

1 ***Genetic influences on the shape of brain ventricular and subcortical structures***

2

3 **Running title: GWAS of brain ventricular and subcortical shapes**

4

5 Bingxin Zhao^{1,2}, Tengfei Li^{3,4}, Xiaochen Yang², Juan Shu², Xifeng Wang⁵, Tianyou Luo⁵, Yue
6 Yang⁵, Zhenyi Wu², Zirui Fan^{1,2}, Zhiwen Jiang⁵, Jie Chen⁵, Yue Shan⁵, Jiarui Tang⁵, Di Xiong⁵,
7 Ziliang Zhu⁵, Mufeng Gao⁵, Wyliena Guan⁵, Chalmer E. Tomlinson⁵, Qunxi Dong⁶, Yun
8 Li^{5,7,8}, Jason L. Stein^{7,9}, Yalin Wang^{6*}, and Hongtu Zhu^{4,5,7,8,10*}

9

10 ¹Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, PA 19104, USA.

11 ²Department of Statistics, Purdue University, West Lafayette, IN 47907, USA.

12 ³Department of Radiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

13 ⁴Biomedical Research Imaging Center, School of Medicine, University of North Carolina at Chapel Hill,
14 Chapel Hill, NC 27599, USA.

15 ⁵Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

16 ⁶School of Computing, Informatics, and Decision Systems Engineering, Arizona State University,
17 Tempe, AZ, USA

18 ⁷Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

19 ⁸Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599,
20 USA.

21 ⁹UNC Neuroscience Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

22 ¹⁰Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel
23 Hill, NC 27599, USA.

24 ****Corresponding authors:***

25 Yalin Wang

26 School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ
27 85287-8809

28 E-mail address: ylwang@asu.edu Phone: (480) 965-6871

29

30 Hongtu Zhu

31 3105C McGavran-Greenberg Hall, 135 Dauer Drive, Chapel Hill, NC 27599.

32 E-mail address: htzhu@email.unc.edu Phone: (919) 966-7250

1 **Abstract**

2 Brain ventricular and subcortical structures are heritable both in size and shape. Genetic
3 influences on brain region size have been studied using conventional volumetric
4 measures, but little is known about the genetic basis of ventricular and subcortical shapes.
5 Here we developed pipelines to extract seven complementary shape measures for lateral
6 ventricles, subcortical structures, and hippocampal subfields. Based on over 45,000
7 subjects in the UK Biobank and ABCD studies, 60 genetic loci were identified to be
8 associated with brain shape features ($P < 1.09 \times 10^{-10}$), 19 of which were not detectable
9 by volumetric measures of these brain structures. Ventricular and subcortical shape
10 features were genetically related to cognitive functions, mental health traits, and multiple
11 brain disorders, such as the attention-deficit/hyperactivity disorder. Vertex-based shape
12 analysis was performed to precisely localize the brain regions with these shared genetic
13 influences. Mendelian randomization suggests brain shape causally contributes to
14 neurological and neuropsychiatric disorders, including Alzheimer's disease and
15 schizophrenia. Our results uncover the genetic architecture of brain shape for ventricular
16 and subcortical structures and prioritize the genetic factors underlying disease-related
17 shape variations.

18

19 **Keywords:** ABCD; Brain disorders; GWAS; Hippocampus subfields; Mental health;
20 Subcortical and ventricular shapes; UK Biobank.

21

22

23

24

25

26

27

28

29

30

31

32

1 Human brain ventricular and subcortical structures are involved in complex brain
2 activities and have important roles in the regulation of cognitive, emotional, and motor
3 functions¹⁻⁷. Morphometric variations of these brain structures can be quantified in-vivo
4 by structural magnetic resonance imaging (MRI). MRI-based volumetric measures (such
5 as regional brain volumes) can estimate a region's overall size, providing a conventional
6 measure of the gross variation of the structure. However, such aggregate measures may
7 not be sensitive to within-region local changes and may not fully capture the complexity
8 of structural deformations. Shape analysis has gained increasing attention to overcome
9 these limitations and characterize brain morphometry beyond simple volumetric traits<sup>8-
10 11</sup>. Recent studies have found that shape features can precisely localize shape
11 deformations in brain structures, providing finer-grained information of the location and
12 pattern of morphological variations, which may not be detectable in traditional volume
13 analysis^{12,13}. For example, shape analysis of the ventricular and subcortical structures has
14 provided sensitive biomarkers for healthy aging¹⁴ and the onset and progression of a wide
15 range of brain diseases, including Alzheimer's disease (AD)^{15,16}, schizophrenia^{17,18},
16 epilepsy¹⁹, major depressive disorder (MDD)^{20,21}, 22q11.2 deletion syndrome²², and
17 bipolar disorder²³.

18

19 Both size (volume) and shape of brain ventricular and subcortical structures have been
20 found to be heritable in family studies^{24,25} and general populations^{13,26-28}. For example,
21 the narrow sense single-nucleotide polymorphism (SNP) heritability estimates for the
22 volume of ventricular and subcortical structures were all higher or close to 40%^{26,27} in the
23 UK Biobank²⁹ (UKB) studies, and the highest SNP heritability estimates for shape features
24 ranged from 32.7% to 53.3% across structures in the Rotterdam Study²⁸. Genome-wide
25 association studies (GWAS) have been conducted to uncover the genetic basis of
26 ventricular and subcortical volumes³⁰⁻⁴⁰, yielding hundreds of associated genetic variants
27 and shared genetic influences with brain disorders and complex traits. However, there is
28 no large-scale GWAS on ventricular and subcortical shape features and their genetic
29 architecture has yet to be determined.

30

31 Using raw MRI from over 45,000 subjects in the UKB and Adolescent Brain Cognitive
32 Development⁴¹ (ABCD) studies, we developed pipelines to extract ventricular and

1 subcortical shape features and characterized their genetic architectures. We identified 60
2 novel genetic loci that contributing to the shape variations, 19 of which cannot be
3 identified in previous GWAS of volumetric measures of these brain structures using the
4 same datasets. We found ventricular and subcortical shape features had shared genetic
5 influences with many cognitive traits and major brain disorders. We further revealed the
6 localized pattern of genetic effects in vertex-wise analysis and identified causal genetic
7 links between brain shape and disorders using Mendelian randomization analysis. The
8 results of this shape study demonstrated genetic effects on ventricular and subcortical
9 structures at a finer spatial resolution than that of traditional volumetric analysis. Our
10 GWAS results will be available through the Brain Imaging Genetics Knowledge Portal (BIG-
11 KP) <https://bigkp.org/>.

12

13 **RESULTS**

14 **Generating reproducible ventricular and subcortical shape features**

15 We developed pipelines to extract shape features from raw structural MRI images for 8
16 ventricular and subcortical structures, including the left/right lateral ventricles, nucleus
17 accumbens, amygdala, caudate, hippocampus, pallidum, putamen, and thalamus.
18 Furthermore, we studied 7 subfields of the hippocampus, namely the left/right cornu
19 ammonis 1 (CA1), CA3, fimbria, hippocampus-amygdala-transition-area (HATA),
20 hippocampal tail, presubiculum, and subiculum (**Fig.1A**). An overview of our workflow can
21 be found in **Fig. S1**. Shape deformations are usually decomposed into two components:
22 one within the surfaces, and the other along the normal axis of the surfaces. For each
23 vertex in the shape image, we calculated 7 complementary shape statistics, including the
24 radial distance from the medial model¹⁴ (referred to as the radial distance); the (log)
25 determinant and two eigenvalues of the Jacobian matrix from the surface tensor-based
26 morphometry (TBM) model⁴² (referred to as the determinant, eigenvalue1, and
27 eigenvalue2); and three features from the multivariate surface TBM (mTBM) model^{10,11}
28 (referred to as the mTBM1, mTBM2, and mTBM3) (**Figs.1B** and **S2**). Briefly, the radial
29 distance describes morphometric changes along the surface normal direction¹¹ and is also
30 called the radial thickness¹⁰. On the other hand, the 6 surface TBM and mTBM model
31 features capture surface deformations perpendicular to the surface normal axis (such as
32 rotation, dilation, and shear within the surfaces). For example, the determinant of

1 Jacobian matrix is analogous to a surface area⁴³, measuring local area dilation or
2 contraction. It quantifies the surface dilation ratio between the given template and the
3 study subject by matching a small surface patch around a particular point of the subject
4 surface to the corresponding point on the template. The three advanced surface mTBM
5 features analyze the full surface tensor using log-Euclidean metrics and can capture more
6 complicated surface deformations^{10,11}.

7

8 After extracting vertex-wise maps, we aggregated them and generated region-specific
9 summary-level features for downstream genetic analyses. For each shape statistics, we
10 have two groups of features. The first group includes 210 structure-averaged shape
11 features in regions or subfields by taking the mean across all the vertices within the
12 structure (7 shape measurements \times (8 shape structures + 7 hippocampal subfields) \times 2
13 hemispheres). In the second group, we applied principal component analysis to extract
14 1,120 region-specific principal components (PCs) by taking the top 10 PCs of the vertex-
15 wise map for each of the 7 statistics in the 8 ventricular and subcortical structures
16 (left/right, $7 \times 8 \times 2 \times 10$) (**Methods**). Principal component analysis is a well-established
17 method for dimension reduction with a wide range of neuroimaging applications. In shape
18 analysis, the top-ranked PCs can characterize the strongest variation components of
19 shape statistics within each structure, which can provide more microstructural details
20 about shape deformations omitted by structure-averaged measures, while alleviating
21 multiple testing burdens (**Fig.1C**). Clinically, variations represented by these PCs may
22 localize shape changes that are more relevant to specific brain-related complex traits or
23 diseases.

24

25 We evaluated the intra-subject reproducibility of the above region or subfield-specific
26 shape features using the repeat scans from the UKB repeat imaging visit (average time
27 between visits = 2 years, average $n = 2,788$). Specifically, we quantified individual-level
28 differences between the two visits by calculating the intraclass correlation coefficient
29 (ICC) of each shape feature between two observations from all revisited individuals. The
30 average ICC was 0.443 (standard error = 0.234) across the 1,330 (210 + 1,120) shape
31 features (**Table S1**). There were 457 shape features with ICC > 0.5, including 117 features
32 for lateral ventricles, 246 for subcortical structures, and 94 for hippocampal subfields (**Fig.**

1 **S3A**, mean ICC = 0.722, standard error = 0.132). The average ICC of the 7 shape statistics
2 ranged from 0.684 to 0.752 (**Fig. S3B**), and the lateral ventricles had the highest mean ICC
3 across all the 8 structures (**Fig. S3C**, mean ICC = 0.816, standard error = 0.129). Our later
4 genetic analyses focused primarily on these 457 reproducible shapes features (ICC > 0.5,
5 363 region-level and 94 subfield-level traits) (**Table S2**).

6 7 **Heritability and associated genetic loci of ventricular and subcortical shape features**

8 We estimated SNP heritability (h^2) for these 457 reproducible shapes features using the
9 UKB individuals of white British ancestry via GCTA⁴⁴ (average $n = 32,631$, phases 1-3
10 release). Most heritability estimates (456/457) were significant after adjusting for
11 multiple comparisons using the Benjamini-Hochberg procedure to control the false
12 discovery rate (FDR) at 0.05 level (**Fig. S4A** and **Table S3**). The mean heritability ranged
13 from 22.1% to 27.3% (standard error = 1.85%) across the 7 groups of shape statistics,
14 suggesting different shape deformation measures were under comparable genetic
15 controls (**Fig. S4B**). The highest heritability reported in each structure ranged from 51.2%
16 (for lateral ventricles) to 19.3% (for amygdala) (**Fig. S2C**). On average, lateral ventricles
17 had higher heritability than that of subcortical structures (32.2% vs. 21.6%, $P < 2.2 \times 10^{-16}$).
18 Subfield analysis provided more information for genetic influences on different parts
19 of the hippocampus. For example, we found the CA3 subfield had the highest heritability
20 among the 7 subfields, while the lowest heritability was observed on the HATA subfield
21 (**Fig. S4D**). In addition, the heritability estimates were largely consistent in females and
22 males (**Fig. S5**, correlation = 0.83).

23
24 It is known that the volumetric measures of these structures were also heritable²⁶⁻²⁸. To
25 quantify the genetic effects additionally contributing on shape measures, we compared
26 the estimates of genetic variance for 363 region-level shape features before and after
27 adjusting for their corresponding regional volumes as covariates (**Table S4**). For the 129
28 mean and first PC (PC1) shape features, we found that the average genetic variance
29 reduced from 0.241 to 0.145, indicating that 39.8% (0.096/0.241) genetic variations were
30 shared by regional brain volumes and shape features (**Fig. 2A**). In this group of shape
31 features, the proportion varied greatly from region to region. For example, the largest
32 proportion was observed in the lateral ventricles (79.2%) and the least was observed in

1 the amygdala (3.4%). As both mean and PC1 features captured the major variations in the
2 brain region⁴⁵, these results quantified the overlapping genetic influences between shape
3 and volumetric measures. On the other hand, for other PCs (other than the PC1s), the
4 genetic variance estimates were much more consistent before and after adjusting for
5 regional volumes (**Fig. 2B**). It was common for these PCs to capture more local variations
6 that are not captured by the mean or PC1s⁴⁵. Specifically, the average proportion of
7 reduction was 11.9% and the reductions were small for the majority of shape features.
8 There was a substantial reduction of shape features in a few second PCs (PC2) on the
9 lateral ventricles (70.9%), which can be explained by the fact that the ventricles were large
10 and therefore the second PC still mainly reflected global variations. Overall, these results
11 suggest that local PCs of shape features can detect genetic influences which are largely
12 independent of those found in regional volumes or aggregated shape measures.

13
14 We performed GWAS for the 457 shape features using the UKB individuals of white British
15 ancestry (average $n = 32,631$, **Methods**). The average intercept in linkage disequilibrium
16 score regression (LDSC)⁴⁶ was 1.0072 (range = (0.982, 1.034)), indicating no genomic
17 inflation of summary statistics because of confounding factors. At a stringent significance
18 level 1.09×10^{-10} ($5 \times 10^{-8}/457$, additionally adjusted for the number of shape features),
19 622 independent (linkage disequilibrium [LD] $r^2 < 0.1$) significant shape-variant
20 associations⁴⁷ were identified, which were distributed across 60 genomic regions
21 (cytogenetic bands). There were 38 regions associated with the lateral ventricles, 10 with
22 hippocampal subfields, 9 with hippocampus, 7 with putamen, 6 with nucleus accumbens,
23 6 with caudate, 4 with pallidum, and 4 with thalamus (**Fig. 2C**). **Table S5** summarizes the
24 list of index genetic variants and their associated shape features. Among the 60 regions,
25 19 were not identified by ventricular and subcortical regional volumes (at the $5 \times 10^{-8}/30$
26 significance level) in the same dataset. The genetic effects were highly consistent
27 between males and females in the sex-specific GWAS (**Fig. S6**, correlation = 0.966), where
28 analysis was conducted for females and males separately.

