nature genetics

ARTICLES

https://doi.org/10.1038/s41588-019-0511-y

Genetic architecture of subcortical brain structures in 38,851 individuals

Subcortical brain structures are integral to motion, consciousness, emotions and learning. We identified common genetic variation related to the volumes of the nucleus accumbens, amygdala, brainstem, caudate nucleus, globus pallidus, putamen and thalamus, using genome-wide association analyses in almost 40,000 individuals from CHARGE, ENIGMA and UK Biobank. We show that variability in subcortical volumes is heritable, and identify 48 significantly associated loci (40 novel at the time of analysis). Annotation of these loci by utilizing gene expression, methylation and neuropathological data identified 199 genes putatively implicated in neurodevelopment, synaptic signaling, axonal transport, apoptosis, inflammation/infection and susceptibility to neurological disorders. This set of genes is significantly enriched for *Drosophila* orthologs associated with neurodevelopmental phenotypes, suggesting evolutionarily conserved mechanisms. Our findings uncover novel biology and potential drug targets underlying brain development and disease.

Subcortical brain structures are essential for the control of autonomic and sensorimotor functions[1,2], the modulation of processes involved in learning, memory and decision-making[3,4], and in emotional reactivity[5,6] and consciousness[7]. They often act through networks influencing input to and output from the cerebral cortex[8,9]. The pathology of many cognitive, psychiatric and movement disorders is restricted to, begins in or predominantly involves subcortical brain structures and related circuitries[10]. For instance, tau pathology has shown to manifest itself early in the brainstem of individuals with Alzheimer's disease before spreading to cortical areas through efferent networks[11]. Similarly, the formation of Lewy bodies and Lewy neurites in Parkinson's disease appears early in the lower brainstem (and olfactory structures) before affecting the substantia nigra[12].

Recent investigations have identified genetic loci influencing the volumes of the putamen, caudate and pallidum, which pointed to genes controlling neurodevelopment and learning, apoptosis and the transport of metals[13,14]. However, a larger study combining these samples and including individuals of a broad age range across diverse studies would enable increased power to identify additional novel genetic variants contributing to variability in subcortical structures, and further improve our understanding of brain development and disease.

We sought to identify novel genetic variants influencing the volumes of seven subcortical structures (the nucleus accumbens, amygdala, caudate nucleus, putamen, globus pallidus, thalamus and brainstem (including the mesencephalon, pons and medulla oblongata)), through genome-wide association (GWA) analyses in almost 40,000 individuals from 53 study samples (Supplementary Tables 1–3) from the Cohorts of Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium, the Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) consortium and UK Biobank.

Results

Heritability. To examine the extent to which genetic variation accounts for variation in subcortical brain volumes, we estimated their heritability in two family-based cohorts: the Framingham Heart Study (FHS) and the Austrian Stroke Prevention Study Family Study (ASPS-Fam). Our analyses were in line with previous studies conducted in twins[15], suggesting that variability in subcortical volumes is moderately to highly heritable. The structures with the highest heritability in the FHS and ASPS-Fam were the brainstem (ranging from 79–86%), caudate nucleus (71–85%), putamen (71–79%) and nucleus accumbens (66%), followed by the globus pallidus (55–60%), thalamus (47–54%) and amygdala (34–59%) (Fig. 1 and Supplementary Table 4). We additionally estimated single-nucleotide polymorphism (SNP)-based heritability (h^2_g) using genome-wide complex trait analysis (GCTA) in the Rotterdam Study, and linkage disequilibrium score regression (LDSC) in the full European sample. As expected, SNP-based heritability estimates were somewhat lower, ranging from 17% for the amygdala to 47% for the thalamus using GCTA, and ranging from 9% for the amygdala to 33% for the brainstem using LDSC. These values are consistent with heritability estimates reported by UK Biobank[14].

Genome-wide associations. We undertook a GWA analysis on the magnetic resonance imaging (MRI)-derived volumes of subcortical structures using the 1000 Genomes Project[16] reference panel (phase 1; version 3) for imputation of missing variants in CHARGE and ENIGMA. UK Biobank performed imputation of variants using the Haplotype Reference Consortium (HRC) reference panel[17] (see details on image acquisition and genotyping in Supplementary Tables 5 and 6, respectively). Our sample comprised up to $n = 37{,}741$ individuals of European ancestry from 48 study samples across CHARGE, ENIGMA and UK Biobank. Additionally, we included three samples for generalization in African Americans (up to $n = 769$) and two for generalization in Asians ($n = 341$). Details on the population characteristics, definition of the outcome and genotyping are provided in Supplementary Tables 2–5. Each study examined the association between genetic variants with a minor allele frequency (MAF) of ≥1% and the volumes of subcortical structures (average volume for bilateral structures) using additive genetic models adjusted for sex, age and total intracranial volume (or total brain volume in UK Biobank), as well as age^2, population structure, psychiatric diagnosis (ENIGMA cohorts), and study site when applicable. After quality control, we conducted meta-analyses per ethnicity combining all samples using sample-size-weighted fixed-effects methods in METAL[18]. An analysis of genetic correlations (r_g) showed consistency of associations across the CHARGE and ENIGMA consortia (combined) and UK Biobank ($r_g > 0.94$;

A full list of affiliations appears at the end of the paper.

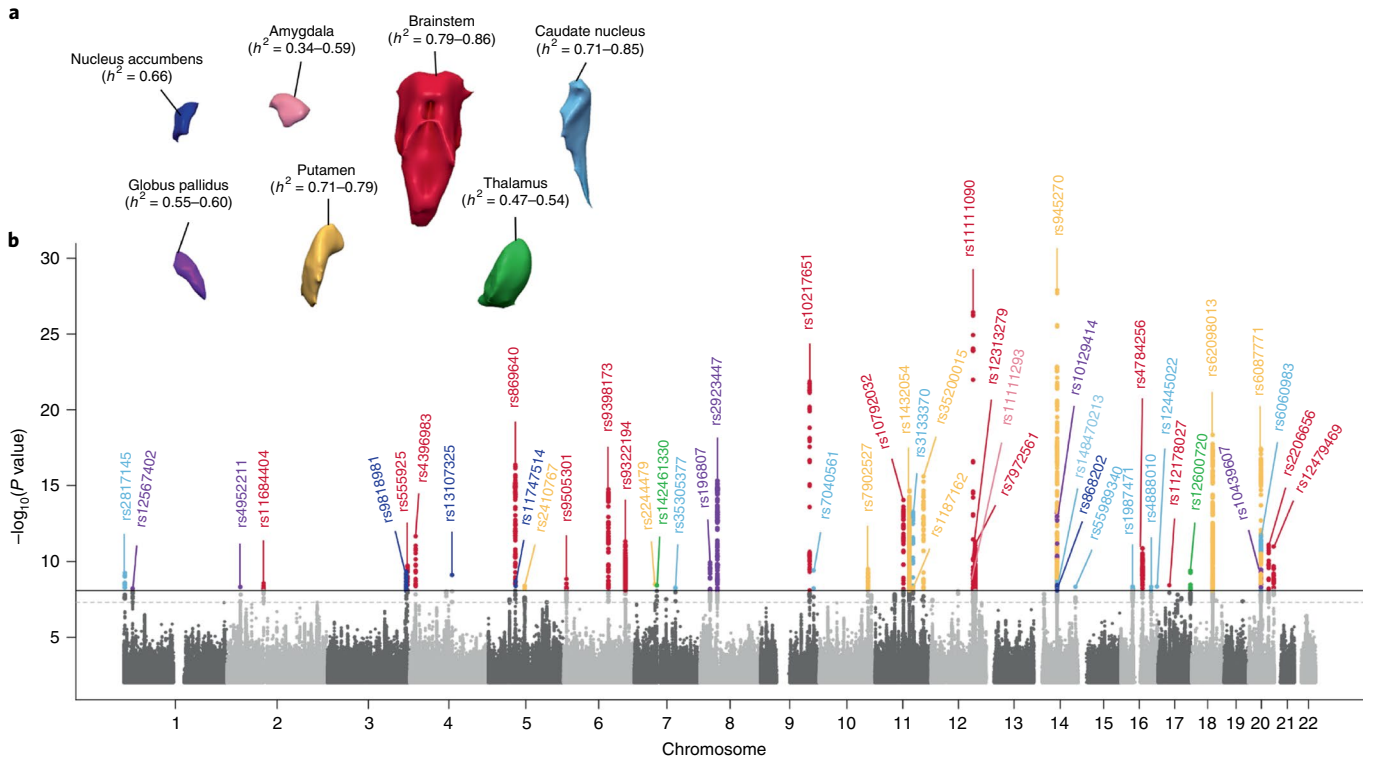

Fig. 1 | Heritability and Manhattan plot of genetic variants associated with subcortical brain volumes in the European sample. a, Family-based heritability (h^2) estimates were performed with SOLAR in the FHS ($n = 895$) and ASPS-Fam ($n = 370$). **b**. Combined Manhattan plot highlighting the most significant SNPs across all subcortical structures (nucleus accumbens, $n = 32{,}562$; amygdala, $n = 34{,}431$; brainstem, $n = 28{,}809$; caudate, $n = 37{,}741$; pallidum, $n = 34{,}413$; putamen, $n = 37{,}571$; thalamus, $n = 34{,}464$). Variants are colored differently for each structure as in **a**. Linear regression models were adjusted for sex, age, age^2, total intracranial volume (CHARGE and ENIGMA) or total brain volume (UK Biobank), and population stratification. The solid horizontal line denotes genome-wide significance, as set in this study after additional Bonferroni correction for six independent traits ($P < 5 \times 10^{-8}/6 = 8.3 \times 10^{-9}$ for two-sided tests). The dashed horizontal line denotes the classic genome-wide threshold of $P < 5 \times 10^{-8}$. Individual Manhattan plots are provided in the Supplementary Note.

$P < 1.46 \times 10^{-15}$), showing the similar genetic architecture of subcortical volumes in these two datasets.

We identified 48 independent genome-wide significant SNPs across all seven subcortical structures, 40 of which were novel at the time of analysis (Table 1). Among these, 26 SNPs were located within genes (one missense; 25 intronic) and 22 were located in intergenic regions. Most of the inflation observed in the quantile plots (Supplementary Fig. 1) was due to polygenic effects. We carried forward these 48 SNPs for in silico generalization in African American and Asian samples, and performed a combined meta-analysis of all samples (Supplementary Table 7). Of the 46 SNPs present in the generalization samples, the direction of association was the same for 13 across all ethnicities and for an additional six SNPs in either the African American or the Asian samples. In the combined meta-analysis, 43 of the 48 associations remained significant, and for 21 SNPs, the strength of association increased when all samples were combined. Although we did not find significant associations for most SNPs at the generalization sample level (probably due to their limited sample size), the sign test for the direction of effect suggested that a large proportion of the SNPs associated with subcortical volumes in the European sample were also associated in the African American and Asian samples at the polygenic level ($P < 1 \times 10^{-4}$; Supplementary Table 8).

To functionally annotate the 48 SNPs identified in the European sample, we used Locus Zoom[19], investigated expression quantitative trait loci (eQTLs) and methylation QTLs (meQTLs) in postmortem brains from the Religious Order Study and the Rush Memory and Aging Project (ROSMAP), and queried *cis*- and *trans*-eQTL datasets in brain and non-brain tissues for the top 48 SNPs or their proxies (linkage disequilibrium $r^2 > 0.8$), using the European population reference (Supplementary Tables 9–12). Lead variants and their proxies were annotated to genes based on the combination of physical proximity, eQTLs and meQTLs, which in some instances assigned more than one gene to a single SNP. Most of our index SNPs had genes assigned based on more than one functional source. This strategy allowed us to identify 199 putatively associated genes (Supplementary Table 13). More details are provided in the Supplementary Note.

Associations with cognition and neuropathology. Although individual SNPs were not related to neuropathological traits or cognitive function in ROSMAP (Supplementary Table 14), we found that the cortical messenger RNA expression of 12 of our putatively associated genes was associated with neuropathological alterations typically observed in Alzheimer's disease (Supplementary Table 15). These included β-amyloid load/the presence of neuritic plaques (*APOBR*, *FAM65C*, *KTN1*, *NUPR1* and *OPA1*) and tau density/neurofibrillary tangles (*FAM65C*, *MEPCE*, *OPA1* and *STAT1*). Many of these genes—together with *ANKRD42*, *BCL2L1*, *RAET1G*, *SGTB* and *ZCCHC14*—were also related to cognitive function.

