \dots, q\}$ denotes the latent region to which the voxel i belongs; $\mu_k \in \mathbb{R}$ denotes the mean SNP heritability for the region k ; and $\sigma^2 \in \mathbb{R}$ is the within-region heritability variance. Given that previous studies showed that the variable variance needs strong priors for parameter estimation⁷⁰, we assume a fixed variance across all regions. The number of regions, q , is determined by heritability information.

We assign priors to the mean and variance parameters, μ_k and σ^2 , as follows:

$$\mu_k \sim N(\mu_0, \sigma_0^2) \quad (2)$$

$$\sigma^2 \sim \text{InvGamma}(\alpha, \beta) \quad (3)$$

where μ_0 and σ_0^2 are hyperparameters that control the mean and variance of μ_k . In practice, we set μ_0 as the mean of all voxel-level heritability. For the choice of σ_0^2 , we first initialize μ_k to be the within-region mean heritability according to the input parcellation initialization (i.e., the masked anatomically defined atlas). Then, we set the variance of μ_k to be σ_0^2 . Moreover, α and β are two hyperparameters for the variance parameter σ^2 . By default, we set $\alpha = 1$ and $\beta = 0.01$ to provide a weak prior for σ^2 .

To incorporate the spatial information, we assign the Ising prior to the latent region parameter z_i :

$$p(z_i) = \exp\left(\frac{\gamma}{|\langle i, j \rangle|} \times 2 \sum_{\langle i, j \rangle} I(z_i = z_j)\right) \quad (4)$$

Here, $\langle i, j \rangle$ denotes all voxels j that are neighbors of voxel i . In our framework, the neighborhood information is modeled either by coordinate-based Euclidean distance or by coordinate-based step distance. Specifically, by specifying a hyperparameter r , the coordinate-based Euclidean distance definition will treat all voxels j within the 3D sphere centered at the voxel i with radius r as the neighbors of voxel i ; the coordinate-based step definition will treat all voxels j that can be reached by “walking” r steps from the center voxel i as the neighbors of voxel i . The $I(\cdot)$ represents the indicator function. Intuitively, the Ising prior assigns a higher probability for a voxel i belonging to a specific region k if more of its neighbors j belong to the region k . The smoothing hyperparameter γ controls the weight of spatial information. Larger γ means a higher probability that the center voxel $v_{i,j,l}$ will belong to the regions that most of its neighbors belong to. Inappropriate large γ will encourage all voxels to belong to the same region; inappropriate small γ will encourage more densely distributed region assignments. In practice, the γ parameter needs to be tuned. An existing anatomical brain atlas is not required for initialization, but it is recommended to improve the convergence of the MCMC algorithm.

We generate posterior samples for our Bayesian model using an efficient MCMC algorithm, Gibbs sampling. We tune the

hyperparameters using the CH score³⁶ to quantify the within-region and between-region heritability variation difference. The detailed derivation and hyperparameter tuning strategy of our Gibbs sampling algorithm are provided in the Supplementary Methods 2 and 3.

Evaluation of the heritability-aware brain parcellation model through simulations. We conducted two simulations to evaluate the performance of the heritability-aware brain parcellation model. Our simulations were designed to randomly generate heritability for each voxel based on the existing brain atlas. We then applied our heritability-aware brain parcellation algorithm to the generated heritability brain maps and attempted to recover the original atlases. To compare the effectiveness of our approach, we also applied the Watershed algorithm⁷¹ to the same data (Supplementary Methods 5). A detailed description of our simulations is provided in Supplementary Methods 4.

To generate voxel-level heritability, we used the AAL atlas with 116 ROIs and MUSE atlas with 145 ROIs (gray matter, white matter, and ventricular regions) in our two simulation studies. We evaluated the performance of the best-tuned results by comparing the resulting parcellation with the ground truth parcellation using several metrics, including the adjusted Rand index (ARI)^{72–74}, adjusted mutual information based score (AMI)⁷⁵, homogeneity score, completeness score, V-measure score (V-M), and Fowlkes Mallows score (FM) (Supplementary Methods 6).

To determine the optimal smoothing parameter γ for our heritability-aware brain parcellation algorithm, we conducted a hyperparameter tuning with $\gamma \in \{0.5, 1, 1.5, \dots, 39, 39.5, 40\}$ using the CH score as our evaluation metric. Our simulation results demonstrate that our heritability-aware brain parcellation algorithm achieves the best recovery of the original ground truth atlas (Supplementary Fig. 2) and outperforms the Watershed algorithm⁷¹ in terms of ARI, AMI, V-M, and FM scores (Supplementary Data 7).

