LETTER

doi:10.1038/nature14101

Common genetic variants influence human subcortical brain structures

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

The highly complex structure of the human brain is strongly shaped by genetic influences[1]. Subcortical brain regions form circuits with cortical areas to coordinate movement[2], learning, memory[3] and motivation[4], and altered circuits can lead to abnormal behaviour and disease[2]. To investigate how common genetic variants affect the structure of these brain regions, here we conduct genome-wide association studies of the volumes of seven subcortical regions and the intracranial volume derived from magnetic resonance images of 30,717 individuals from 50 cohorts. We identify five novel genetic variants influencing the volumes of the putamen and caudate nucleus. We also find stronger evidence for three loci with previously established influences on hippocampal volume[5] and intracranial volume[6]. These variants show specific volumetric effects on brain structures rather than global effects across structures. The strongest effects were found for the putamen, where a novel intergenic locus with replicable influence on volume (rs945270; $P = 1.08 \times 10^{-33}$; 0.52% variance explained) showed evidence of altering the expression of the *KTN1* gene in both brain and blood tissue. Variants influencing putamen volume clustered near developmental genes that regulate apoptosis, axon guidance and vesicle transport. Identification of these genetic variants provides insight into the causes of variability in human brain development, and may help to determine mechanisms of neuropsychiatric dysfunction.

At the individual level, genetic variations exert lasting influences on brain structures and functions associated with behaviour and predisposition to disease. Within the context of the Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) consortium, we conducted a collaborative large-scale genetic analysis of magnetic resonance imaging (MRI) scans to identify genetic variants that influence brain structure. Here, we focus on volumetric measures derived from a measure of head size (intracranial volume, ICV) and seven subcortical brain structures corrected for the ICV (nucleus accumbens, caudate, putamen, pallidum, amygdala, hippocampus and thalamus). To ensure data homogeneity within the ENIGMA consortium, we designed and implemented standardized protocols for image analysis, quality assessment, genetic imputation (to 1000 Genomes references, version 3) and association (Extended Data Fig. 1 and Methods).

After establishing that the volumes extracted using our protocols were substantially heritable in a large sample of twins ($P < 1 \times 10^{-4}$; see Methods and Extended Data Fig. 11a), with similar distributions to previous studies[1], we sought to identify common genetic variants contributing to volume differences by meta-analysing site-level genome-wide association study (GWAS) data in a discovery sample of 13,171 subjects of European ancestry (Extended Data Fig. 2). Population stratification was controlled for by including, as covariates, four population components derived from standardized multidimensional scaling analyses of genome-wide genotype data conducted at each site (see Methods). Site-level GWAS results and distributions were visually inspected to check for statistical inflation and patterns indicating technical artefacts (see Methods).

Meta-analysis of the discovery sample identified six genome-wide significant loci after correcting for the number of variants and traits analysed ($P < 7.1 \times 10^{-9}$; see Methods): one associated with the ICV, two associated with hippocampal volume, and three with putamen volume. Another four loci showed suggestive associations ($P < 1 \times 10^{-7}$) with putamen volume (one locus), amygdala volume (two loci), and caudate volume (one locus; Table 1, Fig. 1 and Supplementary Table 5). Quantile–quantile plots showed no evidence of population stratification or cryptic relatedness (Extended Data Fig. 4a). We subsequently attempted to replicate the variants with independent data from 17,546 individuals.

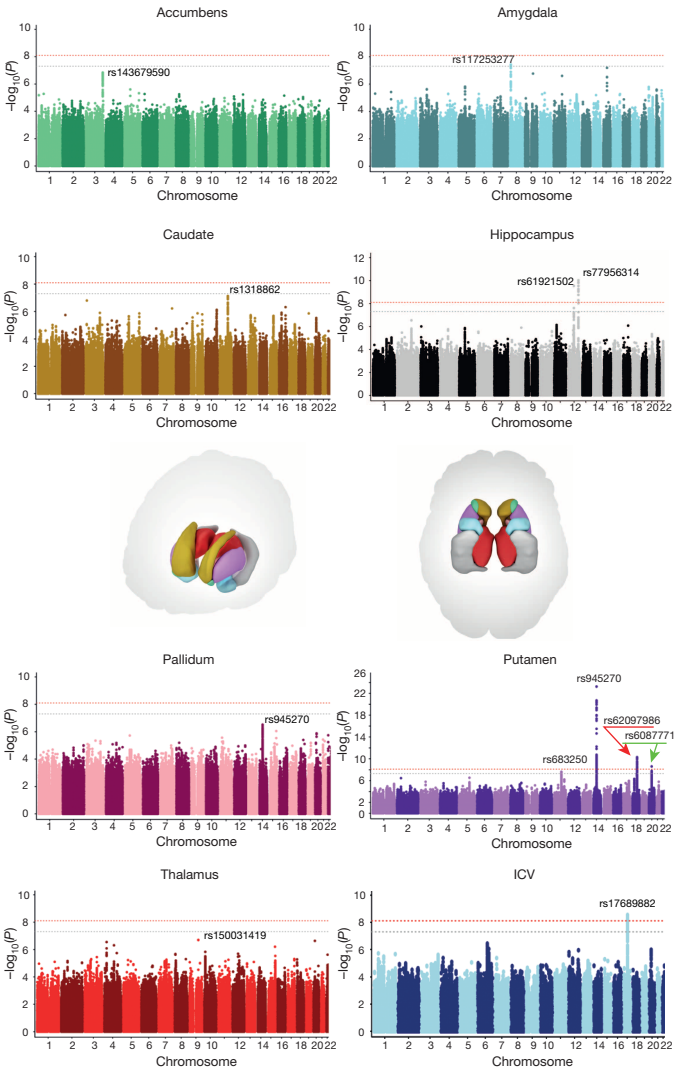


Figure 1 | Common genetic variants associated with subcortical volumes and the ICV. Manhattan plots coloured with a scheme that matches the corresponding structure (middle) are shown for each subcortical volume studied. Genome-wide significance is shown for the common threshold of $P = 5 \times 10^{-8}$ (grey dotted line) and also for the multiple comparisons-corrected threshold of $P = 7.1 \times 10^{-9}$ (red dotted line). The most significant SNP within an associated locus is labelled.

Table 1 | Genetic variants at eight loci were significantly associated with putamen, hippocampus, caudate nucleus and ICV

					Discovery cohort			Replication cohort			Discovery + replication cohorts				
Trait	Marker	A1	A2	Frq	Effect (se)	*P* value	Sample size	Effect (se)	*P* value	Sample size	Effect (se)	*P* value	Total sample size	Variance explained (%)	Diff./allele (%)
Putamen	rs945270	C	G	0.58	60.64 (6.00)	5.43×10^{-24}	13,145	39.15 (5.46)	7.81×10^{-13}	15,130	48.89 (4.04)	1.08×10^{-33}	28,275	0.52	0.94
Putamen	rs62097986	A	C	0.44	39.53 (6.01)	4.86×10^{-11}	13,145	22.46 (5.53)	4.89×10^{-5}	14,891	30.28 (4.07)	1.01×10^{-13}	28,036	0.20	0.58
Putamen	rs6087771	T	C	0.71	40.72 (6.82)	2.42×10^{-9}	11,865	26.97 (6.57)	4.02×10^{-5}	13,675	33.58 (4.73)	1.28×10^{-12}	25,540	0.20	0.64
Putamen	rs683250	A	G	0.63	−33.97 (6.08)	2.33×10^{-8}	13,145	−22.30 (5.89)	1.50×10^{-4}	13,113	−27.95 (4.23)	3.94×10^{-11}	26,258	0.17	0.51
Caudate	rs1318862	T	C	0.58	26.27 (4.89)	7.54×10^{-8}	13,171	31.82 (14.23)	0.025	1,860	26.86 (4.62)	6.17×10^{-9}	15,031	0.22	0.74
Hip.	rs77956314	T	C	0.91	−54.21 (8.37)	9.33×10^{-11}	13,163	−57.43 (12.69)	6.04×10^{-6}	4,027	−55.18 (6.99)	2.82×10^{-15}	17,190	0.36	1.40
Hip.	rs61921502	T	G	0.84	43.40 (6.89)	2.92×10^{-10}	13,163	26.81 (13.32)	0.044	3,046	39.90 (6.12)	6.87×10^{-11}	16,209	0.26	1.01
ICV	rs17689882	A	G	0.22	−15,335.88 (2,582.20)	2.87×10^{-9}	10,944	−5,202.15 (5,428.60)	0.337	1,878	−13,460.47 (2,331.05)	7.72×10^{-9}	12,822	0.26	0.96

The allele frequency (frq) and effect size are given with reference to allele 1 (A1). Effect sizes are given in units of mm^3 per effect allele. Results are provided for the discovery samples and the combined meta-analysis of the discovery and replication cohorts (all European ancestry). Additional validation was attempted in non-European ancestry generalization samples (shown in Supplementary Table 6). The variance explained gives the percentage variance explained by a given SNP after correcting for covariates (see Methods for additional details). The percentage difference in volume per effect allele (Diff./allele) is based on the absolute value of the final combined effect divided by a weighted average of the brain volume of interest across all sites in the discovery sample and then multiplied by 100. Hip, hippocampus.

All subcortical genome-wide significant variants identified in the discovery sample were replicated (Table 1). The variant associated with the ICV did not replicate in a smaller independent sample, but was genome-wide significant in a previously published independent study[6], providing strong evidence for its association with the ICV. Moreover, two suggestive variants associated with putamen and caudate volumes exceeded genome-wide significance after meta-analysis across the discovery and replication data sets (Table 1). Effect sizes were similar across cohorts ($P > 0.1$, Cochran's Q test; Extended Data Fig. 4b). Effect sizes remained consistent after excluding patients diagnosed with anxiety, Alzheimer's disease, attention-deficit/hyperactivity disorder, bipolar disorder, epilepsy, major depressive disorder or schizophrenia (21% of the discovery participants). Correlation in effect size with and without patients was very high ($r > 0.99$) for loci with $P < 1 \times 10^{-5}$, indicating that these effects were unlikely to be driven by disease (Extended Data Fig. 5a). The participants' age range covered most of the lifespan (9–97 years), but only one of the eight significant loci showed an effect related to the mean age of each cohort ($P = 0.002$; rs6087771 affecting putamen volume; Extended Data Fig. 5b), suggesting that nearly all effects are stable across the lifespan. In addition, none of these loci showed evidence of sex effects (Extended Data Fig. 5c).

In our cohorts, significant loci were associated with 0.51–1.40% differences in volume per risk allele, explaining 0.17–0.52% of the phenotypic variance (Table 1); such effect sizes are similar to those of common variants influencing other complex quantitative traits such as height[7] and body mass index[8]. The full genome-wide association results explained 7–15% of phenotypic variance after controlling for the effects of covariates (Extended Data Fig. 11). Notably, the genome-wide significant variants identified here showed specific effects on single brain structures rather than pleiotropic effects across multiple structures, despite similar developmental origins as in the case of caudate and putamen (Extended Data Fig. 6a). Nevertheless, when we subjected the subcortical meta-analysis results to hierarchical clustering, genetic determinants of the subcortical structures were mostly grouped into larger circuits according to their developmental and functional subdivisions (Extended Data Fig. 6b). Genetic variants may therefore have coherent effects on functionally associated subcortical networks. Multivariate cross-structure[9] analyses confirmed the univariate results, but no additional loci reached genome-wide significance (Extended Data Fig. 6c). The clustering of results into known brain circuits in the absence of individually significant genetic variants found in the cross-structure analysis suggests variants of small effect may have similar influences across structures. Most variants previously reported to be associated with brain structure and/or function showed little evidence of large-scale volumetric effects (Supplementary Table 8). We detected an intriguing association with hippocampal volume at a single nucleotide polymorphism (SNP) with a genome-wide significant association with schizophrenia[10] (rs2909457; $P = 2.12 \times 10^{-6}$; where the A allele is associated with decreased risk for schizophrenia and decreased hippocampal volume). In general, however, we detected no genome-wide significant association with brain structure for genome-wide significant loci that contribute risk for neuropsychiatric illnesses (Supplementary Table 9).

Of the four loci influencing putamen volume, we identified an intergenic locus 50 kilobases (kb) downstream of the *KTN1* gene (rs945270; 14q22.3; $n = 28{,}275$; $P = 1.08 \times 10^{-33}$), which encodes the protein kinectin, a receptor that allows vesicle binding to kinesin and is involved in organelle transport[11]. Second, we identified an intronic locus within *DCC* (rs62097986; 18q21.2; $n = 28{,}036$; $P = 1.01 \times 10^{-13}$), which encodes a netrin receptor involved in axon guidance and migration, including in the developing striatum[12] (Extended Data Fig. 3b). Expression of *DCC* throughout the brain is highest in the first two trimesters of prenatal development[13] (Extended Data Fig. 8b), suggesting that this variant may influence brain volumes early in neurodevelopment. Third, we identified an intronic locus within *BCL2L1* (rs6087771; 20q11.21; $n = 25{,}540$; $P = 1.28 \times 10^{-12}$), which encodes an anti-apoptotic factor that inhibits programmed cell death of immature neurons throughout the brain[14] (Extended Data Fig. 3c). Consistent with this, expression of *BCL2L1* in the striatum strongly decreases at the end of neurogenesis (24–38 post-conception weeks (PCW); Extended Data Fig. 8c), a period marked by increased apoptosis in the putamen[13,15]. Fourth, we identified an intronic locus within *DLG2* (rs683250; 11q14.1; $n = 26{,}258$; $P = 3.94 \times 10^{-11}$), which encodes the postsynaptic density 93 (PSD-93) protein (Extended Data Fig. 3d). PSD-93 is a membrane-associated guanylate kinase involved in organizing channels in the postsynaptic density[16]. *DLG2* expression increases during early mid-fetal development in the striatum[13] (Extended Data Fig. 8d). Genetic variants in *DLG2* affect learning and cognitive flexibility[17] and are associated with schizophrenia[18]. Notably, SNPs associated with variation in putamen volume showed enrichment of genes involved in apoptosis and axon guidance pathways (Extended Data Fig. 7 and Supplementary Table 7).