29
30 Using 5 independent European and non-European datasets, we replicated the genomic
31 loci identified in our discovery GWAS. First, we performed GWAS on a UKB European
32 dataset, which includes European individuals in the new UKB phase 4 data (early 2021

1 release) and individuals of white but non-British ancestry in the UKB phases 1-3 data
2 (UKBE, removed the relatives of the discovery sample, average $n = 4,596$). Of the 622
3 identified independent ($LD\ r^2 < 0.1$) shape-variant associations, 410 (65.9%) from 37
4 (61.7%, 37/60) loci passed the 0.05 nominal significance level in UKBE. Their genetic
5 effects all had concordant directions in the UKBE and original discovery GWAS and were
6 highly similar (correlation = 0.980). Second, we repeated GWAS on a validation dataset
7 with UKB non-European subjects (UKBNE, average $n = 1,224$). In this dataset, 19 loci can
8 be validated, and most of their associations (83/86) had the same directions with those
9 in the discovery GWAS and UKBNE. The validated genetic effects were highly consistent
10 between the white British discovery GWAS and non-European GWAS (correlation =
11 0.931). These results suggest similar genetic effects on subjects from different ancestries
12 in the same cohort.

13

14 Next, we carried out GWAS on 3 ABCD validation datasets: the ABCD European (ABCDE,
15 average $n = 3,177$), ABCD Hispanic (ABCDH, average $n = 662$), and ABCD Black (ABCDB,
16 average $n = 1,002$). In ABCDE, 23 of the 60 genomic loci were significant at nominal
17 significance level and had the same effect direction as in the UKB discovery GWAS.
18 The ABCDH and ABCDB had 13 and 15 validated loci, respectively. Interestingly, we found
19 that the genetic effects of 3q28 locus on lateral ventricles were much larger in the three
20 ABCD datasets than those of the UKB discovery sample, especially for the non-European
21 subjects in ABCDH (mean absolute genetic effects 0.082 vs. 0.3367, $P = 1.35 \times 10^{-7}$) and
22 ABCDB (mean absolute genetic effects 0.082 vs. 0.401, $P = 1.67 \times 10^{-4}$). The 3q28 locus
23 was reported to have the strongest associations with lateral ventricular volume³¹ and was
24 widely associated with Alzheimer's disease risk and biomarkers⁴⁸. Larger 3q28 effects in
25 ABCD may suggest that the genetic effects on lateral ventricles were stronger for younger
26 subjects and/or non-European subjects. Overall, 43 of the 60 loci can be validated in at
27 least one of the 5 datasets, 30 loci can be validated in more than one dataset, and 5 loci
28 can be consistently validated in all datasets, including 3q28, 17q24.1, 14q32.11, 12q14.3,
29 and 10q26.13. The validated loci (such as 3q28) may have higher genetic effect sizes in
30 ABCD than in UKB discovery GWAS. These validation results were summarized in **Figure**
31 **S7** and **Table S5**.

32

1 Finally, we detected gene-level associations using MAGMA⁴⁹ and FUMA⁴⁷. MAGMA
2 reported 127 significant genes with 1,343 associations ($P < 5.82 \times 10^{-9}$, adjusted for 457
3 phenotypes), covering all the ventricular and subcortical structures (**Fig. S8** and **Table S6**).
4 Among the 127 significant genes, 59 were identified by regional brain volumes³⁵, 53 were
5 observed in DTI parameters⁴⁵, and 12 overlapped with functional MRI (fMRI) traits⁵⁰. Eight
6 genes were associated with all the 4 brain imaging modalities, including *FAM175B*,
7 *FAM53B*, *METTL10*, and *RP11-12J10.3* (*METTL10-FAM53B* readthrough) in the 10q26.13
8 region, as well as *EPHA3* (3p11.1), *ZIC1* (3q24), *ZIC4* (3q24), and *DAAM1* (14q23.1). The
9 *FAM175B*, *FAM53B*, and *METTL10* genes had important functions in ribosomal translation
10 and cell regeneration⁵¹, and have been mapped to cocaine dependence⁵² and subjective
11 well-being⁵³. The *EPHA3* was involved in axon guidance⁵⁴ and was highly expressed in
12 mesenchymal subtype glioblastoma⁵⁵. The *DAAM1* was an important part of the planar
13 cell polarity signaling in neural development⁵⁶ and was highly expressed in human
14 cerebral cortex⁵⁷. In addition, the *ZIC* genes were important components in patterning
15 the cerebellum⁵⁸. Overall, these 127 MAGMA-significant genes showed gene ontology
16 enrichments⁵⁹ in “cell morphogenesis involved in differentiation (GO:0000904)” and “T
17 cell receptor signaling pathway (GO:0050852)” biological processes at FDR 0.05 level ($P <$
18 2.76×10^{-6}). We also used FUMA⁴⁷ to map significant variants ($P < 1.09 \times 10^{-10}$) to genes
19 through a combination of their base pair location, gene expression, and 3D chromatin (Hi-
20 C) interaction. FUMA reported 383 associated genes, 313 of which were not discovered
21 in MAGMA (**Table S7**). These results demonstrate the polygenic genetic architecture of
22 shape features and prioritize important genes involved in the biological pathways of brain
23 functions and diseases.

24

25 **The shared genetic influences with complex brain traits and disorders.**

26 We took a further look into the 43 validated regions, providing variant annotations and
27 details of shared genetic influences with other complex traits and diseases. For all the
28 independent (LD $r^2 < 0.1$) significant variants (and variants in LD, $r^2 \geq 0.6$) detected in these
29 validated regions, we searched for their GWAS signals reported in the NHGRI-EBI GWAS
30 catalog⁶⁰. Shape deformation of ventricular and subcortical structures has substantial
31 regional and local genetic overlaps with complex traits and clinical endpoints. The full
32 information was presented in **Table S8**, and below we highlighted some regions and their

1 reported variants and genes reported for brain structures/functions, neurological
2 disorders, psychiatric disorders, psychological traits, migraine, cognitive traits,
3 educational attainment, sleep/physical activity, osteoarthritis/pain, Alzheimer's Disease
4 biomarkers, diabetes/kidney diseases, blood traits, blood pressure, smoking/drinking,
5 lung/liver, and lipoprotein cholesterol.

6
7 Our results were concordant with previous GWAS results for regional brain volumes and
8 cortical thickness in many genomic loci, including 3q24 (index variant rs2279829, the
9 nearest gene *ZIC4*), 8q24.12 (rs10283100, *ENPP2*), 9q31.3 (rs734250, *LPAR1*), 9q33.1
10 (rs10983205, *ASTN2*), 11q14.3 (rs1531249, *FAT3*), 11q23.1 (rs34077344, *LINC02550*),
11 11q23.3 (rs10892133, *DSCAML1*), 12q14.3 (rs61921502, *MSRB3*), 12q23.3 (rs12369969,
12 *NUAK1*), 12q24.22 (rs7132910, *HRK*), 14q22.3 (rs945270, *KTN1*), and 16q22.3 (rs7193665,
13 *ZFH3*) (**Figs. S9-S20**). For example, rs7132910 was identified to be associated with
14 hippocampal volume⁶¹. In the current study, we found rs7132910-associations with
15 multiple shape features of the hippocampus, particularly the subiculum subfield.
16 Additionally, rs2279829 and rs7132910 were expression quantitative trait loci (eQTLs) of
17 *ZIC4* and *TESC* in human brain tissues⁶², suggesting that these shape-associated variants
18 were known to affect gene expression levels in human brain. Among these regions, we
19 also tagged genetic variants (LD $r^2 \geq 0.6$) reported for risk-taking⁶³, alcohol consumption,
20 smoking initiation⁶⁴, lung function⁶⁵, chronic obstructive pulmonary disease (COPD)⁶⁶,
21 blood pressure⁶⁵, and Alzheimer's disease pathologies⁶⁷.

22
23 Our shape GWAS results frequently tagged regions reported for white matter
24 microstructure, including 17q24.1 (rs62072157, *GNA13*), 2p13.2 (rs34754475, *DYSF*),
25 3q28 (3:190672426_CT_C, *GMNC*), 5q14.2 (rs12187334, *ATP6AP1L*), 7p21.1 (rs4329170,
26 *TWISTNB*), 7p22.2 (rs1183079, *GNA12*), 7p22.3 (rs368699386, *AMZ1*), 14q32.12
27 (rs529889896, *CCDC88C*), 16q24.2 (rs56023709, *C16orf95*), and 16q24.3 (rs8404, *CDK10*)
28 (**Figs. 3A and S21-S30**). For example, rs62072157 was associated with (the ninth PC of)
29 the radial distance of the right lateral ventricle, and it also had significant associations
30 with the fornix (column and body) and anterior corona radiata white matter tracts ($P < 4$
31 $\times 10^{-9}$). Rs62072157 was an eQTL of *GNA13* and *RGS9* in brain tissues⁶². The *GNA13* was a
32 core gene involved in early brain development regulations⁶⁸ and has been implicated in

1 brain diseases such as schizophrenia⁶⁹. In addition, rs1183079 was a brain eQTL of *AMZ1*,
2 which was associated with mTBM2 and radial distance of the right lateral ventricle, as
3 well as multiple white matter tracts, such as the body of corpus callosum, anterior corona
4 radiata, retrolenticular part of internal capsule, posterior corona radiata, and superior
5 corona radiata ($P < 5.5 \times 10^{-11}$). Furthermore, rs8404 was a brain eQTL of *CDK10*, *SPATA33*,
6 *VPS9D1*, *MC1R*, and *ACSF3*. Rs8404 was associated with multiple shape features of the
7 left hippocampus and affected the integrity of the retrolenticular part of internal capsule
8 and superior longitudinal fasciculus tracts ($P < 2 \times 10^{-8}$). The *CDK10* was important for
9 neural development⁷⁰. In addition to Alzheimer's disease biomarkers⁴⁸ and COPD⁶⁶, we
10 found shared genetic influences with type 2 diabetes⁷¹ and blood traits (such as plasma
11 homocysteine levels⁷² and platelet distribution width⁷³) on these white matter-
12 overlapping genomic loci.

13

14 In 6p22.1 and 18q21.2 regions, we observed the shared genetic influences between shape
15 features and multiple psychiatric disorders and psychological traits. For example, we
16 tagged rs7766356 (nearest gene *ZSCAN23*, 6p22.1) and rs11665242 (*DCC*, 18q21.2), which
17 have been implicated with schizophrenia^{74,75} (**Figs. S31-S32**). Rs7766356 was an eQTL of
18 *ZSCAN23*, *ZSCAN31*, *ZKSCAN3*, and *ZSCAN26*, which might represent drug targets for
19 schizophrenia⁷⁶. We also tagged risk variants for bipolar disorder⁷⁷ (e.g., rs144447022)
20 and MDD⁷⁸ (e.g., rs926552) in these two regions and for neuroticism⁷⁹ in 5q14.3 (e.g.,
21 rs16902900, *TMEM161B*) (**Fig. S33**). In 2q24.2, 6q22.32, 8p11.21 regions (**Fig. 3B** and **Figs.**
22 **S34-S36**), as well as the 6p22.1 and 18q21.2 regions related to brain disorders, we found
23 shared genetic influences with a wide range of cognitive and educational traits, such as
24 intelligence⁸⁰ (e.g., rs2268894, *DPP4*, 2q24.2), educational attainment⁸¹ (e.g.,
25 rs11759026, *CENPW*, 6q22.32; rs2974312, *SMIM19*, 8p11.21), self-reported math ability⁸¹
26 (e.g., rs71559051, *H2BC15*, 6p22.1), and verbal-numerical reasoning⁸² (e.g., rs62100026,
27 *DCC*, 18q21.2). Finally, we tagged sleep-related variants in 10p12.31 and 11q14.1 regions,
28 such as insomnia⁸³ (e.g., rs12251016, *MLLT10*, 10p12.31; and rs667730, *DLG2*, 11q14.1,
29 **Figs. S37-S38**). These results suggest the shape features could be used as imaging
30 biomarkers to study the etiologic study of brain-related diseases and complex traits.

31

1 Our results also help us to better understand the genetic links between brain atrophy and
2 the health of other organs. For example, the index rs10901814 (*FAM53B*, 10q26.13) was
3 associated with the shape features in the hippocampus and lateral ventricles and it was a
4 brain eQTL for *EEF1AKMT2* and *LHPP* genes. In this region, we tagged risk variant (index
5 variant rs4962691 for estimated glomerular filtration rate (eGFR)⁸⁴, which was a clinical
6 biomarker for kidney function and disease (**Fig. 3C**). Brain and kidney had similar
7 hemodynamic mechanisms and shared physiological links⁸⁵. Cognitive impairment and
8 accompanied brain structural changes (such as hippocampus volume) have been
9 frequently reported in chronic kidney disease^{86,87}. To precisely localize the pattern of
10 genetic effects on brain shapes, we took the eGFR lead index rs4962691 and performed
11 vertex-wise analysis on spatial maps of hippocampus and lateral ventricles. We found that
12 the rs4962691-related shape deformation was mainly localized to specific areas of the
13 hippocampus and lateral ventricles, such as the hippocampal tail and CA1, as well as the
14 atrium and posterior horn of lateral ventricles (**Fig. 3D**). These genetic overlaps and local
15 structural variations may represent the mediated brain changes related to the cognitive
16 impairment in chronic kidney disease.

17

18 **Genetic correlations with complex traits and clinical outcomes**

19 We explored genetic correlations (GC) between shape features and a wide range of other
20 complex traits. First, we used LDSC⁸⁸ to examine pairwise genetic correlation between the
21 457 shape features and 211 brain structural traits, including 101 regional brain volumes³⁵
22 and 110 diffusion tensor imaging (DTI) parameters⁴⁵. Among the 96,427 (457 × 211) tests,
23 16.22% were significant at the FDR 5% level (**Fig. S39** and **Table S9**). Both regional brain
24 volumes and DTI parameters had significant genetic correlations with ventricular and
25 subcortical shape features. For DTI parameters, the strongest associations were observed
26 on the lateral ventricles, and the top 5 associated white matter tracts included the fornix,
27 body of corpus callosum, superior corona radiata, posterior corona radiata, and posterior
28 limb of internal capsule ($P < 1.49 \times 10^{-55}$). Meanwhile, either the fornix or body of corpus
29 callosum tracts consistently had the strongest associations with the subcortical structures,
30 such as the hippocampus ($P = 9.67 \times 10^{-25}$), nucleus accumbens ($P = 2.18 \times 10^{-23}$), amygdala
31 ($P = 4.51 \times 10^{-8}$), caudate ($P = 1.32 \times 10^{-19}$), pallidum ($P = 9.36 \times 10^{-16}$), putamen ($P = 2.73$
32 $\times 10^{-15}$), and thalamus ($P = 8.95 \times 10^{-38}$). Our subfield analysis further revealed that the

1 fornix and body of corpus callosum associations were mainly localized in the
2 presubiculum subfield of hippocampus ($P < 3.68 \times 10^{-14}$). The fornix located below the
3 corpus callosum and connected the hippocampus to subcortical structures⁸⁹. Our results
4 suggested that the fornix integrity and shape deformations had strong genetic overlaps.
5 For regional brain volumes, we found associations for volumes of both cortical and
6 subcortical structures. As expected, the ventricular and subcortical structures had strong
7 genetic correlations with subcortical volumes. In addition, genetic correlations were also
8 widely observed between cortical structures and ventricular and subcortical shapes. Top-
9 ranked cortical volumes included the left/right insula (with putamen, $P < 7.61 \times 10^{-18}$),
10 left/right isthmus cingulate (with ventricle and hippocampus presubiculum, $P < 9.77 \times 10^{-10}$),
11 left/right lingual (with ventricle, $P < 2.70 \times 10^{-14}$), left/right pericalcarine (with
12 ventricle, $P < 3.41 \times 10^{-12}$), and left/right cuneus (with ventricle, $P < 2.37 \times 10^{-10}$). Overall,
13 these results suggest that ventricular and subcortical shape features are genetically
14 related to white matter integrity and structural variations of cortical regions.