Method 2: Study Populations

The GIANT study explores individual-level genotyping and T1-weighted MRI data obtained from the UK Biobank⁷⁶. The UK Biobank is a population-based registry that recruited 500,000 UK adults with ethical approval from the National Research Ethics Service Committee North West–Haydock (reference 11/NW/0382). All participants provided informed consent and were aged approximately between 40 and 69 years old at enrollment. Participants completed questionnaires, physical assessments, and provided socio-demographic, cognitive, and medical data. In 2014, a subset of the sample underwent MRI and the data used in our study were acquired between 2014 and 2019. The T1-weighted MRI images were acquired using a 3T Siemens Skyra machine (MPRAGE) with an image resolution of $1 \times 1 \times 1$ mm and a time to echo (TE) of 2000 ms⁷⁷. Further information about the image protocols can be found at http://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/brain_mri.pdf.

We constructed an imaging-genetics cohort from UKBB by including all subjects with both T1-weighted MRI data and imputed genotyping data. The UKBB imaging-genetics cohort comprises 38,290 subjects (20,199 females and 18,091 males), including 35,181 white British individuals (18,503 females and 16,678 males) and 3,109 nonwhite British subjects (1,696 females and 1,413 males).

We validate GIANT using the ADNI data. The individual-level genotyping and T1-weighted MRI data used in the preparation of this article were obtained from the ADNI database (<http://adni.loni.usc.edu>)^{78–82}. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive

impairment (MCI) and early AD. Up-to-date information about the ADNI is available at www.adni-info.org.

We created an imaging-genetics cohort from ADNI by preserving all subjects with both T1-weighted MRI data and imputed genotyping data. The ADNI imaging-genetics cohort comprises 1809 subjects (989 females and 820 males) including 678 cases (mild cognitive impairment patients or Alzheimer's disease patients) and 1131 controls.

In our study, we constructed the genetically informed brain atlas using randomly selected 5000 white British UKBB individuals. We used 30,181 white British UKBB individuals as the discovery cohort for the subsequent neuroanatomical validations and imaging genomics studies. The subjects in the discovery cohort have no overlap with the cohort we used for atlas generation. We used two independent replication cohorts for the validations. (1) We used 3,109 non-white-British UKBB subjects as the first replication cohort to assess the generalizability and robustness of GIANT in populations with different genetic ancestries; (2) We used 1,809 ADNI subjects as the second replication cohort to assess the generalizability and robustness of GIANT in populations with different age range and brain conditions.

Method 3: Neuroimage data preprocessing

T1-weighted MRI is downloaded from the UKBB study⁷⁶ and ADNI study^{78–82}. Raw 3D T1-weighted MRIs were first quality checked (QC) for motion, image artifacts, or restricted field-of-view. Another QC was performed as follows: 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 automated procedure automatically flagged images based on outlying values of quantified metrics (i.e., ROI values), and those flagged images were re-evaluated. The quality-controlled images are first corrected for magnetic field intensity inhomogeneity⁷⁶. Voxel-wise regional volumetric maps, RAVENS, for each tissue volume⁸³ are generated by spatially aligning the skull-stripped images to a template residing in the MNI-space⁸⁴. For the conventional atlas, a multi-atlas parcellation method (MUSE)²¹ was then used to extract 139 ROIs from gray matter and white matter tissue maps. Finally, we downsampled the image from $1 \times 1 \times 1$ mm to $2 \times 2 \times 2$ mm resolution for the consideration of computational expense.

Method 4: Genotyping data preprocessing

We download the raw imputed genotyping data from UKBB⁷⁶ (UKBB Category 263) and ADNI study^{78–82}.

In UKBB, raw genetic data (Version 3) was downloaded from the UKBB website (<https://www.ukbiobank.ac.uk/enable-your-research/about-our-data/genetic-data>) in July 2021. The imputation was performed by the original UKBB genetics study⁴. In our QC steps, we filtered out the (1) multiallelic variants, (2) variants with missing call rates greater than 0.03, (3) variants with minor allele frequencies smaller than 0.01, 4) variants with Hardy-Weinberg equilibrium exact test p-value below the $1e-10$ threshold. Next, we filtered out the subjects (1) with missing call rate exceeding 0.03, (2) with heterozygosity rate outside 5 standard deviations of the population heterozygosity rate. Finally, we match the QCed imputed genotyping cohort with the QCed imaging cohort. All the QC steps are done using the PLINKv2.0⁸⁵ and R. After the harmonization of the QCed imputed genotyping data from both UKBB and ADNI, the imputed genetic data comprises 6,965,659 SNPs and 38,290 subjects, which were used in our GWAS analysis. We further derived the first 10 genetic principle components (PCs) using the SmartPCA from the EIGENSOFT^{86–89}.