Phenotypic and genetic correlations. We explored both phenotypic (Supplementary Table 16) and genetic (Supplementary Table 17) correlations among subcortical volumes. We also investigated genetic correlations of subcortical volumes with traits

Table 1 | Genome-wide association results for subcortical brain volumes in Europeans from the CHARGE and ENIGMA consortia and UK Biobank

SNP	Chromosome	Position	Function	A1/A2	A1 frequency	Weight (SNP *n*)	*Z* score	*P* value[a]	Direction[b]	I^{2c}
Nucleus accumbens (*n* = 32,562)										
rs9818981[d]	3	190,602,087	Intergenic	A/G	0.09	32,282	−6.23	4.70×10^{-10}	- - -	63.2
rs13107325	4	103,188,709	Missense	T/C	0.06	32,283	6.15	7.74×10^{-10}	+++	76.2
rs11747514[d]	5	65,839,259	Intronic	T/G	0.22	32,562	−5.99	2.11×10^{-9}	- - -	0
rs868202[d]	14	56,195,762	Intergenic	T/C	0.56	32,562	5.90	3.55×10^{-9}	+++	0
Amygdala (*n* = 34,431)										
rs11111293[d]	12	102,921,296	Intergenic	T/C	0.78	34,313	6.25	4.16×10^{-10}	+++	0
Brainstem (*n* = 28,809)										
rs11111090	12	102,326,461	Intergenic	A/C	0.52	28,809	10.79	3.70×10^{-27}	+++	0
rs10217651[d]	9	118,923,652	Intronic	A/G	0.39	28,809	9.78	1.40×10^{-22}	+++	0
rs869640[d]	5	65,015,128	Intronic	A/C	0.72	28,809	−8.40	4.36×10^{-17}	- - -	9.5
rs9398173[d]	6	109,000,316	Intronic	T/C	0.33	28,809	−7.95	1.80×10^{-15}	- - -	19.0
rs10792032[d]	11	68,984,602	Intergenic	A/G	0.49	28,648	7.75	9.08×10^{-15}	+++	39.4
rs4396983[d]	4	15,132,604	Intergenic	A/G	0.44	28,809	−7.02	2.27×10^{-12}	- - -	73.6
rs9322194[d]	6	149,920,249	Intronic	T/C	0.34	28,156	6.91	4.94×10^{-12}	+++	0
rs7972561[d]	12	107,139,983	Intronic	A/T	0.33	28,809	6.90	5.05×10^{-12}	+++	0
rs2206656[d]	20	49,130,119	Intronic	C/G	0.61	28,809	6.83	8.26×10^{-12}	+++	0
rs12479469[d]	20	61,145,196	Intergenic	A/G	0.33	25,822	−6.80	1.08×10^{-11}	- - -	65.6
rs4784256[d]	16	52,814,559	Intergenic	A/G	0.40	28,809	6.76	1.41×10^{-11}	+++	0
rs555925[d]	3	193,544,359	intergenic	T/G	0.41	27,934	6.37	1.88×10^{-10}	+++	62.9
rs12313279[d]	12	102,846,504	Intronic	A/G	0.29	28,809	6.21	5.39×10^{-10}	+++	24.9
rs9505301[d]	6	7,887,131	Intronic	A/G	0.89	28,691	−6.05	1.41×10^{-9}	- - -	43.2
rs11684404[d]	2	88,924,622	Intronic	T/C	0.66	28,809	−5.95	2.73×10^{-9}	- - -	0
rs112178027[d]	17	27,564,013	Intergenic	T/C	0.17	28,809	−5.90	3.67×10^{-9}	- - -	0
Caudate nucleus (*n* = 37,741)										
rs3133370	11	92,026,446	Intergenic	T/C	0.67	37,741	7.52	5.59×10^{-14}	+++	44.9
rs6060983[d]	20	30,420,924	Intronic	T/C	0.70	37,741	7.04	1.95×10^{-12}	+++	0
rs7040561[d]	9	128,528,978	Intronic	A/T	0.85	34,049	−6.26	3.84×10^{-10}	- - -	0
rs2817145[d]	1	3,133,422	Intronic	A/T	0.19	35,598	6.20	5.71×10^{-10}	+++	65.3
rs148470213[d]	14	56,193,700	Intergenic	T/C	0.54	29,429	6.18	6.48×10^{-10}	++?	0
rs1987471[d]	16	28,825,866	Intergenic	T/G	0.63	37,741	5.87	4.40×10^{-9}	+++	0
rs12445022[d]	16	87,575,332	Intergenic	A/G	0.33	37,741	5.87	4.45×10^{-9}	+++	0
rs55989340[d]	14	100,635,222	Intergenic	A/G	0.74	37,741	−5.86	4.62×10^{-9}	- - -	52.0
rs4888010[d]	16	73,895,046	Intergenic	A/G	0.47	37,741	5.86	4.67×10^{-9}	+++	74.9
rs35305377[d]	7	99,938,955	Intronic	A/G	0.55	33,429	−5.84	5.36×10^{-9}	- - -	47.8
Globus pallidus (*n* = 34,413)										
rs2923447	8	42,439,848	Intergenic	T/G	0.59	34,413	8.11	4.88×10^{-16}	+++	34.0
rs10129414[d]	14	56,193,272	Intergenic	A/G	0.44	34,413	−7.53	5.11×10^{-14}	- - -	0
rs196807[d]	8	24,682,649	Intergenic	A/G	0.18	34,295	6.44	1.17×10^{-10}	+++	21.1
rs10439607[d]	20	30,258,541	Intronic	A/G	0.30	34,413	−6.28	3.35×10^{-10}	- - -	0
rs4952211[d]	2	32,611,512	Intronic	T/C	0.43	34,252	−5.86	4.72×10^{-9}	- - -	61.9
rs12567402[d]	1	21,870,213	Intronic	T/C	0.33	34,214	5.81	6.17×10^{-9}	+++	0
Putamen (*n* = 37,571)										
rs945270	14	56,200,473	Intergenic	C/G	0.58	37,571	15.03	5.02×10^{-51}	+++	57.3
rs62098013	18	50,863,861	Intronic	A/G	0.38	37,571	8.92	4.59×10^{-19}	+++	33.9
rs6087771	20	30,306,724	Intronic	T/C	0.71	36,291	8.69	3.75×10^{-18}	+++	7.5
rs35200015[d]	11	117,383,215	Intronic	A/G	0.19	37,571	−8.19	2.51×10^{-16}	- - -	0

Continued

Table 1 | Genome-wide association results for subcortical brain volumes in Europeans from the CHARGE and ENIGMA consortia and UK Biobank (continued)

SNP	Chromosome	Position	Function	A1/A2	A1 frequency	Weight (SNP *n*)	*Z* score	*P* value[a]	Direction[b]	I^{2c}
rs1432054	11	83,260,225	Intronic	A/G	0.64	37,571	−7.94	2.10×10^{-15}	- - -	0
rs7902527[d]	10	118,715,399	Intronic	A/G	0.24	37,108	6.29	3.13×10^{-10}	+++	0
rs2244479[d]	7	50,738,987	Intronic	T/C	0.65	36,291	−5.92	3.17×10^{-9}	- - -	32.1
rs2410767[d]	5	87,705,268	Intronic	C/G	0.78	37,571	5.88	3.99×10^{-9}	+++	0
rs1187162[d]	11	92,011,126	Intergenic	T/C	0.42	37,571	5.84	5.14×10^{-9}	+++	0
Thalamus (*n* = 34,464)										
rs12600720[d]	17	78,448,640	Intronic	C/G	0.69	33,023	6.25	4.06×10^{-10}	+++	0
rs142461330[d]	7	55,012,097	Intergenic	T/C	0.92	34,185	−5.90	3.69×10^{-9}	- - -	0

Linear regression models are adjusted for sex, age, age^2, total intracranial volume (CHARGE and ENIGMA) or total brain volume (UK Biobank), and population stratification. [a]*P* values are two tailed. Significance was set at $P<8.3\times10^{-9}$ after additional Bonferroni correction for six independent traits ($5\times10^{-8}/6$). [b]Direction of association, ordered as CHARGE, ENIGMA, and UK Biobank. [c]Heterogeneity as estimated proportion of total variance. [d]Novel SNPs. A1, coded allele; A2, non-coded allele.

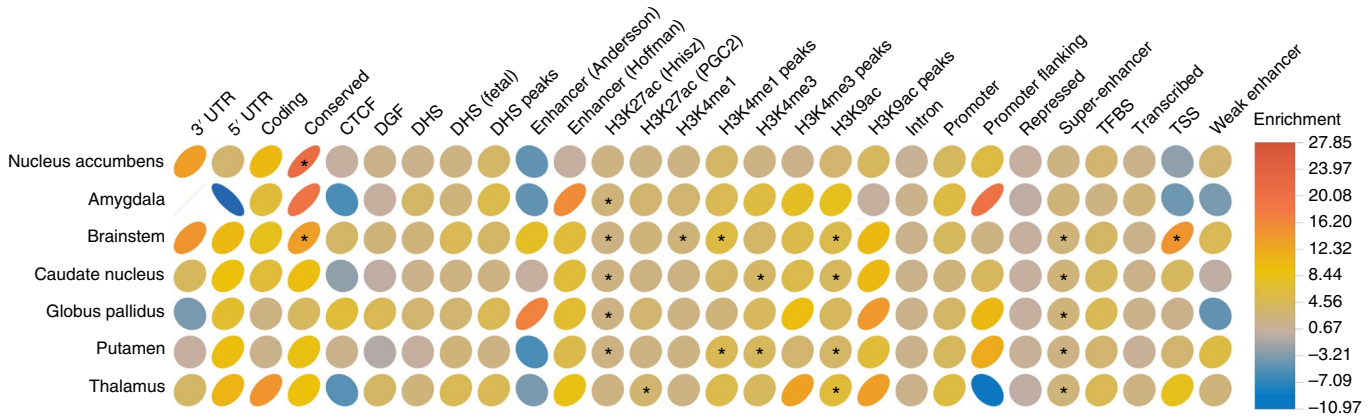

Fig. 2 | Partitioning heritability by functional annotation categories. Analyses performed in the European sample (nucleus accumbens, *n* = 32,562; amygdala, *n* = 34,431; brainstem, *n* = 28,809; caudate, *n* = 37,741; pallidum, *n* = 34,413; putamen, *n* = 37,571; thalamus, *n* = 34,464). Plotted ellipses represent enrichment (proportion of h^2_g explained/proportion of SNPs in a given functional category) for subcortical structures (*y* axis) across 28 functional categories (*x* axis). The color bar indicates the magnitude and direction of enrichment. Starred pairs denote significant over-representation after Bonferroni correction for 168 tests (28 annotation categories and six independent traits; $P<3\times10^{-4}$). CTCF, CCCTC-binding factor; DGF, digital genomic footprint; DHS, DNase I hypersensitivity site; PGC2, Psychiatric Genomics Consortium; TFBS, transcription-factor-binding sites; TSS, transcription start site; UTR, untranslated region. Sources represented on the *x* axis are described in ref. [30].

previously examined in the CHARGE and ENIGMA consortia, including MRI-defined brain volumes[20–22], stroke subtypes[23], anthropometric traits[24], general cognitive function[25], Alzheimer's disease[26], Parkinson's disease[27], bipolar disorder and schizophrenia[28], and attention deficit/hyperactivity disorder (ADHD)[29]. We observed strong phenotypic and genetic overlap among most subcortical structures using LDSC methods, consistent with our finding that many of the loci identified have pleiotropic effects on the volumes of several subcortical structures.

As expected, we found strong genetic correlations among the nuclei composing the striatum—particularly between the nucleus accumbens and the caudate nucleus ($P=9.83\times10^{-19}$) and between the nucleus accumbens and the putamen ($P=1.02\times10^{-17}$). The genetic architecture of thalamic volume highly overlapped with that of most subcortical volumes, except for the caudate nucleus. In contrast, there were no significant genetic correlations for the volume of the brainstem with that of most structures, with the exception of very strong correlations with volumes of the thalamus ($P=1.56\times10^{-22}$) and the globus pallidus ($P=1.52\times10^{-21}$). Individual-level analyses using GCTA in the Rotterdam Study (*n* = 3,486) showed similar correlations despite the smaller sample.

We also observed strong genetic correlations for hippocampal volumes with amygdalar and thalamic volumes. Height correlated with thalamic volumes, and the volume of the brainstem was inversely correlated with ADHD. Notably, caudate nucleus volumes correlated with white matter hyperintensity burden.

Cross-species analysis. To investigate for potential evolutionarily conserved requirements of our gene set in neurodevelopment, neuronal maintenance or both, we examined the available genetic and phenotypic data from the fruit fly *Drosophila melanogaster*. Importantly, compared with mammalian models, the fly genome has been more comprehensively interrogated for roles in the nervous system. We found that a large proportion of candidate genes for human subcortical volumes are strongly conserved in the *Drosophila* genome (59%), and many of these genes appear to have conserved nervous system requirements (Supplementary Table 18). To examine whether this degree of conservation was greater than that expected by chance, we leveraged systematic, standardized phenotype data based on FlyBase annotations using controlled vocabulary terms. Indeed, 22% of the conserved fly homologs are documented to cause 'neuroanatomy-defective' phenotypes in flies,

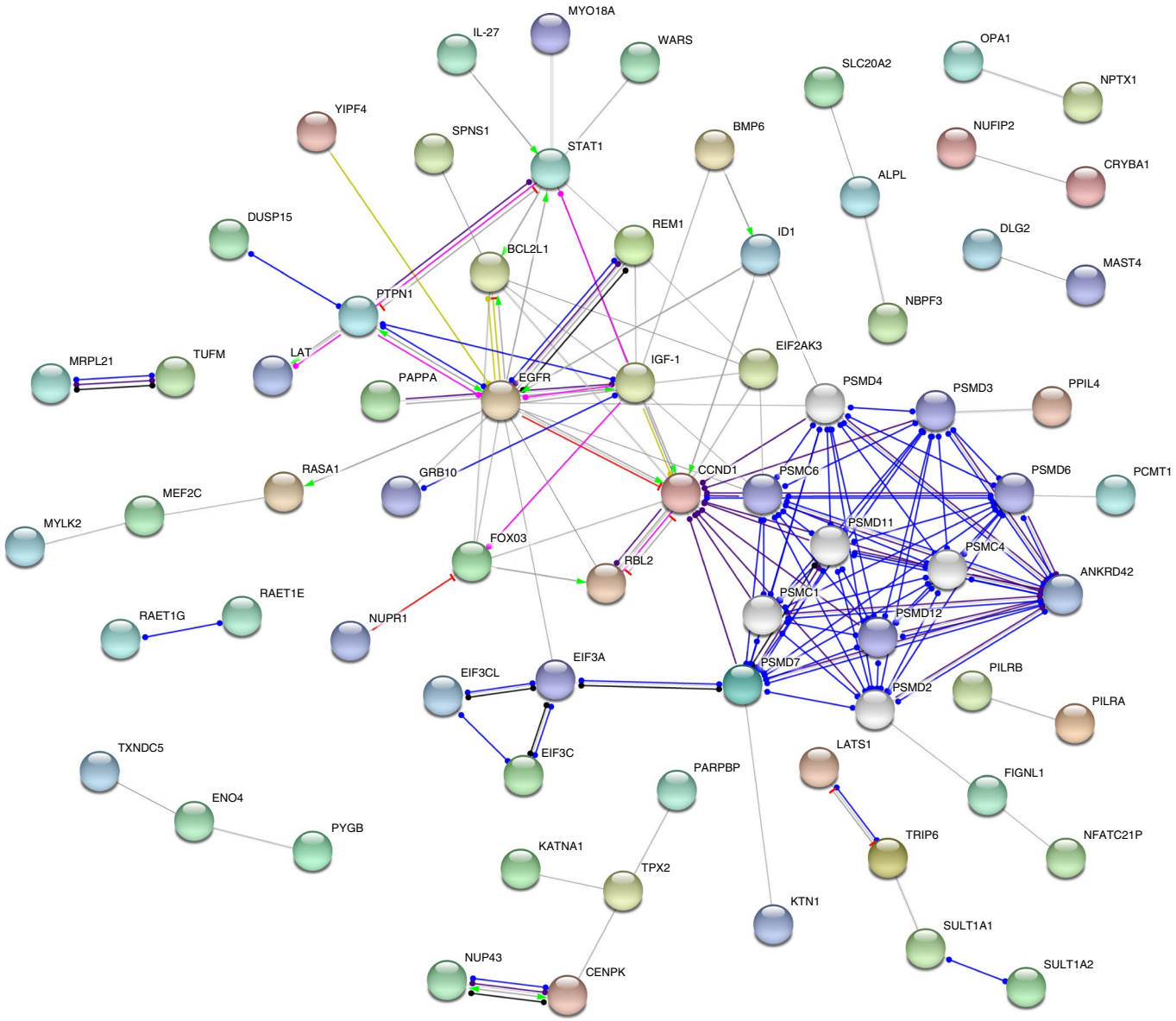

Fig. 3 | Protein–protein interaction network of 148 genes enriched for common variants influencing the volume of subcortical structures. The arrowheads represent protein–protein associations, where the edge color indicates the predicted mode of action (bright green, activation; pink, post-translational modification; red, inhibition; dark blue, binding, purple, catalysis; light blue, phenotype; black, reaction; yellow, transcriptional regulation) and the arrowhead shape represents the predicted action effects (pointed arrow, positive; flat arrow, negative; oval arrow, unspecified). Colored nodes represent the queried proteins and first shell of interactors (five maximum), whereas white nodes represent the second shell of interactors (five maximum).

representing a significant ($P=7.3\times10^{-4}$), nearly twofold enrichment compared with 12.9% representing all *Drosophila* genes associated with such phenotypes (Supplementary Table 19).

