nature human behaviour

Article

https://doi.org/10.1038/s41562-023-01722-6

Genetic architectures of cerebral ventricles and their overlap with neuropsychiatric traits

Received: 23 April 2023

Accepted: 12 September 2023

Published online: 19 October 2023

Check for updates

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

The cerebral ventricles are recognized as windows into brain development and disease, yet their genetic architectures, underlying neural mechanisms and utility in maintaining brain health remain elusive. Here we aggregated genetic and neuroimaging data from 61,974 participants (age range, 9 to 98 years) in five cohorts to elucidate the genetic basis of ventricular morphology and examined their overlap with neuropsychiatric traits. Genome-wide association analysis in a discovery sample of 31,880 individuals identified 62 unique loci and 785 candidate genes associated with ventricular morphology. We replicated over 80% of loci in a well-matched cohort of lateral ventricular volume. Gene set analysis revealed enrichment of ventricular-trait-associated genes in biological processes and disease pathogenesis during both early brain development and degeneration. We explored the age-dependent genetic associations in cohorts of different age groups to investigate the possible roles of ventricular-trait-associated loci in neurodevelopmental and neurodegenerative processes. We describe the genetic overlap between ventricular and neuropsychiatric traits through comprehensive integrative approaches under correlative and causal assumptions. We propose the volume of the inferior lateral ventricles as a heritable endophenotype to predict the risk of Alzheimer's disease, which might be a consequence of prodromal Alzheimer's disease. Our study provides an advance in understanding the genetics of the cerebral ventricles and demonstrates the potential utility of ventricular measurements in tracking brain disorders and maintaining brain health across the lifespan.

Emerging evidence suggests that the cerebral ventricles play critical roles in brain development and degeneration[1]. On the one hand, the enlargement of the cerebral ventricles partly reflects age-related brain parenchymal atrophy[2]. Elucidating the genetic mechanisms of ventricular enlargement would thus help us better understand the biological processes of normal ageing and neurodegenerative diseases, such as Alzheimer's disease (AD) and cognitive impairment[3–5]. On the other hand, newborn neurons in mammals are not generated locally in the grey matter but originate from the proliferative neural stem cells that line the developing ventricles, suggesting the involvement of the ventricles in normal brain development[6,7]. Hence, in addition to brain atrophy, the abnormal morphology of the cerebral ventricles may indicate impaired neurogenesis or circulation in the cerebrospinal fluid (CSF) during brain development[8]. In support of this hypothesis, dysmorphologies (for example, dilation and asymmetry) of the cerebral ventricles have recently been implicated in a spectrum of

e-mail: wcheng@fudan.edu.cn; jintai_yu@fudan.edu.cn

developmental neuropsychiatric disorders, such as attention-deficit/hyperactivity disorder (ADHD)[5], autism spectrum disorder (ASD)[9,10], bipolar disorder (BP)[11], major depressive disorder (MDD)[12] and schizophrenia (SCZ)[13,14].

Given the observed associations of ventricular morphology with brain development and degeneration, uncovering the genetic basis of ventricular morphology might provide invaluable insights into the shared mechanism of the ventricles with neurodevelopment and neurodegenerative disorders. It could, in turn, facilitate managing brain disorders and maintaining brain health across the lifespan. On the assumption that ventricular volume is a highly polygenic trait, many common variants influencing ventricular volume undoubtedly remain to be identified. Two previous genome-wide association studies (GWASs) have identified several genetic loci associated with ventricular volumes in middle-aged to elderly adults[15,16]. However, thus far, it remains undetermined whether these identified ventricular-trait-associated loci contribute to brain development or degeneration. In addition, the causes and consequences of ventricular morphology and neuropsychiatric disorders are still poorly characterized. Furthermore, most previous studies focused solely on the volume of the lateral ventricles and ignored other parts of the ventricular system, yet their physiologies, genetics and implications for brain health may substantially differ[17]. For example, the inferior horns of the lateral ventricles are adjacent to the hippocampus; we therefore hypothesized that they are more closely linked to AD.

Our study has two primary objectives. First, we aim to elucidate the genetic basis of six ventricular phenotypes, including the volumes of the lateral, inferior lateral, third and fourth ventricles as well as asymmetries of the lateral and inferior lateral ventricles. To this end, we conducted GWASs to identify ventricular-trait-associated loci, genes and their biological properties in UK Biobank (UKB) and sought replication in an independent cohort (CHARGE) of middle- to older-aged adults. We also seek to generalize the original findings in children (ABCD), adolescents (IMAGEN) and younger adults (HCP), as well as the longitudinal ventricular changes in younger and older participants, to identify genetic determinants contributing to brain development and ageing. Second, we aim to leverage genetic data to estimate the value of ventricular traits in monitoring common neuropsychiatric disorders and promoting brain health across the lifespan. In this regard, we explored the genetic overlap of brain ventricles with neuropsychiatric traits using comprehensive integrative approaches under correlative and causal assumptions. As the volume of the inferior lateral ventricles showed robust predictive performance and is easily accessible in clinical practice, we performed post hoc analyses to determine whether it could be an endophenotype of AD.

Results

Genetic associations of ventricular morphology

Our study design is summarized in Fig. 1a and Extended Data Fig. 1. Altogether, we aggregated individual-level genetic and neuroimaging data from 38,441 participants in four cohorts, ranging in age from children to the elderly (9–82 years, Supplementary Table 1). Our discovery GWAS sample includes 31,880 white British UKB individuals free of prevalent dementia, stroke, hydrocephalus and intracranial tumours, with the quality control procedures detailed in the Methods. All tested ventricular traits were heritable in our common (minor allele frequency (MAF) ≥ 1%) variant-based heritability analysis on UKB individuals (Fig. 1b), and the volumetric metrics showed higher heritability (from 18% to 40%) than the asymmetric metrics (from 3% to 9%). Rare (MAF < 1%) variants could explain 2–5% of the total heritability of the ventricular traits in UKB individuals. Additional family-based heritability analysis on HCP individuals showed even higher heritability for all traits (from 51% to 85% for volumetric traits and from 29% to 30% for asymmetric traits, Supplementary Table 2).

We performed a univariate GWAS of common variants for the six ventricular traits mentioned above in the discovery UKB dataset. The univariate GWAS detected 42 loci harbouring 152 independent significant variants associated with any cerebral ventricular traits (Bonferroni-corrected $P < 8.33 \times 10^{-9}$, Fig. 1c and Supplementary Tables 3 and 4). The top signal, rs35565319, was a likely deleterious[18] (combined annotation-dependent depletion score, 18.29) exonic variant within *PAPPA* and was identified to be significantly associated with the volume of the fourth ventricle in our study ($P = 1.30 \times 10^{-39}$). This locus has been previously reported to be associated with the volume of a region adjacent to the fourth ventricle[19], the cerebellum (Supplementary Table 5). The second-strongest GWAS signal (rs2088882) was an intergenic variant near *GMNC* ($P = 2.12 \times 10^{-36}$). The locus *GMNC* was the top hit of the lateral ventricular volume as well as the inferior lateral ventricular volume. Intriguingly, this locus might also be involved in tau pathology[20,21], an established biomarker of AD (Supplementary Table 5).

To improve the power to detect genetic associations and cross-validate the results, we next conducted a multivariate GWAS of ventricular morphology using the most commonly used method in neuroimaging GWASs—namely, the multivariate omnibus statistical test (MOSTest)[19,22,23]. Our multivariate GWAS identified 56 loci associated with ventricular morphology, of which 20 were not identified by any univariate GWAS, contributing to a total of 62 ventricular-trait-associated loci (Figs. 1c and 2a and Supplementary Tables 3 and 6). The top hit in the multivariate GWAS was rs7247748, an intronic variant in *GPATCH1*, which also showed a significant association with the volume of the third ventricle in our univariate GWAS analysis (Fig. 1c). Moreover, the multivariate GWAS identified several new ventricular-trait-associated loci that had not been identified by univariate analysis and exhibited pleiotropic effects with other neuropsychiatric traits, such as locus 1q21.3 with multiple cognitive traits and locus 1q31.3 with various psychiatric disorders (Supplementary Table 5).

We sought to replicate the independent significant variants ($P < 8.33 \times 10^{-9}$) in the CHARGE consortium, whose demographics (mean age, 65 years; 56% female; $N = 23{,}533$) were similar to those in our discovery dataset (mean age, 64 years; 54% female; $N = 31{,}880$). Most (83%) independent significant variants were replicated in CHARGE after Bonferroni correction (Fig. 2b and Supplementary Table 7). Notably, we replicated all seven loci (3q28, 7p22.3, 10p12.31, 11q23.1, 12q23.3, 16q24.2 and 22q13.1) associated with the lateral ventricular volume previously reported by CHARGE[15]. Furthermore, we detected six new loci (1p34.2, 1q25.3, 9q31.3, 14q32.12, 17q24.1 and 20q11.22) associated with the lateral ventricular volume, which had never been reported in other cohorts. These new loci are possibly involved in immunological and metabolic mechanisms (Supplementary Table 5). For other ventricular traits, the comparison with previous studies was unable to perform due to the lack of published data.

Additional sensitivity analyses were performed to validate the robustness of independent significant single nucleotide polymorphisms (SNPs). We first asked whether the associations could be affected by collider bias introduced by intracranial volume (ICV). In genetic association analysis without ICV as a covariate, almost all (97%) independent significant signals identified in the discovery GWAS remained genome-wide significant. In the association analysis further excluding samples involved in the longitudinal analysis, 86% of independent significant variants identified in the discovery stage passed the genome-wide significance threshold. We also conducted a sensitivity test using SAIGE software and validated 85% of the initially identified independent significant associations. All the directions of the genetic effects of the independent significant signals in three sensitivity tests were consistent with our initial analysis, suggesting the robustness of our results (Supplementary Tables 8 and 9).

Gene-level analysis and functional enrichment

Gene-level analyses were carried out using genome-wide gene-based association analyses (GWGAS) as well as the positional, eQTL and 3D chromatin interaction mapping strategies through FUMA (Methods)[24].

https://doi.org/10.1038/s41562-023-01722-6

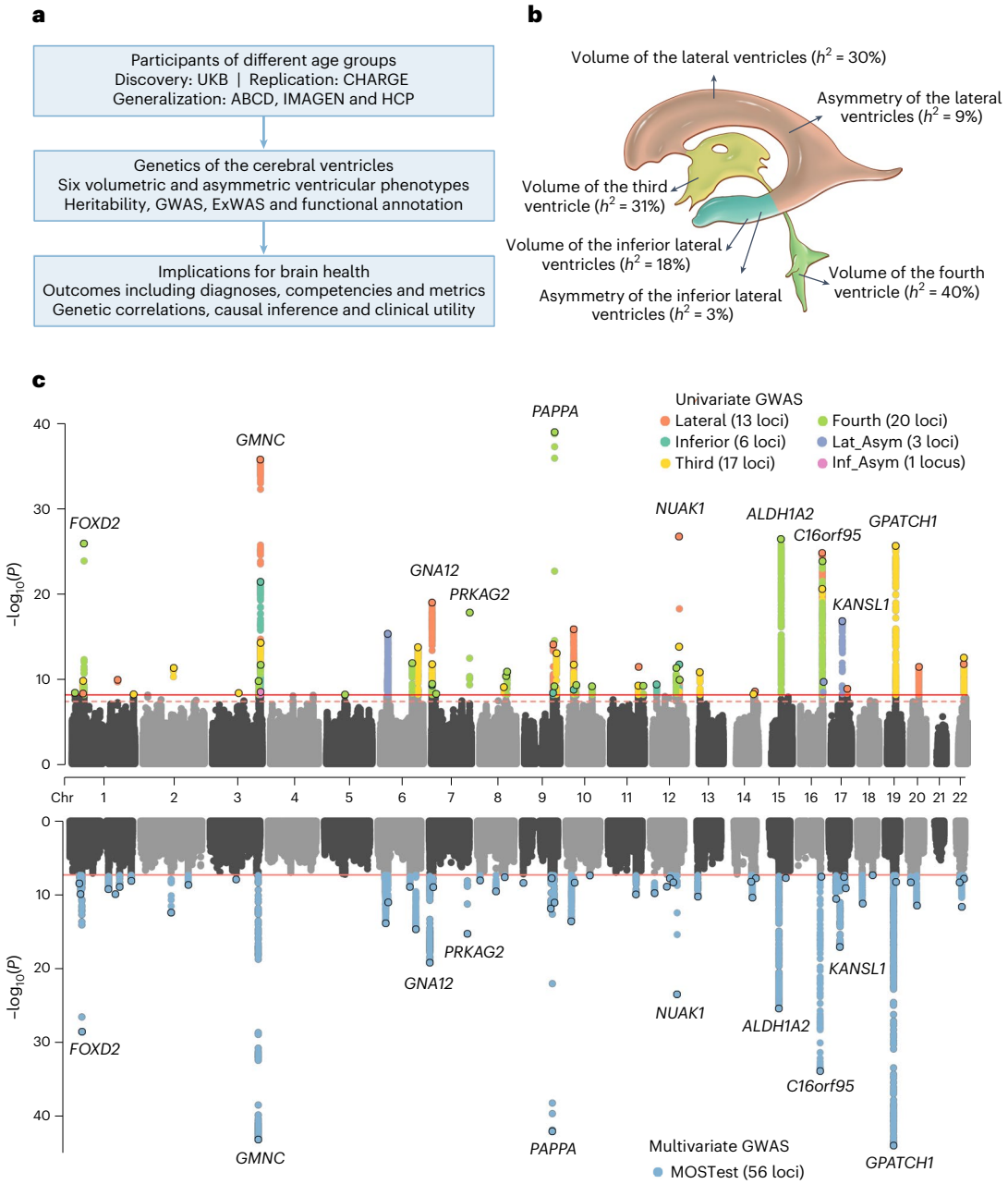


Fig. 1 | Analysis overview. a, Schematic diagram of the study design. A more detailed version can be found in Extended Data Fig. 1. **b**, Ventricular traits and their common-variant-based heritability. **c**, Genetic loci identified through univariate and multivariate common-variant GWAS. The orange lines indicate the genome-wide significance threshold (5.0×10^{-8}), and the red line indicates the Bonferroni-corrected threshold for the univariate GWAS (8.3×10^{-9}). Significant signals are coloured by their traits, and the top signals are annotated by their nearest genes. Chr, chromosome; ExWAS, exome-wide association study.