Hippocampal volume showed an intergenic association near the *HRK* gene (rs77956314; 12q24.22; $n = 17{,}190$; $P = 2.82 \times 10^{-15}$; Extended Data Fig. 3g) and with an intronic locus in the *MSRB3* gene (rs61921502; 12q14.3; $n = 16{,}209$; $P = 6.87 \times 10^{-11}$; Extended Data Fig. 3h), supporting our previous analyses[5,19] of smaller samples imputed to HapMap3 references. Caudate volume was associated with an intergenic locus 80 kb from *FAT3* (rs1318862; 11q14.3; $n = 15{,}031$; $P = 6.17 \times 10^{-9}$; Extended Data Fig. 3e). This gene encodes a cadherin specifically expressed in the

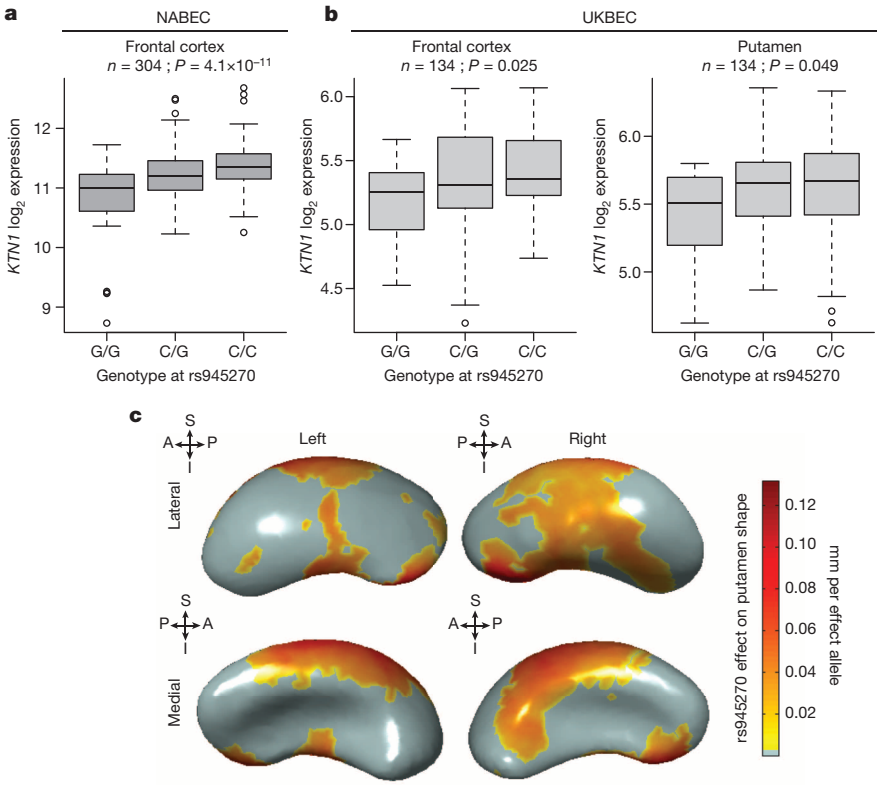


Figure 2 | Effect of rs945270 on *KTN1* expression and putamen shape. **a**, **b**, Expression quantitative trait loci study in brain tissue demonstrates the effect of rs945270 on *KTN1* gene expression in frontal cortex tissue from 304 subjects from the North American Brain Expression Cohort (NABEC[25]) (**a**) and in an independent sample of 134 subjects from the UK Brain Expression Cohort (UKBEC) (**b**), sampled from both frontal cortex and putamen. Boxplot dashed bars mark the twenty-fifth and seventy-fifth percentiles. **c**, Surface-based analysis demonstrates that rs945270 has strong effects on the shape of superior and lateral portions of the putamen in 1,541 subjects. Each copy of the rs945270-C allele was significantly associated with increased width in coloured areas (false discovery rate corrected at $q = 0.05$), and the degree of deformation is labelled by colour, with red indicating greater deformation. Orientation is indicated by arrows. A, anterior; I, inferior; P, posterior, S, superior.

nervous system during embryonic development that influences neuronal morphology through cell–cell interactions[20]. The ICV was associated with an intronic locus within *CRHR1* that tags the chromosome 17q21 inversion[21], which has been previously found to influence ICV[6] (rs17689882; 17q21.31; $n = 12{,}822$; $P = 7.72 \times 10^{-9}$; Extended Data Fig. 3f). Another previously identified variant with association to ICV (rs10784502)[5,19] did not survive genome-wide significance in this analysis but did show a nominal effect in the same direction ($P = 2.05 \times 10^{-3}$; $n = 11{,}373$). None of the genome-wide significant loci in this study were in linkage disequilibrium with known functional coding variants, splice sites, or 3′/5′ untranslated regions, although several of the loci had epigenetic markings suggesting a regulatory role (Extended Data Fig. 3).

Given the strong association with putamen volume, we further examined the rs945270 locus. Epigenetic markers suggest insulator functionality near the locus as this is the lone chromatin mark in the intergenic region[22] (Extended Data Fig. 3a). Chromatin immunoprecipitation followed by sequencing (ChIP-seq) indicate that a variant (rs8017172) in complete linkage disequilibrium with rs945270 ($r^2 = 1.0$) lies within a binding site of the CTCF (CCCTC-binding factor) transcription regulator[23] (Extended Data Fig. 9) in embryonic stem cells. To assess potential functionality in brain tissue, we tested for association with gene expression 1 megabase (Mb) up/downstream. We identified and replicated an effect of rs945270 on the expression of the *KTN1* gene. The C allele, associated with larger putamen volume, also increased expression of *KTN1* in the frontal cortex (discovery sample: 304 neuropathologically normal controls[24] ($P = 4.1 \times 10^{-11}$); replication sample: 134 neuropathologically normal controls ($P = 0.025$)), and putamen (sample: 134 neuropathologically normal controls[25] ($P = 0.049$); Fig. 2a, b). In blood, rs945270 was also strongly associated with *KTN1* expression[26] ($P = 5.94 \times 10^{-31}$; $n = 5{,}311$). After late fetal development, *KTN1* is expressed in the human thalamus, striatum and hippocampus; it is more highly expressed in the striatum than the cortex[13] (Extended Data Fig. 8a). *KTN1* encodes the kinectin receptor facilitating vesicle binding to kinesin, and is heavily involved in organelle transport[11]. Kinectin is only found in the dendrites and soma of neurons, not their axons; neurons with more kinectin have larger cell bodies[27], and kinectin knockdown strongly influences cell shape[28]. The volumetric effects identified here may therefore reflect genetic control of neuronal cell size and/or dendritic complexity. Using three-dimensional surface models of putamen segmentations in MRI scans of 1,541 healthy adolescent subjects, we further localized the allelic effects of rs945270 to regions along the superior and lateral putamen bilaterally, independent of chosen segmentation protocol (Fig. 2c and Extended Data Fig. 10). Each copy of the C allele was associated with an increase in volume along anterior superior regions receiving dense cortical projections from dorsolateral prefrontal cortex and supplementary motor areas[29,30].

In summary, we discovered several common genetic variants underlying variation in different structures within the human brain. Many seem to exert their effects through known developmental pathways including apoptosis, axon guidance and vesicle transport. All structure volumes showed high heritability, but individual genetic variants had diverse effects. The strongest effects were found for putamen and hippocampal volumes, whereas other structures delineated with similar reliability such as the thalamus showed no association with these or other loci (Supplementary Table 4). Discovery of common variants affecting the human brain is now feasible using collaborative analysis of MRI data, and may determine genetic mechanisms driving development and disease.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 April; accepted 19 November 2014.
Published online 21 January 2015.

1. Blokland, G. A., de Zubicaray, G. I., McMahon, K. L. & Wright, M. J. Genetic and environmental influences on neuroimaging phenotypes: a meta-analytical perspective on twin imaging studies. *Twin Res. Hum. Genet.* **15,** 351–371 (2012).
2. Kravitz, A. V. *et al.* Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. *Nature* **466,** 622–626 (2010).
3. Poldrack, R. A. *et al.* Interactive memory systems in the human brain. *Nature* **414,** 546–550 (2001).
4. Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J. & Frith, C. D. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* **442,** 1042–1045 (2006).

5. Stein, J. L. *et al.* Identification of common variants associated with human hippocampal and intracranial volumes. *Nature Genet.* **44,** 552–561 (2012).
6. Ikram, M. A. *et al.* Common variants at 6q22 and 17q21 are associated with intracranial volume. *Nature Genet.* **44,** 539–544 (2012).
7. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467,** 832–838 (2010).
8. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genet.* **42,** 937–948 (2010).
9. van der Sluis, S., Posthuma, D. & Dolan, C. V. TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies. *PLoS Genet.* **9,** e1003235 (2013).
10. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511,** 421–427 (2014).
11. Kumar, J., Yu, H. & Sheetz, M. P. Kinectin, an Essential Anchor for Kinesin-Driven Vesicle Motility. *Science* **267,** 1834–1837 (1995).
12. Hamasaki, T., Goto, S., Nishikawa, S. & Ushio, Y. A role of netrin-1 in the formation of the subcortical structure striatum: repulsive action on the migration of late-born striatal neurons. *J. Neurosci.* **21,** 4272–4280 (2001).
13. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478,** 483–489 (2011).
14. Motoyama, N. *et al.* Massive cell death of immature hematopoietic cells and neurons in Bcl-x-deficient mice. *Science* **267,** 1506–1510 (1995).
15. Itoh, K. *et al.* Apoptosis in the basal ganglia of the developing human nervous system. *Acta Neuropathol.* **101,** 92–100 (2001).
16. Scannevin, R. H. & Huganir, R. L. Postsynaptic organization and regulation of excitatory synapses. *Nature Rev. Neurosci.* **1,** 133–141 (2000).
17. Nithianantharajah, J. *et al.* Synaptic scaffold evolution generated components of vertebrate cognitive complexity. *Nature Neurosci.* **16,** 16–24 (2013).
18. Kirov, G. *et al. De novo* CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol. Psychiatry* **17,** 142–153 (2012).
19. Bis, J. C. *et al.* Common variants at 12q14 and 12q24 are associated with hippocampal volume. *Nature Genet.* **44,** 545–551 (2012).
20. Deans, M. R. *et al.* Control of neuronal morphology by the atypical cadherin Fat3. *Neuron* **71,** 820–832 (2011).
21. Stefansson, H. *et al.* A common inversion under selection in Europeans. *Nature Genet.* **37,** 129–137 (2005).
22. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nature Methods* **9,** 215–216 (2012).
23. Ziebarth, J. D., Bhattacharya, A. & Cui, Y. CTCFBSDB 2.0: a database for CTCF-binding sites and genome organization. *Nucleic Acids Res.* **41,** D188–D194 (2013).
24. Hernandez, D. G. *et al.* Integration of GWAS SNPs and tissue specific expression profiling reveal discrete eQTLs for human traits in blood and brain. *Neurobiol. Dis.* **47,** 20–28 (2012).
25. Ramasamy, A. *et al.* Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nature Neurosci.* **17,** 1418–1428 (2014).
26. Westra, H. J. *et al.* Systematic identification of *trans* eQTLs as putative drivers of known disease associations. *Nature Genet.* **45,** 1238–1243 (2013).
27. Toyoshima, I. & Sheetz, M. P. Kinectin distribution in chicken nervous system. *Neurosci. Lett.* **211,** 171–174 (1996).
28. Zhang, X. *et al.* Kinectin-mediated endoplasmic reticulum dynamics supports focal adhesion growth in the cellular lamella. *J. Cell Sci.* **123,** 3901–3912 (2010).
29. Cohen, M. X., Schoene-Bake, J. C., Elger, C. E. & Weber, B. Connectivity-based segregation of the human striatum predicts personality characteristics. *Nature Neurosci.* **12,** 32–34 (2009).
30. Parent, A. & Hazrati, L. N. Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. *Brain Res. Brain Res. Rev.* **20,** 91–127 (1995).

Supplementary Information is available in the online version of the paper.

Acknowledgements Funding sources for contributing sites and acknowledgments of contributing consortia authors can be found in Supplementary Note 3.

Author Contributions Individual author contributions are listed in Supplementary Note 4.

Author Information Summary statistics from GWAS results are available online using the ENIGMA-Vis web tool: http://enigma.ini.usc.edu/enigma-vis/. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.M.T. (pthomp@usc.edu) or S.E.M. (Sarah.Medland@qimrberghofer.edu.au).