Partitioning heritability. We further investigated enrichment for functional categories of the genome using stratified LDSC methods[30] (Fig. 2). Super-enhancers were significantly enriched in most subcortical structures, with 17% of SNPs explaining 43% of SNP heritability in the brainstem, 39% in the caudate, 44% in the pallidum, 37% in the putamen and 38% in the thalamus. Similarly, strong enrichment was observed for regular enhancers (H3K27ac annotations from Hnisz et al.[31]) in several subcortical structures, explaining over 60% of their SNP heritability. Conserved regions were enriched in the nucleus accumbens and the brainstem, with 2.6% of SNPs explaining 53 and 35% of their SNP heritability, respectively. Finally, only the brainstem showed enrichment for transcription start sites, with 1.8% of SNPs explaining 26% of this structure SNP heritability. The full results are presented in Supplementary Table 20.

Protein–protein interactions. To explore potential functional relationships between proteins encoded by our set of genes, we conducted protein–protein interaction analyses in STRING[32]. Our results showed enrichment of genes involved in brain-specific pathways (that is, regulation of neuronal death and neuronal apoptosis), as well as immune-related (that is, antigen processing and Epstein–Barr virus infection) and housekeeping processes (that is, proteasome, cell differentiation and signaling). Figure 3 shows the protein network, and the detailed pathways are presented in Supplementary Table 21.

Discussion

We undertook a large GWA meta-analysis of variants associated with MRI-derived volumes of the nucleus accumbens, amygdala,

brainstem, caudate nucleus, globus pallidus, putamen and thalamus, including almost 40,000 individuals from 53 study samples worldwide. Our analyses identified a set of 199 candidate genes influencing the volume of these subcortical brain structures, most of which have relevant roles in the nervous system.

Our results show wide overlap of genetic variants determining the volume of subcortical structures, as elucidated from genetic correlations and individual look-ups among structures. We found that 26 candidate genes may influence more than one structure. For instance, significant SNPs near *KTN1* are also associated with the volume of the nucleus accumbens, caudate nucleus and globus pallidus, suggesting that this genomic region may have an important role in determining multiple subcortical brain volumes during development. Furthermore, 14 of the candidate genes were associated with the caudate, globus pallidus and putamen, supporting the shared genetic architecture of the functionally defined corpus striatum.

We identified genes implicated in neurodevelopment. We confirm that the 11q14.3 genomic region near the *FAT3* gene, which was previously associated with the caudate nucleus[13], additionally associated with the putamen in our analysis. This gene encodes a conserved cellular adhesion molecule implicated in neuronal morphogenesis and cell migration, based on mouse genetic studies[33]. SNPs near *PBX3* were associated with caudate volume. *PBX3* is robustly expressed in the developing caudate nucleus of the non-human primate *Macaca fuscata*, consistent with a role in striatal neurogenesis[34].

We found several genes involved in insulin/insulin-like growth factor 1 (IGF-1) signaling, including *IGF1*, *PAPPA*, *GRB10*, *SH2B1* and *TXNDC5*, across the amygdala, brainstem, caudate and putamen. *PAPPA* encodes a secreted metalloproteinase that cleaves IGF-binding proteins, thereby releasing bound IGF. Although IGF may be beneficial in early- and midlife, its effects may be detrimental during aging. Studies of pregnancy-associated plasma protein A similarly support antagonistic pleiotropy. Low circulating pregnancy-associated plasma protein A levels are a marker for adverse outcomes in human embryonic development[35], but in later life, higher levels have been associated with acute coronary syndromes and heart failure[36,37]. Furthermore, Grb10 and SH2B1 act as regulators of insulin/IGF-1 signaling through their SH2 domains[38]. Finally, TXNDC5 has been suggested to increase IGF-1 activity by inhibiting the expression of IGF-binding protein 1 in the context of rheumatoid arthritis[39].

Additional genes related to neurodevelopment include *PTPN1* (brainstem), *ALPL* and *NBPF3* (both related to the globus pallidus) and *SLC20A2* (nucleus accumbens). In studies of both human and mouse embryonic stem cells, PTPN1 was implicated as a critical regulator of neural differentiation[40]. In addition, *PTPN1* encodes a target for the transcriptional regulator encoded by *MECP2*, which causes the neurodevelopmental disorder Rett syndrome, and inhibition of PTPB1 is being explored as a therapeutic strategy in mouse Rett models[41]. *ALPL* mediates neuronal differentiation early during development and postnatal synaptogenesis in transgenic mouse models[42]. ALPL may also help propagate the neurotoxicity induced by tau[43], and its activity increases in Alzheimer's disease[44] and cognitive impairment[45]. *NBPF3* belongs to the neuroblastoma breakpoint family, which encodes domains of the autism- and schizophrenia-related DUF1220 protein[46]. *SLC20A2*, related to the globus pallidus and the thalamus, encodes an inorganic phosphate transporter for which more than 40 mutations have been described in association with familial idiopathic basal ganglia calcification (Fahr's syndrome)[47,48]. It is interesting to note that the other three solute carrier genes were identified in this GWA (*SLC12A9*, *SLC25A29* and *SLC39A8*), suggesting that the molecular transport of metals, amino acids and other solutes across the cellular membrane could play an important role in the development of subcortical brain structures.

Several genes were related to synaptic signaling pathways. We found a SNP in *NPTX1* related to the thalamus, a gene expressed in the nervous system. The encoded protein restricts synapse plasticity[49] and induces β-amyloid neurodegeneration in human and mouse brain tissues[50]. Additionally, we identified an intronic SNP in *SGTB* for the brainstem, which was an eQTL for the expression of SGTB in the dorsolateral prefrontal cortex (DLPFC). Experimental rat models showed that βSGT, which is highly expressed in the brain, forms a complex with the cysteine string protein and heat-shock protein cognate complex (CSP/Hsc70) to function as a chaperone guiding the refolding of misfolded proteins near synaptic vesicles[51]. Other experimental studies in *Caenorhabditis elegans*, showed that genetic manipulation of the ortholog *sgt-1* suppresses toxicity associated with expression of the human β-amyloid peptide[52]. Other genes involved in synaptic signaling are *CHPT1* (brainstem), which is involved in phosphatidylcholine metabolism in the brain, *KATNA1* (brainstem), a conserved regulator of neuronal process formation, outgrowth and synaptogenesis[53,54], and *DLG2* (putamen), encoding an evolutionarily conserved scaffolding protein involved in glutamatergic-mediated synaptic signaling and cell polarity[55] that has been associated with schizophrenia[56], cognitive impairment[57] and Parkinson's disease[58].

Another set of SNPs point to genes involved in autophagy and apoptotic processes, such as *DRAM1* and *FOXO3*, both of which are related to brainstem volumes. *DRAM1* encodes a lysosomal membrane protein involved in activating TP53-mediated autophagy and apoptosis[59], and mouse models mimicking cerebral ischemia and reperfusion have found that inhibiting the expression of *DRAM1* worsens cell injury[60]. The top SNP was also associated with a CpG site proximate to active transcription start sites upstream of *DRAM1* in several mature brain tissues. *FOXO3* has recently been identified as pivotal in an astrocyte network conserved across humans and mice involved in stress, sleep and Huntington's disease[61], and has been related to longevity[62]. In *Drosophila*, a *FOXO3* ortholog regulates dendrite number and length in the peripheral nervous system[63], and in the zebrafish *Danio rerio*, *Foxo3a* knockdown led to apoptosis and mispatterning of the embryonic central nervous system[64]. Additional genes involved in apoptotic processes are *BCL2L1* (globus pallidus and putamen), *BIRC6* (globus pallidus) and *OPA1* (brainstem).

Other genes have been implicated in axonal transport. We confirm the association between the 13q22 locus near *KTN1* with putamen volume[13], and expand by showing that this region is also associated with the nucleus accumbens, caudate and the globus pallidus. The most significant SNP (rs945270) is a robust eQTL for *KTN1* in peripheral blood cells. This gene encodes a kinesin-binding protein involved in the transport of cellular components along microtubules[65], and impairment of these molecular motors has been increasingly recognized in neurological diseases with a subcortical component[66]. The 5q12 locus upstream from *MAST4* was associated with nucleus accumbens volume. *MAST4* encodes a member of the microtubule-associated serine/threonine kinases. This gene has been associated with hippocampal volumes[20] and juvenile myoclonic epilepsy[67], and it appears to be differentially expressed in the prefrontal cortex of atypical cases of frontotemporal lobar degeneration[68]. In *Drosophila*, the knockdown of a conserved *MAST4* homolog enhanced the neurotoxicity of human tau[69], which aggregates to form neurofibrillary tangle pathology in Alzheimer's disease. Furthermore, we identified SNPs near *NEFL* and *NEFM* (globus pallidus), where the top SNP was an eQTL for these genes in subcortical brain tissue and esophagus mucosa. *NEFL* encodes the light chain, and *NEFM* the medium chain of the neurofilament. The proteins encoded by these genes determine neuronal caliber and conduction velocity[70]. Mutations in *NEFL* and *NEFM* genes have been related to neuropsychiatric disorders, and both proteins encoded by these genes are increasingly recognized as powerful biomarkers of neurodegeneration[71].

Finally, several of our candidate genes are also involved in inflammation, immunity and infection (*ANKRD42*, *DEFB124*, *IL27*, *NLRC4*, *PILRA/B*, *TRIM23* and *TRIM4*), in line with the

protein–protein interaction analysis highlighting the Kyoto Encyclopedia of Genes and Genomes–Epstein–Barr virus infection pathway. This suggests that immune-related processes may be an important determinant influencing subcortical volumes, as has been shown by other GWA studies of neurologic traits[72,73].

Overall, the loci identified by our study pinpoint candidate genes not only associated with human subcortical brain volumes, but also reported to disrupt invertebrate neuroanatomy when manipulated in *Drosophila* and many other animal models. Thus, our results are in line with the knowledge that the genomic architecture of central nervous system development has been strongly conserved during evolution. Partitioning heritability results suggest the nucleus accumbens and brainstem are particularly enriched in conserved regions.

One of the main limitations of our study was the small size of our generalization samples, which limits the generalizability of our results to non-European ethnicities. However, our analyses suggest significant concordance for the direction of effect across all ethnicities at the polygenic level. We hope diverse samples become increasingly available to further confirm our findings and make new discoveries. Additionally, we have focused on the discovery of common and less frequent variants. Further efforts to also reveal rare variants and epigenetic signatures associated with subcortical structures will provide an even more refined understanding of the underlying mechanisms involved.

In conclusion, we describe multiple genes associated with the volumes of MRI-derived subcortical structures in a large sample, leveraging diverse bioinformatics resources to validate and follow-up our findings. Our analyses indicate that the variability of evolutionarily old subcortical volumes of humans is moderately to strongly heritable, and that their genetic variation is also strongly conserved across different species. The majority of the variants identified in this analysis point to genes involved in neurodevelopment, regulation of neuronal apoptotic processes, synaptic signaling, axonal transport, inflammation/immunity and susceptibility to neurological disorders. We show that the genetic architecture of subcortical volumes overlaps with that of anthropometric measures and neuropsychiatric disorders. In summary, our findings expand the current understanding of the genetic variation related to subcortical structures, which can help in the identification of novel biological pathways of relevance to human brain development and disease.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41588-019-0511-y.

