Supplementary Materials: heritability of regional brain volumes in large-scale neuroimaging and genetic studies

Bingxin Zhao[1], Joseph G. Ibrahim[1,6], Yun Li[1,2,3], Tengfei Li[4], Yue Wang[1], Yue Shan[1], Ziliang Zhu[1], Fan Zhou[1], Jingwen Zhang[1], Chao Huang[1], Huiling Liao[5], Liuqing Yang[6], Paul M. Thompson[7], and Hongtu Zhu[1,4,*]

[1] Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, 27599
[2]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, 27599
[3]Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, 27599
[4]Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, TX, USA, 77230
[5]Department of Statistics, Texas A&M University, College Station, TX, USA, 77843
[6]Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, 27599
[7]Imaging Genetics Center, Mark and Mary Stevens Institute for Neuroimaging & Informatics, University of Southern California, Los Angeles, CA, USA, 90033
[*]Correspondence: hzhu5@mdanderson.org (H.Z)

June 6, 2018

Supplementary Methods

Genotyping

PNC Imputation

From the PNC database, 8722 participants were genotyped on one of the six different platforms: 66 were genotyped on the Affymetrix array 6.0; 722 were genotyped on the Axiom array; 556 were genotyped on the Illumina HumanHap 550 array version1; 1914 were genotyped on the Illumina HumanHap 550 array version 3; 1657 were genotyped on the Illumina HumanHap 610 array; and 3807 were genotyped on the Illumina Human Omni Express array. We applied the quality control steps to each dataset separately, which included removal of subjects with more than 10% missing values, removal of SNPs (i) with more than 5% missing values, (ii) with MAF smaller than 5%, (iii) with Hardy-Weinberg equilibrium test p-value $< 1\times10^{-6}$, and (iv) located on a sex chromosome. We then employed MACH-Admix software (Liu et al., 2013) to perform genotype imputation, using 1000G Phase I Integrated Release Version 3 haplotypes (1000-Genomes-Project-Consortium et al., 2012) as a reference panel.

We also conducted quality control after imputation, excluding markers with (i) low imputation accuracy (a minimum R^2 cutoff was selected to ensure that the average R^2 value across all SNPs is larger than 0.8); and (ii) Hardy-Weinberg equilibrium test p-value $< 1 \times 10^{-6}$. We combined the six datasets and retained the shared SNPs. Finally, 5,354,265 bi-allelic markers (including SNPs and indels) from 8681 subjects remained for further analysis.

ADNI Imputation

Genetic data from ADNI1 (620,901 genetic markers from 818 subjects) and ADNI2/GO (730,525 genetic markers from 432 subjects) were processed separately with the following pipeline. The first-line quality control steps include (i) call rate check per subject and per SNP marker, (ii) gender check, (iii) sibling pair identification, and (iv) population stratification. The second-line preprocessing steps include removal of SNPs (i) with more than 5% missing values, (ii) with MAF smaller than 10%, (iii) with Hardy-Weinberg equilibrium p-value $< 1 \times 10^{-6}$ and (iv) located on a sex chromosome. For further processing, we included 503,778 SNPs from 756 white (Caucasian) subjects from ADNI1 and 516,453 SNPs from 397 white subjects from ADNI2/GO. We employed MACH-Admix software (Liu et al., 2013) to perform genotype imputation, using 1000G Phase I Integrated Release Version 3 haplotypes (1000-Genomes-Project-Consortium et al., 2012) as a reference panel. We conducted quality control after imputation, excluding markers with (i) low imputation accuracy (a minimum R^2 cutoff was selected to ensure that the average R^2 value across all SNPs is larger than 0.8); and (ii) Hardy-Weinberg equilibrium p-value $< 1 \times 10^{-6}$. We then had 7,986,566 bi-allelic markers (including SNPs and indels) from 756 subjects from ADNI1 and 8,218,182 markers from 397 subjects from ADNI2/GO. We combined the two datasets and retained the shared SNPs. Finally, 7,664,643 bi-allelic markers (including SNPs and indels) from 1153 subjects remained for further analysis.

PING Imputation

We applied the following preprocessing technique to the genetic data. The first-line quality control steps included (i) call rate check per subject and per SNP marker, (ii) gender check, and (iii) sibling pair identification. The second-line preprocessing steps included removal of SNPs (i) with more than 5% missing values, (ii) with MAF smaller than 10%, (iii) with Hardy-Weinberg equilibrium p-value $< 1 \times 10^{-6}$, and (iv) located on a sex chromosome. We thus had 539,865 SNPs from 1036 subjects for further processing. We employed MACH-Admix software (Liu et al., 2013) to perform genotype imputation, using 1000G Phase I Integrated Release Version 3 haplotypes (1000-Genomes-Project-Consortium et al., 2012) as a reference panel. We also conducted quality control after imputation, excluding markers with (i) low imputation accuracy (a minimum R^2 cutoff was selected to ensure that the average R^2 value across all SNPs is larger than 0.8), and (ii) Hardy-Weinberg equilibrium p-value $< 1 \times 10^{-6}$. Finally, 10,883,584 bi-allelic markers (including SNPs and indels) from 1036 subjects were retained for data analysis.

Brain functional communities

We clustered 97 brain regions into 18 non-overlapped functional communities (Buckner et al., 2008; Sporns and Betzel, 2016). The brain regions in each community were listed in Table S6. Here we briefly described the brain functional networks involved in each community:

1. C1: large area of prefrontal cortex. The functions span over the frontoparietal (control) network, default mode network and ventral (attention) network. This region may involve in functions such as decision making, complex cognitive behavior, processing of higher information, personal expression, social behavior moderating, attention, memory, and recognizing faces, characters,

2. C2: ventricles that belong to ventricular system. They are enlarged in a number of neurological conditions and are on average larger in patients with and schizophrenia, bipolar disorder, major depressive disorder and Alzheimer's disease.

3. C3: somatomotor network and auditory network. The insular cortex is often bilaterally activated during noxious somatosensory stimulation and has been suggested to play an important role in pain processing. It is believed to be involved in consciousness and plays a role in diverse functions usually linked to emotion or the regulation of the body's homeostasis, such as perception, motor control, self-awareness, cognitive functioning, and inter-personal experience.

4. C4: somatomotor network. Postcentral is the location of primary somatosensory cortex, and it is the main sensory receptive area for the sense of touch. Precentral is the site of the primary motor cortex. Paracentral is on the medial surface of the hemisphere and is the continuation of the precentral and postcentral gyri. Paracentral controls motor and sensory innervations of the contralateral lower extremity. It is also responsible for control of defecation and urination.

5. C5: default mode network, somatomotor network, visual network, ventral (attention) network and language networks. Cingulate region is an integral part of the limbic system, which is involved in emotion motor imagery and shifting attention between motor targets. Precuneus covers the sensorimotor anterior region, cognitive/associative central region, visual posterior regio. It also has the functions such as self-consciousness, memory, motor imagery and shifting attention between motor targets. The angular gyrus is part of the brain associated with complex language functions and has been associated with spatiovisual attention toward salient features.

6. C6: default mode network and frontoparietal (control) network. Superior frontal gyrus is involved in self-awareness, working memory and laughter.

7. C7: somatomotor network. This community composes the Basal Ganglia area, functions include control of voluntary motor movements, routine behaviors such as bruxism, eye movements, procedural learning, cognition and emotion.

