Genomic Studies Across the Lifespan Point to Early Mechanisms Determining Subcortical Volumes

Supplement

Supplementary methods

Study Population

Heritability analyses

Population-based cohort studies of unrelated individuals

The Internet-based Students HeAlth Research Enterprise (i-Share) study is an ongoing prospective population-based cohort of French-speaking students in higher education institutions (HEI) in France (www.i-share.fr), aiming at evaluating students' health and at exploring early mechanisms contributing to the occurrence of common diseases later in life (1). The bio-Share ancillary study is a biological platform comprising a collection of blood samples from a subset of the i-Share cohort. MRi-Share is a brain imaging ancillary study of i-Share, consisting of a brain MRI and a battery of cognitive tests. Participants were recruited simultaneously for bio-Share and MRi-Share between October 2015 and December 2017, among i-Share participants studying at the University of Bordeaux or other HEIs in Bordeaux and surroundings. Briefly, i-Share participants were eligible for these ancillary studies if they had completed the baseline self-administered online questionnaire, were registered with the national health insurance system, and had signed a written informed consent. To participate in MRi-Share, i-Share students had to be aged between 18 and 35 years and have no contraindication to brain MRI or pregnancy. Participants received an indemnity of 40 euros to compensate for the expenses induced by the participation in each of the two ancillary studies (2). The bio-Share and MRi-Share studies were approved by the regional Ethics Committee (Comité de Protection des Personnes Sud-Ouest et Outre-Mer). For the present study, the population comprised i-Share participants taking part in both MRi-Share and bio-Share and for whom brain MRI and genome-wide genotype data passed quality control filters. Of 1,954 participants who took part in MRi-Share, 1,856 had a usable (after quality control, QC) brain MRI including measures of subcortical brain structures. Genome-wide genotype data were available for 1,862 individuals. In total, 1,777 participants had both high quality brain MRI and genome-wide genotype data available (mean age: 22.1±2.3 years; 71.9% women).

The Three-City Dijon (3C-Dijon) Study is a community-based cohort study comprising 4,931 participants aged 65 years and older, non-institutionalized and randomly selected from the electoral rolls of the city of Dijon between 1999 and 2001 (3). Participants under 80 years of age and enrolled between June 1999 and September 2000 (n=2,763) were invited to undergo a brain MRI. Of these, 2,285 (82.7%) responded favorably, but due to financial restrictions, only

1,924 MRI scans were performed. Of 1,924 participants with an MRI at baseline, 1,623 had usable (after QC) volumetric measurements. After exclusion of participants with prevalent dementia (n=4), history of stroke (n=27), or brain tumors (n=6) at baseline, the remaining sample comprised 1,591 participants with subcortical volumes, of whom 1,440 with genome-wide genotype data (mean age: 72.6±4.0 years; 63.0% women).

Family-based cohort studies

The Framingham Heart Study is a single-site, community-based, prospective cohort study that was initiated in 1948 to investigate risk factors for cardiovascular disease. It comprises 3 generations of participants: the Original cohort, followed since 1948; their offspring and spouses of the offspring, followed since 1971 (Offspring cohort); and children from the largest offspring families enrolled in 2000 (Third-generation cohort). We excluded participants with prevalent stroke (n=110) or with other neurologic disorders that might confound the assessment of brain volumes (n=99) at the time of MRI examination. An additional n=306 participants did not have genotype data and we excluded those aged 35 years or younger (n=114). The present study includes 1,999 participants aged 36 to 64 years (mean age: 53.3±7.5 years; 50.3% women) and 1,828 aged 65 years and older (mean age: 75.8±7.2 years; 55.6% women) with brain MRI and genome-wide genotype data.

Analyses of shared genetic variation across the lifespan

To analyze genetic associations with subcortical volumes in young adults we used the aforementioned 1,777 i-Share participants with high quality brain MRI and genome-wide genotype data. To derive genome-wide significant associations with subcortical structures in middle-aged to older adults we used summary statistics of the largest published meta-analyses of subcortical volumes GWAS (N=38,851, mean age: 54 years, 52% women for all subcortical volumes except hippocampal volume: N=33,536, mean age: 55 years, 55% women) (4,5). Participants with prevalent dementia or stroke at the time of MRI, with presence of large brain infarcts or other neurological pathologies potentially influencing brain measurements were excluded for all subcortical volumes (4) except hippocampal volume (5). In secondary analyses we also used smaller meta-analyses without a subset of cohorts comprising younger age groups. For hippocampal volume we used a previously published GWAS meta-analysis (N=9,232, mean age: 67 years, 55% women) (6). For other subcortical volumes we generated a secondary meta-analysis without the contribution of cohorts participating in the ENIGMA consortium (N=15,981), the LIFE-Adult (N=1,718) and FHS-2 (N= 3,303) cohorts, leading to a sample size of 19,555 participants (mean age: 64 years, 56% women). Of note, only a small

subset of ENIGMA cohorts comprised younger adults, hence these exclusions were conservative.

MRI Acquisition and Phenotyping

For i-Share, MRI acquisitions were performed on a 3 Tesla Siemens Prisma scanner (Erlangen, Germany) at the Bordeaux bio-imaging platform. The MRI protocol lasted about 40 minutes and three types of acquisitions were performed: structural MRI, diffusion MRI and functional MRI. In the present study, we focused on structural MRI. The structural imaging acquisition included high resolution T1-weighted (3D MPRAGE, sagittal acquisition, TR/TE/TI = 2000/2.0/880 ms, repeat x2, 1 mm3 isotropic) and T2-FLAIR (3D SPACE, sagittal acquisition; TR/TE/TI = 5000/394/1800 ms; repeat x2; 1 mm3 isotropic voxel) sequences. Both acquired volumes were entered in the FreeSurfer package v6.0 (http://surfer.nmr.mgh.harvard.edu/) that provided subcortical structure volumes (2).

