Genome-wide association analysis of the human thalamus identifies novel genomic loci and reveals genetic overlap with distinct cortical regions and ten brain disorders

Torbjørn Elvsåshagen, MD, PhD[1,2,3]; Alexey Shadrin, PhD[1,3]; Oleksandr Frei, PhD[1,3]; Dennis van der Meer, PhD[1,4]; Shahram Bahrami, PhD[1,3]; Vinod Jangir Kumar, PhD[5]; Olav Smeland, MD, PhD[1,3]; Lars T. Westlye, PhD[1,6]; Ole A. Andreassen, MD, PhD[1,3]; Tobias Kaufmann, PhD[1,3]

[1] NORMENT, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway
[2] Department of Neurology, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway
[3] Institute of Clinical Medicine, University of Oslo, Oslo, Norway
[4] School of Mental Health and Neuroscience, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands
[5] Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[6] Department of Psychology, University of Oslo, Oslo, Norway

@ Corresponding authors:
Torbjørn Elvsåshagen, MD, PhD & Tobias Kaufmann, PhD
E-mails: torbjorn.elvsashagen@medisin.uio.no, tobias.kaufmann@medisin.uio.no
Postal address: Norwegian Centre for Mental Disorders Research, Oslo University Hospital, PoBox 4956 Nydalen, Norway
Telephone: +47 23 02 73 50, Fax: +47 23 02 73 33

Counts:
Introduction: 169 words
Main text body: 1529 words
References: 56 in the main text
Figures: 5 in the main text

Keywords:
Thalamus, thalamic nuclei, GWAS, neurological disorders, psychiatric disorders

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[1-8]. Structural and functional thalamic alterations are involved in the pathogenesis of common brain disorders[9-12], yet the genetic architecture of the thalamus remains largely unknown. Here, using brain scans and genotype data from 30,114 individuals, we identified 42 (41 novel) genetic loci and 392 genes associated with volumes of the thalamus and its nuclei. In an independent validation sample (*n* = 5,173) 96% of the loci showed the same effect direction (sign test, *P* = 8.6e-14). We mapped the genetic relationship between thalamic nuclei and 180 cerebral cortical areas and found overlapping genetic architectures consistent with thalamocortical connectivity. Pleiotropy analyses between thalamic volumes and ten psychiatric and neurological disorders revealed shared variants for all disorders. Together, these analyses identify the first genetic loci linked to thalamic nuclei and substantiate the emerging view of the thalamus having central roles in cortical functioning and common brain disorders.

Recent studies indicated that the thalamus has a broader role in cognition than previously assumed[1,2]. Cognitive neuroscience is therefore shifting focus to how the thalamus regulates cortical activity and supports higher-order functions such as working memory[6], attentional control[7], and visual processing[8]. There is also a growing appreciation of thalamic contributions to the pathogenesis of neurological and psychiatric disorders, including dementias[9], 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 remains 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 one[13] and two[14] genetic loci associated with whole thalamus volume, yet there is no

genome-wide association study (GWAS) of thalamic nuclei. To further identify the genetic architectures of the thalamus and its nuclei and to assess their relationships with the cerebral cortex and brain disorders, we accessed brain magnetic resonance imaging (MRI) data from n = 30,432 genotyped white British from the UK Biobank[15]. The MRI data was 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 PLINK[18] on whole thalamus and the six nuclei volumes accounting for age, age-orthogonalized age squared, sex, scanning site, intracranial volume, and the first twenty 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 LD score regression[21] was 25% for whole thalamus and 18% to 32% for the six nuclei groups (Fig. 1b). We found genome-wide significant hits for all seven volumes (Bonferroni-corrected $P < 5e\text{-}8/7 = 7e\text{-}9$) and identified a total of 55 lead SNPs in 50 genomic loci (Figs. 1c-d and Supplementary Table 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 volumes, respectively. Eight loci were associated with more than one volume, resulting in 42 unique thalamus-associated loci. One of these (rs76928645 on chromosome 7) was associated

with whole thalamus volume in a recent study[13], whereas the others are novel. Notably, 37 of the 42 genetic loci were associated with only one volume.

Three lead SNPs had combined annotation-dependent depletion (CADD) score > 15, which indicates deleterious protein effects[22]. 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 rs12146713 on chromosome 12 (associated with medial nuclei volume; score=19.6) is *NUAK1*, which regulates Tau protein level[25]. Cerebral Tau accumulation is a defining characteristic of Alzheimer's disease (AD) and other neurodegenerative disorders[26]. rs951366 on chromosome 1 was significant for posterior nuclei volume (score=15.7) and is associated with PD[27]. Further GWAS results are provided in Supplementary Figs. 1-4 and Supplementary Tables 2-8.

For GWAS replication, we used MRI and SNP data from an additional 5,173 white British (51% females; age range 46-81 years) from the UK Biobank[15]. We found that 96% of the lead SNPs from the discovery GWAS had the same effect direction in the replication (sign test; P = 8.6e-14). Moreover, 58% of the discovery lead SNPs had uncorrected $P < 0.05$ in the replication (Supplementary Table 9).