8. C8: visual recognition. Both occipital and parahippocampal are related to recognition. Occipital involves in color processing, within-category identification and face/body/word recognition. Parahippocampal plays an important role in memory encoding and retrieval.

9. C9: default mode network and motion network. Hippocampal plays a role in memory, spatial navigation and control of attention

10. C10: emotional memory and olfaction.

11. C11: visual network. Lingual gyrus is a brain structure involves in processing vision, especially for letters.

12. C12: language network. Superior parietal is related to the manipulation and rearrangement of information in working memory, and patial orientation. Angular and supramarginal are involved with language perception and processing.

13. C13: auditory network and language network. The superior temporal gyrus contains the primary auditory cortex, which is responsible for processing sounds. It also includes the Wernicke's area and involves in the comprehension of language. Supramarginal is related to language perception and processing.

14. C14: memory, such as declarative (autobiographical/episodic/semantic) memories, memory formation, memory consolidation, and memory optimization in sleep.

15. C15: somatosensory system. Thalamus has multiple functions. It is generally believed to act as a relay between different subcortical areas and the cerebral cortex.

16. C16: the medial, cortico-nuclear zone of the cerebellum. Vermis is associated with bodily posture and locomotion. Vermis is included within the spinocerebellum and receives somatic sensory input from the head and proximal body parts via ascending spinal pathways.

17. C17: located to the front of and below the striatum. These structures are important in the production of acetylcholine, which is then distributed widely throughout the brain. Basal forebrain is considered to be the major cholinergic output of the central nervous system (CNS).

18. C18: mostly phylogenetically older regions. They are involved in many critical functions for life such as automated behaviors, physiologic functions, operant conditioning, and motor functions but also play some role in higher order functions as well.

PING Methods

Part of the data used in the preparation of this article were obtained from the Pediatric Imaging, Neurocognition and Genetics (PING) Study database (`http://ping.chd.ucsd.edu/`). PING was launched in 2009 by the National Institute on Drug Abuse (NIDA) and the Eunice Kennedy Shriver National Institute Of Child Health & Human Development (NICHD) as a 2-year project of the American Recovery and Reinvestment Act. The primary goal of PING has been to create a data resource of highly standardized and carefully curated magnetic resonance imaging (MRI) data, comprehensive genotyping data, and developmental and neuropsychological assessments for a large cohort of developing children aged 3 to 20 years. The scientific aim of the project is, by openly sharing these data, to amplify the power and productivity of investigations of healthy and disordered development in children, and to increase understanding of the origins of variation in neurobehavioral phenotypes. For up-to-date information, see `http://ping.chd.ucsd.edu/`.

ADNI Methods

Data used in the preparation of this article were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (`http://adni.loni.usc.edu`). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a 60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimers disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see `www.adni-info.org`.

References

E. Y. Liu, M. Li, W. Wang, Y. Li, Mach-admix: genotype imputation for admixed populations, Genetic epidemiology 37 (2013) 25–37.

1000-Genomes-Project-Consortium, et al., An integrated map of genetic variation from 1,092 human genomes, Nature 491 (2012) 56.

R. L. Buckner, J. R. Andrews-Hanna, D. L. Schacter, The brain's default network, Annals of the New York Academy of Sciences 1124 (2008) 1–38.

O. Sporns, R. F. Betzel, Modular brain networks, Annual review of psychology 67 (2016) 613–640.

Pediatric Imaging, Neurocognition and Genetics (PING) Authors

Connor McCabe[1], Linda Chang[2], Natacha Akshoomoff[3], Erik Newman[1], Thomas Ernst[2], Peter Van Zijl[4], Joshua Kuperman[5], Sarah Murray[6], Cinnamon Bloss[6], Mark Appelbaum[1], Anthony Gamst[1], Wesley Thompson[3], Hauke Bartsch[5].

Alzheimer's Disease Neuroimaging Initiative (ADNI) Authors

Michael Weiner[7], Paul Aisen[1], Ronald Petersen[8], Clifford R. Jack Jr[8], William Jagust[9], John Q. Trojanowki[10], Arthur W. Toga[11], Laurel Beckett[12], Robert C. Green[13], Andrew J. Saykin[14], John Morris[15], Leslie M. Shaw[10], Zaven Khachaturian[16], Greg Sorensen[17], Maria Carrillo[18], Lew Kuller[19], Marc Raichle[15], Steven Paul[20], Peter Davies[21], Howard Fillit[22], Franz Hefti[23], Davie Holtzman[15], M. Marcel Mesulman[24], William Potter[25], Peter J. Snyder[26], Adam Schwartz[27], Tom Montine[28], Ronald G. Thomas[1], Michael Donohue[1], Sarah Walter[1], Devon Gessert[1], Tamie Sather[1], Gus Jiminez[1], Danielle Harvey[12], Matthew Bernstein[8], Nick Fox[29], Paul Thompson[11], Norbert Schuff[7], Charles DeCarli[12], Bret Borowski[8], Jeff Gunter[8], Matt Senjem[8], Prashanthi Vemuri[8], David Jones[8], Kejal Kantarci[8], Chad Ward[8], Robert A. Koeppe[30], Norm Foster[31], Eric M. Reiman[32], Kewei Chen[32], Chet Mathis[19], Susan Landau[9], Nigel J. Cairns[15], Erin Householder[15], Lisa Taylor-Reinwald[15], Virginia M.Y. Lee[10], Magdalena Korecka[10], Michal Figurski[10], Karen Crawford[11], Scott Neu[11], Tatiana M. Foroud[14], Steven Potkin[33], Li Shen[14], Kelley Faber[14], Sungeun Kim[14], Kwangsik Nho[14], Leon Thal[1], Richard Frank[34], Neil Buckholtz[35], Marilyn Albert[36], John Hsiao[35].

[1]UC San Diego, La Jolla, CA 92093, USA. [2]U Hawaii, Honolulu, HI 96822, USA. [3]Department of Psychiatry, University of California, San Diego, La Jolla, California 92093, USA. [4]Kennedy Krieger Institute, Baltimore, MD 21205, USA. [5]Multimodal Imaging Laboratory, Department of Radiology, University of California San Diego, La Jolla, California 92037, USA. [6]Scripps Translational Science Institute, La Jolla, CA 92037, USA. [7]UC San Francisco, San Francisco, CA 94143, USA. [8]Mayo Clinic, Rochester, MN 55905, USA. [9]UC Berkeley, Berkeley, CA 94720-5800, USA. [10]U Pennsylvania, Philadelphia, PA 19104, USA. [11]USC, University of Southern California, Los Angeles, CA 90033, USA. [12]UC Davis, Davis, CA 95616, USA. [13]Brigham and Women s Hospital/Harvard Medical School, Boston MA 02115, USA. [14]Indiana University, Indianapolis, IN 46202-5143, USA. [15]Washington University St. Louis, St. Louis, MO 63130, USA. [16]Prevent Alzheimers Disease 2020, Rockville, MD 20850, USA. [17]Siemens [18]Alzheimers Association, Chicago, IL 60601, USA. [19]University of Pittsburgh, Pittsburgh, PA 15260, USA. [20]Cornell University, Ithaca, NY 14850, USA. [21]Albert Einstein College of Medicine of Yeshiva University, Bronx, NY 10461, USA. [22]AD Drug Discovery Foundation, New York, NY 10019, USA. [23]Acumen Pharmaceuticals, Livermore, California 94551, USA. [24]Northwestern University, Evanston, IL 60208, USA. [25]National Institute of Mental Health, Bethesda, MD 20892-9663, USA. [26]Brown University, Providence, RI 02912, USA. [27]Eli Lilly, Indianapolis, Indiana 46285, USA. [28]University of Washington, Seattle, WA 98195, USA. [29]University of London, London WC1E 7HU, UK. [30]University of Michigan, Ann Arbor, MI 48109, USA. [31]University of Utah, Salt Lake City, UT 84112, USA. [32]Banner Alzheimers Institute, Phoenix, AZ 85006, USA. [33]UC Irvine, Irvine, CA 92697, USA. [34]General Electric [35]National Institute on Aging/National Institutes of Health, Bethesda, MD 20892, USA. [36]The Johns Hopkins University, Baltimore, MD 21218, USA.