In ADNI, we downloaded genotyping data from ADNI 1, GO, 2, and 3 studies. We aligned and integrated the downloaded data using the Homo sapiens (human) genome assembly NCBI37 (hg19) genome builder. We performed the strand alignment according to 1000 Genome phase 3⁹⁰ using McCarthy Group Tools (<https://www.well.ox.ac.uk/~wrayner/tools/>). We imputed the genotyping data using the

Michigan Imputation Server⁹¹ with 1000 Genome phase 3 reference panel of European ancestry. We annotated our imputed genotyping data using ANNOVAR⁹². After alignment and imputation, we performed the quality control (QC) using the following criteria: 1) genotyping call rate greater than 98%, 2) minor allele frequency greater than 0.1%, 3) Hardy-Weinberg Equilibrium greater than 1e-6, 4) missingness per individual less than 5%. All the QC and recoding were performed using PLINK1.9⁸⁵. After data preprocessing, we matched the common subjects in genotyping, neuroimaging, and demographic data. After the harmonization of the QCed imputed genotyping data from both UKBB and ADNI, our QCed ADNI imputed genetic data comprises 6,965,659 SNPs and 1809 subjects.

Method 5: Heritability-aware brain atlas framework

We separately apply the heritability-aware brain parcellation algorithm (Method 1) to gray matter and white matter using 5000 randomly selected individuals from the UKBB imaging-genomics cohort, with the MUSE atlas for initialization (Supplementary Methods 1). We tune the region-smoothing hyperparameter γ on each tissue type, ranging from 0.5 to 40 in increments of 0.5. The best-tuned hyperparameters are selected based on the highest CH score³⁶, which are found to be $\gamma = 6$ for gray matter parcellation and $\gamma = 14$ for white matter parcellation. We combine the best-tuned gray matter and white matter parcellations to create the GIANT atlas.

Method 6: A brain atlas annotation strategy

We annotate the GIANT using the existing anatomically defined brain atlas as a reference atlas. To do this, we count the number of voxels belonging to different ROIs in the reference atlas for each ROI in the GIANT. We then name each ROI in the GIANT based on the proportion of voxels that belong to different ROIs in the reference atlas. To be specific, we calculate the percentage of voxels belonging to each ROI in the reference atlas over the total number of voxels in the ROI being annotated and use this value to name the ROIs in the GIANT.

Method 7: Neuroanatomical Validation

Test-retest reliability evaluation. We conducted the test-retest reliability assessment using ICC on the longitudinal cohorts of both UKBB and ADNI, comprising 1356 and 1917 subjects respectively. For each individual, regional-level gray matter and white matter densities were derived from both the initial and final visits. We calculated six different ICC coefficients as defined by Shrout and Fleiss (1979)²⁵ for each brain region. All calculations were performed using the “psych” package in R⁹³.

Homogeneity evaluation. We evaluated the homogeneity of gray and white matter densities within GIANT across three population cohorts: the UKBB White-British discovery cohort, the UKBB non-White-British replication cohort, and the ADNI replication cohort. Following the methodology of Schaefer et al.²⁸, we measured homogeneity by calculating the weighted standard deviation of regional densities using the formula:

$$\frac{\sum_{k=1}^K sd_k |k|}{\sum_{k=1}^K |k|} \quad (5)$$

where sd_k is the standard deviation of gray matter or white matter densities for the region k , and $|k|$ is the number of voxels in the region k . Lower standard deviations indicate greater homogeneity within each brain region.

Architectonic evaluation. To quantify the alignment of GIANT's architectonic boundaries with those of reference atlases, we employed the AMI score. AMI measures the similarity between two labeled sets, indicating how well a specific voxel can be identified as belonging to a

particular region based on another region. AMI is not dependent on a region's label, and is computed as follows:

$$H(A) = - \sum_{k=1}^K P_A(k) \cdot \log[P_A(k)] \quad (6)$$

$$MI(A, B) = \sum_{a \in A, b \in B} P_{A,B}(a, b) \cdot \log \left(\frac{P_{A,B}(a, b)}{P_A(a)P_B(b)} \right) \quad (7)$$

$$AMI(A, B) = \frac{MI(A, B) - E[MI(A, B)]}{\max(H(A), H(B)) - E[MI(A, B)]} \quad (8)$$

where $H(A)$ represents the entropy for the partitioning A ; $P_A(i)$ denotes the probability that a voxel randomly selected from the set A will belong to the brain region k ; $P_{A,B}(a, b)$ is the probability that a point belongs to both brain region $a \in A$ and $b \in B$; and $E[\cdot]$ means the expectation operator. Higher AMI scores indicate greater similarity between the two brain atlases.

Method 8: Genome-wide association analysis

Genome-wide association analysis with individual-level data. We conducted GWAS on 50 ROIs defined by GIANT, 139 ROIs defined by the MUSE atlas, and 100 ROIs defined by Watershed-based atlas. Each ROI represents a brain region-level quantitative trait measuring brain gray/white matter densities. The GWAS analyses were performed using imputed genotyping data from the UKBB white-British imaging-genomics discovery cohort (30,181 subjects), the UKBB non-white-British replication cohort (3109 subjects), and the ADNI imaging-genomics cohort (1809 subjects). We fit a linear mixed effect regression model for each ROI-SNP pair by treating imaging volumetric quantitative trait as the response variable and common-variant autosomal individual SNP as the independent variable. Our model was adjusted for age, sex, first 10 principal components, and AD-by-proxy/AD as covariates. The genome-wide significant threshold was set as 5×10^{-8} . All the GWAS were performed using Scalable and Accurate Implementation of Generalized mixed model (SAIGE)⁹⁴. We performed post-GWAS analysis using functional mapping and annotation^{42,43}. AD-by-proxy was based on parental diagnosis and exhibited a strong genetic correlation with AD⁹⁵.