Derrek P. Hibar[1]*, Jason L. Stein[1,2]*, Miguel E. Renteria[3]*, Alejandro Arias-Vasquez[4,5,6,7]*, Sylvane Desrivières[8]*, Neda Jahanshad[1], Roberto Toro[9,10,11], Katharina Wittfeld[12,13], Lucija Abramovic[14], Micael Andersson[15], Benjamin S. Aribisala[16,17,18], Nicola J. Armstrong[19,20], Manon Bernard[21], Marc M. Bohlken[14], Marco P. Boks[14], Janita Bralten[4,6,7], Andrew A. Brown[22,23], M. Mallar Chakravarty[24,25], Qiang Chen[26], Christopher R. K. Ching[1,27], Gabriel Cuellar-Partida[3], Anouk den Braber[28], Sudheer Giddaluru[29,30], Aaron L. Goldman[26], Oliver Grimm[31], Tulio Guadalupe[32,33], Johanna Hass[34], Girma Woldehawariat[35], Avram J. Holmes[36,37], Martine Hoogman[4,7], Deborah Janowitz[13], Tianye Jia[8], Sungeun Kim[38,39,40], Marieke Klein[4,7], Bernd Kraemer[41], Phil H. Lee[37,42,43,44], Loes M. Olde Loohuis[45], Michelle Luciano[46], Christine Macare[8], Karen A. Mather[19], Manuel Mattheisen[47,48,49], Yuri Milaneschi[50], Kwangsik Nho[38,39,40], Martina Papmeyer[51], Adaikalavan Ramasamy[52,53], Shannon L. Risacher[38,40], Roberto Roiz-Santiañez[54,55], Emma J. Rose[56,57], Alireza Salami[15,58], Philipp G. Sämann[59], Lianne Schmaal[50], Andrew J. Schork[60,61], Jean Shin[21], Lachlan T. Strike[3,62,63], Alexander Teumer[64], Marjolein M. J. van Donkelaar[4,7], Kristel R. van Eijk[14], Raymond K. Walters[65,66], Lars T. Westlye[23,67], Christopher D. Whelan[1], Anderson M. Winkler[68,69], Marcel P. Zwiers[7], Saud Alhusaini[70,71], Lavinia Athanasiu[22,23], Stefan Ehrlich[34,37,72], Marina M. H. Hakobjan[4,7], Cecilie B. Hartberg[22,73], Unn K. Haukvik[22], Angelien J. G. A. M. Heister[4,7], David Hoehn[59], Dalia Kasperaviciute[74,75], David C. M. Liewald[46], Lorna M. Lopez[46], Remco R. R. Makkinje[4,7], Mar Matarin[76], Marlies A. M. Naber[4,7], D. Reese McKay[69,77], Margaret Needham[56], Allison C. Nugent[35], Benno Pütz[59], Natalie A. Royle[16,46,18], Li Shen[38,39,40], Emma Sprooten[51,69,77], Daniah Trabzuni[53,78], Saskia S. L. van der Marel[4,7], Kimm J. E. van Hulzen[4,7], Esther Walton[34], Christiane Wolf[59], Laura Almasy[79,80], David Ames[81,82], Sampath Arepalli[83], Amelia A. Assareh[19], Mark E. Bastin[16,18,46,84], Henry Brodaty[19], Kazima B. Bulayeva[85], Melanie A. Carless[79], Sven Cichon[86,87,88,89], Aiden Corvin[56], Joanne E. Curran[79], Michael Czisch[59], Greig I. de Zubicaray[62], Allissa Dillman[83], Ravi Duggirala[79], Thomas D. Dyer[79,80], Susanne Erk[90], Iryna O. Fedko[28], Luigi Ferrucci[91], Tatiana M. Foroud[40,92], Peter T. Fox[80,93], Masaki Fukunaga[94], J. Raphael Gibbs[53,82], Harald H. H. Göring[79], Robert C. Green[95,96], Sebastian Guelfi[53], Narelle K. Hansell[3], Catharina A. Hartman[97], Katrin Hegenscheid[98], Andreas Heinz[89], Dena G. Hernandez[53,82], Dirk J. Heslenfeld[99], Pieter J. Hoekstra[97], Florian Holsboer[59], Georg Homuth[100], Jouke-Jan Hottenga[28], Masashi Ikeda[101], Clifford R. Jack Jr[102], Mark Jenkinson[103], Robert Johnson[104], Ryota Kanai[105,106], Maria Keil[41], Jack W. Kent Jr[79], Peter Kochunov[107], John B. Kwok[108,109], Stephen M. Lawrie[51], Xinmin Liu[35,110], Dan L. Longo[111], Katie L. McMahon[63], Eva Meisenzahl[112], Ingrid Melle[22,23], Sebastian Mohnke[90], Grant W. Montgomery[3], Jeanette C. Mostert[4,7], Thomas W. Mühleisen[87,88,89], Michael A. Nalls[83], Thomas E. Nichols[103,113], Lars G. Nilsson[15], Markus M. Nöthen[87,89], Kazutaka Ohi[114], Rene L. Olvera[92], Rocio Perez-Iglesias[55,115], G. Bruce Pike[116,117], Steven G. Potkin[118], Ivar Reinvang[67], Simone Reppermund[19], Marcella Rietschel[31], Nina Romanczuk-Seiferth[90], Glenn D. Rosen[119,120], Dan Rujescu[112], Knut Schnell[121], Peter R. Schofield[108,109], Colin Smith[122], Vidar M. Steen[29,30], Jessika E. Sussmann[51], Anbupalam Thalamuthu[19], Arthur W. Toga[123], Bryan J. Traynor[83], Juan Troncoso[124], Jessica A. Turner[125], Maria C. Valdés Hernández[84], Dennis van 't Ent[28], Marcel van der Brug[126], Nic J. A. van der Wee[127], Marie-Jose van Tol[128], Dick J. Veltman[50], Thomas H. Wassink[129], Eric Westman[130], Ronald H. Zielke[104], Alan B. Zonderman[131], David G. Ashbrook[132], Reinmar Hager[132], Lu Lu[133,134,135], Francis J. McMahon[35], Derek W. Morris[56,136], Robert W. Williams[133,134], Han G. Brunner[4,7,137], Randy L. Buckner[37,138], Jan K. Buitelaar[6,7,139], Wiepke Cahn[14], Vince D. Calhoun[140,141], Gianpiero L. Cavalleri[71], Benedicto Crespo-Facorro[54,55], Anders M. Dale[142,143], Gareth E. Davies[144], Norman Delanty[71,145], Chantal Depondt[146], Srdjan Djurovic[22,147], Wayne C. Drevets[35,148], Thomas Espeseth[23,67], Randy L. Gollub[37,72,96], Beng-Choon Ho[149], Wolfgang Hoffmann[12,64], Norbert Hosten[98], René S. Kahn[14], Stephanie Le Hellard[29,30], Andreas Meyer-Lindenberg[31], Bertram Müller-Myhsok[59,150,151], Matthias Nauck[152], Lars Nyberg[15], Massimo Pandolfo[146], Brenda W. J. H. Penninx[50], Joshua L. Roffman[37], Sanjay M. Sisodiya[74], Jordan W. Smoller[37,42,43,96], Hans van Bokhoven[4,7], Neeltje E. M. van Haren[14], Henry Völzke[64], Henrik Walter[90], Michael W. Weiner[153], Wei Wen[19], Tonya White[154,155], Ingrid Agartz[22,73,156], Ole A. Andreassen[22,23], John Blangero[79,80], Dorret I. Boomsma[28], Rachel M. Brouwer[14], Dara M. Cannon[35,157], Mark R. Cookson[83], Eco J. C. de Geus[28], Ian J. Deary[46], Gary Donohoe[56,136], Guillén Fernández[6,7], Simon E. Fisher[7,32], Clyde Francks[7,32], David C. Glahn[69,77], Hans J. Grabe[13,158], Oliver Gruber[41,59], John Hardy[53], Ryota Hashimoto[159], Hilleke E. Hulshoff Pol[14], Erik G. Jönsson[22,156], Iwona Kloszewska[160], Simon Lovestone[161,162], Venkata S. Mattay[26,163], Patrizia Mecocci[164], Colm McDonald[157], Andrew M. McIntosh[46,51], Roel A. Ophoff[14,45], Tomas Paus[165,166], Zdenka Pausova[21,167], Mina Ryten[53,52], Perminder S. Sachdev[19,168], Andrew J. Saykin[38,40,90], Andy Simmons[169,170,171], Andrew Singleton[83], Hilkka Soininen[172,173], Joanna M. Wardlaw[16,18,46,84], Michael E. Weale[52], Daniel R. Weinberger[26,174], Hieab H. H. Adams[155,175], Lenore J. Launer[176], Stephan Seiler[177], Reinhold Schmidt[177], Ganesh Chauhan[178], Claudia L. Satizabal[179,180], James T. Becker[181,182,183], Lisa Yanek[184], Sven J. van der Lee[175], Maritza Ebling[72,185], Bruce Fischl[72,185,186], W. T. Longstreth Jr[187], Douglas Greve[72,185], Helena Schmidt[188], Paul Nyquist[189], Louis N. Vinke[72,185], Cornelia M. van Duijn[175], Luting Xue[190], Bernard Mazoyer[191], Joshua C. Bis[192], Vilmundur Gudnason[193], Sudha Seshadri[179,181], M. Arfan Ikram[155,175], The Alzheimer's Disease Neuroimaging Initiative†, The CHARGE Consortium†, EPIGEN†, IMAGEN†, SYS†, Nicholas G. Martin[3]§, Margaret J. Wright[3,62]§, Gunter Schumann[8]§, Barbara Franke[4,5,7]§, Paul M. Thompson[1]§ & Sarah E. Medland[3]§

[1]Imaging Genetics Center, Institute for Neuroimaging & Informatics, Keck School of Medicine of the University of Southern California, Los Angeles, California 90292, USA. [2]Neurogenetics Program, Department of Neurology, UCLA School of Medicine, Los Angeles, California 90095, USA. [3]QIMR Berghofer Medical Research Institute, Brisbane 4006, Australia. [4]Department of Human Genetics, Radboud university medical center, Nijmegen 6500 HB, The Netherlands. [5]Department of Psychiatry, Radboud university medical center, Nijmegen 6500 HB, The Netherlands. [6]Department of Cognitive Neuroscience, Radboud university medical center, Nijmegen 6500 HB, The Netherlands. [7]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen 6500 GL, The Netherlands. [8]MRC-SGDP Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London SE5 8AF, UK. [9]Laboratory of Human Genetics and Cognitive Functions, Institut Pasteur, Paris 75015, France. [10]Centre Nationale de Recherche Scientifique (CNRS) Unité de Recherche Associée (URA) 2182 Genes, Synapses and Cognition, Institut Pasteur, Paris 75015, France. [11]Université Paris

Diderot, Sorbonne Paris Cité, Paris 75015, France. [12]German Center for Neurodegenerative Diseases (DZNE) Rostock/Greifswald, Greifswald 17487, Germany. [13]Department of Psychiatry, University Medicine Greifswald, Greifswald 17489, Germany. [14]Brain Center Rudolf Magnus, Department of Psychiatry, University Medical Center Utrecht, Utrecht, 3584 CX, The Netherlands. [15]Umeå Centre for Functional Brain Imaging (UFBI), Umeå University, Umeå 901 87, Sweden. [16]Brain Research Imaging Centre, University of Edinburgh, Edinburgh EH4 2XU, UK. [17]Department of Computer Science, Lagos State University, Lagos, Nigeria. [18]Scottish Imaging Network, A Platform for Scientific Excellence (SINAPSE) Collaboration, Department of Neuroimaging Sciences, University of Edinburgh, Edinburgh EH4 2XU, UK. [19]Centre for Healthy Brain Ageing, School of Psychiatry, University of New South Wales, Sydney 2052, Australia. [20]School of Mathematics and Statistics, University of Sydney, Sydney 2006, Australia. [21]The Hospital for Sick Children, University of Toronto, Toronto M5G 1X8, Canada. [22]NORMENT - KG Jebsen Centre, Institute of Clinical Medicine, University of Oslo, Oslo N-0316, Norway. [23]NORMENT - KG Jebsen Centre, Division of Mental Health and Addiction, Oslo University Hospital, Oslo 0424, Norway. [24]Cerebral Imaging Centre, Douglas Mental Health University Institute, Montreal H4H 1R3, Canada. [25]Department of Psychiatry and Biomedical Engineering, McGill University, Montreal H3A 2B4, Canada. [26]Lieber Institute for Brain Development, Baltimore, Maryland 21205, USA. [27]Interdepartmental Neuroscience Graduate Program, UCLA School of Medicine, Los Angeles, California 90095, USA. [28]Biological Psychology, Neuroscience Campus Amsterdam & EMGO Institute for Health and Care Research, VU University & VU Medical Center, Amsterdam 1081 BT, The Netherlands. [29]NORMENT - KG Jebsen Centre for Psychosis Research, Department of Clinical Science, University of Bergen, 5021 Bergen, Norway. [30]Dr. Einar Martens Research Group for Biological Psychiatry, Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen 5021, Norway. [31]Central Institute of Mental Health, Medical Faculty Mannheim, University Heidelberg, Mannheim 68159, Germany. [32]Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen 6525 XD, The Netherlands. [33]International Max Planck Research School for Language Sciences, Nijmegen 6525 XD, The Netherlands. [34]Department of Child and Adolescent Psychiatry, Faculty of Medicine of the TU Dresden, Dresden 01307 Germany. [35]Human Genetics Branch and Experimental Therapeutics and Pathophysiology Branch, National Institute of Mental Health Intramural Research Program, Bethesda, Maryland 20892, USA. [36]Department of Psychology, Yale University, New Haven, Connecticut 06511, USA. [37]Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts 02115, USA. [38]Center for Neuroimaging, Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. [39]Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. [40]Indiana Alzheimer Disease Center, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. [41]Center for Translational Research in Systems Neuroscience and Psychiatry, Department of Psychiatry and Psychotherapy, University Medical Center, Goettingen 37075, Germany. [42]Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02115, USA. [43]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Boston, Massachusetts 02141, USA. [44]Department of Psychiatry, Harvard Medical School, Boston, Massachusetts 02115, USA. [45]Center for Neurobehavioral Genetics, University of California, Los Angeles, California 90095, USA. [46]Centre for Cognitive Ageing and Cognitive Epidemiology, Psychology, University of Edinburgh, Edinburgh EH8 9JZ, UK. [47]Department of Biomedicine, Aarhus University, Aarhus DK-8000, Denmark. [48]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Aarhus and Copenhagen DK-8000, Denmark. [49]Center for integrated Sequencing, iSEQ, Aarhus University, Aarhus DK-8000, Denmark. [50]Department of Psychiatry, Neuroscience Campus Amsterdam, VU University Medical Center/GGZ inGeest, Amsterdam 1081 HL, The Netherlands. [51]Division of Psychiatry, Royal Edinburgh Hospital, University of Edinburgh, Edinburgh EH10 5HF, UK. [52]Department of Medical and Molecular Genetics, King's College London, London SE1 9RT, UK. [53]Reta Lila Weston Institute and Department of Molecular Neuroscience, UCL Institute of Neurology, London WC1N 3BG, UK. [54]Department of Psychiatry, University Hospital Marqués de Valdecilla, School of Medicine, University of Cantabria-IDIVAL, Santander 39008, Spain. [55]Cibersam (Centro Investigación Biomédica en Red Salud Mental), Madrid 28029, Spain. [56]Neuropsychiatric Genetics Research Group and Department of Psychiatry, Trinity College Institute of Psychiatry, Trinity College Dublin, Dublin 2, Ireland. [57]Center for Translational Research on Adversity, Neurodevelopment and Substance Abuse (C-TRANS), Department of Psychiatry, University of Maryland School of Medicine, Baltimore, Maryland 21045, USA. [58]Aging Research Center, Karolinska Institutet and Stockholm University, 11330 Stockholm, Sweden. [59]Max Planck Institute of Psychiatry, Munich 80804, Germany. [60]Multimodal Imaging Laboratory, Department of Neurosciences, University of California, San Diego, California 92093, USA. [61]Department of Cognitive Sciences, University of California, San Diego, California 92161, USA. [62]School of Psychology, University of Queensland, Brisbane 4072, Australia. [63]Centre for Advanced Imaging, University of Queensland, Brisbane 4072, Australia. [64]Institute for Community Medicine, University Medicine Greifswald, Greifswald D-17475, Germany. [65]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [66]Medical and Population Genetics Program, Broad Institute of Harvard and MIT, Boston, Massachusetts 02142, USA. [67]Department of Psychology, University of Oslo, Oslo 0373, Norway. [68]The Oxford Centre for Functional MRI of the Brain, Nuffield Department of Clinical Neurosciences, Oxford University, Oxford OX3 9DU, UK. [69]Department of Psychiatry, Yale School of Medicine, New Haven, Connecticut 06511, USA. [70]Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montreal H3A 2B4, Canada. [71]Molecular and Cellular Therapeutics, The Royal College of Surgeons, Dublin 2, Ireland. [72]The Athinoula A.Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA. [73]Department of Psychiatric Research and Development, Diakonhjemmet Hospital, Oslo 0319, Norway. [74]UCL Institute of Neurology, London, United Kingdom and Epilepsy Society, London WC1N 3BG, UK. [75]Department of Medicine, Imperial College London, London W12 0NN, UK. [76]Department of Clinical and Experimental Epilepsy, UCL Institute of Neurology, London WC1N 3BG, UK. [77]Olin Neuropsychiatric Research Center, Institute of Living, Hartford Hospital, Hartford, Connecticut 06106, USA. [78]Department of Genetics, King Faisal Specialist Hospital and Research Centre, Riyadh 11211, Saudi Arabia. [79]Texas Biomedical Research Institute, San Antonio, Texas 78245, USA. [80]University of Texas Health Science Center, San Antonio, Texas 78229, USA. [81]National Ageing Research Institute, Royal Melbourne Hospital, Melbourne 3052, Australia. [82]Academic Unit for Psychiatry of Old Age, University of Melbourne, Melbourne 3101, Australia. [83]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland 20892, USA. [84]Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh EH4 2XU, UK. [85]N.I. Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow 119333, Russia. [86]Division of Medical Genetics, Department of Biomedicine, University of Basel, Basel 4055, Switzerland. [87]Institute of Human Genetics, University of Bonn, Bonn, D-53127, Germany. [88]Institute of Neuroscience and Medicine (INM-1), Research Centre Jülich, Jülich, D-52425, Germany. [89]Department of Genomics, Life & Brain Center, University of Bonn, Bonn D-53127, Germany. [90]Department of Psychiatry and Psychotherapy, Charité Universitätsmedizin Berlin, CCM, Berlin 10117, Germany. [91]Clinical Research Branch, National Institute on Aging, Baltimore, Maryland 20892, USA. [92]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. [93]South Texas Veterans Health Care System, San Antonio, Texas 78229, USA. [94]Biofunctional Imaging, Immunology Frontier Research Center, Osaka University, Osaka 565-0871, Japan. [95]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. [96]Harvard Medical School, Boston, Massachusetts 02115, USA. [97]Department of Psychiatry, University of Groningen, University Medical Center Groningen, 9713 GZ Groningen, The Netherlands. [98]Institute of Diagnostic Radiology and Neuroradiology, University Medicine Greifswald, Greifswald 17475, Germany. [99]Departments of Cognitive and Clinical Neuropsychology, VU University Amsterdam, 1081 BT Amsterdam, The Netherlands [100]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald 17489, Germany. [101]Department of Psychiatry, Fujita Health University School of Medicine, Toyoake 470-1192, Japan. [102]Radiology, Mayo Clinic, Rochester, Minnesota 55905, USA. [103]FMRIB Centre, University of Oxford, Oxford OX3 9DU, UK. [104]NICHD Brain and Tissue Bank for Developmental Disorders, University of Maryland Medical School, Baltimore, Maryland 21201, USA. [105]School of Psychology, University of Sussex, Brighton BN1 9QH, UK. [106]Institute of Cognitive Neuroscience, University College London, London WC1N 3AR, UK. [107]Department of Psychiatry, Maryland Psychiatric Research Center, University of Maryland, Baltimore, Maryland 21201, USA. [108]Neuroscience Research Australia, Sydney 2031, Australia. [109]School of Medical Sciences, UNSW, Sydney 2052, Australia. [110]Department of Pathology and Cell Biology, Columbia University Medical Center, New York 10032, USA. [111]Lymphocyte Cell Biology Unit, Laboratory of Genetics, National Institute on Aging, National Institutes of Health, Baltimore, Maryland 21224, USA. [112]Department of Psychiatry, Ludwig-Maximilians-Universität, Munich 80336, Germany. [113]Department of Statistics & WMG, University of Warwick, Coventry CV4 7AL, UK. [114]Department of Psychiatry, Osaka University Graduate School of Medicine, Osaka 565-0871, Japan. [115]Institute of Psychiatry, King's College London, London SE5 8AF, UK. [116]Department of Neurology, University of Calgary, Calgary T2N 2T9, Canada. [117]Department of Clinical Neuroscience, University of Calgary, Calgary T2N 2T9, Canada. [118]Psychiatry and Human Behavior, University of California, Irvine, California 92617, USA. [119]Department of Neurology, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA. [120]Harvard Medical School, Boston, Massachusetts 02115, USA. [121]Department of General Psychiatry, Heidelberg University Hospital, Heidelberg 69115, Germany. [122]Department of Neuropathology, MRC Sudden Death Brain Bank Project, University of Edinburgh, Edinburgh EH8 9AG, UK. [123]Laboratory of Neuro Imaging, Institute for Neuroimaging and Informatics, Keck School of Medicine of the University of Southern California, Los Angeles, California 90033, USA. [124]Department of Pathology, Johns Hopkins University, Baltimore, Maryland 21287, USA. [125]Psychology Department and Neuroscience Institute, Georgia State University, Atlanta, Georgia 30302, USA. [126]Genentech, South San Francisco, California 94080, USA [127]Psychiatry and Leiden Institute for Brain and Cognition, Leiden University Medical Center, Leiden 2333 ZA, The Netherlands. [128]Neuroimaging Centre, University of Groningen, University Medical Center Groningen, Groningen 9713 AW, The Netherlands. [129]Department of Psychiatry, Carver College of Medicine, University of Iowa, Iowa City, Iowa 52242, USA. [130]Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm SE-141 83, Sweden. [131]Behavioral Epidemiology Section, National Institute on Aging Intramural Research Program, Baltimore, Maryland 20892, USA. [132]Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, UK. [133]Center for Integrative and Translational Genomics, University of Tennessee Health Science Center, Memphis, Tennessee 38163, USA. [134]Department of Genetics, Genomics, and Informatics, University of Tennessee Health Science Center, Memphis, Tennessee 38163, USA. [135]Jiangsu Province Key Laboratory for Inflammation and Molecular Drug Target, Medical College of Nantong University, Nantong 226001, China. [136]Cognitive Genetics and Therapy Group, School of Psychology & Discipline of Biochemistry, National University of Ireland Galway, Galway, Ireland. [137]Department of Clinical Genetics, Maastricht University Medical Center, Maastricht 6200 MD, The Netherlands. [138]Department of Psychology, Center for Brain Science, Harvard University, Boston, Massachusetts 02138, USA. [139]Karakter Child and Adolescent Psychiatry, Radboud university medical center, Nijmegen 6500 HB, The Netherlands. [140]The Mind Research Network & LBERI, Albuquerque, New Mexico 87106, USA. [141]Department of ECE, University of New Mexico, Albuquerque, New Mexico 87131, USA. [142]Center for Translational Imaging and Personalized Medicine, University of California, San Diego, California 92093, USA. [143]Departments of Neurosciences, Radiology, Psychiatry, and Cognitive Science, University of California, San Diego, California 92093, USA. [144]Avera Institute for Human Genetics, Sioux Falls, South Dakota, 57108, USA. [145]Neurology Division, Beaumont Hospital, Dublin 9, Ireland. [146]Department of Neurology, Hopital Erasme, Universite Libre de Bruxelles, Brussels 1070, Belgium. [147]Department of Medical Genetics, Oslo University Hospital, Oslo 0450, Norway. [148]Janssen Research & Development, Johnson & Johnson, Titusville, New Jersey 08560, USA. [149]Department of

