







The genetic architecture of the human thalamus and its overlap with ten common brain disorders

Torbjørn Elvsåshagen ^{1,2,3}✉, Alexey Shadrin ^{1,3}, Oleksandr Frei^{1,3,4}, Dennis van der Meer^{1,5}, Shahram Bahrami^{1,3}, Vinod Jangir Kumar⁶, Olav Smeland ^{1,3}, Lars T. Westlye ^{1,7,8}, Ole A. Andreassen ^{1,3,8} & Tobias Kaufmann ^{1,3,9}✉

The thalamus is a vital communication hub in the center of the brain and consists of distinct nuclei critical for consciousness and higher-order cortical functions. Structural and functional thalamic alterations are involved in the pathogenesis of common brain disorders, yet the genetic architecture of the thalamus remains largely unknown. Here, using brain scans and genotype data from 30,114 individuals, we identify 55 lead single nucleotide polymorphisms (SNPs) within 42 genetic loci and 391 genes associated with volumes of the thalamus and its nuclei. In an independent validation sample ($n = 5173$) 53 out of the 55 lead SNPs of the discovery sample show the same effect direction (sign test, $P = 8.6e-14$). We map the genetic relationship between thalamic nuclei and 180 cerebral cortical areas and find overlapping genetic architectures consistent with thalamocortical connectivity. Pleiotropy analyses between thalamic volumes and ten psychiatric and neurological disorders reveal shared variants for all disorders. Together, these analyses identify genetic loci linked to thalamic nuclei and substantiate the emerging view of the thalamus having central roles in cortical functioning and common brain disorders.

¹NORMENT, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. ²Department of Neurology, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway. ³Institute of Clinical Medicine, University of Oslo, Oslo, Norway. ⁴Center for Bioinformatics, Department of Informatics, University of Oslo, Oslo, Norway. ⁵Faculty of Health, Medicine and Life Sciences, School of Mental Health and Neuroscience, Maastricht University, Maastricht, The Netherlands. ⁶Max Planck Institute for Biological Cybernetics, Tübingen, Germany. ⁷Department of Psychology, University of Oslo, Oslo, Norway. ⁸K.G. Jebsen Center for Neurodevelopmental Disorders, University of Oslo, Oslo, Norway. ⁹Tübingen Center for Mental Health, Department of Psychiatry and Psychotherapy, University of Tübingen, Tübingen, Germany. ✉email: torbjorn.elvsashagen@medisin.uio.no; tobias.kaufmann@medisin.uio.no

Recent studies indicated that the thalamus has a broader role in cognition than previously assumed^{1–5}. Cognitive neuroscience is therefore shifting focus to how the thalamus regulates cortical activity and supports higher-order functions such as working memory⁶, attentional control⁷, and visual processing⁸. There is also a growing appreciation of thalamic contributions to the pathogenesis of neurological and psychiatric disorders, including dementias⁹, Parkinson’s disease (PD)¹⁰, schizophrenia (SCZ)¹¹, and bipolar disorder (BD)¹². Despite the importance for human cognition and disease, the genetic architecture of the thalamus and its relationship with cortical structure and common brain disorders remain largely unknown.

The thalamus can be divided into nuclei that mainly relay peripheral information to the cerebral cortex and higher-order nuclei that modulate cortical functions^{1,4}. Two recent studies found one¹³ and two¹⁴ genetic loci associated with whole thalamus volume, yet there is no genome-wide association study (GWAS) of thalamic nuclei. In this work, we further identify genetic loci and candidate genes for volumes of the whole thalamus and its nuclei. We detect overlap between genetic architectures of the thalamic nuclei and 180 cortical areas consistent with thalamocortical structural connectivity. Moreover, pleiotropy analyses between thalamic volumes and ten neurological and psychiatric disorders reveal shared variants for all disorders. Together, these analyses provide insights into the genetic underpinnings of human thalamic nuclei and substantiate the emerging view of the thalamus having central roles in common brain disorders.

Results and discussion

We accessed brain magnetic resonance imaging (MRI) data from $n = 30,432$ genotyped White British from the UK Biobank¹⁵. The MRI data were segmented into the whole thalamus and six thalamic nuclei groups—anterior, lateral, ventral, intralaminar, medial, and posterior—using Bayesian thalamus segmentation^{16,17} (Fig. 1a). We removed data sets with segmentation errors and insufficient data quality ($n = 318$) after visually inspecting the segmentations for each of the 30,432 participants. The resulting 30,114 data sets comprised the discovery sample (52% females; age range 45–82 years).

We conducted GWAS with PLINK¹⁸ on whole thalamus and the six nuclei volumes accounting for age, age-orthogonalized age-squared, sex, scanning site, intracranial volume, and the first 20 genetic principal components. The thalamic nuclei GWAS also accounted for whole thalamus volume, thus revealing genetic signals beyond commonality in volume^{19,20}. The thalamic nuclei GWAS were also run without covarying for whole thalamus volume and are presented in the Supplementary Information.

Single-nucleotide polymorphism (SNP)-based heritability estimated using linkage disequilibrium (LD) score regression²¹ was 25% for whole thalamus and 18–32% for the six nuclei groups (Fig. 1b). We found genome-wide significant hits for all seven volumes (Bonferroni-corrected $P < 5e - 8/7 = 7e - 9$) and identified a total of 55 lead SNPs in 50 genomic loci (Fig. 1c, d and Supplementary Data 1). Seven loci were associated with whole thalamus; 3, 11, and 5 loci with anterior, lateral, and ventral nuclei; and 12, 6, and 6 loci were associated with intralaminar, medial, and posterior nuclei.

