nature human behaviour

Article

https://doi.org/10.1038/s41562-023-01792-6

The genetic architecture of the human hypothalamus and its involvement in neuropsychiatric behaviours and disorders

Received: 6 February 2023

Accepted: 20 November 2023

Published online: 05 January 2024

Check for updates

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

Despite its crucial role in the regulation of vital metabolic and neurological functions, the genetic architecture of the hypothalamus remains unknown. Here we conducted multivariate genome-wide association studies (GWAS) using hypothalamic imaging data from 32,956 individuals to uncover the genetic underpinnings of the hypothalamus and its involvement in neuropsychiatric traits. There were 23 significant loci associated with the whole hypothalamus and its subunits, with functional enrichment for genes involved in intracellular trafficking systems and metabolic processes of steroid-related compounds. The hypothalamus exhibited substantial genetic associations with limbic system structures and neuropsychiatric traits including chronotype, risky behaviour, cognition, satiety and sympathetic–parasympathetic activity. The strongest signal in the primary GWAS, the *ADAMTS8* locus, was replicated in three independent datasets (N = 1,685–4,321) and was strengthened after meta-analysis. Exome-wide association analyses added evidence to the association for *ADAMTS8*, and Mendelian randomization showed lower *ADAMTS8* expression with larger hypothalamic volumes. The current study advances our understanding of complex structure–function relationships of the hypothalamus and provides insights into the molecular mechanisms that underlie hypothalamic formation.

The hypothalamus plays a crucial role in the regulation of vital bodily functions despite being a small portion of the overall brain. Communicating with other brain regions, the hypothalamus regulates multiple metabolic processes and controls activities of the autonomic nervous system, including circadian rhythms, appetite, blood pressure and heart rate[1]. Furthermore, integration of the hypothalamus with the limbic system subserves cognition, emotion and social functions[2,3]. There is also a growing appreciation of contributions by the hypothalamus to the pathogenesis of various psychiatric and neurological disorders such as schizophrenia (SCZ)[4], depression[5], dementias[6] and Parkinson's disease (PD)[7]. Despite their importance in human health and disease, the genetics of the hypothalamus remains markedly understudied.

Owing to its enhanced soft tissue contrast in comparison to computed tomography and ultrasound imaging, magnetic resonance imaging (MRI) is more widely adopted as a non-invasive examination tool for studying the human brain in vivo, for example, cortical and subcortical structural changes in neuropsychiatric diseases[8,9] and genetic influences on brain structure variation[10,11]. Nevertheless, imaging studies of the hypothalamus are scarce, and no GWAS of common variants or exome-wide association studies of rare variants have been conducted to reveal its genetic architecture. Such scarcity partially stems from the lack of automated segmentation tools of hypothalamic subunits, as previous studies were primarily based on segmentations derived from manual delineations[12–14]. These are time-consuming and labour-intensive procedures that can be difficult to reproduce and severely prone to inter-rater and intra-rater variabilities. Recently, an effective automated tool for segmenting the hypothalamus was developed using a deep convolutional neural network[15]. The technique

e-mail: jianfeng64@gmail.com; dong_qiang@fudan.edu.cn; wcheng@fudan.edu.cn; jintai_yu@fudan.edu.cn

https://doi.org/10.1038/s41562-023-01792-6

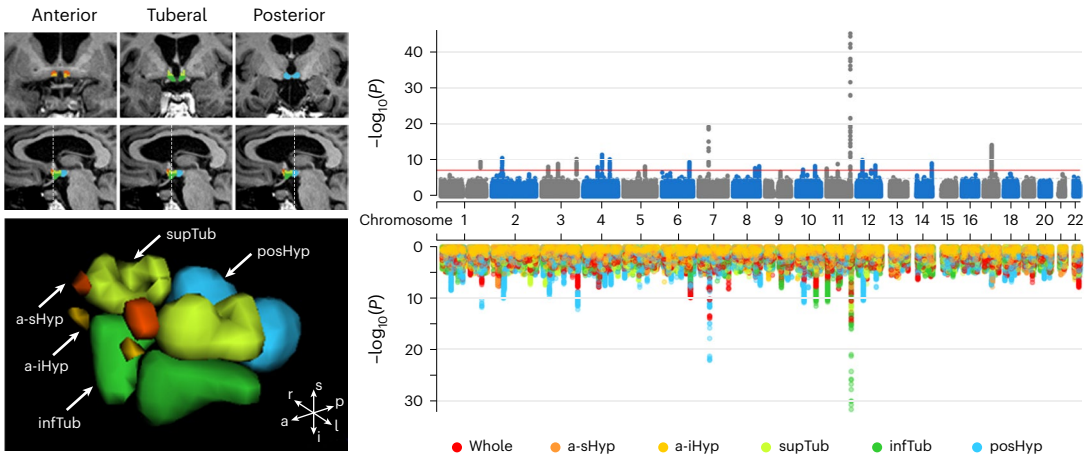

Fig. 1 | The multivariate framework discovered 23 independent loci significantly associated with the hypothalamus. Left, segmentation of the hypothalamus. The hypothalamus is divided into the following five subregions: anterior superior (a-sHyp), anterior inferior (a-iHyp), superior tuberal (supTub), inferior tuberal (infTub) and posterior (posHyp). Top right, Manhattan plot illustrating the $-\log_{10}(P)$ statistic from the mvGWAS based on the MOSTest method. Bottom right, Manhattan plot depicting the $-\log_{10}(P)$ statistics from univariate linear-regression-based GWAS of the whole hypothalamus and single subregions (one colour per region, P values are two-tailed), supporting a distributed genetic architecture across the hypothalamus. s, superior; p, posterior; l, left; i, inferior; a, anterior; r, right.

enables rapid and highly accurate volume estimation of the entire hypothalamus and its subregions from T1-weighted MRI scans. It is scalable to large datasets, which facilitates the detection of the genetic underpinnings of this structure.

In the current study, leveraging large-scale imaging and genetics resources from the UK Biobank (UKB), we aimed to increase our knowledge of the genetic architecture of the hypothalamus. First, one of the largest dataset of hypothalamic imaging samples so far was obtained using the automated segmentation tool. Second, a GWAS of the hypothalamus was conducted to identify common variants and genetic loci linked to the structure with replication in three other independent datasets. Considering that variants probably have distributed effects across hypothalamic regions, we used a multivariable approach, which prevents the need for a multiple-comparison correction or data-reduction strategies. Functional annotation and genetic correlation analyses were subsequently performed to investigate the biological significance of GWAS findings. Third, a rare variant association study (RVAS) was conducted to explore the effect of rare variants on the hypothalamus. Finally, Mendelian randomization (MR) was carried out to probe potential genes that regulate hypothalamic structures. A schematic overview of the study design is illustrated in Supplementary Fig. 1.

Results

Hypothalamus segmentation

A total of 32,956 white British participants from the UKB were included in the discovery GWAS after quality control of imaging and genotyping data. The mean (s.d.) age of the sample dataset was 64.3 (7.5) years, and 52.7% were female. Using T1-weighted MRI scans, the hypothalamic structures were delineated using the automated segmentation tool on the basis of deep convolutional neural networks[15] (Fig. 1). Volumes were estimated for the whole hypothalamus (mean (s.d.) of 826.62 (78.85) mm^3) and its five subregions, including the anterior superior (42.0 (8.32) mm^3), the anterior inferior (31.6 (8.3) mm^3), the superior tuberal (228.2 (25.62) mm^3), the inferior tuberal (281.4 (31.15) mm^3) and the posterior (243.5 (30.89) mm^3) hypothalamus (Supplementary Table 1).

Multivariate GWAS revealed 23 loci associated with the hypothalamus

We performed a multivariate GWAS (mvGWAS) for 8,448,580 single nucleotide polymorphisms (SNPs) using the multivariate omnibus statistical test (MOSTest) method[16]. The analysis was adjusted for age, age squared, sex, scanning site, intracranial volume and the first ten genetic principal components. This method examined each SNP separately for its associations with multiple hypothalamic measures by simultaneously modelling the distributed nature of genetic influences across the hypothalamus (Methods). The functional mapping and annotation (FUMA) of GWAS platform[17] was used to clump mvGWAS results based on linkage disequilibrium (LD) and to identify lead SNPs at each associated locus. As a result, there were 23 unique genomic loci (ranging from 20.4 kb to 1405.4 kb) at the 5×10^{-8} significance level with distributed associations across the hypothalamus and tagged by 26 lead SNPs (Figs. 1 and 2a and Supplementary Tables 2 and 3). The top Manhattan plot depicted in Fig. 1 displays the corresponding multivariate statistics, which emphasized the polygenic nature of the hypothalamus. The bottom Manhattan plot in Fig. 1 illustrates statistics resulting from univariate GWAS analysis of distinct regions in the hypothalamus where the increased univariate statistics of different subregions for certain loci (Supplementary Table 3) supported a distributed genetic architecture across the hypothalamus[18]. Thirteen out of 23 lead SNPs were also implicated in a univariate GWAS (Supplementary Table 3). This result indicates that the MOSTest had an advantage in increasing statistical power by leveraging the distributed effects across hypothalamic subregions compared with the univariate GWAS[16]. For example, the tagging SNP rs3740888 of the 19th loci was also significantly associated with the whole, anterior superior, superior tuberal and inferior tuberal hypothalamus ($\beta = 0.0450–0.0942$, $P_{univariate} = 3.68 \times 10^{-32}–1.83 \times 10^{-8}$).

LD-score regression (LDSC) was used to evaluate inflation from confounding, and all LDSC intercepts of hypothalamic measures were close to one (range of 0.9977–1.0257), which indicated a relatively well-controlled population stratification (Supplementary Table 4). The sensitivity analysis, including additional covariates (that is, batch and array), found almost the same genomic loci (Supplementary Fig. 2a and Supplementary Tables 5 and 6). To focus on hypothalamus-specific associations, the model was further adjusted for the volumes of frontal, parietal, temporal and occipital cortices, and the results also found almost the same loci (Supplementary Fig. 2b and Supplementary Tables 7 and 8). These results indicate the robustness of the original mvGWAS results.

https://doi.org/10.1038/s41562-023-01792-6

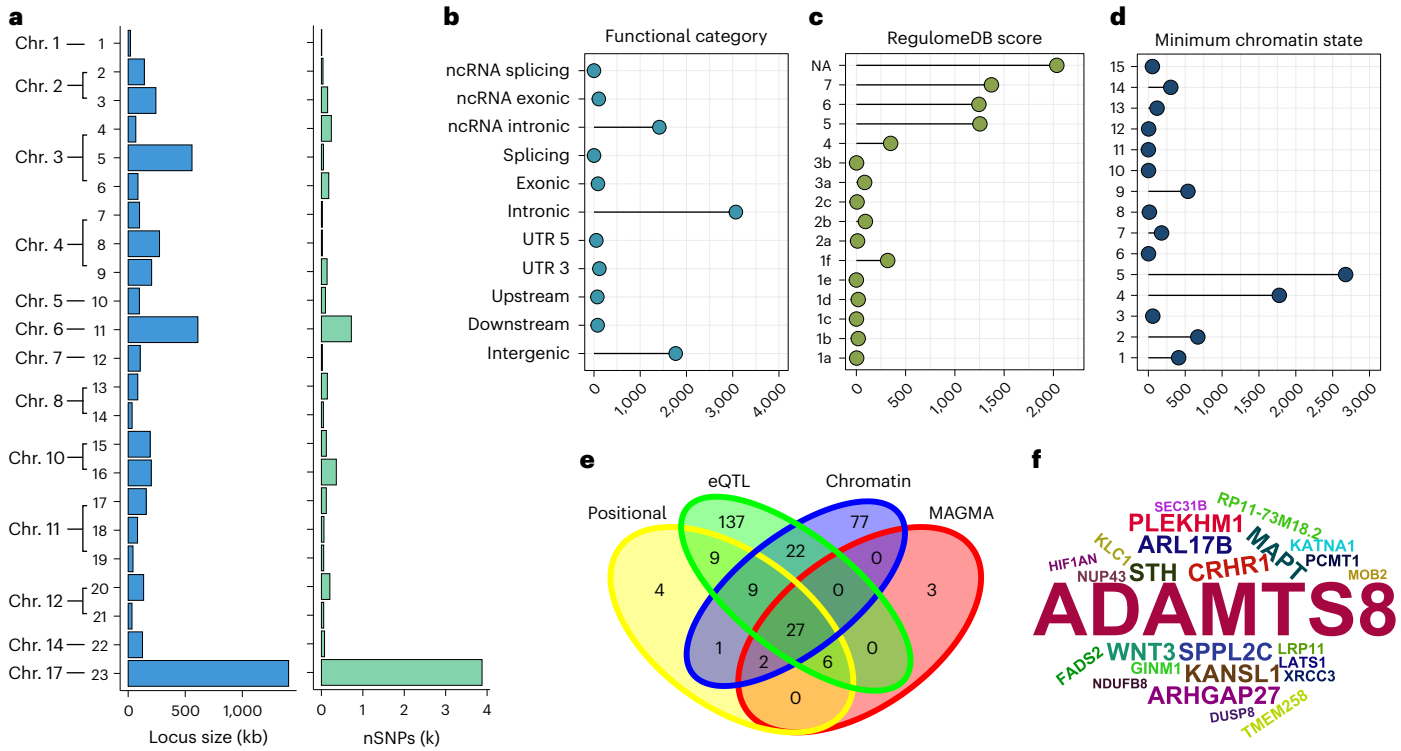

Fig. 2 | Functional annotation and gene mapping of the 23 loci associated with the hypothalamus. a, Overview of the sizes of genomic risk loci and the number of variants. **b**, Distribution of functional consequences of SNPs in significant genomic loci. **c**, Distribution of RegulomeDB scores for SNPs in significant genomic loci. Low scores indicate a higher likelihood of having a regulatory function. **d**, The minimum chromatin state across 127 tissue and cell types for SNPs in significant genomic loci. Lower states indicate higher accessibility, and states 1–7 refer to open chromatin states. **e**, Venn diagram showing the number of genes mapped by the four different strategies. **f**, Wordcloud plot showing 27 genes implicated by all four strategies, with the size of the word representing $-\log_{10}(P)$ statistics from the mvGWAS. Chr., chromosome; ncRNA, non-coding RNA. nSNP, number of SNP; NA, not available; UTR, untranslated regions.