Psychiatry, University of Iowa, Iowa City, Iowa 52242, USA. [150]Munich Cluster for Systems Neurology (SyNergy), Munich 81377, Germany. [151]University of Liverpool, Institute of Translational Medicine, Liverpool L69 3BX, UK. [152]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald 17475, Germany. [153]Center for Imaging of Neurodegenerative Disease, San Francisco VA Medical Center, University of California, San Francisco, California 94121, USA. [154]Department of Child and Adolescent Psychiatry, Erasmus University Medical Centre, Rotterdam 3000 CB, The Netherlands. [155]Department of Radiology, Erasmus University Medical Centre, Rotterdam 3015 CN, The Netherlands. [156]Department of Clinical Neuroscience, Psychiatry Section, Karolinska Institutet, Stockholm SE-171 76, Sweden. [157]Clinical Neuroimaging Laboratory, College of Medicine, Nursing and Health Sciences, National University of Ireland Galway, Galway, Ireland. [158]Department of Psychiatry and Psychotherapy, HELIOS Hospital Stralsund 18435, Germany. [159]Molecular Research Center for Children's Mental Development, United Graduate School of Child Development, Osaka University, Osaka 565-0871, Japan. [160]Medical University of Lodz, Lodz 90-419, Poland. [161]Department of Psychiatry, University of Oxford, Oxford OX3 7JX, UK. [162]NIHR Dementia Biomedical Research Unit, King's College London, London SE5 8AF, UK. [163]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [164]Section of Gerontology and Geriatrics, Department of Medicine, University of Perugia, Perugia 06156, Italy. [165]Rotman Research Institute, University of Toronto, Toronto M6A 2E1, Canada. [166]Departments of Psychology and Psychiatry, University of Toronto, Toronto M5T 1R8, Canada. [167]Departments of Physiology and Nutritional Sciences, University of Toronto, Toronto M5S 3E2, Canada. [168]Neuropsychiatric Institute, Prince of Wales Hospital, Sydney 2031, Australia. [169]Department of Neuroimaging, Institute of Psychiatry, King's College London, London SE5 8AF, UK. [170]Biomedical Research Centre for Mental Health, King's College London, London SE5 8AF, UK. [171]Biomedical Research Unit for Dementia, King's College London, London SE5 8AF, UK. [172]Institute of Clinical Medicine, Neurology, University of Eastern Finland, Kuopio FI-70211, Finland. [173]Neurocentre Neurology, Kuopio University Hospital, Kuopio FI-70211, Finland. [174]Departments of Psychiatry, Neurology, Neuroscience and the Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [175]Department of Epidemiology, Erasmus University Medical Centre, Rotterdam 3015 CN, The Netherlands. [176]Laboratory of Epidemiology and Population Sciences, Intramural Research Program, National Institute on Aging, Bethesda, Maryland 20892, USA. [177]Department of Neurology, Clinical Division of Neurogeriatrics, Medical University Graz, Graz 8010, Austria. [178]INSERM U897, University of Bordeaux, Bordeaux 33076, France. [179]Department of Neurology, Boston University School of Medicine, Boston, Massachusetts 02118, USA. [180]Framingham Heart Study, Framingham, Massachusetts 01702, USA. [181]Department of Neurology, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA. [182]Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA. [183]Department of Psychology, Dietrich School of Arts and Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA. [184]General Internal Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, USA. [185]Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA. [186]Computer Science and AI Lab, Massachusetts Institute of Technology, Boston, Massachusetts 02141, USA. [187]Department of Neurology University of Washington, Seattle, Washington 98195, USA. [188]Institute of Molecular Biology and Biochemistry, Medical University Graz, 8010 Graz, Austria. [189]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [190]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts 02118, USA. [191]Groupe d'Imagerie Neurofonctionnelle, UMR5296 CNRS, CEA and University of Bordeaux, Bordeaux 33076, France. [192]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington 98101, USA. [193]Icelandic Heart Association, University of Iceland, Faculty of Medicine, Reykjavik 101, Iceland.

†A list of authors and affiliations appears in the Supplementary Information.
*These authors contributed equally to this work.
§These authors jointly supervised this work.

METHODS

Details of the GWAS meta-analysis are outlined in Extended Data Fig. 1. All participants in all cohorts in this study gave written informed consent and sites involved obtained approval from local research ethics committees or Institutional Review Boards. The ENIGMA consortium follows a rolling meta-analysis framework for incorporating sites into the analysis. The discovery sample comprises studies of European ancestry (Extended Data Fig. 2) that contributed GWAS summary statistics for the purpose of this analysis on or before 1 October 2013. The deadline for discovery samples to upload their data was made before inspecting the data and was not influenced by the results of the analyses. The meta-analysed results from discovery cohorts were carried forward for secondary analyses and functional validation studies. Additional samples of European ancestry were gathered to provide *in silico* or single genotype replication of the strongest associations as part of the replication sample. A generalization sample of sites with non-European ancestry was used to examine the effects across ethnicities. In all, data were contributed from 50 cohorts, each of which is detailed in Supplementary Tables 1–3.

The brain measures examined in this study were obtained from structural MRI data collected at participating sites around the world. Brain scans were processed and examined at each site locally, following a standardized protocol procedure to harmonize the analysis across sites. The standardized protocols for image analysis and quality assurance are openly available online (http://enigma.ini.usc.edu/protocols/imaging-protocols/). The subcortical brain measures (nucleus accumbens, amygdala, caudate nucleus, hippocampus, pallidum, putamen and thalamus) were delineated in the brain using well-validated, freely available brain segmentation software packages: FIRST[31], part of the FMRIB Software Library (FSL), or FreeSurfer[32]. The agreement between the two software packages has been well documented in the literature[5,33] and was further detailed here (Supplementary Table 4). Participating sites used the software package most suitable for their data set (the software used at each site is given in Supplementary Table 2) without selection based on genotype or the associations present in this study. In addition to the subcortical structures of the brain, we examined the genetic effects of a measure of global head size, the ICV. The ICV was calculated as: 1/(determinant of a rotation-translation matrix obtained after affine registration to a common study template and multiplied by the template volume (1,948,105 mm^3)). After image processing, each image was inspected individually to identify poorly segmented structures. Each site contributed histograms of the distribution of volumes for the left and right hemisphere structures (and a measure of asymmetry) of each subcortical region used in the analysis. Scans marked as outliers (>3 standard deviations from the mean) based on the histogram plots were re-checked at each site to locate any errors. If a scan had an outlier for a given structure, but was segmented properly, it was retained in the analysis. Site-specific phenotype histograms, Manhattan plots and quantile–quantile plots from each participating site are available on the ENIGMA website (http://enigma.ini.usc.edu/publications/enigma-2/).

Each study in the discovery sample was genotyped using commercially available platforms. Before imputation, genetic homogeneity was assessed in each sample using multi-dimensional scaling (MDS) analysis (Extended Data Fig. 2). Ancestry outliers were excluded through visual inspection of the first two components. Quality control filtering was applied to remove genotyped SNPs with low minor allele frequency (<0.01), poor genotype call rate (<95%), and deviations from Hardy–Weinberg equilibrium ($P < 1 \times 10^{-6}$) before imputation. The imputation protocols used MaCH[34] for haplotype phasing and minimac[35] for imputation and are freely available online (http://enigma.ini.usc.edu/protocols/genetics-protocols/). Full details of quality control procedures and any deviations from the imputation protocol are given in Supplementary Table 3.

Genome-wide association scans were conducted at each site for all eight traits of interest including the ICV and bilateral volumes of the nucleus accumbens, amygdala, caudate nucleus, hippocampus, pallidum, putamen and thalamus. For each SNP in the genome, the additive dosage value was regressed against the trait of interest separately using a multiple linear regression framework controlling for age, age^2, sex, 4 MDS components, ICV (for non-ICV phenotypes) and diagnosis (when applicable). For studies with data collected from several centres or scanners, dummy-coded covariates were also included in the model. Sites with family data (NTR-Adults, BrainSCALE, QTIM, SYS, GOBS, ASPSFam, ERF, GeneSTAR, NeuroIMAGE and OATS) used mixed-effects models to control for familial relationships in addition to covariates stated previously. The primary analyses for this paper focused on the full set of subjects including data sets with patients to maximize the power to detect effects. We re-analysed the data excluding patients to verify that detected effects were not due to disease alone (Extended Data Fig. 5a). The protocols used for testing association with mach2qtl (ref. 34) for studies with unrelated subjects and merlin-offline[36] for family-based designs are freely available online (http://enigma.ini.usc.edu/protocols/genetics-protocols/). Full details for the software used at each site are given in Supplementary Table 3.

The GWAS results from each site were uploaded to a centralized server for quality checking and processing. Results files from each cohort were free from genomic inflation in quantile–quantile plots and Manhattan plots (http://enigma.ini.usc.edu/publications/enigma-2/). Poorly imputed SNPs (with $R^2 < 0.5$) and low minor allele count (<10) were removed from the GWAS result files from each site. The resulting files were combined meta-analytically using a fixed-effect, inverse-variance-weighted model as implemented in the software package METAL[37]. The discovery cohorts were meta-analysed first, controlling for genomic inflation. The combined discovery data set (comprised of all meta-analysed SNPs with data from at least 5,000 subjects) was carried forward for the additional analyses detailed below.

To account appropriately for multiple comparisons over the eight traits in our analysis, we first examined the degree of independence between each trait. We generated an 8 × 8 correlation matrix based on the Pearson's correlation between all pair-wise combinations of the mean volumes of each structure in the QTIM study. Using the matSpD software[38] we found that the effective number of independent traits in our analysis was 7. We therefore set a significance criteria threshold of $P < (5 \times 10^{-8}/7) = 7.1 \times 10^{-9}$.

Heritability estimates for mean volumes of each of the eight structures in this study were calculated using structural equation modelling in OpenMx[39]. Twin modelling was performed controlling for age and sex differences on a large sample ($n = 1{,}030$) of healthy adolescent and young adult twins (148 monozygotic and 202 dizygotic pairs) and their siblings from the Queensland Twin Imaging (QTIM) study. Subsequently, a multivariate analysis showed that common environmental factors (C) could be dropped from the model without a significant reduction in the goodness-of-fit ($\Delta\chi^2_{36} = 29.81$; $P = 0.76$). Heritability (h^2) was significantly different from zero for all eight brain measures: putamen ($h^2 = 0.89$; 95% confidence interval 0.85–0.92), thalamus ($h^2 = 0.88$; 0.85–0.92), ICV ($h^2 = 0.88$; 0.84–0.90), hippocampus ($h^2 = 0.79$; 0.74–0.83), caudate nucleus ($h^2 = 0.78$; 0.75–0.82), pallidum ($h^2 = 0.75$; 0.72–0.78), nucleus accumbens ($h^2 = 0.49$; 0.45–0.55), amygdala ($h^2 = 0.43$; 0.39, 0.48) (Extended Data Fig. 11a).