Received: 26 September 2017; Accepted: 5 September 2019;
Published online: 21 October 2019

References

1. Marsden, C. D. The mysterious motor function of the basal ganglia: the Robert Wartenberg Lecture. *Neurology* **32**, 514–539 (1982).
2. Yin, H. H. & Knowlton, B. J. The role of the basal ganglia in habit formation. *Nat. Rev. Neurosci.* **7**, 464–476 (2006).
3. McDonald, A. J. & Mott, D. D. Functional neuroanatomy of amygdalohippocampal interconnections and their role in learning and memory. *J. Neurosci. Res.* **95**, 797–820 (2016).
4. Hikosaka, O., Kim, H. F., Yasuda, M. & Yamamoto, S. Basal ganglia circuits for reward value-guided behavior. *Annu. Rev. Neurosci.* **37**, 289–306 (2014).
5. Salzman, C. D. & Fusi, S. Emotion, cognition, and mental state representation in amygdala and prefrontal cortex. *Annu. Rev. Neurosci.* **33**, 173–202 (2010).
6. Floresco, S. B. The nucleus accumbens: an interface between cognition, emotion, and action. *Annu. Rev. Psychol.* **66**, 25–52 (2015).
7. Fabbro, F., Aglioti, S. M., Bergamasco, M., Clarici, A. & Panksepp, J. Evolutionary aspects of self- and world consciousness in vertebrates. *Front. Hum. Neurosci.* **9**, 157 (2015).
8. Alexander, G. E., DeLong, M. R. & Strick, P. L. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu. Rev. Neurosci.* **9**, 357–381 (1986).
9. Jahanshahi, M., Obeso, I., Rothwell, J. C. & Obeso, J. A. A fronto–striato–subthalamic–pallidal network for goal-directed and habitual inhibition. *Nat. Rev. Neurosci.* **16**, 719–732 (2015).
10. Shepherd, G. M. Corticostriatal connectivity and its role in disease. *Nat. Rev. Neurosci.* **14**, 278–291 (2013).
11. Stratmann, K. et al. Precortical phase of Alzheimer's disease (AD)-related Tau cytoskeletal pathology. *Brain Pathol.* **26**, 371–386 (2016).
12. Del Tredici, K., Rub, U., De Vos, R. A., Bohl, J. R. & Braak, H. Where does Parkinson disease pathology begin in the brain? *J. Neuropathol. Exp. Neurol.* **61**, 413–426 (2002).
13. Hibar, D. P. et al. Common genetic variants influence human subcortical brain structures. *Nature* **520**, 224–229 (2015).
14. Elliott, L. T. et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210–216 (2018).
15. Renteria, M. E. et al. Genetic architecture of subcortical brain regions: common and region-specific genetic contributions. *Genes Brain Behav.* **13**, 821–830 (2014).
16. Clarke, L. et al. The 1000 Genomes Project: data management and community access. *Nat. Methods* **9**, 459–462 (2012).
17. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
18. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
19. Pruim, R. J. et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).
20. Hibar, D. P. et al. Novel genetic loci associated with hippocampal volume. *Nat. Commun.* **8**, 13624 (2017).
21. Adams, H. H. et al. Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nat. Neurosci.* **19**, 1569–1582 (2016).
22. Verhaaren, B. F. et al. Multiethnic genome-wide association study of cerebral white matter hyperintensities on MRI. *Circ. Cardiovasc. Genet.* **8**, 398–409 (2015).
23. Malik, R. et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537 (2018).
24. Yengo, L. et al. Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
25. Davies, G. et al. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat. Commun.* **9**, 2098 (2018).
26. Kunkle, B. W. et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).
27. Simon-Sanchez, J. et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat. Genet.* **41**, 1308–1312 (2009).
28. Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell* **173**, 1705–1715.e16 (2018).
29. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
30. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
31. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
32. Szklarczyk, D. et al. STRINGv10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).
33. Deans, M. R. et al. Control of neuronal morphology by the atypical cadherin Fat3. *Neuron* **71**, 820–832 (2011).
34. Takahashi, K. et al. Expression of *FOXP2* in the developing monkey forebrain: comparison with the expression of the genes *FOXP1*, *PBX3*, and *MEIS2*. *J. Comp. Neurol.* **509**, 180–189 (2008).
35. Kjaer-Sorensen, K. et al. Pregnancy-associated plasma protein A (PAPP-A) modulates the early developmental rate in zebrafish independently of its proteolytic activity. *J. Biol. Chem.* **288**, 9982–9992 (2013).
36. Bayes-Genis, A. et al. Pregnancy-associated plasma protein A as a marker of acute coronary syndromes. *N. Engl. J. Med.* **345**, 1022–1029 (2001).
37. Funayama, A. et al. Serum pregnancy-associated plasma protein A in patients with heart failure. *J. Card. Fail.* **17**, 819–826 (2011).
38. Desbuquois, B., Carre, N. & Burnol, A. F. Regulation of insulin and type 1 insulin-like growth factor signaling and action by the Grb10/14 and SH2B1/B2 adaptor proteins. *FEBS J.* **280**, 794–816 (2013).

39. Li, J. et al. TXNDC5 contributes to rheumatoid arthritis by down-regulating IGFBP1 expression. *Clin. Exp. Immunol.* **192**, 82–94 (2018).
40. Matulka, K. et al. PTP1B is an effector of activin signaling and regulates neural specification of embryonic stem cells. *Cell Stem Cell* **13**, 706–719 (2013).
41. Krishnan, N. et al. PTP1B inhibition suggests a therapeutic strategy for Rett syndrome. *J. Clin. Invest.* **125**, 3163–3177 (2015).
42. Sebastian-Serrano, A. et al. Tissue-nonspecific alkaline phosphatase regulates purinergic transmission in the central nervous system during development and disease. *Comput. Struct. Biotechnol. J.* **13**, 95–100 (2015).
43. Diaz-Hernandez, M. et al. Tissue-nonspecific alkaline phosphatase promotes the neurotoxicity effect of extracellular tau. *J. Biol. Chem.* **285**, 32539–32548 (2010).
44. Vardy, E. R., Kellett, K. A., Cocklin, S. L. & Hooper, N. M. Alkaline phosphatase is increased in both brain and plasma in Alzheimer's disease. *Neurodegener. Dis.* **9**, 31–37 (2012).
45. Kellett, K. A., Williams, J., Vardy, E. R., Smith, A. D. & Hooper, N. M. Plasma alkaline phosphatase is elevated in Alzheimer's disease and inversely correlates with cognitive function. *Int. J. Mol. Epidemiol. Genet.* **2**, 114–121 (2011).
46. Searles Quick, V. B., Davis, J. M., Olincy, A. & Sikela, J. M. DUF1220 copy number is associated with schizophrenia risk and severity: implications for understanding autism and schizophrenia as related diseases. *Transl. Psychiatry* **5**, e697 (2015).
47. Hsu, S. C. et al. Mutations in *SLC20A2* are a major cause of familial idiopathic basal ganglia calcification. *Neurogenetics* **14**, 11–22 (2013).
48. Taglia, I., Bonifati, V., Mignarri, A., Dotti, M. T. & Federico, A. Primary familial brain calcification: update on molecular genetics. *Neurol. Sci.* **36**, 787–794 (2015).
49. Figueiro-Silva, J. et al. Neuronal pentraxin 1 negatively regulates excitatory synapse density and synaptic plasticity. *J. Neurosci.* **35**, 5504–5521 (2015).
50. Abad, M. A., Enguita, M., DeGregorio-Rocasolano, N., Ferrer, I. & Trullas, R. Neuronal pentraxin 1 contributes to the neuronal damage evoked by amyloid-β and is overexpressed in dystrophic neurites in Alzheimer's brain. *J. Neurosci.* **26**, 12735–12747 (2006).
51. Tobaben, S., Varoqueaux, F., Brose, N., Stahl, B. & Meyer, G. A brain-specific isoform of small glutamine-rich tetratricopeptide repeat-containing protein binds to Hsc70 and the cysteine string protein. *J. Biol. Chem.* **278**, 38376–38383 (2003).
52. Fonte, V. et al. Interaction of intracellular β amyloid peptide with chaperone proteins. *Proc. Natl Acad. Sci. USA* **99**, 9439–9444 (2002).
53. Mao, C. X. et al. Microtubule-severing protein katanin regulates neuromuscular junction development and dendritic elaboration in *Drosophila*. *Development* **141**, 1064–1074 (2014).
54. Yu, W. et al. The microtubule-severing proteins spastin and katanin participate differently in the formation of axonal branches. *Mol. Biol. Cell* **19**, 1485–1498 (2008).
55. Zhu, J., Shang, Y. & Zhang, M. Mechanistic basis of MAGUK-organized complexes in synaptic development and signalling. *Nat. Rev. Neurosci.* **17**, 209–223 (2016).
56. Ingason, A. et al. Expression analysis in a rat psychosis model identifies novel candidate genes validated in a large case-control sample of schizophrenia. *Transl. Psychiatry* **5**, e656 (2015).
57. Nithianantharajah, J. et al. Synaptic scaffold evolution generated components of vertebrate cognitive complexity. *Nat. Neurosci.* **16**, 16–24 (2013).
58. Nalls, M. A. et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet.* **46**, 989–993 (2014).
59. Guan, J. J. et al. DRAM1 regulates apoptosis through increasing protein levels and lysosomal localization of BAX. *Cell Death Dis.* **6**, e1624 (2015).
60. Yu, M., Jiang, Y., Feng, Q., Ouyang, Y. & Gan, J. DRAM1 protects neuroblastoma cells from oxygen-glucose deprivation/reperfusion-induced injury via autophagy. *Int. J. Mol. Sci.* **15**, 19253–19264 (2014).
61. Scarpa, J. R. et al. Systems genetic analyses highlight a TGFβ-FOXO3 dependent striatal astrocyte network conserved across species and associated with stress, sleep, and Huntington's disease. *PLoS Genet.* **12**, e1006137 (2016).
62. Donlon, T. A. et al. FOXO3 longevity interactome on chromosome 6. *Aging Cell* **16**, 1016–1025 (2017).
63. Sears, J. C. & Broihier, H. T. FoxO regulates microtubule dynamics and polarity to promote dendrite branching in *Drosophila* sensory neurons. *Dev. Biol.* **418**, 40–54 (2016).
64. Peng, K. et al. Knockdown of FoxO3a induces increased neuronal apoptosis during embryonic development in zebrafish. *Neurosci. Lett.* **484**, 98–103 (2010).
65. Santama, N., Er, C. P., Ong, L. L. & Yu, H. Distribution and functions of kinectin isoforms. *J. Cell Sci.* **117**, 4537–4549 (2004).
66. Liu, X. A., Rizzo, V. & Puthanveettil, S. V. Pathologies of axonal transport in neurodegenerative diseases. *Transl. Neurosci.* **3**, 355–372 (2012).
67. Consortium, E. et al. Genome-wide association analysis of genetic generalized epilepsies implicates susceptibility loci at 1q43, 2p16.1, 2q22.3 and 17q21.32. *Hum. Mol. Genet.* **21**, 5359–5372 (2012).
68. Martins-de-Souza, D. et al. Proteomic analysis identifies dysfunction in cellular transport, energy, and protein metabolism in different brain regions of atypical frontotemporal lobar degeneration. *J. Proteome Res.* **11**, 2533–2543 (2012).
69. Shulman, J. M. et al. Functional screening in *Drosophila* identifies Alzheimer's disease susceptibility genes and implicates Tau-mediated mechanisms. *Hum. Mol. Genet.* **23**, 870–877 (2014).
70. Friede, R. L. & Samorajski, T. Axon caliber related to neurofilaments and microtubules in sciatic nerve fibers of rats and mice. *Anat. Rec.* **167**, 379–387 (1970).
71. Yuan, A., Rao, M. V., Veeranna & Nixon, R. A. Neurofilaments and neurofilament proteins in health and disease. *Cold Spring Harb. Perspect. Biol.* **9**, a018309 (2017).
72. Bis, J. C. et al. Whole exome sequencing study identifies novel rare and common Alzheimer's-associated variants involved in immune response and transcriptional regulation. *Mol Psychiatry* https://doi.org/10.1038/s41380-018-0112-7 (2018).
73. Marioni, R. E. et al. GWAS on family history of Alzheimer's disease. *Transl. Psychiatry* **8**, 99 (2018).

Acknowledgements

We thank all of the study participants for contributing to this research. Full acknowledgements and grant support details are provided in the Supplementary Note.

Author contributions

C.L.S. drafted the manuscript with contributions from H.H.H.A., D.P.H., C.C.W., T.V.L., A.A.-V., S.Ehrlich., A.K.H., M.W.V., D.J., T.G.M.v.E., C.D.W., M.J.W., S.E.F., K.A.M., P.J.H., B.F., H.J.G., A.D.J., O.L.L., S.Debette, S.E.M., J.M.S., P.M.T., S.S. and M.A.I. M.S., N.J., L.R.Y., T.V.L., G.C., L.A., M.E.R., A.d.B., I.K., M.A., S.A., S.E., R.R.-S., A.K.H., H.J.J., A.Stevens., J.B., M.W.V., A.V.W., K.W., N.A., S.H., A.L.G., P.H.L., S.G., S.L.H., D.K., L.Schmaal, S.M.L., I.A., E.W., D.T.-G., J.C.I., L.N.V., R.B., F.C., D.J., O.C., U.K.H., B.S.A., C.-Y.C., A.A.A., M.P.B., A.F.M., S.K.M., P.A., A.J.Schork., D.C.M.L., T.Y.W., L.Shen, P.G.S., E.J.C.d.G., M.T., K.R.v.E., N.J.A.v.d.W., A.M.M., J.S.R., N.R., W.H., M.C.V.H., J.B.J.K., L.M.O.L., A.Hofman, G.H., M.E.B., S.R., J.-J.H., A.Simmons, N.H., P.R.S., T.W.M., P.Maillard, O.Gruber, N.A.G., J.E.S., H.Lemaître, B.M.-M., D.v.R., I.J.D., R.M.B., I.M., R.K., H.v.B., M.J.W., D.v.'t.E., M.M.N., S.E.F., A.S.B., K.A.M., N.R.-S., D.J.H., H.J.G., C.M.v.D., J.M.W., C.DeCarli, P.L.D.J. and V.G. contributed to the preparation of data. C.L.S., H.H.H.A., D.P.H., M.J.K., J.L.S., M.S., M.Sargurupremraj, N.J., G.V.R., A.V.S., J.C.B., X.J., M.Luciano., E.H., A.Teumer, S.J.v.d.L., J.Y., L.R.Y., S.L., K.J.Y., G.C., M.E.R., N.J.A., H.J.J., A.V.W., S.H., N.M.S., S.G., D.T.G., J.S., C.-Y.C., L.M.O.L., Q.Y., A.Thalamuthu, I.O.F., D.v.'t.E., C.Depondt and P.L.D.J. performed the statistical analyses. C.L.S., H.H.H.A., C.C.W., M.J.K., T.V.L., S.L., Y.H., K.J.Y., J.D.E., Q.Y. and A.D.J. carried out the downstream analyses. All authors reviewed the manuscript for intellectual content.