Gene mapping and functional annotation

All the candidate SNPs (n = 6,808) that were defined as the SNPs in LD ($r^2 \geq 0.6$) with any one of the independent significant SNPs in the genetic loci were functionally annotated using the FUMA platform[17] (Supplementary Table 9). Most of these SNPs were intronic (45.1%) or intergenic (26%) and 1.3% were exonic (Fig. 2b). A small proportion of SNPs (6.9%) were in RegulomeDB category 1–2 (Fig. 2c), which implied that they have potential regulatory function. The majority (85%) had minimum chromatin states of 1–7 (Fig. 2d), which indicated their location in open chromatin regions[18]. We used positional, expression quantitative trait loci (eQTL) and chromatin-interaction mappings to map candidate SNPs to genes[19] (Supplementary Table 10). We also performed gene-based analyses by applying the multi-marker analysis of genomic annotation (MAGMA) method[20] and found 38 unique genes (Supplementary Fig. 3 and Supplementary Table 11). These 4 methods together identified 297 unique genes, for which 27 genes were mutually implicated by all 4 methods. The robustness of the 27 genes indicated their essential roles in hypothalamus genetics.

The strongest GWAS hit was mapped to *ADAMTS8* (Fig. 2f), which is important for anti-angiogenesis and blood pressure in vivo[21]. Nine genes were mapped from the locus with the second strongest GWAS signals, including *CRHR1*, *SPPL2C*, *ARL17B*, *MAPT*, *STH*, *WNT3*, *KANSL1*, *PLEKHM1* and *ARHGAP27*. All these genes had high expression levels in the brain, were related to intracellular trafficking and showed relevance to neuropsychiatric functions or disorders to different extents[22–24]. The search tool for the retrieval of interacting genes or proteins (STRING) database discovered 36 known or putative pairwise interactions between the proteins encoded by the 27 genes ($P < 1 \times 10^{-16}$; Supplementary Fig. 4 and Supplementary Table 12) compared with 3 interactions expected for a random set of proteins of the same size from the entire proteome. Gene set enrichment analysis found that many of the significant gene sets reflected processes related to biosynthesis and metabolism of multiple substances, such as steroids, lipids, cholesterol, organic hydroxy compound and other small molecular compounds ($P_{\text{Bonferonni}} = 5.66 \times 10^{-4}$–$3.24 \times 10^{-2}$), organic response to stimuli ($P_{\text{Bonferonni}} = 3.24 \times 10^{-2}$–$4.91 \times 10^{-2}$) and cellular immunity ($P_{\text{Bonferonni}} = 3.66 \times 10^{-5}$–$3.57 \times 10^{-2}$) (Supplementary Table 13). These findings support the validity of our GWAS results and demonstrate that the genes were at least partially biologically connected, as random noise would not result in such functional clustering.

Close genetic correlations with structures of the limbic system

The genome-based restricted maximum likelihood (GREML) method in genome-wide complex trait analysis (GCTA) software was used to estimate the SNP-based heritability (h^2) for the whole hypothalamus and each subregion. Significant heritabilities (false discovery rate (FDR)-corrected $P < 0.05$) were observed for all the hypothalamic volumes, ranging from 10% for the anterior inferior hypothalamus to 30% for the posterior hypothalamus (Fig. 3a). Heritability was also estimated using LDSC, which showed similar patterns despite lower h^2 estimations (Supplementary Fig. 5 and Supplementary Table 14).

Significant and strong correlations were found between each pair of hypothalamic regions by both LDSC ($r_g = 0.16$–0.82, $P_{\text{FDR}} = 3.94 \times 10^{-154}$–$4.98 \times 10^{-2}$; Fig. 3a) and GCTA–GREML ($r_g = 0.84$–0.99, all $P_{\text{FDR}} < 2.23 \times 10^{-308}$) methods (Supplementary Table 15). The results were consistent with the strong phenotypic associations between regions ($r_g = 0.16$–0.83, $P_{\text{FDR}} = < 2.23 \times 10^{-308}$–$1.10 \times 10^{-307}$; Fig. 3a and Supplementary Table 15) and indicated a distributed genetic architecture across the hypothalamus.

We next performed genetic correlation analysis among 6 hypothalamic measures and 101 regional brain volumes and observed significant

https://doi.org/10.1038/s41562-023-01792-6

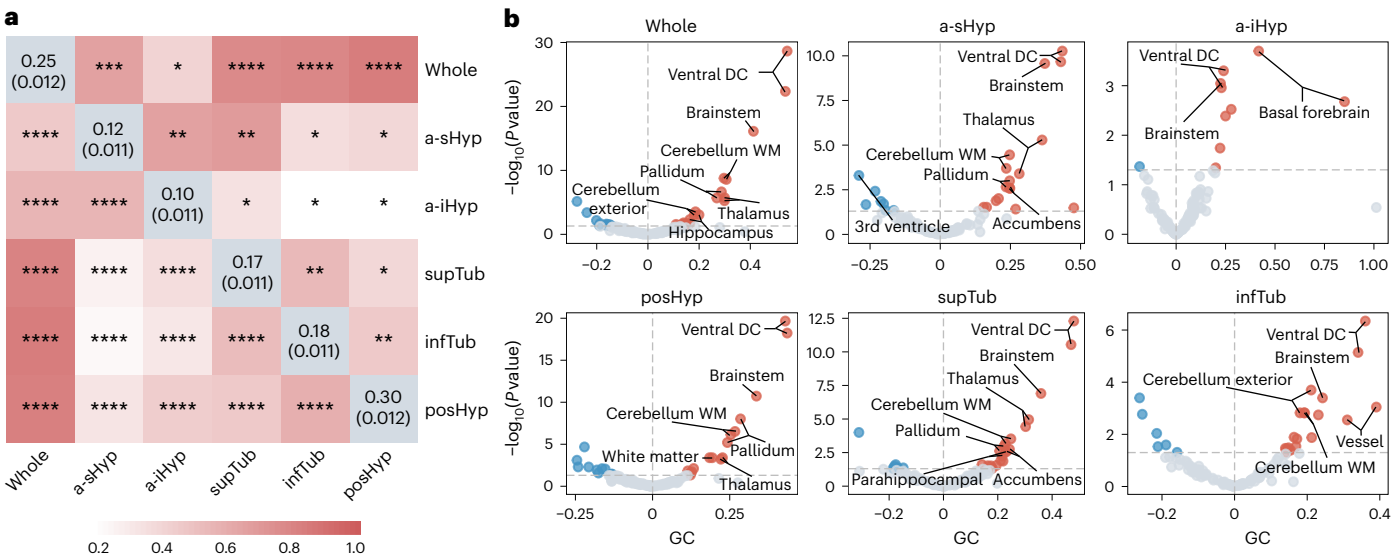

Fig. 3 | Heritability of hypothalamus volumes and their genetic correlation with each other and other regional brain volumes from univariate GWAS. **a**, Heritability estimations and genetic and phenotypic correlations among the hypothalamic regions. The numbers on the diagonal depict heritability estimates (h^2) using the GCTA method, with standard error values in parentheses. LD-score regression-based genetic correlations between each pair of regions are shown above the diagonal with linear regression-based phenotypic correlations off the diagonal. The colour bar indicates the correlation coefficient of each bar. *P* values are two-tailed and reported in the corresponding box with FDR correction. The overall lowest genetic correlations are for the anterior inferior hypothalamus, which is also the structure with the lowest heritability score. Significance: $^{*}P_{FDR} < 0.05$, $^{**}P_{FDR} < 1 \times 10^{-5}$, $^{***}P_{FDR} < 1 \times 10^{-25}$, $^{****}P_{FDR} < 1 \times 10^{-100}$. **b**, Genetic correlation (GC) estimates among 101 regional brain volumes and the whole hypothalamus and its subregions. Red, blue and grey dots indicate positive, negative and non-significant correlation, respectively. Regions with both sides surviving FDR adjustment are labelled. Cerebellum WM, cerebellum white matter; Ventral DC, ventral diencephalon.

correlations in adjacent and connected structures with the hypothalamus (Fig. 3b). When using the LDSC method with FDR corrections for 606 multiple tests, the ventral diencephalon (both sides: $r_g = 0.23$–0.54, $P_{FDR} = 1.35 \times 10^{-26}$–$1.21 \times 10^{-2}$) and the brainstem ($r_g = 0.23$–$0.41$, $P_{FDR} = 9.39 \times 10^{-15}$–$1.07 \times 10^{-2}$) were identified as genetically correlated with all 6 hypothalamic measures. By contrast, significant correlations with at least one hypothalamic measure were found for the thalamus (both sides: $r_g = 0.22$–0.36, $P_{FDR} = 5.91 \times 10^{-5}$–$7.90 \times 10^{-3}$), the pallidum (both sides: $r_g = 0.21$–0.29, $P_{FDR} = 4.15 \times 10^{-7}$–$2.24 \times 10^{-2}$), the accumbens (both sides: $r_g = 0.23$–0.25, $P_{FDR} = 1.79 \times 10^{-2}$–$2.35 \times 10^{-2}$), the hippocampus (both sides: $r_g = 0.16$–0.20, $P_{FDR} = 5.64 \times 10^{-3}$–$4.07 \times 10^{-2}$), the parahippocampus (both sides: $r_g = 0.22$–0.23, $P_{FDR} = 2.46 \times 10^{-2}$–$4.55 \times 10^{-2}$), the basal forebrain (both sides: $r_g = 0.42$–0.85, $P_{FDR} = 3.56 \times 10^{-3}$–$2.05 \times 10^{-2}$) and the cerebellum exterior and white matter (both sides: $r_g = 0.18$–0.31, $P_{FDR} = 8.48 \times 10^{-8}$–$1.57 \times 10^{-2}$) (Supplementary Table 16). When using the GCTA–GREML method, nearly all (594 out of 606) the regional brain volumes were significantly correlated with the hypothalamus ($P_{FDR} \leq 2.23 \times 10^{-308}$–$1.57 \times 10^{-2}$, 557 correlations with $r_g = 0.05$–1.00, 37 correlations with $r_g = -0.83$ to -0.08). Moreover, those identified in LDSC remained within the top strongest correlations and the same directions, including the ventral diencephalon (both sides: $r_g = 0.53$–0.91, $P_{FDR} \leq 2.23 \times 10^{-308}$–$3.94 \times 10^{-82}$), the thalamus (both sides: $r_g = 0.48$–0.85, $P_{FDR} \leq 2.23 \times 10^{-308}$–$1.47 \times 10^{-60}$), the accumbens (both sides: $r_g = 0.10$–0.73, $P_{FDR} \leq 2.23 \times 10^{-308}$–$8.21 \times 10^{-4}$), the hippocampus (both sides: $r_g = 0.55$–0.86, $P_{FDR} \leq 2.23 \times 10^{-308}$–$1.05 \times 10^{-84}$) and the parahippocampus (both sides: $r_g = 0.55$– 0.76, $P_{FDR} \leq 2.23 \times 10^{-308}$–$4.64 \times 10^{-76}$) (Supplementary Table 17). Consistent results from both methods indicated strong spatial colocalization of genetic covariation and a close genetic association with limbic system structures for the hypothalamus.