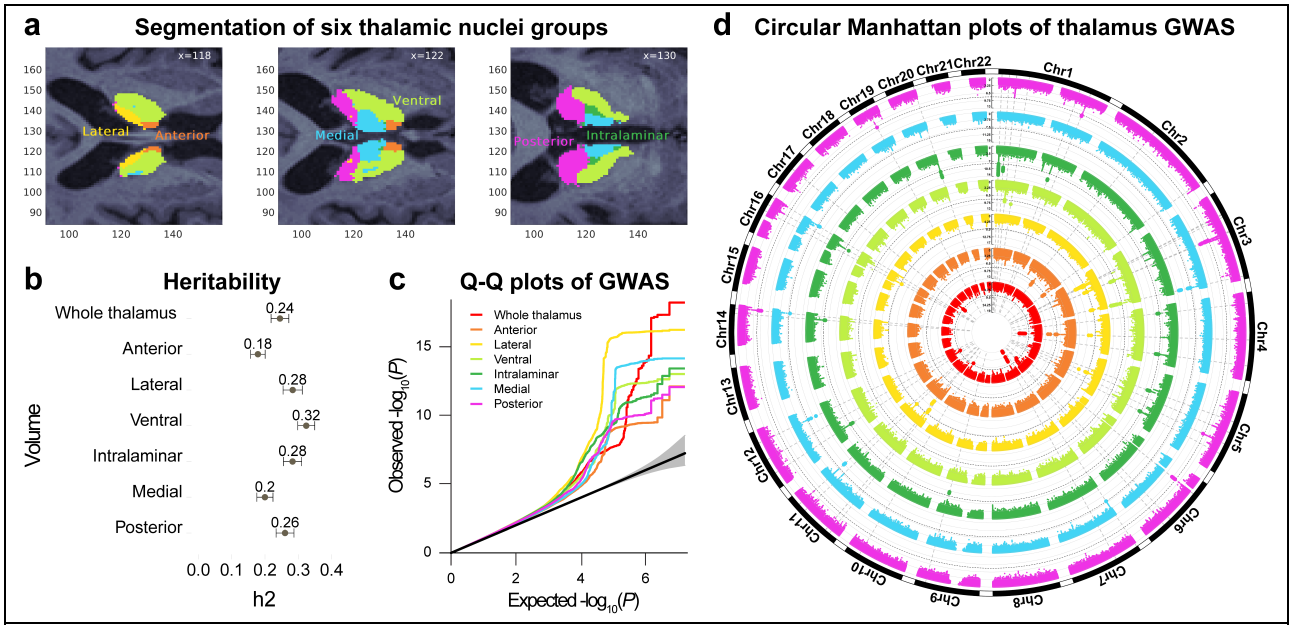

Fig. 1 | Analysis of the GWAS discovery sample identifies 42 loci associated with thalamic volumes. a The thalamus was segmented into six nuclei groups – anterior, lateral, ventral, intralaminar, medial, and posterior nuclei – using Bayesian thalamus segmentation[16,17]. All data sets were assessed by visually inspecting axial view figures of the segmentations for each participant and we removed sets with segmentation errors, insufficient data quality, and pathologies. **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. Error bars, s.e. **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. GWAS; genome-wide association studies.

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)[28] to map candidate SNPs to genes. This identified 336 unique genes across the seven volumes (Supplementary Table 10 and Supplementary Fig. 5). We conducted genome-wide gene-based association analyses (GWGAS) using MAGMA[29] and

detected 126 unique genes across the thalamic volumes (Fig. 2, Supplementary Fig. 6, and Supplementary Table 11). 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 disorders[26], and is linked to general cognitive ability[30]. The most strongly associated GWGAS gene for the thalamic nuclei was *DCDC1*. *DCDC1* is a member of the doublecortin gene family with high expression levels in the fetal brain[31], yet its functions remain largely unknown.

Fifty-six of the GWGAS genes were not mapped by the GWAS analyses, resulting in a total of 392 thalamus-linked genes. Notably, 95% of these were associated with only one volume. We then conducted gene-set analyses using MAGMA[29] and found significant Gene Ontology sets for the intralaminar nuclei, implicating cell-cell adhesion and calcium ion binding, and for the lateral nuclei, involving cell growth and synapse organization (Supplementary Table 12).