The three mapping strategies and GWGAS identified 784 unique genes associated with any ventricular traits, of which 56 genes had the most robust evidence, as they were simultaneously mapped by all four strategies (Fig. 2c,d and Supplementary Tables 10–13). Moreover, we identified an additional gene (*C4A*) according to the gene-based rare variant association analysis (MAF < 1%), whose rare and likely deleterious missense variants were significantly associated with the volume of the third ventricle with relatively large effects (MAF = 5.38×10^{-5}, $\beta = 2.31$, $P_{adj} = 0.02$, Supplementary Tables 14 and 15). Overall, we identified 785 ventricular-trait-associated genes, nearly half of which overlapped among multiple ventricular traits, especially among the volumetric traits (Fig. 2c and Supplementary Table 16).

To investigate the biological mechanisms underlying ventricular-trait-associated genes, we next performed gene set enrichment analysis on the FUMA platform[24]. The FUMA enrichment analysis yielded 262 unique Gene Ontology terms, pathways and GWAS Catalog phenotypes, while the most significant gene sets were driven by the asymmetry of the lateral ventricles (Fig. 2e). The top enriched biological functions involved developmental biology, chromatin organization, gene expression, signal transduction and cell cycle (Extended Data Fig. 2). We also identified several enriched phenotypes related to mood and cognition, such as SCZ, ASD, neuroticism, sociability, intelligence and cognitive ability (Supplementary Tables 17 and 18). Drug enrichment analysis was additionally conducted using ventricular-trait-associated genes through DSigDB (Supplementary Table 19)[25]. Several medications treating mental/mood disorders were enriched, such as azacyclonol (enrichment $P_{adj} = 6.21 \times 10^{-24}$), thioridazine (enrichment $P_{adj} = 5.58 \times 10^{-15}$) and

https://doi.org/10.1038/s41562-023-01722-6

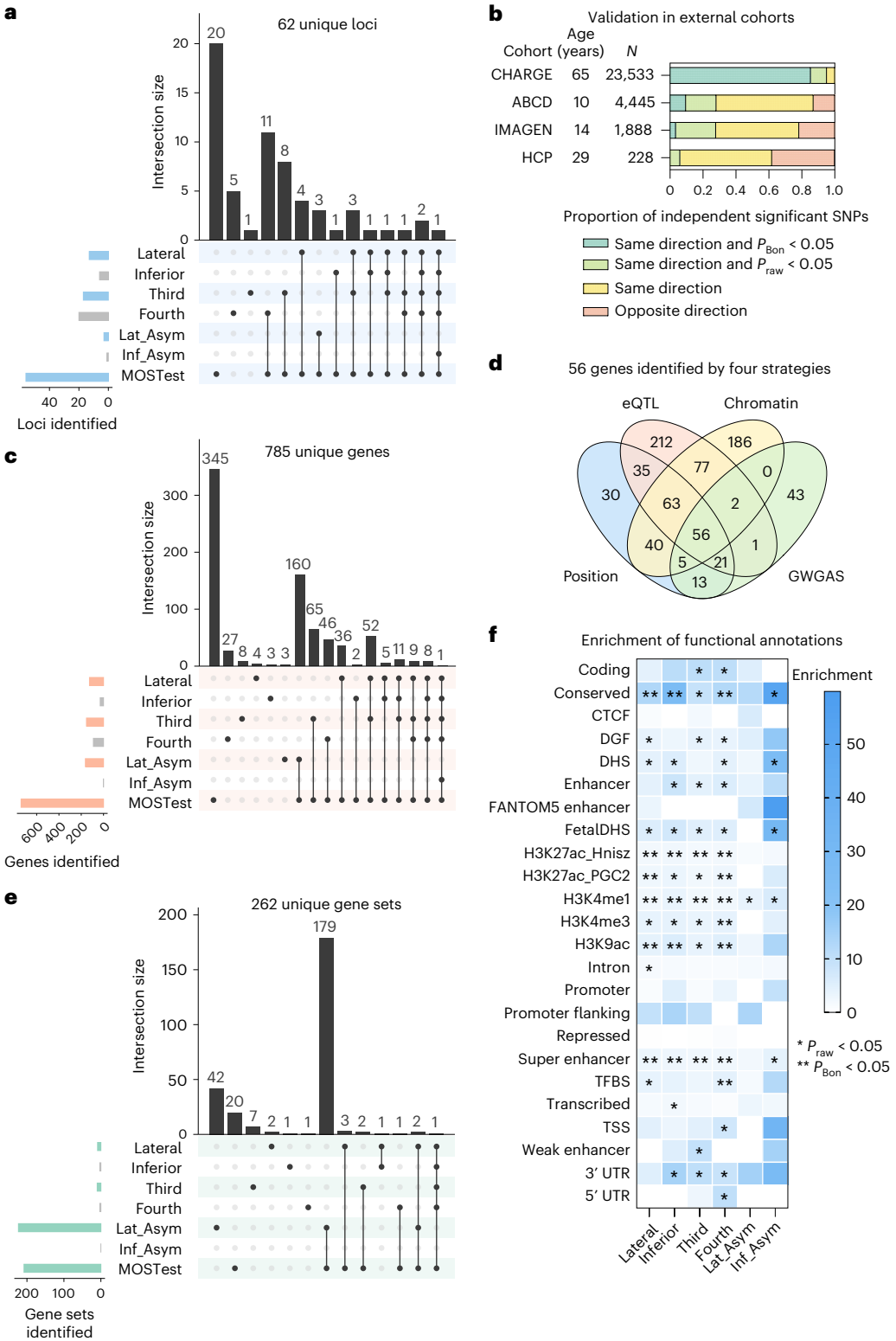


Fig. 2 | Characterization of ventricular-trait-associated loci. a, Upset plot showing the numbers of unique and shared loci among ventricular traits. **b**, Proportion of validated independent significant SNPs in the replication and three generalization datasets. The validation estimates are derived from univariate GWASs. **c**, Upset plot showing the numbers of unique and shared genes among ventricular traits. **d**, Venn diagram displaying 56 robust ventricular-trait-associated genes identified by four strategies. **e**, Upset plot showing the numbers of unique and shared gene sets among ventricular traits. **f**, Heritability enrichment for functional SNP categories of ventricular traits. Colour saturation represents the estimate for enrichment. Significant enrichment was defined as an enrichment estimate >1 with $P < 0.05$ after Bonferroni correction in stratified LDSC. CTCF, CCCTC-binding factor; DGF, digital genomic footprinting; DHS, DNase I hypersensitive site; TFBS, transcription factor binding site; TSS, transcription start site; UTR, untranslated region.

chlorprothixene (enrichment $P_{adj} = 1.17 \times 10^{-13}$), suggesting that the morphology of the cerebral ventricles might be altered by these drugs, especially during brain development.

Cell-type enrichment analysis was carried out to identify which cell types in the brain might be responsible for ventricular development and degeneration. Enriched cell types for the volume of the lateral ventricles included glia (astrocytes, oligodendrocytes and microglia) and endotheliocytes. Astrocytes were also enriched in the volume of the inferior lateral ventricles. Replicating neuronal progenitors were enriched for the volume of the fourth ventricle, suggesting that the volume of the fourth ventricle is more likely to be determined by developmental stages and remain stable in adult stages (Supplementary Table 20).

As different functional categories might contribute to heritability disproportionately, we applied the baseline model in stratified linkage disequilibrium score regression (LDSC) to investigate the enriched functional elements of ventricular traits[26]. We demonstrated a considerable enrichment of heritability in conserved regions in almost all ventricular traits except the asymmetry of the lateral ventricles (estimates for enrichment, 9.34 to 51.44; $P = 5.80 \times 10^{-8}$ to 3.30×10^{-2}), suggesting the biological importance of the conserved regions in maintaining normal ventricular morphology. We also observed significant enrichment of super-enhancers and histone marks (H3K4me1, H3K4me3, H3K9ac and H3K27ac) in volumetric traits, which indicates epigenetic gene regulation (Fig. 2f and Supplementary Table 21).

Concordance and discordance among different age groups

As all volumetric metrics and the asymmetry of the inferior lateral ventricles showed positive associations with age (Extended Data Fig. 3 and Supplementary Table 22), we hypothesized that some genetic loci might exert different influences on ventricular traits across the lifespan. To test this hypothesis, we explored the concordance and discordance of genetic effects on ventricular traits among individuals of different age groups. Concordant genetic results indicate that the genetic association is robust in cohorts across the lifespan. In contrast, discordant genetic effects among different age groups might suggest that the loci are involved in brain development or neurodegeneration. For example, ageing and dementia might be better understood when variants that mainly affect ventricular enlargement in older age are identified.

We first asked whether there were concordant association results of independent signals in cohorts of different age groups, including ABCD, IMAGEN, HCP, CHARGE and UKB (Fig. 2b and Supplementary Tables 7 and 23). We found that variants in locus 15q21.3 were consistently significantly associated with the volume of the fourth ventricle in all tested cohorts (Fig. 3a). Significant variants in locus 15q21.3 were mapped to *ALDH1A2*, a gene encoding an enzyme that converts retinaldehyde to retinoic acid. Retinoic acid is a hormonal signalling molecule involved in developing and adult tissues[27]. Another example of concordance was the top GWAS signal (locus 3q28) associated with the lateral and third ventricular volumes; that is, all cohorts except HCP (the cohort with the smallest sample size) showed consistent significant results with extremely low *P* values (lateral, $P = 5.20 \times 10^{-10}$ to 6.29×10^{-10}; third, $P = 6.07 \times 10^{-11}$ to 1.61×10^{-6}; Fig. 3b). The top signals in locus 3q28 were located near *GMNC*, which regulates DNA replication[28].

Next, the discordant genetic effects among individuals of different age groups were evaluated by meta-regression by age. The genetic effects of independent significant signals were first tested in ten subgroups from four independent cohorts with individual-level data, including ABCD (age range, 9–11 years), IMAGEN (age range, 13–20 years), HCP (age range, 22–36 years) and seven UKB sub-cohorts (with an age gap every five years; Supplementary Table 24). Meta-regression of SNP effects according to age was subsequently conducted using linear and quadratic models after the alignment of alleles. This analysis yielded four variants whose genetic effects were linearly or quadratically associated with age after Bonferroni correction (Fig. 3c–f and Supplementary Table 25). Notably, the top GWAS signal (locus 3q28) of the lateral and the inferior lateral ventricular volume is supposed to be mainly involved in brain development, since its genetic effect attenuated but persisted in older adults (Fig. 3c,d,g,h). When the UKB subgroups were 1:1 matched by sex, 93.4% of the conclusions were not changed (Supplementary Table 26).

We also leveraged longitudinal data on ventricular traits in neurodevelopmental cohorts (ABCD/IMAGEN; combined $N = 4{,}547$; mean age at baseline, 11 years) and the ageing cohort (UKB; $N = 3{,}646$; mean age at baseline, 62 years) to investigate whether the independent significant signals identified by the discovery GWAS could be validated in the analyses of annual change rates for the same ventricular traits. We observed that 9% of the tested variants in ABCD/IMAGEN and 13% of the variants in UKB were nominally significant ($P < 0.05$) in the longitudinal analyses and had consistent direction with those in the discovery GWAS, but none survived the Bonferroni correction (Supplementary Table 27). In both the longitudinal analysis and meta-regression, two loci (7p22.3 and 16q24.2) consistently exerted much stronger genetic influences on lateral ventricular volume in neurodegeneration than in neurodevelopment (Fig. 3h and Supplementary Tables 25 and 27). In addition, their top variants, rs798513 in 7p22.3 and rs4843555 in 16q24.2, were reported to be associated with a wide range of age-related traits, including cognition, insomnia and metabolic features (atlas.ctglab.nl/PheWAS)[29].

Finally, BrainSpan analyses were conducted to provide additional insights into the expression patterns of ventricular-trait-associated genes from the early prenatal period to middle adulthood. We observed significant prenatal upregulation of genes associated with the volume (19 weeks post-conception) and asymmetry (8 weeks post-conception) of the lateral ventricles, suggesting that the genetic influences of ventricular-trait-associated genes on ventricular morphology might start from early prenatal stages (Supplementary Table 28).

Genetic overlap with neuropsychiatric traits

To understand the genetic overlap between ventricular traits and outcomes for brain health, we estimated their genetic correlations by LDSC[30] and assessed causal relationships via bidirectional Mendelian randomization (MR) analyses[31]. Additionally, we conducted polygenic risk score (PRS) analysis to examine whether individuals with a higher genetic liability to brain disorders exhibited statistically different ventricular traits compared with others. Furthermore, we employed a multivariate Cox proportional hazard regression model to validate the genetically inferred causal relationships using clinical data.

We first estimated the genetic correlations between six ventricular traits and outcomes for brain health[32], including clinical diagnoses, competencies/behaviours and metrics of brain structure or function (Fig. 4). LDSC analysis did not identify any significant correlations between ventricles and clinical diagnoses. For competencies and behaviours, significant positive correlations were observed between larger ventricles and longer reaction time, heavier drinking and higher general risk tolerance. Leftward asymmetry of the lateral ventricles was genetically correlated with left-handedness. For metrics of brain structure or function, larger ventricles were significantly correlated with thinner cerebral cortex, more white matter hyperintensities and lower p-tau levels in CSF (Supplementary Table 29). Specifically, the negative genetic correlation between inferior lateral ventricular volume and CSF p-tau ($r_g = -0.77$, $P = 2.00 \times 10^{-3}$) was not a product of collider bias introduced by ICV since we reperformed LDSC using summary statistics without adjusting for ICV and got similar results ($r_g = -0.74$, $P = 1.70 \times 10^{-3}$).