Percentage variance explained by each genome-wide significant SNP was determined based on the final combined discovery data set (Extended Data Fig. 6a) or the discovery combined with the replication samples (Table 1) after correction for covariates using the following equation:

$$R^2_{g|c}/(1-R^2_c) = (t^2/((n-k-1)+t^2)) * 100$$

where the t-statistic is calculated as the beta coefficient for a given SNP from the regression model (controlling for covariates) divided by the standard error of the beta estimate, and where n is the total number of subjects and k is the total number of covariates included in the model ($k = 10$) (ref. 40). $R^2_{g|c}$ is the variance explained by the variant controlling for covariates and R^2_c is the variance explained by the covariates alone. $R^2_{g|c}/(1 - R^2_c)$ gives the variance explained by the genetic variant after accounting for covariate effects. The total variance explained by the GWAS (Extended Data Fig. 11b, c) was calculated by first linkage disequilibrium pruning the results without regard to significance (pruning parameters in PLINK:––indep-pairwise 1000kb 25 0.1). The t-statistics of the regression coefficients from the pruned results are then corrected for the effects of 'winner's curse' and the variance explained by each SNP after accounting for covariate effects is summed across SNPs using freely available code (http://sites.google.com/site/honcheongso/software/total-vg)[40,41]. As the correction for winners curse may be influenced by asymmetry in the distribution of t (arising from the choice of reference allele) we bootstrapped the choice of reference allele (5,000 iterations) to derive the median value and 95% confidence intervals of the estimates of variance explained (Extended Data Fig. 11b, c). The correction for winner's curse corrected for upward biases when estimating the percentage variance explained by each SNP across the genome via simulation[40], but this correction could still allow some bias. Future large studies will be able to evaluate independently the percentage variance explained.

We performed multivariate GWAS using the Trait-based Association Test that uses Extended Simes procedure (TATES)[9]. For the TATES analysis we used GWAS summary statistics from the discovery data set and the correlation matrix created from the eight phenotypes using the QTIM data set (Extended Data Fig. 6c).

We examined the moderating effects of mean age and proportion of females on the effect sizes estimated for the top loci influencing brain volumes (Extended Data Fig. 5b, c) using a mixed-effect meta-regression model such that:

$$\text{effect} = \beta_0 + \beta_{\text{mod}} X_{\text{mod}} + \varepsilon + \eta$$

In this model, the effect and variance at each site are treated as random effects and the moderator X_{mod} (either mean age or proportion of females) is treated as a fixed effect. Meta-regression tests were performed using the metafor package (version 1.9-1) in R.

Hierarchical clustering was performed on the GWAS t-statistics from the discovery data set results using independent SNPs clumped from the TATES results (clumping parameters: significance threshold for index SNP = 0.01, significance

threshold for clumped SNPs = 0.01, $r^2 = 0.25$, physical distance = 1 Mb; Extended Data Fig. 6b). Regions with the strongest genetic similarity were grouped together based on the strength of their pairwise correlations. The results were represented visually using hierarchical clustering with default settings from the gplots package (version 2.12.1) in R.

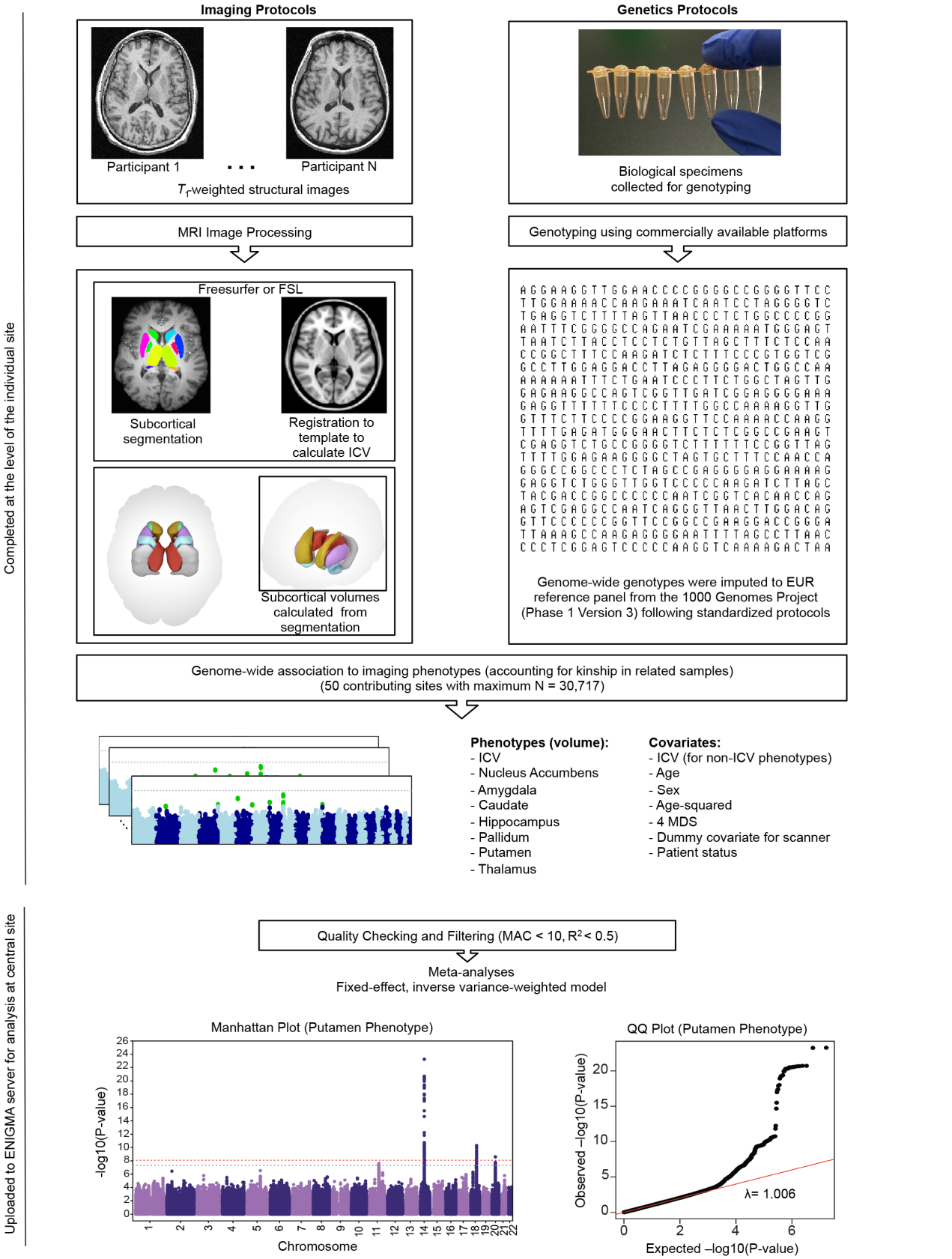
Gene annotation, gene-based test statistics and pathway analysis were performed using the KGG2.5 software package[42] (Supplementary Table 7 and Extended Data Fig. 7). Linkage disequilibrium was calculated based on RSID numbers using the 1000 Genomes Project European samples as a reference (http://enigma.ini.usc.edu/protocols/genetics-protocols/). For the annotation, SNPs were considered 'within' a gene if they fell within 5 kb of the 3′/5′ untranslated regions based on human genome (hg19) coordinates. Gene-based tests were performed using the GATES test[42] without weighting *P* values by predicted functional relevance. Pathway analysis was performed using the hybrid set-based test (HYST) of association[43]. For all gene-based tests and pathway analyses, results were considered significant if they exceeded a Bonferroni correction threshold accounting for the number of pathways and traits tested such that $P_{\text{thresh}} = 0.05/(671 \text{ pathways} \times 7 \text{ independent traits}) = 1.06 \times 10^{-5}$.

Expression quantitative loci were examined in two independent data sets: the NABEC (GSE36192)[24] and UKBEC (GSE46706)[44,45]. Detailed processing and exclusion criteria for both data sets are described elsewhere[24,45]. In brief, the UKBEC consists of 134 neuropathologically normal donors from the MRC Sudden Death Brain Bank in Edinburgh and Sun Health Research Institute; expression was profiled on the Affymetrix Exon 1.0 ST array. The NABEC is comprised of 304 neurologically normal donors from the National Institute of Ageing and expression profiled on the Illumina HT12v3 array. The expression values were corrected for gender and batch effects and probes that contained polymorphisms (seen >1% in European 1000G) were excluded from analyses[44]. Blood expression quantitative trait loci (eQTL) data were queried using the Blood eQTL Browser (http://genenetwork.nl/bloodeqtlbrowser/)[26]. Brain expression over the lifespan was measured from a spatio-temporal atlas of human gene expression and graphed using custom R scripts (GSE25219; details given in[13]).

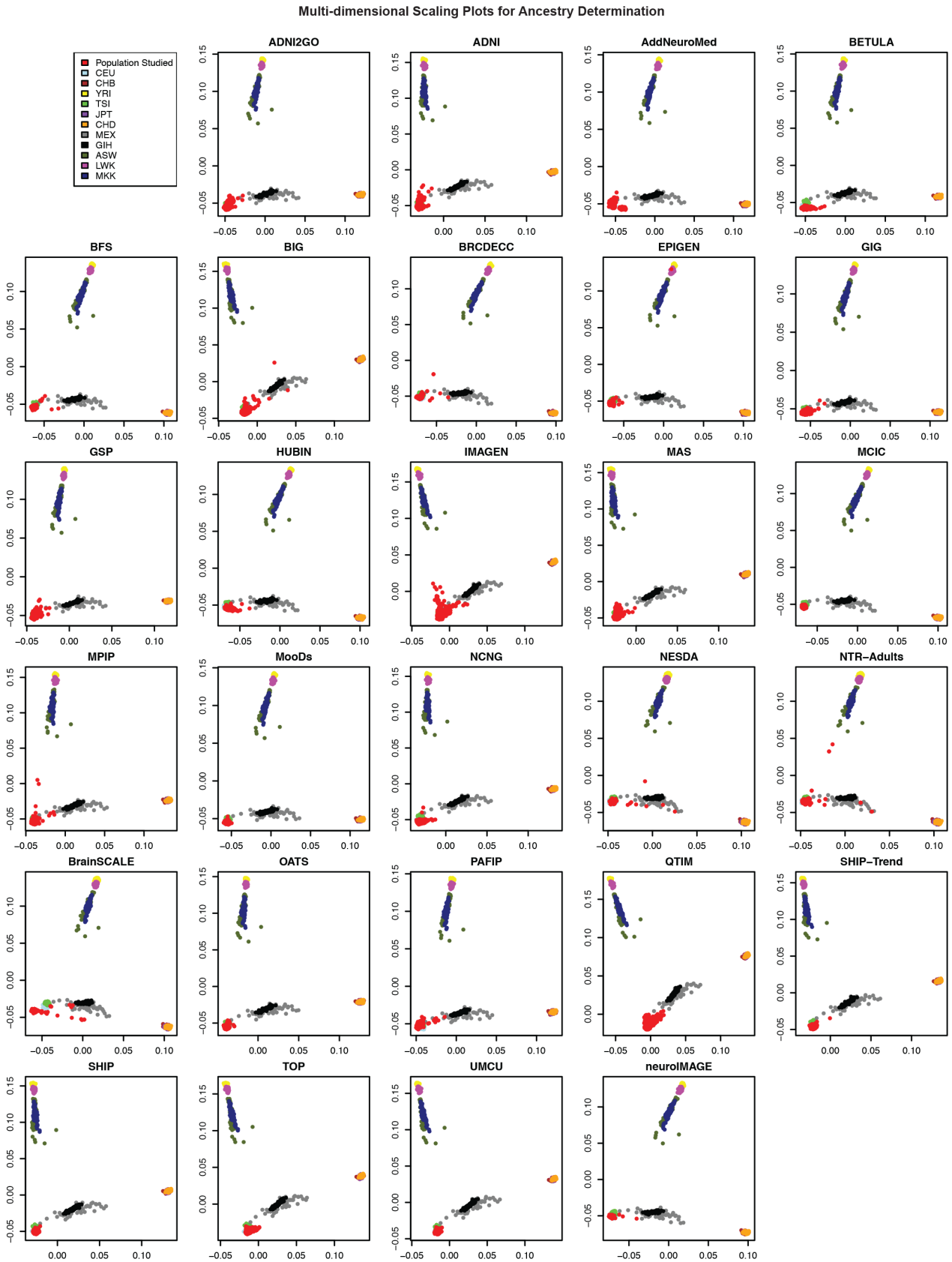
Fine-grained three-dimensional surface mappings of the putamen were generated using a medial surface modelling method[46,47] in 1,541 healthy subjects from the IMAGEN study[48] (Fig. 2c and Extended Data Fig. 10a, b). Putamen volume segmentations from either FSL (Fig. 2c and Extended Data Fig. 10a) or FreeSurfer (Extended Data Fig. 10b) were first converted to three-dimensional meshes and then co-registered to an average template for statistical analysis. The medial core distance was used as a measure of shape and was calculated as the distance from each point on the surface to the centre of the putamen. At each point along the surface of the putamen, an association test was performed using multiple linear regression in which the medial core distance at a given point on the surface was the outcome measure and the additive dosage value of the top SNP was the predictor of interest while including the same covariates that were used for volume including age, sex, age^2, 4 MDS, ICV and site.

In Extended Data Fig. 3, all tracks were taken from the UCSC Genome Browser Human hg19 assembly. SNPs (top 5%) shows the top 5% associated SNPs within the locus and are coloured by their correlation to the top SNP. Genes shows the gene models from GENCODE version 19. Conservation was defined at each base through the phyloP algorithm which assigns scores as $-\log_{10} P$ values under a null hypothesis of neutral evolution calculated from pre-computed genomic alignment of 100 vertebrate species[49]. Conserved sites are assigned positive scores, while faster-than-neutral evolving sites are given negative scores. TFBS conserved shows computationally predicted transcription factor binding sites using the Transfac Matrix Database (v.7.0) found in human, mouse and rat. Brain histone (1.3 year) and brain histone (68 year) show maps of histone trimethylation at histone H3 Lys 4 (H3K4me3), an epigenetic mark for transcriptional activation, measured by ChIP-seq. These measurements were made in neuronal nuclei (NeuN+) collected from prefrontal cortex of post-mortem human brain[50]. CpG methylation was generated using methylated DNA immunoprecipitation and sequencing from postmortem human frontal cortex of a 57-year-old male[51]. DNaseI hypersens displays DNaseI hypersensitivity, evidence of open chromatin, which was evaluated in postmortem human frontal cerebrum from three donors (age 22–35), through the ENCODE consortium[52]. Finally, hES Chrom State gives the predicted chromatin states based on computational integration of ChIP-seq data for nine chromatin marks in H1 human embryonic stem cell lines derived in the ENCODE consortium[53].