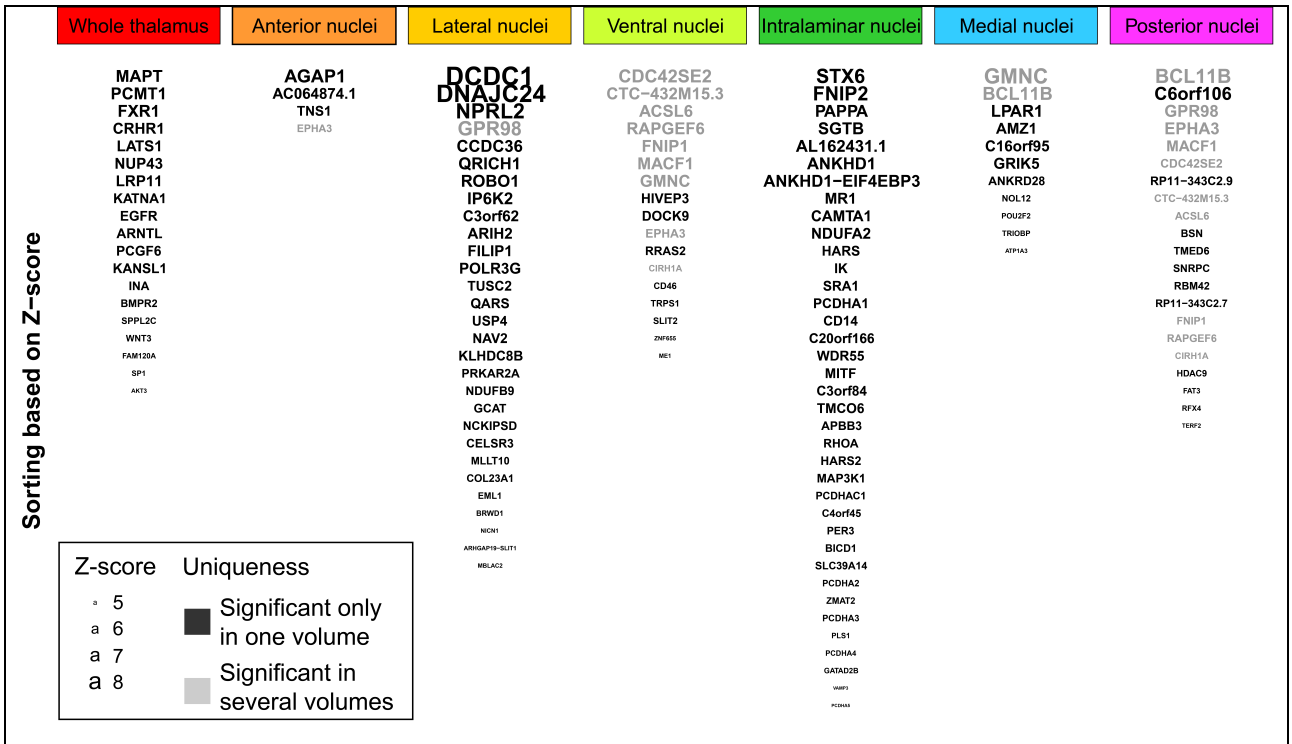

Fig. 2 | GWGAS identifies 126 unique genes associated with thalamic volumes. Nineteen genes were associated with whole thalamus, 4, 29, and 17 genes were associated with volumes of the anterior, lateral, and ventral nuclei, and 37, 11, and 21 genes

were associated with intralaminar, medial, and posterior nuclei volumes, respectively. Larger font size and higher position for gene names indicate greater *Z*-score. Black font designates genes that were significantly associated with only one volume, whereas grey font indicates genes associated with more than one volume. Additional results of the GWGAS are presented in Supplementary Table 11. GWGAS; genome-wide gene-based association analysis.

The above analyses identified the first genetic loci linked to higher-order thalamic nuclei, which project to distinct cortical regions and support cognition[1,4,32]. For example, medial thalamic nuclei are densely interconnected with prefrontal and temporal cortices and regulate working memory and attentional control[6,7], whereas posterior thalamic nuclei project to occipital cortices and support visual processing[8]. Recent studies suggest that well-connected brain regions exhibit stronger genetic correlations than less-connected regions[33,34], 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 recently[35] 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; Supplementary Table 13). Interestingly, we found positive correlations mainly for higher-order thalamic nuclei with cortical distributions consistent with established thalamocortical projections 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[36,37].