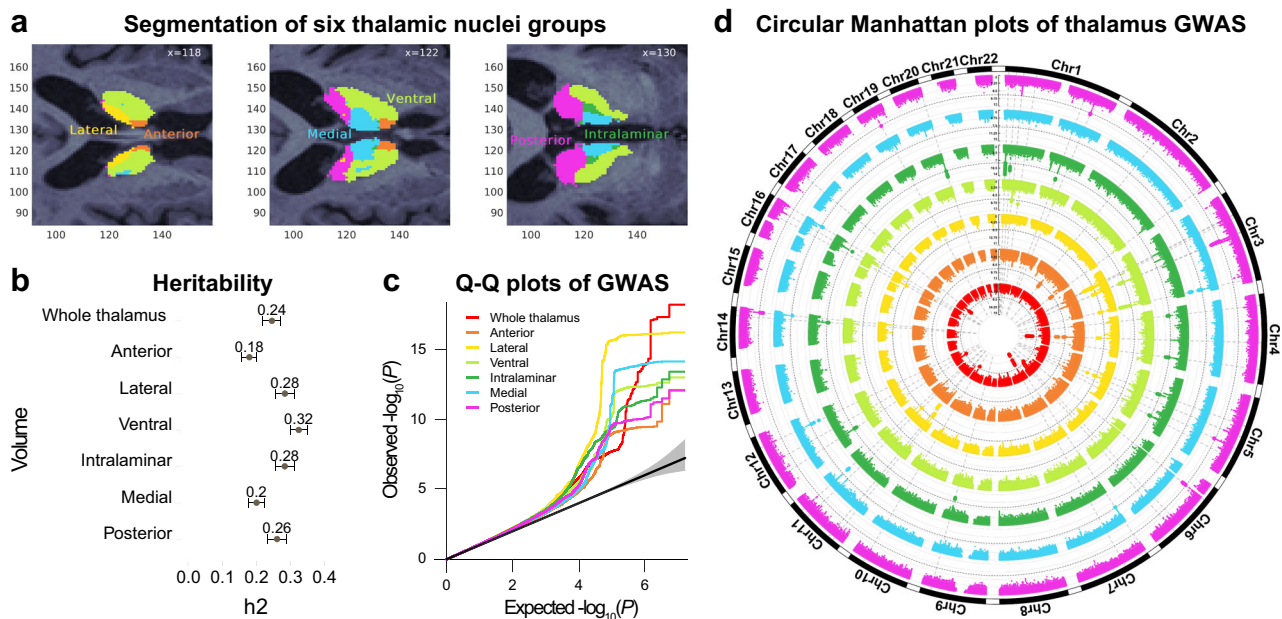


Fig. 1 Analysis of the GWAS discovery sample identifies 42 loci associated with thalamic volumes. **a** The thalamus was segmented into six nuclei groups—anterior, lateral, ventral, intralaminar, medial, and posterior nuclei—using Bayesian thalamus segmentation^{16,17}. All data sets were assessed by visually inspecting axial view figures of the segmentations for each participant and we removed sets with segmentation errors, insufficient data quality, and pathologies. The nuclei volumes in the left and right thalamus were summed and these were used in the analyses. **b** Heritability estimates for the thalamic volumes in the discovery sample of $n = 30,114$ participants from the UK Biobank. All thalamic volumes showed substantial heritability. Data are presented as mean \pm SE. **c** Q-Q plots for the thalamic volumes of the discovery sample. **d** Circular Manhattan plots of GWAS for thalamus volumes of the discovery sample. The innermost plot reflects the GWAS of whole thalamus volume, whereas from center to the periphery, the plots indicate the GWAS of the anterior, lateral, ventral, intralaminar, medial, and posterior nuclei, respectively. Black circular dashed lines indicate genome-wide significance (two-sided $P < 7e - 9$). Horizontal Manhattan plots for the seven volumes are shown in Supplementary Fig. 2. The colors in **a**, **c**, and **d** indicate the same volumes, i.e., red color reflects whole thalamus; orange, yellow, and light green indicate the anterior, lateral, and ventral nuclei, respectively; whereas dark green, blue, and magenta reflect intralaminar, medial, and posterior nuclei volumes, respectively. GWAS, genome-wide association studies; h², heritability.

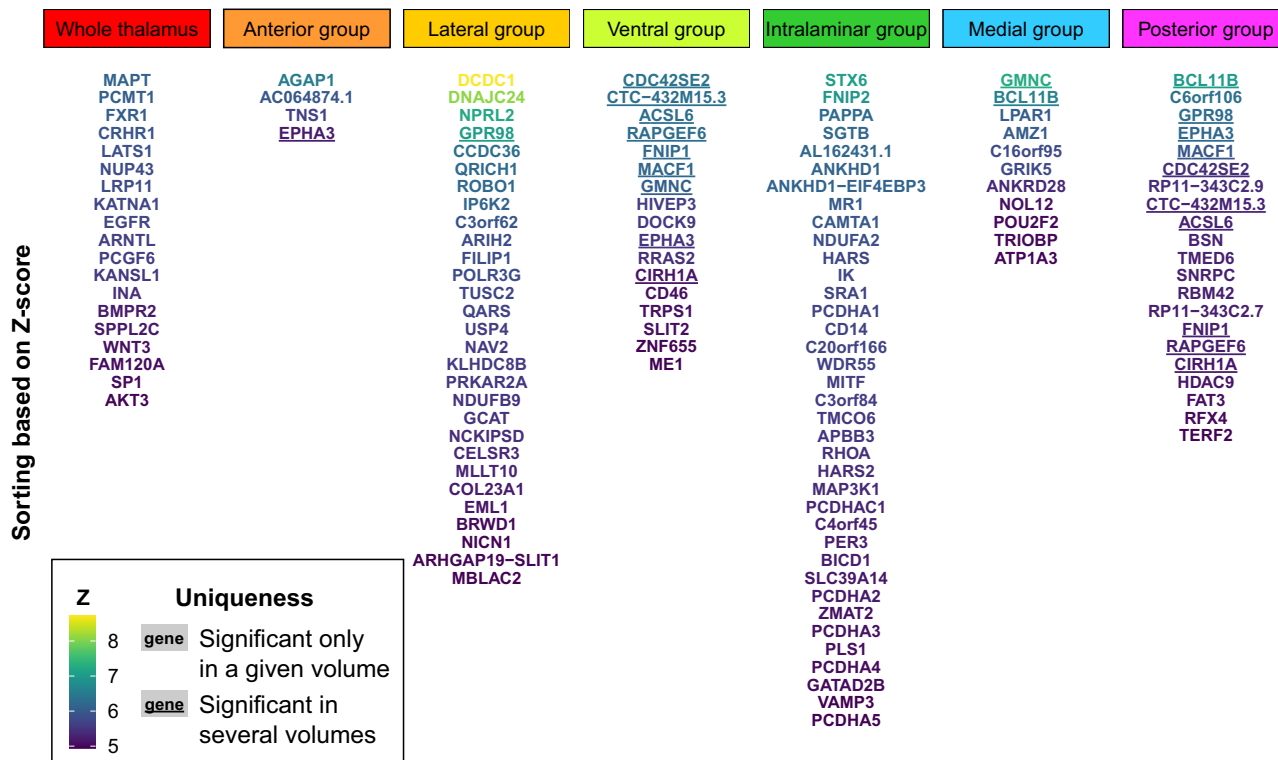


Fig. 2 GWGAS identifies 127 unique genes associated with thalamic volumes. Nineteen genes were associated with whole thalamus; 4, 29, and 17 genes were associated with volumes of the anterior, lateral, and ventral nuclei; and 37, 11, and 21 genes were associated with intralaminar, medial, and posterior nuclei volumes, respectively. Lighter font color and higher position for gene names indicate greater Z-score. Underlined gene names designate genes that were significantly associated with more than one volume, whereas gene names not underlined indicate genes associated with only one volume. Additional results of the GWGAS are presented in Supplementary Data 12. GWGAS, genome-wide gene-based association analysis.

and posterior nuclei volumes, respectively. Eight loci were associated with more than 1 volume, resulting in 42 unique thalamus-associated loci. One of these (rs76928645 on chromosome 7) was associated with whole thalamus volume in a recent study¹³. Notably, 37 of the 42 genetic loci were associated with only one volume. The genomic inflation factors for the seven GWAS are listed in Supplementary Data 2. The strongest genomic inflation was observed for the ventral nuclei ($\lambda = 1.168$), whereas the anterior nuclei showed the weakest inflation ($\lambda = 1.105$).