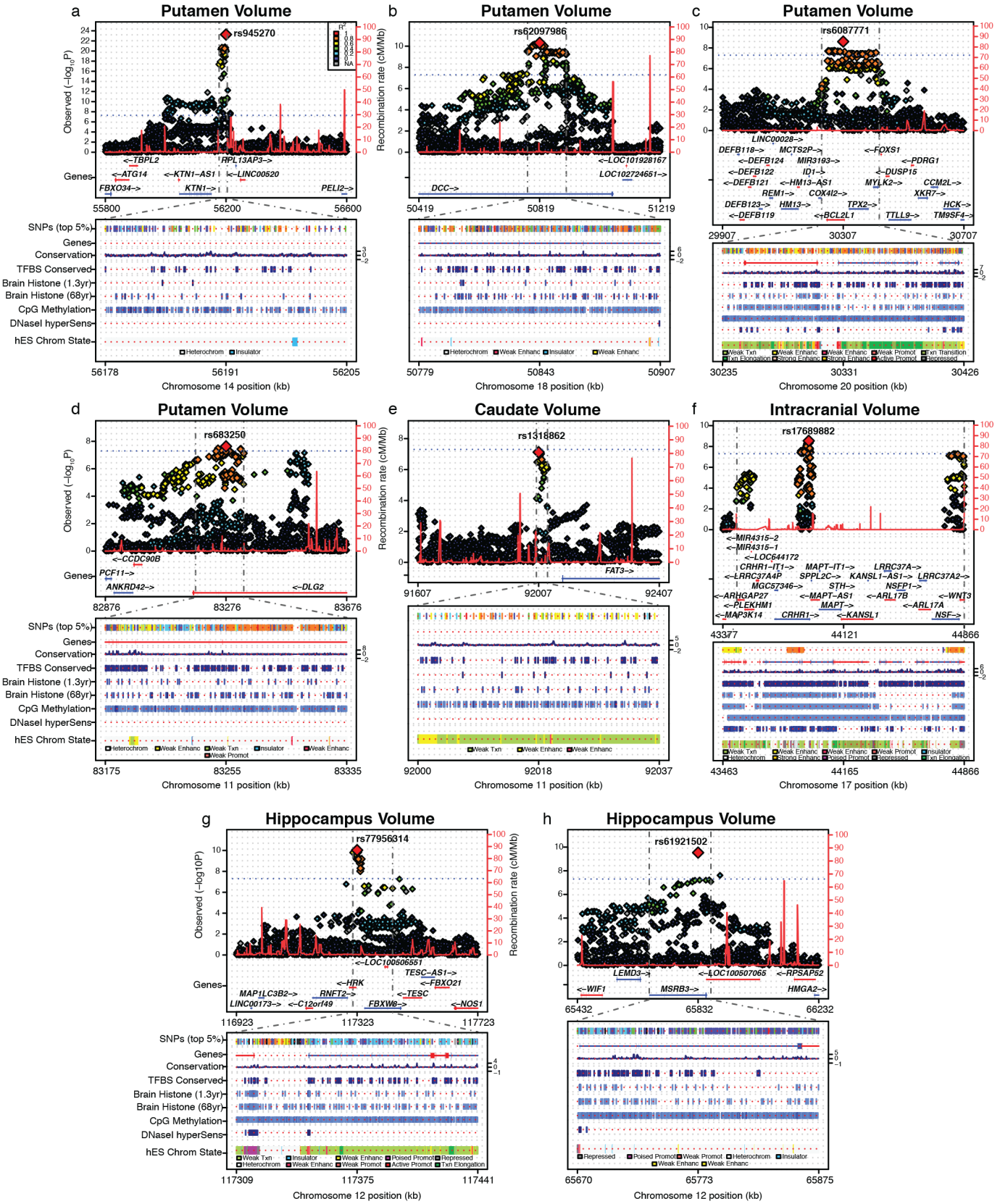
31. Patenaude, B., Smith, S. M., Kennedy, D. N. & Jenkinson, M. A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage* **56,** 907–922 (2011).
32. Fischl, B. *et al.* Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33,** 341–355 (2002).
33. Morey, R. A. *et al.* Scan-rescan reliability of subcortical brain volumes derived from automated segmentation. *Hum. Brain Mapp.* **31,** 1751–1762 (2010).
34. Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34,** 816–834 (2010).
35. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genet.* **44,** 955–959 (2012).
36. Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genet.* **30,** 97–101 (2002).
37. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26,** 2190–2191 (2010).
38. Nyholt, D. R. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am. J. Hum. Genet.* **74,** 765–769 (2004).
39. Boker, S. *et al.* OpenMx: an open source extended structural equation modeling framework. *Psychometrika* **76,** 306–317 (2011).
40. Walters, R., Bartels, M. & Lubke, G. Estimating variance explained by all variants in meta-analysis with heterogeneity. *Behav. Genet.* **43,** 543 (2013).
41. So, H. C., Li, M. & Sham, P. C. Uncovering the total heritability explained by all true susceptibility variants in a genome-wide association study. *Genet. Epidemiol.* **35,** 447–456 (2011).
42. Li, M. X., Gui, H. S., Kwan, J. S. & Sham, P. C. GATES: a rapid and powerful gene-based association test using extended Simes procedure. *Am. J. Hum. Genet.* **88,** 283–293 (2011).
43. Li, M. X., Kwan, J. S. & Sham, P. C. HYST: a hybrid set-based test for genome-wide association studies, with application to protein-protein interaction-based association analysis. *Am. J. Hum. Genet.* **91,** 478–488 (2012).
44. Ramasamy, A. *et al.* Resolving the polymorphism-in-probe problem is critical for correct interpretation of expression QTL studies. *Nucleic Acids Res.* **41,** e88 (2013).
45. Trabzuni, D. *et al.* Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies. *J. Neurochem.* **119,** 275–282 (2011).
46. Gutman, B. A. *et al.* Maximizing power to track Alzheimer's disease and MCI progression by LDA-based weighting of longitudinal ventricular surface features. *Neuroimage* **70,** 386–401 (2013).
47. Gutman, B. A., Wang, Y. L., Rajagopalan, P., Toga, A. W. & Thompson, P. M. Shape matching with medial curves and 1-d group-wise registration. In *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, 716–719 (2012).
48. Schumann, G. *et al.* The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Mol. Psychiatry* **15,** 1128–1139 (2010).
49. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20,** 110–121 (2010).
50. Cheung, I. *et al.* Developmental regulation and individual differences of neuronal H3K4me3 epigenomes in the prefrontal cortex. *Proc. Natl Acad. Sci. USA* **107,** 8824–8829 (2010).
51. Maunakea, A. K. *et al.* Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466,** 253–257 (2010).
52. Boyle, A. P. *et al.* High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132,** 311–322 (2008).
53. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473,** 43–49 (2011).
54. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55,** 997–1004 (1999).
55. Hager, R., Lu, L., Rosen, G. D. & Williams, R. W. Genetic architecture supports mosaic brain evolution and independent brain-body size regulation. *Nat. Commun.* **3,** 1079 (2012).
56. Schmucker, D. & Chen, B. Dscam and DSCAM: complex genes in simple animals, complex animals yet simple genes. *Genes Dev.* **23,** 147–156 (2009).
57. Brunet, A., Datta, S. R. & Greenberg, M. E. Transcription-dependent and -independent control of neuronal survival by the PI3K-Akt signaling pathway. *Curr. Opin. Neurobiol.* **11,** 297–305 (2001).



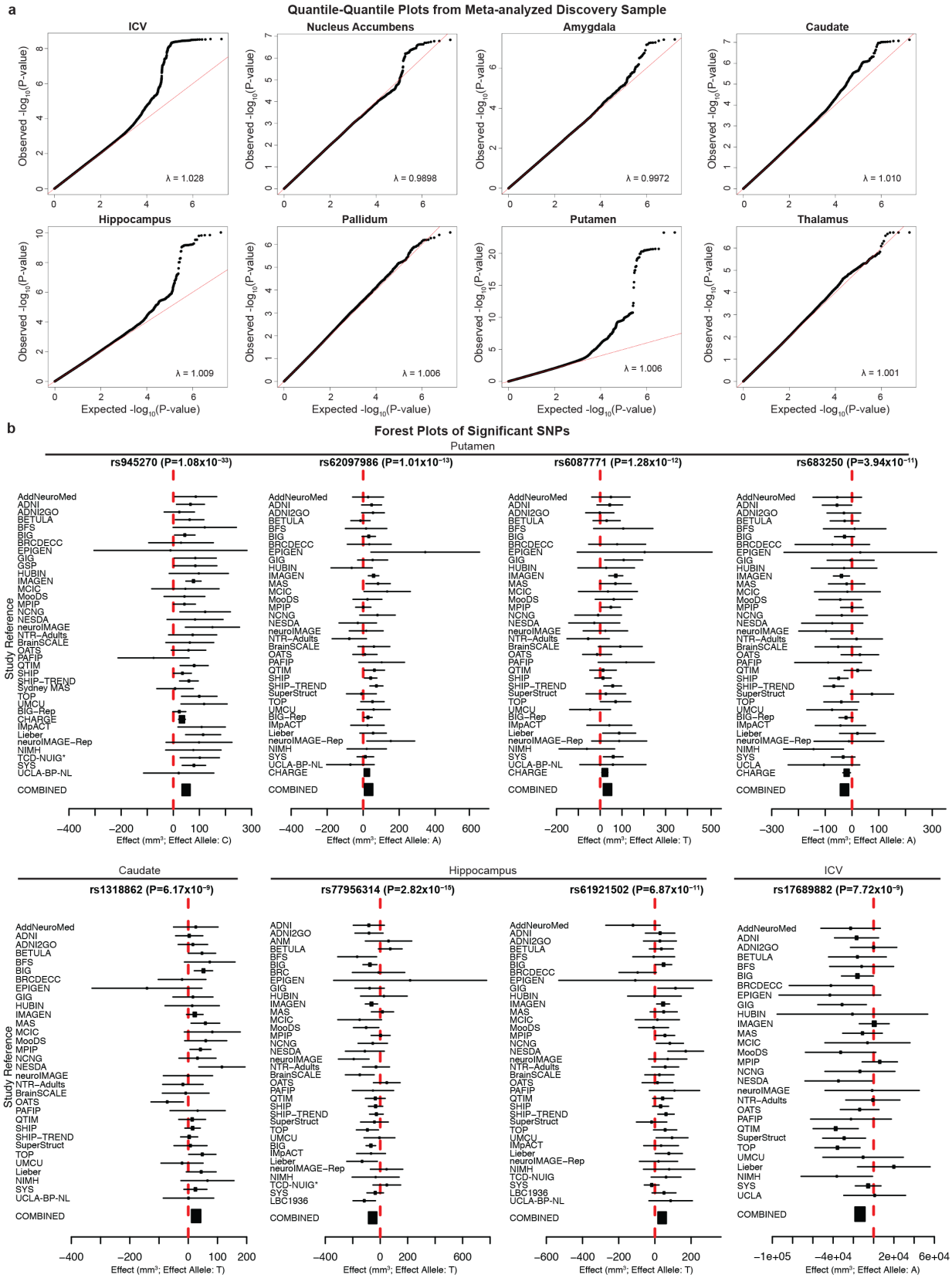
Extended Data Figure 1 | Outline of the genome-wide association meta-analysis. Structural T_1-weighted brain MRI and biological specimens for DNA extraction were acquired from each individual at each site. Imaging protocols were distributed to and completed by each site for standardized automated segmentation of brain structures and calculation of the ICV. Volumetric phenotypes were calculated from the segmentations. Genome-wide genotyping was completed at each site using commercially available chips. Standard imputation protocols to the 1000 Genomes reference panel (phase 1, version 3) were also distributed and completed at each site. Each site completed genome-wide association for each of the eight volumetric brain phenotypes with the listed covariates. Statistical results from GWAS files were uploaded to a central site for quality checking and fixed effects meta-analysis.



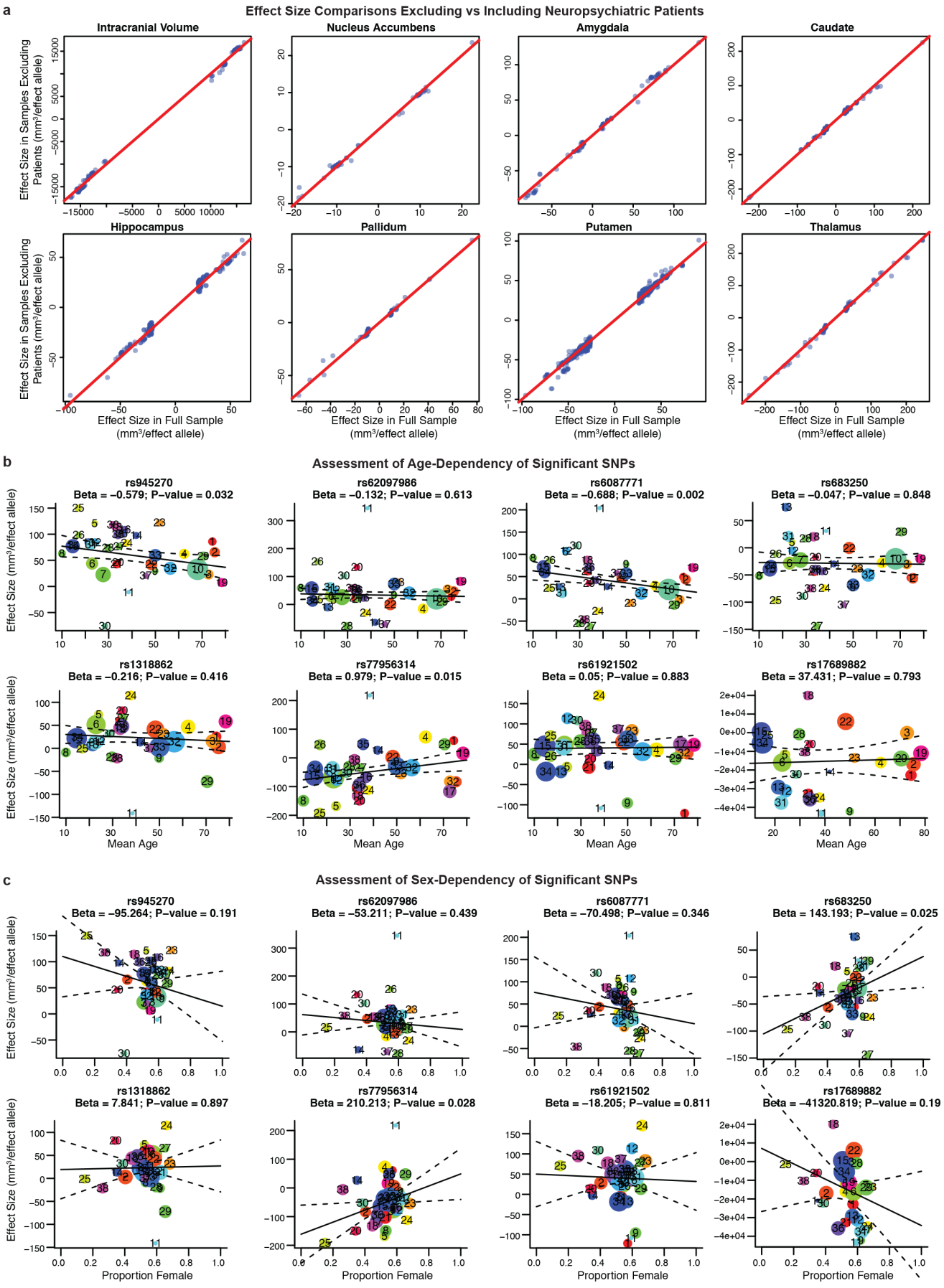
Extended Data Figure 2 | Ancestry inference via multi-dimensional scaling plots. Multi-dimensional scaling (MDS) plots of the discovery cohorts to HapMap III reference panels of known ancestry are displayed. Ancestry is generally homogeneous within each group. In all discovery samples any individuals with non-European ancestry were excluded before association. The axes have been flipped to the same orientation for each sample for ease of comparison. ASW, African ancestry in southwest USA; CEU, Utah residents with northern and western European ancestry from the CEPH collection; CHD, Chinese in metropolitan Denver, Colorado; GIH, Gujarati Indians in Houston, Texas; LWK, Luhya in Webuye, Kenya; MEX, Mexican ancestry in Los Angeles, California; MKK, Maasai in Kinyawa, Kenya; TSI, Tuscans in Italy; YRI, Yoruba in Ibadan, Nigeria.