Genetic correlations with neuropsychiatric traits

To better understand the relationship between hypothalamic structures and functions, we investigated genetic correlations between the hypothalamus and neuropsychiatric traits using both the GCTA–GREML and LDSC methods. Traits were selected on the basis of five categories of physiological processes and functions for which the hypothalamus is engaged in the nervous system[1]. When using GCTA–GREML with individual-level trait data from the UKB as input and adjusting for volumes of frontal, parietal, temporal and occipital cortices to focus on hypothalamus-specific influence, we found that all these traits were significantly correlated with at least one hypothalamic measure, and all correlations of each trait, when $P < 0.05$, had the same directions between hypothalamic regions (Fig. 4 and Supplementary Table 18). Specifically, the traits for sleep and circadian rhythms[25,26] were genetically correlated with the whole hypothalamus and different substrates, including chronotype ($r_g = 0.87$–0.98, $P_{FDR} \leq 2.23 \times 10^{-308}$), daytime napping ($r_g = -0.08$, $P_{FDR} = 5.56 \times 10^{-3}$) and sleep duration ($r_g = -0.11$ to -0.08, $P_{FDR} = 2.82 \times 10^{-3}$–$3.67 \times 10^{-2}$). Significant correlations were also discovered for traits for risky behaviours[2,27], particularly with moderate-to-large estimates, including cigarettes per day ($r_g = -0.48$ to -0.29, $P_{FDR} = 2.67 \times 10^{-2}$–$4.65 \times 10^{-2}$), frequency of drinking ($r_g = 0.11$–0.28, $P_{FDR} = 2.96 \times 10^{-11}$–$2.16 \times 10^{-4}$) and age at first sexual intercourse ($r_g = -0.18$ to -0.14, $P_{FDR} = 1.25 \times 10^{-12}$–$4.84 \times 10^{-6}$). Significant genetic correlations also linked the hypothalamus to traits for learning and cognition[28,29], including incorrect pair matches ($r_g = 0.18$–0.20, $P_{FDR} = 1.72 \times 10^{-2}$–$1.81 \times 10^{-2}$) and reaction time ($r_g = 0.15$–$0.30$, $P_{FDR} = 4.69 \times 10^{-9}$–$5.82 \times 10^{-5}$). Additionally, traits for food intake and satiety[30,31] ($r_g = 0.08$–0.22, $P_{FDR} = 6.69 \times 10^{-14}$–$4.63 \times 10^{-2}$) and parasympathetic–sympathetic activity[32,33] ($r_g = 0.06$–0.27, $P_{FDR} = 8.05 \times 10^{-14}$–$2.73 \times 10^{-2}$) were correlated with hypothalamic measures. The LDSC analysis showed correlations at uncorrected $P < 0.05$ for cigarettes per day ($r_g = -0.11$ to -0.10, $P = 2.50 \times 10^{-2}$–4.00×10^{-2}), poor appetite or overeating ($r_g = 0.14$, $P = 4.74 \times 10^{-2}$), body mass index ($r_g = 0.11$, $P = 4.74 \times 10^{-2}$), systolic blood pressure ($r_g = 0.07$–0.09, $P = 5.90 \times 10^{-3}$–4.43×10^{-2}) and diastolic blood pressure ($r_g = 0.07$–0.09, $P = 5.90 \times 10^{-3}$–4.43×10^{-2}), but these correlations lost significance after corrections for multiple tests (Supplementary Table 19). We also performed GCTA–GREML without adjustment for the four cortices, which showed

https://doi.org/10.1038/s41562-023-01792-6

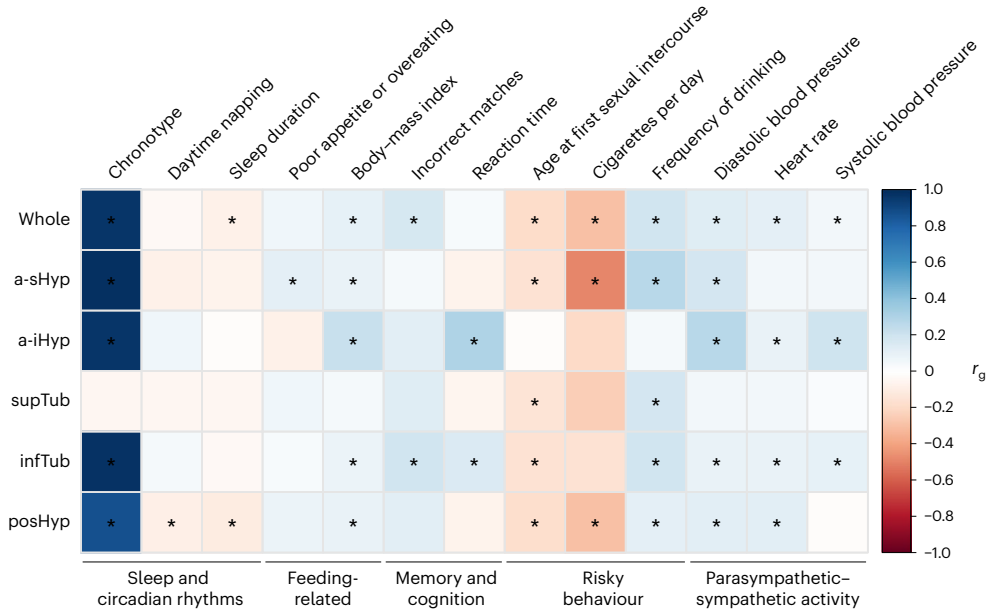

Fig. 4 | Genetic correlations between the hypothalamus and its function-related traits. Genetic correlations significantly differ from zero after FDR correction for the 78 tests in the analysis marked with an asterisk. Blue, positive genetic correlation; red, negative genetic correlation.

similar results despite several associations with inconsistent directions (Supplementary Fig. 6 and Supplementary Table 20). Although these results indicated that correlations with inconsistent directions before adjustment had been confounded by other brain regions, they demonstrated that these relationships are not solely driven by general brain associations. Taken together, these findings provide support for the involvement of the hypothalamus in the neuropsychiatric traits.

Genetic associations with neuropsychiatric disorders

Genetic correlations were also investigated between the hypothalamus and 13 common neuropsychiatric disorders, including SCZ, bipolar disorder (BD), major depression (MD), generalized anxiety disorder (GAD), posttraumatic stress disorder (PSD), attention-deficit/hyperactivity disorder (ADHD), anorexia nervosa (AN), autism spectrum disorder (ASD), obsessive–compulsive disorder (OCD), Tourette's syndrome (TS), stroke, Alzheimer's disease (AD) and PD. Because of rare neuropsychiatric diagnoses in the neuroimaging samples of the UKB, we only used the LDSC method with publicly available summary statistics rather than GCTA–GREML. We observed genetic correlations with uncorrected $P < 0.05$ between the following disorders and brain regions: OCD and the inferior tubular ($r_g = -0.27$, $P = 6.00 \times 10^{-3}$), the superior tubular ($r_g = -0.21$, $P = 2.91 \times 10^{-2}$) and the whole hypothalamus ($r_g = -0.18$, $P = 4.31 \times 10^{-2}$); AD and the inferior anterior hypothalamus ($r_g = -0.24$, $P = 3.43 \times 10^{-2}$); and SCZ and the whole hypothalamus ($r_g = -0.07$, $P = 3.15 \times 10^{-2}$) (Supplementary Table 21). These correlations were not significant after multiple testing corrections. We also performed bidirectional MR analyses on the 13 disorders and the hypothalamus to examine putative causal relationships. In forward MR, both the whole and superior tubular hypothalamus were associated with PSD at uncorrected $P < 0.05$ (inverse variance-weighted (IVW) odds ratio (OR) = 0.38–0.52, $P = 3.81 \times 10^{-3}$–4.32×10^{-2}) and no directional pleiotropy was identified ($P_{\text{MR-EGGER intercept}} = 0.42$–$0.94$) (Supplementary Table 22). In reverse MR, AD was associated with 4 subregions at uncorrected $P < 0.05$, including the anterior superior, the anterior inferior, the tubular superior and the tubular inferior hypothalamus (OR = 0.96–0.97, $P = 2.52 \times 10^{-3}$–2.35×10^{-2}), which showed no significant directional pleiotropy ($P_{\text{MR-EGGER intercept}} = 0.30$–$0.96$) (Supplementary Table 23). However, these associations were not significant after multiple testing corrections, which indicated a lack of evidence of causality.

We then explored the polygenic architecture of the hypothalamus and potential genetic overlap with disorders using conditional FDR and conjunctional FDR analyses with summary statistics of 13 common neuropsychiatric disorders. When conditioning the multivariate statistic of hypothalamus on the disorders, conditional Q-Q plots showed a clear pattern of pleiotropic enrichment (Supplementary Fig. 7). In total, 36 (SCZ), 30 (BD), 34 (ASD), 31 (MD), 28 (PD), 33 (PSD), 31 (OCD), 26 (stroke), 38 (ADHD), 31 (AN), 29 (AD), 29 (TS) and 25 (GAD) genetic loci were identified below the conditional FDR level of 0.01 and associated with the hypothalamus (Supplementary Table 24), which were all more than the 23 loci identified in the original mvGWAS for the hypothalamus. Of these, we identified 20 loci that significantly overlapped (conjunctional FDR level of <0.05) with SCZ, 4 with BD, 3 with MD, 1 with PSD, 3 with ASD, 1 with OCD, 1 with stroke and 2 with PD (Supplementary Fig. 8 and Supplementary Table 25). We mapped the identified loci in conjunctional FDR analysis to genes using positional, eQTL and chromatin-interaction mapping (Supplementary Table 26) and found that multiple genes were implicated for more than two conditions (Supplementary Table 27). Notably, a strong overlap was found among SCZ, BD and MD, for which 19 genes overlapped between the hypothalamus and all three disorders, including *KLC1*, *RP11-73M18.2* and *XRCC3*. These 3 genes were among the aforementioned 27 genes that were robustly associated with the hypothalamus. These results suggest that genetic pleiotropy exists among neuropsychiatric disorders that, to some extent, share common underlying hypothalamic pathological mechanisms. When considered together with other evidence from genetic correlation analyses and MR, these findings provide support for the involvement of the hypothalamus in neuropsychiatric disorders.

Replication and meta-analysis

Multivariate replications were conducted in 3 independent samples, including 3,284 UKB participants who were excluded from the discovery analyses as they were not of white British descent, 4,321 participants from the Adolescent Brain Cognitive Development (ABCD) study[34] and 1,685 participants from the IMAGEN study[35]. Sample descriptions are

https://doi.org/10.1038/s41562-023-01792-6

provided in Supplementary Table 1. With the replication threshold set at 0.05, 7, 11 and 11 loci were replicated for the UKB, ABCD and IMAGEN samples, respectively (Supplementary Table 2). The strongest locus in the discovery sample, the *ADATMS8* locus on chromosome 11, was replicated in all three datasets (Fig. 5 and Supplementary Tables 2 and 28–30). A meta-analysis combining the four datasets was further performed using the software METAL. Summary statistics from univariate GWAS of the entire hypothalamic volume were used because the MOSTest method did not provide effect estimations or directions for mvGWAS. Consequently, associations with increased significance were detected for the variants in the *ADAMTS8* locus after meta-analysis. Moreover, all the associations across samples, when $P < 0.05$, had consistent effect directions (Supplementary Table 31). For example, the significance increased from $P = 4.62 \times 10^{-15}$ to $P = 3.66 \times 10^{-15}$ with the same direction for rs3740888 in the *ADAMTS8* locus.

RVAS of the hypothalamus

To explore the effects of rare variants on hypothalamic structures, a RVAS using the SKAT-O test was performed among 28,988 unrelated British participants with both whole-exome sequencing and hypothalamic imaging data available[36]. The analysis was restricted to variants with a minor allele frequency value of <0.01, which were grouped into two categories (loss of function, and loss of function and probable deleterious missense) by protein-coding gene. No significant association was identified under the *P* threshold (2.29×10^{-7}) of Bonferroni correction for all 18,748 genes, 2 variant functional groups and 6 hypothalamic measures, whereas several subthreshold associations were observed (Supplementary Fig. 9 and Supplementary Table 32). The gene *PPT2*, the strongest signal in RVAS, which encodes a member of the palmitoyl-protein thioesterase family, was associated with volumes of both the whole hypothalamus ($P_{SKAT\text{-}O} = 5.20 \times 10^{-7}$) and the superior tuberal hypothalamus ($P_{SKAT\text{-}O} = 2.38 \times 10^{-7}$). Associations were also observed for *CHCHD1* and *CHCHD4*, two evolutionarily conserved genes, with the superior tuberal hypothalamus ($P_{SKAT\text{-}O} = 3.07 \times 10^{-7}$) and the posterior hypothalamus ($P_{SKAT\text{-}O} = 6.61 \times 10^{-7}$), respectively. Further single variant analysis denoted that the signals were contributed by a burden of multiple ultrarare variants (Supplementary Table 33). Through browsing the GWAS ATLAS resource (https://atlas.ctglab.nl)[37], *PPT2*, *CHCHD1* and *CHCHD4* have been previously linked to metabolism, blood pressure and immunity, which were closely related to hypothalamic functions. Notably, we found associations of the *ADAMTS8* gene with volumes of the whole hypothalamus and its subregions ($P_{SKAT\text{-}O} = 8.09 \times 10^{-4}$–$4.10 \times 10^{-2}$).

MR for *ADAMTS8* and the hypothalamus

Given the strong and robust signals from *ATAMTS8*, we proposed that the expression of this gene would affect hypothalamic structures. MR analyses were therefore performed with eQTLGen summary statistics (the largest eQTL data) as the exposure input and our univariate GWAS of the whole hypothalamic volume as the outcome input. After LD pruning and examining instrumental SNPs associated with potential confounders (Methods), three independent genome-wide SNPs in chromosome 11 (rs3740888, rs747250 and rs11606448) were used for MR. The IVW method, the primary MR analytical approach, showed that *ADAMTS8* expression levels were significantly associated with the volume of the whole hypothalamus ($\beta_{IVW} = 0.92$, $P = 1.72 \times 10^{-5}$; Fig. 6a). Sensitivity analyses, including weighted median and weighted mode methods, also presented significant results, and different methods showed consistent estimates (Fig. 6b), which suggested that *ADAMTS8* expression had an influence on hypothalamic structures. MR-Egger intercepts indicated no directional pleiotropy ($P = 0.22$). Moreover, leave-one-out analysis and single SNP analysis showed that the associations were not driven by a single SNP (Supplementary Tables 34 and 35).