We then conducted brain-wide LDSC analyses to estimate the genetic correlations between ventricular traits and brain subregions (Fig. 5). Genetically predicted larger ventricles were correlated

https://doi.org/10.1038/s41562-023-01722-6

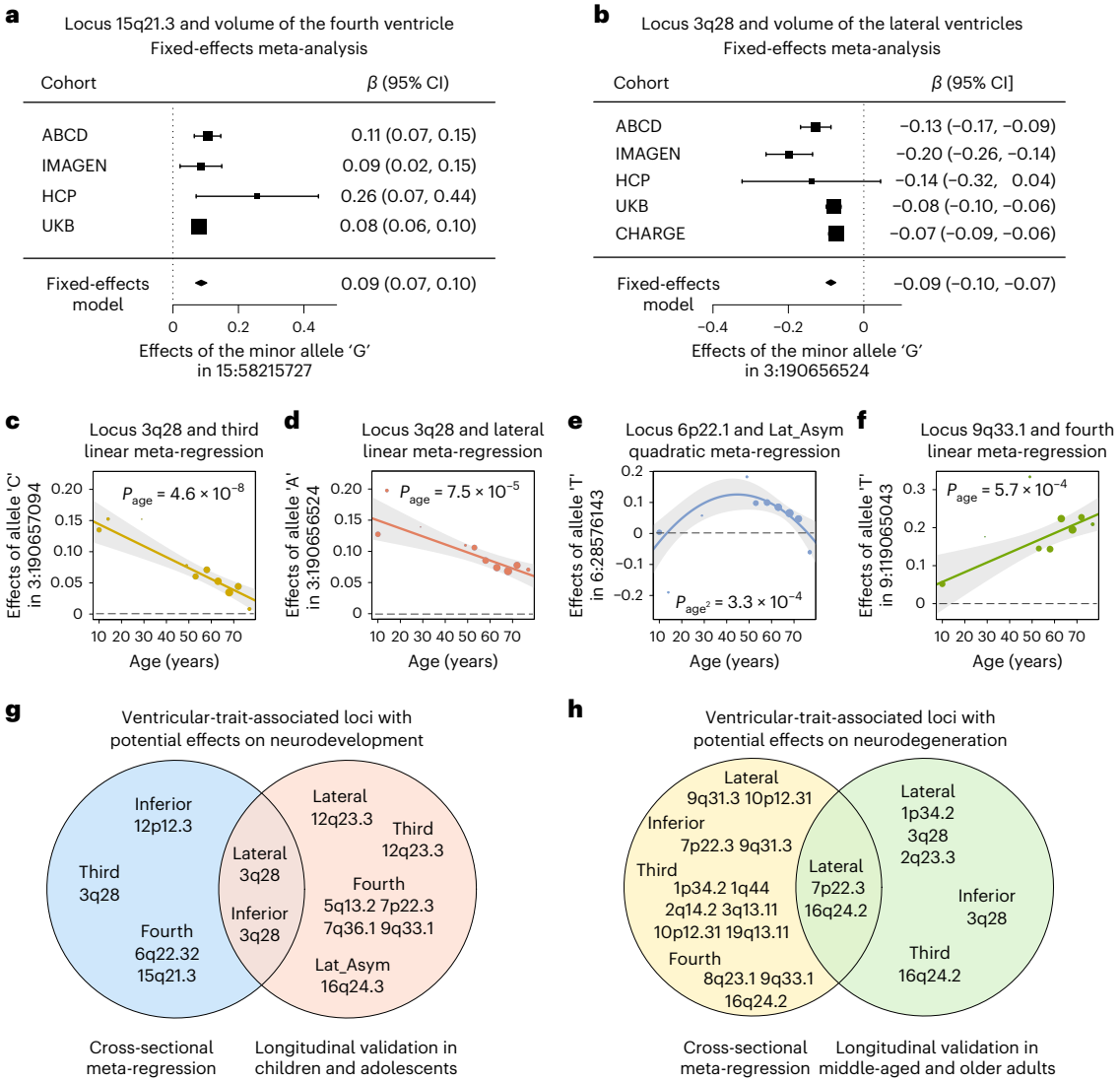


Fig. 3 | Concordance and discordance among individuals of different age groups. a, Fixed-effects meta-analysis of rs4775006 (15:58215727) in locus 15q21.3 for the volume of the fourth ventricle (N = 38,441 biologically independent samples). The effects were directionally consistent and significant among all cohorts. The data are presented as β values with 95% CIs. **b**, Fixed-effects meta-analysis of the top GWAS signal in the volume of the lateral ventricles—that is, rs2088882 (3:190656524) in locus 3q28 (N = 61,974 biologically independent samples). The data are presented as β values with 95% CIs. **c**–**f**, Meta-regression of SNP effects according to age was conducted using linear and quadratic models. The plots show the results for four variants that passed the Bonferroni correction. The x axis represents the median age of each cohort, and the y axis denotes the genetic effect of the minor allele on cross-sectional ventricular traits. The dot size indicates the $-\log_{10} P$ value of the genetic effect of the minor allele on cross-sectional ventricular traits. The error band around the regression line indicates the 95% CI. **g**,**h**, Identification of ventricular-trait-associated loci with potential effects (nominal $P < 0.05$) on neurodevelopment (**g**) or neurodegeneration (**h**). We employed two complementary methods, including meta-regression of cross-sectional genetic association results by age and longitudinal validation analysis in children/adolescents (ABCD and IMAGEN) and an ageing cohort (UKB). We found that one locus (3q28) had potential effects on neurodevelopment, and two loci (7p22.3 and 16q24.2) had potential effects on neurodegeneration, by combining the results of the two methods.

with smaller volumes of subcortical regions, including the thalamus, accumbens and putamen. Larger ventricles were also genetically correlated with thinner superior temporal, lingual and cuneus cortexes. Genetically predicted larger lateral ventricles were correlated with a smaller pericalcarine area but a larger paracentral area. Suggestive correlations between ventricular asymmetries and metrics of their adjacent brain regions were also observed, but none of these estimates survived correction for multiple testing (Supplementary Table 30).

Due to data limitation, MR, PRS and Cox analyses were restricted to major neuropsychiatric diseases, which play fundamental roles in brain health[32]. Notably, reverse causality of AD on inferior lateral ventricular volume was observed in all MR approaches (inverse-variance-weighted β (β_{IVW}) 0.03, $P = 9.50 \times 10^{-3}$; $\beta_{Egger} = 0.04$, $P = 9.45 \times 10^{-3}$; $\beta_{weighted\ median} = 0.04$, $P = 6.29 \times 10^{-3}$; $\beta_{weighted\ mode} = 0.04$, $P = 3.84 \times 10^{-3}$), with no heterogeneity ($P = 0.25$) or pleiotropy ($P = 0.21$) detected (Extended Data Fig. 4 and Supplementary Tables 31–34). Moreover, individuals with higher PRS to AD also showed larger inferior lateral ventricular volumes in the fully adjusted model ($\beta = 0.02$, $P = 4.76 \times 10^{-5}$, N = 31,843, Supplementary Table 35). Multivariate Cox regression analyses further demonstrated that each s.d. increase of the inferior lateral ventricular volume exhibited hazard ratios (HRs) for incident AD and dementia of 5.04 (95% confidence interval (CI), 2.56–9.91; $P = 2.78 \times 10^{-6}$; $N_{case} = 15$; $N_{total} = 15{,}198$) and 2.92 (95% CI, 2.02–4.22; $P = 1.20 \times 10^{-8}$; $N_{case} = 44$; $N_{total} = 15{,}198$; Supplementary Table 36). The risks remained significant after restricting analysis in individuals with

https://doi.org/10.1038/s41562-023-01722-6

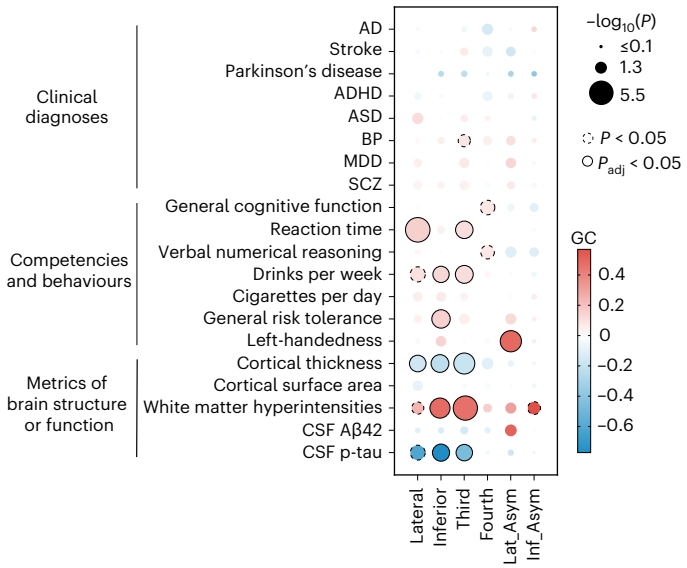


Fig. 4 | Genetic correlations between cerebral ventricles and outcomes for brain health. The genetic correlations were estimated using LDSC. Outcomes for brain health include clinical diagnoses, competencies and behaviours, and metrics of brain structure or function. The dot colour represents the magnitude of positive (red) or negative (blue) genetic correlation, while the dot size denotes the $-\log_{10} P$ value of the genetic correlation. Correlation estimates with nominal $P < 0.05$ are annotated with dashed circles, and those that survived false discovery rate corrections are annotated with solid circles. GC, genetic correlation.

a follow-up period of more than three years (for AD, HR = 5.40; 95% CI, 2.49–11.70; $P = 1.93 \times 10^{-5}$; for dementia, HR = 2.89; 95% CI, 1.81–4.61; $P = 8.05 \times 10^{-6}$).

Inferior lateral ventricular volume as an endophenotype for AD

Although our analyses also demonstrated other associations between ventricles and neuropsychiatric disorders (Fig. 6), we considered that only the association between the inferior lateral ventricular volume and AD was extremely unlikely to be falsely positive due to the consistency among multiple analytic approaches and the strong statistical significance. Since the inferior lateral ventricles are adjacent to the hippocampus, which is well known as the affected brain region in AD, we reperformed Cox regression analyses of incident AD and dementia, including the volume of the hippocampus as an additional covariate. Surprisingly, the inferior lateral ventricular volume showed the strongest significance and largest effect size among all included brain regions (Extended Data Fig. 4 and Supplementary Table 37). We also examined the association between the inferior lateral ventricular volume and parental history of AD or related dementias (ADD) among individuals free of dementia. Multivariate logistic analysis showed that each s.d. increase of the inferior lateral ventricular volume was associated with higher odds of ADD family history (odds ratio (OR), 1.05; $P = 2.08 \times 10^{-3}$), and the association became stronger when adjusting for parental age at death (OR = 1.07, $P = 7.03 \times 10^{-4}$).

To validate the value of the inferior lateral ventricular volume for predicting AD risk, we replicated the multivariate Cox proportional hazard regression analysis in the ADNI cohort, which was specifically designed to detect AD early and track its progression. During a mean follow-up time of 4.5 years, 343 of 1,633 participants developed incident AD. We found that a one-s.d. increase in the volume of the inferior lateral ventricles was associated with a 62% increase in the risk of incident AD in the base model (HR = 1.62; 95% CI, 1.40–1.87; $P = 5.54 \times 10^{-11}$). Again, the association was independent of the hippocampus (additionally adjusted for hippocampus, HR = 1.28; 95% CI, 1.08–1.52; $P = 3.69 \times 10^{-3}$). The association remained significant after the restriction in individuals with a follow-up period of more than three years (base model, HR = 1.53; 95% CI, 1.30–1.79; $P = 1.93 \times 10^{-7}$; additionally adjusted for hippocampus, HR = 1.26; 95% CI, 1.05–1.52; $P = 2.05 \times 10^{-2}$; Supplementary Table 38).

The above results reveal that the increased volume of the inferior lateral ventricles might be a consequence of prodromal AD and an endophenotype for predicting AD risk. To better understand the common etiological mechanism shared between the two traits, we looked up all genes associated with the inferior lateral ventricular volume on the AD Knowledge Portal for their general relevance to AD[33]. We identified two genes (*GNA12* and *LPAR1*) that are robustly associated with the inferior lateral ventricular volume and present substantial supporting evidence from genomics, genetics and the existing literature on AD (Supplementary Table 39). *GNA12* encodes G-protein subunit alpha-12, involved as the modulator or transducer in various transmembrane signalling systems. The integral membrane protein encoded by *LPAR1* is lysophosphatidic acid receptor 1 and belongs to the G-protein coupled receptor 1 family. Collectively, both genes were enriched in brain tissues and oligodendrocyte precursor cells/oligodendrocytes, the myelinating glia of the central nervous system[34].

Discussion

Here we demonstrate the genetic architecture of the cerebral ventricular morphology among 61,974 participants from 9 to 98 years old. We identified 62 genomic loci and 785 candidate genes associated with the cerebral ventricles, of which some exhibited changing effects during brain development and ageing. Enrichment analyses implicated biological pathways and neuropsychiatric disorders across developmental and neurodegenerative stages. Follow-up analyses demonstrated the broad genetic overlap of ventricular morphology with various neuropsychiatric traits. Moreover, we identified the volume of the inferior lateral ventricles as a consequence of prodromal AD and a robust endophenotype predicting AD risk through multiple analytic approaches.

Genetic studies have made considerable efforts to investigate the genetic architectures of cortical structures and subcortical nuclei[35–37], yet the ventricular system has been largely ignored. More than cavities containing protective fluid cushions and sinks for waste, the cerebral ventricles are critical to neocortical evolution, and their morphology is tightly linked to pathology in other parts of the central nervous system[6,38]. The common-variant-based heritability estimates of ventricular traits (18% to 40% for volumes and 3% to 9% for asymmetries) are comparable to or slightly higher than those of the cortex (14% to 37% for volumes and 2% to 9% for asymmetries)[36,39], strengthening the hypothesis that they share a common origin in evolution. Furthermore, through state-of-the-art analysis, we discovered that rare variants may account for 0.3% to 2.3% of the heritabilities for the ventricular traits. However, there is still a large gap between family-based heritabilities and SNP-based heritabilities. The still-missing heritability could be explained by different demographics of the two cohorts used for estimating heritabilities or by some rare and structural variants not well captured by the current sequencing technologies[40]. The latter could be addressed by advanced sequencing methods capable of providing deeper coverage of the human genome in the future[41].