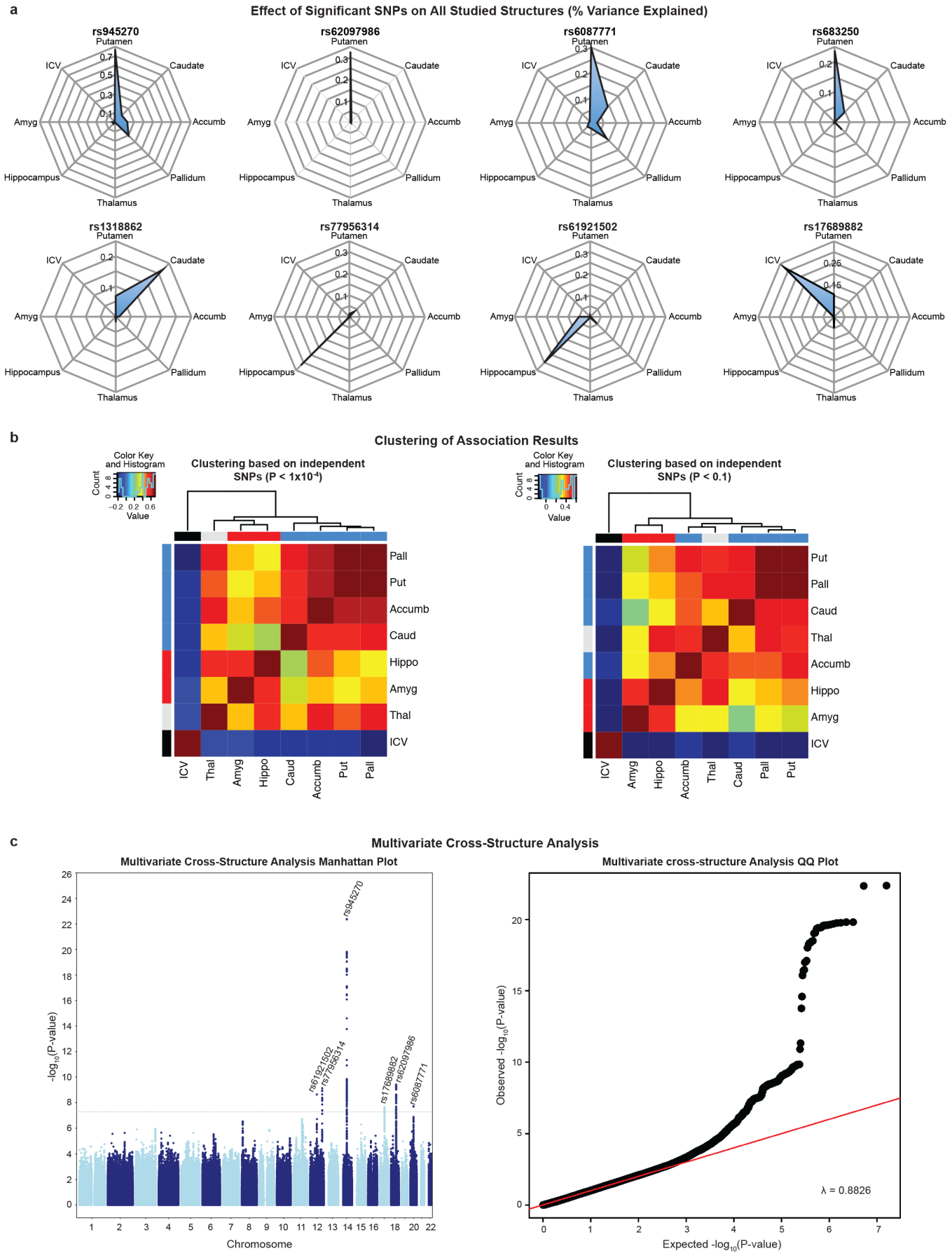
Extended Data Figure 3 | Genomic function is annotated near novel genome-wide significant loci. a–h, For each panel, zoomed-in Manhattan plots (±400 kb from top SNP) are shown with gene models below (GENCODE version 19). Plots below are zoomed to highlight the genomic region that probably contains the causal variant(s) ($r^2 > 0.8$ from the top SNP). Genomic annotations from the UCSC browser and ENCODE are displayed to indicate potential functionality (see Methods for detailed track information). SNP coverage is low in **f** owing to a common genetic inversion in the region. Each plot was made using the LocusTrack software (http://gump.qimr.edu.au/general/gabrieC/LocusTrack/).



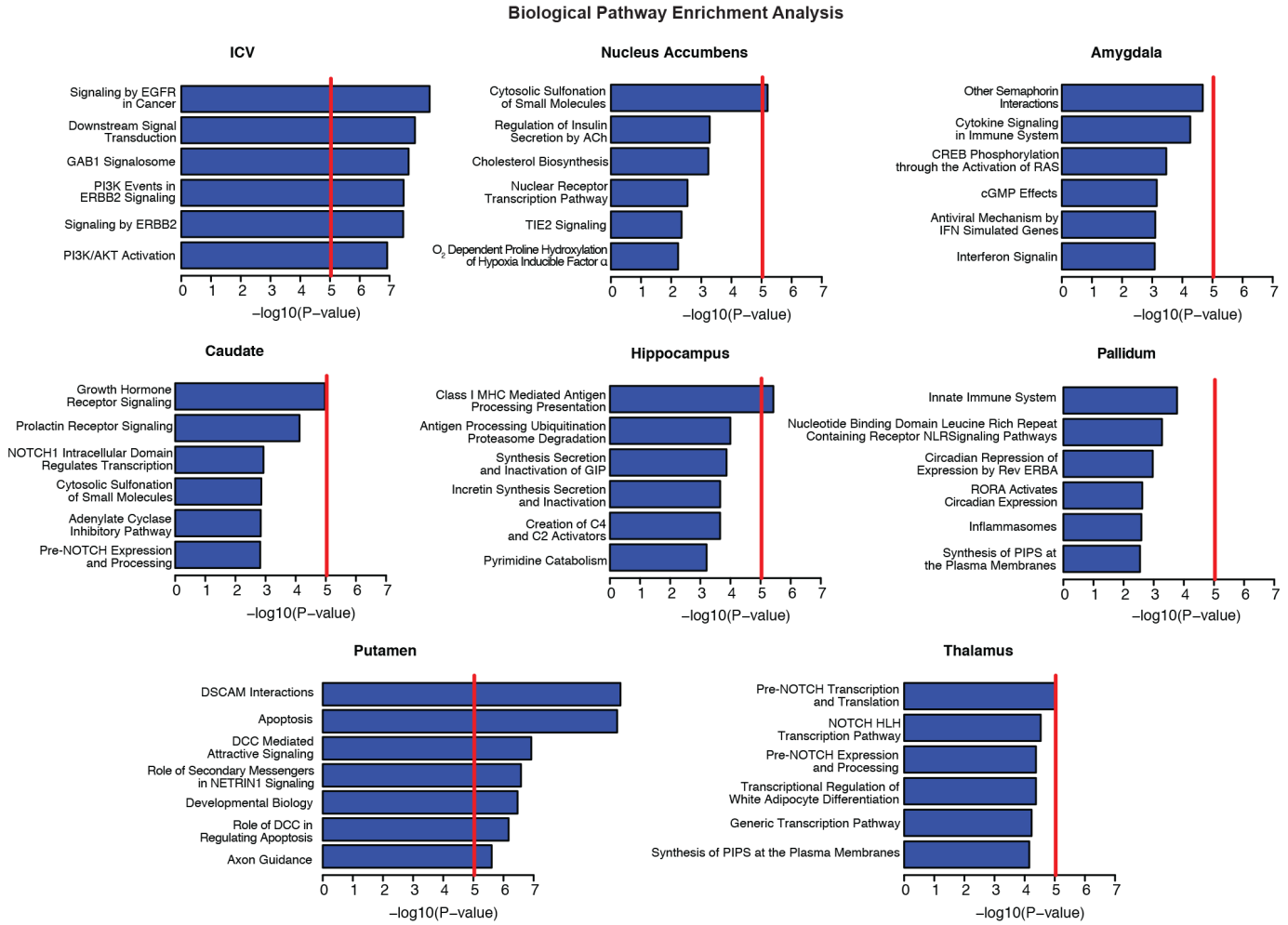
Extended Data Figure 4 | Quantile–quantile and forest plots from meta-analysis of discovery cohorts. **a**, Quantile–quantile plots show that the observed *P* values only deviate from the expected null distribution at the most significant values, indicating that population stratification or cryptic relatedness are not unduly inflating the results. This is quantified through the genomic control parameter (λ; which evaluates whether the median test statistic deviates from expected)[54]. λ values near 1 indicate that the median test statistic is similar to those derived from a null distribution. Corresponding meta-analysis Manhattan plots can be found in Fig. 1. **b**, Forest plots show the effect at each of the contributing sites to the meta-analysis. The size of the dot is proportional to the sample size, the effect is shown by the position on the *x* axis, and the standard error is shown by the line. Sites with an asterisk indicate the genotyping of a proxy SNP (in perfect linkage disequilibrium calculated from 1000 Genomes) for replication.



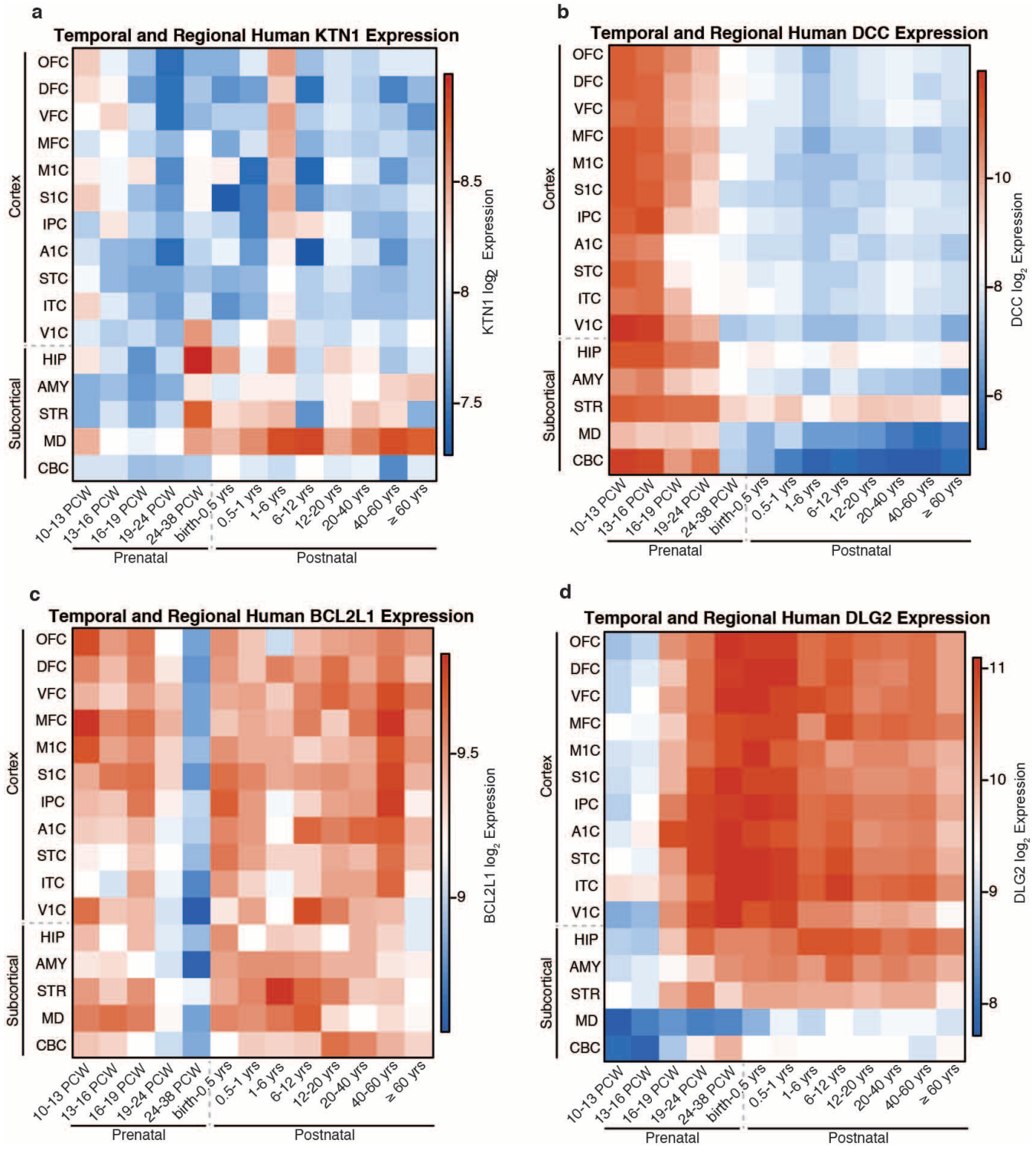
Extended Data Figure 5 | Influence of patients with neuropsychiatric disease, age and gender on association results. **a**, Scatterplot of effect sizes including and excluding patients with neuropsychiatric disorders for nominally significant SNPs. For each of the eight volumetric phenotypes, SNPs with $P < 1 \times 10^{-5}$ in the full discovery set meta-analysis were also evaluated excluding the patients. The beta values from regression, a measure of effect size, are plotted (blue dots) along with a line of equivalence between the two conditions (red line). The correlation between effect sizes with and without patients was very high ($r > 0.99$), showing that the SNPs with significant effects on brain structure are unlikely to be driven by the diseased individuals. **b**, Meta-regression comparison of effect size with mean age at each site. Each site has a corresponding number and coloured dot in each graph. The size of each dot is based on the standard error such that bigger sites with more definitive estimates have larger dots (and more influence on the meta-regression). The age range of participants covered most of the lifespan (9–97 years), but only one of these eight loci showed a significant relationship with the mean age of each cohort (rs608771 affecting putamen volume). **c**, Meta-regression comparison of effect size with the proportion of females at each site. No loci showed evidence of moderation by the proportion of females in a given sample. However, the proportion of females at each site has a very restricted range, so results should be interpreted with caution. Plotted information follows the same convention as described in **b**. The sites are numbered in the following order: (1) AddNeuroMed, (2) ADNI, (3) ADNI2GO, (4) BETULA, (5) BFS, (6) BIG, (7) BIG-Rep, (8) BrainSCALE, (9) BRCDECC, (10) CHARGE, (11) EPIGEN, (12) GIG, (13) GSP, (14) HUBIN, (15) IMAGEN, (16) IMpACT, (17) LBC1936, (18) Lieber, (19) MAS, (20) MCIC, (21) MooDS, (22) MPIP, (23) NCNG, (24) NESDA, (25) neuroIMAGE, (26) neuroIMAGE-Rep, (27) NIMH, (28) NTR-Adults, (29) OATS, (30) PAFIP, (31) QTIM, (32) SHIP, (33) SHIP-TREND, (34) SYS, (35) TCD-NUIG, (36) TOP, (37) UCLA-BP-NL and (38) UMCU.