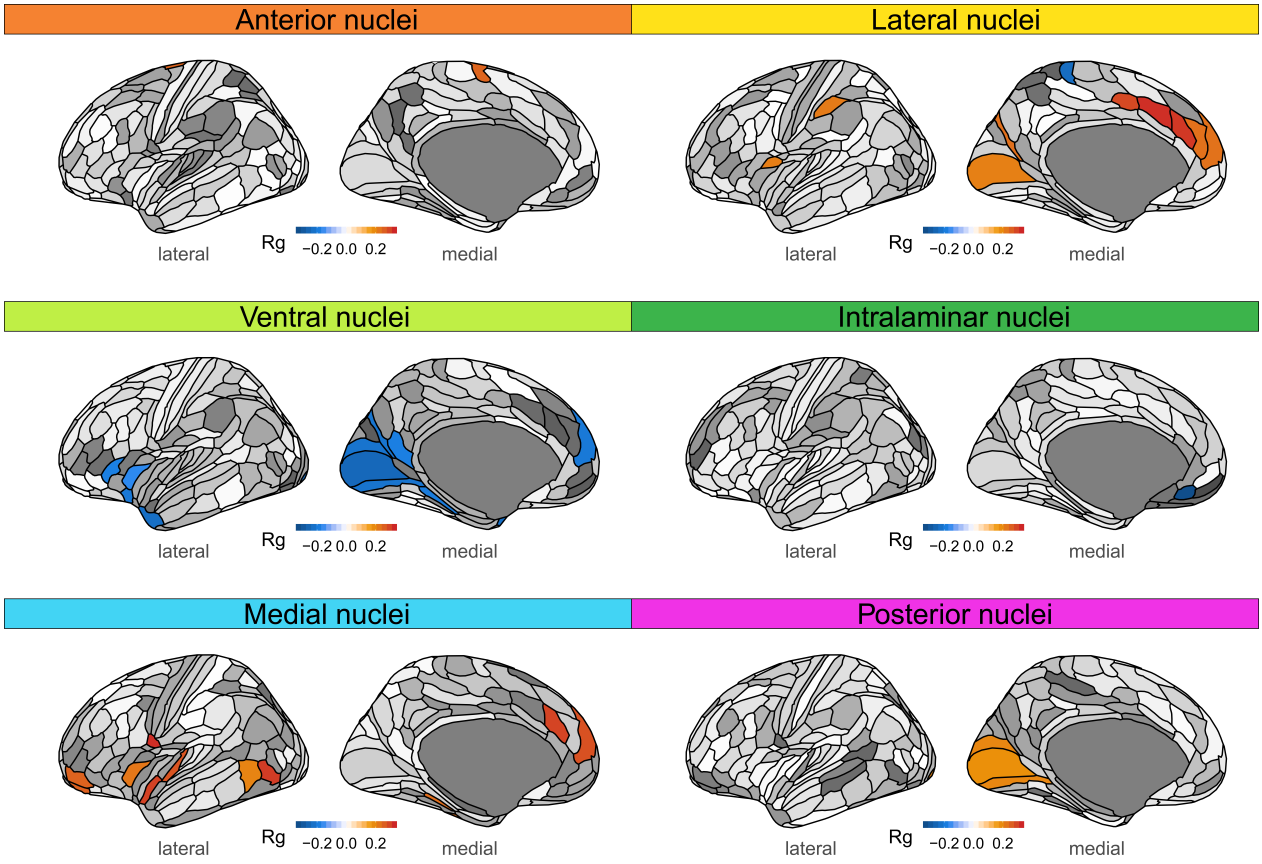

Fig. 3 | Thalamocortical genetic relationships. We found significant genetic correlations between all thalamic nuclei and distinct cortical regions. There were significant positive genetic correlations between anterior nuclei and medial premotor cortex, between lateral nuclei and mainly medial prefrontal, 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 (P < FDR). Corresponding statistics are provided in Supplementary Table 13. Rg; genetic correlation. FDR; false discovery rate.

Thalamic alterations are reported in a growing number of psychiatric and neurological disorders[9-12,38,39], yet the genetic relationships between the thalamus and disorders have not been clarified. We used GWAS summary statistics for attention-deficit/hyperactivity disorder (ADHD)[40], autism spectrum disorder (ASD)[41], BD[42], major depression (MD)[43,44], SCZ[45], AD[46],

multiple sclerosis (MS)[47], PD[27,48], and generalized and focal epilepsy (GEP/FEP)[49] and detected significant genetic correlations between volumes and PD, BD, and MS (Fig. 4).