For 3C-Dijon, MRI acquisition was performed on a 1.5-Tesla Siemens Magnetom scanner. All participants were placed in the same position in the scanner following the orbitomeatal line as reference (7). High-resolution T1-weighted brain volume was acquired using a 3D inversion recovery fast spoiled-gradient echo sequence (3D SPGR; TR = 9.7 ms; TE = 4 ms; TI = 600 ms; coronal acquisition). The axially reoriented 3D volume matrix size was 256 x 192 x 256 mm^3, with a 1.0 x 0.98 x 0.98 mm^3 voxel size. T2- and PD-weighted brain volumes were acquired using the same 2D fast spin echo sequence with two echo times (TR = 4400 ms; TE1 = 16 ms; TE2 = 98 ms). T2 and PD acquisitions consisted of 35 axial slices of 3.5 mm thickness (0.5 mm between slices), having a 256 x 256 mm^2 matrix size and a 0.98 x 0.98 mm^2 in-plane resolution. The T1- and T2-weighted images of each subject were processed with SPM99 that provided brain tissue probability maps (8). Subcortical structure volumes were automatically computed by integrating the voxel intensities of the SPM (https://www.fil.ion.ucl.ac.uk/spm/) grey matter segmented volume in the Automated Anatomical Labeling (AAL) atlas subcortical regions of interest (9).

The Framingham Heart Study brain MRI protocol has been described in detail previously (10,11). Briefly, brain MRI was performed using a 1-, 1.5-, or 3-Tesla Siemens Magnetom scanner, which included a 3-dimensional T1-weighted coronal spoiled gradient-recalled echo sequence. Total intracranial volume was derived from T1 images after removal of non-brain tissues by an atlas-based method, followed by manual editing as needed (12). Subcortical volumes were generated from T1-weighted images using FreeSurfer version 6 (13,14).

MRI acquisition parameters and phenotyping methods in other cohorts participating in the published meta-analyses have been described in detail (4–6).

For data from Satizabal et al (4), each study investigated the volumes of seven subcortical structures: the nucleus accumbens, amygdala, brainstem, caudate nucleus, globus pallidus, putamen and thalamus. These phenotypes were defined as the mean volume (cm^3) of the left and right hemispheres. Each study contributed MRI data obtained using diverse scanners, field strengths and acquisition protocols. The estimations of volumes for the seven subcortical brain structures and total intracranial volume were generated following freely available and inhouse segmentation methods that were previously described and validated (4).

For data from Hibar et al (5), hippocampal volumes were estimated using the automated and previously validated segmentation algorithms, FSL FIRST47 from the FMRIB Software Library (FSL) and FreeSurfer. Hippocampal segmentations were visually examined at each site, and poorly segmented scans were excluded. Sites also generated histogram plots to identify any volume outliers. Individuals with a volume more than three standard deviations away from the mean were visually inspected to verify proper segmentation. Statistical outliers were included in analysis if they were properly segmented; otherwise, they were removed. Average bilateral hippocampal volume was highly correlated across automated procedures used to measure it (5).

For data from Bis et al (6), each study evaluated the total hippocampal volume using 1T, 1.5T or 3T MRI and either operator-defined, manually traced boundaries drawn on serial coronal sections or automated methods according to previously described reading protocols (6).

Genotyping, quality control, and imputation in the i-Share study

Genome-wide genotyping of 1,872 i-Share participants was performed using the Affymetrix Precision Medicine Axiom Array at McGill Genome Center (Canada). After quality control, genotype data were available for 1,862 participants (7 participants were removed due to sex discrepancies, 2 participants who appeared to be duplicates but not twins, and one participant with a kinship coefficient >0.0625 (third degree related) with more than 20 other participants, suggesting a possible sample contamination (KING software) (15). Using Plink 1.9 we performed a principal component (PC) analysis of population stratification, including in addition to i-Share, 3C-Dijon participants and the European sample from the 1000 genomes reference panel, both representing European populations: we used the two first PCs to flag outliers lying beyond the mean ± 4 standard deviations of the corresponding PC distribution (N=156) (16). Using the KING software (15), we estimated kinship coefficients for each pair

of individuals and used two kinship thresholds to flag individuals as related: kinship >0.10 for association analyses and kinship >0.0625 (3rd degree related) for analyses requiring more stringent exclusion criteria, such as heritability analyses. For each threshold, we first identified multi-related samples (individuals with more than 1 related participant), flagging and removing (only for the rest of relatedness analyses) iteratively the multi-related sample with the highest missingness rate until no multi-related sample remained. Then, for each pair of related individuals, we flagged the one with the highest missingness (kinship>0.10: N=48; kinship>0.0625: N=66). Flagged participants for relatedness and population stratification were not systematically removed because some statistical methods, such as the linear mixed models, allow to account for potential biases due to the population structure. After applying standard quality control procedures (SNP call rate <98%, Hardy-Weinberg Equilibrium <0.001), we imputed the genotypes on the Haplotype Reference Consortium (HRC) reference panel.

For 3C-Dijon participants, genome-wide genotyping was performed at the Centre National de Génotypage in Evry (France) using the Illumina Human610Iquad BeadChips. FHS used the Affymetrix 500K mapping array plus Affymetrix 50K supplemental array. Genotypes in 3C-Dijon were also imputed to HRC and FHS used the 1000 Genomes imputation panel.

Only Single Nucleotide Polymorphisms (SNPs) with minor allele frequencies (MAF) $\geq$0.01 and an imputation score>0.5 were retained for the present analyses.

Genome-wide genotyping, quality control, and imputation in other cohorts participating in the published meta-analyses have been described in detail previously (4–6). The published GWAS predominantly used 1000 Genomes (phase 1 version 3) European sample (4,5), except for studies using data from the UK Biobank imputed on the HRC reference panel (4) and Bis et al which used HapMap CEU population (6).

Statistical analyses

Heritability analyses

Population-based cohort studies of unrelated individuals

To estimate heritability for each subcortical volume in i-Share and 3C-Dijon, we used GCTA (v1.26.0) to estimate the proportion of variance explained by genome-wide SNPs and did not constrain estimations (17). Only SNPs with a MAF>0.01 and an imputation score>0.9 were considered. In i-Share, we removed participants flagged for non-European ancestry as well as those with genetic relatedness>0.05, using a genetic relationship matrix (GRM) as implemented in GCTA (n=1,528) (17). In 3C, we removed related participants with GRM>0.05

(n=1,396). These analyses were adjusted for age, sex, total intracranial volume, and the first four principal components of population stratification.