Supplementary Tables

Table S1: Demographical information of participants in MRI datasets after quality control

Dataset	Sample size	Mean age (s.d.)	Age range	Gender proportion (Female/Male)
UK Biobank	9868	64.26 (7.48)	(47 80)	0.525/0.475
ADNI	1248	74.18 (7.16)	(55 92)	0.434/0.566
PNC	1492	21.14 (3.71)	(14 29)	0.525/0.475
PING	924	12.38(4.98)	(3 21)	0.482/0.518

Table S2: Sample size of each dataset after conducting all quality control procedures

Dataset	Sample Size	Number of SNPs
UK Biobank	9031	461,488 (unimputed)
ADNI	1023	7,368,446
PNC	599	5,151,137
PING	508	8,612,247

Table S3: Heritability estimates of major brain regions in UK Biobank

Region	SNP heritability	Adjusted P-value	SE
BV	0.659	1.48E-19	0.071
CSF (adjusted for BV)	0.393	1.51E-08	0.069
GM Volume (adjusted for BV)	0	NA	NA
WM Volume (adjusted for BV)	0.628	4.64E-18	0.071
GM Volume (not adjusted for BV)	0.675	1.93E-20	0.071
WM Volume (not adjusted for BV)	0.620	8.58E-18	0.071

Table S4: Heritability estimates of 97 brain regions in UK Biobank. Provided in Excel Data Set.

Table S5: List of the top regions with largest gender disparity, as measured by absolute difference in point heritability estimates

ROI ID	ROI names	Difference in heritability
R35	right.putamen	0.437
L35	left.putamen	0.411
L09	left.cuneus	0.350
L25	left.paracentral	0.335
L34	left.precuneus	0.327
R14	right.inferior.parietal	0.310
R13	right.inferior.lateral.ventricle	0.303
R45	right.vessel	0.300
L42	left.thalamus.proper	0.270
L29	left.pars.triangularis	0.263
R12	right.hippocampus	0.256
RE51	X3rd.ventricle	0.241
L40	left.superior.temporal	0.230
R19	right.lateral.orbitofrontal	0.226
L22	left.medial.orbitofrontal	0.225

Table S6: Brain functional communities and their main functions

Community ID	Size	Main functions
C1	16	frontoparietal (control), default mode, ventral (attention)
C2	6	ventricular system
C3	4	somatomotor and auditory
C4	6	somatomotor
C5	11	default mode, somatomotor, visual, ventral (attention),language
C6	2	default mode ,frontoparietal (control)
C7	6	somatomotor
C8	6	visual recognition
C9	2	default mode ,motion
C10	2	emotional memory, olfaction
C11	6	visual
C12	6	language
C13	5	auditory, language
C14	2	memory
C15	2	somatosensory system
C16	3	bodily posture, locomotion
C17	2	production of acetylcholine
C18	10	other (mostly phylogenetically older regions)

Table S7: Heritability estimates and raw p-values for the top 10 PCs with and without adjusting for BV, UK Biobank

		Adjusted for BV			Not adjusted for BV		
	PC	Heritability	P-value	SE	Heritability	P-value	SE
1	PC1	0	NA	NA	0.687	1.001E-22	0.070
2	PC2	0.179	4.987E-03	0.069	0.179	4.986E-03	0.069
3	PC3	0.501	3.453E-13	0.070	0.508	1.512E-13	0.070
4	PC4	0.627	4.297E-20	0.069	0.633	2.052E-20	0.069
5	PC5	0.551	9.248E-16	0.069	0.556	4.913E-16	0.069
6	PC6	0.641	2.948E-20	0.070	0.635	6.148E-20	0.070
7	PC7	0.440	1.571E-10	0.070	0.444	1.304E-10	0.070
8	PC8	0.294	1.361E-05	0.070	0.293	1.475E-05	0.070
9	PC9	0.398	7.335E-09	0.070	0.398	7.601E-09	0.070
10	PC10	0.411	2.314E-09	0.070	0.413	1.815E-09	0.070

Table S8: Heritability estimates of each brain region in ADNI. Provided in Excel Data Set.

Table S9: Heritability estimates of each brain region in PING. Provided in Excel Data Set.

Table S10: Heritability estimates of each brain region in PNC. Provided in Excel Data Set.

Table S11: Heritability estimates and raw p-values for the top 10 PCs with and without adjusting for BV, ADNI.

		Adjusted for BV			Not adjusted for BV		
	PC	Heritability	P-value	SE	Heritability	P-value	SE
1	PC1	0.000	NA	NA	0.337	1.955E-01	0.393
2	PC2	0.528	8.985E-02	0.394	0.445	1.283E-01	0.392
3	PC3	0.684	4.458E-02	0.402	0.720	3.673E-02	0.402
4	PC4	0.551	7.384E-02	0.380	0.690	3.399E-02	0.378
5	PC5	0.153	3.514E-01	0.400	0.317	2.125E-01	0.397
6	PC6	0.233	2.853E-01	0.410	0.174	3.355E-01	0.409
7	PC7	0.000	5.000E-01	0.393	0.000	5.000E-01	0.393
8	PC8	0.636	4.705E-02	0.380	0.575	6.525E-02	0.380
9	PC9	0.554	8.000E-02	0.394	0.553	8.038E-02	0.394
10	PC10	0.254	2.627E-01	0.400	0.225	2.878E-01	0.402

Table S12: Heritability estimates and raw p-values for the top 10 PCs with and without adjusting for BV, PNC.

		Adjusted for BV			Not adjusted for BV		
	PC	Heritability	P-value	SE	Heritability	P-value	SE
1	PC1	0.000	5.000E-01	0.569	0.527	1.661E-01	0.544
2	PC2	0.000	5.000E-01	0.551	0.000	5.000E-01	0.550
3	PC3	0.873	3.428E-02	0.479	0.779	5.545E-02	0.488
4	PC4	0.000	5.000E-01	0.517	0.000	5.000E-01	0.520
5	PC5	1.000	1.214E-02	0.444	1.000	1.221E-02	0.444
6	PC6	0.237	3.319E-01	0.545	0.262	3.143E-01	0.541
7	PC7	0.000	5.000E-01	0.540	0.000	5.000E-01	0.538
8	PC8	0.000	5.000E-01	0.535	0.000	5.000E-01	0.535
9	PC9	0.000	5.000E-01	0.565	0.000	5.000E-01	0.563
10	PC10	0.399	2.306E-01	0.541	0.251	3.277E-01	0.563

Table S13: Heritability estimates and raw p-values for the top 10 PCs with and without adjusting for BV, PING.