With the understanding of age-related changes in the human brain[5], the importance of age-dependent genetic effects on neuroimaging traits has been highlighted[42]. Our study aggregated individual-level data from four independent cohorts, nearly covering the whole lifespan (age range, 9–82 years). It remains unclear why the percentage of females in UKB dropped in higher age classes even though life expectancy is higher for females, though it barely influenced our results. Consistent evidence from two complementary methods (longitudinal

https://doi.org/10.1038/s41562-023-01722-6

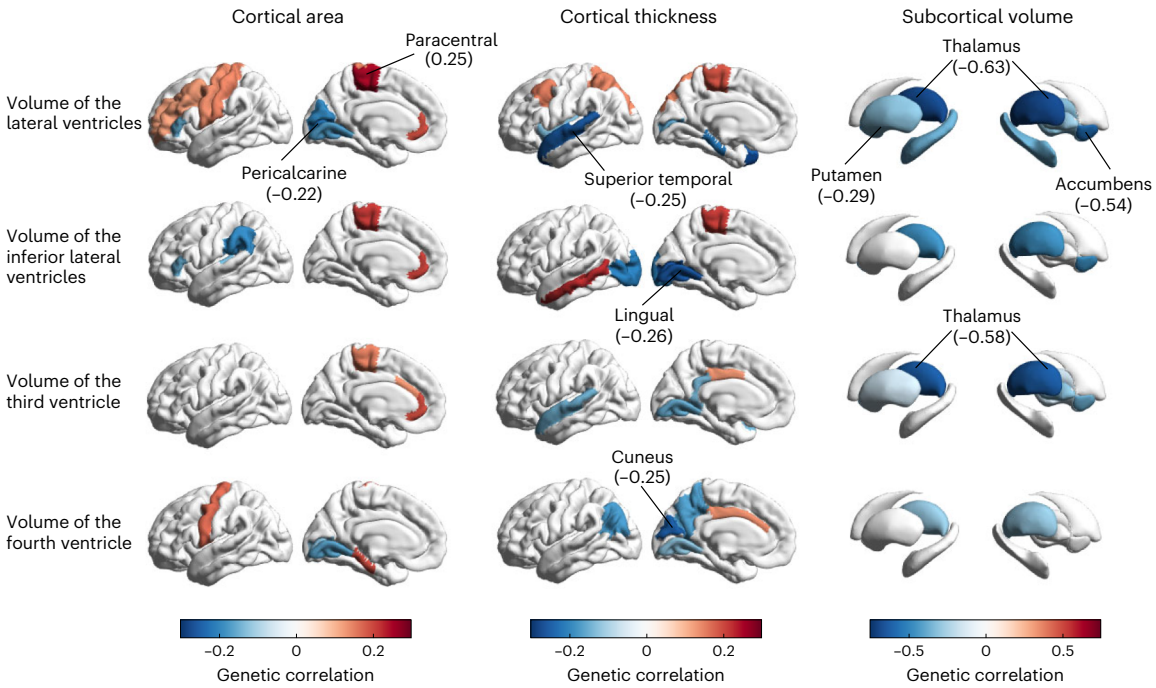


Fig. 5 | Genetic correlations between cerebral ventricles and brain subregions. The genetic correlations were estimated using LDSC. The outcomes include cortical thickness and surface area of 34 regions and volumes of six subcortical structures. All regions that are genetically correlated with ventricular traits at nominal $P < 0.05$ are coloured on the basis of the magnitude of positive (red) or negative (blue) correlation estimates. Regions with genetic correlation estimates that survived false discovery rate corrections are annotated with text. The genetic correlations between asymmetries of the lateral and inferior lateral ventricles and brain subregions are not shown because no correlations are significant in the LDSC analyses.

analyses and meta-regression of cross-sectional analyses) showed that locus 7p22.3 (*AMZ1*) and locus 16q24.2 (*C16orf95*) mainly influence the enlargement of the cerebral ventricles in older adults. This result could help us better understand the genetics of brain ageing, and the epigenetic regulation of the two loci may serve as therapeutic targets for neurogenerative disorders in the future. Moreover, locus 3q28 (*GMNC*) is shown to mainly influence brain development, while the genetic effect attenuated but persisted in older adults. Interestingly, the C allele of the lead variant rs13066753 in 3q28 is associated with larger inferior ventricular volume, lower t-tau and p-tau levels in CSF[20,21], and a lower risk for AD[43], and our LDSC analysis also demonstrated a negative correlation between inferior ventricular volume and CSF p-tau. On the basis of our results in the lifespan analysis, we hypothesize that genetically determined large ventricular volumes driven by locus 3q28 in one's early life may lead to more CSF in adulthood. As a result, CSF tau protein levels would be relatively decreased[44], and subsequently, the AD risk would be reduced. Another probable explanation is the compensatory hypothesis. That is, those with higher genetic liability to AD might develop small ventricles in preclinical AD stages for the maintenance of normal cognition[45].

The volume of the lateral ventricles is the most widely studied ventricular phenotype in previous neuroimaging publications[5,42] and accounts for the largest part of the ventricular system. As we hypothesized that the inferior lateral ventricles are more closely linked to AD and might have different genetic architectures from other parts of the lateral ventricles, the volume of the lateral ventricles in our study did not include the volume of their inferior horns. Given our successful replication of all loci identified by CHARGE and the validation of 85% of our loci in the CHARGE cohort, we believe that the impact of the slightly different phenotype definitions between our study and CHARGE was relatively small. As expected, we observed that enlarged lateral ventricles were significantly correlated with thinner cortex and worse cognition. However, the statistical significance of the association between lateral ventricular volume and neuropsychiatric diseases is modest and thus cannot rule out false positives. The genetic associations of the third ventricular volume were similar to those of the lateral ventricular volume. The volume of the fourth ventricle exhibited the highest heritability (40% in SNP-based analysis and 85% in twin analysis) among all ventricular traits, and our results indicate that the fourth ventricle was less likely to be involved in the biology of ageing or neuropsychiatric diseases.

One of our key findings is the reverse causal relationship between enlarged inferior lateral ventricular volume and AD. This relationship was revealed by MR and PRS and was validated in time-to-event and family history analyses. As the highest-contributing feature in the multivariate Cox model involving the volume of the hippocampus as a covariate[46], the inferior lateral ventricular volume might be more sensitive than traditional neuroimaging metrics to predict the risk of developing AD[47]. The predictive value of the inferior lateral ventricular volume has been noticed by some clinical studies with shorter follow-up periods[48,49], but we propose it as an endophenotype for AD since we have thoroughly examined the eligibility criteria of an endophenotype: (1) it is heritable, (2) it associates with illness in the population, (3) it manifests in individuals independent of the illness state and (4) it co-segregates with the illness in families[50]. The post hoc investigation suggests that the dysfunction of oligodendrocytes, which form the myelin sheath around axons, might mediate inferior ventricular enlargement and AD[51]. Moreover, the volume of the inferior lateral ventricles can be easily visualized in non-invasive structural MRI images, demonstrating substantial feasibility and clinical utility.

Besides volumetric metrics, we also reported the genetic determinants of ventricular asymmetries, which have been suggested to be involved in mental disorders in observational studies[9,13,52–54] but had never been investigated in genetic studies. The asymmetry of the lateral ventricles tends to be stable with ageing, and the BrainSpan analyses suggested that the related genes were differentially expressed in early prenatal stages instead of subsequent stages. In gene set analysis, we observed tremendous enrichment of

https://doi.org/10.1038/s41562-023-01722-6

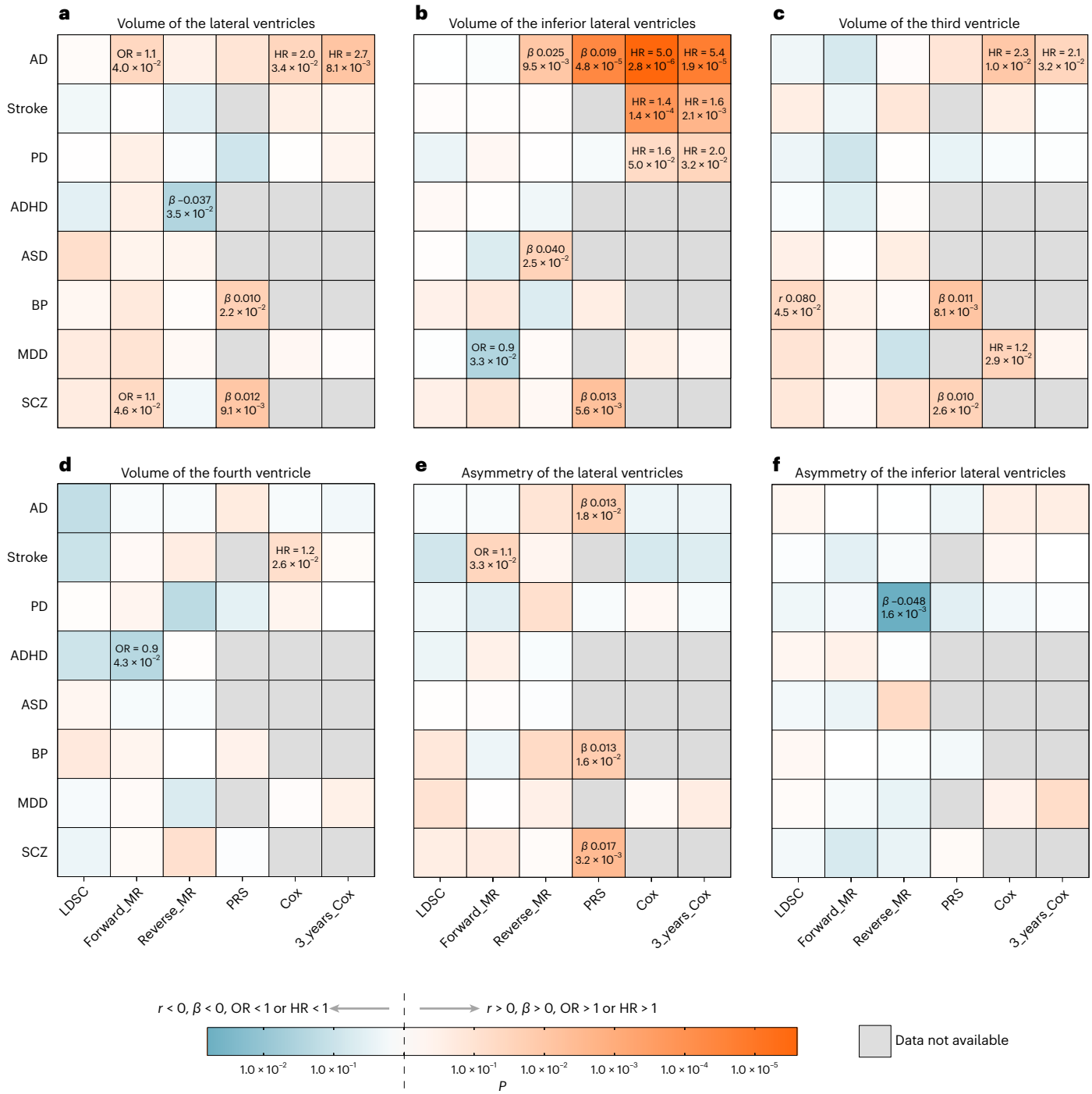


Fig. 6 | Causes and consequences of cerebral ventricular morphology and major neuropsychiatric disorders. a–**f**, Results of the LDSC, bidirectional MR, PRS, time-to-event and sensitivity analyses to investigate the associations of eight major neuropsychiatric disorders with the volume of the lateral ventricles (**a**), the volume of the inferior lateral ventricles (**b**), the volume of the third ventricle (**c**), the volume of the fourth ventricle (**d**), the asymmetry of the lateral ventricles (**e**) and the asymmetry of the inferior lateral ventricles (**f**). Red indicates a positive association (coefficient >0 or log-transformed coefficient >1), blue denotes a negative association (coefficient <0 or log-transformed coefficient <1) and grey represents unavailable data for analysis. Colour saturation reflects the $-\log_{10} P$ value of the related analysis. Associations with nominal $P < 0.05$ are annotated with their coefficients and raw P values. PD, Parkinson's disease.

lateral-ventricular-asymmetry-associated genes for a spectrum of neuropsychiatric disorders. However, the follow-up analyses suggested that the enrichment might be a product of the pleiotropic effect due to polygenicity, as correlation and causality analyses yielded few significant results. In addition, we observed increasing leftward asymmetry of the inferior lateral ventricles during ageing in older adults. The phenomenon might be attributed to normal ageing, as the asymmetry of the inferior lateral ventricles did not show a significant association with major neuropsychiatric diseases.

We acknowledge some limitations in our study. First, in the strict sense, our study did not include participants across the entire lifespan due to the lack of data from prenatal and early postnatal stages. Instead, we conducted BrainSpan analyses to investigate the spatiotemporal expression pattern of ventricular-trait-associated genes from early

prenatal stages to middle adulthood. Meanwhile, our sample size in middle adulthood was relatively small, which may lead to imprecise estimates. Second, we restricted our analysis to individuals of European ancestry to reduce population stratification; the generalization of results to other ethnicities thus needs to be validated in the future. Third, little age overlap was observed in the discovery and generalization cohorts. As a result, it is possible that the age-specific analyses could be influenced by other cohort-specific factors, even though we applied similar quality control and analytic procedures across these cohorts. Fourth, although we performed strict quality control before fitting the GWAS model, more advanced models (such as BOLT-LMM[55], SAIGE[56] and Regenie[57]) should be considered to better account for population structure and relatedness. Fifth, it remains unclear how well inferior ventricular volume would differentiate between AD and other dementias. Finally, as our imaging study excluded individuals with contraindications to MRI and those who resided far from imaging assessment centres, this selective approach may have compromised the representativeness of our sample, resulting in potential participation bias.