Single variant analyses and Genetic Risk Score approaches

We generated genetic risk scores (GRS) for subcortical volumes in young adults (i-Share) by summing the number of independent risk alleles identified as genome-wide significant ($p<5.0\times10^{-8}$) in published GWAS meta-analyses in middle-aged to older adults, weighting each risk allele by the regression coefficient for the corresponding SNP in the published GWAS. SNPs were clumped using Plink 1.9 (LD-r^2>0.10 and distance<250kb). We removed participants flagged for non-European ancestry and relatedness (kinship coefficient>0.10) (n=1,586). All associations were tested using linear regression models in R v3.6.1 and adjusted for age, sex, total intracranial volume, and the first four principal components of population stratification. To account for multiple testing, we corrected for four independent phenotypes ($p<1.25\times10^{-2}$). As a sensitivity analysis, GRS analyses were repeated using summary statistics of the subcortical volumes GWAS conducted after removing cohorts with young participants as described above.

Transcriptome-wide association study

To explore genes underlying genetic associations with subcortical volumes across the lifespan we performed transcriptome-wide association studies (TWAS) using TWAS-Fusion (18).

First, we used summary statistics from the aforementioned published GWAS meta-analyses of subcortical volumes (4,5) and 17 publicly available gene expression reference panels (expression quantitative trait loci [eQTL] reference panels) from blood (Netherlands Twin Registry, NTR; Young Finns Study, YFS) (18,19), brain (GTEx, CommonMind Consortium, CMC) (20,21) and peripheral nerve tissues (GTEx) (20). We downloaded precomputed SNP-expression weights from the TWAS-Fusion website for each gene in the reference panel (http://gusevlab.org/projects/fusion/). TWAS-Fusion was then used to estimate the TWAS association statistics between predicted expression and each subcortical volume by integrating information from expression reference panels (SNP-expression weights), GWAS summary statistics (SNP-subcortical volumes effect estimates), and LD reference panels (SNP correlation matrix) (18). Transcriptome-wide significant genes were determined using a Bonferroni correction in each tissue expression panel, based on the average number of features (3793.5 genes) and 4 independent phenotypes ($p<3.30\times10^{-6}$ [0.05/(4*3793.5)]) tested across all tissues. Transcriptome-wide significant genes were then tested in conditional analysis as implemented in the Fusion software (18). Conditionally significant genes were subsequently

tested in a colocalization analysis to estimate the posterior probability of a shared causal variant between the gene expression and trait association (PP4), using a prior probability of $p<3.30 \times 10^{-6}$ for the subcortical volumes association (22). Genes presenting a PP4≥0.75, for which eQTLs did not reach genome-wide significance in association with subcortical volumes, and were not in LD ($r^2<0.01$) with any of the lead SNPs of genome-wide significant risk loci for subcortical volumes, were considered as novel.

Next, in order to test colocalized associations in a younger population, we used the individual-level prediction of the gene expression option implemented in the Fusion software to generate expression-weights in young adults (i-Share data). These individual-level predictors were used in Plink 1.9 to generate gene expression scores for each gene-expression (16). These scores were then tested for each corresponding subcortical volume in young adults using linear regression models in R v3.6.1 (lm function), adjusting for age, sex, total intracranial volume, and the first four principal components of population stratification, and removing participants flagged for non-European ancestry and relatedness (kinship coefficient>0.10) (n=1,586). The significance threshold accounted for the number of genes colocalized in ≥1 tissue for each phenotype (accumbens: $p<5.00 \times 10^{-2}$ (0.05/1); amygdala: $p<5.00 \times 10^{-2}$ (0.05/1); caudate: $p<2.17 \times 10^{-3}$ (0.05/23); hippocampus: $p<1.67 \times 10^{-2}$ (0.05/3); pallidum: $p<2.94 \times 10^{-3}$ (0.05/17); putamen: $p<3.85 \times 10^{-3}$ (0.05/13); thalamus: $p<5.00 \times 10^{-2}$ (0.05/1)).

Clinical correlates

We tested whether genetically predicted Alzheimer's disease (AD) (23) or Parkinson's disease (PD) (24) have an impact on subcortical volumes in the general population in young, middle-aged and older adults. We used the generalised summary data-based Mendelian randomization (GSMR) tool implemented in GCTA, with the summary statistics of the latest, largest published GWAS meta-analyses of AD (23) and PD (24) and of the same subcortical volumes GWAS described above (17,25). We selected independent ($r^2<0.05$) genome-wide significant SNPs ($p<5\times10^{-8}$) as instruments for each exposure. SNPs that had pleiotropic effects on both exposure and outcome were removed using the HEIDI-outlier method (pHEIDI<0.01) implemented in GSMR. The threshold for significance accounted for four independent subcortical volumes and two diseases ($p<6.25\times10^{-3}$).

Supplementary references

1. Montagni I, Guichard E, Kurth T (2016): Association of screen time with self-perceived attention problems and hyperactivity levels in French students: a cross-sectional study. *BMJ Open* 6: e009089.

2. Tsuchida A, Laurent A, Crivello F, Petit L, Joliot M, Pepe A, *et al.* (2020): *The MRi-Share Database: Brain Imaging in a Cross-Sectional Cohort of 1,870 University Students*. bioRxiv. [Preprint]. https://doi.org/10.1101/2020.06.17.154666

3. 3C Study Group (2003): Vascular Factors and Risk of Dementia: Design of the Three-City Study and Baseline Characteristics of the Study Population. *Neuroepidemiology* 22: 316–325.

4. Satizabal CL, Adams HHH, Hibar DP, White CC, Knol MJ, Stein JL, *et al.* (2019): Genetic architecture of subcortical brain structures in 38,851 individuals. *Nat Genet* 51: 1624–1636.

5. Hibar DP, Adams HHH, Jahanshad N, Chauhan G, Stein JL, Hofer E, *et al.* (2017): Novel genetic loci associated with hippocampal volume. *Nat Commun* 8: 13624.

6. Bis JC, DeCarli C, Smith AV, van der Lijn F, Crivello F, Fornage M, *et al.* (2012): Common variants at 12q14 and 12q24 are associated with hippocampal volume. *Nat Genet* 44: 545–551.