		Adjusted for BV			Not adjusted for BV		
	PC	Heritability	P-value	SE	Heritability	P-value	SE
1	PC1	0.000	NA	NA	0.014	4.909E-01	0.624
2	PC2	0.224	3.610E-01	0.629	0.219	3.739E-01	0.644
3	PC3	1.000	5.578E-02	0.628	0.943	8.175E-02	0.625
4	PC4	1.000	4.704E-02	0.597	1.000	3.132E-02	0.596
5	PC5	1.000	4.947E-02	0.606	1.000	4.993E-02	0.606
6	PC6	0.385	2.634E-01	0.608	0.429	2.363E-01	0.603
7	PC7	0.761	9.018E-02	0.568	0.861	5.598E-02	0.562
8	PC8	0.283	3.078E-01	0.564	0.282	2.954E-01	0.563
9	PC9	0.849	7.662E-02	0.594	0.806	8.824E-02	0.592
10	PC10	0.605	1.504E-01	0.585	0.598	1.491E-01	0.585

Table S14: Percentage of missing values of all ROIs in the UK Biobank, ADNI, PNC and PING datasets. Provided in Excel Data Set.

Supplementary Figures

Figure S1. SNP heritability and adjusted p-values ranked by estimates in UK Biobank.

Figure S2. Relationship between the SNP heritability estimate and the average volume of each brain region. Genetic contributions have large variation among regions with comparable mean volume sizes.

Figure S3. Heritability estimate of each brain region by each chromosome. The chromosomes are ordered by their length. In UK Biobank, longer chromosomes tend to have larger heritability estimates than shorter ones.

Figure S4. Heritability estimates of top 10 PCs by each chromosome. The chromosomes are ordered from left to right by their lengths. In UK Biobank, longer chromosomes tend to have larger heritability estimates than shorter ones.

Figure S5. Aggregated heritability of top 10 PCs by each chromosome. In UK Biobank and ADNI, heritability explained by each chromosome is highly correlated with chromosome length.

Figure S6. SNP heritability and adjusted p-values grouped by PC loadings in UK Biobank. The top 10 loadings corresponding to each of the top 10 PCs were classified into one module. Regions classified in modules corresponding to PCs that explain more volume variation do not necessarily have higher heritability estimates.

Figure S7. Heritability estimates from ADNI, PNC and PING datasets have much larger variance and more extreme values than those from the UK Biobank.

Figure S8. Comparing heritability estimates from UK biobank and those from other datasets. Each point represents pair of point estimates by all common autosomal SNPs (MAF > 0.01).

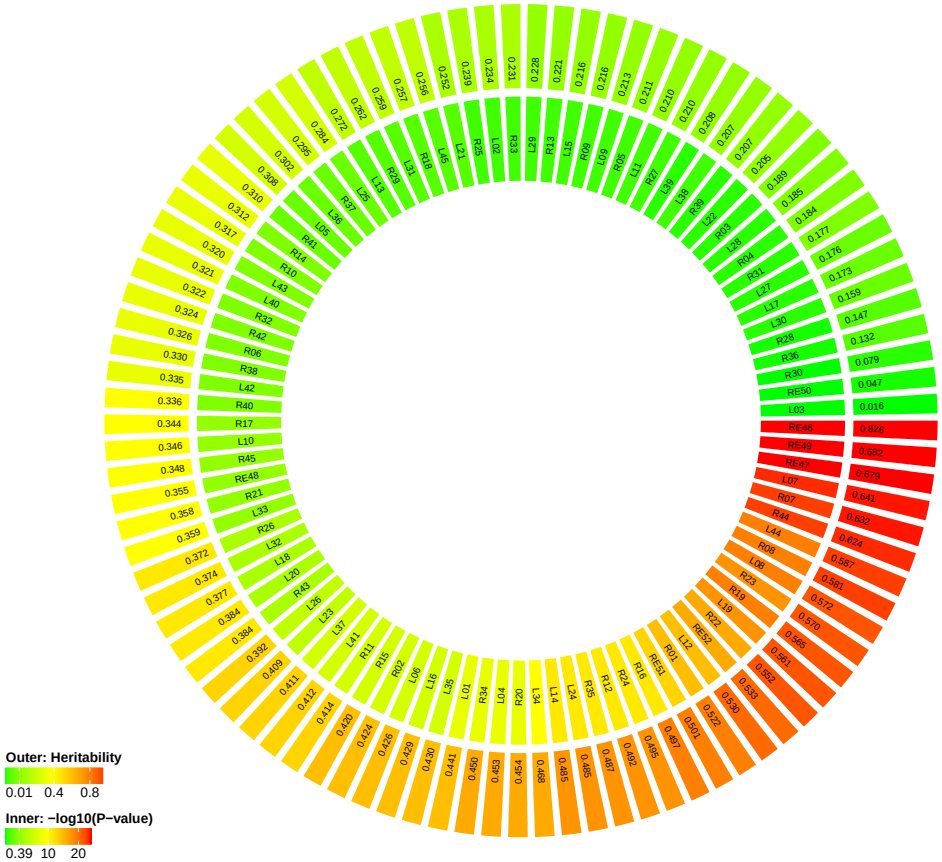
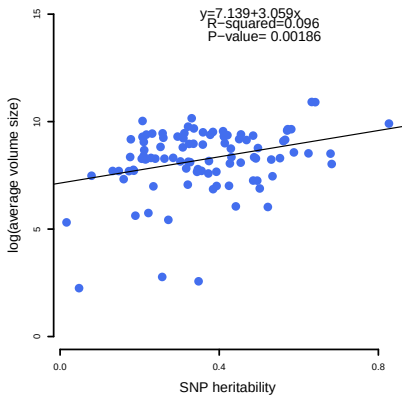