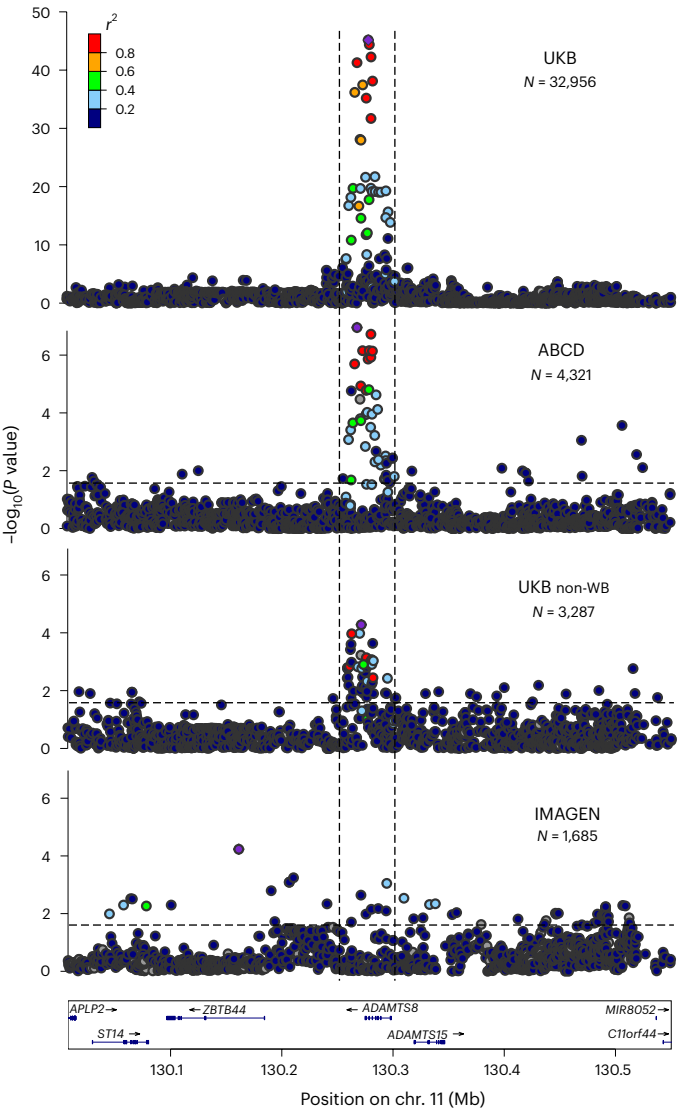

Fig. 5 | Regional association plots of the *ATAMTS8* locus across discovery and replication samples. Associations of mvGWAS in different samples are shown for the region covering ±0.25 Mb of the *ATAMTS8* locus. The vertical dashed lines define the *ATAMTS8* locus. The raw two-sided *P* values are presented without corrections for multiple comparisons. The horizontal dashed lines indicate the replication threshold of 0.05. Variants with the smallest *P* values are shown in purple dots. Dot colours represent different levels of LD. non-WB, non-white British.

Discussion

In the current study, we conducted a large genome-wide and exome-wide association study of the hypothalamus and revealed genetic architectures of hypothalamic structures. GWAS using a multivariable approach identified 23 loci linked to the structure and a moderate heritability for hypothalamic volumes. Functional analyses implicated genes involved in intracellular trafficking and biological processes enriched in biosynthesis and metabolism. We observed genetic associations between the hypothalamus and adjacent brain structures of limbic systems. We also discovered that hypothalamic structures were genetically correlated with neuropsychiatric traits such as chronotype, risky behaviours, cognitive levels, satiety and sympathetic–parasympathetic activity. Additionally, genetic overlaps between the hypothalamus and neuropsychiatric disorders were recognized, and the identification of shared genes implicated for disorders

https://doi.org/10.1038/s41562-023-01792-6

a

Method	nSNP	β	s.e.	95% CI	P
IVW	3	0.92	0.21	0.50–1.34	1.72×10^{-5}
MR-Egger	3	0.67	0.46	−0.23–1.57	2.83×10^{-1}
Weighted median	3	0.81	0.14	0.54–1.08	4.55×10^{-9}
Weighted mode	3	0.7	0.13	0.44–0.96	3.41×10^{-2}

b

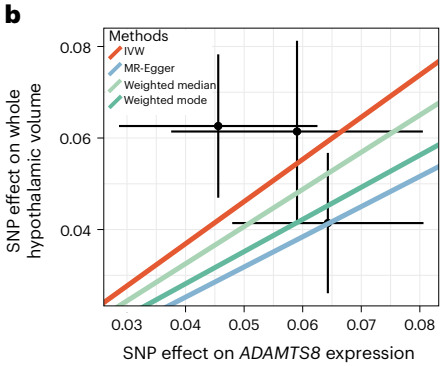

Fig. 6 | MR analyses for *ADAMTS8* expression versus whole hypothalamic volumes. a, Summary statistic for the different MR models analysed. **b**, Scatterplot illustrating genetic associations with whole hypothalamic volumes (N = 32,956) against genetic associations with *ADAMTS8* expression (N = 31,684), with dots indicating β values of effects and error bars representing 95% confident intervals (CI) for the associations. Different lines represent different MR methods.

suggested common hypothalamic pathological mechanisms. We successfully replicated the *ADAMTS8* locus, the top signal in discovery GWAS, in three separate samples and found increased significance in meta-analysis. Although no significant associations were detected in RVAS, the analysis implied that multiple rare variants also contributed to the association between *ATAMTS8* and the hypothalamus. MR analyses further supported a biological connection between *ATAMTS8* gene expression levels and hypothalamic structures.

Our findings agreed with those reported in the literature. The most significant locus tagged by rs3740888 was previously found to be associated with the ventral diencephalon and nearby structures such as the parahippocampus, the brainstem and the ventricles[7]. The variant was also identified in large-scale GWAS of blood pressure[8–10], which aligned with our results of a genetic correlation between hypothalamic measures and blood pressure and with the observation that the top hit signal was mainly detected from the analysis of the tuberal hypothalamus, a region responsible for blood pressure control[38]. The second identified locus tagged by rs199534 mapped to genes closely associated with hypothalamic functions. For instance, *CRHR1* was specific to regulation of the hypothalamic–pituitary–adrenal pathway. *SPPL2C*, *ARL17B*, *MAPT*, *STH*, *WNT3*, *KANSL1*, *PLEKHM1* and *ARHGAP27* were clustered by STRING analysis (Supplementary Fig. 3) and were shown to be all engaged in intracellular transport system (Results). Together with the gene ontology analysis indicating biological processes enriched in synthesis and decomposition of biochemical compounds, this evidence accords with the major role of the hypothalamus in regulating body composition and energy metabolism[1].

The study of the hypothalamus and its topological variability provides valuable insights into underlying physiological and pathological processes. Consistent genetic associations with three chronotypic phenotypes were in line with common knowledge that the hypothalamus has an essential role in circadian rhythms[39]. Additionally, the findings on genetic associations of the hypothalamus with regional brain structures mutually agreed with those with functional traits. The hypothalamus, the hippocampus, the thalamus, the pallidum, the accumbens and the basal forebrain constitute the key components of two different networks in the limbic system: one essential for memory and learning, the other one for social–emotional functions[1]. Our analyses provided genetic evidence of both structural and functional associations within these networks. Furthermore, the genetic correlations both with adjacent structures and with functional traits complemented the genetic overlap analysis in confirming a role of the hypothalamus in neuropsychiatric disease.

It seemed that heritability and regional volume correlated with the hypothalamus. Although larger hypothalamic subunits such as the posterior hypothalamus had a h^2 value of 0.30, smaller subunits such as the anterior inferior hypothalamus only had a h^2 value of 0.10. A potential explanation for this result is that measured values of small regions included noise signals, which might have influenced estimations of heritability. The other possibility is that the low h^2 values of small regions are partially attributed to their low heritability per se. For instance, the posterior hypothalamus (243.45 mm^3) was smaller than the inferior tuberal hypothalamus (281.44 mm^3), whereas the posterior hypothalamus showed a higher h^2 (0.30) than the inferior tuberal hypothalamus (0.18). Similar results were obtained for the posterior hypothalamus (243.45 mm^3) and the whole hypothalamus (826.62 mm^3). The h^2 of the posterior hypothalamus (0.30) was higher than the whole hypothalamus (0.25).

Strong evidence was found for *ADAMTS8*, the role of which in hypothalamus had not been investigated before. Consistent association results were found from four separate imaging cohorts. In particular, the associations between *ADAMTS8* and hypothalamic structures were discovered in the ABCD and IMAGEN cohorts, representing individuals at an early stage of life, which strengthened the validity of our findings and suggested an age-independent effect of *ADAMTS8*. Additionally, results of exome-wide scans in the UKB also supported associations for *ADAMTS8*. Furthermore, the association between lower *ADAMTS8* expression levels and larger hypothalamic sizes (Fig. 6; note the phenotypic variables were inverse-rank-transformed before mvGWAS using the MOSTest, Methods) was in line with existing research of *ADAMTS8*. This gene encodes a secreted proteinase anchored to the extracellular matrix and can disrupt angiogenesis in vivo[40]. The inhibiting effect on vessel formation may explain the adverse effects on volumetric measures of the hypothalamus. Further experimental research is required to reveal causal molecular association mechanisms of *ADAMTS8* with hypothalamic structures.

Our findings implied that several genes contribute to the pathology of the hypothalamus in different psychiatric and neurological conditions. These findings support the idea of pleiotropy across a spectrum of neuropsychiatric disorders and may be worth further investigation for the identification of disorder-independent drug targets[16]. For example, we discovered that *KLC1* overlapped between the hypothalamus and SCZ, MD and BD. *KLC1* encodes a member of the kinesin light chain family, which is highly expressed in brain tissue. Kinesin is a tetrameric molecule composed of two heavy chains and two light chains. It transports various cargos along microtubules towards their plus ends such as vesicles, mitochondria and the Golgi complex[41]. Although the heavy chains provide motor activity, the light chains bind to various cargo. Researchers recently found that dysfunction of kinesin proteins impairs vesicle transport of NMDA

receptor subunit 2A and the adenomatous polyposis coli complex, cause disrupted synaptic plasticity and lead to behavioural and cognitive defects, which underlie SCZ pathogenesis[42,43]. Previous studies also found that kinesin-mediated antimanic and antidepressant effects of glycogen synthase kinase 3 inhibitors interfere with dissociation of vesicles containing AMPA[44]. Therefore, the kinesin cargo system may provide valuable targets for developing new therapeutics for psychiatric disorders.

The segmentation tool we used was more effective than ones used in earlier studies. The mean absolute volume (±s.d.) of the segmented hypothalamus was 826.6 (78.9) mm^3, within the range of reported results from other MRI volumetric studies[13,45,46]. By contrast, previously reported volumes featured a high degree of variability, ranging from 600 mm^3 to 1,100 mm^3. This was primarily owing to the absence of established guidelines for the delineation of the hypothalamus. On the one hand, there was considerable variation in the structures that were defined as 'hypothalamic' or included in the hypothalamus across studies. On the other hand, manual segmentation of such a delicate structure was severely prone to inter-rater and intra-rater variability and resulted in restriction of large sample usage. Nomenclature and definition of the hypothalamus used in a previous study[13] were adopted in the current study, which provided identifiable anatomical landmarks for segmentation. The introduction of the automated delineation method in our research contributed to reliable phenotypic measurement and GWAS analyses.

The multivariate approach was also a strength in our study. MOSTest is not repeating univariate analyses, nor is it recapturing the univariate associations from the most significant univariate results. Take the lead SNP rs55938136, for example. Although the *P* values of all univariate results were not less than 5×10^{-8}, the result of MOSTest was significant ($P = 5.36 \times 10^{-13}$). The same conditions applied to other lead SNPs such as rs6752635, rs17048681, rs1375875, among others. In addition, although only 7, 1, 0, 1, 4 and 9 loci were detected in univariate GWAS of the 6 hypothalamic measures, 23 were identified in mvGWAS. A multivariate approach can boost statistical power while maintaining data scale compared with separate univariate analyses of individual phenotypes[16]. A single set of genome-wide association results, pertaining simultaneously to multiple regions of the hypothalamus, was then fed into functional annotation and downstream analyses, such as gene mapping and testing for genetic overlaps with diseases[47]. Without the noise inherent in repeat univariate testing, the multivariate approach enabled the exploitation and interpretation of important features of the genetic architecture of specific brain structures.

Our GWAS had several limitations. First, the study was mainly conducted among participants of British ancestry with a specific age range. We noted that there were associations with inconsistent effect directions but no significance in UKB non-white British and ABCD cohorts, which suggested that sample heterogeneity due to ancestry and age may have influenced effect estimation. Thus, the generalization of these findings to other ethnicities or specific cohorts should be treated with caution. We anticipate that large-scale diverse samples will become more readily available to further verify our findings and make new discoveries. Second, the automated segmentation tool parcellated the hypothalamus into only five subdivisions instead of individual nuclei. The small size of the hypothalamus, the low power of the magnetic field and insufficient identifiable landmarks in MRI were the main reasons for this limitation. Accordingly, analyses of hypothalamic nuclei could not be conducted, and the genetic correlations between the hypothalamic subdivisions and functional traits in our analysis may not be interpreted as direct evidence for specific nucleus–function relationships. A more detailed segmentation tool is needed. Third, there were 5% overlapping individuals from the UKB in the summary statistics of AN. Unlike LDSC, sample overlap is principally not allowed in MR and conditional FDR and conjunctional FDR methods. Nevertheless, previous research has shown that this small percentage of sample overlap is not sufficient to bias the analyses[48,49]. Fourth, although a large number of hypothalamic samples were analysed in our study, the sample size may be relatively insufficient for RVAS. This may partially explain the lack of significant associations in RVAS and the fact that associations with the strongest signals were all contributed by ultra-rare variants that had minor allele counts of fewer than ten. Finally, the specificity of the hypothalamus to neuropsychiatric traits needs further investigation. Our research provided multiple lines of evidence from functional annotation, genetic correlations and genetic overlap collectively supporting that the hypothalamus is genetically involved in neuropsychiatry. Additionally, sensitivity analyses controlling the four different cortices showed that the associations were not solely driven by general associations in the brain. However, it cannot be completely excluded that the association between the hypothalamus and neuropsychiatric traits was due to other brain regions. More detailed research is needed to untangle associations of the hypothalamus and other brain regions with neuropsychiatric traits.