15

16 Next, we examined genetic correlations between the 457 shape features and 48 complex
17 traits and diseases. At the FDR 5% level (457×48 tests), we found the shape features were
18 associated with brain disorders (such as attention-deficit/hyperactivity disorder (ADHD),
19 schizophrenia, and anorexia nervosa), cognitive traits (such as cognitive function,
20 intelligence, and reaction time), sleep traits (such as snoring, insomnia, extreme
21 chronotype), neuroticism, risk-taking, metabolic traits, and cardiovascular diseases (such
22 as hypertension and coronary artery disease (CAD)) (**Fig. S40** and **Table S10**). For example,
23 ADHD was positively correlated with shape features of the left hippocampus and lateral
24 ventricles ($|GC| > 0.184$, $P < 5.69 \times 10^{-4}$, **Figs. 4A-4B**). In ADHD, there have been reports
25 of abnormalities in hippocampal structures, possibly as a result of the brain's efforts to
26 compensate for disruptions of time perception and a tendency to avoid waiting⁹⁰. The
27 Rotterdam Study reported that ADHD-related genetic variants were associated with
28 structural brain changes in the lateral ventricles⁹¹. In addition, schizophrenia was
29 genetically associated with shape features of the thalamus and hippocampus ($|GC| >$
30 0.133 , $P < 6.14 \times 10^{-4}$). Both the thalamus and hippocampus had crucial roles in functional
31 and structural pathways related to schizophrenia⁹² and smaller volumes in the two
32 structures were frequently reported in schizophrenia patients^{93,94}. An earlier study using

1 subcortical brain volumes was not able to detect genetic overlap between schizophrenia
2 risk and subcortical structures⁹⁵. Furthermore, most significant genetic correlations with
3 cognitive function, intelligence, and education were with the lateral ventricles ($|GC| >$
4 0.113 , $P < 5.42 \times 10^{-4}$), while the reaction time additionally had genetic correlations with
5 the thalamus ($|GC| > 0.115$, $P < 7.90 \times 10^{-4}$, **Fig. 4C**). Ventricular enlargement was strongly
6 correlated with cognitive performance decline⁹⁶. The thalamus passed information
7 between the brain and body and anticipatory thalamic activity can predict reaction time⁹⁷.
8 We also observed specific genetic correlations for other traits, such as between insomnia
9 and the caudate ($|GC| > 0.141$, $P < 7.85 \times 10^{-4}$), risk tolerance and the hippocampus ($|GC|$
10 > 0.283 , $P < 8.15 \times 10^{-4}$), and automobile speeding and the lateral ventricles ($|GC| > 0.148$,
11 $P < 3.52 \times 10^{-4}$). Furthermore, we found that CAD was genetically correlated with the
12 hippocampus and lateral ventricles ($|GC| > 0.260$, $P < 6.40 \times 10^{-4}$, **Fig. 4D**). CAD had long-
13 term negative impact on brain health and reduced neural connectivity changes in the
14 hippocampus were observed in CAD and may contribute to cognitive impairment⁹⁸. A
15 significant genetic correlation between AD and the hippocampus was not observed.

16

17 **Phenome-wide association study using shape polygenic risk scores**

18 We tested associations between the 457 shape features and more complex traits and
19 clinical outcomes using polygenic risk scores (PRS) of shape features in the UKB non-
20 imaging cohort (**Methods**). A total of 276 complex traits and clinical outcomes were
21 selected from a variety of categories (**Table S11**). Briefly, we constructed PRS using PRS-
22 CS⁹⁹ for shape features on unrelated UKB subjects without brain MRIs ($n = 379,860$, also
23 removing relatives of imaging subjects). We then focused on the UKB white British
24 subjects and randomly selected 70% individuals (average $n = 202,405$) as discovery
25 sample to test pairwise associations (457×276 tests) and validated these results in a hold-
26 out dataset consisting of the rest of 30% white British subjects (average $n = 86,736$), UKB
27 white non-British subjects (average $n = 20,746$), and UKB non-white subjects (average $n =$
28 $21,587$). Detailed information on the adjusted covariates can be found in the **Methods**
29 section. We prioritized significant associations in the validation sample (at the nominal
30 significance level) with concordant regression coefficients when they passed the
31 Bonferroni significance level in the discovery sample. Among the 457×276 tests, 2,907
32 were significant at Bonferroni significance level in the discovery sample, among which

1 94.60% (2,750) were validated (**Fig. S41** and **Table S12**). We highlighted below the
2 association patterns with clinical outcomes, mental health, cognitive function, physical
3 activity, lifestyle, and biomarkers.

4
5 We observed significant associations between shape PRS and multiple diseases, including
6 diabetes, hyperthyroidism, hypothyroidism, multiple sclerosis, psoriasis, and vascular
7 heart problems. For example, diabetes was significantly associated with PRS for multiple
8 shape features of the left hippocampus and its subfields ($|\beta| > 0.0109, P < 2.47 \times 10^{-7}$).
9 These findings were consistent with previous studies showing patients with long diabetes
10 had higher risk of hippocampal atrophy, and loss of hippocampal neuroplasticity and
11 neurogenesis^{100,101}. In addition, both hyperthyroidism and hypothyroidism were mostly
12 associated with shape PRS of the lateral ventricles and hippocampus (and hippocampal
13 subfields) ($|\beta| > 0.0107, P < 3.87 \times 10^{-7}$). Adults with hypothyroidism were reported to
14 have decreased hippocampal volume¹⁰² and people with hyperthyroidism had smaller
15 grey matter volume in bilateral hippocampus¹⁰³. Hyperthyroidism and hypothyroidism
16 had also been linked to changes of brain ventricle size¹⁰⁴. Multiple sclerosis was mostly
17 associated with shape features of the putamen and hippocampus ($|\beta| > 0.0108, P < 3.42$
18 $\times 10^{-7}$). Decreased putamen and hippocampal volumes have been reported in multiple
19 sclerosis patients^{105,106}. We also found associations between vascular heart problems and
20 shape PRS of the hippocampus ($|\beta| > 0.0105, P < 3.71 \times 10^{-7}$), consistent with previous
21 studies showing decreased hippocampal volume among patients with vascular heart
22 problems¹⁰⁷.

23
24 There were significant associations between shape PRS and multiple mental health traits
25 related to anxiety, depression, and neuroticism ($|\beta| > 0.0106, P < 1.09 \times 10^{-11}$). Most of
26 these mental health traits were associated with shape features of the hippocampus and
27 its subfields. These observed associations were consistent with recent findings on
28 reduced hippocampal volume in subiculum¹⁰⁸ and fimbria¹⁰⁹ among patients with MDD.
29 Significant associations were also found between mental health and shape PRS of other
30 brain structures, such as between nervous feelings and the nucleus accumbens,
31 neuroticism and the putamen, as well as tiredness/lethargy and the caudate.

32

1 The blood biochemistry biomarkers were widely associated with the shape PRS of all
2 structures, with the majority of associations involving the hippocampus, hippocampal
3 subfields, and lateral ventricles. Examples of associated biomarkers included
4 apolipoprotein A, aspartate aminotransferase, glycated haemoglobin (HbA1c), high-
5 density lipoprotein (HDL) cholesterol, insulin-like growth factor 1 (IGF-1), total protein,
6 and urate ($|\beta| > 0.009$, $P < 4.28 \times 10^{-42}$). It is known that urate caused hippocampal
7 infection, which in turn induced cognitive dysfunction¹¹⁰. HbA1c measured the blood
8 sugar level and was used in diagnosis of diabetes. Our results were consistent with a
9 recent study that higher level of HbA1c was associated with smaller hippocampal
10 volume¹¹¹. In summary, shape PRS uncovered the links between shape features and a
11 wide range of complex traits and diseases. As the shape PRS were genetically predicted
12 traits, these observed associations also indicate the widespread underlying shared
13 genetic influences.

14

15 **Causal relationships with clinical endpoints detected by Mendelian randomization.**

16 To explore the causes and consequences of shape deformations, Mendelian
17 randomization (MR) was used to identify potential causal relationships between the 457
18 shape features and 288 clinical endpoints collected by FinnGen¹¹² and the Psychiatric
19 Genomics Consortium¹¹³ (**Table S13**). We tested 14 different MR methods¹¹⁴⁻¹¹⁷, and the
20 detailed implementation information can be found in the **Methods** section.

21

22 At the Bonferroni significance level ($P < 3.92 \times 10^{-8}$), we found strong evidence of genetic
23 causal effects from shape features to brain disorders, including AD, schizophrenia, and
24 cross disorders (five major psychiatric disorders¹¹⁸) (**Fig. 5A** and **Table S14**). For example,
25 multiple ventricular shape statistics had significant genetic causal effects on Alzheimer's
26 disease ($|\beta| > 0.474$, $P < 2.73 \times 10^{-8}$), most of which were from the left lateral ventricle.
27 AD can affect gray and white matter structures surrounding the ventricles, and ventricular
28 enlargement and expansion have been frequently identified in AD^{9,119-121}. Our findings
29 provide further evidence for the causal genetic pathway underlying brain structural
30 variations and Alzheimer's disease. There was a consistent sign of causal genetic effects
31 across different MR methods. In addition, shape features of the putamen and
32 hippocampus were causally linked to schizophrenia ($|\beta| > 0.339$, $P < 1.97 \times 10^{-8}$). In

1 schizophrenia, neuropsychological impairments have been associated with hippocampal
2 structures⁹⁴. It is well known that the putamen is associated with both increased
3 dopamine synthesis capacity and frontostriatal dysconnectivity in schizophrenia, as well
4 as with antipsychotic treatment effects¹²². Additionally, we did not detect causal genetic
5 links to schizophrenia by analyzing the volumes of putamen and hippocampus structures
6 ($|\beta| < 0.293$, $P > 0.0011$). Furthermore, there were significant genetic causal relationships
7 between shape features and other diseases of the nervous system, such as carpal tunnel
8 syndrome ($|\beta| > 0.249$, $P < 1.26 \times 10^{-8}$) and migraine ($|\beta| > 0.256$, $P < 7.61 \times 10^{-14}$). All
9 of the above results passed the MR-Egger intercept test, indicating the absence of
10 horizontal pleiotropy.

11

12 Additionally, we identified causal relationships where sleep traits were the exposure and
13 brain structural traits were the outcome at Bonferroni significance level ($P < 1.12 \times 10^{-8}$). A
14 large proportion of the significant findings (26/42) were related to the diseases of the
15 circulatory system, such as aortic aneurysm, calcific aortic valvular stenosis, heart failure,
16 and hypertensive heart disease (**Fig. 5B** and **Table S14**). For example, calcific aortic
17 valvular stenosis had causal genetic links to the shape of the pallidum ($|\beta| > 0.059$, $P <$
18 1.02×10^{-8}), and hypertensive heart disease may lead to the changes of the fimbria
19 subfield ($|\beta| > 0.075$, $P < 4.28 \times 10^{-9}$). Multiple brain disorders, such as stroke¹²³ and
20 dementia¹²⁴, can be caused by diseases of the circulatory system. We also found that
21 COPD was causally related to the shape of hippocampus and lateral ventricles ($|\beta| >$
22 0.064 , $P < 9.87 \times 10^{-9}$). Cognitive impairments and related alterations of hippocampus had
23 been found in COPD patients¹²⁵. Overall, our MR results suggest that genetic causal
24 pathways may exist between brain shapes and brain disorders, such as Alzheimer's
25 disease. These findings also reveal possible genetic mechanisms of non-brain diseases
26 (e.g., heart diseases) that underlie brain health.

27

28 **DISCUSSION**

29 Brain volumes are commonly used to analyze brain ventricular and subcortical structures.
30 Increasing evidence, however, indicates that volumetric measurements can only partially
31 capture structural complexity. As a result, shape analysis has been performed to uncover
32 deformations that are not visible in volumetric analysis. Using over 45,000 MRI scans in

1 the UKB and ABCD studies, our study uncovered the genetic architecture of shape
2 features within the ventricular and subcortical regions. We identified genetic connections
3 between shape features and a wide range of complex traits and clinical outcomes.
4 Leveraging shape features identified new loci that could not be identified by brain
5 volumes and provided fine details for localizing the genetic effect patterns within brain
6 structures. The results of our study improved the spatial resolution for identifying
7 genetically important brain areas that influenced clinical outcomes. For example,
8 rs4962691, a risk variant for eGFR, had genetic effects in specific parts of the
9 hippocampus and lateral ventricles. These shape features may be used as
10 endophenotypes of the cognitive impairment in kidney disease. In summary, as one of
11 the first large-scale studies to examine the genetic architecture of ventricular and
12 subcortical shape features, our results provide specific shape biomarkers that can be used
13 in clinical research questions.

14

15 The present study has a few limitations. First, the current analysis mainly used data from
16 the European UKB subjects, which may limit the generalizability of our research findings.
17 In the validation analysis, we have observed that European-significant genetic variants
18 tended to have larger effects in non-European cohorts of the ABCD study. The inclusion
19 of more global samples and identification of cross-population components of genetic
20 effects on brain shape will be of great interest in future studies. Second, we used PCA to
21 extract low-rank features from vertex-wise maps, which can capture local shape
22 deformations while mitigating multiple testing burden for genome-wide testing. PCA is a
23 powerful statistical tool to extract linear structures in the data. Nevertheless, PCA may
24 not be the most efficient method for reducing dimensions because complicated shape
25 variations can be non-linear. It might be possible to generate more powerful shape
26 features using variational autoencoder¹²⁶ and transfer learning¹²⁷. Lastly, we studied
27 hippocampal subfields, which may reflect specific biological processes and cognitive
28 functions^{128,129}. Other subcortical structures (such as amygdala¹³⁰) and lateral ventricles¹²
29 can also be divided into different portions or subregions with distinct functions. The
30 development and application of automated segmentation methods in large-scale MRI
31 datasets will enable the discovery of genetic influences at the subfield level for more brain
32 structures.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

METHODS

Methods are available in the *Methods* section.

Note: One supplementary information pdf file, one supplementary figure pdf file, and one supplementary table zip file are available.