This study uncovered the genomic loci, genes and biological functions associated with ventricular morphology; examined the concordance and discordance among individuals of different age groups; and described the correlative and causal relationships between the cerebral ventricles and outcomes for brain health. We identified the volume of the inferior lateral ventricles as an endophenotype for predicting AD risk and a consequence of prodromal AD. This study helps elucidate the shared mechanisms between ventricular morphology and neuropsychiatric disorders, and the identified endophenotype could inform future clinical practice.

Methods

Samples across the lifespan

This genetic study used individual-level data from four independent cohorts with diverse age groups (Supplementary Table 1-1), including ABCD (mean age, 10 years; range, 9–11 years), IMAGEN (mean, 14; range, 13–20), HCP (mean, 29; range, 22–36), UKB (mean, 64; range, 45–82) and summary-level data from the CHARGE consortium (mean, 65; range, 20–98). All study sites were approved by their local research ethics committees or institutional review boards. All adult participants or participants' guardians for children or adolescents gave written informed consent before study enrolment.

For the discovery GWAS, we utilized brain MRI data from the imaging visit of UKB (N ~ 40,000). UKB is a large-scale prospective cohort to uncover the genetic and non-genetic determinants of health-related outcomes in middle and old age[58]. The UKB sample was restricted to participants with white British ancestry according to the sample quality control file 'ukb_sqc_v2.txt' to reduce bias due to population stratification and brain disorders. We excluded 1,146 participants with prevalent dementia, stroke, hydrocephalus or intracranial tumours on the basis of ICD-10 codes (Supplementary Table 40). The CHARGE consortium was a collaboration of prospective population-based cohorts designed to uncover the genetic basis of age-related traits in individuals of mainly European descent[59]. The CHARGE cohort has similar demographics to UKB, and its summary statistics were applied for replication analysis. The details of the CHARGE cohort regarding the ventricular GWAS (N = 23,533) were published previously[15].

For generalization in other age groups across the lifespan, we also used data from three cohorts whose age range did not overlap with UKB. The ABCD study is the largest longitudinal neuroimaging cohort (N = 11,876) tracking brain development from childhood through adolescence[60]. We extracted ABCD subjects of European descent through principal component analysis according to the 1000 G reference panel[61]. IMAGEN is a European multicentre genetic-neuroimaging project of healthy youth, which recruits more than 2,000 adolescents around age 14 and follows them longitudinally. The project was designed to be ethnically homogenous—that is, all participants were required to have four Western European grandparents[62]. HCP is a twin-family study aiming to examine brain connectivity and its variability in healthy young adults[63]. We downloaded data from the HCP 1200 Subject Release and excluded individuals with relatedness or not reported as non-Hispanic/Latino white ethnicity for genetic association analysis. In the post hoc analysis, the ADNI cohort was added to validate the association between inferior lateral ventricular volume and incident AD. The ADNI cohort was specifically designed to develop markers for early detection and tracking of AD[64]. Further information about the ADNI cohort can be found in the Supplementary Methods.

Imaging of the cerebral ventricles

The image acquisition and processing pipeline of the discovery dataset are available on the UKB website (https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/brain_mri.pdf). Briefly, the T1-weighted MPRAGE images were collected on a 3T Siemens Skyra scanner with $1\times1\times1$ mm^3 resolution, a $208\times256\times256$-pixel field of view, an acceleration factor of 2 and pre-scan normalization. Pre-processing quality control included aligning modalities and removing artefacts. The raw ventricular traits (the volumes of the left and right lateral ventricles, the left and right inferior lateral ventricles, the third ventricle, and the fourth ventricle) were extracted using FreeSurfer's aseg tool (v.6.0), followed by semi-automated quality control according to the image-processing pipeline developed and run on behalf of UKB, such as the detection of a bad field of view, head motion, noise, severe lesions and big ventricles[65,66]. Imaging details of the replication and generalization cohorts are provided in the Supplementary Methods and Supplementary Table 41.

We calculated six ventricular traits in each cohort: the volume of the lateral ventricles (the sum of both sides, including the left and right inferior lateral ventricles), the volume of the inferior lateral ventricles (the sum of both sides), the volume of the third ventricle, the volume of the fourth ventricle, the asymmetry index of the lateral ventricles and the asymmetry index of the inferior lateral ventricles. The asymmetry was measured by a widely used formula[39,67]:

$$\text{Asymmetry index} = \frac{\text{left} - \text{right}}{(\text{left} + \text{right})/2}$$

According to the definition, a positive asymmetry index indicates that the volume of the left side is greater than that of the right (that is, leftward asymmetry). Additionally, we excluded 996 individuals whose raw ventricular values were outliers greater than five times median absolute deviations (MAD). We normalized the six ventricular phenotypic values for GWASs using rank-based inverse normal transformation before conducting the analyses.

Genotyping and quality control

All genetic analyses were carried out with GRCh37/hg19 genomic coordinates. Genotype calling, imputation and quality control have been conducted centrally by the UKB team[68]. Briefly, participants in UKB were genotyped with two similar arrays, the UKB Axiom array and the UK BiLEVE Axiom Array. Phasing was achieved via SHAPEIT3 along with the 1000 Genomes Project[69], and imputation was conducted through IMPUTE4 with Haplotype Reference Consortium as the main reference panel. After imputation, poor-quality markers and samples were identified on the basis of tests such as batch effects, plate effects, heterozygosity and missing rates[68]. We applied additional quality control procedures to exclude markers with call rate <0.95, MAF < 0.01, Hardy–Weinberg $P < 10^{-6}$ and imputation quality score <0.5, and samples with missing rate >0.05, sex mismatches, abnormal sex chromosome aneuploidy, heterozygosity rate outliers, and those that were not in the white British ancestry subset or had more than ten putative third-degree relatives. The quality control was performed with PLINK v.2.0 (ref. 70), and all quality control procedures resulted

in 8.5 million autosomal variants and 31,880 individuals for the final discovery GWAS analysis. We performed rare-variant-based heritability analysis, gene-based rare variant association analysis and ExWAS using whole-exome sequencing data in UKB; their detailed quality control procedures are shown in the Supplementary Methods.

For the replication GWAS dataset, cohorts in the CHARGE consortium performed genotyping, quality control and imputation separately[15]. We imputed the genetic data of ABCD, IMAGEN and HCP with validated tools, including IMPUTE and Michigan Imputation Server. The quality control procedures in the three generalization datasets were similar to those in UKB (Supplementary Table 41). The final sample sizes for the cross-sectional generalization analyses were 4,445 for ABCD, 1,888 for IMAGEN and 228 for HCP.

Heritability estimation and genetic association analyses

We calculated the common- and rare-variant-based and family-based heritability estimates for all six ventricular traits[71]. Common-variant-based heritability analysis was performed with the genome-based restricted maximum likelihood approach as implemented in Genome-wide Complex Trait Analysis (GCTA) v.1.93.2, adjusting for age, age^2, sex, ICV, imaging centre and the first ten genetic principal components (PCs)[72,73]. Rare-variant-based heritability was calculated using the burden heritability regression method[74], with summary statistics from single variant ExWAS (see below) as input. Burden heritability explained by the rare variants was calculated from the regression slope using the aggregate mode via the R package bhr[75,76]. Family-based heritability was calculated in 105 monozygotic and 70 dizygotic twin pairs in the HCP project. Twin zygosity was determined by the genotyped data when possible; otherwise, the self-reported data were employed. The classic ACE model was fit using the twinlm() function in the mets R package[75,77].

Univariate GWAS for six ventricular traits was performed using linear regression under the additive model in PLINK v.2.0 (ref. 70). Rare-variant univariate ExWAS was conducted under the linear mixed model via SAIGE-GENE+ software[78]. Variants with minor allele count <10 were considered ultra-rare variants and were collapsed into the same group. MOSTest on matlabR2018b was used to conduct the multivariate GWAS, which implements permutation testing to identify genetic association across multiple traits and shows tremendous power when genetic effects are shared across traits[19]. For all association analyses, we regressed out the same covariates as the GCTA heritability analyses. To examine the potential impact of collider bias, we further dropped ICV from the covariates and reperformed the association analyses[79,80]. As the subsequent longitudinal analysis included a subset of the discovery GWAS sample, we validated the genetic associations after removing those participants. We also validated our initial results through the SAIGE software, whose full genetic relationship matrix was used for fitting the model in Step 1 to account for population structure and relatedness[56]. The significance threshold was $P < 5.00 \times 10^{-8}$ for the multivariate association analyses and Bonferroni-corrected ($P < 8.33 \times 10^{-9}$) for the univariate association analyses (see the Supplementary Methods for genomic risk loci identification details).

GWGAS was performed using MAGMA v.1.08 implemented in FUMA[24,81]. The GWGAS was conducted among 18,879 protein-coding genes. The significance threshold was $P < 0.05/18{,}879 = 2.65 \times 10^{-6}$ for the summary measure of ventricular morphology (multivariate-GWAS-based) and $P < 4.41 \times 10^{-7}$ for the six ventricular volumetric and asymmetric traits (univariate-GWAS-based and Bonferroni-corrected). The MHC region was excluded before the analysis. Positional (up to 10 kb apart), eQTL (brain data from psychENCODE[82], xQTLServer[83], ComminMind[84], BRAINEAC[85] and GTEx v.8 (ref. 86) and blood data from eQTLGen[87] and GTEx v.8 (ref. 86)) and 3D chromatin interaction (adult and fetal human brain data) mapping strategies were also utilized to assign candidate variants to genes associated with ventricles. The gene-based rare-variant association analysis for ventricular traits was calculated using Burden test, SKAT and SKAT-O through SAIGE-GENE+[78], adjusting for age, age^2, sex, ICV, imaging centre and the first ten genetic PCs. SKAT-O was chosen as the primary approach as it combines the Burden test and SKAT and considers various degrees of effect heterogeneity; thus, it may better capture the associations between rare variants and traits[88]. The significance threshold was set to $P < 4.45 \times 10^{-7}$ (that is, 0.05/18,731 genes and six ventricular traits).

The methods for functional follow-up of genetic associations, including phenome-wide association scan, gene-set enrichment, cell type and partitioned heritability analyses, are available in the Supplementary Methods.

Replication and generalization

Data from the CHARGE consortium were used for replication as its demographics were similar to those in the discovery dataset. For independent significant variants associated with the lateral ventricular volume identified in the discovery GWAS, we extracted their genetic associations in the CHARGE summary data. If any variant was not identified in CHARGE, the genetic association from its closest proxy (LD $r^2 > 0.8$) was used. A variant was defined as replicated when it was Bonferroni-corrected significant (that is, $P < 0.05$ divided by the number of independent significant SNPs) in CHARGE, and its direction of effect in CHARGE was consistent with that in the discovery dataset. The CHARGE consortium did not provide data for the other cerebral ventricles or asymmetric metrics, so replication analyses in the other ventricular traits were not conducted.

To explore the role of ventricular-trait-associated loci in brain development and degeneration, we conducted a generalization analysis of the independent significant signals of six ventricular traits in children (ABCD, $N = 4{,}445$), adolescents (IMAGEN, $N = 1{,}888$) and young adults (HCP, $N = 228$) using linear regression under the additive model in PLINK v.2.0. We calculated the proportion of nominal and Bonferroni-corrected validated signals ($P < 0.05$ with the same direction of effects) in each cohort and investigated whether there were concordant results among all cohorts, which indicates robustness and stability among all age groups. The discordance of genetic effects among samples across the lifespan was quantified by fixed-effects meta-regression, with the median age of each subgroup as the independent variable and the genetic effect as the dependent variable. To better explore the associations between age and genetic effects in older adults, we split the UKB sample ($N = 31{,}880$) every five years and reran the genetic association analysis within each group. These procedures generated ten subgroups with median ages of 10, 14, 29, 49, 53, 58, 63, 68, 72 and 77 years (Supplementary Table 1-1). After allele alignment, IVW fixed-effects meta-regression of independent signals was performed using linear and quadratic models with the R package metafor[75,89]. We performed sensitivity analysis after 1:1 matching of males and females in each UKB subgroup because we observed that the percentage of females dropped in higher age classes in UKB.

We hypothesized that genetic loci associated with cross-sectional ventricular traits might also influence the longitudinal change of cerebral ventricles. To test this hypothesis, we leveraged longitudinal imaging data from children and adolescents (ABCD, with a sample for longitudinal analysis of 3,409; and IMAGEN, with a sample for analysis of 1,138) and older adults (UKB, with a sample for analysis of 3,646; Supplementary Table 1-2). For longitudinal neuroimaging data, participants in IMAGEN and UKB were followed up only once, while ABCD participants were followed up two and/or four years after recruitment. If neuroimaging data from more than one follow-up were available, we chose the data with the longest follow-up period for longitudinal analysis. For the independent significant variants identified in the discovery GWAS, their genetic associations with annual change rates of the same ventricular traits were calculated using linear regression under the addictive model in PLINK v.2.0, including age, age^2, sex, imaging site and the first ten PCs as covariates. ICV was not corrected in the longitudinal analysis because associations

between ICV and brain changes over time have been shown to be small[42]. The longitudinal genetic analyses were performed in each cohort separately, and then the genetic effects in the two younger cohorts were meta-analysed using METAL to enhance statistical power[90]. Variants that passed the Bonferroni threshold ($P = 0.05/152$ variants) were considered significant.

Moreover, we acknowledge that we did not include prenatal samples or samples in their early postnatal life due to data limitations. To compensate for this gap, we conducted BrainSpan analyses using data spanning from early prenatal (8 weeks post-conception) to middle adulthood (40 years) via FUMA[24]. The significant enrichment for developmental ages was defined as a Bonferroni-corrected $P < 0.05$ in upregulated, downregulated or both-sided differentially expressed gene sets.