In conclusion, our findings improved our understanding of the intricate relationships between topographical features, physiological functions and pathological processes, and provide new insights into key molecular mechanism that underlie the structural formation of the hypothalamus.

Methods

Participants

This study is based on publicly available data with different levels of accessibility. Individual-level data from three independent cohorts were used, including the UKB (http://www.ukbiobank.ac.uk/), the ABCD study (https://abcdstudy.org/) and the IMAGEN study (https://imagen-europe.com/). This study was approved by the Institutional Review Boards (IRBs) of all participating institutions and was carried out in accordance with the approved protocols. The UKB was approved by the National Health Service National Research Ethics (reference 11/NW/0382). The ABCD study was approved by the central IRB at the University of California, San Diego. The IMAGEN study was approved by the institutional ethics committees of King's College London, the University of Nottingham, Trinity College Dublin, the University of Heidelberg, Technische Universität Dresden, Commissariat à l'Energie Atomique et aux Energies Alternatives and the University Medical Center at the University of Hamburg. Written informed consent was obtained from all participants and/or their parents or guardians, and children in the ABCD study consented before participation.

We used brain MRI data from ~40,000 genotyped individuals from the UKB under accession number 19542 (https://biobank.ndph.ox.ac.uk/showcase/index.cgi). Our primary analytical sample was restricted to white British individuals and data availability and processing (see below) led to a final sample of 32,956 participants (mean age of 64.3 years) in the discovery GWAS and 28,988 participants in the RVAS. A dataset of 3,287 participants (mean age of 62.8 years) also from the UKB, who identified as white but not British (see below), were included as GWAS replication samples. Additionally, we used the ABCD study and the IMAGEN study for GWAS replication[50]. Data of 4,321 children (mean age 9.90 years) in the ABCD study and of 859 young adolescents (mean age 14.0 years) in the IMAGEN study with complete genetic data and T1 MRI scans were included. Characteristics of the participants are summarized in Supplementary Table 1. The current study was not preregistered. No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to or larger than those reported in previous publications on neuroimaging GWAS[10,11,47].

Quality control of genetic data

For GWAS, we followed a standard quality control procedure to that of the UKB v.3 imputed genetic data[51]. Individuals with missing genotype rates exceeding 0.05, a mismatch between self-reported and genetic sex, putative sex chromosome aneuploidy, heterozygosity rate outliers,

more than ten putative third-degree relatives were first removed. The primary analyses were restricted to individuals with 'white British ancestry', as previously defined[51], based on the combination of self-report and genetic principal component analysis (UKB Field ID: 220006). Variants with a call rate of <0.95, a minor allele frequency of <0.01, a Hardy–Weinberg P value of $<10^{-6}$ or an imputation quality score of <0.5 were excluded. We applied the same procedure to obtain the independent GWAS replication dataset of a white population from the UKB, except that the participants did not self-identify as British nor did they fall within the genetic cluster bounds of 'white British ancestry' as previously defined[51]. For ABCD and IMAGEN, non-European participants were removed and similar quality control processes for genotyping data were applied (see details in Supplementary Table 36).

For RAVS of the UKB, whole-exome sequencing data from ~450,000 participants were used. Details of sample preparation and sequencing have been previously described[52]. In this study, we utilized the OQFE whole-exome sequencing pVCF files provided by the UKB and performed additional quality control similar to the previous study[53]. In brief, multiallelic sites were split into biallelic sites, and all calls that had low genotype quality or extreme low or high genotype depth were set to no-call. We also removed variants with a call rate of <90%, a Hardy–Weinberg P value of $<10^{-15}$ and location in Ensembl low-complexity regions. Samples withdrawn from the study, duplicates, samples with discordance between self-reported and genetically inferred sex, samples for which the transition/transversion ratio, the heterozygote/homozygote ratio, single nucleotide variant/insertion and deletion ratio and number of singletons exceeded eight standard deviations from the mean, samples with genetic relation to one another at second degree or closer, and non-British were removed.

Preparation of imaging data and segmentation of the hypothalamus

UKB T1-weighted MRI scans were collected from three scanning sites throughout the United Kingdom on identically configured Siemens Skyra 3T MRI scanners with 32-channel radiofrequency receive head coil. The full scanning protocols and procedures can be found in UKB documents (https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/brain_mri.pdf). ABCD T1-weighted scans were obtained from 21 imaging sites using Siemens Prisma, GE and Philips 3T scanners. Scanning protocols were harmonized across sites, and full details of imaging protocols used in the ABCD study have been previously outlined[34]. IMAGEN T1-weighted MRI scans were acquired using Siemens TrioTim, GE and Philips 3T scanners at eight sites using compatible scanning parameters and the same scanning protocol at all sites. Full details of MRI acquisition, quality checks and standardization across scanners have been previously described for the IMAGEN protocol[50].

With T1-weighted MRI scans from three cohorts, the whole hypothalamus and its subregions, including the anterior superior, the anterior inferior, the superior tuberal, the inferior tuberal and the posterior hypothalamus, were then delineated using a previously developed automated segmentation tool[15]. The fully automated segmentation method was based on a deep convolutional neural network. It was established with aggressive data augmentation to make the model robust to T1-weighted scans from different sources and has previously been validated in three independent datasets, attributing reliable performance and replicable results, thus without any need for preprocessing[15]. For each of the six hypothalamic measures, we calculated the summed volumes of both sides as phenotype input in GWAS and removed data points deviating more than 5 median absolute deviation from the median[54].

mvGWAS

MOSTest can leverage the distributed nature of genetic effects across spatially distributed brain phenotypes while accounting for their covariances, which can boost statistical power to discover potential variant–phenotype associations. Specifically, the multivariate correlation structure is estimated on randomly permuted genotype data. MOSTest then calculates the Mahalanobis norm as the sum of squared decorrelated z values across univariate GWAS summary statistics to integrate effects across phenotypes into a multivariate z statistic for each genetic variant before the gamma cumulative density function is used to fit an analytical form for the null distribution[16]. This allows the extrapolation of the null distribution below the 5×10^{-8} significance threshold without performing an unfeasible number of permutations. The mathematical details and the soft implementation procedure can be found in ref. 16.

In the current study, the volumes of the total hypothalamus and each of the five subregions were residualized for age, age squared, sex, scanning site, intracranial volume and the first ten genetic principal components. The resulting residuals were then jointly fed into the MOSTest analysis. By default, all phenotypic variables were inverse-rank transformed into a normal distribution before GWAS analysis. MOSTest performs permutation testing to identify genetic effects across multiple phenotypes and generates mvGWAS summary statistics across all the six hypothalamic measures. As mvGWAS did not provide effect estimates and directions, the univariate GWAS of each hypothalamic measure was also extracted from the univariate stream of MOSTest for post-GWAS analyses, including genetic correlations, MR, meta-analysis and heritability estimation. The LDSC intercepts calculated using LDSC (v.1.0.1) were used to evaluate inflation and confounding bias. To increase the robustness of the mvGWAS results, we additionally added batch and array as well as volumes of frontal, parietal, temporal and occipital cortices in covariates and re-ran the MOSTest in the discovery sample.

MOSTest was also performed for replication with samples from the non-white British group of the UKB and the ABCD and IMAGEN studies. The software METAL[55] was used to perform meta-analysis on four samples using univariate GWAS summary statistics of the whole hypothalamus derived from MOSTest.

Genomic loci characterization and gene mapping

FUMA (v.1.3.6)[17], an online platform for post-GWAS analysis, was applied to characterize genomic loci, perform functional annotation and implement gene mapping. Default parameters were applied in the process unless otherwise specified.

In brief, using the 1000 Genomes Phase 3 EUR as a reference panel, SNPs with genome-wide significant mvGWAS $P < 5 \times 10^{-8}$ that had LD $r^2 < 0.6$ with any others were identified as independent significant SNPs. A fraction of the independent significant SNPs in approximate linkage equilibrium with each other at $r^2 < 0.1$ were considered as a lead SNP. LD blocks of significant SNPs located within 250 kb of each other were merged into one genomic locus. Candidate SNPs were defined as the SNPs that had LD $r^2 \geq 0.6$ with any of the independent significant SNPs in the genetic loci. The variants included those from the reference panel that might not have been included in the GWAS. The major histocompatibility complex region on chromosome 6 was excluded by default. The associated SNPs were annotated by FUMA based on functional categories, including ANNOVAR categories[56], combined annotation-dependent depletion scores[57], RegulomeDB scores[58] and chromatin state[59].

Gene mapping was performed using different mapping strategies, including positional mapping (physical distance within 10 kb), eQTL mapping (four brain-expression data repositories: PsychENCORE[60], CommonMind Consortium[61], BRAINEAC[62] and GTEx v8 Brain[63] with the significant level set at FDR of 0.05) and chromatin-interaction mapping (seven brain-related Hi-C chromatin conformation capture datasets: PsychENCORE EP link (one way)[60], PsychENCORE promoter anchored loops[61], HiC adult cortex[47], HiC foetal cortex[47], HiC dorsolateral prefrontal cortex[64], HiC hippocampus[64] and HiC neural progenitor cell[64]; 17 brain-related repositories in the Roadmap Epigenomics Project[65]: E053, E054, E067, E068, E069, E070, E071, E072, E073, E074, E081, E082, E003, E008, E007, E009 and E010). Genome-wide gene-based association analysis was also used to map genes by performing MAGMA

(v.1.08)[20] implemented in FUMA with complete mvGWAS summary statistics as input. MAGMA conducted multiple linear regression to obtain gene-based *P* values and the Bonferroni-corrected significant threshold was set at $P = 0.05/18877$.

Protein–protein interaction network

Protein network analysis was performed using STRING (http://string-db.org) with 27 significant genes appearing in all four mapping strategies as input. Information about protein–protein interactions is included in the STRING dataset from a variety of sources, including experimental data, publications and computational prediction techniques. Only links with a confidence score of at least medium confidence (confidence score of >0.4; default parameter) were kept.

Gene set enrichment analysis

All the mapped genes arising from mvGWAS was used as input for gene set enrichment analysis, using the 7,343 gene ontology biological process gene sets[66], on the basis of the hypergeometric test as implemented in GENE2FUNC of FUMA[17].

Heritability estimation and genetic correlations between neuroimaging measures

SNP heritability was estimated using the GREML method implemented in GCTA (v.1.93.2)[67]. Covariates were the same as those in mvGWAS. Heritability was also estimated using LDSC (v.1.0.1)[68] with univariate GWAS summary statistics as input for which genomic inflation factors (λ_{GC}) and LDSC intercepts were also calculated.

Genetic correlations among six hypothalamic measures and among hypothalamic measures and regional brain volumes were analysed using both LDSC and bivariate GCTA–GREML methods. When performing LDSC, a large-scale imaging GWAS of 19,629 UKB participants provided summary statistics of 101 cortical, subcortical and ventricular volumetric imaging phenotypes[69]. We used the Benjamini and Hochberg FDR method to adjust for multiple testing. Pre-computed European LD scores from the 1000 Genomes Project Phase 3 in the LDSC package were used. As recommended, we restricted our analyses to Hapmap3 SNPs and excluded the major histocompatibility region. When performing GCTA–GREML, individual-level data of the 101 imaging-derived phenotypes extracted using FreeSurefer aparc (UKB Category ID = 192) and aseg (UKB Category ID = 190) atlases were used (Supplementary Table 15).

Genetic correlations between the hypothalamus and neuropsychiatric traits and disorders

Both LDSC and bivariate GCTA–GREML were used to explore the genetic correlations among hypothalamic measures and function-related traits. Five different categories of traits were tested with summary statistics downloaded from the corresponding published GWAS for LDSC and with individual-level data acquired from certain UKB Field IDs for GCTA–GREML, including the following: sleep and circadian rhythms (chronotype[70], Field ID: 1180; daytime napping[71], Field ID: 1190; sleep duration[71], Field ID: 1160); risky behaviours (cigarettes per day[72], Field IDs: 3456 and 2887; frequency of drinking[72], Field ID: 1558; age at first sexual intercourse[73], Field ID: 2139); learning and cognition (reaction time[74], Field ID: 20023; incorrect pair matches in round[75], Field ID: 399); food intake and satiety (recent poor appetite or overeating[75], Field ID: 20511; body mass index[76], Field ID 21001); and parasympathetic–sympathetic activity (systolic blood pressure[77], Field ID: 4080; diastolic blood pressure[77], Field ID: 4079; heart rate[78], Field ID: 102) (Supplementary Tables 37 and 38). To investigate the hypothalamus-specific influence, the GCTA–GREML data were re-run adjusting for frontal, parietal, temporal and occipital cortices that were derived from the FreeSurefer aparc atlas[79]. We corrected for multiple testing through an FDR of 0.05 in both methods, and brain regions with *P* values passing FDR corrections of both sides were considered to be significantly correlated with the hypothalamus.

The genetic correlations between hypothalamic measures and neuropsychiatric disorders were evaluated using LDSC with publicly available GWAS summary statistics for 13 neuropsychiatric disorders: SCZ[80], BD[81], MD[82], GAD[83], PSD[84], ADHD[85], AN[86], ASD[87], OCD[88], TS[89], stroke[90], AD[91] and PD[92]. The sample sizes ranged from 9,725 to 446,696 individuals with European ancestry. Detailed information is summarized in Supplementary Table 39. Correlations with *P* values surviving FDR correction were considered significant.