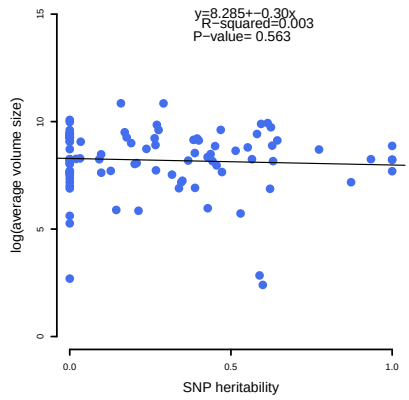
Figure S1

A



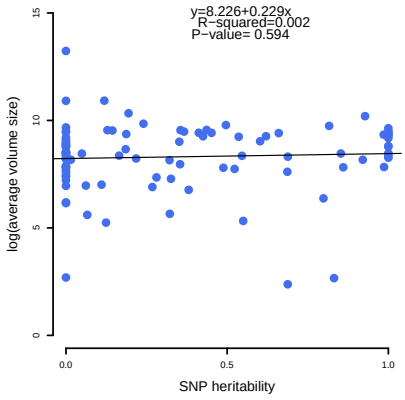
UK Biobank

B



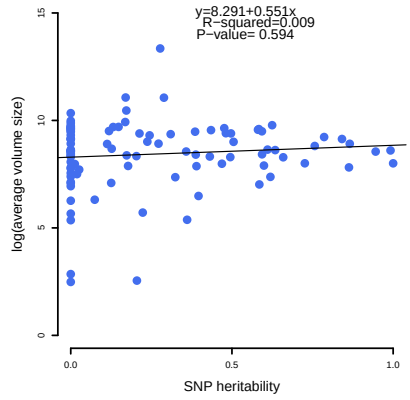
ADNI

C



PNC

D



PING

Figure S2

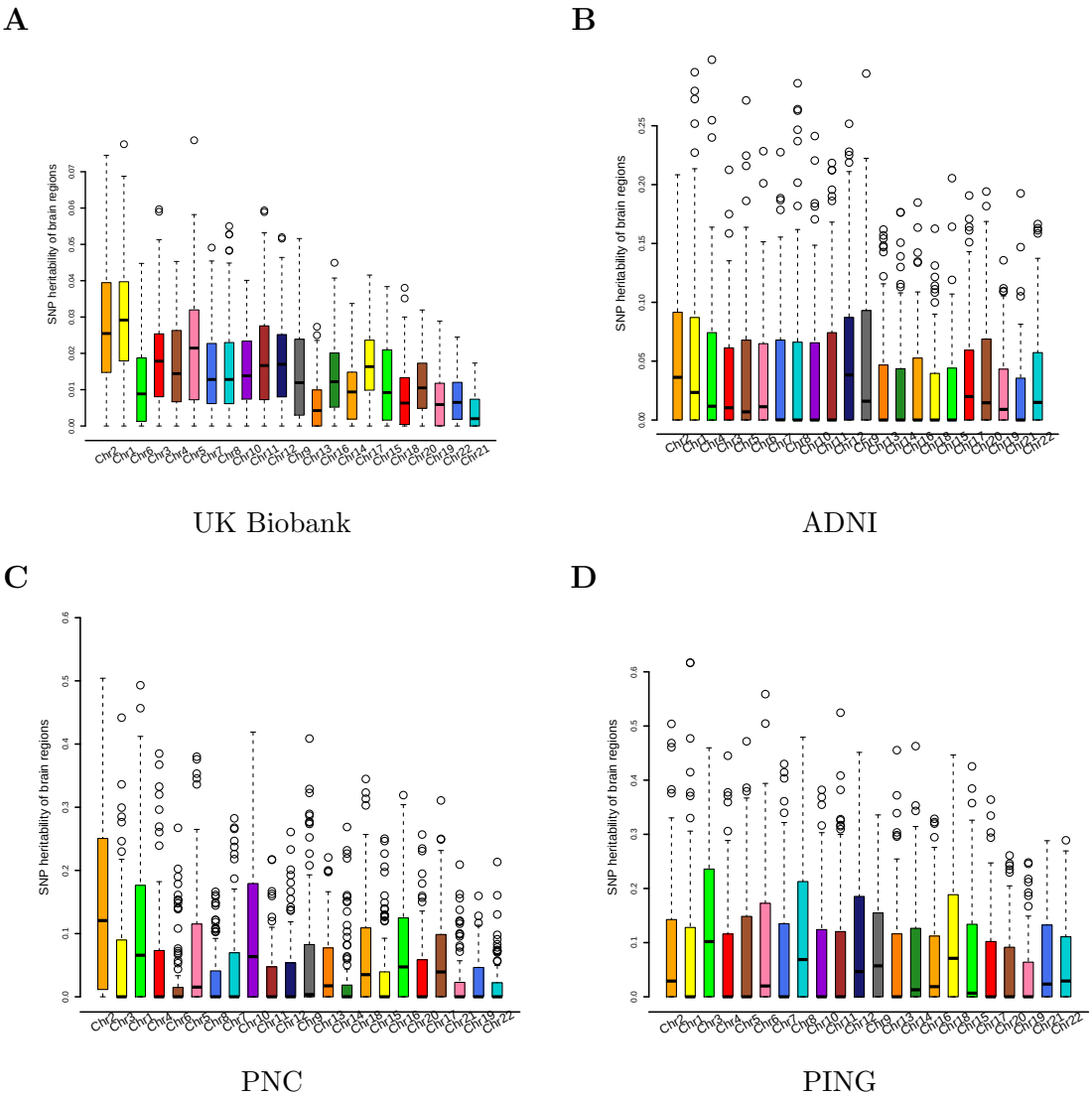

A
B
C
D
SNP heritability of brain regions
UK Biobank
ADNI
PNC
PING

Figure S3

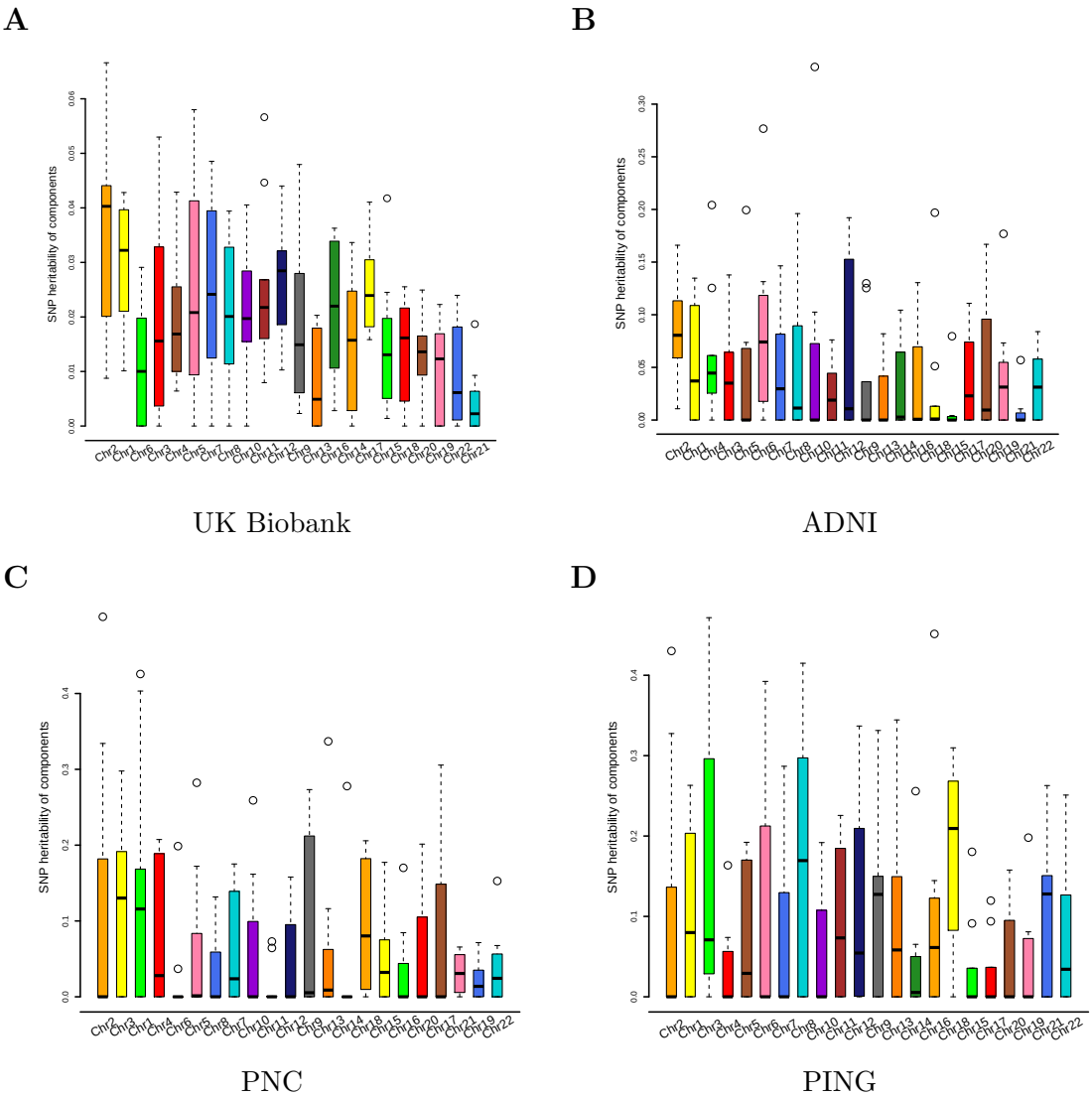

A
SNP heritability of components
UK Biobank
B
SNP heritability of components
ADNI
C
SNP heritability of components
PNC
D
SNP heritability of components
PING

Figure S4

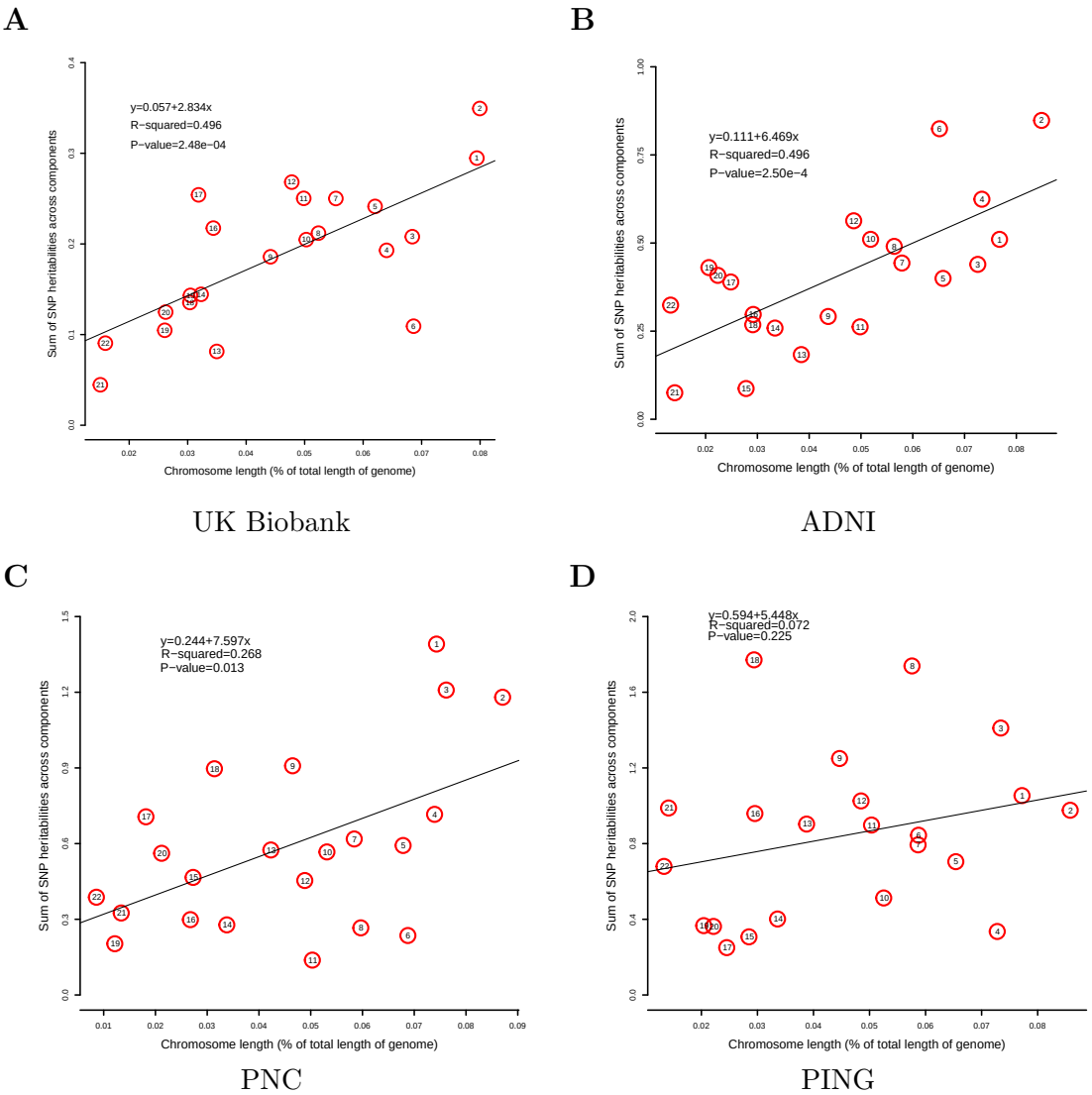

A
y=0.057+2.834x
R−squared=0.496
P−value=2.48e−04
Sum of SNP heritabilities across components
Chromosome length (% of total length of genome)
UK Biobank
B
y=0.111+6.469x
R−squared=0.496
P−value=2.50e−4
Sum of SNP heritabilities across components
Chromosome length (% of total length of genome)
ADNI
C
y=0.244+7.597x
R−squared=0.268
P−value=0.013
Sum of SNP heritabilities across components
Chromosome length (% of total length of genome)
PNC
D
y=0.594+5.448x
R−squared=0.072
P−value=0.225
Sum of SNP heritabilities across components
Chromosome length (% of total length of genome)
PING