7. Maillard P, Delcroix N, Crivello F, Dufouil C, Gicquel S, Joliot M, *et al.* (2008): An automated procedure for the assessment of white matter hyperintensities by multispectral (T1, T2, PD) MRI and an evaluation of its between-centre reproducibility based on two large community databases. *Neuroradiology* 50: 31–42.

8. Ashburner J, Friston KJ (2000): Voxel-based morphometry--the methods. *NeuroImage* 11: 805–821.

9. Godin O, Tzourio C, Rouaud O, Zhu Y, Maillard P, Pasquier F, *et al.* (2010): Joint Effect of White Matter Lesions and Hippocampal Volumes on Severity of Cognitive Decline: The 3C-Dijon MRI Study. *J Alzheimers Dis* 20: 453–463.

10. Debette S, Beiser A, Hoffmann U, DeCarli C, O'Donnell CJ, Massaro JM, *et al.* (2010): Visceral fat is associated with lower brain volume in healthy middle-aged adults. *Ann Neurol* 68: 136–144.

11. DeCarli C, Massaro J, Harvey D, Hald J, Tullberg M, Au R, *et al.* (2005): Measures of brain morphology and infarction in the framingham heart study: establishing what is normal. *Neurobiol Aging* 26: 491–510.

12. Aljabar P, Heckemann RA, Hammers A, Hajnal JV, Rueckert D (2009): Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *NeuroImage* 46: 726–738.

13. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, *et al.* (2002): Whole Brain Segmentation. *Neuron* 33: 341–355.

14. Fischl B, Salat DH, van der Kouwe AJW, Makris N, Ségonne F, Quinn BT, Dale AM (2004): Sequence-independent segmentation of magnetic resonance images. *NeuroImage* 23: S69–S84.

15. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M (2010): Robust relationship inference in genome-wide association studies. *Bioinformatics* 26: 2867–2873.

16. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, *et al.* (2007): PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* 81: 559–575.

17. Yang J, Lee SH, Goddard ME, Visscher PM (2011): GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am J Hum Genet* 88: 76–82.

18. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, *et al.* (2016): Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* 48: 245–252.

19. Wright FA, Sullivan PF, Brooks AI, Zou F, Sun W, Xia K, *et al.* (2014): Heritability and genomics of gene expression in peripheral blood. *Nat Genet* 46: 430–437.

20. GTEx Consortium (2017): Genetic effects on gene expression across human tissues. *Nature* 550: 204–213.

21. Fromer M, Roussos P, Sieberts SK, Johnson JS, Kavanagh DH, Perumal TM, *et al.* (2016): Gene Expression Elucidates Functional Impact of Polygenic Risk for Schizophrenia. *Nat Neurosci* 19: 1442–1453.

22. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V (2014): Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet* 10: e1004383.

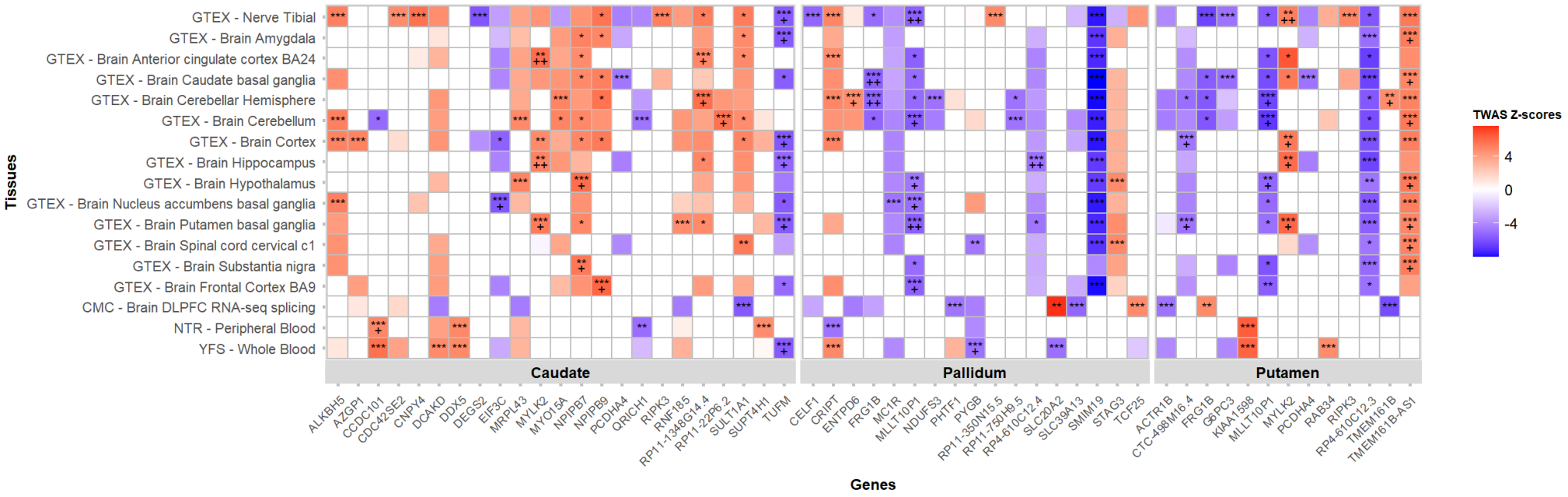
23. Schwartzentruber J, Cooper S, Liu JZ, Barrio-Hernandez I, Bello E, Kumasaka N, *et al.* (2021): Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. *Nat Genet* 53: 392–402.

24. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, *et al.* (2019): Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol* 18: 1091–1102.

25. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, *et al.* (2018): Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun* 9: 224.