Genetic overlap between the hypothalamus and neuropsychiatric disorders

The pleiotropy-informed conditional FDR method was used to identify genetic overlaps between the hypothalamus and common neuropsychiatric disorders, for which the conditional Q-Q plot, the conditional FDR method and the conjunctional FDR method were sequentially applied. The publicly available GWAS summary statistics for the 13 neuropsychiatric disorders were also used here as input. The analyses were run in Matlab R2018b and Python 3.7.7. A detailed description of genetic overlap analyses is provided in the Supplementary Information.

RVAS

RVAS was performed to further identify genes and rare genetic variants associated with the volume of the whole hypothalamus and each subregion using SKAT-O test through SAIGE-GENE+ (ref. 36) among unrelated British individuals. Variants were filtered to the rare ones (minor allele frequency < 0.01) and were annotated using SnpEff[93], and the most severe consequence was kept for each variant. Then variants were grouped into loss of function (annotated as stop gained, start lost, splice donor, splice acceptor, stop lost or frameshift) and probable deleterious missense (predicted as deleteriousness in SIFT[94]; PolyPhen2 HDIV and PolyPhen2 HVAR[95]; LRT[96]; and MutationTaster[97]) and were collapsed for each protein-coding gene. Similar to the GWAS analysis, age, age squared, sex, scanning site, intracranial volume and the first ten genetic principal components (calculated using whole-exome sequencing data) were adjusted. All phenotypic variables underwent inverse normalization and a relative coefficient cut-off of 0.05 for the sparse GRM for the variance ratio estimation was used. Owing to an insufficient number of common variants to calculate the variance ratio, the parameter --isCateVarianceRatio was set to 'false' in the RVAS. Bonferroni correction was used, and $P < 2.29 \times 10^{-7}$ (that is, 0.05/218219, all genes, all variant functional groups and all phenotypes together) was considered significant. For several genes of interest, we further performed single variant association analysis for the rare variants (minor allele counts ≥ 10) and the ultrarare variants, and variants with minor allele counts < 10 were collapsed into one group in the single variant association analysis.

MR analyses

Bidirectional MR analyses were conducted among the 6 hypothalamic measures and 13 neuropsychiatric disorders. Genetic instruments were selected at the primary *P* threshold of 5×10^{-8} and the secondary *P* threshold of 1×10^{-6}. The SNPs were clumped with a window size of 10,000 kb and a LD r^2 of 0.001 using the 1000 Genomes EUR reference panel. The IVW was conducted as the primary method given its high sensitivity for causal inference[98]. Another three methods were performed to complement and enhance the reliability of the results, including MR-Egger, the weighted median and the weighted mode. MR-Egger uses the slope coefficient of the Egger regression and can provide a robust estimate even if all instruments are invalid[99]. When the 'no measurement error' assumption in MR-Egger was violated, as indicated by an I^2_{GX} statistic of less than 0.90, the simulation extrapolation method was used for bias adjustment[100]. The weighted median method can generate reliable results when up to 50% instruments are invalid[101]. The weighted mode method presents less bias and lower type-I error rates under relaxed instruments assumptions[102]. When only

one genetic instrument was available, the Wald ratio method was used. We performed FDR corrections for 78 comparisons in both forward and reverse MR. Associations with $P_{IVW/Wald\ ratio}$ values surviving FDR corrections were considered significant. The MR-Egger intercept indicating the mean pleiotropic effect of all genetic variants was used to examine the potential bias of directional pleiotropy. Heterogeneity in the IVW estimates was assessed using Cochran's *Q*-test. The *F*-statistics and power calculation were based on previously described methods[103,104]. The R packages TwoSampleMR (v.0.5.5), simex (v.1.8) and ieugwasr (v.0.1.5) were used to implement these methods.

In MR investigating the association between gene expression with the strongest signal and the hypothalamic volume, we used univariate GWAS summary statistics of the whole hypothalamic volume as the outcome dataset. Summary statistics from eQTLGen datasets were extracted as the exposure dataset because of its largest sample size among the current eQTL datasets[105]. We set a LD pruning window size of 10,000 kb, an r^2 threshold of 0.5 and a *P* value threshold of 1×10^{-6} using the 1000 Genomes EUR reference panel. As above, IVW was the primary method used to conduct MR with three other MR methods. The MR-Egger intercept was also used for detecting directional pleiotropy. Leave-one-out and single-SNP analyses were performed to examine effects of outlying SNPs on MR estimates. To avoid confounder bias that interferes with the pathway between gene expression and the hypothalamus, we examined the associations between the instruments and potential confounders. Scarce evidence was found for established confounding factors phenotypically associated with both *ADAMTS8* levels and hypothalamus. Previous studies identified associations of *ADAMTS8* levels with pulmonary arterial hypertension and right ventricular function[40] that possibly linked to the hypothalamus[106]. In addition, socioeconomic status, education, drinking and smoking behaviour were reported to influence neuropsychiatric disorders and brain structure alterations[107], which may have a role in confounding the association between gene expression and the hypothalamus. Hence, we referred to the PhenoScanner v.2 database (http://www.phenoscanner.medschl.cam.ac.uk/), the NHGRI-EBI GWAS catalogue (https://www.ebi.ac.uk/gwas/docs/file-downloads/) and the recent GWAS summary statistics of these traits[72,108–112]. No confounder SNPs were identified.

Reporting summary

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

Data availability

The full summary statistics of multivariate and univariate GWAS for the hypothalamus can be found at the figshare website (https://figshare.com/projects/GWAS_summary_data_of_hypothalamus/165589). Summary statistics of regional brain measures are available at https://www.med.unc.edu/bigs2/data/gwas-summary-statistics/. Summary statistics of neuropsychiatric traits and disorders for genetic correlation analyses are summarized in Supplementary Tables 38 and 39. Summary statistics of eQTL were obtained through the eQTLGen website (https://molgenis26.gcc.rug.nl/downloads/eqtlgen/cis-eqtl/2019-12-11-cis-eQTLsFDR-ProbeLevel-CohortInfoRemoved-BonferroniAdded.txt.gz). The individual-level imaging and genetic data used in the current study are available through the UKB (https://biobank.ndph.ox.ac.uk/showcase/index.cgi, accession number 19542), the ABCD (https://nda.nih.gov/data_dictionary.html?source=ABCD%2BRelease%2B4.0&submission=ALL) and the IMAGEN (https://imagen2.cea.fr/account_manager). Data were used under licence and can be accessed through applications.

Code availability

This study used openly available software and codes, including MOSTest (https://github.com/precimed/mostest), PLINK (v.2.0; https://www.cog-genomics.org/plink/), FUMA (v.1.3.6; https://fuma.ctglab.nl/), MAGMA (v.1.08; https://ctg.cncr.nl/software/magma/, also implemented in FUMA), GCTA (v.1.93.2; http://cnsgenomics.com/software/gcta/), LDSC (v.1.0.1; https://github.com/bulik/ldsc/), STRING (https://www.stringdb.org/), Michigan Imputation Server (https://imputationserver.sph.umich.edu/), cFDR (https://github.com/precimed/pleiofdr/), METAL (v.2011-03-25; http://www.sph.umich.edu/csg/abecasis/Metal/), SAIGE-GENE+ (https://saigegit.github.io/SAIGE-doc/), R (v.4.0.3), Matlab R2018b and Python (v.3.10). Custom scripts for the analyses in this paper can be found at GitHub (https://github.com/wbs-whuer/GWAS_hypothalamus/).

References

1. Müller, H. L. et al. Hypothalamic syndrome. *Nat. Rev. Dis. Prim.* **8**, 24 (2022).
2. Yang, T. et al. Social control of hypothalamus-mediated male aggression. *Neuron* **95**, 955–970.e4 (2017).
3. Hu, R. K. et al. An amygdala-to-hypothalamus circuit for social reward. *Nat. Neurosci.* **24**, 831–842 (2021).
4. Riecher-Rössler, A. Oestrogens, prolactin, hypothalamic–pituitary–gonadal axis, and schizophrenic psychoses. *Lancet Psychiatry* **4**, 63–72 (2017).
5. Zheng, Z. et al. Hypothalamus–habenula potentiation encodes chronic stress experience and drives depression onset. *Neuron* **110**, 1400–1415.e6 (2022).
6. Ishii, M. & Iadecola, C. Metabolic and non-cognitive manifestations of Alzheimer's disease: the hypothalamus as both culprit and target of pathology. *Cell Metab.* **22**, 761–776 (2015).
7. Breen, D. P. et al. Hypothalamic volume loss is associated with reduced melatonin output in Parkinson's disease. *Mov. Disord.* **31**, 1062–1066 (2016).
8. van Erp, T. G. et al. Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. *Mol. Psychiatry* **21**, 547–553 (2016).
9. Hibar, D. P. et al. Cortical abnormalities in bipolar disorder: an MRI analysis of 6503 individuals from the ENIGMA Bipolar Disorder Working Group. *Mol. Psychiatry* **23**, 932–942 (2018).
10. Hibar, D. P. et al. Novel genetic loci associated with hippocampal volume. *Nat. Commun.* **8**, 13624 (2017).
11. Satizabal, C. L. et al. Genetic architecture of subcortical brain structures in 38,851 individuals. *Nat. Genet.* **51**, 1624–1636 (2019).
12. Bocchetta, M. et al. Detailed volumetric analysis of the hypothalamus in behavioral variant frontotemporal dementia. *J. Neurol.* **262**, 2635–2642 (2015).
13. Makris, N. et al. Volumetric parcellation methodology of the human hypothalamus in neuroimaging: normative data and sex differences. *NeuroImage* **69**, 1–10 (2013).
14. Baroncini, M. et al. MRI atlas of the human hypothalamus. *NeuroImage* **59**, 168–180 (2012).
15. Billot, B. et al. Automated segmentation of the hypothalamus and associated subunits in brain MRI. *NeuroImage* **223**, 117287 (2020).
16. van der Meer, D. et al. Understanding the genetic determinants of the brain with MOSTest. *Nat. Commun.* **11**, 3512 (2020).
17. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
18. Bahrami, S. et al. Distributed genetic architecture across the hippocampal formation implies common neuropathology across brain disorders. *Nat. Commun.* **13**, 3436 (2022).
19. Ou, Y. N. et al. The genetic architecture of human amygdala volumes and their overlap with common brain disorders. *Transl. Psychiatry* **13**, 90 (2023).
20. 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).

21. Wain, L. V. et al. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat. Genet.* **43**, 1005–1011 (2011).
22. Wang, X. D. et al. Nectin-3 links CRHR1 signaling to stress-induced memory deficits and spine loss. *Nat. Neurosci.* **16**, 706–713 (2013).
23. He, L. et al. Exome-wide age-of-onset analysis reveals exonic variants in ERN1 and SPPL2C associated with Alzheimer's disease. *Transl. Psychiatry* **11**, 146 (2021).
24. Allen, M. et al. Gene expression, methylation and neuropathology correlations at progressive supranuclear palsy risk loci. *Acta Neuropathol.* **132**, 197–211 (2016).
25. Mignot, E., Taheri, S. & Nishino, S. Sleeping with the hypothalamus: emerging therapeutic targets for sleep disorders. *Nat. Neurosci.* **5**, 1071–1075 (2002).
26. Jones, S. E. et al. Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. *Nat. Commun.* **10**, 343 (2019).
27. Aydogan, G. et al. Genetic underpinnings of risky behaviour relate to altered neuroanatomy. *Nat. Hum. Behav.* **5**, 787–794 (2021).
28. Floresco, S. B. Learning is a matter of history and relevance for lateral hypothalamus. *Nat. Neurosci.* **24**, 295–296 (2021).
29. Averbeck, B. B. & Murray, E. A. Hypothalamic interactions with large-scale neural circuits underlying reinforcement learning and motivated behavior. *Trends Neurosci.* **43**, 681–694 (2020).
30. Li, M. M. et al. The paraventricular hypothalamus regulates satiety and prevents obesity via two genetically distinct circuits. *Neuron* **102**, 653–667.e6 (2019).
31. Horvath, T. L. The hardship of obesity: a soft-wired hypothalamus. *Nat. Neurosci.* **8**, 561–565 (2005).
32. Zhou, J. J., Shao, J. Y., Chen, S. R. & Pan, H. L. Calcineurin controls hypothalamic NMDA receptor activity and sympathetic outflow. *Circ. Res.* **131**, 345–360 (2022).
33. Gruber, T. et al. Obesity-associated hyperleptinemia alters the gliovascular interface of the hypothalamus to promote hypertension. *Cell Metab.* **33**, 1155–1170.e10 (2021).
34. Casey, B. J. et al. The Adolescent Brain Cognitive Development (ABCD) study: imaging acquisition across 21 sites. *Dev. Cogn. Neurosci.* **32**, 43–54 (2018).
35. Mascarell Maričić, L. et al. The IMAGEN study: a decade of imaging genetics in adolescents. *Mol. Psychiatry* **25**, 2648–2671 (2020).
36. Zhou, W. et al. SAIGE-GENE+ improves the efficiency and accuracy of set-based rare variant association tests. *Nat. Genet.* **54**, 1466–1469 (2022).
37. Watanabe, K. et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **51**, 1339–1348 (2019).
38. Hall, J. E. & Hall, M. E. *Guyton and Hall Textbook of Medical Physiology* (Elsevier, 2020).
39. Yamagata, T. et al. The hypothalamic link between arousal and sleep homeostasis in mice. *Proc. Natl Acad. Sci. USA* **118**, e2101580118 (2021).
40. Omura, J. et al. ADAMTS8 promotes the development of pulmonary arterial hypertension and right ventricular failure: a possible novel therapeutic target. *Circ. Res.* **125**, 884–906 (2019).
41. Morihara, T. et al. Transcriptome analysis of distinct mouse strains reveals kinesin light chain-1 splicing as an amyloid-β accumulation modifier. *Proc. Natl Acad. Sci. USA* **111**, 2638–2643 (2014).
42. Alsabban, A. H., Morikawa, M., Tanaka, Y., Takei, Y. & Hirokawa, N. Kinesin *Kif3b* mutation reduces NMDAR subunit NR2A trafficking and causes schizophrenia-like phenotypes in mice. *EMBO J.* **39**, e101090 (2020).
43. Yin, X., Takei, Y., Kido, M. A. & Hirokawa, N. Molecular motor KIF17 is fundamental for memory and learning via differential support of synaptic NR2A/2B levels. *Neuron* **70**, 310–325 (2011).
44. Du, J. et al. A kinesin signaling complex mediates the ability of GSK-3β to affect mood-associated behaviors. *Proc. Natl Acad. Sci. USA* **107**, 11573–11578 (2010).
45. Gabery, S. et al. Volumetric analysis of the hypothalamus in Huntington disease using 3T MRI: the IMAGE-HD Study. *PLoS ONE* **10**, e0117593 (2015).
46. Goldstein, J. M. et al. Hypothalamic abnormalities in schizophrenia: sex effects and genetic vulnerability. *Biol. Psychiatry* **61**, 935–945 (2007).
47. Sha, Z. et al. The genetic architecture of structural left–right asymmetry of the human brain. *Nat. Hum. Behav.* **5**, 1226–1239 (2021).
48. Burgess, S., Davies, N. M. & Thompson, S. G. Bias due to participant overlap in two-sample Mendelian randomization. *Genet. Epidemiol.* **40**, 597–608 (2016).
49. LeBlanc, M. et al. A correction for sample overlap in genome-wide association studies in a polygenic pleiotropy-informed framework. *BMC Genomics* **19**, 494 (2018).
50. Schumann, G. et al. The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Mol. Psychiatry* **15**, 1128–1139 (2010).
51. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
52. Backman, J. D. et al. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* **599**, 628–634 (2021).
53. Jurgens, S. J. et al. Analysis of rare genetic variation underlying cardiometabolic diseases and traits among 200,000 individuals in the UK Biobank. *Nat. Genet.* **54**, 240–250 (2022).
54. Wu, B. S. et al. Genome-wide association study of cerebellar white matter microstructure and genetic overlap with common brain disorders. *NeuroImage* **269**, 119928 (2023).
55. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
56. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
57. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (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. Wang, D. et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, eaat8464 (2018).
61. Fromer, M. et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).
62. Ramasamy, A. et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat. Neurosci.* **17**, 1418–1428 (2014).
63. Consortium, G. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
64. Schmitt, A. D. et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* **17**, 2042–2059 (2016).
65. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
66. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
67. Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
68. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).