Three lead SNPs had a combined annotation-dependent depletion (CADD) score > 15, which indicates deleterious protein effects²². rs13107325 (score = 23.1) was associated with anterior and lateral nuclei volumes, is located within the metal ions transporter gene *SLC39A8* on chromosome 4, and is linked to cognitive performance, PD, and SCZ^{23,24}. The gene nearest to rs12146713 on chromosome 12 (associated with medial nuclei volume; score = 19.6) is *NUAK1*, which regulates the Tau protein level²⁵. Cerebral Tau accumulation is a defining characteristic of Alzheimer’s disease (AD) and other neurodegenerative disorders²⁶. rs951366 on chromosome 1 was significant for posterior nuclei volume (score = 15.7) and is associated with PD²⁷. Further GWAS results are provided in Supplementary Figs. 1–4 and Supplementary Data 3–9.

For GWAS replication, we used MRI and SNP data from an additional 5173 White British (51% females; age range 46–81 years) from the UK Biobank¹⁵. We found that 53 out of the 55 lead SNPs from the discovery GWAS had the same effect direction in the replication (sign test; $P = 8.6e - 14$). Moreover, 32 of the discovery lead SNPs had uncorrected $P < 0.05$, whereas 23 had uncorrected $P > 0.05$, in the replication (Supplementary Data 10).

To assess the tissue and cell-type specificity of the thalamic GWAS findings, we used LD-score regression applied to

specifically expressed genes (LDSC-SEGs)²⁸. We found that volumes of the whole thalamus, and intralaminar and medial nuclei groups were enriched for central nervous system (CNS) tissues and cell types, yet none of these associations remained significant after Bonferroni correction (Supplementary Fig. 5).

To further examine the biological significance of the GWAS results, we used positional, expression quantitative trait loci (eQTL), and chromatin interaction mapping in Functional Mapping and Annotation of GWAS (FUMA)²⁹ to map candidate SNPs to genes. This identified 336 unique genes across the 7 volumes (Supplementary Data 11 and Supplementary Fig. 6). We conducted genome-wide gene-based association analyses (GWGAS) using MAGMA³⁰ and detected 127 unique genes across the thalamic volumes (Fig. 2, Supplementary Fig. 7, and Supplementary Data 12). The GWGAS gene most strongly associated with whole thalamus volume was *MAPT*. *MAPT* codes for Tau protein in neurons, is implicated in the pathogenesis of neurodegenerative disorders²⁶, and is linked to general cognitive ability³¹. The most strongly associated GWGAS gene for the thalamic nuclei was *DCDC1*, which was linked to lateral nuclei volume. *DCDC1* is a member of the doublecortin gene family with high expression levels in the fetal brain³², yet its functions remain largely unknown.

Fifty-five of the GWGAS genes were not mapped by the GWAS analyses, resulting in a total of 391 thalamus-linked genes. Notably, 95% of these were associated with only one volume. We conducted gene-set analyses using MAGMA³⁰ and found significant Gene Ontology sets for the intralaminar nuclei, implicating cell–cell adhesion, and for the lateral nuclei, involving cell growth and synapse organization (Supplementary Data 13).

We then performed protein–protein interaction analyses to explore the functional relationships between proteins encoded by

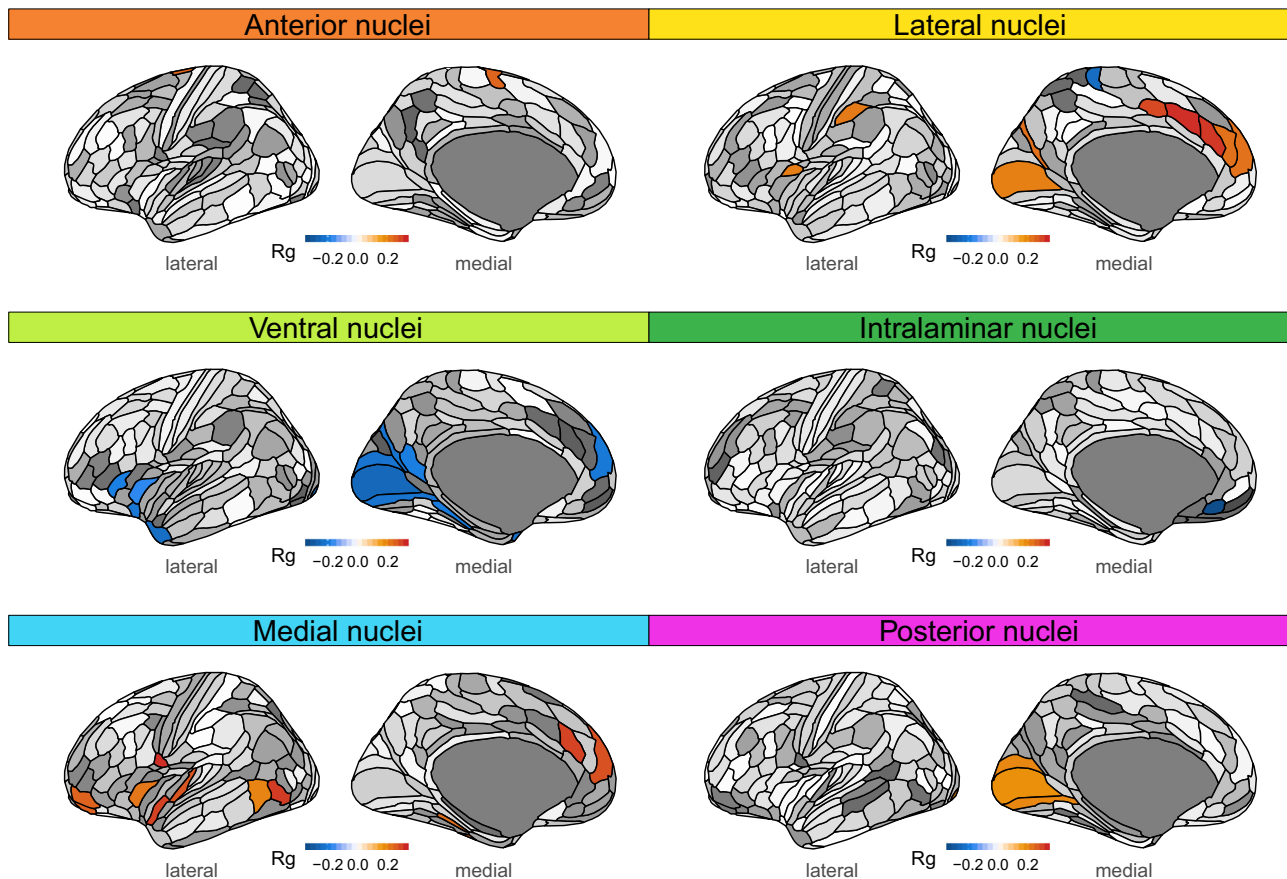


Fig. 3 Thalamocortical genetic relationships. We found significant genetic correlations between all thalamic nuclei and distinct cortical regions. There were significant positive genetic correlations between anterior nuclei and medial premotor cortex; between lateral nuclei and mainly medial prefrontal, anterior cingulate, parietal, and visual cortices; between medial nuclei and prefrontal and temporal cortices; and between posterior nuclei and visual cortices. We also found significant negative genetic associations between ventral nuclei and visual, prefrontal, and temporal cortical regions, and between intralaminar nuclei and rostral medial prefrontal cortex. Warm and cool colors in cortical regions indicate significant positive and negative genetic correlations, respectively, after adjusting for analyses of 180 cortical regions and 6 volumes (two-sided $P < \text{FDR}$). Corresponding statistics are provided in Supplementary Data 15. FDR, false discovery rate; Rg, genetic correlation.

the 391 candidate genes and detected a network with significantly more interactions than expected by chance (protein–protein interaction enrichment: $P < 1e - 16$; Supplementary Data 14 and Supplementary Fig. 8). The most central network nodes were EGFR, RHOA, KANSL1, and NFKB1.