Figure S5

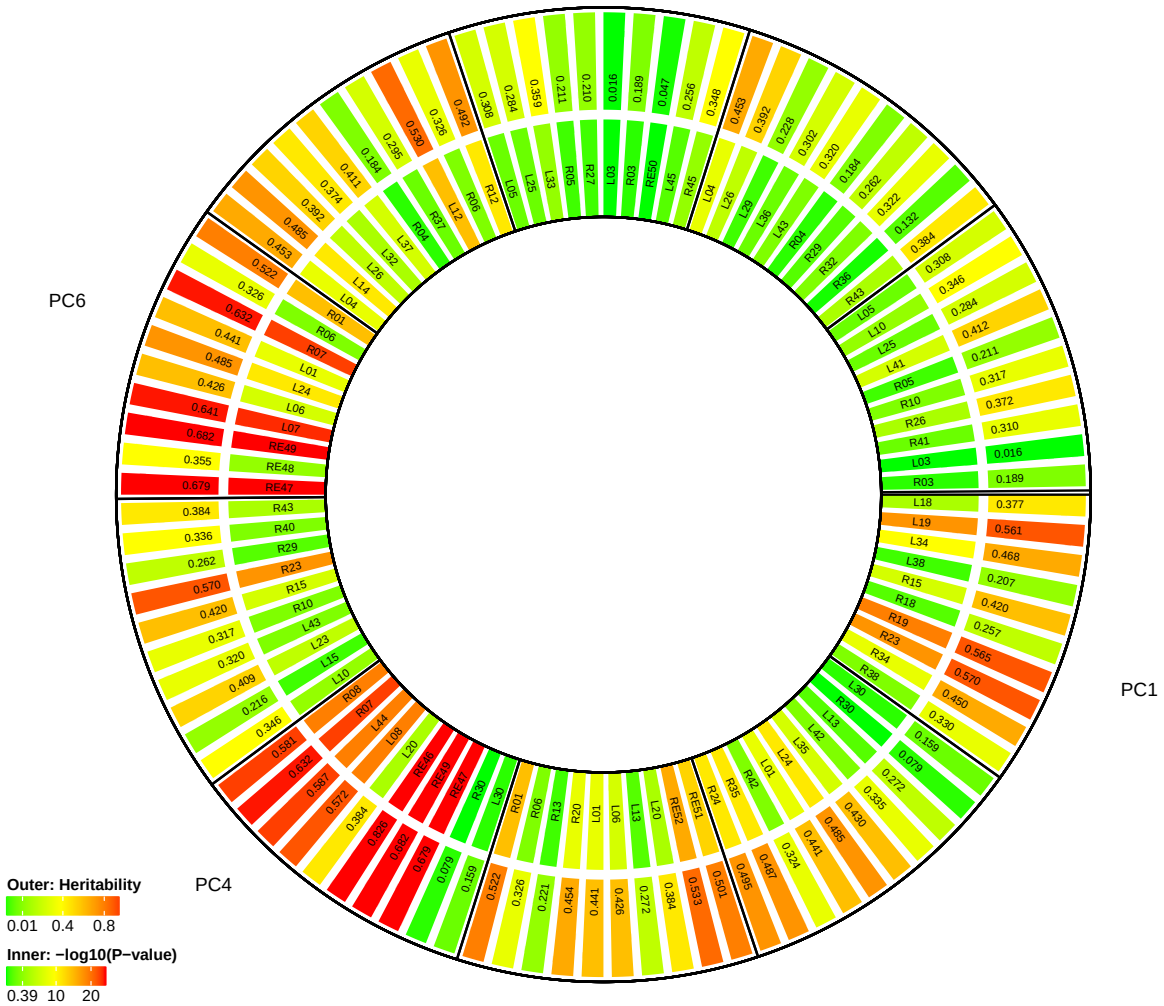

PC6
PC1
PC4
Outer: Heritability
0.01 0.4 0.8
Inner: −log10(P−value)
0.39 10 20

Figure S6

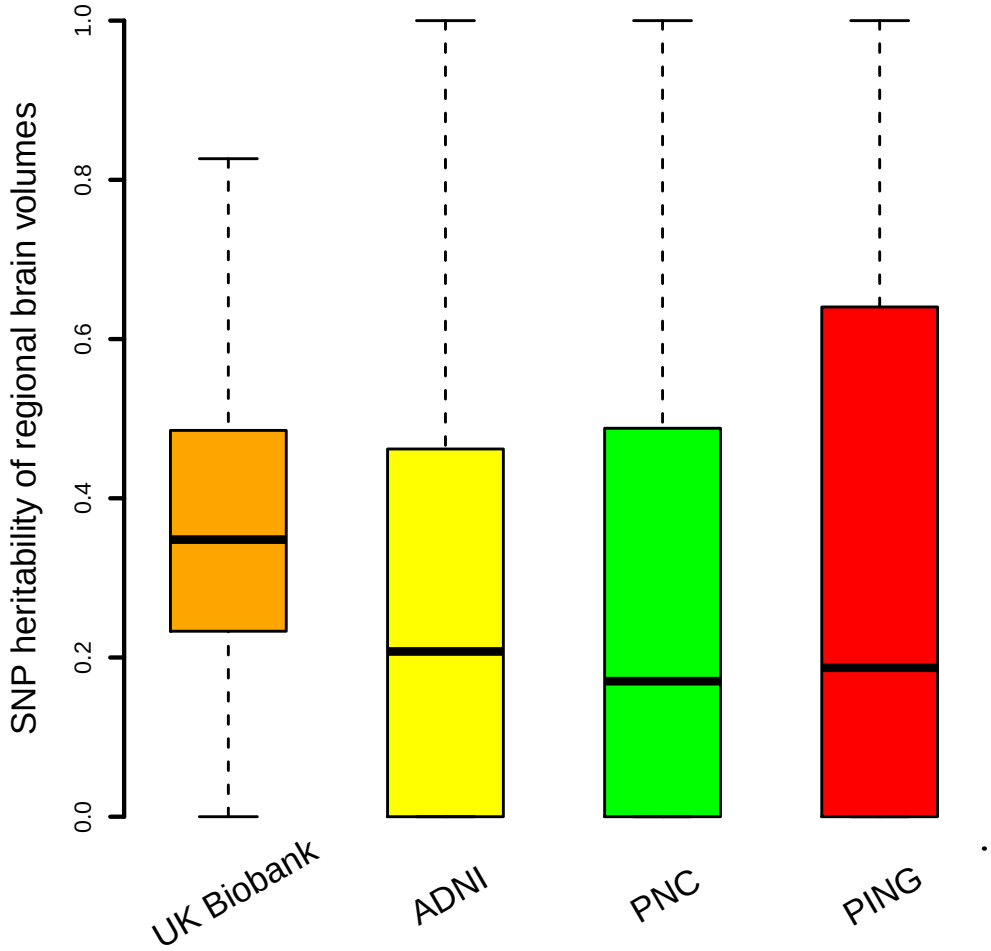
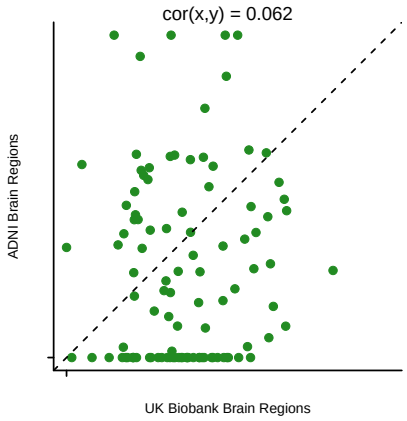

SNP heritability of regional brain volumes
0.0
0.2
0.4
0.6
0.8
1.0
UK Biobank
ADNI
PNC
PING