ACKNOWLEDGEMENTS

This research was partially supported by U.S. NIH grants MH086633 (HT.Z.), MH116527 (TF.L.), and U01HG011720 (Y.L.). We thank the individuals represented in the UK Biobank and ABCD studies for their participation and the research teams for their work in collecting, processing and disseminating these datasets for analysis. We thank Doug Crabill for helpful conversations on computing. We would like to thank the University of North Carolina at Chapel Hill and Purdue University and their research computing group for providing computational resources and support that have contributed to these research results. We gratefully acknowledge all the studies and databases that made GWAS summary data available. This research has been conducted using the UK Biobank resource (application number 22783), subject to a data transfer agreement. Part of the data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (<https://abcdstudy.org>), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children age 9-10 and follow them over 10 years into early adulthood. The ABCD Study is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, U24DA041147, U01DA041093, and U01DA041025. A full list of supporters is available at <https://abcdstudy.org/federal-partners.html>. A listing of participating sites and a complete listing of the study investigators can be found at <https://abcdstudy.org/scientists/workgroups/>. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD

1 consortium investigators. Assistance for this project was provided by the UNC Intellectual
2 and Developmental Disabilities Research Center (NICHD; P50 HD103573).

4 **AUTHOR CONTRIBUTIONS**

5 B.Z. and H.Z. designed the study. B.Z., TF.L., X.Y., J.S., X.W., TY. L, Y.Y., Z.W., Z.F., and Z. J.
6 analyzed the data. TF. L., X.W., TY. L, Y.Y., J.C., Y.S., J.T., D.X, Z.Z., M.G., W.G., and C.T.
7 processed the MRI data. Y.W. and Q.D. helped on the shape feature pipelines. Y.L. and
8 J.L.S. provided feedback on study design and results interpretations. B.Z. wrote the
9 manuscript with feedback from all authors.

11 **CORRESPONDENCE AND REQUESTS FOR MATERIALS** should be addressed to H.Z.

13 **COMPETETING FINANCIAL INTERESTS**

14 The authors declare no competing financial interests.

16 **REFERENCES**

- 17 1. Scelsi, C. *et al.* The lateral ventricles: a detailed review of anatomy, development,
18 and anatomic variations. *American Journal of Neuroradiology* **41**, 566-572
19 (2020).
- 20 2. Crosson, B. Subcortical Functions in Cognition. Vol. 31 419-421 (Springer, 2021).
- 21 3. Johnson, M.H. Subcortical face processing. *Nature Reviews Neuroscience* **6**, 766-
22 774 (2005).
- 23 4. Utter, A.A. & Basso, M.A. The basal ganglia: an overview of circuits and function.
24 *Neuroscience & Biobehavioral Reviews* **32**, 333-342 (2008).
- 25 5. Berridge, K.C. & Kringelbach, M.L. Pleasure systems in the brain. *Neuron* **86**, 646-
26 664 (2015).
- 27 6. Bressler, S.L. & Menon, V. Large-scale brain networks in cognition: emerging
28 methods and principles. *Trends in cognitive sciences* **14**, 277-290 (2010).
- 29 7. Doyon, J. & Benali, H. Reorganization and plasticity in the adult brain during
30 learning of motor skills. *Current opinion in neurobiology* **15**, 161-167 (2005).

- 1 8. Tang, X. *et al.* A fully-automated subcortical and ventricular shape generation
2 pipeline preserving smoothness and anatomical topology. *Frontiers in*
3 *neuroscience* **12**, 321 (2018).
- 4 9. Thompson, P.M. *et al.* Mapping hippocampal and ventricular change in
5 Alzheimer disease. *Neuroimage* **22**, 1754-1766 (2004).
- 6 10. Wang, Y., Chan, T.F., Toga, A.W. & Thompson, P.M. Multivariate tensor-based
7 brain anatomical surface morphometry via holomorphic one-forms. in
8 *International conference on medical image computing and computer-assisted*
9 *intervention* 337-344 (Springer, 2009).
- 10 11. Wang, Y. *et al.* Surface-based TBM boosts power to detect disease effects on the
11 brain: an N= 804 ADNI study. *Neuroimage* **56**, 1993-2010 (2011).
- 12 12. Thompson, P.M. *et al.* ENIGMA and global neuroscience: A decade of large-scale
13 studies of the brain in health and disease across more than 40 countries.
14 *Translational psychiatry* **10**, 1-28 (2020).
- 15 13. Ge, T. *et al.* Multidimensional heritability analysis of neuroanatomical shape.
16 *Nature communications* **7**, 13291 (2016).
- 17 14. Ching, C.R. *et al.* Sex-dependent age trajectories of subcortical brain structures:
18 analysis of large-scale percentile models and shape morphometry. in *16th*
19 *International Symposium on Medical Information Processing and Analysis* Vol.
20 11583 44-54 (SPIE, 2020).
- 21 15. Tang, X. *et al.* Shape abnormalities of subcortical and ventricular structures in
22 mild cognitive impairment and Alzheimer's disease: detecting, quantifying, and
23 predicting. *Human brain mapping* **35**, 3701-3725 (2014).
- 24 16. Cong, S. *et al.* Surface-based morphometric analysis of hippocampal subfields in
25 mild cognitive impairment and Alzheimer's disease. in *2015 IEEE 58th*
26 *International Midwest Symposium on Circuits and Systems (MWSCAS)* 1-4 (IEEE,
27 2015).
- 28 17. Jamea, A.A. *et al.* Volumetric and Shape Analysis of the Subcortical Regions in
29 Schizophrenia Patients: A Pilot Study. *Journal of Clinical Imaging Science* **9**(2019).
- 30 18. Gutman, B.A. *et al.* A meta-analysis of deep brain structural shape and
31 asymmetry abnormalities in 2,833 individuals with schizophrenia compared with

- 1 3,929 healthy volunteers via the ENIGMA Consortium. *Human brain mapping* **43**,
2 352-372 (2022).
- 3 19. Kim, J.B., Suh, S.-I. & Kim, J.H. Volumetric and shape analysis of hippocampal
4 subfields in unilateral mesial temporal lobe epilepsy with hippocampal atrophy.
5 *Epilepsy research* **117**, 74-81 (2015).
- 6 20. Lu, Y. *et al.* The volumetric and shape changes of the putamen and thalamus in
7 first episode, untreated major depressive disorder. *NeuroImage: Clinical* **11**, 658-
8 666 (2016).
- 9 21. Ho, T.C. *et al.* Subcortical shape alterations in major depressive disorder:
10 Findings from the ENIGMA major depressive disorder working group. *Human*
11 *brain mapping* **43**, 341-351 (2022).
- 12 22. Ching, C.R. *et al.* Mapping subcortical brain alterations in 22q11. 2 deletion
13 syndrome: Effects of deletion size and convergence with idiopathic
14 neuropsychiatric illness. *American Journal of Psychiatry* **177**, 589-600 (2020).
- 15 23. Quigley, S.J. *et al.* Volume and shape analysis of subcortical brain structures and
16 ventricles in euthymic bipolar I disorder. *Psychiatry Research: Neuroimaging* **233**,
17 324-330 (2015).
- 18 24. Roalf, D.R. *et al.* Heritability of subcortical and limbic brain volume and shape in
19 multiplex-multigenerational families with schizophrenia. *Biological psychiatry* **77**,
20 137-146 (2015).
- 21 25. Styner, M. *et al.* Morphometric analysis of lateral ventricles in schizophrenia and
22 healthy controls regarding genetic and disease-specific factors. *Proceedings of*
23 *the National Academy of Sciences* **102**, 4872-4877 (2005).
- 24 26. Biton, A. *et al.* Polygenic architecture of human neuroanatomical diversity.
25 *bioRxiv*, 592337 (2019).
- 26 27. Zhao, B. *et al.* Heritability of regional brain volumes in large-scale neuroimaging
27 and genetic studies. *Cerebral Cortex* **29**, 2904-2914 (2018).
- 28 28. Roshchupkin, G.V. *et al.* Heritability of the shape of subcortical brain structures
29 in the general population. *Nature communications* **7**(2016).
- 30 29. Miller, K.L. *et al.* Multimodal population brain imaging in the UK Biobank
31 prospective epidemiological study. *Nature Neuroscience* **19**, 1523-1536 (2016).

- 1 30. Satizabal, C.L. *et al.* Genetic architecture of subcortical brain structures in 38,851
2 individuals. *Nature genetics* **51**, 1624-1636 (2019).
- 3 31. Vojinovic, D. *et al.* Genome-wide association study of 23,500 individuals
4 identifies 7 loci associated with brain ventricular volume. *Nature*
5 *communications* **9**, 1-11 (2018).
- 6 32. Hibar, D.P. *et al.* Novel genetic loci associated with hippocampal volume. *Nature*
7 *Communications* **8**, 13624 (2017).
- 8 33. Hibar, D.P. *et al.* Common genetic variants influence human subcortical brain
9 structures. *Nature* **520**, 224-229 (2015).
- 10 34. Van der Meer, D. *et al.* Brain scans from 21,297 individuals reveal the genetic
11 architecture of hippocampal subfield volumes. *Molecular psychiatry* **25**, 3053-
12 3065 (2020).
- 13 35. Zhao, B. *et al.* Genome-wide association analysis of 19,629 individuals identifies
14 variants influencing regional brain volumes and refines their genetic co-
15 architecture with cognitive and mental health traits. *Nature genetics* **51**, 1637-
16 1644 (2019).
- 17 36. Medland, S.E. *et al.* Ten years of enhancing neuro-imaging genetics through
18 meta-analysis: An overview from the ENIGMA Genetics Working Group. *Human*
19 *Brain Mapping* **43**, 292-299 (2022).
- 20 37. Elliott, L.T. *et al.* Genome-wide association studies of brain imaging phenotypes
21 in UK Biobank. *Nature* **562**, 210-216 (2018).
- 22 38. van der Meer, D. *et al.* Understanding the genetic determinants of the brain with
23 MOSTest. *Nature communications* **11**, 1-9 (2020).
- 24 39. Jansen, P.R. *et al.* Genome-wide meta-analysis of brain volume identifies
25 genomic loci and genes shared with intelligence. *Nature communications* **11**, 1-
26 12 (2020).
- 27 40. Smith, S.M. *et al.* An expanded set of genome-wide association studies of brain
28 imaging phenotypes in UK Biobank. *Nature neuroscience* **24**, 737-745 (2021).
- 29 41. Charani, B. *et al.* Baseline brain function in the preadolescents of the ABCD
30 Study. *Nature Neuroscience*, 1-11 (2021).
- 31 42. Gutman, B.A. *et al.* Medial demons registration localizes the degree of genetic
32 influence over subcortical shape variability: an N= 1480 meta-analysis. in 2015

1 *IEEE 12th International Symposium on Biomedical Imaging (ISBI) 1402-1406*
2 (IEEE, 2015).

3 43. Couvy-Duchesne, B. *et al.* Parsimonious model for mass-univariate vertexwise
4 analysis. *Journal of Medical Imaging* **9**, 052404 (2022).

5 44. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide
6 complex trait analysis. *The American Journal of Human Genetics* **88**, 76-82
7 (2011).

8 45. Zhao, B. *et al.* Common genetic variation influencing human white matter
9 microstructure. *Science* **372**(2021).

10 46. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from
11 polygenicity in genome-wide association studies. *Nature genetics* **47**, 291-295
12 (2015).

13 47. Watanabe, K., Taskesen, E., Bochoven, A. & Posthuma, D. Functional mapping
14 and annotation of genetic associations with FUMA. *Nature Communications* **8**,
15 1826 (2017).

16 48. Cruchaga, C. *et al.* GWAS of cerebrospinal fluid tau levels identifies risk variants
17 for Alzheimer's disease. *Neuron* **78**, 256-268 (2013).

18 49. de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: generalized
19 gene-set analysis of GWAS data. *PLoS Computational Biology* **11**, e1004219
20 (2015).

21 50. Zhao, B. *et al.* Genetic influences on the intrinsic and extrinsic functional
22 organizations of the cerebral cortex. *medRxiv* (2021).

23 51. Tissink, E. *et al.* Specificity and overlap in the genetic architectures of functional
24 and structural connectivity within cerebral resting-state networks. *bioRxiv*
25 (2022).

26 52. Gelernter, J. *et al.* Genome-wide association study of cocaine dependence and
27 related traits: FAM53B identified as a risk gene. *Molecular psychiatry* **19**, 717-
28 723 (2014).

29 53. Okbay, A. *et al.* Genetic variants associated with subjective well-being,
30 depressive symptoms, and neuroticism identified through genome-wide
31 analyses. *Nature Genetics* **48**, 624–633 (2016).

- 1 54. Nishikimi, M., Oishi, K., Tabata, H., Torii, K. & Nakajima, K. Segregation and
2 pathfinding of callosal axons through EphA3 signaling. *Journal of Neuroscience*
3 **31**, 16251-16260 (2011).
- 4 55. Day, B.W. *et al.* EphA3 maintains tumorigenicity and is a therapeutic target in
5 glioblastoma multiforme. *Cancer cell* **23**, 238-248 (2013).
- 6 56. Tissir, F. & Goffinet, A.M. Planar cell polarity signaling in neural development.
7 *Current opinion in neurobiology* **20**, 572-577 (2010).
- 8 57. Shin, J. *et al.* Global and regional development of the human cerebral cortex:
9 Molecular architecture and occupational aptitudes. *Cerebral Cortex* **30**, 4121-
10 4139 (2020).
- 11 58. Gaston-Massuet, C., Henderson, D.J., Greene, N.D. & Copp, A.J. Zic4, a zinc-finger
12 transcription factor, is expressed in the developing mouse nervous system.
13 *Developmental dynamics: an official publication of the American Association of*
14 *Anatomists* **233**, 1110-1115 (2005).
- 15 59. Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P.D. PANTHER version
16 14: more genomes, a new PANTHER GO-slim and improvements in enrichment
17 analysis tools. *Nucleic acids research* **47**, D419-D426 (2019).
- 18 60. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide
19 association studies, targeted arrays and summary statistics 2019. *Nucleic Acids*
20 *Research* **47**, D1005-D1012 (2018).
- 21 61. Stein, J.L. *et al.* Identification of common variants associated with human
22 hippocampal and intracranial volumes. *Nature Genetics* **44**, 552-561 (2012).
- 23 62. de Klein, N. *et al.* Brain expression quantitative trait locus and network analysis
24 reveals downstream effects and putative drivers for brain-related diseases.
25 *bioRxiv* (2021).
- 26 63. Linnér, R.K. *et al.* Genome-wide association analyses of risk tolerance and risky
27 behaviors in over 1 million individuals identify hundreds of loci and shared
28 genetic influences. *Nature Genetics* **51**, 245-257 (2019).
- 29 64. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights
30 into the genetic etiology of tobacco and alcohol use. *Nature genetics* **51**, 237-244
31 (2019).