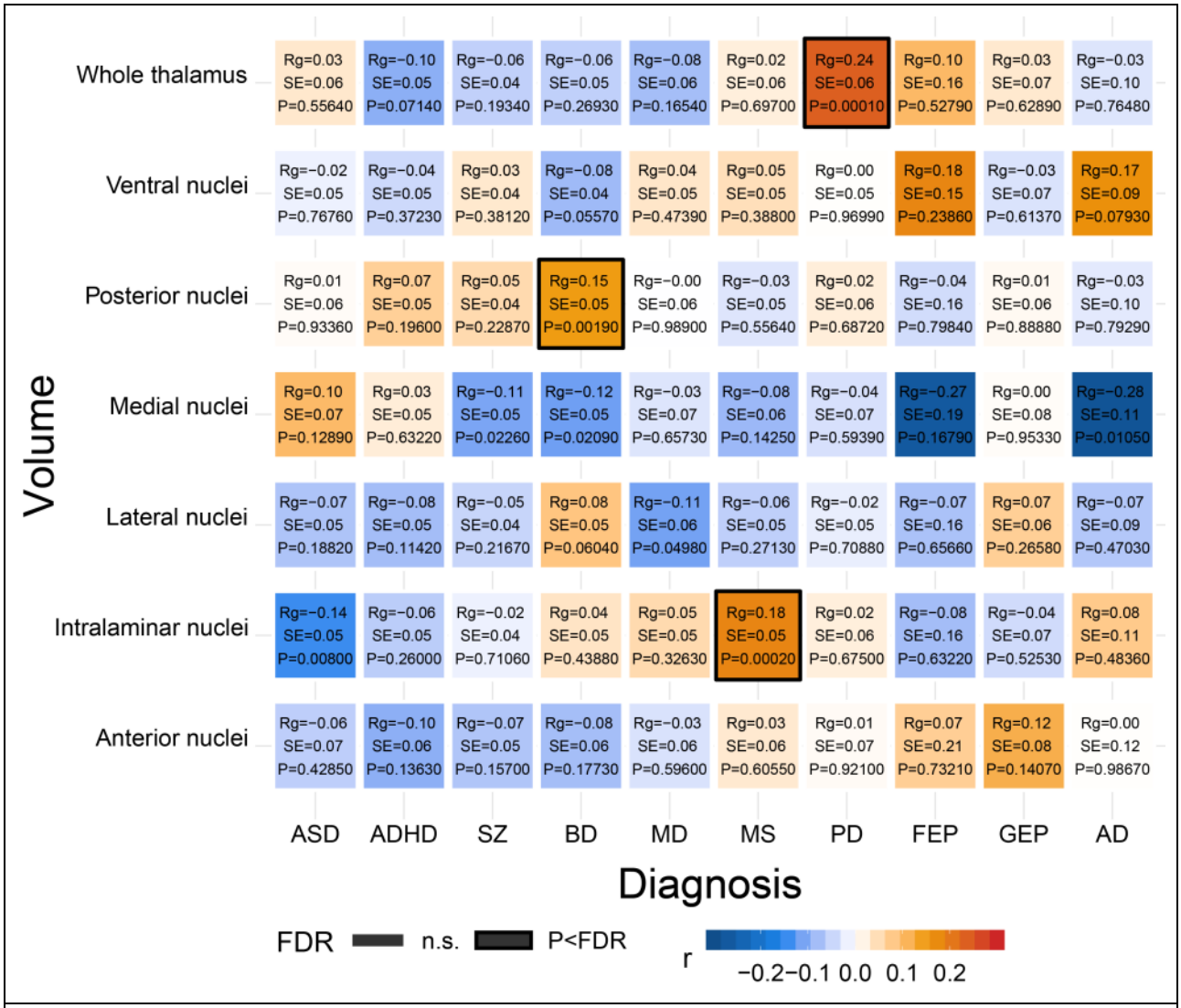

Fig. 4 | LD-score regression based genetic correlations between thalamic volumes and ten brain disorders. There were significant positive correlations between the whole thalamus and PD, between posterior nuclei and BD, and between intralaminar nuclei and MS, after FDR-correcting across all 70 analyses (7 volumes x 10 disorders). AD; Alzheimer's disease. ADHD; attention deficit hyperactivity disorder. ASD; autism spectrum disorder. BD; bipolar disorder. FEP; focal epilepsy. GEP; generalized epilepsy. MD; major depression. MS; multiple sclerosis. PD; Parkinson's disease. SCZ; schizophrenia. FDR; false discovery rate.

To further examine the polygenic overlap between thalamic volumes and the ten disorders, we performed conjunctional FDR analyses, which enable detection of genetic loci shared between traits[50-53]. Notably, we identified jointly associated loci across volumes and disorders (Fig. 5; Supplementary Table 14) and found the largest number of overlapping loci for SCZ (66), PD (26), and BD (15), when applying a conjunctional 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 conjunctional FDR threshold of 0.01, there were jointly associated loci for all disorders, except GEP and FEP (Supplementary Table 14).

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 rs2693698 is *BCL11B* (significant for the latter volumes also in the GWGAS) and encodes a transcription factor expressed in the fetal and adult brain[54]. *BCL11B* is thought to regulate neuron development and its mutations are associated with intellectual disability and neuropsychiatric disorders[54,55]. rs5011432 was associated with medial nuclei volume in MD and AD and the nearest gene is *TMEM106B*, which regulates lysosome functions[56]. rs13107325, within *SLC39A8* as discussed above, was significant for both SCZ and PD and associated with anterior, lateral, and intralaminar nuclei volumes.

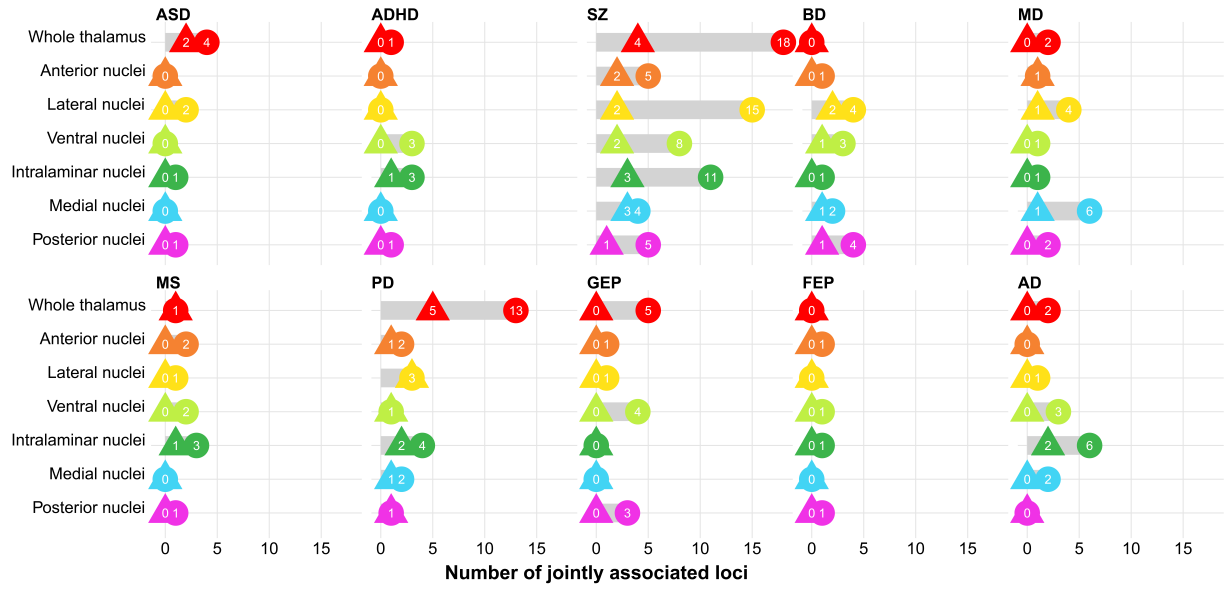

Fig. 5 | Genetic loci shared between thalamic volumes and ten brain disorders. Conjunctional FDR analysis detected shared genetic loci across thalamic volumes and the 10 disorders. The figure shows results with FDR thresholds of both 0.05 (circles) and 0.01 (triangles). We found largest number of overlapping loci for SCZ (66), PD (26), and BD (15), when applying a conjunctional 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 conjunctional 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. FDR; false discovery rate. AD; Alzheimer's disease. ADHD; attention deficit hyperactivity disorder. ASD; autism spectrum disorder. BD; bipolar disorder. FEP; focal epilepsy. GEP; generalized epilepsy. MD; major depression. MS; multiple sclerosis. PD; Parkinson's disease. SCZ; schizophrenia.

In summary, our study provides new insights into the genetic underpinnings of the human thalamus and identifies the first 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 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 findings 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 the first 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.