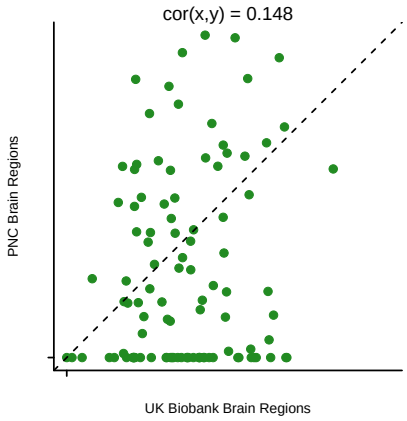
Figure S7

A



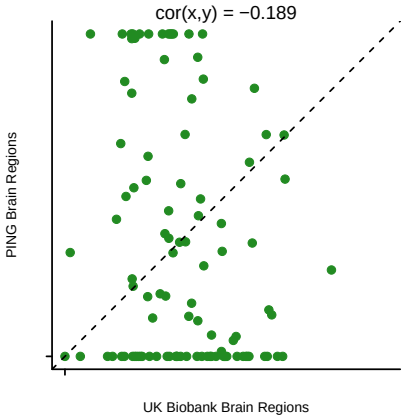
UK Biobank v.s. ADNI

B



UK Biobank v.s. PNC

C



UK Biobank v.s. PING

Figure S8

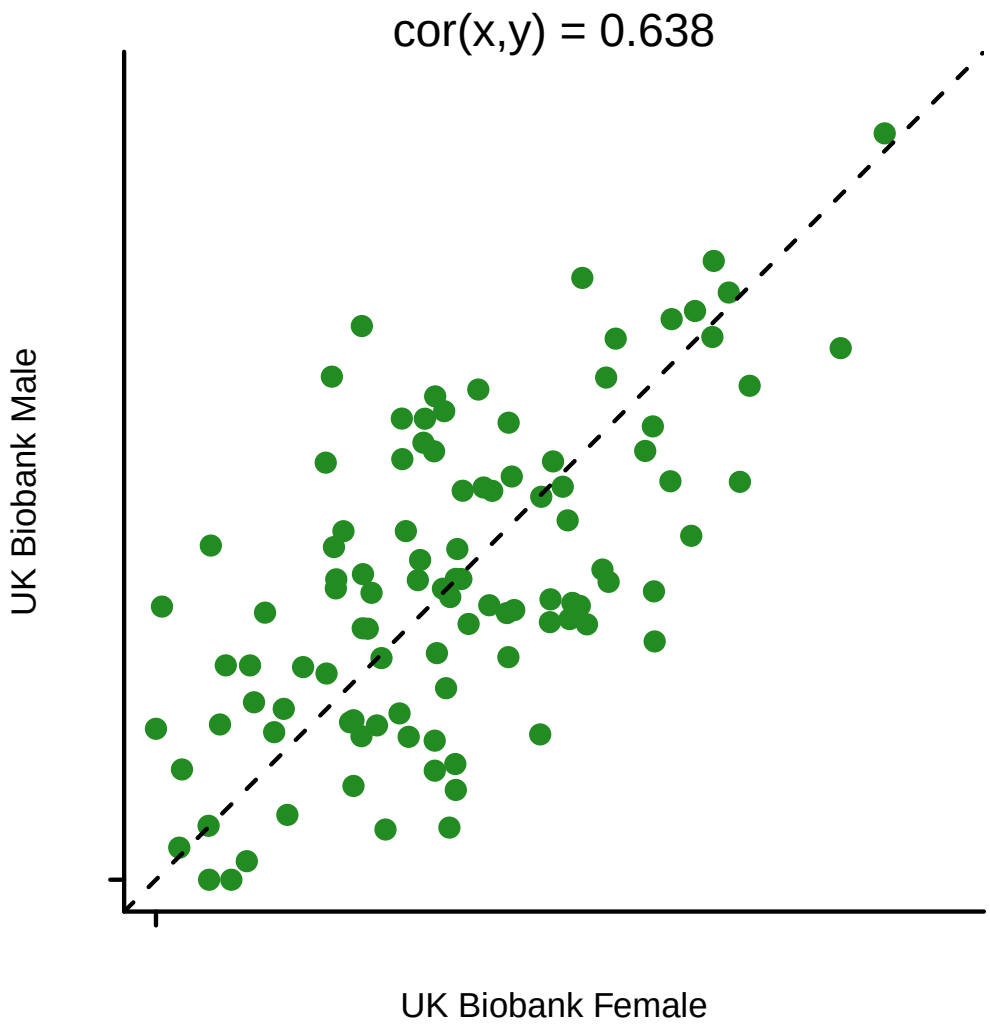


Figure S9

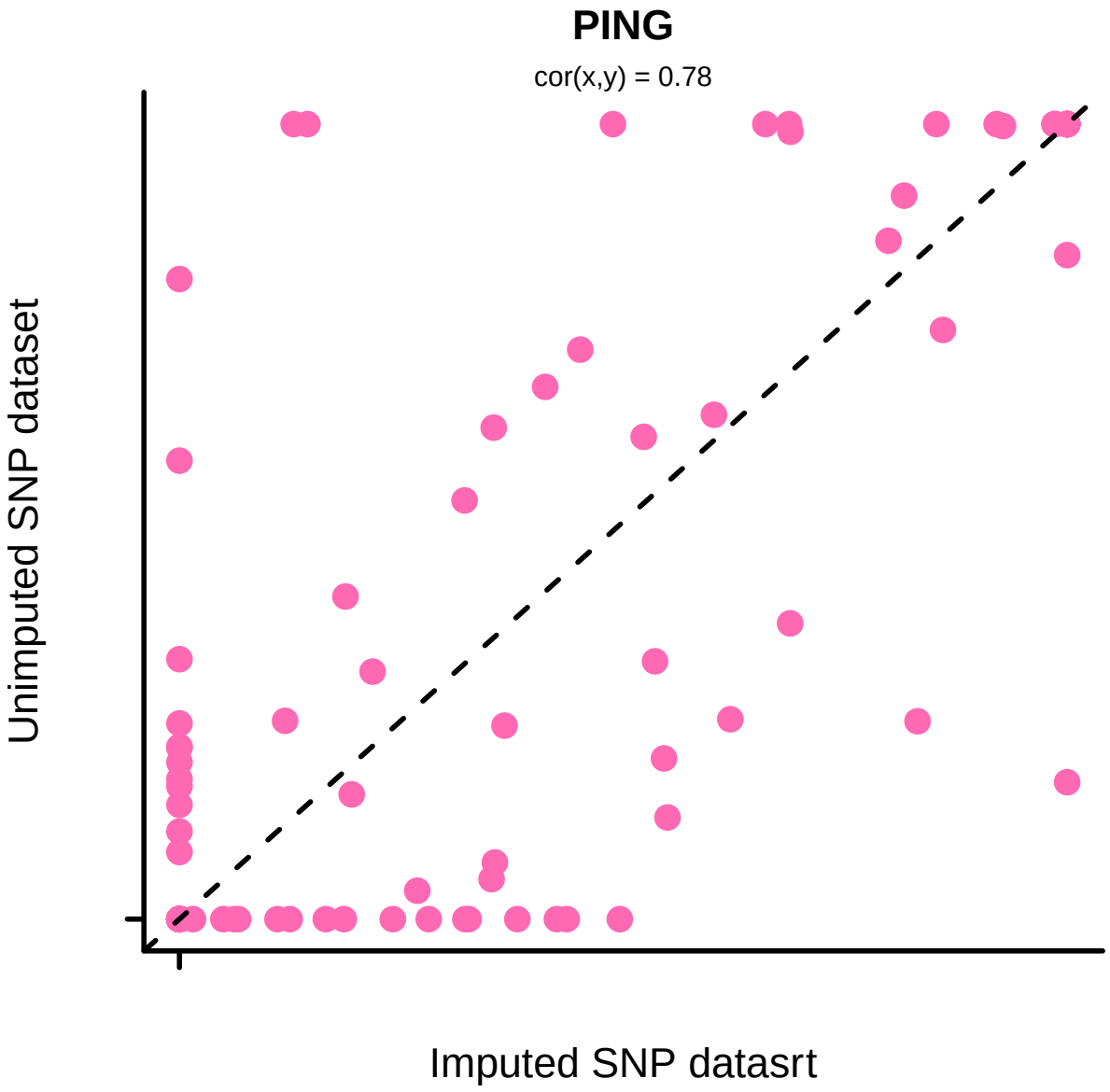


Figure S10