- 1 65. Kichaev, G. *et al.* Leveraging polygenic functional enrichment to improve GWAS
2 power. *The American Journal of Human Genetics* **104**, 65-75 (2019).
- 3 66. Sakornsakolpat, P. *et al.* Genetic landscape of chronic obstructive pulmonary
4 disease identifies heterogeneous cell-type and phenotype associations. *Nature*
5 *genetics* **51**, 494-505 (2019).
- 6 67. Wang, H. *et al.* Genome-wide interaction analysis of pathological hallmarks in
7 Alzheimer's disease. *Neurobiology of aging* **93**, 61-68 (2020).
- 8 68. Xiang, B. *et al.* The role of genes affected by human evolution marker GNA13 in
9 schizophrenia. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*
10 **98**, 109764 (2020).
- 11 69. Hirayama-Kurogi, M. *et al.* Downregulation of GNA13-ERK network in prefrontal
12 cortex of schizophrenia brain identified by combined focused and targeted
13 quantitative proteomics. *Journal of Proteomics* **158**, 31-42 (2017).
- 14 70. Düster, R., Ji, Y., Pan, K.-T., Urlaub, H. & Geyer, M. Functional characterization of
15 the human Cdk10/Cyclin Q complex. *Open Biology* **12**, 210381 (2022).
- 16 71. Vujkovic, M. *et al.* Discovery of 318 new risk loci for type 2 diabetes and related
17 vascular outcomes among 1.4 million participants in a multi-ancestry meta-
18 analysis. *Nature genetics* **52**, 680-691 (2020).
- 19 72. Shane, B. *et al.* The 677C→ T variant of MTHFR is the major genetic modifier of
20 biomarkers of folate status in a young, healthy Irish population. *The American*
21 *journal of clinical nutrition* **108**, 1334-1341 (2018).
- 22 73. Vuckovic, D. *et al.* The polygenic and monogenic basis of blood traits and
23 diseases. *Cell* **182**, 1214-1231. e11 (2020).
- 24 74. Bigdeli, T.B. *et al.* Genome-wide association studies of schizophrenia and bipolar
25 disorder in a diverse cohort of US veterans. *Schizophrenia bulletin* **47**, 517-529
26 (2021).
- 27 75. Periyasamy, S. *et al.* Association of schizophrenia risk with disordered niacin
28 metabolism in an Indian genome-wide association study. *JAMA psychiatry* **76**,
29 1026-1034 (2019).
- 30 76. Lago, S.G. & Bahn, S. The druggable schizophrenia genome: from repurposing
31 opportunities to unexplored drug targets. *NPJ genomic medicine* **7**, 1-13 (2022).

- 1 77. Wu, Y. *et al.* Multi-trait analysis for genome-wide association study of five
2 psychiatric disorders. *Translational psychiatry* **10**, 1-11 (2020).
- 3 78. Cai, N. *et al.* Minimal phenotyping yields genome-wide association signals of low
4 specificity for major depression. *Nature Genetics* **52**, 437-447 (2020).
- 5 79. Nagel, M. *et al.* Meta-analysis of genome-wide association studies for
6 neuroticism in 449,484 individuals identifies novel genetic loci and pathways.
7 *Nature Genetics* **50**, 920 (2018).
- 8 80. Savage, J.E. *et al.* Genome-wide association meta-analysis in 269,867 individuals
9 identifies new genetic and functional links to intelligence. *Nature Genetics* **50**,
10 912-919 (2018).
- 11 81. Lee, J.J. *et al.* Gene discovery and polygenic prediction from a genome-wide
12 association study of educational attainment in 1.1 million individuals. *Nature*
13 *Genetics* **50**, 1112–1121 (2018).
- 14 82. de la Fuente, J., Davies, G., Grotzinger, A.D., Tucker-Drob, E.M. & Deary, I.J. A
15 general dimension of genetic sharing across diverse cognitive traits inferred from
16 molecular data. *Nature Human Behaviour* **5**, 49-58 (2021).
- 17 83. Jansen, P.R. *et al.* Genome-wide analysis of insomnia in 1,331,010 individuals
18 identifies new risk loci and functional pathways. *Nature Genetics* **51**, 394-403
19 (2019).
- 20 84. Morris, A.P. *et al.* Trans-ethnic kidney function association study reveals putative
21 causal genes and effects on kidney-specific disease aetiologies. *Nature*
22 *communications* **10**, 1-14 (2019).
- 23 85. Afsar, B. *et al.* Brain-kidney cross-talk: Definition and emerging evidence.
24 *European journal of internal medicine* **36**, 7-12 (2016).
- 25 86. Chang, C.Y. *et al.* Cognitive impairment and hippocampal atrophy in chronic
26 kidney disease. *Acta Neurologica Scandinavica* **136**, 477-485 (2017).
- 27 87. Vemuri, P. *et al.* Association of kidney function biomarkers with brain MRI
28 findings: the BRINK study. *Journal of Alzheimer's Disease* **55**, 1069-1082 (2017).
- 29 88. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases
30 and traits. *Nature Genetics* **47**, 1236-1241 (2015).

- 1 89. Choi, Y.J., Lee, E.J. & Lee, J.E. The fornix: functional anatomy, normal
2 neuroimaging, and various pathological conditions. *Investigative Magnetic*
3 *Resonance Imaging* **25**, 59-75 (2021).
- 4 90. Plessen, K.J. *et al.* Hippocampus and amygdala morphology in attention-
5 deficit/hyperactivity disorder. *Archives of general psychiatry* **63**, 795-807 (2006).
- 6 91. Vilor-Tejedor, N. *et al.* Aging-Dependent Genetic Effects Associated to ADHD
7 Predict Longitudinal Changes of Ventricular Volumes in Adulthood. *Frontiers in*
8 *psychiatry* **11**, 574 (2020).
- 9 92. Pergola, G., Selvaggi, P., Trizio, S., Bertolino, A. & Blasi, G. The role of the
10 thalamus in schizophrenia from a neuroimaging perspective. *Neuroscience &*
11 *Biobehavioral Reviews* **54**, 57-75 (2015).
- 12 93. Csernansky, J.G. *et al.* Abnormalities of thalamic volume and shape in
13 schizophrenia. *American Journal of Psychiatry* **161**, 896-902 (2004).
- 14 94. Harrison, P.J. The hippocampus in schizophrenia: a review of the
15 neuropathological evidence and its pathophysiological implications.
16 *Psychopharmacology* **174**, 151-162 (2004).
- 17 95. Franke, B. *et al.* Genetic influences on schizophrenia and subcortical brain
18 volumes: large-scale proof of concept. *Nature Neuroscience* **19**, 420-431 (2016).
- 19 96. Apostolova, L.G. *et al.* Hippocampal atrophy and ventricular enlargement in
20 normal aging, mild cognitive impairment and Alzheimer's disease. *Alzheimer*
21 *disease and associated disorders* **26**, 17 (2012).
- 22 97. Nikulin, V.V. *et al.* Anticipatory activity in the human thalamus is predictive of
23 reaction times. *Neuroscience* **155**, 1275-1283 (2008).
- 24 98. Niu, J. *et al.* Neural Dysconnectivity in the Hippocampus Correlates With White
25 Matter Lesions and Cognitive Measures in Patients With Coronary Artery
26 Disease. *Frontiers in Aging Neuroscience* **14**(2022).
- 27 99. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C.A. & Smoller, J.W. Polygenic prediction via
28 Bayesian regression and continuous shrinkage priors. *Nature Communications*
29 **10**, 1776 (2019).
- 30 100. Hirabayashi, N. *et al.* Association between diabetes and hippocampal atrophy in
31 elderly Japanese: the Hisayama study. *Diabetes care* **39**, 1543-1549 (2016).

- 1 101. Ho, N., Sommers, M.S. & Lucki, I. Effects of diabetes on hippocampal
2 neurogenesis: links to cognition and depression. *Neuroscience & Biobehavioral*
3 *Reviews* **37**, 1346-1362 (2013).
- 4 102. Cooke, G.E., Mullally, S., Correia, N., O'Mara, S.M. & Gibney, J. Hippocampal
5 volume is decreased in adults with hypothyroidism. *Thyroid* **24**, 433-440 (2014).
- 6 103. Zhang, W. *et al.* Grey matter abnormalities in untreated hyperthyroidism: a
7 voxel-based morphometry study using the DARTEL approach. *European journal*
8 *of radiology* **83**, e43-e48 (2014).
- 9 104. Oatridge, A. *et al.* Changes in brain size with treatment in patients with hyper-or
10 hypothyroidism. *American journal of neuroradiology* **23**, 1539-1544 (2002).
- 11 105. Krämer, J. *et al.* Early and degressive putamen atrophy in multiple sclerosis.
12 *International journal of molecular sciences* **16**, 23195-23209 (2015).
- 13 106. Boscheron, J. *et al.* Insights on the Relationship Between Hippocampal
14 Connectivity and Memory Performances at the Early Stage of Multiple Sclerosis.
15 *Frontiers in neurology* **12**, 667531 (2021).
- 16 107. Barekatin, M. *et al.* The relationship between regional brain volumes and the
17 extent of coronary artery disease in mild cognitive impairment. *Journal of*
18 *Research in Medical Sciences: The Official Journal of Isfahan University of*
19 *Medical Sciences* **19**, 739 (2014).
- 20 108. Roddy, D.W. *et al.* The hippocampus in depression: more than the sum of its
21 parts? Advanced hippocampal substructure segmentation in depression.
22 *Biological Psychiatry* **85**, 487-497 (2019).
- 23 109. Xu, J. *et al.* Left fimbria atrophy is associated with hippocampal metabolism in
24 female major depressive disorder patients. in *2018 40th Annual International*
25 *Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 1136-
26 1139 (IEEE, 2018).
- 27 110. Shao, X. *et al.* Uric acid induces cognitive dysfunction through hippocampal
28 inflammation in rodents and humans. *Journal of Neuroscience* **36**, 10990-11005
29 (2016).
- 30 111. Garfield, V. *et al.* HbA1c and brain health across the entire glycaemic spectrum.
31 *Diabetes, Obesity and Metabolism* **23**, 1140-1149 (2021).

- 1 112. Kurki, M.I. *et al.* FinnGen: Unique genetic insights from combining isolated
2 population and national health register data. *medRxiv* (2022).
- 3 113. Sullivan, P.F. *et al.* Psychiatric genomics: an update and an agenda. *American*
4 *Journal of Psychiatry* **175**, 15-27 (2017).
- 5 114. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference
6 across the human phenome. *elife* **7**(2018).
- 7 115. Zhao, Q., Wang, J., Hemani, G., Bowden, J. & Small, D.S. Statistical inference in
8 two-sample summary-data Mendelian randomization using robust adjusted
9 profile score. *The Annals of Statistics* **48**, 1742-1769 (2020).
- 10 116. Wang, J. *et al.* Causal inference for heritable phenotypic risk factors using
11 heterogeneous genetic instruments. *PLoS genetics* **17**, e1009575 (2021).
- 12 117. Ye, T., Shao, J. & Kang, H. Debiased inverse-variance weighted estimator in two-
13 sample summary-data Mendelian randomization. *The Annals of statistics* **49**,
14 2079-2100 (2021).
- 15 118. Consortium, C.-D.G.o.t.P.G. Identification of risk loci with shared effects on five
16 major psychiatric disorders: a genome-wide analysis. *The Lancet* **381**, 1371-1379
17 (2013).
- 18 119. Nestor, S.M. *et al.* Ventricular enlargement as a possible measure of Alzheimer's
19 disease progression validated using the Alzheimer's disease neuroimaging
20 initiative database. *Brain* **131**, 2443-2454 (2008).
- 21 120. Adamo, S. *et al.* Ventricular expansion, white matter hyperintensities, and global
22 cognition in Alzheimer's disease and normal aging. *medRxiv* (2020).
- 23 121. Ferrarini, L. *et al.* Shape differences of the brain ventricles in Alzheimer's disease.
24 *Neuroimage* **32**, 1060-1069 (2006).
- 25 122. Luo, Q. *et al.* Association of a schizophrenia-risk nonsynonymous variant with
26 putamen volume in adolescents: a voxelwise and genome-wide association
27 study. *JAMA psychiatry* **76**, 435-445 (2019).
- 28 123. Gardener, H., Wright, C.B., Rundek, T. & Sacco, R.L. Brain health and shared risk
29 factors for dementia and stroke. *Nature Reviews Neurology* **11**, 651-657 (2015).
- 30 124. Broce, I.J. *et al.* Dissecting the genetic relationship between cardiovascular risk
31 factors and Alzheimer's disease. *Acta neuropathologica* **137**, 209-226 (2019).

- 1 125. Li, J. & Fei, G.-H. The unique alterations of hippocampus and cognitive
2 impairment in chronic obstructive pulmonary disease. *Respiratory research* **14**,
3 1-9 (2013).
- 4 126. Wang, D. & Gu, J. VASC: dimension reduction and visualization of single-cell RNA-
5 seq data by deep variational autoencoder. *Genomics, proteomics &*
6 *bioinformatics* **16**, 320-331 (2018).
- 7 127. Kirchler, M. *et al.* transferGWAS: GWAS of images using deep transfer learning.
8 *Bioinformatics* **38**, 3621-3628 (2022).
- 9 128. Vilor-Tejedor, N. *et al.* Genetic Influences on Hippocampal Subfields: An
10 Emerging Area of Neuroscience Research. *Neurology Genetics* **7**(2021).
- 11 129. Sämann, P.G. *et al.* FreeSurfer-based segmentation of hippocampal subfields: A
12 review of methods and applications, with a novel quality control procedure for
13 ENIGMA studies and other collaborative efforts. *Human brain mapping* **43**, 207-
14 233 (2022).
- 15 130. Nakamura, Y. *et al.* The association between amygdala subfield-related
16 functional connectivity and stigma reduction 12 months after social contacts: a
17 functional neuroimaging study in a subgroup of a randomized controlled trial.
18 *Frontiers in human neuroscience* **14**, 356 (2020).
- 19 131. Alfaro-Almagro, F. *et al.* Image processing and Quality Control for the first 10,000
20 brain imaging datasets from UK Biobank. *NeuroImage* **166**, 400-424 (2018).
- 21 132. Casey, B.J. *et al.* The adolescent brain cognitive development (ABCD) study:
22 imaging acquisition across 21 sites. *Developmental cognitive neuroscience* **32**,
23 43-54 (2018).
- 24 133. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic
25 data. *Nature* **562**, 203-209 (2018).
- 26 134. Chen, C.-Y. *et al.* Improved ancestry inference using weights from external
27 reference panels. *Bioinformatics* **29**, 1399-1406 (2013).
- 28 135. Avants, B.B. *et al.* A reproducible evaluation of ANTs similarity metric
29 performance in brain image registration. *Neuroimage* **54**, 2033-2044 (2011).
- 30 136. Jiang, L. *et al.* A resource-efficient tool for mixed model association analysis of
31 large-scale data. *Nature genetics* **51**, 1749 (2019).

1 137. Dey, R. *et al.* An efficient and accurate frailty model approach for genome-wide
2 survival association analysis controlling for population structure and relatedness
3 in large-scale biobanks. *bioRxiv* (2020).

4

5 **METHODS**

6 **Shape features and imaging datasets.**

7 The raw structural MRI data from the UKB and ABCD studies were used in this study. The
8 UKB study's ethics approval was obtained from the North West Multicentre Research
9 Ethics Committee (approval number: 11/NW/0382). The procedures of the ABCD study
10 were approved by the institutional review boards at ABCD collection sites (approval
11 numbers: 201708123 and 160091). The image collection and processing procedures can
12 be found in Alfaro-Almagro, et al. ¹³¹ for the UKB study and Casey, et al. ¹³² for the ABCD
13 study.