References

1. Acsády L. The thalamic paradox. *Nat Neurosci* 2017; **20:** 901-902.

2. Nakajima M, Halassa MM. Thalamic control of functional cortical connectivity. *Curr Opin Neurobiol* 2017; **44:** 127-131.

3. Hwang K, Bertolero MA, Liu WB, D'Esposito M. The Human Thalamus Is an Integrative Hub for Functional Brain Networks. *J Neurosci* 2017; **37:** 5594-5607.

4. Sherman SM. Thalamus plays a central role in ongoing cortical functioning. *Nat Neurosci* 2016; **19:** 533-541.

5. Guo ZV, Inagaki HK, Daie K, Druckmann S, Gerfen CR, Svoboda K. Maintenance of persistent activity in a frontal thalamocortical loop. *Nature* 2017; **545:** 181-186.

6. Bolkan SS, Stujenske JM, Parnaudeau S, Spellman TJ, Rauffenbart C, Abbas AI *et al.* Thalamic projections sustain prefrontal activity during working memory maintenance. *Nat Neurosci* 2017; **20:** 987-996.

7. Schmitt LI, Wimmer RD, Nakajima M, Happ M, Mofakham S, Halassa MM. Thalamic amplification of cortical connectivity sustains attentional control. *Nature* 2017; **545:** 219-223.

8. Arcaro MJ, Pinsk MA, Chen J, Kastner S. Organizing principles of pulvino-cortical functional coupling in humans. *Nat Commun* 2018; **9:** 5382.

9. Aggleton JP, Pralus A, Nelson AJ, Hornberger M. Thalamic pathology and memory loss in early Alzheimer's disease: moving the focus from the medial temporal lobe to Papez circuit. *Brain* 2016; **139:** 1877-1890.

10. Parker PR, Lalive AL, Kreitzer AC. Pathway-Specific Remodeling of Thalamostriatal Synapses in Parkinsonian Mice. *Neuron* 2016; **89:** 734-740.

11. Woodward ND, Karbasforoushan H, Heckers S. Thalamocortical dysconnectivity in schizophrenia. *Am J Psychiatry* 2012; **169:** 1092-1099.

12. Hibar DP, Westlye LT, van Erp TG, Rasmussen J, Leonardo CD, Faskowitz J *et al.* Subcortical volumetric abnormalities in bipolar disorder. *Mol Psychiatry* 2016; **21:** 1710-1716.

13. Zhao B, Luo T, Li T, Li Y, Zhang J, Shan Y *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* 2019; **51:** 1637-1644.

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

15. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J *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* 2015; **12:** e1001779.

16. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C *et al.* Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 2002; **33:** 341-355.

17. Iglesias JE, Insausti R, Lerma-Usabiaga G, Bocchetta M, Van Leemput K, Greve DN *et al.* A probabilistic atlas of the human thalamic nuclei combining ex vivo MRI and histology. *Neuroimage* 2018; **183:** 314-326.

18. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* 2007; **81:** 559-575.

19. 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* 2018; doi: 10.1038/s41380-41018-40262-41387.

20. Elvsashagen T, Bahrami S, van der Meer D, Agartz I, Alnaes D, Barch DM *et al.* The genetic architecture of human brainstem structures and their involvement in common brain disorders. *Nat Commun* 2020; **11:** 4016.

21. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR *et al.* An atlas of genetic correlations across human diseases and traits. *Nature genetics* 2015; **47:** 1236-1241.

22. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014; **46:** 310-315.

23. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* 2018; **50:** 1112-1121.

24. Pickrell JK, Berisa T, Liu JZ, Segurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet* 2016; **48:** 709-717.

25. Lasagna-Reeves CA, de Haro M, Hao S, Park J, Rousseaux MW, Al-Ramahi I *et al.* Reduction of Nuak1 Decreases Tau and Reverses Phenotypes in a Tauopathy Mouse Model. *Neuron* 2016; **92:** 407-418.

26. Wang Y, Mandelkow E. Tau in physiology and pathology. *Nature Reviews Neuroscience* 2015; **17:** 22.

27. Chang D, Nalls MA, Hallgrimsdottir IB, Hunkapiller J, van der Brug M, Cai F *et al.* A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet* 2017; **49:** 1511-1516.

28. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 2017; **8:** 1826.

29. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 2015; **11:** e1004219.

30. Lam M, Trampush JW, Yu J, Knowles E, Davies G, Liewald DC *et al.* Large-Scale Cognitive GWAS Meta-Analysis Reveals Tissue-Specific Neural Expression and Potential Nootropic Drug Targets. *Cell reports* 2017; **21:** 2597-2613.

31. Zeng L, Gu S, Li Y, Zhao E, Xu J, Ye X *et al.* Identification of a novel human doublecortin-domain-containing gene (DCDC1) expressed mainly in testis. *J Hum Genet* 2003; **48:** 393-396.

32. Sherman SM, Guillery RW. *Exploring the thalamus and its role in cortical function, 2nd ed*. MIT Press: Cambridge, MA, US, 2006, xxi, 484-xxi, 484pp.

33. Hawrylycz M, Miller JA, Menon V, Feng D, Dolbeare T, Guillozet-Bongaarts AL *et al.* Canonical genetic signatures of the adult human brain. *Nat Neurosci* 2015; **18:** 1832-1844.

34. Richiardi J, Altmann A, Milazzo AC, Chang C, Chakravarty MM, Banaschewski T *et al.* BRAIN NETWORKS. Correlated gene expression supports synchronous activity in brain networks. *Science* 2015; **348:** 1241-1244.

35. Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* 2016; **536:** 171-178.