Competing interests

D.P.H. is currently an employee at Genentech. D.J. has received travel and speaker's honoraria from Janssen–Cilag, as well as research funding from DFG. R.L.B. is a consultant for Pfizer and Roche. P.A. is a scientific adviser for Genoscreen. T.Y.W. is a consultant and advisory board member for Allergan, Bayer, Boehringer–Ingelheim, Genentech, Merck, Novartis, Oxurion (formerly ThromboGenics) and Roche, and is a co-founder of Plano and EyRiS. A.M.M. has received grant support from Eli Lilly, Janssen, Pfizer and the Sackler Trust. B.M.P. serves on the steering committee of the Yale Open Data Access Project funded by Johnson & Johnson. A.M.-L. is a member of the advisory board for the Lundbeck International Neuroscience Foundation and Brainsway, a member of the editorial board for the American Association for the Advancement of Science and Elsevier, a faculty member of the Lundbeck International Neuroscience Foundation and a consultant for Boehringer Ingelheim. W.J.N. is the founder and scientific lead of Quantib BV, in addition to being a shareholder. M.M.N. is a shareholder of Life & Brain, receives a salary from Life & Brain, has received support from Shire for attending conferences and has received financial remuneration from the Lundbeck Foundation, Robert Bosch Foundation and Deutsches Ärzteblatt for participation in scientific advisory boards. B.F. has received educational speaking fees from Shire and Medice. H.J.G. has received travel grants and speaker's honoraria from Fresenius Medical Care, Neuraxpharm and Janssen–Cilag, as well as research funding from Fresenius Medical Care.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0511-y.

Correspondence and requests for materials should be addressed to C.L.S. or M.A.I.

Reprints and permissions information is available at www.nature.com/reprints.

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

Claudia L. Satizabal[1,2,3,4,266*], Hieab H. H. Adams[5,6,7,266], Derrek P. Hibar[8,266], Charles C. White[9,10,266], Maria J. Knol[5], Jason L. Stein[8,11,12], Markus Scholz[13,14], Muralidharan Sargurupremraj[15], Neda Jahanshad[8], Gennady V. Roshchupkin[5,6,16], Albert V. Smith[17,18,19], Joshua C. Bis[20], Xueqiu Jian[21], Michelle Luciano[22], Edith Hofer[23,24], Alexander Teumer[25], Sven J. van der Lee[5], Jingyun Yang[26,27], Lisa R. Yanek[28], Tom V. Lee[29], Shuo Li[30], Yanhui Hu[31], Jia Yu Koh[32], John D. Eicher[33], Sylvane Desrivières[34], Alejandro Arias-Vasquez[35,36,37,38], Ganesh Chauhan[15,39], Lavinia Athanasiu[40,41], Miguel E. Rentería[42], Sungeun Kim[43,44,45], David Hoehn[46], Nicola J. Armstrong[47], Qiang Chen[48], Avram J. Holmes[49,50], Anouk den Braber[51,52,53,54], Iwona Kloszewska[55], Micael Andersson[56,57], Thomas Espeseth[40,58], Oliver Grimm[59], Lucija Abramovic[60], Saud Alhusaini[61,62], Yuri Milaneschi[63], Martina Papmeyer[64,65], Tomas Axelsson[66], Stefan Ehrlich[50,67,68], Roberto Roiz-Santiañez[69,70,71], Bernd Kraemer[72], Asta K. Håberg[73,74], Hannah J. Jones[75,76,77], G. Bruce Pike[78,79], Dan J. Stein[80,81], Allison Stevens[68], Janita Bralten[36,38], Meike W. Vernooij[5,6], Tamara B. Harris[82], Irina Filippi[83], A. Veronica Witte[84,85], Tulio Guadalupe[86,87], Katharina Wittfeld[88,89], Thomas H. Mosley[90], James T. Becker[91,92,93], Nhat Trung Doan[41], Saskia P. Hagenaars[22], Yasaman Saba[94], Gabriel Cuellar-Partida[95], Najaf Amin[5], Saima Hilal[96,97], Kwangsik Nho[43,44,45], Nazanin Mirza-Schreiber[46,98], Konstantinos Arfanakis[26,99,100], Diane M. Becker[28], David Ames[101,102], Aaron L. Goldman[48], Phil H. Lee[50,103,104,105,106], Dorret I. Boomsma[51,52,53,107], Simon Lovestone[108,109], Sudheer Giddaluru[110,111], Stephanie Le Hellard[110,111], Manuel Mattheisen[112,113,114,115,116], Marc M. Bohlken[60], Dalia Kasperaviciute[117,118], Lianne Schmaal[119,120], Stephen M. Lawrie[64], Ingrid Agartz[41,115,121], Esther Walton[67,122], Diana Tordesillas-Gutierrez[71,123], Gareth E. Davies[124], Jean Shin[125], Jonathan C. Ipser[80], Louis N. Vinke[126], Martine Hoogman[36,38], Tianye Jia[34], Ralph Burkhardt[14,127], Marieke Klein[36,38], Fabrice Crivello[128], Deborah Janowitz[88], Owen Carmichael[129], Unn K. Haukvik[40,130], Benjamin S. Aribisala[131,132], Helena Schmidt[94], Lachlan T. Strike[95,133], Ching-Yu Cheng[32,134], Shannon L. Risacher[44,45], Benno Pütz[46], Debra A. Fleischman[26,27,135], Amelia A. Assareh[136], Venkata S. Mattay[48,137,138], Randy L. Buckner[50,139], Patrizia Mecocci[140], Anders M. Dale[141,142,143,144,145], Sven Cichon[146,147,148], Marco P. Boks[60], Mar Matarin[117,149,150], Brenda W. J. H. Penninx[63], Vince D. Calhoun[151,152,153], M. Mallar Chakravarty[154,155], Andre F. Marquand[38,156], Christine Macare[34], Shahrzad Kharabian Masouleh[84,157], Jaap Oosterlaan[158,159], Philippe Amouyel[160,161,162,163], Katrin Hegenscheid[164], Jerome I. Rotter[165], Andrew J. Schork[166,167], David C. M. Liewald[22], Greig I. de Zubicaray[168,169], Tien Yin Wong[32,170], Li Shen[171], Philipp G. Sämann[46], Henry Brodaty[136,172], Joshua L. Roffman[50], Eco J. C. de Geus[51,52,53,107], Magda Tsolaki[173], Susanne Erk[174], Kristel R. van Eijk[175], Gianpiero L. Cavalleri[176], Nic J. A. van der Wee[177,178], Andrew M. McIntosh[22,64], Randy L. Gollub[50,68,103], Kazima B. Bulayeva[179], Manon Bernard[125], Jennifer S. Richards[35,38,180], Jayandra J. Himali[3,4,30], Markus Loeffler[13,14], Nanda Rommelse[37,38,181], Wolfgang Hoffmann[89,182], Lars T. Westlye[40,41], Maria C. Valdés Hernández[131,183], Narelle K. Hansell[95,133], Theo G. M. van Erp[184,185], Christiane Wolf[186], John B. J. Kwok[187,188,189], Bruno Vellas[190,191], Andreas Heinz[192], Loes M. Olde Loohuis[193], Norman Delanty[61,194], Beng-Choon Ho[195], Christopher R. K. Ching[8,196], Elena Shumskaya[36,38,156], Baljeet Singh[197], Albert Hofman[5,198], Dennis van der Meer[40,41,199], Georg Homuth[200], Bruce M. Psaty[20,201,202,203], Mark E. Bastin[131,183], Grant W. Montgomery[204], Tatiana M. Foroud[45,205], Simone Reppermund[136,206], Jouke-Jan Hottenga[51,52,53,107], Andrew Simmons[207,208,209], Andreas Meyer-Lindenberg[59], Wiepke Cahn[60], Christopher D. Whelan[8,61], Marjolein M. J. van Donkelaar[36,38], Qiong Yang[30], Norbert Hosten[164], Robert C Green[103,210], Anbupalam Thalamuthu[136], Sebastian Mohnke[174],

Hilleke E. Hulshoff Pol[60], Honghuang Lin[3,211], Clifford R. Jack Jr[212], Peter R. Schofield[188,213], Thomas W. Mühleisen[148,214,215], Pauline Maillard[197], Steven G. Potkin[216], Wei Wen[136], Evan Fletcher[197], Arthur W. Toga[217], Oliver Gruber[72], Matthew Huentelman[167], George Davey Smith[76], Lenore J. Launer[82], Lars Nyberg[56,57,218], Erik G. Jönsson[41,115], Benedicto Crespo-Facorro[70,71], Nastassja Koen[80,81], Douglas N. Greve[68,219], André G. Uitterlinden[5,220], Daniel R. Weinberger[48,137,221,222,223], Vidar M. Steen[110,111], Iryna O. Fedko[51,52,107], Nynke A. Groenewold[80], Wiro J. Niessen[6,16,224], Roberto Toro[225], Christophe Tzourio[226], William T. Longstreth Jr[202,227], M. Kamran Ikram[5,228], Jordan W. Smoller[50,103,105,106], Marie-Jose van Tol[229], Jessika E. Sussmann[64], Tomas Paus[230,231,232], Hervé Lemaître[83], Matthias L. Schroeter[14,84,233], Bernard Mazoyer[128], Ole A. Andreassen[40,41], Florian Holsboer[46,234], Chantal Depondt[235], Dick J. Veltman[63], Jessica A. Turner[152,153,236], Zdenka Pausova[125], Gunter Schumann[34], Daan van Rooij[35,38,180], Srdjan Djurovic[110,237], Ian J. Deary[22], Katie L. McMahon[168,169], Bertram Müller-Myhsok[46,238,239], Rachel M. Brouwer[60], Hilkka Soininen[240,241], Massimo Pandolfo[235], Thomas H. Wassink[195], Joshua W. Cheung[8], Thomas Wolfers[36,38], Jean-Luc Martinot[83], Marcel P. Zwiers[38,156], Matthias Nauck[242,243], Ingrid Melle[40,41], Nicholas G. Martin[95], Ryota Kanai[244,245,246], Eric Westman[247], René S. Kahn[60,248], Sanjay M. Sisodiya[117,118], Tonya White[6,249], Arvin Saremi[8], Hans van Bokhoven[36,38], Han G. Brunner[36,38,250,251], Henry Völzke[25,243], Margaret J. Wright[133,252], Dennis van 't Ent[51,52,53,107], Markus M. Nöthen[147,253], Roel A. Ophoff[193,254], Jan K. Buitelaar[35,38,181], Guillén Fernández[35,38], Perminder S. Sachdev[136,255], Marcella Rietschel[59], Neeltje E. M. van Haren[60,249], Simon E. Fisher[38,87], Alexa S. Beiser[3,4,30], Clyde Francks[38,87], Andrew J. Saykin[44,45,205], Karen A. Mather[136,188], Nina Romanczuk-Seiferth[192], Catharina A. Hartman[180], Anita L. DeStefano[3,30], Dirk J. Heslenfeld[256], Michael W. Weiner[257,258], Henrik Walter[174], Pieter J. Hoekstra[180], Paul A. Nyquist[28], Barbara Franke[36,37,38], David A. Bennett[26,27], Hans J. Grabe[88,89], Andrew D. Johnson[33], Christopher Chen[96,97], Cornelia M. van Duijn[5,259], Oscar L. Lopez[92,93], Myriam Fornage[21,260], Joanna M. Wardlaw[22,183,261], Reinhold Schmidt[23], Charles DeCarli[262], Philip L. De Jager[9,10], Arno Villringer[84,85], Stéphanie Debette[4,15,226], Vilmundur Gudnason[18,19], Sarah E. Medland[95,267], Joshua M. Shulman[29,263,264,265,267], Paul M. Thompson[8,267], Sudha Seshadri[3,4,267] and M. Arfan Ikram[5,6,267]*

[1]Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, UT Health San Antonio, San Antonio, TX, USA. [2]Department of Population Health Sciences, UT Health San Antonio, San Antonio, TX, USA. [3]The Framingham Heart Study, Framingham, MA, USA. [4]Department of Neurology, Boston University School of Medicine, Boston, MA, USA. [5]Department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands. [6]Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, the Netherlands. [7]Department of Clinical Genetics, Erasmus MC, Rotterdam, the Netherlands. [8]Imaging Genetics Center, USC Mark and Mary Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. [9]Cell Circuits Program, Broad Institute, Cambridge, MA, USA. [10]Center for Translational and Computational Neuroimmunology, Department of Neurology, Columbia University Medical Center, New York, NY, USA. [11]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. [12]UNC Neuroscience Center, University of North Carolina, Chapel Hill, NC, USA. [13]Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany. [14]LIFE: The Leipzig Research Center for Civilization Diseases, University of Leipzig, Leipzig, Germany. [15]University of Bordeaux, Inserm, Bordeaux Population Health Research Center, Team VINTAGE, UMR 1219, Bordeaux, France. [16]Department of Medical Informatics, Erasmus MC, Rotterdam, the Netherlands. [17]Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA. [18]Faculty of Medicine, University of Iceland, Reykjavík, Iceland. [19]Icelandic Heart Association, Kopavogur, Iceland. [20]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA. [21]The Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX, USA. [22]Department of Psychology, Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK. [23]Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, Graz, Austria. [24]Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria. [25]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [26]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA. [27]Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, USA. [28]GeneSTAR Research Program, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [29]Department of Neurology, Baylor College of Medicine, Houston, TX, USA. [30]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA. [31]Department of Genetics, Harvard Medical School, Boston, MA, USA. [32]Singapore Eye Research Institute, Singapore National Eye