The above analyses identified genetic loci linked to higher-order thalamic nuclei, which project to distinct cortical regions and support cognition^{1,4,33}. For example, medial thalamic nuclei are densely interconnected with prefrontal and temporal cortices, and regulate working memory and attentional control^{16,7}, whereas posterior thalamic nuclei project to occipital cortices and support visual processing⁸. Recent studies suggest that well-connected brain regions exhibit stronger genetic correlations than less-connected regions^{34,35}, yet whether this principle applies to thalamocortical connectivity remains unknown. Thus, we ran GWAS in the discovery sample for volumes for each of the 180 cortical regions defined recently³⁶ and examined genetic correlations with the six thalamic nuclei volumes. These analyses revealed significant associations between specific cortical regions and each of the thalamic nuclei (Fig. 3 and Supplementary Data 15). Interestingly, we found positive correlations mainly for higher-order thalamic nuclei with cortical distributions consistent with established thalamocortical projection patterns, i.e., medial nuclei correlated with prefrontal and temporal cortices, and posterior nuclei with the visual cortex. We also found significant positive

associations between the higher-order lateral nuclei^{17,34} and cortical regions, consistent with their connections with medial prefrontal, anterior cingulate, parietal, and visual cortices^{37,38}.

Thalamic alterations are reported in a growing number of psychiatric and neurological disorders^{9–12,39,40}, yet the genetic relationships between the thalamus and disorders have not been clarified. We used GWAS summary statistics for attention-deficit/hyperactivity disorder (ADHD)⁴¹, autism spectrum disorder (ASD)⁴², BD⁴³, major depression (MD)^{44,45}, SCZ⁴⁶, AD⁴⁷, multiple sclerosis (MS)⁴⁸, PD^{27,49}, and generalized and focal epilepsy (GEP/FEP)⁵⁰, and detected significant genetic correlations between volumes and PD, BD, and MS (Fig. 4).

To further examine the polygenic overlap between thalamic volumes and the ten disorders, we performed conjunctive false discovery rate (FDR) analyses, which enable detection of genetic loci shared between traits^{51–54}. Notably, we identified jointly associated loci across volumes and disorders (Fig. 5 and Supplementary Data 16), and found the largest number of overlapping loci for SCZ (66), PD (26), and BD (15), when applying a conjunctive FDR threshold of 0.05. ASD, ADHD, MD, MS, GEP, FEP, and MS had 8, 8, 17, 10, 14, 4, and 14 loci jointly associated with thalamic volumes, respectively. When using a more stringent conjunctive FDR threshold of 0.01, there were jointly associated loci for all disorders, except GEP and FEP (Supplementary Data 16).