Supplementary tables and figures

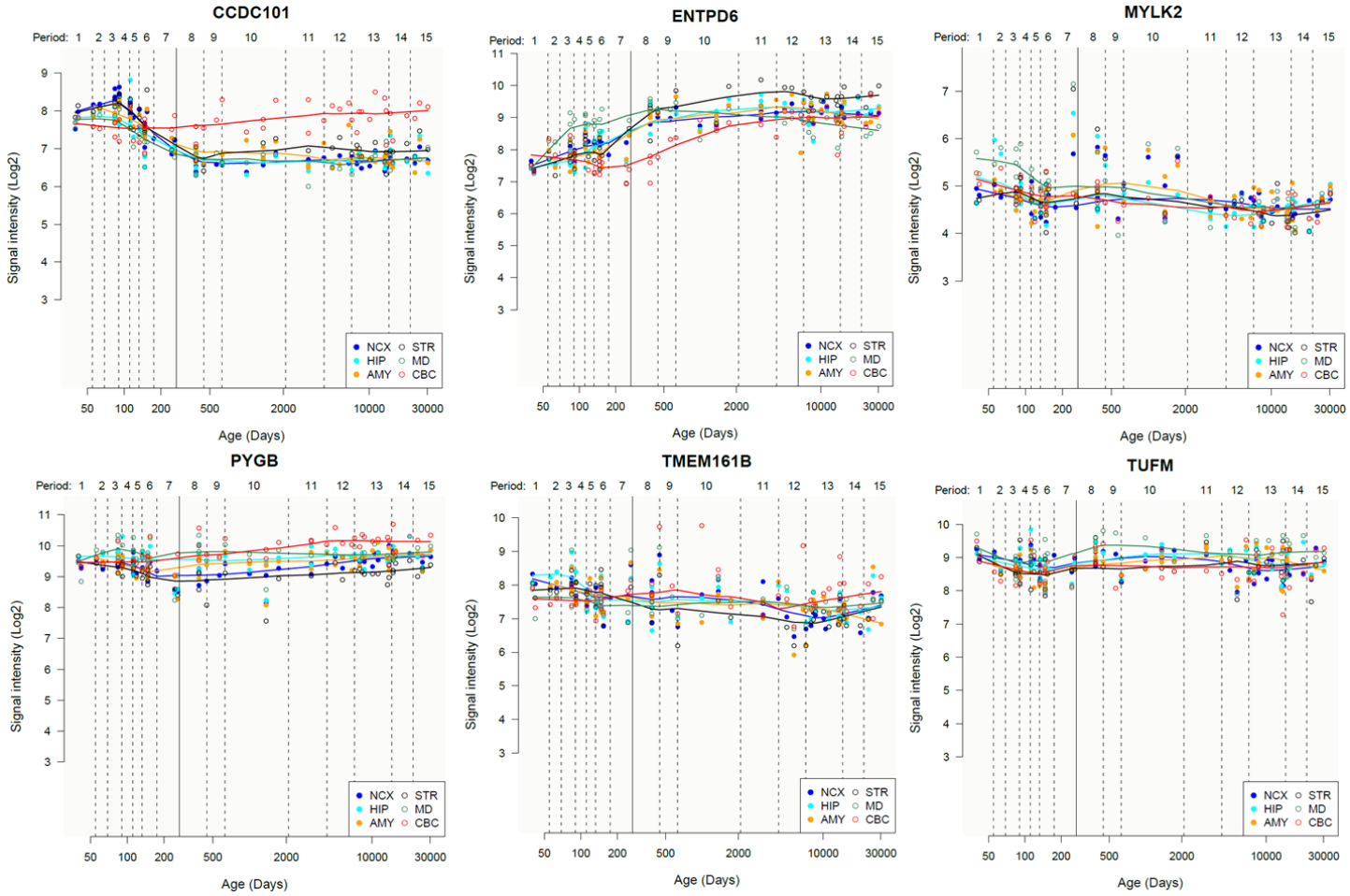
Figure S1: Heatmaps of the transcriptome-wide association studies of the caudate nucleus, putamen and pallidum reaching transcriptome wide significance and colocalized in older persons (Satizabal et al, Nat Genet 2019 and Hibar et al, Nat Commun 2017)



Legend: * : TWAS Significant ($p<3.30\times10^{-6}$) ** : Conditionally significant ($p<0.05$) *** : COLOC PP4 > 0.75

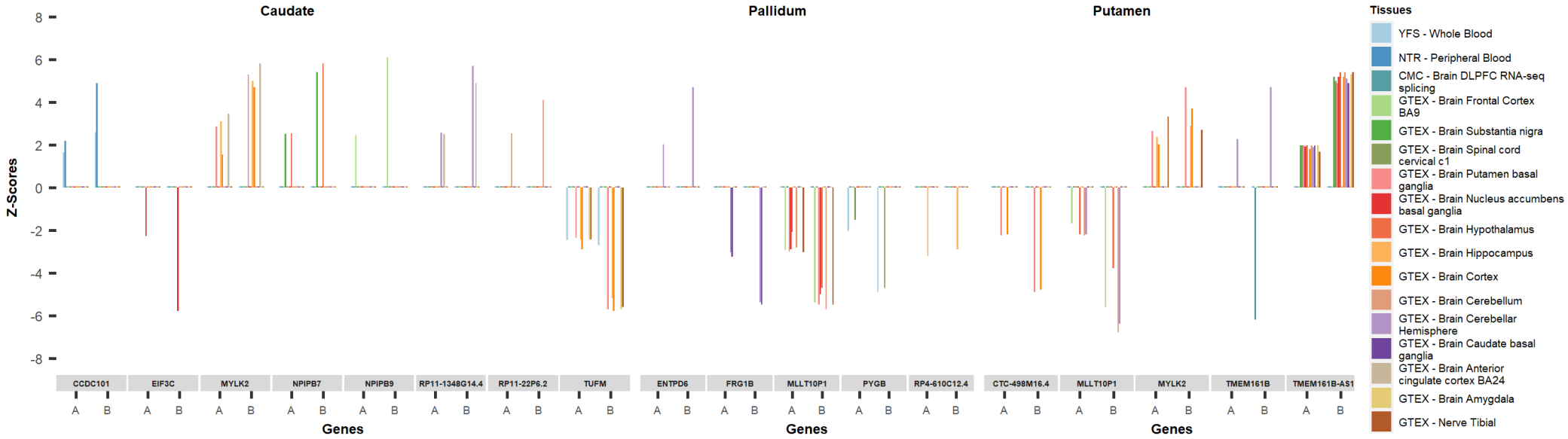
\+ : Nominally significant in i-Share ++ : Significant in i-Share (after multiple-testing correction)

Figure S2: Lifetime brain gene expression profile of genes reaching transcriptome wide significance colocalized in older persons and at least nominal significance in young adults with a correspondence in the Human Brain Transcriptome database