Centre, Singapore, Singapore. [33]Division of Intramural Research, Population Sciences Branch, National Heart, Lung and Blood Institute, Framingham, MA, USA. [34]MRC-SGDP Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. [35]Department of Cognitive Neuroscience, Radboud University Medical Center, Nijmegen, the Netherlands. [36]Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands. [37]Department of Psychiatry, Radboud University Medical Center, Nijmegen, the Netherlands. [38]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands. [39]Centre for Brain Research, Indian Institute of Science, Bangalore, India. [40]CoE NORMENT, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. [41]CoE NORMENT, Institute of Clinical Medicine, University of Oslo, Oslo, Norway. [42]Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [43]Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN, USA. [44]Center for Neuroimaging, Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN, USA. [45]Indiana Alzheimer Disease Center, Indiana University School of Medicine, Indianapolis, IN, USA. [46]Max Planck Institute of Psychiatry, Munich, Germany. [47]Mathematics and Statistics, Murdoch University, Perth, Western Australia, Australia. [48]Lieber Institute for Brain Development, Baltimore, MD, USA. [49]Department of Psychology, Yale University, New Haven, CT, USA. [50]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [51]Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [52]Netherlands Twin Register, Vrije Universiteit, Amsterdam, the Netherlands. [53]Amsterdam Neuroscience, Amsterdam, the Netherlands. [54]Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, VU Amsterdam, Amsterdam UMC, Amsterdam, the Netherlands. [55]Medical University of Lodz, Lodz, Poland. [56]Department of Integrative Medical Biology, Umeå University, Umeå, Sweden. [57]Umeå Centre for Functional Brain Imaging (UFBI), Umeå University, Umeå, Sweden. [58]Department of Psychology, University of Oslo, Oslo, Norway. [59]Central Institute of Mental Health, Medical Faculty Mannheim, University Heidelberg, Mannheim, Germany. [60]Department of Psychiatry, UMC Brain Center, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands. [61]The Royal College of Surgeons in Ireland, Dublin, Ireland. [62]Department of Neurology, Yale School of Medicine, New Haven, CT, USA. [63]Department of Psychiatry, Amsterdam Neuroscience, VU University Medical Center, Amsterdam, the Netherlands. [64]Division of Psychiatry, Royal Edinburgh Hospital, University of Edinburgh, Edinburgh, UK. [65]Division of Systems Neuroscience of Psychopathology, Translational Research Center, University Hospital of Psychiatry, University of Bern, Bern, Switzerland. [66]Department of Medical Sciences, Molecular Medicine and Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [67]Division of Psychological and Social Medicine and Developmental Neurosciences, Faculty of Medicine, TU Dresden, Dresden, Germany. [68]Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA. [69]Department of Psychiatry, University Hospital Marqués de Valdecilla, School of Medicine, University of Cantabria-IDIVAL, Santander, Spain. [70]Department of Medicine, University Hospital Marqués de Valdecilla, School of Medicine, University of Cantabria-IDIVAL, Santander, Spain. [71]Centro Investigación Biomédica en Red Salud Mental, Santander, Spain. [72]Section for Experimental Psychopathology and Neuroimaging, Department of General Psychiatry, Heidelberg University, Heidelberg, Germany. [73]Department of Neuroscience, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway. [74]Department of Radiology, St. Olav's Hospital, Trondheim University Hospital, Trondheim, Norway. [75]Centre for Academic Mental Health, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. [76]MRC Integrative Epidemiology Unit, Bristol Medical School, University of Bristol, Bristol, UK. [77]NIHR Bristol Biomedical Research Centre, University Hospitals Bristol NHS Foundation Trust and University of Bristol, Bristol, UK. [78]Department of Radiology, University of Calgary, Calgary, Alberta, Canada. [79]Department of Clinical Neurosciences, University of Calgary, Calgary, Alberta, Canada. [80]Department of Psychiatry and Mental Health, University of Cape Town, Cape Town, South Africa. [81]South African Medical Research Council Unit on Risk and Resilience in Mental Disorders, Cape Town, South Africa. [82]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, National Institutes of Health, Bethesda, MD, USA. [83]INSERM, Research Unit 1000 'Neuroimaging and Psychiatry', Paris Saclay University and Paris Descartes University—DIGITEO Labs, Gif sur Yvette, France. [84]Department of Neurology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany. [85]Faculty of Medicine, CRC 1052 'Obesity Mechanisms', University of Leipzig, Leipzig, Germany. [86]International Max Planck Research School for Language Sciences, Nijmegen, the Netherlands. [87]Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands. [88]Department of Psychiatry, University Medicine Greifswald, Greifswald, Germany. [89]German Center for Neurodegenerative Diseases, Greifswald, Germany. [90]Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA. [91]Department of Psychology, University of Pittsburgh, Pittsburgh, PA, USA. [92]Department of Neurology, University of Pittsburgh, Pittsburgh, PA, USA. [93]Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA. [94]Research Unit-Genetic Epidemiology, Gottfried Schatz Research Centre for Cell Signaling, Metabolism and Aging, Molecular Biology and Biochemistry, Medical University of Graz, Graz, Austria. [95]QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [96]Department of Pharmacology, National University of Singapore, Singapore, Singapore. [97]Memory Aging and Cognition Center, National University Health System, Singapore, Singapore. [98]Institute of Neurogenomics, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg, Germany. [99]Department of Biomedical Engineering, Illinois Institute of Technology, Chicago, IL, USA. [100]Department of Diagnostic Radiology and Nuclear Medicine, Rush University Medical Center, Chicago, IL, USA. [101]Academic Unit for Psychiatry of Old Age, University of Melbourne, Melbourne, Victoria, Australia. [102]National Ageing Research Institute, Royal Melbourne Hospital, Melbourne, Victoria, Australia. [103]Harvard Medical School, Boston, MA, USA. [104]Lurie Center for Autism, Massachusetts General Hospital, Harvard Medical School, Lexington, MA, USA. [105]Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. [106]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Boston, MA, USA. [107]Amsterdam Public Health Research Institute, VU Medical Center, Amsterdam, the Netherlands. [108]Department of Psychiatry, University of Oxford, Oxford, UK. [109]NIHR Dementia Biomedical Research Unit, King's College London, London, UK. [110]NORMENT, Department of Clinical Science, University of Bergen, Bergen, Norway. [111]Dr Einar Martens Research Group for Biological Psychiatry, Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway. [112]Centre for integrated Sequencing, Aarhus University, Aarhus, Denmark. [113]Department of Biomedicine, Aarhus University, Aarhus, Denmark. [114]The Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH), Aarhus, Denmark. [115]Department of Clinical Neuroscience, Centre for Psychiatric Research, Karolinska Institutet, Stockholm, Sweden. [116]Stockholm Health Care Services, Stockholm County Council, Stockholm, Sweden. [117]UCL Queen Square Institute of Neurology, London, UK. [118]Chalfont Centre for Epilepsy, Bucks, UK. [119]Centre for Youth Mental Health, The University of Melbourne, Melbourne, Victoria, Australia. [120]Orygen, The National Centre of Excellence in Youth Mental Health, Melbourne, Victoria, Australia. [121]Department of Research and Development, Diakonhjemmet Hospital, Oslo, Norway. [122]Department of Psychology, University of Bath, Bath, UK. [123]Neuroimaging Unit, Technological Facilities, Valdecilla Biomedical Research Institute IDIVAL, Santander, Spain. [124]Avera Institute for Human Genetics, Sioux Falls, SD, USA. [125]Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada. [126]Center for Systems Neuroscience, Boston University, Boston, MA, USA. [127]Institute of Clinical Chemistry and Laboratory Medicine, University Hospital Regensburg, Regensburg, Germany. [128]Neurodegeneratives Diseases Institute, CNRS UMR 5293, Université de Bordeaux, Bordeaux, France. [129]Pennington Biomedical Research Center, Baton Rouge, LA, USA. [130]Department of Adult Psychiatry, Institute for Clinical Medicine, University of Oslo, Oslo, Norway. [131]Brain Research Imaging Centre, University of Edinburgh, Edinburgh, UK. [132]Department of Computer Science, Lagos State University, Ojo, Nigeria. [133]Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia. [134]Ophthalmology and Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore, Singapore. [135]Department of Behavioral Sciences, Rush University Medical Center, Chicago, IL, USA. [136]Centre for Healthy Brain Ageing, School of Psychiatry, University of New South Wales, Sydney, New South Wales, Australia. [137]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [138]Department of Radiology, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [139]Department of

Psychology, Center for Brain Science, Harvard University, Cambridge, MA, USA. [140]Section of Gerontology and Geriatrics, Department of Medicine, University of Perugia, Perugia, Italy. [141]Center for Multimodal Imaging and Genetics, University of California, San Diego, San Diego, CA, USA. [142]Department of Cognitive Sciences, University of California, San Diego, San Diego, CA, USA. [143]Department of Neurosciences, University of California. San Diego, San Diego, CA, USA. [144]Department of Psychiatry, University of California, San Diego, San Diego, CA, USA. [145]Department of Radiology, University of California, San Diego, San Diego, CA, USA. [146]Division of Medical Genetics, Department of Biomedicine, University of Basel, Basel, Switzerland. [147]Institute of Human Genetics, University of Bonn, Bonn, Germany. [148]Institute for Neuroscience and Medicine: Structural and Functional Organisation of the Brain (INM-1), Research Centre Jülich, Jülich, Germany. [149]Reta Lila Weston Institute, UCL Institute of Neurology, London, UK. [150]Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK. [151]Department of ECE, University of New Mexico, Albuquerque, NM, USA. [152]The Mind Research Network and LBERI, Albuquerque, NM, USA. [153]Tri-institutional Center for Translational Research in Neuroimaging and Data Science, Georgia State University, Atlanta, GA, USA. [154]Cerebral Imaging Centre, Douglas Mental Health University Institute, Montreal, Québec, Canada. [155]Departments of Psychiatry and Biological and Biomedical Engineering, McGill University, Montreal, Québec, Canada. [156]Donders Centre for Cognitive Neuroimaging, Radboud University, Nijmegen, the Netherlands. [157]Institute for Neuroscience and Medicine: Brain and Behaviour (INM-7), Research Centre Jülich, Jülich, Germany. [158]Clinical Neuropsychology Section, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [159]Emma Neuroscience Group, Department of Pediatrics, Emma Children's Hospital, Amsterdam Reproduction & Development, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands. [160]LabEx DISTALZ—U1167, RID-AGE-Risk Factors and Molecular Determinants of Aging-Related Diseases, University of Lille, Lille, France. [161]Inserm U1167, Lille, France. [162]Centre Hospitalier Universitaire Lille, Lille, France. [163]Institut Pasteur de Lille, Lille, France. [164]Institute of Diagnostic Radiology and Neuroradiology, University Medicine Greifswald, Greifswald, Germany. [165]Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute and Pediatrics at Harbor-UCLA Medical Center, Torrance, CA, USA. [166]Institute of Biological Psychiatry, Mental Health Center Sct. Hans, Roskilde, Denmark. [167]Neurogenomics Division, Translational Genomics Research Institute, Phoenix, AZ, USA. [168]Faculty of Health, Queensland University of Technology, Brisbane, Queensland, Australia. [169]Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia. [170]Academic Medicine Research Institute, Duke-NUS Medical School, Singapore, Singapore. [171]Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA, USA. [172]Dementia Centre for Research Collaboration, UNSW, Sydney, New South Wales, Australia. [173]1st Department of Neurology, AHEPA University Hospital, Aristotle University of Thessaloniki, Thessaloniki, Greece. [174]Division of Mind and Brain Research, D, Corporate member of Freie Universität Berliepartment of Psychiatry and Psychotherapy, Charité—Universitätsmedizin Berlin, Humboldt-Universität zu Berlin and Berlin Institute of Health, Berlin, Germany. [175]Brain Center Rudolf Magnus, Human Neurogenetics Unit, UMC Utrecht, Utrecht, the Netherlands. [176]Department of Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin, Ireland. [177]Department of Psychiatry, Leiden University Medical Center, Leiden, the Netherlands. [178]Leiden Institute for Brain and Cognition, Leiden University Medical Center, Leiden, the Netherlands. [179]Department of Evolution and Genetics, Dagestan State University, Makhachkala, Russia. [180]University Medical Center Groningen, Department of Psychiatry, University of Groningen, Groningen, the Netherlands. [181]Karakter Child and Adolescent Psychiatry University Center, Nijmegen, the Netherlands. [182]Section Epidemiology of Health Care and Community Health, Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [183]Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK. [184]Clinical Translational Neuroscience Laboratory, Department of Psychiatry and Human Behavior, University of California, Irvine, Irvine, CA, USA. [185]Center for the Neurobiology of Learning and Memory, University of California, Irvine, Irvine, CA, USA. [186]Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany. [187]Brain and Mind Centre, University of Sydney, Sydney, New South Wales, Australia. [188]Neuroscience Research Australia, Sydney, New South Wales, Australia. [189]University of New South Wales, Sydney, New South Wales, Australia. [190]Department of Internal Medicine, INSERM U 1027, University of Toulouse, Toulouse, France. [191]Department of Geriatric Medicine, INSERM U 1027, University of Toulouse, Toulouse, France. [192]Department of Psychiatry and Psychotherapy, Charité Universitätsmedizin Berlin, Berlin, Germany. [193]Center for Neurobehavioral Genetics, University of California, Los Angeles, Los Angeles, CA, USA. [194]Neurology Division, Beaumont Hospital, Dublin, Ireland. [195]Department of Psychiatry, Carver College of Medicine, University of Iowa, Iowa City, IA, USA. [196]Interdepartmental Neuroscience Graduate Program, UCLA School of Medicine, Los Angeles, CA, USA. [197]Imaging of Dementia and Aging Laboratory, Department of Neurology, University of California, Davis, Davis, CA, USA. [198]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [199]School of Mental Health and Neuroscience, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, the Netherlands. [200]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany. [201]Kaiser Permanent Washington Health Research Institute, Seattle, WA, USA. [202]Department of Epidemiology, University of Washington, Seattle, WA, USA. [203]Department of Health Services, University of Washington, Seattle, WA, USA. [204]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia. [205]Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA. [206]Department of Developmental Disability Neuropsychiatry, School of Psychiatry, UNSW Medicine, Sydney, New South Wales, Australia. [207]Biomedical Research Unit for Dementia, King's College London, London, UK. [208]Department of Neuroimaging, Institute of Psychiatry, King's College London, London, UK. [209]Division of Clinical Geriatrics, Department of Neurobiology, Care Sciences and Society, Karolinska Institute, Stockholm, Sweden. [210]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. [211]Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA. [212]Department of Radiology, Mayo Clinic, Rochester, MN, USA. [213]School of Medical Sciences, UNSW, Sydney, New South Wales, Australia. [214]Department of Biomedicine, University of Basel, Basel, Switzerland. [215]Cécile and Oskar Vogt Institute for Brain Research, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. [216]Department of Psychiatry and Human Behavior, University of California, Irvine, Irvine, CA, USA. [217]Laboratory of Neuro Imaging, USC Mark and Mary Stevens Neuroimaging and Informatics Institute, Keck School of Medicine of the University of Southern California, Los Angeles, CA, USA. [218]Radiation Sciences, Umeå University, Umeå, Sweden. [219]Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. [220]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, the Netherlands. [221]Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [222]Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [223]Department of Psychiatry, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [224]Imaging Physics, Faculty of Applied Sciences, Delft University of Technology, Delft, the Netherlands. [225]Institut Pasteur, Paris, France. [226]Department of Neurology, CHU de Bordeaux, Bordeaux, France. [227]Department of Neurology, University of Washington, Seattle, WA, USA. [228]Department of Neurology, Erasmus MC, Rotterdam, the Netherlands. [229]Cognitive Neuroscience Center, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands. [230]Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, Toronto, Ontario, Canada. [231]Department of Psychology, University of Toronto, Toronto, Ontario, Canada. [232]Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada. [233]Clinic for Cognitive Neurology, University Clinic Leipzig, Leipzig, Germany. [234]HMNC Brain Health, Munich, Germany. [235]Department of Neurology, Hopital Erasme, Universite Libre de Bruxelles, Brussels, Belgium. [236]Department of Psychology, Georgia State University, Atlanta, GA, USA. [237]Department of Medical Genetics, Oslo University Hospital, Oslo, Norway. [238]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany. [239]Institute of Translational Medicine, University of Liverpool, Liverpool, UK. [240]Institute of Clinical Medicine—Neurology, University of Eastern Finland, Kuopio, Finland. [241]Neurocentre Neurology, Kuopio University Hospital, Kuopio, Finland. [242]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany. [243]German Center for Cardiovascular Research (partner site Greifswald), Greifswald, Germany. [244]Department of Neuroinformatics, Araya, Tokyo, Japan. [245]Institute of Cognitive Neuroscience, University College London, London, UK. [246]School of Psychology, University of Sussex, Brighton, UK.