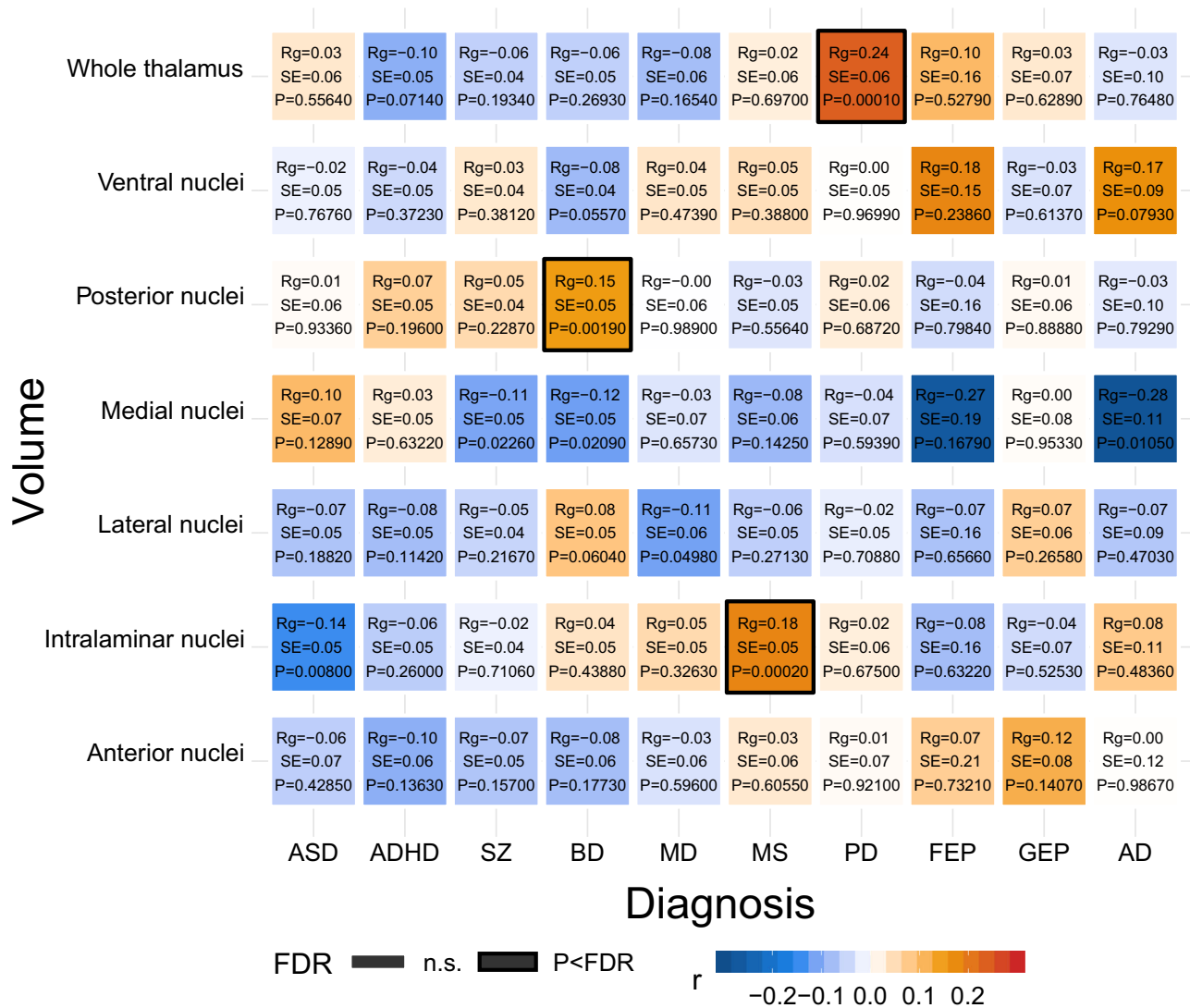


Fig. 4 LD-score regression-based genetic correlations between thalamic volumes and ten brain disorders. We assessed genetic correlations between thalamic volumes and ten brain disorders using LD-score regression. Warm and cool colors indicate positive and negative genetic associations, respectively. There were significant positive correlations between the whole thalamus and PD, between posterior nuclei and BD, and between intralaminar nuclei and MS, as indicated by a black frame, after FDR correcting across all 70 analyses (7 volumes x 10 disorders; two-sided $P < \text{FDR}$). AD; Alzheimer’s disease. ADHD; attention-deficit hyperactivity disorder. ASD, autism spectrum disorder; BD, bipolar disorder; FDR, false discovery rate; FEP, focal epilepsy; GEP, generalized epilepsy; MD, major depression; MS, multiple sclerosis; PD, Parkinson’s disease; SCZ, schizophrenia.

Further studies of how the overlapping genetic regions influence thalamus structure and disorder risk are warranted, yet several of the shared loci are noteworthy. rs2693698 was significant for SCZ and BD, and jointly associated with medial and posterior nuclei volumes. The gene nearest to rs2693698 is *BCL11B* (significant for the latter volumes also in the GWAS) and encodes a transcription factor expressed in the fetal and adult brain⁵⁵. *BCL11B* is thought to regulate neuron development and its mutations are associated with intellectual disability and neuropsychiatric disorders^{55,56}. rs5011432 was associated with medial nuclei volume in MD and AD, and the nearest gene is *TMEM106B*, which regulates lysosome functions⁵⁷. rs13107325, within *SLC39A8* as discussed above, was significant for both SCZ and PD, and associated with anterior, lateral, and intralaminar nuclei volumes.

In summary, our study provides insights into the genetic underpinnings of the human thalamus and identifies genetic loci linked to thalamic nuclei. We found that the majority of

thalamus-linked loci and genes were associated with only one of the seven volumes, which may suggest at least partly independent genetic architectures. This demonstrates the importance of targeting individual nuclei rather than studying the thalamus as a whole, in line with similar benefits observed for other brain structures^{19,20}. In addition, we found genetic correlations between the thalamic nuclei and distinct cortical regions. Notably, the positive associations were mainly found for higher-order thalamic nuclei and the cortical distributions are consistent with thalamocortical connectivity. These results are in line with emerging findings, suggesting that functionally connected brain regions exhibit stronger genetic correlations than less-connected regions^{35,36}. Finally, we identified genetic loci shared between thalamus volumes and ten neurological and psychiatric disorders. Further mechanistic studies are required to clarify how the thalamus contributes to the pathogenesis of brain disorders and the pleiotropic loci identified by our analyses could inform such experiments.

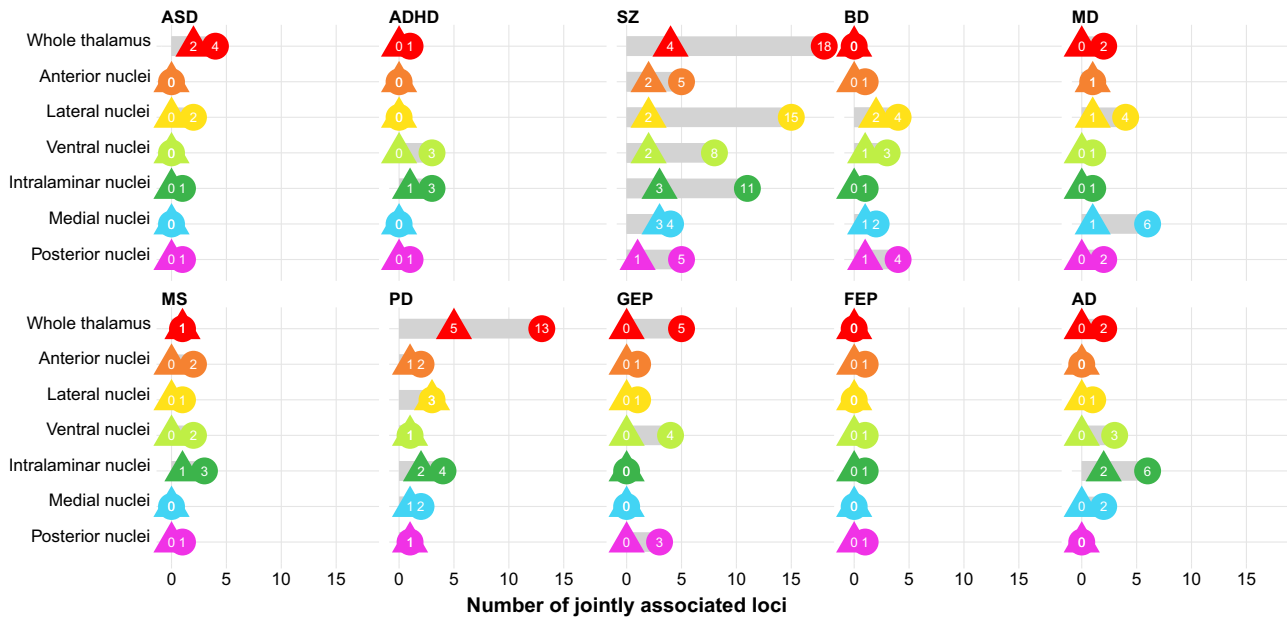


Fig. 5 Genetic loci shared between thalamic volumes and ten brain disorders. Conjunctural FDR analysis detected shared genetic loci across thalamic volumes and the ten disorders. The figure shows results with FDR thresholds of both 0.05 (circles) and 0.01 (triangles). We found the largest number of overlapping loci for SCZ (66), PD (26), and BD (15), when applying a conjunctural FDR threshold of 0.05. For ASD, ADHD, MD, MS, GEP, FEP, and MS, there were 8, 8, 17, 10, 14, 4, and 14 genetic loci jointly associated with thalamic volumes and disorders when using a threshold of 0.05, respectively. When using a conjunctural FDR threshold of 0.01, there were overlapping loci associated with thalamic volumes and SCZ (17), PD (14), BD (5), ASD (2), ADHD (1), MDD (3), MS (2), and AD (2), and no shared locus for GEP or FEP. AD, Alzheimer’s disease; ADHD, attention-deficit hyperactivity disorder; ASD, autism spectrum disorder; BD, bipolar disorder; FDR, false discovery rate; FEP, focal epilepsy; GEP, generalized epilepsy; MD, major depression; MS, multiple sclerosis; PD, Parkinson’s disease; SCZ, schizophrenia.