Genetic correlation

Genetic correlation analyses were performed between ventricular traits and outcomes for brain health using cross-trait LDSC, which was robust to sample overlap between trait pairs[30]. The analyses utilized summary statistics from univariate ventricular GWAS, GWAS of outcomes for brain health[32] and the precomputed LD scores from the European population of the 1000 Genomes Project as a reference panel. The outcomes for brain health included clinical diagnoses (AD[43], stroke[91], PD[92], ADHD[93], ASD[94], MDD[95], BP[96] and SCZ[97]), competencies and behaviours (general cognitive function[98], verbal numerical reasoning[98], reaction time[98], drinking[99], smoking[99], general risk tolerance[100] and handedness[101]), and metrics of brain structure or function (total cortical area and mean cortical thickness[102], white matter hyperintensities[103], CSF Aβ42, and p-tau levels[49]). As recommended, the LDSC analyses were restricted to Hapmap3 SNPs, and the MHC region was excluded before analyses. To uncover the genetic association between ventricular traits and brain structures of specific regions, we further conducted brain-wide LDSC analyses. The summary statistics of 34 cortical area traits, 34 cortical thickness traits and 6 subcortical volumetric traits were downloaded from the ENIGMA consortium[102,104].

Bidirectional MR

Bidirectional two-sample MR was employed to infer the causal relationship between six ventricular traits and eight major neuropsychiatric diseases. MR analyses were not performed for competencies/behaviours and metrics since these GWAS samples overlapped with our ventricular GWAS. MR only utilized variants strongly associated with exposure to reduce the weak instrument bias[105]. We selected instrument variables using genetic variants associated with the exposure on the basis of $P < 5 \times 10^{-6}$ due to the limited number of variants that passed the genome-wide significance threshold[106,107]. LD clumping was performed by excluding variants that have an $r^2 > 0.001$ with another variant according to the European 1000 G reference panel. We employed four MR analytical methods under different causal scenarios, including IVW, MR-Egger, weighted median and weighted mode, where IVW was the primary approach to present results[31]. The heterogeneity and pleiotropy were also checked in case of violating fundamental MR assumptions[108]. All MR analyses were carried out using the TwoSampleMR package in R[75,109].

PRS analyses

To evaluate the association between genetic liability to neuropsychiatric disorders and ventricular traits, we leveraged the standard PRS for AD, PD, BP and SCZ provided by UKB in the PRS Release (Category 301)[110]. Analyses for stroke, ADHD, ASD and MDD were not performed due to the unavailability of data. A detailed description of PRS in UKB can been found in the Supplementary Methods. We restricted our regression analyses to white British ancestry and excluded individuals with prevalent dementia, stroke, hydrocephalus and intracranial tumours. After we merged covariates and excluded outliers of ventricular traits exceeding 5 MAD, the final sample size for regression analysis was 31,843. Multiple linear regression analyses were performed to investigate the association between inverse-normal-transformed ventricular traits and PRS for brain diseases with the adjustment of age, age^2, sex, imaging site, ICV and the first ten genetic PCs. The analyses were achieved with the lm function in R.

Time-to-event validation

To validate the predictive capacity of ventricular traits on risks of neuropsychiatric disorders, we conducted Cox proportional hazards analyses with clinical data in the UKB cohort. Neuropsychiatric disorders for analysis were restricted to those with at least ten incident cases after all quality control procedures, including AD, stroke, PD and MDD. Dementia was also chosen as an outcome for analysis to support the findings in AD due to its larger number of incident cases. Incident cases with neuropsychiatric disorders were ascertained from the first occurrences data with ICD-10 codes (Supplementary Table 40). Participants who died, withdrew from the study before the end of follow-up or had not developed any neuropsychiatric disorder by the end of follow-up (12 November 2021) were censored. Prior to analyses, we excluded individuals with prevalent dementia, stroke, hydrocephalus, intracranial tumours or incomplete sets of covariates, leaving 35,168 participants. For each neuropsychiatric disorder, we subsequently excluded prevalent cases, individuals without follow-up data from the first imaging visit and outliers of ventricular traits exceeding 5 MAD. The final sample for analysis ranged from 13,089 to 15,200, and incident cases ranged from 15 to 225. After normalization of ventricular traits using the inverse normal transformation method, Cox proportional hazard regression models were fitted to estimate HRs and 95% CIs for the ventricular traits on risks of incident neuropsychiatric disorders. Adjusted covariates included age, age^2, sex, imaging site and ICV. For AD and dementia, additional covariates included education and the number of APOE ε4 alleles.

To test the robustness of the primary findings, we conducted sensitivity analyses with the same covariates and excluded individuals with follow-up times shorter than three years. The mean follow-up time increased from 2.44 years to 4.42 years. In sensitivity analysis, the sample size ranged from 4,694 to 5,562, and the number of incident cases ranged from 12 to 93. For the association between the inferior lateral ventricular volume and AD/dementia, we performed additional analysis by adding the volume of the hippocampus as a covariate to investigate the potential confounding effect driven by the hippocampus being near the inferior lateral ventricles.

Post hoc analyses

As we observed significant associations between the volume of the inferior lateral ventricles and incident AD risk, we sought to validate this association in an external cohort[111]. A total of 1,633 ADNI subjects free of prevalent dementia passed the ventricular imaging quality control procedures, 343 of whom developed AD during a mean follow-up time of 4.5 years. Cox proportional hazard regression analysis was conducted after adjusting age, age^2, sex, education, APOE ε4, baseline cognitive diagnosis and ICV in the base model. Moreover, we reperformed Cox analysis after additionally adjusting for the volume of the hippocampus to examine whether the association was independent of the adjacent hippocampus. We also conducted sensitivity analyses excluding individuals with follow-up times shorter than three years. The sensitivity analysis included 1,035 subjects free of dementia at baseline, 258 of whom developed incident AD.

To investigate whether larger inferior lateral ventricular volumes and ADD tend to be inherited together within families, we applied the logistic regression model to explore the association between the inferior lateral ventricular volume and parental history of ADD among UKB participants free of prevalent dementia, stroke, hydrocephalus and intracranial tumours. We excluded individuals who lacked data regarding ADD parental history or any covariate, or whose parents died

before age 65, and we excluded outliers of ventricular traits exceeding 5 MAD. A total of 30,723 individuals were included in the final analysis, of which 8,044 had a parental history of ADD. After normalization of the ventricular traits, multivariate logistic regression was performed with ADD parental history as the dependent variable and the volume of the inferior lateral ventricles as the independent variable. Adjusted factors included age, age^2, sex, education, APOE ε4, imaging site and ICV. Parental age at death was additionally adjusted in exploratory analysis. The results are presented as ORs with 95% CIs.

Further post hoc analyses were conducted to identify potential shared biological mechanisms between the inferior lateral ventricular volume and AD. We searched the Agora Resources via the AD knowledge portal to identify genetic, genomic and literature evidence of all genes associated with inferior lateral ventricular volumes[33]. We also used the Human Protein Atlas to investigate the tissue and cell type specificity of the critical genes[34].

Ethical approval

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. UKB was approved by the National Health Service National Research Ethics (ref. no. 11/NW/0382). The ABCD study was approved by the central Institutional Review Board at the University of California, San Diego. 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. The HCP was reviewed and approved by the Institutional Ethics Committee of Washington University in St. Louis, Missouri.

Reporting summary

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

Data availability

Our GWAS summary statistics for cerebral ventricles can be found at https://doi.org/10.6084/m9.figshare.21529491. The GWAS results are also available on the FUMA website (https://fuma.ctglab.nl/browse/; ID 619–625). The individual-level data used in the present study were obtained from UKB (https://www.ukbiobank.ac.uk/), ABCD (https://abcdstudy.org/), IMAGEN (http://imagen-project.org/), HCP (http://www.humanconnectome.org/) and ADNI (https://adni.loni.usc.edu/). The summary-level data for the CHARGE consortium were obtained from dbGaP (phs000930.v9.p1). The GWAS Catalog resource can be found at https://www.ebi.ac.uk/gwas/. The GWAS atlas can be found at https://atlas.ctglab.nl/PheWAS/. The DSigDB database can be found at http://dsigdb.tanlab.org/DSigDBv1.0/. Agora can be found at https://agora.adknowledgeportal.org/. The Human Protein Atlas can be found at https://www.proteinatlas.org/.

Code availability

This study used openly available software and code, specifically R (https://www.r-project.org/), PLINK (https://www.cog-genomics.org/plink/), GCTA (http://cnsgenomics.com/software/gcta/), IMPUTE (https://mathgen.stats.ox.ac.uk/impute/impute_v2.html), Michigan Imputation Server (https://imputationserver.sph.umich.edu/), MOSTest (https://github.com/precimed/mostest), METAL (http://csg.sph.umich.edu/abecasis/metal/), FUMA (https://fuma.ctglab.nl/), MAGMA (https://ctg.cncr.nl/software/magma/, also implemented in FUMA), SAIGE-GENE+ (https://saigegit.github.io/SAIGE-doc/), Enrichr (https://maayanlab.cloud/Enrichr/) and LDSC (https://github.com/bulik/ldsc/). Custom scripts for the analyses in this paper are available through GitHub (https://github.com/yjge/cerebral_ventricles).

References

1. Duy, P. Q. et al. Brain ventricles as windows into brain development and disease. *Neuron* **110**, 12–15 (2022).
2. de Mélo Silva Júnior, M. L., Diniz, P. R. B., de Souza Vilanova, M. V., Basto, G. P. T. & Valença, M. M. Brain ventricles, CSF and cognition: a narrative review. *Psychogeriatrics* **22**, 544–552 (2022).
3. Sapkota, S., McFall, G. P., Masellis, M., Dixon, R. A. & Black, S. E. Differential cognitive decline in Alzheimer's disease is predicted by changes in ventricular size but moderated by apolipoprotein E and pulse pressure. *J. Alzheimers Dis.* **85**, 545–560 (2022).
4. West, N. A. et al. Neuroimaging findings in midlife and risk of late-life dementia over 20 years of follow-up. *Neurology* **92**, e917–e923 (2019).
5. Bethlehem, R. A. I. et al. Brain charts for the human lifespan. *Nature* **604**, 525–533 (2022).
6. Lui, J. H., Hansen, D. V. & Kriegstein, A. R. Development and evolution of the human neocortex. *Cell* **146**, 18–36 (2011).
7. Chojnacki, A. K., Mak, G. K. & Weiss, S. Identity crisis for adult periventricular neural stem cells: subventricular zone astrocytes, ependymal cells or both? *Nat. Rev. Neurosci.* **10**, 153–163 (2009).
8. Duy, P. Q. et al. Impaired neurogenesis alters brain biomechanics in a neuroprogenitor-based genetic subtype of congenital hydrocephalus. *Nat. Neurosci.* **25**, 458–473 (2022).
9. Richards, R. et al. Increased hippocampal shape asymmetry and volumetric ventricular asymmetry in autism spectrum disorder. *NeuroImage Clin.* **26**, 102207 (2020).
10. Prigge, M. B. D. et al. A 16-year study of longitudinal volumetric brain development in males with autism. *NeuroImage* **236**, 118067 (2021).
11. McWhinney, S. R. et al. Association between body mass index and subcortical brain volumes in bipolar disorders—ENIGMA study in 2735 individuals. *Mol. Psychiatry* **26**, 6806–6819 (2021).
12. Schmaal, L. et al. Subcortical brain alterations in major depressive disorder: findings from the ENIGMA Major Depressive Disorder working group. *Mol. Psychiatry* **21**, 806–812 (2016).
13. Okada, N. et al. Abnormal asymmetries in subcortical brain volume in schizophrenia. *Mol. Psychiatry* **21**, 1460–1466 (2016).
14. Brugger, S. P. & Howes, O. D. Heterogeneity and homogeneity of regional brain structure in schizophrenia: a meta-analysis. *JAMA Psychiatry* **74**, 1104–1111 (2017).
15. Vojinovic, D. et al. Genome-wide association study of 23,500 individuals identifies 7 loci associated with brain ventricular volume. *Nat. Commun.* **9**, 3945 (2018).
16. Smith, S. M. et al. An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nat. Neurosci.* **24**, 737–745 (2021).
17. Scelsi, C. L. et al. The lateral ventricles: a detailed review of anatomy, development, and anatomic variations. *AJNR Am. J. Neuroradiol.* **41**, 566–572 (2020).
18. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
19. van der Meer, D. et al. Understanding the genetic determinants of the brain with MOSTest. *Nat. Commun.* **11**, 3512 (2020).
20. Deming, Y. et al. Genome-wide association study identifies four novel loci associated with Alzheimer's endophenotypes and disease modifiers. *Acta Neuropathol.* **133**, 839–856 (2017).
21. Jansen, I. E. et al. Genome-wide meta-analysis for Alzheimer's disease cerebrospinal fluid biomarkers. *Acta Neuropathol.* **144**, 821–842 (2022).
22. Sha, Z., Schijven, D., Fisher, S. E. & Francks, C. Genetic architecture of the white matter connectome of the human brain. *Sci. Adv.* **9**, eadd2870 (2023).
23. Bahrami, S. et al. Distributed genetic architecture across the hippocampal formation implies common neuropathology across brain disorders. *Nat. Commun.* **13**, 3436 (2022).

24. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
25. Yoo, M. et al. DSigDB: drug signatures database for gene set analysis. *Bioinformatics* **31**, 3069–3071 (2015).
26. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
27. Chen, Y. et al. Structural basis of ALDH1A2 inhibition by irreversible and reversible small molecule inhibitors. *ACS Chem. Biol.* **13**, 582–590 (2018).
28. Piergiovanni, G. & Costanzo, V. GEMC1 is a novel TopBP1-interacting protein involved in chromosomal DNA replication. *Cell Cycle* **9**, 3662–3666 (2010).
29. Watanabe, K. et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **51**, 1339–1348 (2019).
30. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
31. Burgess, S. et al. Guidelines for performing Mendelian randomization investigations [version 2; peer review: 2 approved]. *Wellcome Open Research* https://doi.org/10.12688/wellcomeopenres.15555.2 (2020).
32. Gorelick, P. B. et al. Defining optimal brain health in adults: a presidential advisory from the American Heart Association/American Stroke Association. *Stroke* **48**, e284–e303 (2017).
33. Greenwood, A. K. et al. The AD Knowledge Portal: a repository for multi-omic data on Alzheimer's disease and aging. *Curr. Protoc. Hum. Genet.* **108**, e105 (2020).
34. Uhlén, M. et al. Proteomics: tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
35. van der Meer, D. et al. The genetic architecture of human cortical folding. *Sci. Adv.* **7**, eabj9446 (2021).
36. Makowski, C. et al. Discovery of genomic loci of the human cerebral cortex using genetically informed brain atlases. *Science* **375**, 522–528 (2022).
37. Satizabal, C. L. et al. Genetic architecture of subcortical brain structures in 38,851 individuals. *Nat. Genet.* **51**, 1624–1636 (2019).
38. Fame, R. M. & Lehtinen, M. K. Emergence and developmental roles of the cerebrospinal fluid system. *Dev. Cell* **52**, 261–275 (2020).
39. Sha, Z. et al. The genetic architecture of structural left–right asymmetry of the human brain. *Nat. Hum. Behav.* **5**, 1226–1239 (2021).
40. Wainschtein, P. et al. Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. *Nat. Genet.* **54**, 263–273 (2022).
41. Girirajan, S. Missing heritability and where to find it. *Genome Biol.* **18**, 89 (2017).
42. Brouwer, R. M. et al. Genetic variants associated with longitudinal changes in brain structure across the lifespan. *Nat. Neurosci.* **25**, 421–432 (2022).
43. 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).
44. Hansson, O. et al. The genetic regulation of protein expression in cerebrospinal fluid. *EMBO Mol. Med.* **15**, e16359 (2023).
45. Zhang, X. et al. Bridging Integrator 1 (BIN1) genotype effects on working memory, hippocampal volume, and functional connectivity in young healthy individuals. *Neuropsychopharmacology* **40**, 1794–1803 (2015).
46. Genon, S., Eickhoff, S. B. & Kharabian, S. Linking interindividual variability in brain structure to behaviour. *Nat. Rev. Neurosci.* **23**, 307–318 (2022).
47. Jack, C. R. Jr. et al. NIA-AA Research Framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement.* **14**, 535–562 (2018).
48. Macdonald, K. E., Bartlett, J. W., Leung, K. K., Ourselin, S. & Barnes, J. The value of hippocampal and temporal horn volumes and rates of change in predicting future conversion to AD. *Alzheimer Dis. Assoc. Disord.* **27**, 168–173 (2013).
49. Coupé, P. et al. Hippocampal-amygdalo-ventricular atrophy score: Alzheimer disease detection using normative and pathological lifespan models. *Hum. Brain Mapp.* **43**, 3270–3282 (2022).
50. Lee Gregory, M., Burton, V. J. & Shapiro, B. K. in *Neurobiology of Brain Disorders* (eds Zigmond, M. J. et al.) 18–41 (Academic Press, 2015).
51. Coleman, J. Young brain fluid improves memory in old mice. *Nature* https://doi.org/10.1038/d41586-022-01282-1 (2022).
52. Sasabayashi, D. et al. Subcortical brain volume abnormalities in individuals with an at-risk mental state. *Schizophr. Bull.* **46**, 834–845 (2020).
53. Lewis, M. M. et al. Asymmetrical lateral ventricular enlargement in Parkinson's disease. *Eur. J. Neurol.* **16**, 475–481 (2009).
54. Kuo, F. & Massoud, T. F. Structural asymmetries in normal brain anatomy: a brief overview. *Ann. Anat.* **241**, 151894 (2022).
55. Loh, P. R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
56. Zhou, W. et al. Efficiently controlling for case–control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).
57. Mbatchou, J. et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* **53**, 1097–1103 (2021).
58. Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
59. Psaty, B. M. et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ. Cardiovasc. Genet.* **2**, 73–80 (2009).
60. Casey, B. J. et al. The Adolescent Brain Cognitive Development (ABCD) study: imaging acquisition across 21 sites. *Dev. Cogn. Neurosci.* **32**, 43–54 (2018).
61. Lee, P. H. et al. Genetic association of attention-deficit/hyperactivity disorder and major depression with suicidal ideation and attempts in children: the Adolescent Brain Cognitive Development Study. *Biol. Psychiatry* **92**, 236–245 (2022).
62. Schumann, G. et al. The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Mol. Psychiatry* **15**, 1128–1139 (2010).
63. Van Essen, D. C. et al. The Human Connectome Project: a data acquisition perspective. *NeuroImage* **62**, 2222–2231 (2012).
64. Hendrix, J. A. et al. The Worldwide Alzheimer's Disease Neuroimaging Initiative: an update. *Alzheimers Dement.* **11**, 850–859 (2015).
65. Alfaro-Almagro, F. et al. Image processing and quality control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage* **166**, 400–424 (2018).
66. Miller, K. L. et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* **19**, 1523–1536 (2016).
67. Kong, X. Z. et al. Mapping cortical brain asymmetry in 17,141 healthy individuals worldwide via the ENIGMA Consortium. *Proc. Natl Acad. Sci. USA* **115**, E5154–e5163 (2018).
68. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
69. O'Connell, J. et al. Haplotype estimation for biobank-scale data sets. *Nat. Genet.* **48**, 817–820 (2016).

70. Chang et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, s13742-015-0047-8 (2015).
71. Yang, J., Zeng, J., Goddard, M. E., Wray, N. R. & Visscher, P. M. Concepts, estimation and interpretation of SNP-based heritability. *Nat. Genet.* **49**, 1304–1310 (2017).
72. Yang, J., Hong Lee, S., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
73. Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
74. Weiner, D. J. et al. Polygenic architecture of rare coding variation across 394,783 exomes. *Nature* **614**, 492–499 (2023).
75. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2020).
76. Weiner, D. J. et al. Polygenic architecture of rare coding variation across 394783 exomes. *Nature* **614**, 492–499 (2023).
77. Holst, K. K., Scheike, T. H. & Hjelmborg, J. B. The liability threshold model for censored twin data. *Comput. Stat. Data Anal.* **93**, 324–335 (2016).
78. Zhou, W. et al. SAIGE-GENE+ improves the efficiency and accuracy of set-based rare variant association tests. *Nat. Genet.* **54**, 1466–1469 (2022).
79. Aschard, H., Vilhjálmsson, B. J., Joshi, A. D., Price, A. L. & Kraft, P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *Am. J. Hum. Genet* **96**, 329–339 (2015).
80. Fürtjes, A. E. et al. General dimensions of human brain morphometry inferred from genome-wide association data. *Hum. Brain Mapp.* **44**, 3311–3323 (2023).
81. 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).
82. Wang, D. et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, eaat8464 (2018).
83. Ng, B. et al. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat. Neurosci.* **20**, 1418–1426 (2017).
84. Fromer, M. et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).
85. 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).
86. GTEx Consortium The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
87. 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).
88. Lee, S., Wu, M. C. & Lin, X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* **13**, 762–775 (2012).
89. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* **36**, 1–48 (2010).
90. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
91. 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).
92. Kurki, M. I. et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
93. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
94. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
95. 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).
96. Stahl, E. A. et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).
97. 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).
98. Davies, G. et al. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat. Commun.* **9**, 2098 (2018).
99. 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).
100. Karlsson Linnér, R. et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat. Genet.* **51**, 245–257 (2019).
101. de Kovel, C. G. F. & Francks, C. The molecular genetics of hand preference revisited. *Sci. Rep.* **9**, 5986 (2019).
102. Grasby, K. L. et al. The genetic architecture of the human cerebral cortex. *Science* **367**, eaay6690 (2020).
103. Traylor, M. et al. Genetic variation in PLEKHG1 is associated with white matter hyperintensities (*n*=11,226). *Neurology* **92**, e749–e757 (2019).
104. Hibar, D. P. et al. Common genetic variants influence human subcortical brain structures. *Nature* **520**, 224–229 (2015).
105. Burgess, S., Thompson, S. G. & CRP CHD Genetics Collaboration Avoiding bias from weak instruments in Mendelian randomization studies. *Int J. Epidemiol.* **40**, 755–764 (2011).
106. Andrews, S. J., Fulton-Howard, B., O'Reilly, P., Marcora, E. & Goate, A. M. Causal associations between modifiable risk factors and the Alzheimer's phenome. *Ann. Neurol.* **89**, 54–65 (2021).
107. Mavromatis, L. A. et al. Association between brain structure and alcohol use behaviors in adults: a Mendelian randomization and multiomics study. *JAMA Psychiatry* **79**, 869–878 (2022).
108. Davies, N. M., Holmes, M. V. & Davey Smith, G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *Br. Med. J.* **362**, k601 (2018).
109. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, e34408 (2018).
110. Thompson, D. J. et al. UK Biobank release and systematic evaluation of optimised polygenic risk scores for 53 diseases and quantitative traits. Preprint at *medRxiv* https://doi.org/10.1101/2022.06.16.22276246 (2022).
111. Petersen, R. C. et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology* **74**, 201–209 (2010).

Acknowledgements

We thank all participants and cooperating institutions. The UKB analyses were conducted using the UKB Resource under application no. 19542. This study was supported by grants from the Science and Technology Innovation 2030 Major Projects (grant no. 2022ZD0211600 to J.-T.Y.), the National Natural Science Foundation of China (grant nos 92249305 and 82071201 to J.-T.Y., 82071997 to W.C. and 81971032 to L.T.), the Shanghai Municipal Science and Technology Major Project (grant no. 2018SHZDZX01 to J.-F.F.), the Research Start-up Fund of Huashan Hospital (grant no. 2022QD002 to J.-T.Y.), the Shanghai Rising-Star Program (grant no. 21QA1408700 to W.C.), the 111 Project (grant no. B18015 to J.-F.F.), and the ZHANGJIANG LAB,

the Tianqiao and Chrissy Chen Institute, and the State Key Laboratory of Neurobiology and Frontiers Center for Brain Science of Ministry of Education, Fudan University. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

J.-T.Y. designed the study. Y.-J.G. and B.-S.W. organized the data, carried out the statistical analysis and participated in writing the first draft of the manuscript. Y.-J.G., B.-S.W. and Y.Z. designed and drew the figures. S.-D.C., J.-J.K., Y.-T.D., X.-Y.H., Y.-L.Z. and Q.M. organized and analysed the data. Y.Z., S.-D.C., Y.-R.Z., Y.-T.D., Y.-N.O., X.-Y.H., K.K., H.L., T.P., J.-F.F., Q.D., L.T., G.S., W.C. and J.-T.Y. critically revised the manuscript. IMAGEN provided data used for this study. All authors read and approved the final manuscript.

Competing interests

T.B. served in an advisory or consultancy role for eye level, Infectopharm, Lundbeck, Medice, Neurim Pharmaceuticals, Oberberg GmbH, Roche and Takeda. He received conference support or a speaker's fee from Janssen, Medice and Takeda. He received royalties from Hogrefe, Kohlhammer, CIP Medien and Oxford University Press; the present 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 a speaker's fee from Shire. She received royalties from Hogrefe, Kohlhammer and Schattauer. The present work is unrelated to the above grants and relationships. The other authors report no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41562-023-01722-6.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-023-01722-6.

Correspondence and requests for materials should be addressed to Wei Cheng or Jin-Tai Yu.

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

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

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

Yi-Jun Ge[1,32], Bang-Sheng Wu[1,32], Yi Zhang[1,32], Shi-Dong Chen[1], Ya-Ru Zhang[1], Ju-Jiao Kang[2], Yue-Ting Deng[1], Ya-Nan Ou[3], Xiao-Yu He[1], Yong-Li Zhao[3], Kevin Kuo[1], Qing Ma[2], Tobias Banaschewski[4], Gareth J. Barker[5], Arun L. W. Bokde[6], Sylvane Desrivières[7], Herta Flor[8,9], Antoine Grigis[10], Hugh Garavan[11], Penny Gowland[12], Andreas Heinz[13], Rüdiger Brühl[14], Jean-Luc Martinot[15], Marie-Laure Paillère Martinot[15,16], Eric Artiges[15,17], Frauke Nees[4,8,18], Dimitri Papadopoulos Orfanos[10], Herve Lemaitre[10,19], Tomáš Paus[20,21], Luise Poustka[22], Sarah Hohmann[4], Sabina Millenet[4], Juliane H. Fröhner[23], Michael N. Smolka[23], Nilakshi Vaidya[24], Henrik Walter[13], Robert Whelan[25], IMAGEN Consortium*, Jian-Feng Feng[2,26,27,28,29], Lan Tan[3], Qiang Dong[1], Gunter Schumann[24,30], Wei Cheng[1,2,26,27,31] ✉ & Jin-Tai Yu[1] ✉