69. 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).
70. Maukonen, M. et al. Genetic associations of chronotype in the Finnish general population. *J. Biol. Rhythms* **35**, 501–511 (2020).
71. Dashti, H. S. et al. Genetic determinants of daytime napping and effects on cardiometabolic health. *Nat. Commun.* **12**, 900 (2021).
72. Liu, M. et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat. Genet.* **51**, 237–244 (2019).
73. Mills, M. C. et al. Identification of 371 genetic variants for age at first sex and birth linked to externalising behaviour. *Nat. Hum. Behav.* **5**, 1717–1730 (2021).
74. Davies, G. et al. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat. Commun.* **9**, 2098 (2018).
75. Jiang, L., Zheng, Z., Fang, H. & Yang, J. A generalized linear mixed model association tool for biobank-scale data. *Nat. Genet.* **53**, 1616–1621 (2021).
76. Yengo, L. et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
77. Evangelou, E. et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat. Genet.* **50**, 1412–1425 (2018).
78. Zhu, Z. et al. Genetic overlap of chronic obstructive pulmonary disease and cardiovascular disease-related traits: a large-scale genome-wide cross-trait analysis. *Respir. Res.* **20**, 64 (2019).
79. Desikan, R. S. et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**, 968–980 (2006).
80. Pardiñas, A. F. et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).
81. Stahl, E. A. et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).
82. 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).
83. Meier, S. M. et al. Genetic variants associated with anxiety and stress-related disorders: a genome-wide association study and mouse-model study. *JAMA Psychiatry* **76**, 924–932 (2019).
84. Duncan, L. E. et al. Largest GWAS of PTSD (N=20 070) yields genetic overlap with schizophrenia and sex differences in heritability. *Mol. Psychiatry* **23**, 666–673 (2018).
85. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
86. Watson, H. J. et al. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat. Genet.* **51**, 1207–1214 (2019).
87. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
88. International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OCGAS). Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol. Psychiatry* **23**, 1181–1188 (2018).
89. Yu, D. et al. Interrogating the genetic determinants of Tourette's syndrome and other tic disorders through genome-wide association studies. *Am. J. Psychiatry* **176**, 217–227 (2019).
90. Malik, R. et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537 (2018).
91. Kunkle, B. W. et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).
92. Blauwendraat, C. et al. Parkinson's disease age at onset genome-wide association study: defining heritability, genetic loci, and α-synuclein mechanisms. *Mov. Disord.* **34**, 866–875 (2019).
93. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
94. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nat. Protoc.* **11**, 1–9 (2016).
95. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **Chapter 7**, Unit7.20 (2013).
96. Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes. *Genome Res.* **19**, 1553–1561 (2009).
97. Schwarz, J. M., Rödelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **7**, 575–576 (2010).
98. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).
99. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
100. Bowden, J. et al. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. *Int. J. Epidemiol.* **45**, 1961–1974 (2016).
101. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
102. Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **46**, 1985–1998 (2017).
103. Burgess, S. & Thompson, S. G. Avoiding bias from weak instruments in Mendelian randomization studies. *Int. J. Epidemiol.* **40**, 755–764 (2011).
104. Burgess, S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. *Int. J. Epidemiol.* **43**, 922–929 (2014).
105. Võsa, U. et al. Large-scale *cis*- and *trans*-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* **53**, 1300–1310 (2021).
106. Sehgal, P. B., Yang, Y. M. & Miller, E. J. Hypothesis: neuroendocrine mechanisms (hypothalamus-growth hormone–STAT5 axis) contribute to sex bias in pulmonary hypertension. *Mol. Med.* **21**, 688–701 (2015).
107. Guo, J. et al. Mendelian randomization analyses support causal relationships between brain imaging-derived phenotypes and risk of psychiatric disorders. *Nat. Neurosci.* **25**, 1519–1527 (2022).
108. 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).
109. Jiang, L. et al. A resource-efficient tool for mixed model association analysis of large-scale data. *Nat. Genet.* **51**, 1749–1755 (2019).
110. Walters, R. K. et al. Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat. Neurosci.* **21**, 1656–1669 (2018).
111. Rhodes, C. J. et al. Genetic determinants of risk in pulmonary arterial hypertension: international genome-wide association studies and meta-analysis. *Lancet. Respir. Med.* **7**, 227–238 (2019).

112. Aung, N. et al. Genome-wide association analysis reveals insights into the genetic architecture of right ventricular structure and function. *Nat. Genet.* **54**, 783–791 (2022).

Acknowledgements

This study was supported by grants from the Science and Technology Innovation 2030 Major Projects (2022ZD0211600 to J.-T.Y.), the National Natural Science Foundation of China (82071201 to J.-T.Y., 82071997 to W.C.), the Shanghai Municipal Science and Technology Major Project (2018SHZDZX01 to J.-F.F.), the Postdoctoral Innovation Talents Support Program (BX20230087 to S.-D.C.), the Research Start-up Fund of Huashan Hospital (2022QD002 to J.-T.Y.), the Excellence 2025 Talent Cultivation Program at Fudan University (3030277001 to J.-T.Y.), the Shanghai Talent Development Funding for The Project (2019074 to J.-T.Y.), the Shanghai Rising-Star Program (21QA1408700 to W.C.), the 111 Project (B18015 to J.-F.F.), the ZhangJiang Laboratory, the Tianqiao and Chrissy Chen Institute, the State Key Laboratory of Neurobiology and Frontiers Center for Brain Science of Ministry of Education, and the Shanghai Center for Brain Science and Brain-Inspired Technology, Fudan University. The funders had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; preparation, review or approval of the manuscript; and decision to submit the manuscript for publication. Data acquisition and analyses were conducted using the UKB Resource under approved project number 19542. Participation of the UKB participants is gratefully appreciated. We also thank the UKB team for collecting and preparing data for analyses. The ABCD Study is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123 and U24DA041147. A full list of supporters is available at https://abcdstudy.org/federalpartners.html. A list of participating sites and a complete list of the study investigators can be found at https://abcdstudy.org/consortium_members/. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators. IMAGEN received support from the following sources: the European Union-funded FP6 Integrated Project IMAGEN (Reinforcement-related behaviour in normal brain function and psychopathology) (LSHM-CT- 2007-037286), the Horizon 2020 funded ERC Advanced Grant 'STRATIFY' (Brain network based stratification of reinforcement-related disorders) (695313), Human Brain Project (HBP SGA 2, 785907, and HBP SGA 3, 945539), the Medical Research Council Grant 'c-VEDA' (Consortium on Vulnerability to Externalizing Disorders and Addictions) (MR/N000390/1), the National Institutes of Health (NIH) (R01DA049238, a decentralized macro and micro gene-by-environment interaction analysis of substance use behaviour and its brain biomarkers), the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, the Bundesministeriumfür Bildung und Forschung (BMBF grants 01GS08152; 01EV0711; Forschungsnetz AERIAL 01EE1406A, 01EE1406B; Forschungsnetz IMAC-Mind 01GL1745B), the Deutsche Forschungsgemeinschaft (DFG grants SM 80/7-2, SFB 940, TRR 265, NE 1383/14-1), the Medical Research Foundation and Medical Research Council (grants MR/R00465X/1 and MR/S020306/1), the NIH-funded ENIGMA (grants 5U54EB020403-05 and 1R56AG058854-01), NSFC grant 82150710554 and European Union funded project 'environMENTAL', grant number 101057429. Further support was provided by grants from the ANR (ANR-12-SAMA-0004, AAPG2019-GeBra), the Eranet Neuron (AF12-NEUR0008-01-WM2NA; and ANR-18-NEUR00002-01-ADORe), the Fondation de France (00081242), the Fondation pour la Recherche Médicale (DPA20140629802), the Mission Interministérielle de Lutte-contre-les-Drogues-et-les-Conduites-Addictives (MILDECA), the Assistance-Publique-Hôpitaux-de-Paris and INSERM (interface grant), Paris Sud University IDEX 2012, the Fondation de l'Avenir (grant AP-RM-17-013), the Fédération pour la Recherche sur le Cerveau; the National Institutes of Health, Science Foundation Ireland (16/ERCD/3797), USA (Axon, Testosterone and Mental Health during Adolescence; RO1 MH085772-01A1) and by NIH Consortium grant U54 EB020403, supported by a cross-NIH alliance that funds Big Data to Knowledge Centres of Excellence.

Author contributions

Concept and design: J.-T.Y., W.C., Q.D. and J.-F.F. Acquisition, analysis or interpretation of data: all authors. Drafting of the manuscript: S.-D.C., B.-S.W., W.Z. and J.Y. Critical revision of the manuscript for important intellectual content: all authors. Statistical analyses: S.-D.C., B.-S.W., J.Y. and W.Z. Obtained funding: J.-T.Y., W.C., J.-F.F. and S.-D.C. Administrative, technical or material support: W.C., J.-F.F., Q.D. and J.-T.Y. Supervision: J.-F.F., W.C., J.-T.Y. and Q.D.

Competing interests

T.B. served in an advisory or consultancy role for eye level, Infectopharm, Lundbeck, Medice, Neurim Pharmaceuticals, Oberberg, Roche and Takeda. He received conference support or speaker's fees from Janssen, Medice and Takeda. He received royalties from Hogrefe, Kohlhammer, CIP Medien and Oxford University Press; the current work is unrelated to these relationships. G.J.B. has received honoraria from General Electric Healthcare for teaching on scanner programming courses. L.P. served in an advisory or consultancy role for Roche and Viforpharm and received speaker's fees from Shire. She received royalties from Hogrefe, Kohlhammer and Schattauer. The current work is unrelated to the above grants and relationships. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-023-01792-6.

Correspondence and requests for materials should be addressed to Jian-Feng Feng, Qiang Dong, Wei Cheng or Jin-Tai Yu.