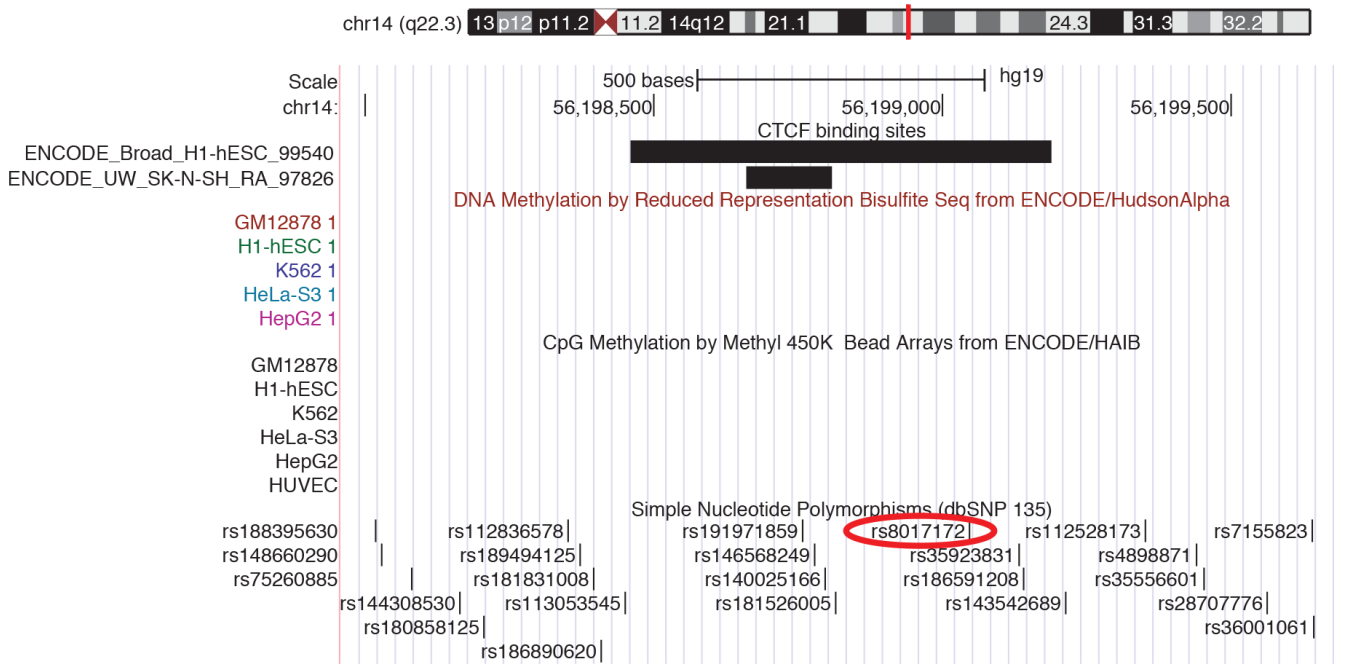
Extended Data Figure 6 | Cross-structure analyses. **a**, Radial plots of effect sizes from the discovery sample for all genome-wide significant SNPs identified in this study. Plots indicate the effect of each genetic variant, quantified as percentage variance explained, on the eight volumetric phenotypes studied. As expected, the SNPs identified with influence on a phenotype show the highest effect size for that phenotype: putamen volume (rs945270, rs62097986, rs608771 and rs683250), hippocampal volume (rs77956314 and rs61921502), caudate volume (rs1318862) and ICV (rs17689882). In general much smaller effects are observed on other structures. **b**, Correlation heat map of GWAS test statistics (*t*-values) and hierarchical clustering[55]. Independent SNPs were chosen within an linkage disequilibrium block based on the highest association in the multivariate cross-structure analysis described in Extended Data Fig. 6c. Two heat maps are shown taking only independent SNPs with either $P < 1 \times 10^{-4}$ (left) or $P < 0.01$ (right) in the multivariate cross-structure analysis. Different structures are labelled in developmentally similar regions by the colour bar on the top and side of the heat map including basal ganglia (putamen, pallidum, caudate and accumbens; blue), amygdalo–hippocampal complex (hippocampus and amygdala; red), thalamus (turquoise) and ICV (black). Hierarchical clustering showed that developmentally similar regions have mostly similar genetic influences across the entire genome. The low correlation with the ICV is owing to it being used as a covariate in the subcortical structure GWAS associations. **c**, A multivariate cross-structure analysis of all volumetric brain traits. A Manhattan plot (left) and corresponding quantile–quantile plot (right) of multivariate GWAS analysis of all traits (volumes of the accumbens, amygdala, caudate, hippocampus, pallidum, putamen, thalamus, and ICV) in the discovery data set using the TATES method[9] is shown. Multivariate cross-structure analysis confirmed the univariate analyses (see Table 1), but did not reveal any additional loci achieving cross-structure levels of significance.



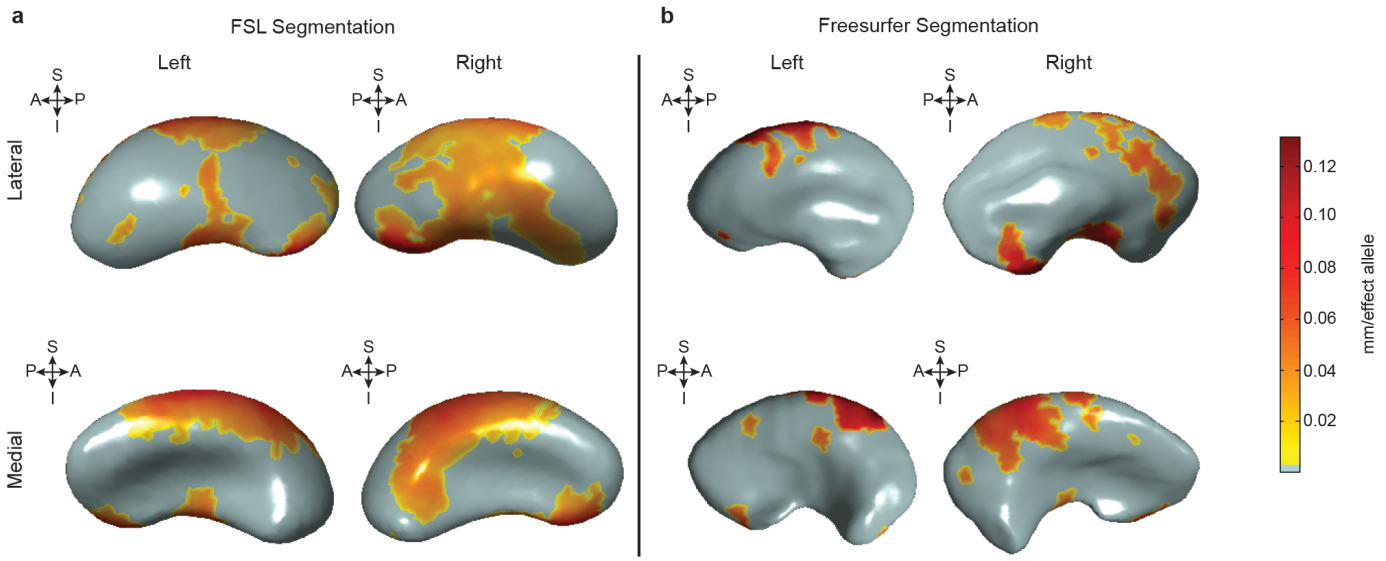
Extended Data Figure 7 | Pathway analysis of GWAS results for each brain structure. A pathway analysis was performed on each brain volume GWAS using KGG[42] to conduct gene-based tests and the Reactome database for pathway definition[43]. Pathway-wide significance was calculated using a Bonferroni correction threshold accounting for the number of pathways and traits tested such that $P_{thresh} = 0.05/(671 \text{ pathways} \times 7 \text{ independent traits}) = 1.06 \times 10^{-5}$ and is shown here as a red line. The number of independent traits was calculated by accounting for the non-independence of each of the eight traits examined (described in the Methods). Variants that influence the putamen were clustered near genes known to be involved in DSCAM interactions, neuronal arborization and axon guidance[56]. Variants that influence intracranial volume are clustered near genes involved in EGFR and phosphatidylinositol-3-OH kinase (PI(3)K)/AKT signalling pathways, known to be involved in neuronal survival[57]. All of these represent potential mechanisms by which genetic variants influence brain structure. It is important to note that the hybrid set-based test (HYST) method for pathway analysis used here can be strongly influenced by a few highly significant genes, as was the case for putamen hits in which *DCC* and *BCL2L1* were driving the pathway results.



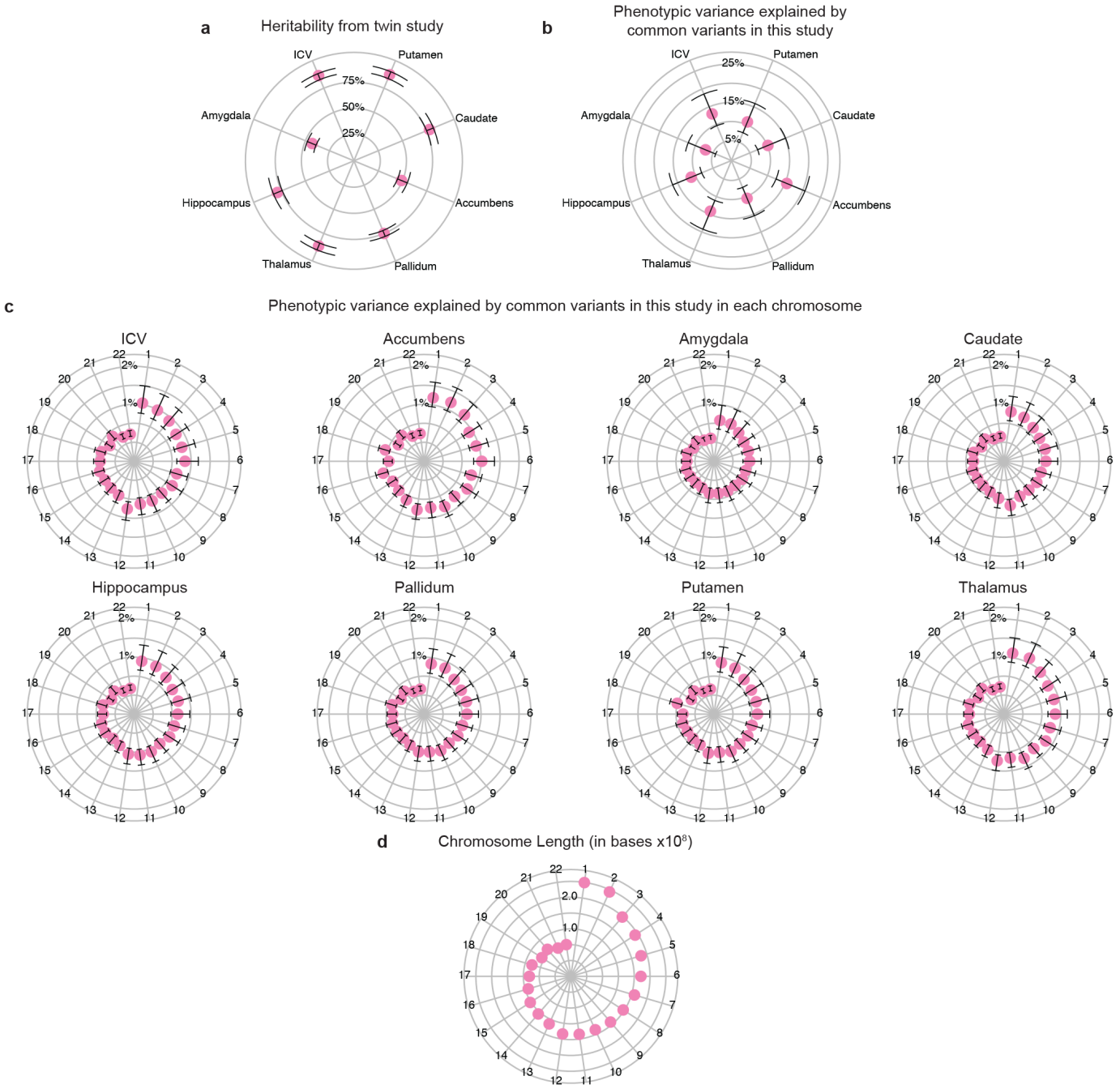
Extended Data Figure 8 | Spatio-temporal maps showing expression of genes near the four significant putamen loci over time and throughout regions of the brain. Spatio-temporal gene expression[13] was plotted as normalized $\log_2$ expression. Different areas of the neocortex (A1C, primary auditory cortex; DFC, dorsolateral prefrontal cortex; IPC, posterior inferior parietal cortex; ITC, inferior temporal cortex; MFC, medial prefrontal cortex; M1C, primary motor cortex; OFC, orbital prefrontal cortex; STC, superior temporal cortex; S1C, primary somatosensory cortex; VFC, ventrolateral prefrontal cortex; V1C, primary visual cortex) as well as subcortical areas (AMY, amygdala; CBC, cerebellar cortex; HIP, hippocampus; MD, mediodorsal nucleus of the thalamus; STR, striatum) are plotted from 10 post-conception weeks (PCW) to more than 60 years old. Genes that probably influence putamen volume are expressed in the striatum at some point during the lifespan. After late fetal development, *KTN1* is expressed in the human thalamus, striatum and hippocampus and is more highly expressed in the striatum than the cortex. Most genes seem to have strong gradients of expression across time, with *DCC* most highly expressed during early prenatal life, and *DLG2* most highly expressed at mid-fetal periods and throughout adulthood. *BCL2L1*, which inhibits programmed cell death, has decreased striatal expression at the end of neurogenesis (24–38 PCW), a period marked by increased apoptosis in the putamen[15].



Extended Data Figure 9 | CTCF-binding sites in the vicinity of the putamen locus marked by rs945270. CTCF-binding sites from the ENCODE project are displayed from the database CTCFBSDB 2.0 (ref. 23) from two different cell types: embryonic stem cells (track ENCODE_Broad_H1-hESC_99540) and a neuroblastoma cell line differentiated with retinoic acid (ENCODE_UW_SK-N-SH_RA_97826). A proxy SNP to the top hit within the locus, rs8017172 ($r^2 = 1.0$ to rs945270), lies within a CTCF-binding site called based on ChIP-seq data in the embryonic stem cells and near the binding site in neural SK-N-SH cells. As this is the lone chromatin mark in the intergenic region (see Extended Data Fig. 3), it suggests that the variant may disrupt a CTCF-binding site and thereby influence transcription of surrounding genes.



Extended Data Figure 10 | Shape analysis in 1,541 young healthy subjects shows consistent deformations of the putamen regardless of segmentation protocol. **a**, **b**, The distance from a medial core to surfaces derived from FSL FIRST (**a**; identical to Fig. 2c) or FreeSurfer (**b**) segmentations was derived in the same 1,541 subjects. Each copy of the rs945270-C allele was significantly associated with an increased width in coloured areas (false discovery rate corrected at $q = 0.05$) and the degree of deformation is labelled by colour. The orientation is indicated by arrows. A, anterior; I, inferior; P, posterior; S, superior. Shape analysis in both software suites gives statistically significant associations in the same direction. Although the effects are more widespread in the FSL segmentations, FreeSurfer segmentations also show overlapping regions of effect, which appears strongest in anterior and superior sections.



Extended Data Figure 11 | The phenotypic variance explained by all common variants in this study. **a**, Twin-based heritability (with 95% confidence intervals), measuring additive genetic influences from both common and rare variation, is shown for comparison with common variant based heritability (see Methods). **b**, The median estimated percentage of phenotypic variance explained by all SNPs (and 95% confidence interval) is given for each brain structure studied[41]. The full genome-wide association results from common variants explain approximately 7–15% of variance depending on the phenotype. **c**, The median estimated variance explained by each chromosome is shown for each phenotype. **d**, Some chromosomes explain more variance than would be expected by their length, for example chromosome 18 in the case of the putamen, which contains the *DCC* gene.