[247]Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden. [248]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [249]Department of Child and Adolescent Psychiatry/Psychology, Erasmus MC-Sophia Children's Hospital, Rotterdam, the Netherlands. [250]Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, the Netherlands. [251]GROW School for Oncology and Developmental Biology, Maastricht, the Netherlands. [252]Centre for Advanced Imaging, University of Queensland, Brisbane, Queensland, Australia. [253]Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany. [254]Department of Psychiatry, Erasmus Medical Center, Rotterdam, the Netherlands. [255]Neuropsychiatric Institute, Prince of Wales Hospital, Sydney, New South Wales, Australia. [256]Department of Psychology, VU University Amsterdam, Amsterdam, the Netherlands. [257]Center for Imaging of Neurodegenerative Disease, San Francisco VA Medical Center, University of California, San Francisco, San Francisco, CA, USA. [258]Department of Radiology and Biomedical Imaging, University of California, San Francisco, San Francisco, CA, USA. [259]Leiden Academic Centre for Drug Research, Leiden University, Leiden, the Netherlands. [260]Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA. [261]UK Dementia Research Institute, University of Edinburgh, Edinburgh, UK. [262]Department of Neurology, Center for Neuroscience, University of California, Davis, Sacramento, CA, USA. [263]Department of Neuroscience, Baylor College of Medicine, Houston, TX, USA. [264]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA. [265]Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX, USA. [266]These authors contributed equally: Claudia L. Satizabal, Hieab H. H. Adams, Derrek P. Hibar, Charles C. White. [267]These authors jointly supervised this work: Sarah E. Medland, Joshua M. Shulman, Paul M. Thompson, Sudha Seshadri, M. Arfan Ikram. *e-mail: satizabal@uthscsa.edu; m.a.ikram@erasmusmc.nl

Methods

Study population. The present effort included 53 study samples from the CHARGE consortium[74], ENIGMA consortium[75] and UK Biobank[76]. Briefly, the CHARGE consortium is a collaboration of predominantly population-based cohort studies investigating the genomics of age-related complex diseases, including those of the brain (https://depts.washington.edu/chargeco/wiki/). The ENIGMA consortium brings together various studies, approximately 75% of which are population based, with the remainder using case control designs for various neuropsychiatric or neurodegenerative diseases (http://enigma.ini.usc.edu/). UK Biobank is a large-scale prospective epidemiological study of over 500,000 individuals aged 40–69 years from the United Kingdom, which was established to investigate the genetic and non-genetic determinants of middle- and old-age diseases (https://www.ukbiobank.ac.uk/).

Our sample consisted of up to $n = 37{,}741$ individuals of European ancestry. We additionally included three generalization samples of African Americans (up to $n = 769$) and two generalization samples of Asians ($n = 341$). All participants provided written informed consent and the investigators on the participating studies obtained approval from their institutional review board or equivalent organization. The institutional review boards of Boston University and the University of Southern California, as well as the local ethics board of Erasmus University Medical Center approved this study.

Exclusion criteria comprised prevalent dementia or stroke at the time of the MRI scan and, when available, the presence of large brain infarcts or other neurological pathologies seen during MRI that could substantially influence the measurement of brain volumes (for example, brain tumor or trauma). Individual studies applied the exclusion criteria before analysis.

Definition of phenotypes. Our 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 (in cm^3) of the left and right hemispheres, with the exception of the brainstem, for which the total volume (in cm^3) was used. 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 in-house segmentation methods that were previously described and validated. The summary statistics for subcortical brain volumes in CHARGE study samples are presented in Supplementary Table 3. The study-specific MRI protocols and software are described in Supplementary Table 5. We recently published results describing the genetic variation associated with hippocampal volumes[20]; therefore, we have not included the hippocampus in this report.

Genotyping. Genotyping was performed using a variety of commercial arrays across the participating studies. Study samples and genetic variants underwent similar quality control procedures based on the genetic homogeneity, call rate, MAF and Hardy–Weinberg equilibrium. Good-quality variants were used as input for imputation to the 1000 Genomes Project (phase 1; version 3) reference panel[16], or the HRC (version 1.1)[17] in UK Biobank, using validated software packages. A detailed description of the genotyping and quality control carried out by each study is described in Supplementary Table 6.

Heritability. Heritability of subcortical brain volumes was estimated in the FHS[77] and ASPS-Fam[78]—two population-based cohorts with family structure. We used SOLAR[79] to determine the ratio of the genetic variance to the phenotypic variance, including variance component models that were adjusted for age, sex and total intracranial volume, as well as age^2 and principal components if required, in the same way as described for the GWA analysis. We also estimated the variance of subcortical structures explained by SNPs in a sample of $n = 3{,}486$ unrelated participants from the Rotterdam Study using GCTA[80], and additionally in the full European sample using LDSC methods[81]. Supplementary Table 4 provides family- and SNP-based heritabilities for subcortical structures.

Genome-wide associations and meta-analysis. For CHARGE and ENIGMA, each study undertook a GWA analysis on the volumes of seven MRI subcortical brain structures (or those that were available to each study), according to a common predefined analysis plan. Studies including unrelated participants performed linear regression analyses, whereas those including related participants conducted linear mixed models to account for familial relationships. Models assumed additive genetic effects and were adjusted for age, sex, total intracranial volume and, if applicable, age^2, principal components to account for population stratification, psychiatric diagnosis (ENIGMA cohorts), and study site. Individual studies shared summary statistics to a centralized, secured computing space. Analysis in the UK Biobank sample followed a similar approach in $n = 8{,}312$ unrelated participants, although the genetic data used for these analyses used only those variants imputed using the HRC[17] reference panel. As the data released by UK Biobank did not include total intracranial volume, linear regression models in this sample were adjusted for age, age^2, sex, total brain volume and principal components. We used LDSC methods[81] to investigate the genetic correlations for all subcortical structures between the CHARGE and ENIGMA consortia combined and UK Biobank. There was no evidence suggesting differences in the genetic architecture of both samples.

Before meta-analysis, we performed quality control on the summary statistics from each study sample by using a series of quality checks implemented in EasyQC[82]. Filters were set to remove SNPs with poor imputation ($R^2 < 0.5$), rare SNPs (MAF < 0.1%) or SNPs with an effective allele count ($2 \times \text{MAF} \times \text{study sample size} \times \text{imputation quality}$) of <20. Finally, we only considered variants present in at least 70% of the total European sample for each structure.

Fixed-effects meta-analyses weighting for sample size was performed using METAL[18], given that not all samples used the same methods for acquisition and post-processing of brain images. We used the LDSC intercept to correct for population stratification and cryptic relatedness[81]. Quantile and Manhattan plots are presented for each subcortical structure in Supplementary Fig. 1. To correct for multiple comparisons across our seven traits, we calculated the Pearson's correlation among subcortical structures, adjusting for age, sex and intracranial volume in $n = 4{,}459$ participants from the Rotterdam Study. After 1,000 permutations, the resulting number of independent traits was six, leading to the definition of a significant threshold as $P < (5 \times 10^{-8}/6) = 8.3 \times 10^{-9}$. To select our top independent SNPs in the European meta-analysis, we ran a multi-SNP-based conditional and joint association analysis (GCTA-COJO)[80] using $n = 6{,}921$ participants from the Rotterdam Study as the reference sample. In secondary analyses, we looked for associations of our index SNPs (the most significant variant in each locus) with the other six subcortical structures.

We conducted separate meta-analyses by ancestry, and further performed a combined meta-analysis including all samples. Forest plots were created to explore the contribution of participating studies to each of the significant SNPs (Supplementary Fig. 4). To assess signal overlap with African American and Asian samples, we first clumped variants with $P < 1 \times 10^{-4}$ in the European sample, and then ran binomial sign tests for the correlation of the direction of association across ethnic groups.

Functional annotations. We used Locus Zoom[19], based on the hg19 UCSC Genome Browser assembly, for the visualization of the nearest genes within a ±500-kilobase genomic region. We also investigated *cis* (1-megabase) eQTLs and meQTLs for our index SNPs in postmortem brains from ROSMAP. In ROSMAP, the DLPFC was selected for initial multi-omics data generation, as it is relevant to multiple common neuropathologies and cognitive phenotypes in the aging population[83]. RNA was extracted from the gray matter of DLPFC, and next-generation RNA sequencing was done on the Illumina HiSeq for samples with an RNA integrity score of >5 and a quantity threshold of >5 µg, as previously described[83,84]. We quantile-normalized the fragments per kilobase of transcript per million fragments mapped, correcting for batch effect with Combat[84,85]. These adjusted fragments per kilobase of transcript per million fragments mapped values were used for analysis. A subset of $n = 407$ participants had quality-controlled RNA sequencing data and were included in the eQTL analysis.

DNA methylation levels from the gray matter of the DLPFC were measured using the Illumina HumanMethylation450 BeadChip, and the measurements underwent quality control processing as previously described (that is, detection $P < 0.01$ for all samples)[83], yielding $n = 708$ participants with 415,848 discrete CpG dinucleotide sites with methylation measurement. Any missing methylation levels from any of quality-controlled CpG dinucleotide sites were imputed using a k-nearest neighbor algorithm for $k = 100$ (ref. [83]). A subset of $n = 488$ participants in our study had quality-controlled genome-wide methylation data and were included in the *cis*-meQTL analysis. Finally, the associations between our index SNPs and CpG sites were plotted along Roadmap Epigenomics chromatin states for ten brain tissues[86].

We further queried *cis*- and *trans*-eQTLs in non-brain and brain tissues from additional eQTL repositories[87]. We searched for proxies to our index SNPs with linkage disequilibrium $r^2 > 0.8$, using the European population reference in rAggr (1,000 G; phase 1; March 2012), then queried index and proxy SNPs against eQTLs from diverse databases[88]. Blood cell-related eQTL studies included: fresh lymphocytes and leukocytes; leukocyte samples in individuals with celiac disease; whole blood samples; lymphoblastoid cell lines (LCLs) derived from asthmatic children; HapMap LCLs from three populations; a separate study on HapMap Utah Residents with Northern and Western European Ancestry (CEU) LCLs; LCL population samples; neutrophils; $CD19^+$ B cells; primary phytohemagglutinin-stimulated T cells; $CD4^+$ T cells; peripheral blood monocytes; long non-coding RNAs in $CD14^+$ monocytes purified from white blood cells and $CD14^+$ monocytes before and after stimulation with lipopolysaccharide or interferon-γ; $CD11^+$ dendritic cells before and after *Mycobacterium tuberculosis* infection; a separate study of dendritic cells before or after stimulation with lipopolysaccharide, influenza or interferon-β; micro-RNA QTLs, DNase I QTLs, histone acetylation QTLs and ribosomal occupancy QTLs queried for LCLs; and splicing QTLs and micro-RNA QTLs queried in whole blood. Non-blood cell tissue eQTL searches included: omental and subcutaneous adipose; visceral fat stomach; endometrial carcinomas; ER^+ and ER^- breast cancer tumor cells; liver; osteoblasts; intestine; normal and cancerous colon; skeletal muscle; breast tissue (normal and cancerous); lung; skin; primary fibroblasts; sputum; pancreatic islet cells; prostate; rectal mucosa; and arterial wall and heart tissue from left ventricles and left and right

atria. Micro-RNA QTLs were also queried for gluteal and abdominal adipose and liver. MeQTLs were queried in pancreatic islet cells. Further messenger RNA and micro-RNA QTLs were queried from ER+ invasive breast cancer samples, as well as colon, kidney renal clear, lung and prostate adenocarcinoma samples. Brain eQTL studies included: brain cortex; cerebellar cortex; cerebellum; frontal cortex; gliomas; hippocampus; inferior olivary nucleus (from medulla); intralobular white matter; occipital cortex; parietal lobe; pons; prefrontal cortex; putamen (at the level of the anterior commissure); substantia nigra; temporal cortex; thalamus; and visual cortex. eQTL data were integrated from online sources, including ScanDB[89], the GTEx Portal[90] and the Pritchard Lab[91]. Cerebellum, parietal lobe and liver eQTL data were downloaded from ScanDB. *Cis*-eQTLs were limited to those with $P < 1.0 \times 10^{-6}$ and *trans*-eQTLs were limited to those with $P < 5.0 \times 10^{-8}$. The results for GTEx Analysis version 6 for 48 tissues were downloaded from the GTEx Portal (https://www.gtexportal.org). For all gene-level eQTLs, if at least one SNP passed the tissue-specific empirical threshold in GTEx, the best SNP for that eQTL was always retained.