The spatio-temporal gene expression level is plotted as log2-transformed exon array signal intensity (y-axis) against the post conception days (x-axis) as provided by the Human Brain Transcriptom project database. Periods of human development and adulthood are indicated by vertical dashed lines: 4-8 post conception weeks [PCW] (period 1), 8-10 PCW (period 2), 10-13 PCW (period 3), 13-16 PCW (period 4), 16-19 PCW (period 5), 19-24 PCW (period 6), 24-38 PCW (period 7), birth- 6 postnatal months (period 8), 6-12 postnatal months (period 9), 1- 6 years (period 10), 6-12 years (period 11), 12-20 years (period 12), 20-40 years (period 13), 40-60 years (period 14), and 60 years+ (period 15). The boundary between pre- and postnatal periods is indicated by the solid vertical line. Each colored point represents the expression level of each gene across 16 anatomical brain regions and ages. Brain structure includes 11 neocortical areas (NCX, blue), and 5 subcortical regions: hippocampus (HIP, cyan), amygdala (AMY, orange), striatum (STR, black), mediodorsal nucleus of thalamus (MD, dark green), and cerebellar cortex (CBC, red).

Figure S3: Barplots of the transcriptome-wide association studies of the caudate nucleus, putamen and pallidum reaching transcriptome wide significance and colocalized in older persons (Satizabal et al, Nat Genet 2019 and Hibar et al, Nat Commun 2017) and at least nominal significance in young adults (i-Share cohort)



Legend: A: Young adults; B: Older adults

Table S1: Heritability of subcortical volumes in young adults, middle-aged adults and in older adults in both unrelated and family-based population.

	Unrelated population						Family-based population					
	Young adults (18-35y) i-Share Cohort* *(N=1,528)*			Older adults (65+) 3C-Dijon Cohort* *(N=1,396)*			Middle-aged adults (36-64y) Framingham Heart Study** *(N=1,999)*			Older adults (65+) Framingham Heart Study** *(N=1,828)*		
Subcortical volume	h^2	SE	p	h^2	SE	p	h^2	SE	p	h^2	SE	p
Accumbens	0.02	0.21	4.63E-01				0.51	0.07	5.71E-14	0.50	0.09	1.62E-08
Amygdala	0.61	0.21	2.45E-03	0.00	0.22	3.85E-01	0.75	0.07	4.11E-29	0.23	0.09	3.07E-03
Caudate Nucleus	0.78	0.20	3.71E-05	0.06	0.22	3.93E-01	0.81	0.06	3.11E-40	0.68	0.08	4.18E-17
Hippocampus	0.94	0.20	2.38E-06	0.22	0.23	1.55E-01	0.76	0.06	1.14E-31	0.49	0.08	5.60E-09
Pallidum	0.85	0.20	1.20E-05	0.36	0.23	5.31E-02	0.61	0.06	1.77E-22	0.62	0.08	2.70E-14
Putamen	0.63	0.21	1.38E-03	0.31	0.22	6.62E-02	0.78	0.06	1.47E-35	0.63	0.07	2.72E-16
Thalamus	0.23	0.20	1.25E-01	0.28	0.22	8.65E-02	0.60	0.06	4.54E-22	0.57	0.09	4.89E-09

* SNP-heritability estimates obtained with GCTA GREML

** Family-based heritability estimates obtained with SOLAR

Table S2: Association of genome-wide significant variants (individually and aggregated in genetic risk scores) for subcortical volumes in older adults with the same volumes in young adults

SNP*	Chr	Position	A1	Freq	Nearest genes	Lead SNP from GWAS	Single variant analysis			TWAS analyses
							i-Share Cohort†			Gene colocalized in the locus‡
							Beta	SE	p	
Accumbens										
rs9827516	3q28	190,619,643	T	0.08	SNAR-I,OSTN	rs9818981	-0.01	0.01	1.93E-01	
rs13105581	4q24	103,228,830	T	0.09	SLC39A8	rs13107325	0.02	0.01	5.07E-02	
rs72757294	5q12.3	65,827,816	A	0.22	SREK1,MAST4	rs11747514	0.01	0.01	3.58E-01	
rs9671291	14q22.3	56,181,695	G	0.44	KTN1,RPL13AP3	rs868202	-0.01	0.01	4.29E-02	
GRS from full GWAS**							-0.04	0.04	3.56E-01	
GRS from GWAS old only§							-0.05	0.08	4.69E-01	
Amygdala										
rs11111293	12q23.2	102,921,296	C	0.18	IGF1,LINC00485	rs11111293	-0.04	0.01	**4.50E-03**	
GRS from full GWAS**							-0.61	0.14	**9.47E-06**	
GRS from GWAS old only§							-0.23	0.15	1.29E-01	
Caudate nucleus										
rs10909901	1p36.32	3,131,235	T	0.28	PRDM16	rs2817145	0.13	0.03	**8.03E-06**	
rs35305377	7q22.1	99,938,955	G	0.46	PMS2P1,PILRB	rs35305377	0.03	0.02	1.85E-01	AZGP1, CNPY4
rs7040561	9q33.3	128,528,978	T	0.13	PBX3	rs7040561	0.03	0.04	4.27E-01	
rs10830894	11q14.3	92,018,778	T	0.41	MIR4490,FAT3	rs3133370	-0.08	0.02	**5.98E-04**	
rs1953353	14q22.3	56,189,751	A	0.32	KTN1,RPL13AP3	rs148470213	-0.10	0.03	**1.92E-04**	
rs3783330	14q32.2	100,572,954	G	0.33	EVL	rs55989340	0.04	0.02	7.33E-02	DEGS2
rs4115668	16p11.2	28,607,532	A	0.27	SULT1A2	rs1987471	-0.10	0.03	**1.45E-04**	**CCDC101**, **NPIPB7**, **NPIPB9**, SULT1A1, **TUFM, RP11-1348G14.4, RP11-22P6.2**
rs4888921	16q22.3	73,899,994	G	0.48	LOC100506172,PSMD7	rs4888010	0.01	0.02	6.78E-01	
rs12445022	16q24.2	87,575,332	A	0.35	ZCCHC14,JPH3	rs12445022	0.01	0.02	6.52E-01	
rs1062794	20q11.21	30,381,758	C	0.34	TPX2	rs6060983	-0.08	0.02	1.59E-03	**MYLK2 (++)**
GRS from full GWAS**							-0.73	0.14	**9.76E-08**	
GRS from GWAS old only§							-0.99	0.21	**2.31E-06**	
Hippocampus										
rs12474587	2q24.2	162,802,993	T	0.40	SLC4A10	rs2268894	0.03	0.02	1.70E-01	