Methods

Samples, thalamus segmentations, and quality control procedures. We included raw T1-weighted three-dimensional brain magnetic MRI data from $n = 30,432$ genotyped White British from the UK Biobank¹⁵ in the GWAS discovery sample. The results from the discovery GWAS were also used in the analyses of genetic overlap with brain disorders and the discovery sample therefore excluded individuals with CNS diagnoses. Additional data became available to us after we had performed the main discovery analysis and we combined this data with the previously excluded data of individuals with CNS diagnoses, to maximize the sample size for a replication sample. The replication sample therefore comprised MRI and SNP data from an additional 5266 White British of the general population in the UK Biobank¹⁵, irrespective of diagnoses. The National Health Service National Research Ethics Service (ref. 11/112 NW/0382) has approved the UK Biobank and all participants gave signed informed consent before study inclusion. The projects from which summary statistics of GWAS were used for genetic overlap analysis were each approved by the local ethics committees and informed consent was obtained from all participants^{27,41–47,49,50}. The Norwegian Regional Committees for Medical and Health Research Ethics (REC South East) evaluated our pipelines that use summary statistics from published works for genetic analysis as performed in the current study and found that no additional institutional approval is needed.

The MRI data for all individuals was stored and analyzed locally at the University of Oslo. Using Bayesian thalamus segmentation based on ex-vivo MRI and histology in Freesurfer 6.0^{16,17}, we segmented the MRI data into the whole thalamus and six thalamic nuclei groups, i.e., anterior, lateral, ventral, intralaminar, medial, and posterior nuclei groups. These groups, as defined by Iglesias et al.¹⁷, include the following thalamic nuclei: anterior group—the anteriorventral nucleus; lateral group—the laterodorsal and lateral posterior nucleus; ventral group—the ventral anterior, ventral anterior magnocellular, ventral lateral anterior, ventral lateral posterior, ventral posterolateral, and ventromedial nucleus; intralaminar group—the central medial, central lateral, paracentral, centromedian, and parafascicular nucleus; medial group—the paratenial, reuniens, mediodorsal medial magnocellular, and mediodorsal lateral parvocellular nucleus; and the posterior group—the lateral geniculate, medial geniculate, limitans, pulvinar anterior, pulvinar medial, pulvinar lateral, and pulvinar inferior. The segmentations of these thalamic nuclei groups have high test–retest reliability, show agreement with histological studies of thalamic nuclei, and are robust to differences in MRI contrast¹⁷. For each nuclei group, we summed the volumes in the left and right thalamus, and these were used in all analyses.

We then manually assessed the quality and delineations of the discovery and replication sample MRI data sets by visually inspecting axial view figures of the segmentations for each participant. This procedure excluded 318 data sets from the discovery sample (due to tumors and other lesions (6%), cysts (12%), ventricle abnormalities (18%), segmentation errors (14%), and insufficient data quality (50%)) and 93 data sets from the replication sample (due to tumors and other lesions (4%), cysts (13%), ventricle abnormalities (13%), segmentation errors (9%), and insufficient data quality (61%)). Thus, the final sizes of the discovery and replication GWAS samples were $n = 30,114$ and $n = 5173$, respectively.

GWAS studies for thalamic volumes and identification of genomic loci. We performed GWAS on the thalamus volumes and genotype data from the participants in the GWAS discovery and replication samples, and due to data availability restricted the analyses to White British individuals. We used standard quality control procedures to the UK Biobank v3 imputed genetic data and removed SNPs with an imputation quality score < 0.5 , a minor allele frequency < 0.05 , missing in more than 5% of individuals, and failing the Hardy–Weinberg equilibrium tests at a $P < 1e - 6$.

GWAS was conducted for the seven thalamic volumes, i.e., volumes of the whole thalamus and the anterior, lateral, ventral, intralaminar, medial, and posterior nuclei groups, using PLINK v2.0¹⁸. All GWAS accounted for age, age-orthogonalized age-squared, sex, scanning site, intracranial volume, and the first 20 genetic principal components to control for population stratification. The resulting P -values were Bonferroni-corrected for analyses of seven volumes. In addition, the GWAS for the six thalamic nuclei groups was run, both with and without whole thalamus as an additional covariate. The main text presents results for thalamic nuclei groups when accounting for whole thalamus volume, whereas GWAS results for the nuclei volumes are provided in the Supplementary Information.

We identified genetic loci related to thalamic volumes using the FUMA platform v1.3.5²⁹. The settings and results of the FUMA analyses can be found at <https://fuma.ctglab.nl/browse> (FUMA ID 135-141). For these analyses, we used the UKB release2b White British as the reference panel. Independent significant SNPs were identified by the genome-wide significant threshold ($P < 7e - 9$) and by their independency ($r^2 \leq 0.6$ within a 1 mb window). We defined independent significant SNPs with $r^2 < 0.1$ within a 1 mb window as lead SNPs and genomic loci were identified by merging lead SNPs closer than 250 kb. All SNPs in LD ($r^2 \geq 0.6$) with one of the independent significant SNPs in the genetic loci were defined as candidate SNPs. We used the minimum r^2 to determine the borders of the genomic risk loci.

Functional annotation, gene-based association, gene-set, tissue, and cell specificity, and protein–protein interaction analyses. We functionally annotated all candidate SNPs of thalamus volumes using FUMA²⁹, which is based on information from 18 biological repositories and tools. FUMA prioritizes the most likely causal SNPs and genes by employing positional, eQTL, and chromatin interaction mapping²⁹. The platform annotates significantly associated SNPs with functional categories, combined CADD scores²², RegulomeDB scores⁵⁸, and chromatin states²⁹. A CADD score > 12.37 indicates a deleterious protein effect²², whereas the RegulomeDB score suggests the regulatory functionality of SNPs based on eQTLs and chromatin marks. Chromatin states show the genomic region's accessibility for every 200 bp with 15 categorical states predicted by ChromHMM 1.00 based on five histone modification marks (H3K4me3, H3K4me1, H3K36me3, H3K27me3, and H3K9me3) for 127 epigenomes⁵⁹. Lower scores reflect higher accessibility in the chromatin state and to a more open state. Roadmap suggests the following 15-core chromatin states: 1 = Active Transcription Start Site (TSS); 2 = Flanking Active TSS; 3 = Transcription at Gene 5' and 3'; 4 = Strong Transcription; 5 = Weak Transcription; 6 = Genic Enhancers; 7 = Enhancers; 8 = Zinc Finger Genes and Repeats; 9 = Heterochromatic; 10 = Bivalent/Poised TSS; 11 = Flanking Bivalent/Poised TSS/Enh; 12 = Bivalent Enhancer; 13 = Repressed PolyComb; 14 = Weak Repressed PolyComb; and 15 = Quiescent/Low⁶⁰.

We performed GWAS and gene-set analyses with MAGMA v1.07³⁰ in FUMA on the complete GWAS input data. We excluded the major histocompatibility complex region before running the MAGMA-based analyses. MAGMA runs multiple linear regression to obtain gene-based *P*-values and the Bonferroni-corrected significant threshold was $P = 0.05/18158 \text{ genes}/7 \text{ volumes} = 3.9e - 7$. We ran gene-set analyses with hypergeometric tests for curated gene sets and GO terms obtained from the MsigDB⁶¹.