To evaluate the robustness of our GWAS results, we conducted sensitivity analyses on the β coefficients of significant GWAS signals derived from the white British discovery cohort. These analyses included: (1) comparing the Pearson correlation of β coefficients between the discovery cohort and replication cohorts, and (2) assessing the concordance of β coefficient signs between the discovery and replication cohorts. The concordance was measured as the proportion of matching signs relative to the total number of significant GWAS signals for each phenotype. We then reported the weighted average Pearson correlations and the proportion of matching signs across all GIANT ROIs, weighted by the number of significant GWAS signals identified in each ROI.

Genome-wide association meta-analysis with GWAS summary statistics. We conducted a GWAS respectively in the UKBB white-British imaging genomics discovery cohort, the UKBB non-white-British replication cohort, and the ADNI replication cohort. We use METAL (version released on 2020-05-05) software⁴⁰ to combine p-values across the three GWAS summary statistics taking into account the sample size and effect directions of each study. To track the effect allele frequency across different studies, we reported the mean, minimum, and maximum effect allele frequency to monitor the inconsistent naming of reference alleles across different studies. Our meta-analysis was performed using the random effect model. We

left all the other parameters we did not mention above as default parameters.

Post-GWAS study with Functional Mapping and Annotation of GWAS (FUMA). For each ROI volumetric imaging QT, we performed the post-GWAS analysis using FUMA^{42,43}. FUMA is a web-based platform using information from multiple biological resources to facilitate functional annotation of GWAS results. We used the FUMA analysis protocol from Wen et al.¹⁵. We constructed LD blocks by tagging all variants with minor allele frequency greater than or equal to 0.0005 and with at least one of the independent significant variants. Of note, the LD blocks are constructed from the 1000 Genomes as reference panels, which may not be overlapped with the variants in the current study. Finally, FUMA merges the LD blocks of independent significant variants into a single genomic locus if they are within 250 kilobases from the closest boundary variants of LD blocks. We used the default parameters settings on FUMA online platform for the other unmentioned parameters.

Method 9: Associations of GIANT regions with other phenotypic traits

We evaluated the associations of GIANT regions significantly enriched for other phenotypic traits using the NHGRI-EBI GWAS Catalog database v1.0.3.1⁵⁴ through FUMA. We identified phenotypic associations by examining traits from the NHGRI-EBI GWAS Catalog that share significant genetic signals with our imaging-genomics findings. Our analysis began with a GENE2FUNC analysis in FUMA, where we input the significant genes identified through FUMA's SNP2GENE analysis of GWAS summary statistics. This analysis highlighted phenotypes significantly enriched by brain volumetric traits defined by the GIANT atlas. To ensure robustness, we filtered out phenotypes with fewer than 15 overlapping genetic signals between the GWAS Catalog and the GIANT data. Next, we categorized the phenotypic traits by phenotype categories and calculated the number of significantly enriched traits within each category for each GIANT region, applying a false discovery rate (FDR) correction for multiple comparisons. For each region, we selected the top three phenotype categories with the highest number of significantly enriched traits. Finally, we visualized the associations between GIANT regions and phenotype categories using a Sankey diagram.

Method 10: Genetic correlation analysis

We estimate the pairwise genetic correlations for region-level brain variations defined by both GIANT and the MUSE atlas in both the UKBB and ADNI imaging-genomics cohorts. To estimate the genetic correlations, we use the GWAS summary statistics obtained from our previous GWAS analyses. We perform the analysis using LDAK⁵⁵, which extends the LD score regression model^{96–98} by assuming the LDAK model and accounting for confounding inflation that is multiplicative⁵⁵.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The individual level data from UK Biobank (UKBB) and Alzheimer's Disease Neuroimaging Initiative (ADNI) are available under restricted access. ADNI data are available at <https://adni.loni.usc.edu/data-samples/access-data/> pending application approval and compliance with the data usage agreement. Researchers can apply to use the UK Biobank resource for health-related research that is in the public interest (<https://www.ukbiobank.ac.uk/register-apply/>). The GWAS summary statistics generated in this study have been deposited in the Zenodo database under accession code <https://doi.org/10.5281/zenodo.14549178>.

The summary-level data generated in this study are provided in the Supplementary Information/Source Data file and Zenodo database. Source data are provided with this paper.

Code availability

The source code for atlas generation, the template atlas in NIFTI format, the atlas annotation, the GWAS summary statistics, and the post-GWAS FUMA analysis results are all available through GitHub (<https://github.com/JingxuanBao/GIANT>).