SNP*	Chr	Position	A1	Freq	Nearest genes	Lead SNP from GWAS	Single variant analysis			TWAS analyses
							i-Share Cohort†			Gene colocalized in the locus‡
							Beta	SE	p	
rs34257018	5q12.3	66,080,929	T	0.37	MAST4	rs2289881	-0.03	0.02	1.10E-01	
rs6962499	7q36.3	155,808,233	G	0.23	SHH,LOC285889	rs11979341	0.04	0.02	7.88E-02	
rs2416560	9q33.1	119,252,207	C	0.35	ASTN2	rs7020341	0.02	0.02	3.65E-01	
rs17178006	12q14.3	65,718,299	G	0.09	MSRB3	rs61921502	-0.15	0.03	**1.72E-06**	
rs113205216	12q24.22	117,326,943	A	0.09	HRK,FBXW8	rs77956314	0.12	0.03	**1.41E-04**	
GRS from full GWAS**							-0.46	0.09	**2.08E-07**	
GRS from GWAS old only§							-1.12	0.21	**7.74E-08**	
Pallidum										
rs4654960	1p36.12	21,863,495	G	0.44	ALPL	rs12567402	0.02	0.01	1.37E-01	
rs176415	2p22.3	32,649,778	A	0.35	BIRC6	rs4952211	-0.02	0.01	7.14E-02	
rs196814	8p21.2	24,716,594	G	0.22	ADAM7,NEFM	rs196807	0.03	0.01	2.39E-02	
rs6474403	8p11.21	42,410,936	G	0.40	SMIM19,CHRNB3	rs2923447	0.01	0.01	5.16E-01	SLC20A2, SMIM19
rs945270	14q22.3	56,200,473	G	0.44	KTN1,RPL13AP3	rs10129414	-0.04	0.01	**2.29E-04**	
rs113818546	20q11.21	30,369,090	T	0.24	TPX2	rs10439607	-0.03	0.01	8.46E-03	**ENTPD6**, **FRG1B (++)**, **MLLT10P1 (++)**, **PYGB**
GRS from full GWAS**							-0.28	0.07	**6.15E-05**	
GRS from GWAS old only§							-0.16	0.10	1.20E-01	
Putamen										
rs7445169	5q14.3	87,703,099	G	0.25	TMEM161B-AS1	rs2410767	-0.08	0.03	1.25E-02	**CTC-498M16.4**, **TMEM161B**, **TMEM161B-AS1**
rs2715135	7p12.1	50,750,128	T	0.34	GRB10	rs2244479	0.03	0.03	3.61E-01	
rs10886017	10q25.3	118,672,531	A	0.27	KIAA1598	rs7902527	0.04	0.03	1.59E-01	KIAA1598
rs2512662	11q14.1	83,130,781	A	0.30	CCDC90B,DLG2	rs1432054	0.08	0.03	9.62E-03	
rs3133370	11q14.3	92,026,446	C	0.37	MIR4490,FAT3	rs1187162	0.05	0.03	6.52E-02	
rs12800264	11q23.3	117,396,269	A	0.18	DSCAML1	rs35200015	-0.16	0.04	**1.30E-05**	
rs8017172	14q22.3	56,199,048	A	0.44	KTN1,RPL13AP3	rs945270	-0.15	0.03	**1.43E-07**	
rs17488580	18q21.2	50,757,261	T	0.40	DCC	rs62098013	0.09	0.03	2.34E-03	
rs6060954	20q11.21	30,383,187	T	0.34	TPX2	rs6087771	-0.08	0.03	5.05E-03	**FRG1B**, **MLLT10P1**, **MYLK2 (++)**
GRS from full GWAS**							-0.67	0.10	**5.04E-11**	
GRS from GWAS old only§							-1.05	0.21	**5.92E-07**	
Thalamus										
rs74504435	7p11.2	54,949,256	G	0.10	SEC61G,EGFR	rs142461330	0.09	0.05	5.17E-02	
rs12600720	17q25.3	78,448,640	G	0.31	NPTX1	rs12600720	0.04	0.03	2.31E-01	
GRS from full GWAS**							-0.21	0.31	5.01E-01	
GRS from GWAS old only§							0.07	0.54	8.98E-01	

P-values in **bold** significant results after multiple testing correction.
Genes in bold significant in i-Share TWAS and (++) after multiple testing correction.

* For each locus associated at genome-wide significant level with at least one subcortical structure in Satizabal et al and Hibar et al, associations of the lead SNP and nearby variants (±250 kb) in moderate to high LD (LD-r^2>0.5) with the corresponding phenotype were tested in young adults. Only the top SNP of each locus is presented in this table.
† TWAS analyses based on the summary statistics from Satizabal et al, Nat Genet 2019 for subcortical volumes (except hippocampal volume) and from Hibar et al, Nat Commun 2017 for hippocampal volume and results from the i-Share TWAS (n=1,586)
‡ A gene was consired as in the same locus than the top SNP from the GWAS if at least one of its eQTLs was in LD (r^2 > 0.01) with the top SNP
** GRS generated using the SNPs with p<5e-08 from the summary statistics of the GWAS of subcortical volumes from Satizabal et al. Nat Genet 2019 and from Hibar et al, Nat Commun 2017 for hippocampal volume
§ GRS generated using the SNPs with p<5e-08 from the summary statistics of the GWAS of subcortical volumes from Satizabal et al. Nat Genet 2019 after excluding cohorts containing young participants and from Bis et al, Nat Genet 2012 for hippocampal volume

Table S3: List of the colocalized (COLOC PP4 > 0.75) genes identified in transcriptome-wide association study using the GWAS summary statistics for subcortical volumes