We used LDSC-SEGs to assess the tissue and cell-type specificity of the GWAS findings (<https://github.com/bulik/ldsc/wiki/Cell-type-specific-analyses>). Following the analyses of Finucane et al.²⁸, we classified 205 tissues and cell types into nine categories for visualization, i.e., “Other,” “Musculoskeletal-connective,” “Liver,” “Endocrine,” “Digestive,” “CNS,” “Cardiovascular,” “Blood/Immune,” and “Adipose.” We used a Bonferroni-corrected significance threshold of $P = 0.05/205 \text{ tissues and cell types}/7 \text{ thalamic GWAS} = 3.5e - 5$.

We conducted protein–protein interaction analysis using STRING to explore the functional relationships between protein-encoded thalamus-linked genes and exported the network to Cytoscape 3.8.2. for further analyses. We assessed protein–protein interaction enrichment and node degrees of the resulting network.

Analyses of genetic correlation and overlap between thalamus volumes, cortical volumes, and ten brain disorders. We derived volumes of 180 cortical regions in each hemisphere based on a multimodal parcellation of the cerebral cortex³⁶. We then ran GWAS for volumes of each region summed for the two hemispheres with total cortical volume as a covariate and followed the implementation performed for the thalamus volumes. Next, we assessed genetic correlation between thalamic nuclei groups and cortical regions using LD-score regression, and adjusted the *P*-values across all performed analyses (7 thalamic volumes × 180 cortical volumes) using FDR correction in R statistics⁶².

Furthermore, we obtained GWAS summary statistics for ADHD⁴¹, ASD, SZ, BD, and MD from the Psychiatric Genomics Consortium^{42–46}, for AD from the International Genomics of Alzheimer's Project⁴⁷, for MS from the International Multiple Sclerosis Genetics Consortium⁴⁸, for PD from the International Parkinson Disease Genetics Consortium and 23andMe^{27,49}, and for GEP and FEP from The International League Against Epilepsy Consortium on Complex Epilepsies⁵⁰. We performed similar analysis of genetic correlations between volumes and disorders as described for the thalamocortical correlations. In addition, we employed conjunctive FDR statistics^{51,53,54,63,64} to assess polygenic overlap between the seven thalamic volumes and the ten brain disorders. The conjunctive FDR analyses were run using in-house software, available at <https://github.com/precimed/pleiofdr/>, MATLAB 2017a, and Python 3.7.4.

A review and mathematical description of this method when applied to neurological and psychiatric disorders were recently published by Smeland et al.⁵⁴. In brief, the conjunctive FDR is defined by the maximum of the two conditional FDR values for a specific SNP. The method calculates a posterior probability that an SNP is null for either trait or both at the same time, given that the *P*-values for both phenotypes are as small, or smaller, than the *P*-values for each trait individually. The empirical null distribution in GWASs is affected by global variance inflation and all *P*-values were therefore corrected for inflation using a genomic inflation control procedure. All analyses were performed after excluding SNPs in the major extended histocompatibility complex (hg19 location chromosome 6: 25119106–33854733) and 8p23.1 regions (hg19 location chromosome 8: 7242715–12483982) for all cases, and *MAPT* and *APOE* regions for PD and AD, respectively, as complex correlations in regions with intricate LD can bias the FDR estimation.

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

Data availability

Data used in this study were obtained from the UK Biobank, from the Psychiatric Genomics Consortium, 23andMe, the International Genomics of Alzheimer's Project, the International Multiple Sclerosis Genetics Consortium, the International Parkinson Disease Genetics Consortium, and from The International League Against Epilepsy Consortium on Complex Epilepsies. The PD GWAS partly included data from the 23andMe Consortium and is available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Please visit research.23andme.com/collaborate/#publication for more information and to apply to access the data. GWAS results are also available on the FUMA website; ID 135-141. The summary statistics for thalamic volumes of the present study are publicly available on GitHub.

Code availability

This study used openly available software and code, specifically Freesurfer, Plink, LD-score regression, LDSC-SEG, Cytoscape, and conjunctive FDR.

Received: 2 October 2020; Accepted: 16 April 2021;

Published online: 18 May 2021

References

- Acsády, L. The thalamic paradox. *Nat. Neurosci.* **20**, 901–902 (2017).
- Nakajima, M. & Halassa, M. M. Thalamic control of functional cortical connectivity. *Curr. Opin. Neurobiol.* **44**, 127–131 (2017).
- Hwang, K., Bertolero, M. A., Liu, W. B. & D'Esposito, M. The human thalamus is an integrative hub for functional brain networks. *J. Neurosci.* **37**, 5594–5607 (2017).
- Sherman, S. M. Thalamus plays a central role in ongoing cortical functioning. *Nat. Neurosci.* **19**, 533–541 (2016).
- Guo, Z. V. et al. Maintenance of persistent activity in a frontal thalamocortical loop. *Nature* **545**, 181–186 (2017).
- Bolkan, S. S. et al. Thalamic projections sustain prefrontal activity during working memory maintenance. *Nat. Neurosci.* **20**, 987–996 (2017).
- Schmitt, L. I. et al. Thalamic amplification of cortical connectivity sustains attentional control. *Nature* **545**, 219–223 (2017).
- Arcaro, M. J., Pinsk, M. A., Chen, J. & Kastner, S. Organizing principles of pulvino-cortical functional coupling in humans. *Nat. Commun.* **9**, 5382 (2018).
- Aggleton, J. P., Pralus, A., Nelson, A. J. & Hornberger, M. Thalamic pathology and memory loss in early Alzheimer's disease: moving the focus from the medial temporal lobe to Papez circuit. *Brain* **139**, 1877–1890 (2016).
- Parker, P. R., Lalive, A. L. & Kreitzer, A. C. Pathway-specific remodeling of thalamostriatal synapses in Parkinsonian mice. *Neuron* **89**, 734–740 (2016).
- Woodward, N. D., Karbasforoushan, H. & Heckers, S. Thalamocortical dysconnectivity in schizophrenia. *Am. J. Psychiatry* **169**, 1092–1099 (2012).
- Hibar, D. P. et al. Subcortical volumetric abnormalities in bipolar disorder. *Mol. Psychiatry* **21**, 1710–1716 (2016).
- Zhao, B. et al. Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. *Nat. Genet.* **51**, 1637–1644 (2019).
- Satizabal, C. L. et al. Genetic architecture of subcortical brain structures in 38,851 individuals. *Nat. Genet.* **51**, 1624–1636 (2019).
- 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).
- Fischl, B. et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341–355 (2002).
- Iglesias, J. E. et al. A probabilistic atlas of the human thalamic nuclei combining ex vivo MRI and histology. *Neuroimage* **183**, 314–326 (2018).
- Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- van der Meer, D., Rokicki, J., Kaufmann, T., Córdova-Palomera, A. & Moberget, T. Brain scans from 21,297 individuals reveal the genetic architecture of hippocampal subfield volumes. *Mol. Psychiatry*, <https://doi.org/10.1038/s41380-41018-40262-41387> (2018).
- Elvsashagen, T. et al. The genetic architecture of human brainstem structures and their involvement in common brain disorders. *Nat. Commun.* **11**, 4016 (2020).

21. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
22. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
23. Lee, J. J. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
24. Pickrell, J. K. et al. Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48**, 709–717 (2016).
25. Lasagna-Reeves, C. A. et al. Reduction of Nuak1 decreases tau and reverses phenotypes in a tauopathy mouse model. *Neuron* **92**, 407–418 (2016).
26. Wang, Y. & Mandelkow, E. Tau in physiology and pathology. *Nat. Rev. Neurosci.* **17**, 22 (2015).
27. Chang, D. et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat. Genet.* **49**, 1511–1516 (2017).
28. Finucane, H. K. & Reshef, Y. A. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
29. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
30. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
31. Lam, M. et al. Large-scale cognitive GWAS meta-analysis reveals tissue-specific neural expression and potential nootropic drug targets. *Cell Rep.* **21**, 2597–2613 (2017).
32. Zeng, L. et al. Identification of a novel human doublecortin-domain-containing gene (DCDC1) expressed mainly in testis. *J. Hum. Genet.* **48**, 393–396 (2003).
33. Sherman, S. M. & Guillery, R. W. *Exploring the Thalamus and its Role in Cortical Function* 2nd edn (MIT Press, Boston, 2006).
34. Hawrylycz, M. et al. Canonical genetic signatures of the adult human brain. *Nat. Neurosci.* **18**, 1832–1844 (2015).
35. Richiardi, J. et al. BRAIN NETWORKS. Correlated gene expression supports synchronous activity in brain networks. *Science* **348**, 1241–1244 (2015).
36. Glasser, M. F. et al. A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
37. Bezdudnaya, T. & Keller, A. Laterodorsal nucleus of the thalamus: a processor of somatosensory inputs. *J. Comp. Neurol.* **507**, 1979–1989 (2008).
38. Conte, W. L., Kamishina, H., Corwin, J. V. & Reep, R. L. Topography in the projections of lateral posterior thalamus with cingulate and medial agranular cortex in relation to circuitry for directed attention and neglect. *Brain Res.* **1240**, 87–95 (2008).
39. Wang, Z. et al. Community-informed connectomics of the thalamocortical system in generalized epilepsy. *Neurology* **93**, e1112–e1122 (2019).
40. Woodward, N. D., Giraldo-Chica, M., Rogers, B. & Cascio, C. J. Thalamocortical dysconnectivity in autism spectrum disorder: an analysis of the Autism Brain Imaging Data Exchange. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **2**, 76–84 (2017).
41. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
42. Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol. Autism* **8**, 21 (2017).
43. Stahl, E. A. et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).
44. Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
45. Hyde, C. L. et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat. Genet.* **48**, 1031–1036 (2016).
46. Ripke, S. et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
47. Lambert, J. C. et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
48. Sawcer, S. et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214–219 (2011).
49. 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).
50. International League Against Epilepsy Consortium on Complex Epilepsies. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. *Nat. Commun.* **9**, 5269 (2018).
51. Andreassen, O. A. et al. Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. *Am. J. Hum. Genet.* **92**, 197–209 (2013).
52. Liu, J. Z. et al. Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. *Nat. Genet.* **45**, 670–675 (2013).
53. Schork, A. J., Wang, Y., Thompson, W. K., Dale, A. M. & Andreassen, O. A. New statistical approaches exploit the polygenic architecture of schizophrenia-implications for the underlying neurobiology. *Curr. Opin. Neurobiol.* **36**, 89–98 (2016).
54. Smeland, O. B. & Frei, O. Discovery of shared genomic loci using the conditional false discovery rate approach. *Hum. Genet.* **139**, 85–94 (2020).
55. Lennon, M. J., Jones, S. P., Lovelace, M. D., Guillemin, G. J. & Brew, B. J. Bcl11b-A critical neurodevelopmental transcription factor-roles in health and disease. *Front. Cell Neurosci.* **11**, 89 (2017).
56. Lessel, D. et al. BCL11B mutations in patients affected by a neurodevelopmental disorder with reduced type 2 innate lymphoid cells. *Brain* **141**, 2299–2311 (2018).
57. Tropea, T. F. et al. TMEM106B effect on cognition in Parkinson disease and frontotemporal dementia. *Ann. Neurol.* **85**, 801–811 (2019).
58. Boyle, A. P. et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012).
59. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
60. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* **12**, 2478–2492 (2017).
61. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
62. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2013).
63. Andreassen, O. A. et al. Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate. *PLoS Genet.* **9**, e1003455 (2013).
64. Nichols, T., Brett, M., Andersson, J., Wager, T. & Poline, J. B. Valid conjunction inference with the minimum statistic. *Neuroimage* **25**, 653–660 (2005).

Acknowledgements

We were funded by the South-Eastern Norway Regional Health Authority 2015-078 (T.E.); 2013-123, 2017-112, and 2019-108 (O.A.A.); 2014-097, 2015-073, and 2016-083 (L.T.W.); by the Research Council of Norway (276082 (T.K.); 262656, 248778, 223273, and 273291 (O.A.A.); 204966 (L.T.W.); EEA 2014-2021, # 6/2019 (O.A.A.); 249795 and 273345 (L.T.W.); 283798 SYNCHIZ (O.A.A.); Stiftelsen Kristian Gerhard Jebsen, the European Research Council (ERCStG 802998 BRAINMINT (L.T.W.)); NVIDIA Corporation GPU Grant (T.K.); 248828 Centre for Digital Life Norway (O.A.A.); the Ebbe Frøland foundation (T.E.); a research grant from Mrs. Throne-Holst (T.E.); and H2020 (# 847776) (O.A.A.). This research has been conducted using the UK Biobank Resource (access code 27412, <https://www.ukbiobank.ac.uk/>) and we thank the research participants and employees of the UK Biobank. We also thank the ADHD, ASD, SCZ, BD, and MD Working Groups of the Psychiatric Genomics Consortium, the International Genomics of Alzheimer's Project, the International Multiple Sclerosis Genetics Consortium, International Parkinson Disease Genomics Consortium, The International League Against Epilepsy Consortium on Complex Epilepsies, and 23andMe, Inc. for granting us access to their GWAS summary statistics, and the many people who provided DNA samples for their studies. This work was performed on the TSD (Tjeneste for Sensitive Data) facilities, owned by the University of Oslo, operated and developed by the TSD service group at the University of Oslo, IT-Department (USIT), and on resources provided by UNINETT Sigma2—the National Infrastructure for High Performance Computing and Data Storage in Norway.

Author contributions

T.E. and T.K. conceived the study, performed quality control, and analyzed and interpreted the data. A.S., O.F., D.v.d.M., S.B., O.S., V.J.K., L.T.W., and O.A.A. contributed with expertise on processing pipelines, conceptual feedback on analyses scopes, and/or interpretation of results. T.E. and T.K. wrote the first draft of the paper, and all authors contributed to and approved the final manuscript.

Competing interests

T.E. is a consultant to BrainWaveBank and received speaker's honoraria from Lundbeck and Janssen Cilag. O.A.A. is a consultant to BrainWaveBank and HealthLytix, and received speaker's honoraria from Lundbeck. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-23175-z>.

Correspondence and requests for materials should be addressed to T.E. or T.K.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021