References

- Thompson, P. M. et al. ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. *Transl. psychiatry* **10**, 100 (2020).
- Thompson, P. M. et al. The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav.* **8**, 153–182 (2014).
- Sudlow, C. et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
- Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
- 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).
- Wang, C. et al. Phenotypic and genetic associations of quantitative magnetic susceptibility in UK Biobank brain imaging. *Nat. Neurosci.* **25**, 818–831 (2022).
- Zhao, B. et al. Common variants contribute to intrinsic human brain functional networks. *Nat. Genet.* **54**, 508–517 (2022).
- Smith, S. M. et al. An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nat. Neurosci.* **24**, 737–745 (2021).
- Mollink, J. et al. The spatial correspondence and genetic influence of interhemispheric connectivity with white matter microstructure. *Nat. Neurosci.* **22**, 809–819 (2019).
- Elliott, L. T. et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210–216 (2018).
- Makowski, C. et al. Discovery of genomic loci of the human cerebral cortex using genetically informed brain atlases. *Science* **375**, 522–528 (2022).
- Wen, J. et al. Novel genomic loci and pathways influence patterns of structural covariance in the human brain. *medRxiv*, 2022.2007.2020.22277727 (2022).
- Yang, Z. et al. A deep learning framework identifies dimensional representations of Alzheimer's Disease from brain structure. *Nat. Commun.* **12**, 7065 (2021).
- Bao, J. et al. Integrative analysis of multi-omics and imaging data with incorporation of biological information via structural Bayesian factor analysis. *Brief. Bioinforma.* **24**, bbad073 (2023).
- Wen, J. et al. Characterizing Heterogeneity in Neuroimaging, Cognition, Clinical Symptoms, and Genetics Among Patients With Late-Life Depression. *JAMA Psychiatry* **79**, 464–474 (2022).
- Shen, L. & Thompson, P. M. Brain Imaging Genomics: Integrated Analysis and Machine Learning. *Proc. IEEE* **108**, 125–162 (2020).
- Zhao, B. et al. Common genetic variation influencing human white matter microstructure. *Science* **372**, eabf3736 (2021).
- Desikan, R. S. et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).
- Chen, C.-H. et al. Hierarchical genetic organization of human cortical surface area. *Science* **335**, 1634–1636 (2012).

20. Yang, Z. et al. Gene-SGAN: discovering disease subtypes with imaging and genetic signatures via multi-view weakly-supervised deep clustering. *Nat. Commun.* **15**, 354 (2024).
21. 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).
22. Li, F. & Zhang, N. R. Bayesian variable selection in structured high-dimensional covariate spaces with applications in genomics. *J. Am. Stat. Assoc.* **105**, 1202–1214 (2010).
23. Speed, D., Holmes, J. & Balding, D. J. Evaluating and improving heritability models using summary statistics. *Nat. Genet.* **52**, 458–462 (2020).
24. Zhang, Q., Privé, F., Vilhjálmsson, B. & Speed, D. Improved genetic prediction of complex traits from individual-level data or summary statistics. *Nat. Commun.* **12**, 1–9 (2021).
25. Shrout, P. E. & Fleiss, J. L. Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* **86**, 420 (1979).
26. McGraw, K. O. & Wong, S. P. Forming inferences about some intraclass correlation coefficients. *Psychol. methods* **1**, 30 (1996).
27. Koo, T. K. & Li, M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **15**, 155–163 (2016).
28. Schaefer, A. et al. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).
29. Lawrence, R. M. et al. Standardizing human brain parcellations. *Sci. Data* **8**, 78 (2021).
30. Tzourio-Mazoyer, N. et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* **15**, 273–289 (2002).
31. Joliot, M. et al. AICHA: An atlas of intrinsic connectivity of homotopic areas. *J. Neurosci. Methods* **254**, 46–59 (2015).
32. Craddock, R. C., James, G. A., Holtzheimer III, P. E., Hu, X. P. & Mayberg, H. S. A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Hum. brain Mapp.* **33**, 1914–1928 (2012).
33. Gousias, I. S. et al. Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest. *NeuroImage* **40**, 672–684 (2008).
34. Talairach, J. & Szikla, G. Application of stereotactic concepts to the surgery of epilepsy. *Advances in Stereotactic and Functional Neurosurgery 4: Proceedings of the 4th Meeting of the European Society for Stereotactic and Functional Neurosurgery*, 35–54 (1980).
35. Yeo, B. T. T. et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology* (2011).
36. Caliński, T. & Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat.-theory Methods* **3**, 1–27 (1974).
37. Makris, N. et al. Decreased volume of left and total anterior insular lobule in schizophrenia. *Schizophr. Res.* **83**, 155–171 (2006).
38. Klein, A. & Tourville, J. 101 labeled brain images and a consistent human cortical labeling protocol. *Front. Neurosci.* **6**, 33392 (2012).
39. Yang, X. et al. Developing and sharing polygenic risk scores for 4,206 brain imaging-derived phenotypes for 400,000 UK Biobank subjects not participating in the imaging study. *medRxiv*, 2023-2004 (2023).
40. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
41. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
42. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
43. Watanabe, K., Umičević Mirkov, M., de Leeuw, C. A., van den Heuvel, M. P. & Posthuma, D. Genetic mapping of cell type specificity for complex traits. *Nat. Commun.* **10**, 3222 (2019).
44. Boyle, A. P. et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012).
45. Berger, M. F. et al. Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat. Biotechnol.* **24**, 1429–1435 (2006).
46. Berger, M. F. et al. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **133**, 1266–1276 (2008).
47. Matys, V. et al. TRANSFAC® and its module TRANSCOMP®: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* **34**, D108–D110 (2006).
48. Bryne, J. C. et al. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res.* **36**, D102–D106 (2007).
49. Badis, G. et al. Diversity and complexity in DNA recognition by transcription factors. *Science* **324**, 1720–1723 (2009).
50. Scharer, C. D. et al. Genome-wide promoter analysis of the SOX4 transcriptional network in prostate cancer cells. *Cancer Res.* **69**, 709–717 (2009).
51. Wei, G. H. et al. Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. *EMBO J.* **29**, 2147–2160 (2010).
52. Boyle, A. P. et al. High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. *Genome Res.* **21**, 456–464 (2011).
53. Pique-Regi, R. et al. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res.* **21**, 447–455 (2011).
54. 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).
55. Speed, D. & Balding, D. J. SumHer better estimates the SNP heritability of complex traits from summary statistics. *Nat. Genet.* **51**, 277–284 (2019).
56. Karczewski, K. J. et al. The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res.* **45**, D840–D845 (2017).
57. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
58. Consortium, G. T. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
59. Fromer, M. et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).
60. Ramasamy, A. et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat. Neurosci.* **17**, 1418–1428 (2014).
61. Wang, D. et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, eaat8464 (2018).
62. Schmitt, A. D. et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* **17**, 2042–2059 (2016).
63. Clough, E. & Barrett, T. The gene expression omnibus database. *Statistical Genomics: Methods and Protocols*, 93–110 (2016).
64. Roadmap Epigenomics Consortium Integrative analysis coordination Kundaje Anshul 1 2 3 Meuleman Wouter 1 2 Ernst Jason 1 2 4 Bilenky, M., Scientific program management Chadwick Lisa, H. & Principal investigators Bernstein Bradley, E. C. J. F. E. J. R. H. M. M. A. M. A. R. B. S. J. A. W. T. K. M. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).