Gene	Chr	Start	Stop
Accumbens			
NBR1	17	41,322,487	41,363,708
Amygdala			
EPS15	1	51,819,934	51,984,995
Caudate nucleus			
QRICH1	3	49,067,141	49,131,504
CDC42SE2	5	130,599,701	130,730,382
PCDHA4	5	140,186,658	140,391,929
AZGP1	**7**	**99,564,349**	**99,573,735**
CNPY4	**7**	**99,717,264**	**99,723,128**
MRPL43	10	102,737,578	102,747,272
RIPK3	**14**	**24,805,226**	**24,809,242**
DEGS2	**14**	**100,612,752**	**100,626,012**
EIF3C	16	28,390,902	28,437,775
NPIPB7	**16**	**28,467,693**	**28,481,868**
CCDC101	**16**	**28,565,248**	**28,603,111**
SULT1A1	**16**	**28,616,907**	**28,634,907**
NPIPB9	**16**	**28,763,755**	**28,784,144**
RP11-1348G14.4	**16**	**28,814,097**	**28,829,149**
TUFM	**16**	**28,853,731**	**28,857,729**
RP11-22P6.2	**16**	**28,873,487**	**28,873,851**
MYO15A	17	18,012,019	18,083,116
ALKBH5	17	18,086,866	18,113,267
DCAKD	17	43,100,705	43,138,477
SUPT4H1	**17**	**56,422,535**	**56,429,599**
DDX5	17	62,494,373	62,502,484
MYLK2	**20**	**30,407,177**	**30,422,500**
RNF185	22	31,556,137	31,603,005
Hippocampus			
METTL10	10	126,447,405	126,480,439
FER1L4	20	34,146,506	34,195,484
RBM12	20	34,236,846	34,252,878
Pallidum			
PHTF1	1	114,239,823	114,301,777
CRIPT	2	46,844,310	46,857,315
STAG3	7	99,775,346	99,812,010
RP11-350N15.5	8	38,200,000	38,200,000
SLC20A2	**8**	**42,273,979**	**42,397,356**
SMIM19	**8**	**42,396,297**	**42,408,140**
RP11-750H9.5	**11**	**47,400,000**	**47,400,000**
SLC39A13	**11**	**47,430,045**	**47,438,051**
CELF1	**11**	**47,487,488**	**47,574,792**
NDUFS3	**11**	**47,600,561**	**47,606,115**
TCF25	16	89,939,993	89,977,792
MC1R	16	89,984,286	89,987,385
ENTPD6	20	25,176,338	25,207,360
PYGB	20	25,228,705	25,278,648
RP4-610C12.4	20	29,513,582	29,521,213

Gene	Chr	Start	Stop
FRG1B	**20**	**29,611,878**	**29,634,007**
MLLT10P1	**20**	**29,637,583**	**29,638,138**
Putamen			
ACTR1B	2	98,272,401	98,280,561
TMEM161B	**5**	**87,485,449**	**87,564,696**
TMEM161B-AS1	**5**	**87,564,698**	**87,732,491**
CTC-498M16.4	**5**	**87,700,000**	**87,800,000**
PCDHA4	5	140,186,658	140,391,929
KIAA1598	**10**	**118,642,887**	**118,886,097**
RIPK3	**14**	**24,805,226**	**24,809,242**
RAB34	17	27,041,298	27,045,286
G6PC3	17	42,148,097	42,153,712
RP4-610C12.3	20	29,521,637	29,522,987
FRG1B	**20**	**29,611,878**	**29,634,007**
MLLT10P1	**20**	**29,637,583**	**29,638,138**
MYLK2	**20**	**30,407,177**	**30,422,500**
Thalamus			
FAIM	3	138,327,541	138,352,213

In **bold**, loci already identified in GWAS (eQTLs genome-wide significant or in LD (r^2>0.01) with the lead SNP of a genome-wide significant locus)

Table S4: Results of the GSMR analyses of subcortical volumes with Alzheimer's and Parkinson's diseases

	Alzheimer's disease				Parkinson's disease			
	N SNPs	Beta	SE	p	N SNPs	Beta	SE	p
Accumbens								
Young adults	82	0.01	0.00	8.18E-02	27	-0.02	0.01	2.73E-02
Middle-aged to older adults	64	-0.05	0.01	**1.05E-09**	27	0.01	0.01	5.46E-01
Older adults	95	-0.05	0.01	**4.76E-06**	28	0.01	0.02	5.10E-01
Amygdala								
Young adults	82	0.00	0.01	8.51E-01	27	0.00	0.02	8.25E-01
Middle-aged to older adults	62	-0.04	0.01	**2.52E-07**	27	-0.02	0.01	1.43E-01
Older adults	95	-0.03	0.01	**2.81E-03**	28	-0.01	0.02	4.30E-01
Caudate Nucleus								
Young adults	81	0.05	0.02	4.39E-02	27	0.05	0.04	1.54E-01
Middle-aged to older adults	59	-0.02	0.01	**3.10E-03**	25	0.00	0.01	9.40E-01
Older adults	94	-0.03	0.01	**5.85E-03**	28	0.00	0.02	8.73E-01
Hippocampus								
Young adults	82	0.00	0.02	8.39E-01	27	0.06	0.03	3.30E-02
Middle-aged to older adults	98	-0.05	0.01	**5.07E-10**	29	0.00	0.01	8.41E-01
Older adults	55	-0.04	0.01	**8.92E-04**	22	0.03	0.02	5.85E-02
Pallidum								
Young adults	82	0.01	0.01	2.30E-01	27	0.02	0.02	2.40E-01
Middle-aged to older adults	63	-0.02	0.01	5.47E-02	27	-0.01	0.01	5.08E-01
Older adults	94	-0.01	0.01	3.30E-01	28	0.00	0.02	7.92E-01
Putamen								
Young adults	82	0.10	0.03	**6.56E-04**	27	-0.07	0.04	9.42E-02
Middle-aged to older adults	60	-0.04	0.01	**3.06E-06**	27	-0.02	0.01	7.77E-02
Older adults	95	-0.04	0.01	**5.04E-04**	28	-0.01	0.02	7.19E-01
Thalamus								
Young adults	82	-0.01	0.03	6.87E-01	27	0.13	0.05	**2.79E-03**
Middle-aged to older adults	63	-0.03	0.01	**7.58E-05**	27	0.01	0.01	2.26E-01
Older adults	95	-0.03	0.01	**2.64E-03**	27	0.00	0.02	7.84E-01

P-values in **bold** significant results after multiple testing correction ($p<6.25\times10^{-3}$)