36. Bezdudnaya T, Keller A. Laterodorsal nucleus of the thalamus: A processor of somatosensory inputs. *J Comp Neurol* 2008; **507:** 1979-1989.

37. Conte WL, Kamishina H, Corwin JV, Reep RL. 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* 2008; **1240:** 87-95.

38. Wang Z, Larivière S, Xu Q, Vos de Wael R, Hong S-J, Wang Z *et al.* Community-informed connectomics of the thalamocortical system in generalized epilepsy. *Neurology* 2019; **93:** e1112-e1122.

39. Woodward ND, Giraldo-Chica M, Rogers B, Cascio CJ. Thalamocortical dysconnectivity in autism spectrum disorder: An analysis of the Autism Brain Imaging Data Exchange. *Biol Psychiatry Cogn Neurosci Neuroimaging* 2017; **2:** 76-84.

40. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature genetics* 2019; **51:** 63-75.

41. 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. *Molecular autism* 2017; **8:** 21.

42. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature genetics* 2019; **51:** 793-803.

43. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature genetics* 2018; **50:** 668-681.

44. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR *et al.* Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature genetics* 2016; **48:** 1031-1036.

45. Ripke S, Neale B, Corvin A, Walters J, Farh K, Holmans P *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511:** 421-427.

46. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013; **45:** 1452-1458.

47. Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L *et al.* Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 2011; **476:** 214-219.

48. Nalls MA, Pankratz N, Lill CM, Do CB, Hernandez DG, Saad M *et al.* Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nature genetics* 2014; **46:** 989-993.

49. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. *Nat Commun* 2018; **9:** 5269.

50. Andreassen OA, Djurovic S, Thompson WK, Schork AJ, Kendler KS, O'Donovan MC *et al.* Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. *Am J Hum Genet* 2013; **92:** 197-209.

51. Liu JZ, Hov JR, Folseraas T, Ellinghaus E, Rushbrook SM, Doncheva NT *et al.* Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. *Nat Genet* 2013; **45:** 670-675.

52. Schork AJ, Wang Y, Thompson WK, Dale AM, Andreassen OA. New statistical approaches exploit the polygenic architecture of schizophrenia--implications for the underlying neurobiology. *Current opinion in neurobiology* 2016; **36:** 89-98.

53. Smeland OB, Frei O. Discovery of shared genomic loci using the conditional false discovery rate approach. *Hum Genet* 2020; **139:** 85-94.

54. Lennon MJ, Jones SP, Lovelace MD, Guillemin GJ, Brew BJ. Bcl11b-A Critical Neurodevelopmental Transcription Factor-Roles in Health and Disease. *Front Cell Neurosci* 2017; **11:** 89.

55. Lessel D, Gehbauer C, Bramswig NC, Schluth-Bolard C, Venkataramanappa S, van Gassen KLI *et al.* BCL11B mutations in patients affected by a neurodevelopmental disorder with reduced type 2 innate lymphoid cells. *Brain* 2018; **141:** 2299-2311.

56. Tropea TF, Mak J, Guo MH, Xie SX, Suh E, Rick J *et al.* TMEM106B Effect on cognition in Parkinson disease and frontotemporal dementia. *Ann Neurol* 2019; **85:** 801-811.

57. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012; **22:** 1790-1797.

58. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nature methods* 2012; **9:** 215-216.

59. Ernst J, Kellis M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* 2017; **12:** 2478-2492.

60. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005; **102:** 15545-15550.

61. R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* 2013.

62. Andreassen OA, Thompson WK, Schork AJ, Ripke S, Mattingsdal M, Kelsoe JR *et al.* Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate. *PLoS Genet* 2013; **9:** e1003455.

63. Nichols T, Brett M, Andersson J, Wager T, Poline JB. Valid conjunction inference with the minimum statistic. *Neuroimage* 2005; **25:** 653-660.

64. Smeland OB, Frei O, Shadrin A, O'Connell K, Fan CC, Bahrami S *et al.* Discovery of shared genomic loci using the conditional false discovery rate approach. *Hum Genet* 2020; **139:** 85-94.

Acknowledgments

The authors were funded by the South-Eastern Norway Regional Health Authority 2015-078 (T.E.), 2013-123 (O.A.A.), 2014-097 (L.T.W.), 2015-073 (L.T.W.) and 2016-083 (L.T.W.), by the Research Council of Norway (276082 LifespanHealth (T.K.), 213837 (O.A.A), 223273 NORMENT (O.A.A.), 204966 (L.T.W.), 229129 (O.A.A.), 249795 (L.T.W.), 273345 (L.T.W.) and 283798 SYNSCHIZ (O.A.A.)), Stiftelsen Kristian Gerhard Jebsen, the European Research Council (ERCStG 802998 BRAINMINT (L.T.W.)), NVIDIA Corporation GPU Grant (T.K.), the Ebbe Frøland foundation (T.E.), and a research grant from Mrs. Throne-Holst (T.E.).

This research has been conducted using the UK Biobank Resource (access code 27412, https://www.ukbiobank.ac.uk/) and we would like to 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 the data. All authors gave conceptual input and interpreted the 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. received speaker's honoraria from Lundbeck and Janssen Cilag. O.A.A. is a consultant to HealthLytix and received speaker's honoraria from Lundbeck.