65. Vickovic, S. et al. Three-dimensional spatial transcriptomics uncovers cell type localizations in the human rheumatoid arthritis synovium. *Commun. Biol.* **5**, 129 (2022).
66. Hagmann, P. From diffusion MRI to brain connectomics. *EPFL* (2005).
67. Panizzon, M. S. et al. Distinct genetic influences on cortical surface area and cortical thickness. *Cereb. Cortex* **19**, 2728–2735 (2009).
68. Besag, J. On the statistical analysis of dirty pictures. *J. R. Stat. Soc. Ser. B: Stat. Methodol.* **48**, 259–279 (1986).
69. Gottardo, R., Besag, J., Stephens, M. & Murua, A. Probabilistic segmentation and intensity estimation for microarray images. *Biostatistics* **7**, 85–99 (2006).
70. Zhao, E. et al. Spatial transcriptomics at subspot resolution with BayesSpace. *Nat. Biotechnol.* **39**, 1375–1384 (2021).
71. Vincent, L. & Soille, P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Mach. Intell.* **13**, 583–598 (1991).
72. Rand, W. M. Objective criteria for the evaluation of clustering methods. *J. Am. Stat. Assoc.* **66**, 846–850 (1971).
73. Hubert, L. & Arabie, P. Comparing partitions. *J. Classif.* **2**, 193–218 (1985).
74. Vinh, N. X., Epps, J. & Bailey, J. Information theoretic measures for clusterings comparison: is a correction for chance necessary? *Proceedings of the 26th Annual International Conference on Machine Learning*, 1073–1080 (2009).
75. Kraskov, A., Stögbauer, H. & Grassberger, P. Estimating mutual information. *Phys. Rev. E* **69**, 066138 (2004).
76. Miller, K. L. et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* **19**, 1523–1536 (2016).
77. Alfaro-Almagro, F. et al. Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage* **166**, 400–424 (2018).
78. Saykin, A. J. et al. Genetic studies of quantitative MCI and AD phenotypes in ADNI: progress, opportunities, and plans. *Alzheimer's Dement.* **11**, 792–814 (2015).
79. Shen, L. et al. Genetic analysis of quantitative phenotypes in AD and MCI: imaging, cognition and biomarkers. *Brain Imaging Behav.* **8**, 183–207 (2014).
80. Shen, L. et al. Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: A study of the ADNI cohort. *Neuroimage* **53**, 1051–1063 (2010).
81. Weiner, M. W. et al. The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. *Alzheimer's Dement.* **9**, e111–e194 (2013).
82. Weiner, M. W. et al. Recent publications from the Alzheimer's Disease Neuroimaging Initiative: Reviewing progress toward improved AD clinical trials. *Alzheimer's Dement.* **13**, e1–e85 (2017).
83. Davatzikos, C., Genc, A., Xu, D. & Resnick, S. M. Voxel-Based Morphometry Using the RAVENS Maps: Methods and Validation Using Simulated Longitudinal Atrophy. *NeuroImage* **14**, 1361–1369 (2001).
84. Ou, Y., Sotiras, A., Paragios, N. & Davatzikos, C. DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting. *Med. Image Anal.* **15**, 622–639 (2011).
85. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, s13742-13015–10047-13748 (2015).
86. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
87. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
88. Abraham, G. & Inouye, M. Fast principal component analysis of large-scale genome-wide data. *PLoS One* **9**, e93766 (2014).
89. Abraham, G., Qiu, Y. & Inouye, M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics* **33**, 2776–2778 (2017).
90. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
91. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
92. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164–e164 (2010).
93. Revelle, W. R. psych: Procedures for psychological, psychometric, and personality research. Northwestern University, Evanston, Illinois. R package version 2.4.12. <https://CRAN.R-project.org/package=psych> (2024).
94. Zhou, W. et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).
95. 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).
96. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
97. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
98. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).