14

15 The shape feature generation pipeline was detailed in the **Supplementary Note** and an
16 overview of the procedures and examples were virtualized in **Figures S1 and S42**. Briefly,
17 we focused on 8 ventricular and subcortical structures, including the left/right lateral
18 ventricles, nucleus accumbens, amygdala, caudate, hippocampus, pallidum, putamen,
19 and thalamus. We constructed the vertex-wise map for 7 different shape statistics,
20 including radial distance, mTBM1, mTBM2, mTBM3, determinant, eigenvalue1, and
21 eigenvalue2. For each shape statistics, we aggregated these vertex-wise data by 1) taking
22 the mean across all vertices for each structure; and 2) generating the top-ranked PCs for
23 each structure (left and right hemispheres separately). Intuitively, the purpose of these
24 PCs was to capture global and local variations within the vertex-wise representation of
25 brain structures. Typically, the first one or two PCs represented the global patterns, which
26 were similar to the mean values. Other PCs, however, captured local variations that mean
27 or top-ranked PCs missed. See **Figure 1C** for an illustration. Additionally, we segmented
28 the hippocampus and calculated the mean of shape statistics for each of the following
29 subfields: the left/right CA1, CA3, fimbria, HATA, hippocampal tail, presubiculum, and
30 subiculum. In total, we had 1,330 (210 mean values and 1,120 PCs) shape features. Based
31 on the UKB revisit data, we calculated the reproducibility (ICC) and selected those with

1 ICC > 0.5, resulting in a final set of 457 shape features (363 at the structure-level and 94
2 at the subfield-level) for genetic analysis. See **Table S1** for their names and descriptions.

3
4 We analyzed the above 457 reproducible traits in the following datasets: 1) the white
5 British discovery dataset, where the data were from white British subjects in UKB phases
6 1 to 3 imaging data (average $n = 32,631$, released up through 2020); 2) the UKB European
7 validation dataset, which included White individuals in the newly released UKB phase 4
8 data and the UKB non-British white individuals in phases 1 to 3 data (UKBE, removed the
9 relatives of the discovery sample, average $n = 4,596$); 3) the UKB non-European validation
10 dataset that consisting of non-White subjects in the UKB phases 1 to 4 data (UKBNE,
11 average $n = 1,224$); 4) the UKB first revisit dataset (average $n = 2,788$); and 5) the ABCD
12 dataset (average $n = 8,496$). The average age (at imaging) of all UKB subjects was 64.2
13 (standard error = 7.73), 51.6% were females; the average age for all ABCD students was
14 9.93 (standard error = 0.62), 48.2% were females. Self-reported ethnicity (Data-Field
15 21000) was used to assign ancestry in UKB, whose accuracy was verified in Bycroft, et al.
16 ¹³³. We assigned ancestry to the ABCD participants based on self-reported ethnic
17 background combined with SNPweights¹³⁴ inferences, see Zhao, et al. ⁴⁵ for more
18 information.

19
20 **Heritability and GWAS analysis.** We downloaded UKB imputed genetics data¹³³ (Data-
21 Category 263) and locally imputed the ABCD genetic data using the Michigan Imputation
22 Server (<https://imputationserver.sph.umich.edu/>) with the 1000 Genomes Phase 3
23 (Version 5) reference panel⁴⁵. In both UKB and ABCD, the following quality controls were
24 performed on imaging subjects with genetics data: 1) removing subjects with > 10%
25 missing genotypes; 2) removing genetic variants with minor allele frequency (MAF) < 0.01;
26 3) removing genetic variants with missing genotype rate > 10%; 4) removing variants that
27 failed the Hardy-Weinberg test at 1×10^{-7} significance level; and 5) removing genetic
28 variants with imputation INFO score < 0.8. We used GCTA⁴⁴ to estimate SNP-based
29 heritability with all autosomal SNPs in the white British discovery dataset (average $n =$
30 32,631). The adjusted covariates include age (at imaging), age-squared, sex, age-sex
31 interaction, age-squared-sex interaction, imaging site, the top 40 genetic PCs¹³³,
32 volumetric scaling, head motion, head motion-squared, brain position, and brain position-

1 squared^{37,40}. For the 363 structure-level shape features, the genetic variance estimates
2 were also extracted from the GCTA and were compared before and after additionally
3 adjusting for the regional volumetric measurements. Specifically, for each of the 7
4 subcortical structures, we additionally controlled the corresponding FIRST subcortical
5 volumes (Category 1102). For the lateral ventricles, we adjusted for the regional volumes
6 estimated from ANTs¹³⁵. The genome-wide association analysis was conducted with a
7 linear mixed effect model using fastGWA¹³⁶, adjusted for the same covariates as the
8 GCTA. We also conducted GWAS separately in validation datasets and adjusted for only
9 the top 10 genetic PCs rather than the top 40. In the ABCD dataset, we performed
10 validation GWAS separately for African American, European, and Hispanic subjects,
11 removing one subject randomly from each twin pair⁴⁵. In all analyses, we removed values
12 greater than five times the median absolute deviation from the median for each
13 continuous phenotype or covariate variable.

14

15 We used FUMA⁴⁷ (version v1.3.8) to characterize genomic loci with European LD files from
16 the 1000 Genomes. To define the LD boundaries, FUMA used independent significant
17 variants, which were genetic variants whose P -value smaller than the predefined
18 threshold (here was 1.09×10^{-10} , $5 \times 10^{-8}/457$) and were independent of other significant
19 variants ($LD\ r^2 < 0.6$). FUMA then constructed LD blocks for these independent significant
20 variants by tagging all variants in LD ($r^2 \geq 0.6$) with at least one independent significant
21 variant with a $MAF \geq 0.0005$. There may have been variants from the 1000 Genomes
22 reference panel that were not included in the GWAS. Moreover, within these significant
23 variants, we defined independent lead variants as those that were independent from
24 each other ($LD\ r^2 < 0.1$). In the case of close LD blocks (<250 kb based on the closest
25 boundary variants of LD blocks), they were merged into one genomic locus. Independent
26 significant variants and all the variants in LD with them ($r^2 \geq 0.6$) were looked up on the
27 NHGRI-EBI GWAS catalog (version e104_2021-09-15) to search for associations ($P < 9 \times$
28 10^{-6}) reported for any traits. For selected colocalized index variants, we also performed
29 association analysis in vertex-wise data to illustrate local association patterns. The
30 significance threshold was set to be $0.05/\text{number of vertices in each structure}$. We
31 adjusted for the same set of covariates as in the above genome-wise analysis.

32

1 Gene-based testing was performed using UKB white British discovery GWAS summary
2 statistics for 18,796 protein-coding genes via MAGMA⁴⁹ (version 1.08). We used default
3 MAGMA settings with zero window size around each gene. We also conducted functional
4 annotation and mapping analysis in FUMA, where genetic variants were annotated with
5 their genomic functionality and then were mapped to 35,808 candidate genes using
6 positional, eQTL, and 3D chromatin interaction information. Brain-related tissues/cells
7 were selected in all options and the default values were used for all other parameters in
8 FUMA. LDSC⁸⁸ (version 1.0.1) was used to infer genetic correlations. LD scores were
9 computed using 1000 Genomes European data provided by LDSC. The major
10 histocompatibility complex (MHC) region was removed from the HapMap3 variants.

11

12 **Polygenic risk scores on UKB non-imaging subjects.** As a first step, we constructed a PRS
13 based on PRS-CS⁹⁹ for each shape feature. We input GWAS summary statistics from the
14 UKB white British discovery dataset (average $n = 32,631$), and randomly selected 1,500
15 subjects from the UKB European validation dataset as validation. We used all default
16 parameters in the PRS-CS software (<https://github.com/getian107/PRScs>) and generated
17 the PRS for all non-imaging individuals in the UKB study (removing relatives of the UKB
18 imaging individuals). The second step was to explore the associations with 276
19 phenotypes across various trait domains using these non-imaging UKB individuals,
20 including 24 mental health traits (Category 100060), 5 cognitive traits (Category 100026),
21 12 physical activity traits (Category 100054), 6 electronic device use traits (Category
22 100053), 8 sun exposure traits (Category 100055), 3 sexual factor traits (Category
23 100056), 3 social support traits (Category 100061), 12 family history of diseases (Category
24 100034), 21 diet traits (Category 100052), 9 alcohol drinking traits (Category 100051), 6
25 smoking traits (Category 100058), 34 blood biochemistry biomarkers (Category 17518), 3
26 blood pressure traits (Category 100011), 3 spirometry traits (Category 100020), 32 early
27 life factors (Categories 135, 100033, 100034, and 100072), 9 greenspace and coastal
28 proximity (Category 151), 2 hand grip strength (Category 100019), 13 residential air
29 pollution traits (Category 114), 5 residential noise pollution traits (Category 115), 2 body
30 composition traits by impedance (Category 100009), 4 health and medical history traits
31 (Category 100036), 3 female specific factors (Category 100069), 1 education trait

1 (Category 100063), and 57 curated disease phenotypes based on Dey, et al.¹³⁷ (**Table**
2 **S11**).

3

4 We used a discovery-validation design and repeated our analysis in two independent
5 samples: 1) the discovery sample, which consisted of 70% randomly selected independent
6 UKB non-imaging subjects of white British ancestry (average $n = 202,405$) and 2) the
7 validation sample, including the left 30% independent UKB white British non-imaging
8 subjects (average $n = 86,736$), white non-British non-imaging subjects (average $n =$
9 $20,746$), and non-white non-imaging subjects (average $n = 21,587$). The adjusted
10 covariates included age, age-squared, sex, age-sex interaction, age-squared-sex
11 interaction, as well as 40 genetic PCs. We reported P values from the two-sided t test and
12 prioritized on the results that were 1) significant after Bonferroni correction in the
13 discovery dataset, 2) significant at nominal significance level (0.05) in the validation
14 dataset; and 3) the regression coefficients had matched directions in the discovery and
15 validation datasets.

16

17 **MR analysis with clinical endpoints.** We examined the genetic causal relationships
18 between the 457 shape features and 288 clinical endpoints, where 275 of them were from
19 FinnGen (https://www.finngen.fi/en/access_results) and 13 were from the PGC
20 (<https://pgc.unc.edu/>). For FinnGen, we selected 275 clinical traits from the latest release
21 (R7) and with more than 5,000 cases. To reduce the potential influence of sample overlap,
22 we avoid PGC studies that have been using data solely from the UKB study. More detailed
23 information can be found in **Table S13**.

24

25 Before running MR methods, we performed standard preprocessing steps on GWAS data.
26 The genetic variants were first selected with significance threshold 5×10^{-8} in the
27 exposure GWAS. To ensure the genetic variants used in the MR were independent, LD
28 clumping was implemented using $r^2 = 0.01$, window size = 10,000, and the 1000
29 Genomes European ancestry data being the reference panel. The harmonization
30 procedure in the TwoSampleMR package (<https://mrcieu.github.io/TwoSampleMR/>)
31 helped us infer the correct allele alignment, therefore the selected variants on the

1 exposure and the reported effect of the same variant on the outcome corresponded to
2 the same allele.

3
4 We tested 14 MR methods¹¹⁴⁻¹¹⁷, including the IVW, IVW multiple random effect model,
5 IVW fixed effect model, MR-Egger, Simple Median, Weighted Median, Penalized
6 Weighted Median, Simple Mode, Simple Mode (NOME), Weighed Mode, Weighted Mode
7 (NOME), DIVW, GRAPPLE, and MR-RAPS. To ensure the reliability of our results, we
8 compared estimates from different methods. After running MR analysis on all pairs
9 between brain shape features and clinical endpoints, we used two steps to select the
10 significant causal results. We first removed all the estimated causal associations with less
11 than 6 variants used in the MR analysis. Then for the remaining estimates, we performed
12 Bonferroni adjustments for multiple testing. Besides comparing estimates across
13 different MR methods, we also tested potential violations in MR analysis to make sure
14 our results were reliable. For example, a significant intercept of MR Egger regression
15 indicated the presence of horizontal pleiotropy. All reported results have passed these
16 tests.

17 18 **Code availability**

19 We made use of publicly available software and tools. The pipelines used in shape feature
20 extractions can be found at https://www.nitrc.org/frs/?group_id=1461. The codes used
21 in other parts of the paper are available upon reasonable request.

22 23 **Data availability**

24 The individual-level data used in the present study can be applied from the UKB
25 (<https://www.ukbiobank.ac.uk/>) and ABCD (<https://abcdstudy.org/>) studies. Our GWAS
26 summary statistics will be shared on Zenodo and at the BIG-KP <https://bigkp.org/>. Our
27 GWAS results will also be available via the interactive web browser at
28 <http://165.227.78.169:443/>.

29 30 **Figure Legends**

31 **Fig. 1 Illustrations of brain structures and their shape features.**

1 **(A)** In the left panel, we illustrate the 8 ventricular and subcortical structures, including
2 the lateral ventricles, nucleus accumbens, amygdala, caudate, hippocampus, pallium,
3 putamen, and thalamus. In the right panel, we present the 7 hippocampal subfields,
4 including the cornu ammonis 1 (CA1), CA3, fimbria, hippocampus-amygdala-transition-
5 area (HATA), hippocampal tail, presubiculum, and subiculum. **(B)** We illustrate the spatial
6 pattern of radial distance in the vertex-wise maps of 8 ventricular and subcortical
7 structures. These maps were generated by averaging the data from 500 randomly
8 selected UKB subjects. See Figure S2 for additional maps of other 6 shape statistics
9 (mTBM1, mTBM2, mTBM3, determinant, eigenvalue1, and eigenvalue2). L, left; R, right.
10 **(C)** Comparison between the mean radial distance (RD) and structure-specific RD principal
11 components (PCs) on the lateral ventricles. (I) illustrates an example vertex-wise RD map
12 within the lateral ventricles after inter-subject centralization; (II) shows the residual RD
13 map after removing the within-subject mean RD; In (III) and (IV), instead of removing the
14 within-subject mean as in (II), we removed the top one and 10 RD PCs, respectively. (V)
15 illustrates the standard deviation across the vertices in residual RD map for each subject
16 in the UKB ($n = 32746$). Comparing (II) with (III), the top one PC can capture more spatial
17 variations than the mean RD and thus reduce the standard deviations of residuals in (V).
18 (V) also shows that the standard deviations are further reduced after removing the top
19 10 RD PCs (in (IV)), suggesting that additional PCs can account for more local spatial
20 variations that are ignored by the mean RD (in (III)) or top one RD PC (in (II)).