Peer review information *Nature Human Behaviour* thanks Tie-Lin Yang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Shi-Dong Chen[1,32], Jia You[2,32], Wei Zhang[2,32], Bang-Sheng Wu[1,32], Yi-Jun Ge[1], Shi-Tong Xiang[2], Jing Du[1,2], Kevin Kuo[1], Tobias Banaschewski[3], Gareth J. Barker[4], Arun L. W. Bokde[5], Sylvane Desrivières[6], Herta Flor[7,8], Antoine Grigis[9], Hugh Garavan[10], Penny Gowland[11], Andreas Heinz[12], Rüdiger Brühl[13], Jean-Luc Martinot[14], Marie-Laure Paillère Martinot[14,15], Eric Artiges[14,16], Frauke Nees[3,7,17], Dimitri Papadopoulos Orfanos[9], Herve Lemaitre[9,18], Tomáš Paus[19,20], Luise Poustka[21], Sarah Hohmann[22], Sabina Millenet[3], Christian Baeuchl[23], Michael N. Smolka[23], Nilakshi Vaidya[24], Henrik Walter[12], Robert Whelan[25], Gunter Schumann[24,26], IMAGEN Consortium*, Jian-Feng Feng[2,27,28,29,30] ✉, Qiang Dong[1] ✉, Wei Cheng[1,2,27,28,31] ✉ & Jin-Tai Yu[1] ✉

[1]Department of Neurology and Institute of Neurology, Huashan Hospital, State Key Laboratory of Medical Neurobiology and MOE Frontiers Center for Brain Science, Shanghai Medical College, Fudan University, National Center for Neurological Disorders, Shanghai, China. [2]Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China. [3]Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [4]Department of Neuroimaging, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK. [5]Discipline of Psychiatry, School of Medicine and Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland. [6]Institute of Psychiatry, Psychology & Neuroscience, Social, Genetic, Developmental Psychiatry Centre, King's College London, London, UK. [7]Institute of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [8]Department of Psychology, School of Social Sciences, University of Mannheim, Mannheim, Germany. [9]NeuroSpin, CEA, Université Paris-Saclay, Gif-sur-Yvette, France. [10]Departments of Psychiatry and Psychology, University of Vermont, Burlington, VT, USA. [11]Sir Peter Mansfield Imaging Centre School of Physics and Astronomy, University of Nottingham, Nottingham, UK. [12]Department of Psychiatry and Psychotherapy CCM, Charité–Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany. [13]Physikalisch-Technische Bundesanstalt (PTB), Braunschweig and Berlin, Germany. [14]Institut National de la Santé et de la Recherche Médicale, INSERM U 1299 "Trajectoires développementales & psychiatrie", University Paris-Saclay, CNRS, Ecole Normale Supérieure Paris-Saclay, Centre Borelli, Gif-sur-Yvette, France. [15]AP-HP, Sorbonne University, Department of Child and Adolescent Psychiatry, Pitié-Salpêtrière Hospital, Paris, France. [16]Psychiatry Department, EPS Barthélémy Durand, Etampes, France. [17]Institute of Medical Psychology and Medical Sociology, University Medical Center Schleswig Holstein, Kiel University, Kiel, Germany. [18]Institut des Maladies Neurodégénératives, UMR 5293, CNRS, CEA, Université de Bordeaux, Bordeaux, France. [19]Departments of Psychiatry and Neuroscience, Faculty of Medicine and Centre Hosptalier Universitaire Sainte-Justine, University of Montreal, Montreal, Quebec, Canada. [20]Departments of Psychiatry and Psychology, University of Toronto, Toronto, Ontario, Canada. [21]Department of Child and Adolescent Psychiatry and Psychotherapy, University Medical Centre Göttingen, Göttingen, Germany. [22]Department of Child and Adolescent Psychiatry, Psychotherapy and Psychosomatics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. [23]Department of Psychiatry and Psychotherapy, Technische Universität Dresden, Dresden, Germany. [24]Centre for Population Neuroscience and Stratified Medicine (PONS), Department of Psychiatry and Neuroscience, Charité Universitätsmedizin Berlin, Berlin, Germany. [25]School of Psychology and Global Brain Health Institute, Trinity College Dublin, Dublin, Ireland. [26]Centre for Population Neuroscience and Precision Medicine (PONS), Institute for Science and Technology of Brain-inspired Intelligence (ISTBI), Fudan University, Shanghai, China. [27]Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence (Fudan University), Ministry of Education, Shanghai, China. [28]Fudan ISTBI–ZJNU Algorithm Centre for Brain-Inspired Intelligence, Zhejiang Normal University, Jinhua, China. [29]MOE Frontiers Center for Brain Science, Fudan University, Shanghai, China. [30]Zhangjiang Fudan International Innovation Center, Shanghai, China. [31]Shanghai Medical College and Zhongshan Hospital Immunotherapy Technology Transfer Center, Shanghai, China. [32]These authors contributed equally: Shi-Dong Chen, Jia You, Wei Zhang, Bang-Sheng Wu. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: jianfeng64@gmail.com; dong_qiang@fudan.edu.cn; wcheng@fudan.edu.cn; jintai_yu@fudan.edu.cn

IMAGEN Consortium

Tobias Banaschewski[3], Gareth J. Barker[4], Arun L. W. Bokde[5], Sylvane Desrivières[6], Herta Flor[7,8], Antoine Grigis[9], Hugh Garavan[10], Penny Gowland[11], Andreas Heinz[12], Rüdiger Brühl[13], Jean-Luc Martinot[14], Marie-Laure Paillère Martinot[14,15], Eric Artiges[14,16], Frauke Nees[3,7,17], Dimitri Papadopoulos Orfanos[9], Herve Lemaitre[9,18], Tomáš Paus[19,20], Luise Poustka[21], Sarah Hohmann[22], Sabina Millenet[3], Christian Baeuchl[23], Michael N. Smolka[23], Nilakshi Vaidya[24], Henrik Walter[12], Robert Whelan[25] & Gunter Schumann[24,26]

nature portfolio

Corresponding author(s): Jin-Tai Yu, Wei Cheng, Qiang Dong, Feng-Jian Feng

Last updated by author(s): Nov 7, 2023

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	No software was involved in data collection (data used is all directly available from UKB,ABCD, and IMAGEN as described in details in the paper)
Data analysis	R version 4.0.3 were used for data cleansing and visulization. MOSTest on matlabR2018b (https://github.com/precimed/mostest) for multivariable GWAS; PLINK version 2.0 (https://www.cog-genomics.org/plink/) for genetic data preparation; Michigan Imputation Server (https://imputationserver.sph.umich.edu/) for impuation; FUMA version v1.3.6(https://fuma.ctglab.n/l) for annotation of genomic locus; MAGMA v1.08 (https://ctg.cncr.nl/software/magma/, also implemented in FUMA) for gene analysis; GCTA v1.93.2(http://cnsgenomics.com/software/gcta/) , LDSC v1.0.1(https://github.com/bulik/ldsc/), condFDR/conjFDR (https://github.com/precimed/pleiofdr/) ,and python v3.10 for genetic association analysis; STRING (https://www.stringdb.org/) for protein-protein interaction; METAL version 2011-03-25 (http://www.sph.umich.edu/csg/abecasis/Metal/) for meta analysis of GWAS summary statistics; SAIGE-GENE+ (https://saigegit.github.io/SAIGE-doc/) for WES analysis.

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

Data

Policy information about availability of data

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The full summary statistics of multivariate and univariate GWAS for hypothalamus can be found at the figshare website (https://figshare.com/projects/GWAS_summary_data_of_hypothalamus/165589). Summary statistics of regional brain measures are available at https://www.med.unc.edu/bigs2/data/gwas-summary-statistics/. Summary statistics of neuropsychiatric traits and disorders for genetic correlation analysis were summarized in Supplementary Data 38, 39. Summary statistics of eQTL were obtained through eQTLGen website (https://molgenis26.gcc.rug.nl/downloads/eqtlgen/cis-eqtl/2019-12-11-cis-eQTLsFDR-ProbeLevel-CohortInfoRemoved-BonferroniAdded.txt.gz). The individual-level imaging and genetic data used in the present study are available through UKB (https://biobank.ndph.ox.ac.uk/showcase/index.cgi, accession number 19542), ABCD (https://nda.nih.gov/data_dictionary.html?source=ABCD%2BRelease%2B4.0&submission=ALL), and IMAGEN (https://imagen2.cea.fr/account_manager). Data were used under license and can be accessed through application.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	We took sex into considerations in our study and our findings could apply to both male and female. Sex in the UK Biobank, ABCD and IMAGENwas determined based on self-reporting data via questionaire, and all included participants gave written informed consent for sharing of individual-level data.
Population characteristics	This study included 32956 white British participants from UKB (17368 [54.2%] female; mean [SD] age, 64.3 [7.5] years), 4321 European participants from ABCD .(2014 [46.6%] female; mean [SD] age, 9.90 [0.71] years) , and 1685 European participants from IMAGEN ((859 [51.1%] female; mean [SD] age, 14.0 [0.46] years)) . Baseline descriptions can be found in S-data1. Statistics were reported with Mean (SD) for continuous variables and number (percentage) for categorical variables.
Recruitment	The UKB enrolled the participants aged 40-69 years between 2006 and 2010 for baseline assessments in 22 centers across the UK. The assessment visits comprised interviews and questionnaires covering lifestyles and health conditions, physical measures, biological samples, imaging, and genotyping. The database is linked to national health datasets, including primary care, hospital inpatient, death, and cancer registration data. The ABCD study is a new and ongoing project of very substantial size and scale involving 21 data acquisition sites, aming to recruit 11,500 children and follow them for ten years with extensive assessments at multiple timepoints.The IMAGEN study is the multicentre genetic-neuroimaging study with comprehensive behavioural and neuropsychological characterization, functional and structural neuroimaging and genome-wide association analyses of 2000 14-year-old adolescents.
Ethics oversight	This study is based on publicly available data with different levels of accessibility. The study was approved by the Institutional Review Boards of all participating institutions and was carried out in accordance with the approved protocols.The UK Biobank was approved by the National Health Service National Research Ethics (ref: 11/NW/0382). The ABCD study was approved by the central Institutional Review Board (IRB) at the University of California, San Diego, and the IMAGEN study was approved by the institutional ethics committee of King's College London, University of Nottingham, Trinity College Dublin, University of Heidelberg, Technische Universität Dresden, Commissariat à l'Energie Atomique et aux Energies Alternatives, and University Medical Center at the University of Hamburg.

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

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	No statistical methods were used to predetermine sample sizes, and eligible participants with both genetic and imaging data were included as much as possible.
Data exclusions	Participants without genetic data, MRI T1 data, and covariates available were excluded.

Replication	Multivariable GWASes were also performed in UKB non-white British population, ABCD, and IAMGEN cohorts to replicate the findings from UKB white British population.
Randomization	Covariates including age, age squared, sex, scanning site, intracranial volume, and the first 10 genetic principal components were adjusted in the study.
Blinding	Blinding was not applicable to this study as this study is observational.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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

Magnetic resonance imaging

Experimental design

Design type	Genome-wide association analysis using phenotypes derived from structural MRI
Design specifications	UK Biobank designed the imaging acquisition protocals including 6 modalities, covering structual, diffusion and functional iamging. In the current study, T1-weighted structural image was used and the image was acquired using straight sagittal orientation for 5 minutes.
Behavioral performance measures	In bivariate GCTA-GREML analysis to explore the genetic correlation between hypothalamic measures and function-related traits. Five different categories of phenotypes were tested including sleep and circadian rhythms (chronotype, field 1180; daytime napping, field 1190; sleep duration, field 1160), risky behaviours (cigarettes per day, field 3456 and 2887; drinks frequency, field 1558; age at first sexual intercourse, field 2139), learning and cognition (reaction time, field 20023; incorrect pair matches in round, field 399), food intake and satiety (recent poor appetite or overeating, field 20511; body mass index, field 21001), and water intake and cardiovascular activity (systolic blood pressure, field 4080; diastolic blood pressure, field 4079; heart rate, field 102).

Acquisition

Imaging type(s)	T1-weighted structural imaging
Field strength	3T
Sequence & imaging parameters	UKB: Resolution: 1×1×1 mm Field-of-view: 208×256×256 matrix Duration: 5 minutes 3D MPRAGE, sagittal, in-plane acceleration iPAT=2, prescan-normalise ABCD: Matrix: 256×256 Slices: 176 (Siemens), 225 (Philips), 208 (GE) FOV: 256×256 (Siemens,GE), 256×240 (Philips) %FOV phase: 100% (Siemens,GE), 93.75% (Philips) Resolution: 1.0×1.0×1.0 mm3 TR: 2500 ms (Siemens,GE), 6.31 ms (Philips) TE: 2.88 ms (Siemens), 2.9 ms (Philips), 2 ms (GE) TI: 1060 ms Flip Angle: 8 deg Parallel Imaging: 2x (Siemens,GE), 1.5 x 2.2 (Philips) MultiBand Acceleration: Off Phase partial Fourier: Off (Siemens,GE), NA (Philips) Acquisition Time: 7min12sec (Siemens), 5min38sec (Philips), 6min09sec (GE)

IMAGEN:
TR: 2300 ms
TE: 2.8 ms
TI: 900 ms
Parallel imaging/factor: N
NSA: 1
Scan duration:~ 09:20
Excitation flip angle (degrees): 8-9
2D/3D: 3D
Matrix freq dirn: 256
Matrix phase dirn: 256
Matrix size(3D): 160,170
FOV frequency: 28.0 cm
FOV phase: 94%
Slice thickness: 1.1 mm
Slice orientation: Sagittal
Slice acquisition direction: Left->Right

Area of acquisition: The whole brain were acquired, while the hypothalamic structures were the analytic targets .

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software: With T1-weighted MRI scans from the three cohorts, the whole hypothalamus and its subregions including anterior superior, anterior inferior, superior tuberal, inferior tuberal, and posterior hypothalamus were then delineated using an automated segmentation tool developed by Billot B. et al. (https://github.com/BBillot/hypothalamus_seg)

Normalization: see above

Normalization template: see above

Noise and artifact removal: see above

Volume censoring: see above

Statistical modeling & inference

Model type and settings: Generalized linear models

Effect(s) tested: The beta were obtained from GWAS for the generalized linear models

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

Anatomical location(s): An automated segmentation tool developed by Billot B. et al. (https://github.com/BBillot/hypothalamus_seg) was used to extract the volumes of hypothalamus and its subregions.

Statistic type for inference (See Eklund et al. 2016): Voxel-wise

Correction: False discovery rate (FDR) correction

Models & analysis

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