Acknowledgements

This work was supported in part by the National Institutes of Health grants R01 AG071470 (L.S.), U01 AG068057 (L.S.), U01 AG066833 (L.S.), RF1 AG063481 (Q.L.), RF1 AG068191 (Y.Z.), and R01 AG071174 (Q.L.). Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). 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.

Author contributions

Study concept and design: J.B., L.S., and Q.L. Statistical analysis: J.B., C.C. Data interpretation: J.B., J.W., L.S., and Q.L. Data collection and processing: J.B., J.W., Y.C., G.E., Z.Y., C.D., L.S., and Q.L. Manuscript critical revision and submission approval: J.B., J.W., C.C., S.M., J.C., M.S., Y.C., G.E., Z.Y., S.Y., Z.W., Y.Z., D.K., D.D., A.J.S., B.Z., C.D., L.S., and Q.L.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-57636-6>.

Correspondence and requests for materials should be addressed to Qi Long or Li Shen.

Peer review information *Nature Communications* thanks Janine Bisterbosch and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

¹Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. ²Graduate Group in Genomics and Computational Biology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. ³Laboratory of AI and Biomedical Science (LABS), Columbia University, New York, NY, USA. ⁴Center for Innovation in Imaging Biomarkers and Integrated Diagnostics (CIMBID), Department of Radiology, Columbia University, New York, NY, USA. ⁵New York Genome Center (NYGC), New York, NY, USA. ⁶Department of Biostatistics and Health Data Science, Indiana University School of Medicine, Indianapolis, IN, USA. ⁷Graduate Group in Bioengineering, University of Pennsylvania, Philadelphia, PA, USA. ⁸Center for Biomedical Image Computing and Analytics (CBICA), University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. ⁹Graduate Group in Applied Mathematics and Computational Science, University of Pennsylvania, Philadelphia, PA, USA. ¹⁰Department of Biostatistics, Yale University School of Public Health, New Haven, CT, USA. ¹¹Department of Mathematics, United States Naval Academy, Annapolis, MD, USA. ¹²Department of Radiology and Imaging Sciences, Indiana University, Indianapolis, IN, USA. ¹³Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, PA, USA. ¹⁴Center for AI and Data Science for Integrated Diagnostics (AI2D), Department of Radiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. ✉e-mail: qlong@penmedicine.upenn.edu; li.shen@penmedicine.upenn.edu