Associations of cognition and neuropathology phenotypes with gene expression in the brain. We further related cognitive function and neuropathological findings to the expression of the 199 genes influencing subcortical volumes in 508 brains from the ROSMAP samples.

Briefly, brain autopsies were performed as previously described, and each brain was inspected for common pathologies relating to loss of cognition in aging populations[92,93]. In this report, we included: neurofibrillary tangles; neuritic plaques; β-amyloid load; tau density; hippocampal sclerosis; Lewy bodies; and neuronal loss in substantia nigra. Neurofibrillary tangles and neuritic plaques were visualized by modified Bielschowsky silver stain, then counted and scaled in five brain regions: mid-frontal; temporal; inferior parietal; entorhinal cortex; and hippocampus CA1. Composite scores for each of these three pathology types were derived by scaling the counts within each of the five regions and taking the square root of the average of the regional scaled values to account for their positively skewed distribution[92–94]. β-amyloid load and tau tangle density were measured by immunohistochemistry and square root transformed as previously described[95]. Lewy bodies were identified using immunohistochemistry, and were further dichotomized as present or absent based on the recommendations of the Report of the Consortium on DLB International Workshop[96]. Hippocampal sclerosis was recorded as either present or absent, as evaluated by hematoxylin and eosin staining. Nigral neuronal loss was assessed in the substantia nigra in the mid- to rostral midbrain near or at the exit of the third nerve using hematoxylin and eosin staining and 6-µm sections and a semiquantitative scale (0–3)[97].

Global cognition was computed as a composite score of 19 (Religious Order Study) and 17 (Rush Memory and Aging Project) cognitive tests performed during annual evaluations, including five cognitive domains: episodic memory; semantic memory; working memory; perceptual speed; and visuospatial ability[92,93]. From these scores, we created normalized summary measures to limit the influence of outliers. We used global cognition proximate to death to derive cognitive reserve. Separately, the residual slope of global cognitive change and the residual slopes of cognitive change in the five cognitive domains were derived through general linear mixed models, controlling for age at enrollment, sex and education.

Phenotypic and genetic correlations. We estimated the Pearson's partial phenotypic correlations among the volumes of subcortical structures in 894 participants from the FHS. Similarly to the GWA, these analyses were corrected for the effects of sex, age, age^2, total intracranial volume and principal component 1.

Genetic correlation analyses were performed using LDSC methods[81]. The GWA meta-analysis results for the seven subcortical brain structures were correlated with each other's, as well as with published GWA studies on the following traits: hippocampal volume[20]; intracranial volume[21]; white matter hyperintensities[22]; stroke subtypes[23]; adult height and body mass index[24]; fat-free mass and whole-body water mass[98]; Alzheimer's disease[26]; Parkinson's disease[27]; general cognitive function[25]; bipolar disorder and schizophrenia[28]; and ADHD[29].

Look-up of functional orthologs in *D. melanogaster*. For the cross-species assessment of gene–phenotype relationships in *Drosophila*, we relied on a similar analytic approach as in previous work[99]. Human genes were mapped to corresponding *Drosophila* orthologs using the *Drosophila* Integrated Ortholog Prediction Tool (https://www.flyrnai.org/diopt)[100], which incorporates 14 distinct algorithms to define orthology. Fly gene orthologs were defined based on a *Drosophila* Integrated Ortholog Prediction Tool score of ≥2, indicating that at least two algorithms were in agreement on the pairing. When more than one of the fly ortholog was predicted, all such genes meeting this threshold were included in our analyses. This resulted in a gene set consisting of 168 *Drosophila* homologs of human candidate genes at subcortical volume susceptibility loci. The resulting 37 genes associated with neuroanatomy-defective phenotypes in *Drosophila* (22%) were annotated based on the controlled vocabulary terms implemented in FlyBase (http://flybase.org/)[101]. Genes causing neuroanatomy-defective phenotypes in *Drosophila* include both loss-of-function and gain-of-function genetic manipulations of fly gene homologs. Loss-of-function studies included both classical mutant alleles (for example, point mutations, gene deletions or transposon insertions) or gene knockdown using RNA interference transgenic strains. Gain-of-function experiments were based on tissue-specific overexpression of the fly gene orthologs. The hypergeometric overlap test was used to assess for enrichment of neuroanatomy-defective phenotypes among the conserved gene set.

Protein–protein interactions and network analysis. We used the human STRING database resource (string-db.org)[32] for the exploration of direct (physical) and indirect (functional) protein–protein interactions based on the gene set derived from the GWA results and functional annotations (Supplementary Table 13). The input parameters included a medium-confidence interaction score (0.4) with first and second shells of a maximum of five interactors. Finally, we generated a protein–protein interaction network based on known and predicted interactions.

Partitioning heritability. Partitioned heritability was estimated with stratified LDSC methods[30]. This method partitions SNP heritability using GWA study summary results and accounting by linkage disequilibrium. We used the meta-analysis results from the European sample to partition SNPs by 28 functional categories, including: coding; intron; promoter; 3′/5′ untranslated region; digital genomic footprint; transcription factor binding site; chromHMM and Segway annotations for six cell lines; DNase I hypersensitivity sites; H3K4me1, H3K4me3 and H3K9ac marks; two sets of H3K27ac marks; super-enhancers; conserved regions in mammals; and FANTOM5 enhancers. Significance was set at $P < (0.05/(28 \times 6)) = 3 \times 10^{-4}$.

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

Data availability

The genome-wide summary statistics that support the findings of this study are available from the CHARGE dbGaP (accession code: phs000930) and ENIGMA (http://enigma.ini.usc.edu/research/download-enigma-gwas-results) websites.

References

74. Psaty, B. M. et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ. Cardiovasc. Genet.* **2**, 73–80 (2009).
75. Thompson, P. M. et al. The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav.* **8**, 153–182 (2014).
76. Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
77. Tsao, C. W. & Vasan, R. S. Cohort Profile: the Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. *Int. J. Epidemiol.* **44**, 1800–1813 (2015).
78. Schmidt, R. et al. Assessment of cerebrovascular risk profiles in healthy persons: definition of research goals and the Austrian Stroke Prevention Study (ASPS). *Neuroepidemiology* **13**, 308–313 (1994).
79. Almasy, L. & Blangero, J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am. J. Hum. Genet.* **62**, 1198–1211 (1998).
80. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
81. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
82. Winkler, T. W. et al. Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
83. Bennett, D. A., Yu, L. & De Jager, P. L. Building a pipeline to discover and validate novel therapeutic targets and lead compounds for Alzheimer's disease. *Biochem. Pharm.* **88**, 617–630 (2014).
84. Chan, G. et al. *CD33* modulates TREM2: convergence of Alzheimer loci. *Nat. Neurosci.* **18**, 1556–1558 (2015).
85. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
86. Roadmap Epigenomics Association et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
87. Eicher, J. D. et al. GRASPv2.0: an update on the Genome-Wide Repository of Associations between SNPs and phenotypes. *Nucleic Acids Res.* **43**, D799–D804 (2015).
88. Zhang, X. et al. Synthesis of 53 tissue and cell line expression QTL datasets reveals master eQTLs. *BMC Genomics* **15**, 532 (2014).
89. Zhang, W. et al. SCAN database: facilitating integrative analyses of cytosine modification and expression QTL. *Database* **2015**, bav025 (2015).
90. Consortium, G. T. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
91. Veyrieras, J. B. et al. High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet.* **4**, e1000214 (2008).

92. Bennett, D. A. et al. Overview and findings from the rush Memory and Aging Project. *Curr. Alzheimer Res.* **9**, 646–663 (2012).
93. Bennett, D. A., Schneider, J. A., Arvanitakis, Z. & Wilson, R. S. Overview and findings from the religious orders study. *Curr. Alzheimer Res.* **9**, 628–645 (2012).
94. Replogle, J. M. et al. A TREM1 variant alters the accumulation of Alzheimer-related amyloid pathology. *Ann. Neurol.* **77**, 469–477 (2015).
95. Barnes, L. L., Schneider, J. A., Boyle, P. A., Bienias, J. L. & Bennett, D. A. Memory complaints are related to Alzheimer disease pathology in older persons. *Neurology* **67**, 1581–1585 (2006).
96. McKeith, I. G. et al. Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): report of the Consortium on DLB International Workshop. *Neurology* **47**, 1113–1124 (1996).
97. Schneider, J. A. et al. Substantia nigra tangles are related to gait impairment in older persons. *Ann. Neurol.* **59**, 166–173 (2006).
98. Zheng, J. et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).
99. Wangler, M. F., Hu, Y. & Shulman, J. M. *Drosophila* and genome-wide association studies: a review and resource for the functional dissection of human complex traits. *Dis. Model Mech.* **10**, 77–88 (2017).
100. Hu, Y. et al. An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinformatics* **12**, 357 (2011).
101. Marygold, S. J., Crosby, M. A., Goodman, J. L. & FlyBase, C. Using FlyBase, a database of *Drosophila* genes and genomes. *Methods Mol. Biol.* **1478**, 1–31 (2016).

nature research

Corresponding author(s): Drs. Satizabal & Ikram NG-A47229R

Last updated by author(s): Aug 28, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software were used
Data analysis	Associations at the individual study level: mach2qtl, probABEL, plink, merlin, SNPtest, RareMetalWorker. Neuroimaging processing: FreeSurfer (versions 3.0.2, 4.0.1, 4.3.0, 4.5.0, 5.0.0, 5.1.0, 5.3.0), SPM99, FSL-FIRST (versions 4.1.5, 4.1.7, 4.1.9, 5.0.4), custom image analysis pipeline based on the MNI software, in-house imaging software, MIPAV. Imputation on 1000 genomes or HRC: beagle, MaCH, IMPUTE, minimac, SHAPEIT. QC, heritability and meta-analysis: EASYQC, SOLAR, R, METAL, GCTA, LDSC, HASE. Functional follow-up: R, rAggr, Locus Zoom, LDSC, DIOPT, FlyBase, STRING.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

We have included the following statement: "The genome-wide summary statistics that support the findings of this study will be made available through the CHARGE dbGaP (accession number phs000930) and ENIGMA (http://enigma.ini.usc.edu/research/download-enigma-gwas-results) websites."

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We did not base our study on a calculated sample size. The sample size of each study sample included in the GWAS was determined by the number of individuals with both brain MRI outcomes and GWAS data. Following current practice for meta-analysis of GWAS, our study includes the largest possible sample from CHARGE and ENIGMA cohorts that were able to contribute data. We used the largest UKBB sample with QC'd imaging and GWAS data that was available to us.
Data exclusions	Our analysis pre-specified the exclusion of persons with prevalent stroke and dementia at the time of MRI, as well as those with neuroimaging abnormalities (i.e. brain tumor, large brain infarcts) that could have influenced the measurement of MRI outcomes.
Replication	We sought replication in three samples of African-American ancestry (up to n=769), and two of Asian (Chinese/Malay) ancestry (n=341). Due to differences in allele frequency in non-European samples, some of our lead variants were not present in one or more of our replication samples. Details are presented in Supplementary Table S7.
Randomization	Randomization was not used in this observational study.
Blinding	During data acquisition, the personnel in charge of processing MRI data were blinded to the genetic participant's data and vice-versa.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Our sample consisted of up to n=38,851 individuals of European ancestry. We additionally included three generalization samples of African-Americans (up to n=769), and two generalization samples of Asians (n=341). Participants' age ranged from 9 to 98 years and the percentage of females ranged between 0 and 73%.
Recruitment	The present effort included 53 study samples from the Cohorts of Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium, the Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) consortium, and the United Kingdom Biobank (UKBB). The CHARGE consortium is a collaboration of predominantly population-based cohort studies investigating the genomics of age-related complex diseases, including those of the brain. The ENIGMA consortium brings together various studies, approximately 75% of which are population-based, with the remainder using case-control designs for various neuropsychiatric or neurodegenerative diseases. The UKBB is a large-scale prospective epidemiological study of individuals aged 40-69 years from the United Kingdom, established to investigate the genetic and non-genetic determinants of middle and old age diseases.
Ethics oversight	The institutional review boards of Boston University and the University of Southern California, as well as the local ethics board of Erasmus University Medical Center approved this study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Magnetic resonance imaging

Experimental design

Design type	GWAS of brain MRI-based subcortical structures.
Design specifications	Each study contributed MRI data obtained using diverse scanners, field strengths, and acquisition protocols as described in Supplementary Table S5.
Behavioral performance measures	No behavioral performance was assessed, only structural MRI.

Acquisition

Imaging type(s)	Structural MRI
Field strength	1T to 4T
Sequence & imaging parameters	Varied per study, specified in Supplementary Table S5. All studies had T1-weighted sequences
Area of acquisition	Whole brain scan, including brainstem

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software	Freesurfer, SPM, FSL-FIRST, MNI, in-house imaging software, MIPAV (Supplementary Table S5)
Normalization	Normalized as in specific approaches detailed above
Normalization template	Varied across cohorts based on software described above
Noise and artifact removal	Scans where volumes could not be assessed due to artifacts were excluded
Volume censoring	Varied across cohorts based on software described above

Statistical modeling & inference

Model type and settings	Fixed-effects meta-analysis assuming additive genetic model
Effect(s) tested	Associations of SNPs across the whole genome with MRI-based subcortical brain volumes

Specify type of analysis: ☒ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016)	Association of genotype allele dosages with subcortical brain volumes
Correction	We applied Bonferroni correction for testing of multiple SNPs and MRI traits

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☒	☐ Multivariate modeling or predictive analysis