21

22 **Fig. 2 Genetic variance and the associated genomic regions of shape features.**

23 **(A-B)** The dots represent the genetic variance estimates of shape features. We compare
24 the original (marginal) genetic variance estimates before adjusting for corresponding
25 volumetric measurements (x axis) and the conditional genetic variance estimates after
26 adjusting for volumes (y axis). The results for the mean values and top one PCs (PC1s) are
27 displayed in the left panel **(A)**, and results for the other PCs are displayed in the right panel
28 **(B)**. We show the significant estimates after controlling the false discovery rate of multiple
29 testing at 5% level. Based on these results, we find that volumes can partially capture the
30 genetic influences on PC1s and mean features (in A), while they cannot capture the
31 majority of genetic influences on other PCs (in B). **(C)** Ideogram of 60 genomic regions

1 influencing shape features ($P < 1.09 \times 10^{-10}$), 19 of which were not identified by the
2 corresponding volumetric measurements. The colors of dots represent the different
3 structures (and hippocampal subfields). Each signal dot indicates that at least one of the
4 shape features of this brain structure is associated with the genomic region. The name of
5 genomic regions replicated in more than one validation datasets or in one validation
6 dataset at the nominal significance level were highlighted in red and brown labels,
7 respectively.

8

9 **Fig. 3 Genetic loci associated with both shape features and other complex traits.**

10 **(A)** In the 17q24.1 region, we observed shared genetic influences ($LD\ r^2 \geq 0.6$) between
11 shape features (e.g., Vent_Right_RD_PC9, index variant rs62072157) and brain white
12 matter microstructure (e.g., FX_MD, index variant rs35122942). Vent_Right_RD_PC9, the
13 ninth PC of the radial distance in right lateral ventricle; FX_MD, the mean diffusivity in the
14 fornix tract (column and body of fornix). **(B)** In the 2q24.2 region, we observed shared
15 genetic influences ($LD\ r^2 \geq 0.6$) between shape features (e.g., Hipp_Right_Eigen1_PC6,
16 index variant rs1014445) and cognitive traits (e.g., intelligence, index variant rs2268894).
17 Hipp_Right_Eigen1_PC6, the sixth PC of the eigenvalue1 in right hippocampus. **(C)** In the
18 10q26.13 region, we observed shared genetic influences ($LD\ r^2 \geq 0.6$) between shape
19 features (e.g., Sub_CA3_Right_RD_Mean, index variant rs10901814) and kidney function
20 biomarker (eGRF, index variant rs4962691). Sub_CA3_Right_RD_Mean, the mean radial
21 distance in the CA3 subfield of right hippocampus; eGRF, estimated glomerular filtration
22 rate. In **(D)**, we illustrate the signed $-\log_{10}(P\text{-value})$ of associations between the
23 rs4962691 variant and vertex-wise data of the hippocampus and lateral ventricles for 4
24 shape statistics, including radial distance, mTBM1, mTBM2, and mTBM3. It was observed
25 that the pattern of genetic effects varied by subregion within each structure.

26

27 **Fig. 4 Selected genetic correlations with complex traits and diseases.**

28 **(A)** We illustrate pairwise genetic correlations between shape features (x axis) and other
29 complex traits and diseases (y axis) estimated by LDSC. The asterisks highlight significant
30 pairs after controlling the FDR at 5% level. The colors represent the genetic correlation

1 estimates. See Table S1 for descriptions of the shape features. **(B)** We illustrate brain
2 structures whose shape features are genetically related to brain disorders, such as
3 schizophrenia, attention-deficit/hyperactivity disorder (ADHD), and anorexia nervosa. **(C)**
4 We illustrate brain structures whose shape features are genetically related to cognitive
5 traits, education, and behavioral traits. **(D)** We illustrate brain structures whose shape
6 features are genetically related to cardiovascular diseases, such as coronary artery
7 disease and hypertension.

8

9 **Fig. 5 Mendelian randomization analysis with clinical outcomes.**

10 **(A)** Significant causal genetic links from brain shape features to clinical endpoints ($P < 3.92$
11 $\times 10^{-8}$). **(B)** Significant causal genetic links from clinical endpoints to brain shape features
12 ($P < 1.12 \times 10^{-8}$). In both A and B, the chord plot in the middle display each causal pair. The
13 first level (out) circle indicates each disease or brain structure. The second level circle
14 indicates the specific diseases with each disease category, or shape features within each
15 brain structure. The third level circle on the shape feature side indicates which diseases
16 the shape feature links to. In addition, the circle plots on the left- and right-hand sides
17 display the number of significant pairs for each exposure and outcome variable,
18 respectively. FG, FinnGen; PGC, Psychiatric Genomics Consortium; COPD, chronic
19 obstructive pulmonary disease; Sub, subfield of hippocampus; HP, hippocampal tail; and
20 CA, cornu ammonis. See Table S1 for descriptions of the shape features.

A Ventricular and subcortical structures

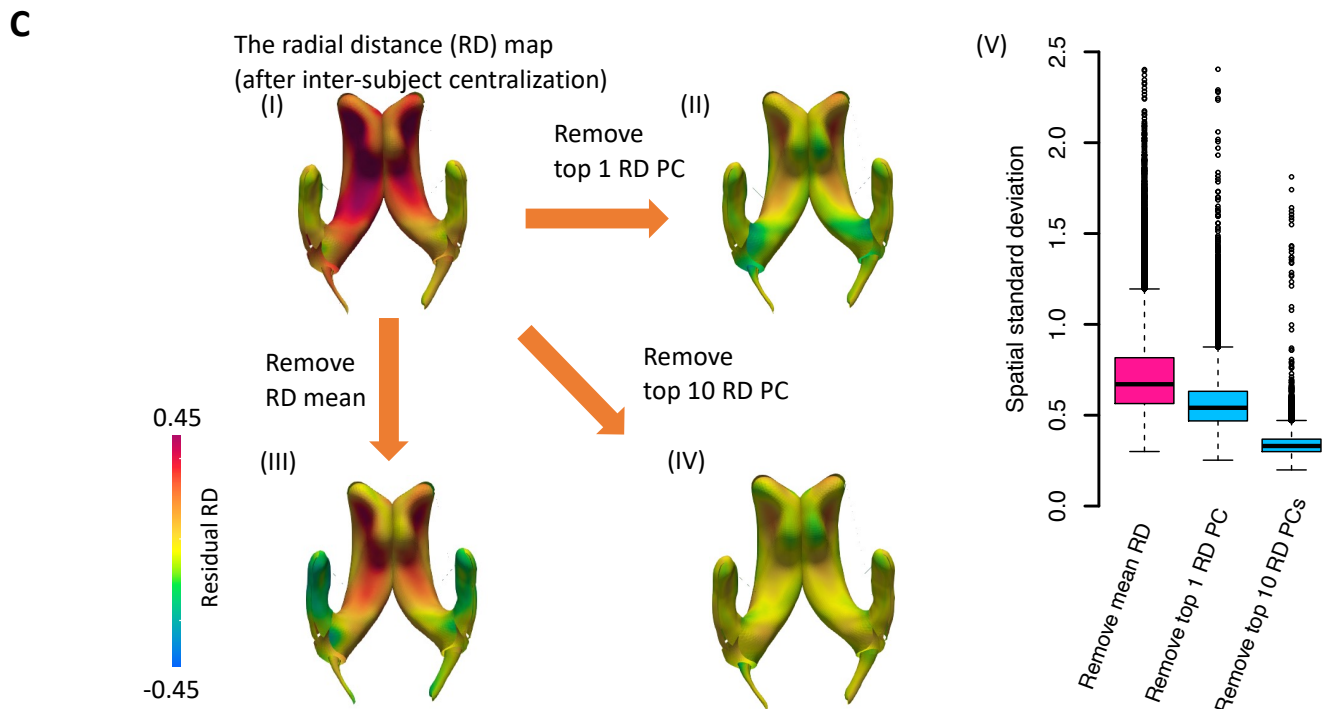
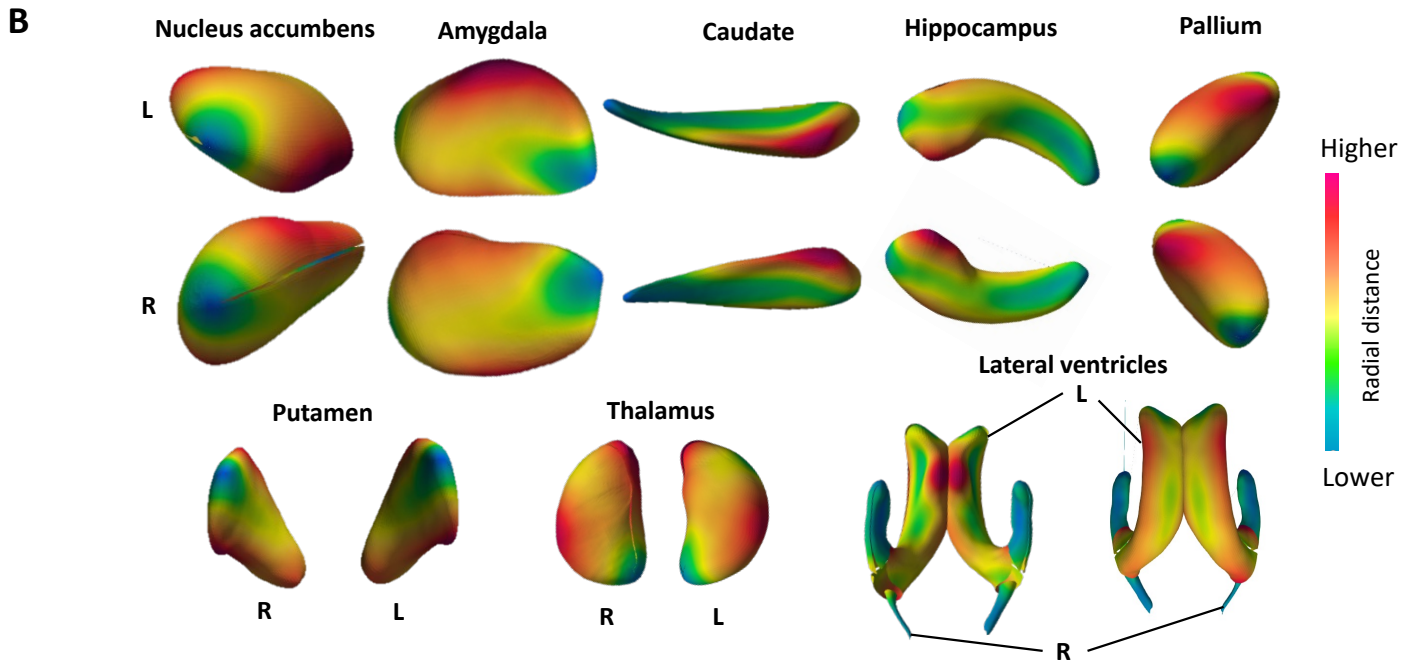
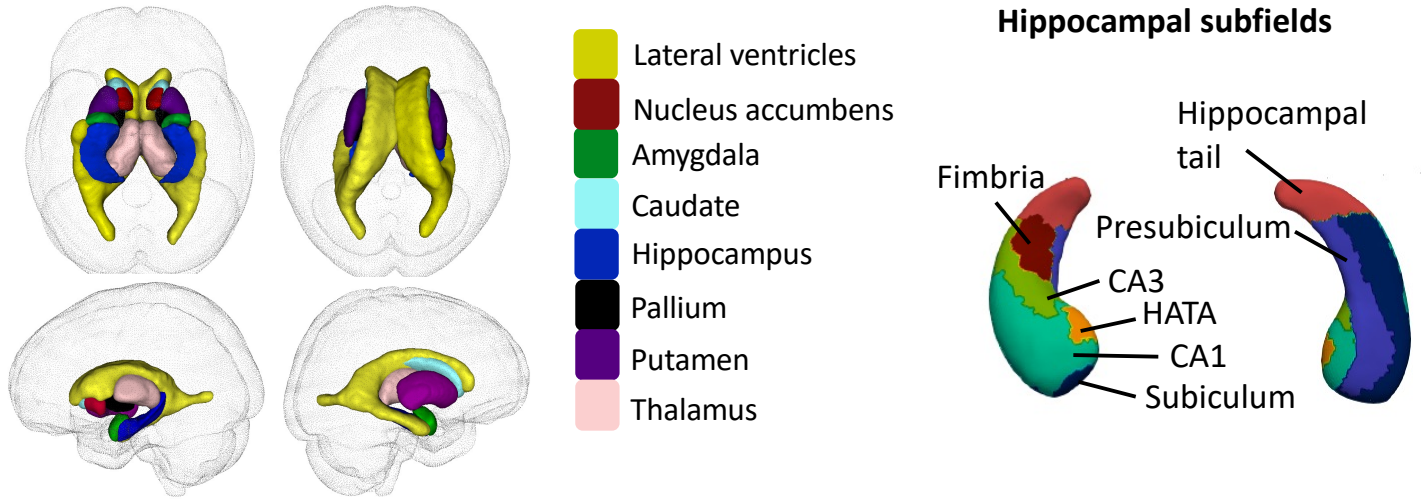