Methods

Samples, thalamus segmentations, and quality control procedures

We included raw T1-weighted 3D brain magnetic MRI data from *n* = 30,432 genotyped white British from the UK Biobank[15] 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. For the replication GWAS, we obtained MRI and SNP data from an additional 5,266 white British of the general population in the UK Biobank[15] 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[27,40-46,48,49]. 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 anteroventral 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 contrast[17].

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 = 5,173, respectively.

Genome-wide association 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 v2.0[18]. All GWAS accounted for age, age-orthogonalized age squared, sex, scanning site,

458 intracranial volume, and the first twenty genetic principal components to control for population
459 stratification. The resulting *P*-values were Bonferroni-corrected for analyses of seven volumes. In
460 addition, the GWAS for the six thalamic nuclei groups was run both with and without whole
461 thalamus as an additional covariate. The main text presents results for thalamic nuclei groups
462 when accounting for whole thalamus volume, whereas GWAS results for the nuclei volumes are
463 provided in the Supplementary Information.

464 We identified genetic loci related to thalamic volumes using the FUMA platform v1.3.5[28].
465 The details of the FUMA analyses will be made public at [https://fuma.ctglab.nl/browse]. For
466 these analyses, we used the UKB release2b White British as reference panel. Independent
467 significant SNPs were identified by the genome-wide significant threshold ($P < 7e{-}9$) and by
468 their independency ($r^2 \leq 0.6$ within a 1 mb window). We defined independent significant SNPs
469 with $r^2 < 0.1$ within a 1 mb window as lead SNPs and genomic loci were identified by merging
470 lead SNPs closer than 250 kb. All SNPs in LD ($r^2 \geq 0.6$) with one of the independent significant
471 SNPs in the genetic loci were defined as candidate SNPs. We used the minimum r^2 to determine
472 the borders of the genomic risk loci.

473

474 *Functional annotation, gene-based association, and gene-set analysis*

475 We functionally annotated all candidate SNPs of thalamus volumes using FUMA[28], which is
476 based on information from 18 biological repositories and tools. FUMA prioritizes the most likely
477 causal SNPs and genes by employing positional, eQTL, and chromatin interaction mapping[28].
478 The platform annotates significantly associated SNPs with functional categories, combined
479 CADD scores[22], RegulomeDB scores[57], and chromatin states[28]. A CADD score > 12.37 indicates
480 a deleterious protein effect[22], whereas the RegulomeDB score suggests the regulatory
481 functionality of SNPs based on eQTLs and chromatin marks. Chromatin states show the genomic

region's accessibility for every 200bp with 15 categorical states predicted by ChromHMM based on five histone modification marks (H3K4me3, H3K4me1, H3K36me3, H3K27me3, H3K9me3) for 127 epigenomes[58]. 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; 15=Quiescent/Low13[59].

We performed genome-wide gene-based association and gene-set analyses with MAGMA v1.07[29] in FUMA on the complete GWAS input data. We excluded the MHC 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 $P = 0.05/18158$ genes/7 volumes = 3.9e-7. We ran gene-set analyses with hypergeometric tests for curated gene sets and GO terms obtained from the MsigDB[60].

Analyses of genetic correlation and overlap between thalamus volumes, cortical volumes, and ten brain disorders

We derived volumes of 180 cortical regions based on a multimodal parcellation of the cerebral cortex[35] and performed GWAS for each region, following 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 x 180 cortical volumes) using FDR correction in R statistics[61].

Furthermore, we obtained GWAS summary statistics for ADHD[40], ASD, SZ, BD, and MD from the Psychiatric Genomics Consortium[41-45], for AD from the International Genomics of Alzheimer's Project[46], for MS from the International Multiple Sclerosis Genetics Consortium[47], for PD from the International Parkinson Disease Genomics Consortium and 23andMe[27,48], and for GEP and FEP from The International League Against Epilepsy Consortium on Complex Epilepsies[49]. We performed similar analysis of genetic correlations between volumes and disorders as described for the thalamocortical correlations. In addition, we employed conjunctional FDR statistics[50,52,53,62,63] to assess polygenic overlap between the seven thalamic volumes and the ten brain disorders. A review and mathematical description of this method when applied to neurological and psychiatric disorders were recently published by Smeland *et al*.[64]. In brief, the conjunctional FDR is defined by the maximum of the two conditional FDR values for a specific SNP. The method calculates a posterior probability that a 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, since complex correlations in regions with intricate LD can bias the FDR estimation.

Data availability

Data used in this article were obtained from the UK Biobank [https://www.ukbiobank.ac.uk/], from the Psychiatric Genomics Consortium [https://www.med.unc.edu/pgc/], 23andMe

529 [https://www.23andme.com/], the International Genomics of Alzheimer's Project
530 [http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php], the International
531 Multiple Sclerosis Genetics Consortium [http://imsgc.net/], the International Parkinson Disease
532 Genomics Consortium [https://pdgenetics.org/], and from The International League Against
533 Epilepsy Consortium on Complex Epilepsies [https://www.ilae.org/]. The PD GWAS partly
534 included data from the 23andMe consortium and is available through 23andMe to qualified
535 researchers under an agreement with 23andMe that protects the privacy of the 23andMe
536 participants. Please visit research.23andme.com/collaborate/#publication for more information
537 and to apply to access the data.
538
539 **Code availability**
540 This study used openly available software and code, specifically Freesurfer
541 [https://surfer.nmr.mgh.harvard.edu/], Plink [https://www.cog-genomics.org/plink/], LD-score
542 regression [https://github.com/bulik/ldsc/], and conjunctional FDR
543 [https://github.com/precimed/pleiofdr/].