The Alzheimer's Disease Neuroimaging Initiative

Michael Weiner^{15,16}, Paul Aisen¹⁷, Ronald Petersen¹⁸, Clifford R. Jack Jr¹⁸, William Jagust¹⁹, Susan Landau¹⁹, Monica Rivera-Mindt²⁰, Ozioma Okonkwo²¹, Leslie M. Shaw²², Edward B. Lee²², Arthur W. Toga²³, Laurel Beckett²⁴, Danielle Harvey²⁴, Robert C. Green²⁵, Andrew J. Saykin¹², Kwangsik Nho²⁶, Richard J. Perrin²⁷, Duygu Tosun¹⁵, Pallavi Sachdev²⁸, Erin Drake²⁵, Tom Montine²⁹, Catherine Conti¹⁶, Rachel Nosheny¹⁵, Diana Truran Sacrey¹⁶, Juliet Fockler¹⁵, Melanie J. Miller¹⁶, Winnie Kwang¹⁵, Chengshi Jin¹⁵, Adam Diaz¹⁶, Miriam Ashford¹⁶, Derek Flenniken¹⁶, Adrienne Kormos¹⁶, Michael Rafii¹⁷, Rema Raman¹⁷, Gustavo Jimenez¹⁷, Michael Donohue¹⁷, Jennifer Salazar¹⁷, Andrea Fidell¹⁷, Virginia Boatwright¹⁷, Justin Robison¹⁷, Caileigh Zimmerman¹⁷, Yuliana Cabrera¹⁷, Sarah Walter¹⁷, Taylor Clanton¹⁷, Elizabeth Shaffer¹⁷, Caitlin Webb¹⁷, Lindsey Hergesheimer¹⁷, Stephanie Smith¹⁷, Sheila Ogowang¹⁷, Olusegun Adegoke¹⁷, Payam Mahboubi¹⁷, Jeremy Pizzola¹⁷, Cecily Jenkins¹⁷, Naomi Saito²⁴, Kedir Adem Hussen¹⁷, Hannatu Amaza²¹, Mai Seng Thao²¹, Shaniya Parkins³⁰, Omobolanle Ayo³⁰, Matt Glittenberg²¹, Isabella Hoang²¹, Kaori Kubo Germano²⁰, Joe Strong²¹, Trinity Weisensel²¹, Fabiola Magana²¹, Lisa Thomas²¹, Vanessa Guzman³⁰, Adeyinka Ajayi³⁰, Joseph Di Benedetto³⁰, Sandra Talavera²⁰, Robert A. Koeppe³¹, Gil Rabinovici¹⁵, Victor Villemagne³², Brian LoPresti³², John Morris²⁷, Erin Franklin²⁷, Virginia M. Y. Lee²², Magdalena Korecka²², Magdalena Brylska²², Yang Wan²², J. Q. Trojanowski²², Karen Crawford²³, Scott Neu²³, Tatiana M. Foroud²⁶, Taeho Jo²⁶, Shannon L. Risacher²⁶, Hannah Craft²⁶, Liana G. Apostolova²⁶, Kelly Nudelman²⁶, Kelley Faber, Zoë Potter²⁶, Kaci Lacy²⁶, Rima Kaddurah-Daouk³³, Li Shen¹² ✉, Jason Karlawish²², Claire Erickson²², Joshua Grill³⁴, Emily Largent²², Kristin Harkins²², Zaven Kachaturian³⁵, Richard Frank³⁶, Peter J. Snyder³⁷, Neil Buckholtz³⁸, John K. Hsiao³⁸, Laurie Ryan³⁸, Susan Molchan³⁸, Zaven Khachaturian³⁹, Maria Carrillo⁴⁰, William Potter⁴¹, Lisa Barnes⁴², Marie Bernard³⁸, Hector González⁴³, Carole Ho⁴⁴, Jonathan Jackson⁴⁵, Eliezer Masliah³⁸, Donna Masterman⁴⁶ & Nina Silverberg³⁸

¹⁵University of California, San Francisco, CA, USA. ¹⁶Northern California Institute for Research and Education, San Francisco, CA, USA. ¹⁷University of Southern California, Los Angeles, CA, USA. ¹⁸Mayo Clinic, Rochester, MN, USA. ¹⁹University of California, Berkeley, CA, USA. ²⁰Fordham University, Bronx, NY, USA. ²¹University of Wisconsin-Madison, Madison, WI, USA. ²²University of Pennsylvania, Philadelphia, PA, USA. ²³University of Southern California School of

Medicine, Los Angeles, CA, USA. ²⁴University of California, Davis, CA, USA. ²⁵Harvard University, Cambridge, MA, USA. ²⁶Indiana University School of Medicine, Indianapolis, IN, USA. ²⁷Washington University, St. Louis, MO, USA. ²⁸Eisai, Nutley, NJ, USA. ²⁹University of Washington, Seattle, WA, USA. ³⁰Mt. Sinai, New York, NY, USA. ³¹University of Michigan, Ann Arbor, MI, USA. ³²University of Pittsburgh, Pittsburgh, PA, USA. ³³Duke University, Durham, NC, USA. ³⁴University of California, Irvine, CA, USA. ³⁵Khachaturian, Radebaugh & Associates (KRA), Potomac, MD, USA. ³⁶General Electric, Schenectady, NY, USA. ³⁷University of Connecticut, Storrs, CT, USA. ³⁸National Institute on Aging, Bethesda, MD, USA. ³⁹Prevent Alzheimer's Disease, 2020 Potomac, MD, USA. ⁴⁰Alzheimer's Association, Chicago, IL, USA. ⁴¹National Institute of Mental Health, Bethesda, MD, USA. ⁴²Rush University, Chicago, IL, USA. ⁴³University of California, San Diego, CA, USA. ⁴⁴Denali Therapeutics, South San Francisco, CA, USA. ⁴⁵Massachusetts General Hospital, Boston, MA, USA. ⁴⁶Biogen, Cambridge, MA, USA.