Figure 1

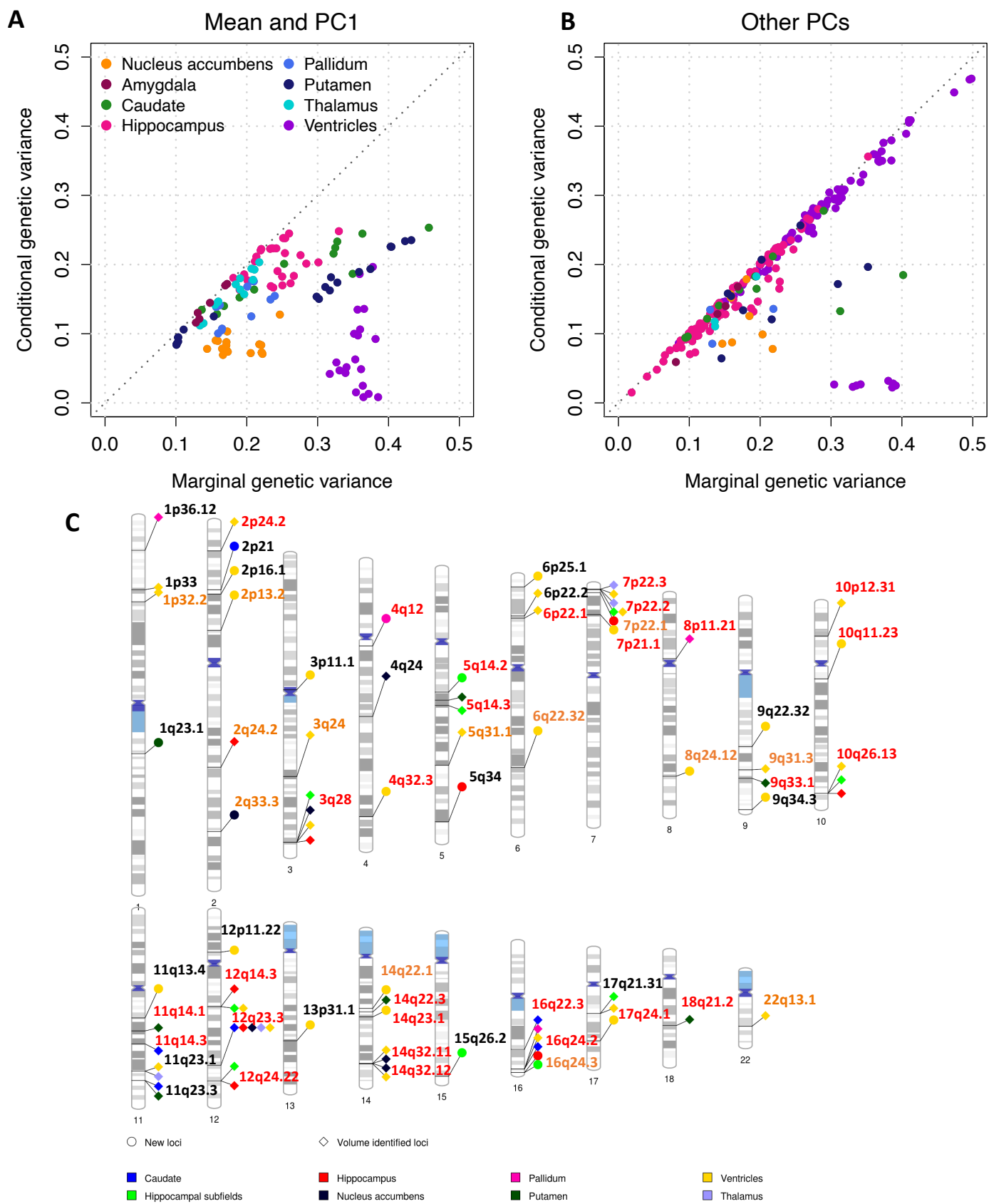


Figure 2

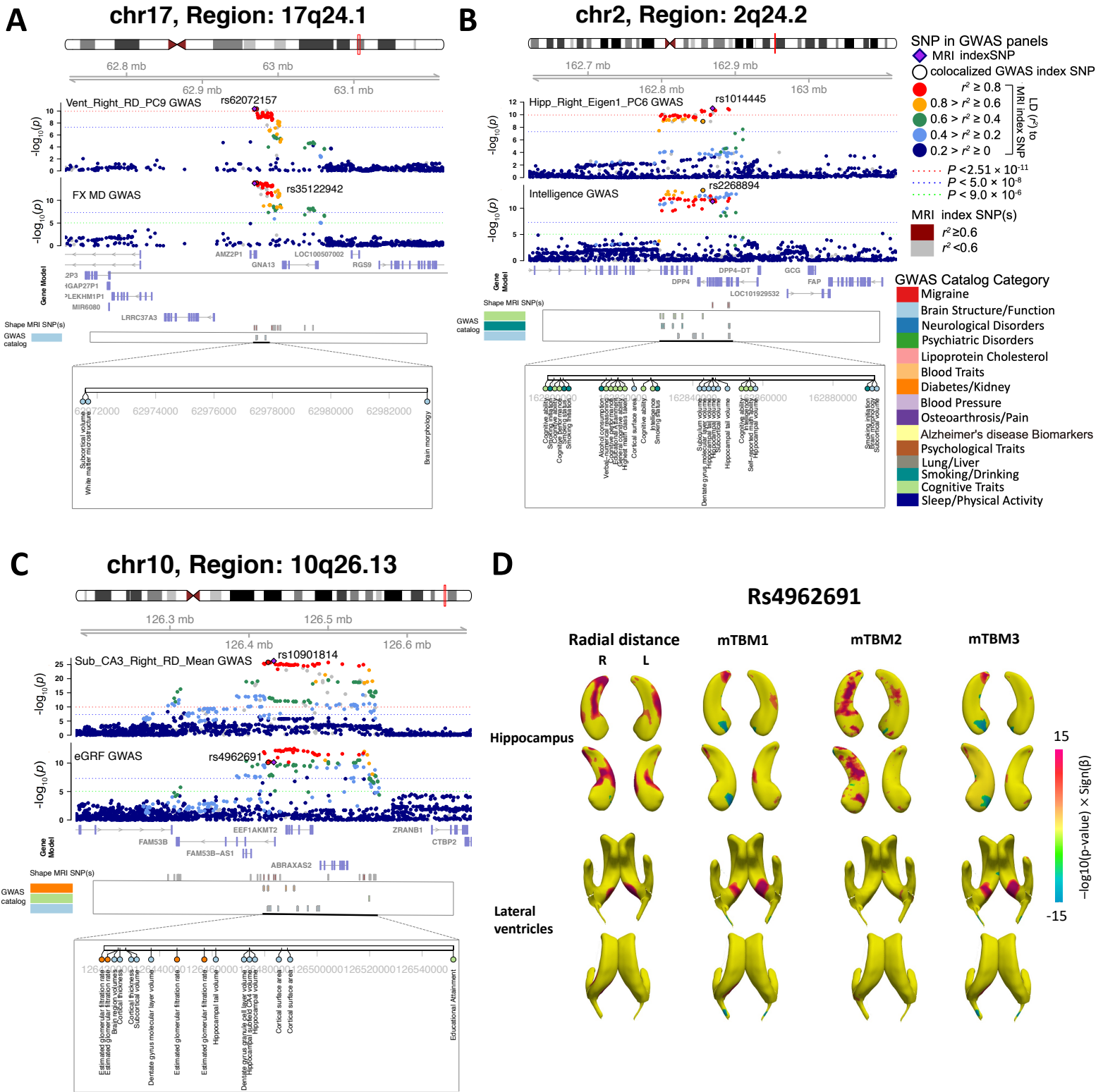


Figure 3

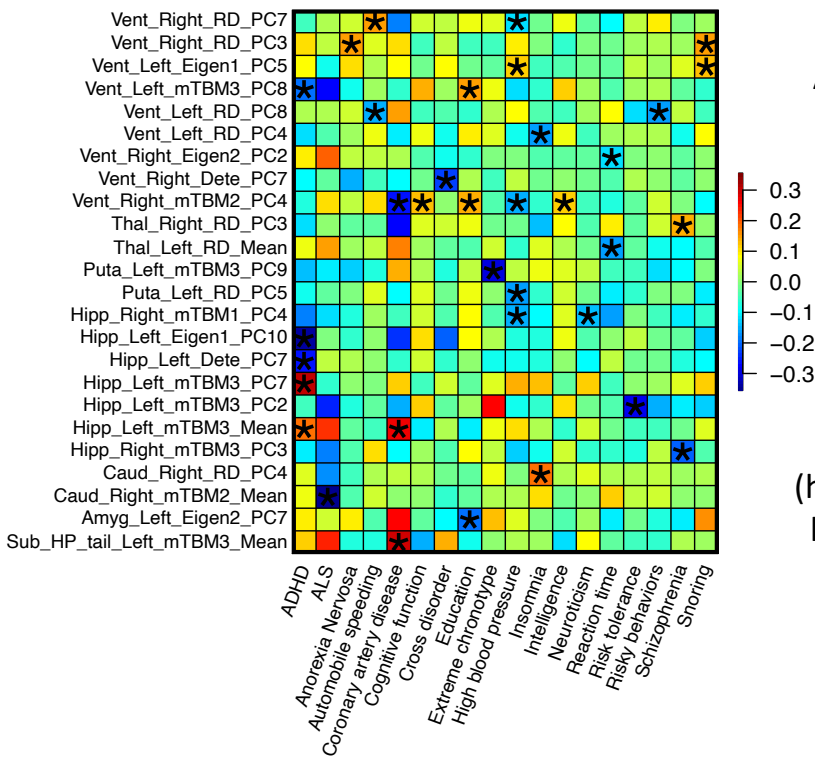
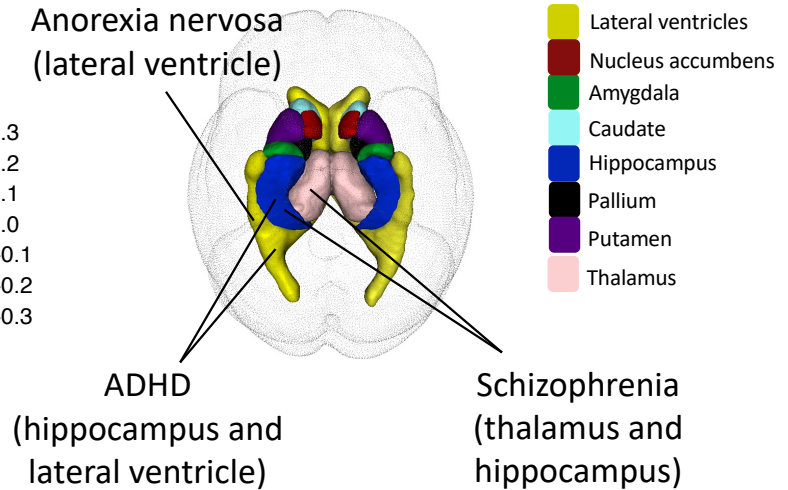
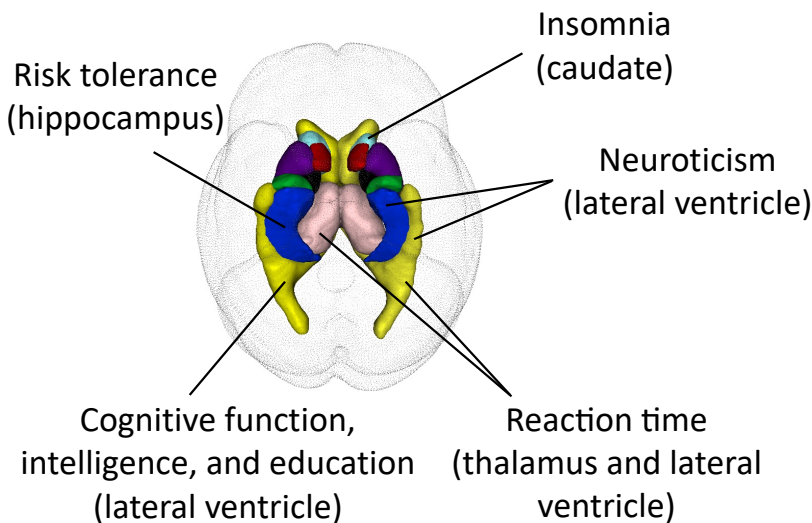
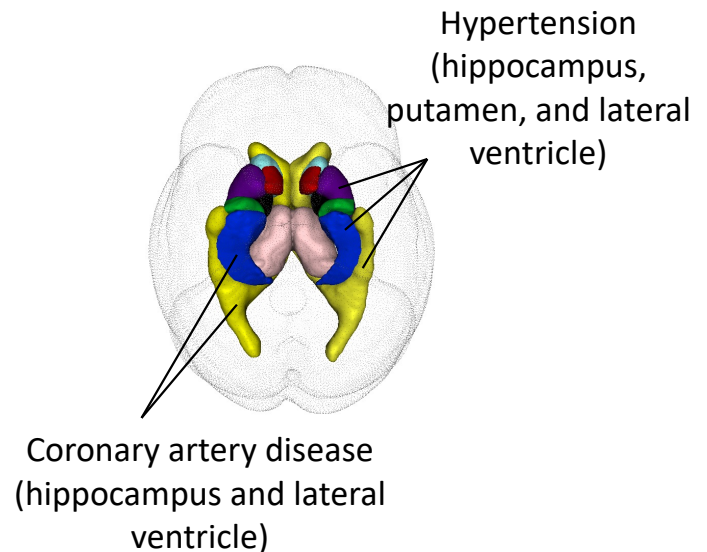
A**B****Brain disorders****C****Cognitive traits, education, and behavioral traits****D****Cardiovascular diseases**

Figure 4

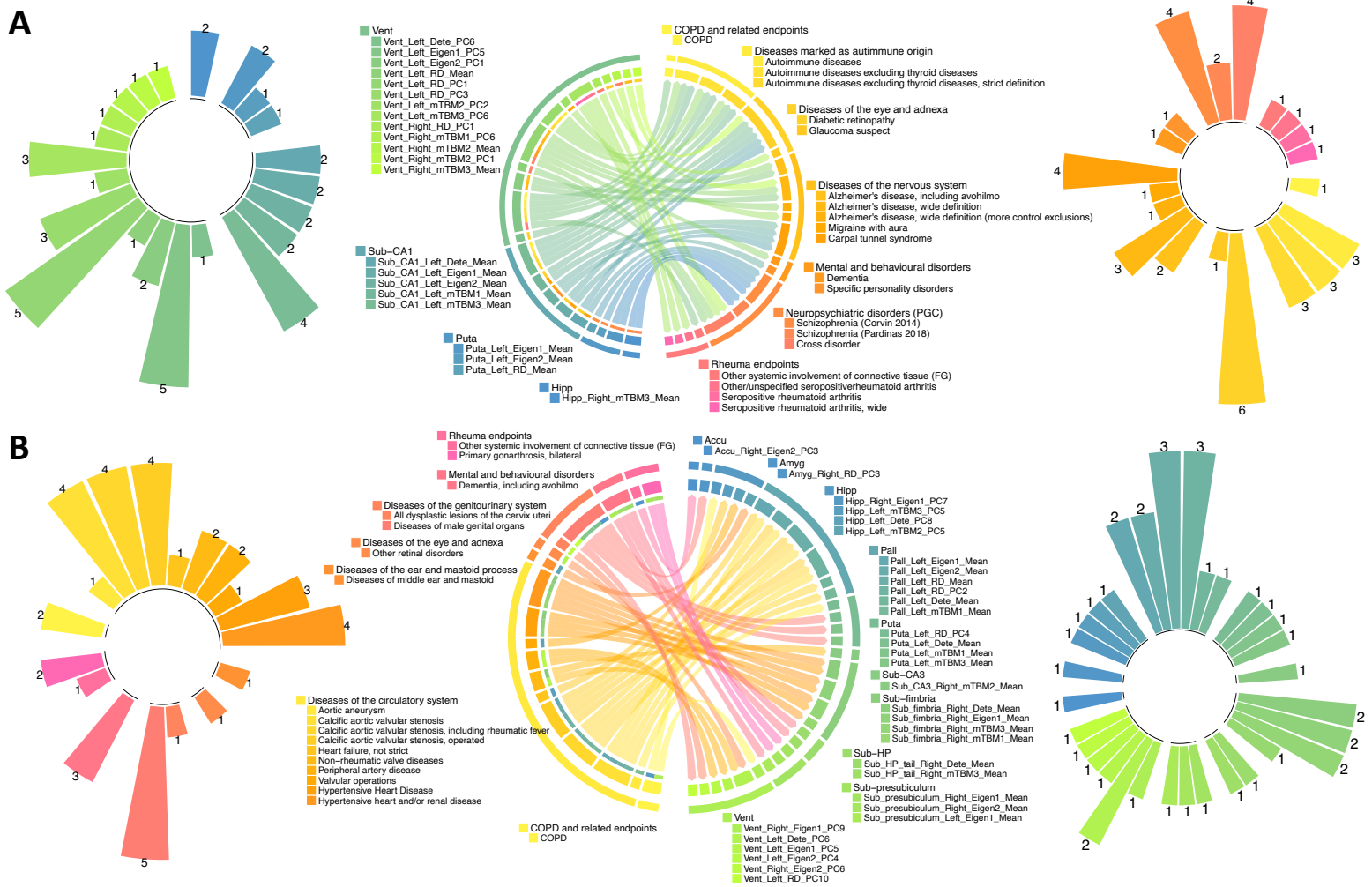


Figure 5