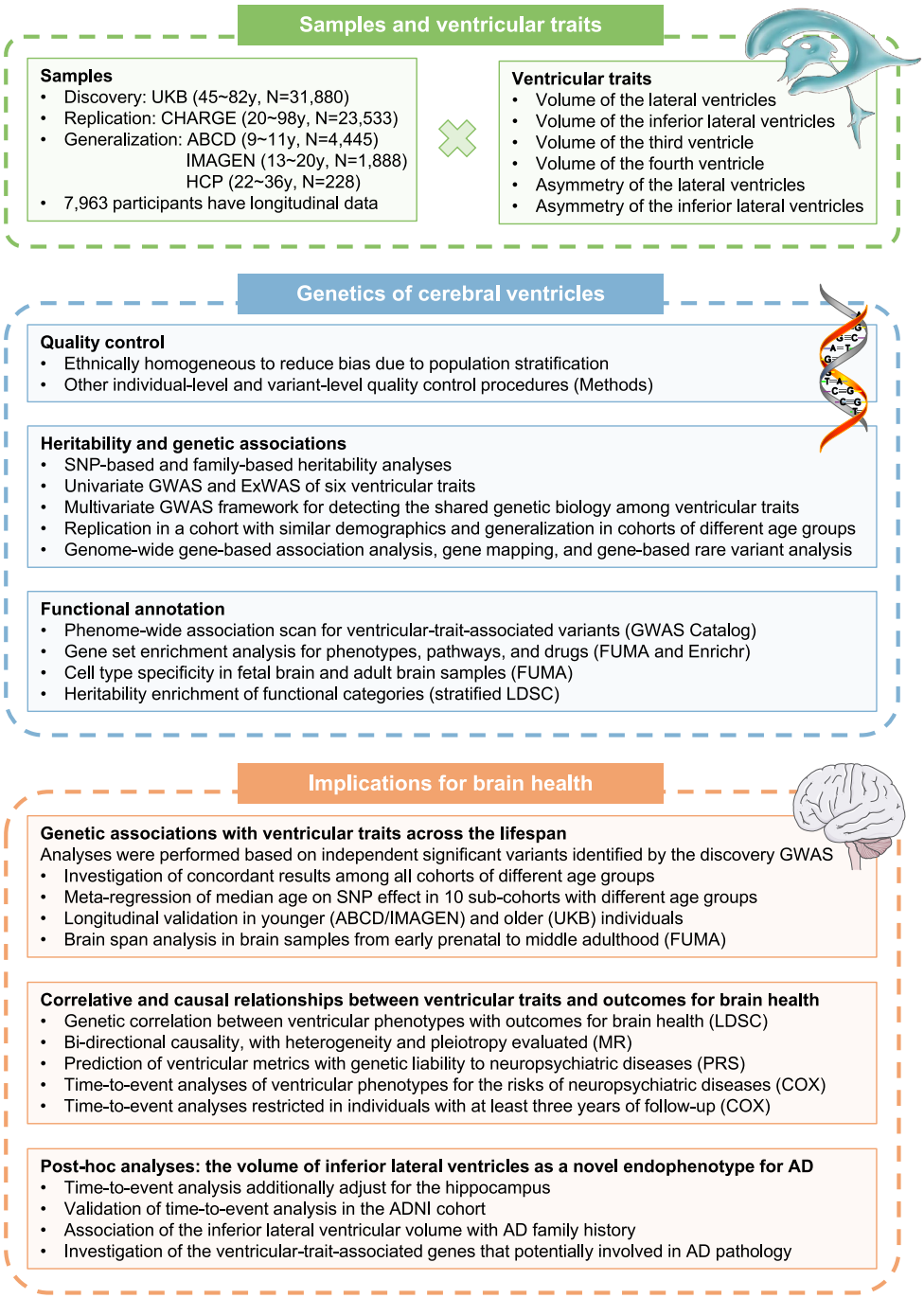
[1]Department of Neurology and National Center for Neurological Disorders, Huashan Hospital, State Key Laboratory of Medical Neurobiology and MOE Frontiers Center for Brain Science, Shanghai Medical College, Fudan University, Shanghai, China. [2]Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China. [3]Department of Neurology, Qingdao Municipal Hospital, Qingdao University, Qingdao, China. [4]Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [5]Department of Neuroimaging, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK. [6]Discipline of Psychiatry, School of Medicine and Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland. [7]Centre for Population Neuroscience and Precision Medicine, Institute of Psychiatry, Psychology & Neuroscience, SGDP Centre, King's College London, London, UK. [8]Institute of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [9]Department of Psychology, School of Social Sciences, University of Mannheim, Mannheim, Germany. [10]NeuroSpin, CEA, Université Paris-Saclay, Gif-sur-Yvette, France. [11]Departments of Psychiatry and Psychology, University of Vermont, Burlington, VT, USA. [12]Sir Peter Mansfield Imaging Centre, School of Physics and Astronomy, University of Nottingham, Nottingham, UK. [13]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. [14]Physikalisch-Technische Bundesanstalt, Braunschweig and Berlin, Germany. [15]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. [16]AP-HP, Sorbonne University, Department of Child and Adolescent Psychiatry, Pitié-Salpêtrière Hospital, Paris, France. [17]Psychiatry Department, EPS Barthélémy Durand, Etampes, France. [18]Institute of Medical Psychology and Medical Sociology, University Medical Center Schleswig Holstein, Kiel University, Kiel, Germany. [19]Institut des Maladies Neurodégénératives, UMR 5293, CNRS, CEA, Université de Bordeaux, Bordeaux, France. [20]Departments of Psychiatry and Neuroscience, Faculty of Medicine and Centre Hospitalier Universitaire Sainte-Justine, University of Montreal, Montreal, Quebec, Canada. [21]Departments of Psychiatry and Psychology, University of Toronto, Toronto, Ontario, Canada. [22]Department of Child and Adolescent Psychiatry and Psychotherapy, University Medical Centre Göttingen, Göttingen, Germany. [23]Department of Psychiatry and Neuroimaging Center, Technische Universität Dresden, Dresden, Germany. [24]Centre for Population Neuroscience and Stratified Medicine, 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]Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence (Fudan University), Ministry of Education, Beijing, China. [27]Fudan ISTBI—ZJNU Algorithm Centre for Brain-Inspired Intelligence, Zhejiang Normal University, Jinhua, China. [28]MOE Frontiers Center for Brain Science, Fudan University, Shanghai, China. [29]Zhangjiang Fudan International Innovation Center, Shanghai, China. [30]Centre for Population Neuroscience and Precision Medicine, Institute for Science and Technology of Brain-Inspired Intelligence, Fudan University, Shanghai, China. [31]Shanghai Medical College and Zhongshan Hospital Immunotherapy Technology Transfer 79 Center, Shanghai, China. [32]These authors contributed equally: Yi-Jun Ge, Bang-Sheng Wu, Yi Zhang. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: wcheng@fudan.edu.cn; jintai_yu@fudan.edu.cn

IMAGEN Consortium

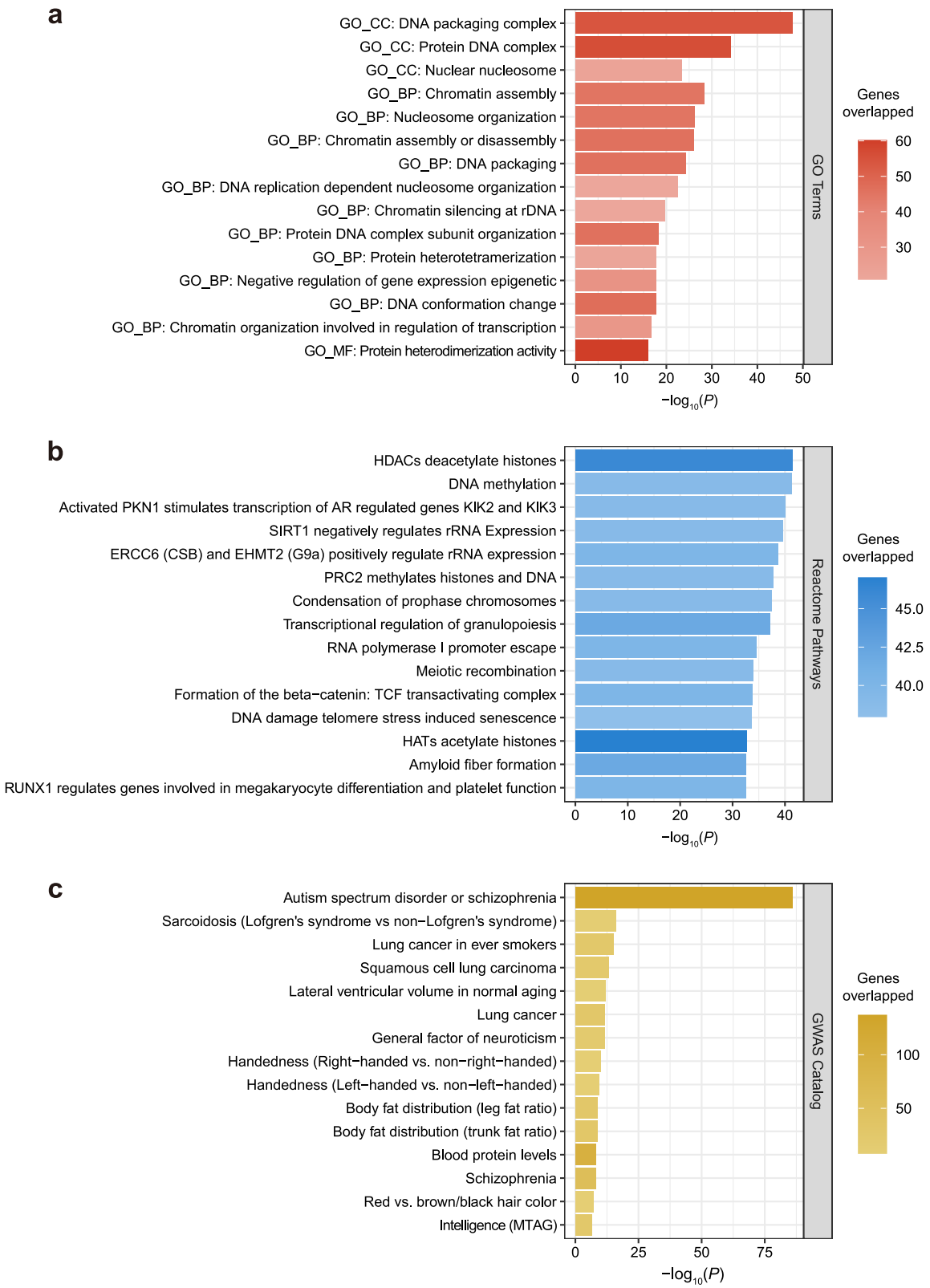
Tobias Banaschewski[4], Gareth J. Barker[5], Arun L. W. Bokde[6], Sylvane Desrivières[7], Herta Flor[8,9], Antoine Grigis[10], Hugh Garavan[11], Penny Gowland[12], Andreas Heinz[13], Rüdiger Brühl[14], Jean-Luc Martinot[15], Marie-Laure Paillère Martinot[15,16], Eric Artiges[15,17], Frauke Nees[4,8,18], Dimitri Papadopoulos Orfanos[10], Herve Lemaitre[10,19], Tomáš Paus[20,21], Luise Poustka[22], Sarah Hohmann[4], Sabina Millenet[4], Juliane H. Fröhner[23], Michael N. Smolka[23], Nilakshi Vaidya[24], Henrik Walter[13], Robert Whelan[25] & Gunter Schumann[24,30]

https://doi.org/10.1038/s41562-023-01722-6



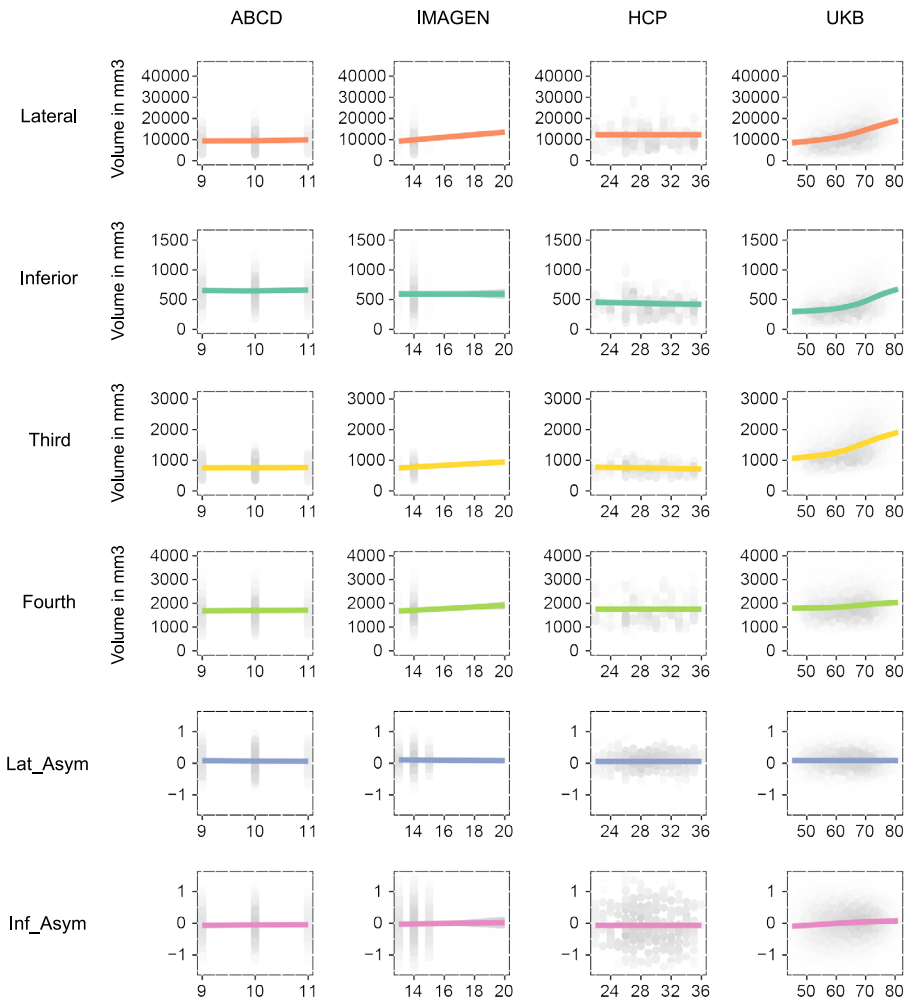
Extended Data Fig. 1 | Schematic diagram of the study design. First, we aim to elucidate the genetic basis of six ventricular phenotypes, including volumes of the lateral, inferior lateral, third, and fourth ventricles, as well as asymmetries of the lateral and inferior lateral ventricles. Second, we aim to leverage genetic data to estimate the value of ventricular traits in monitoring common neuropsychiatric disorders and promoting brain health across the lifespan. The brain and DNA images were produced using Servier Medical Art (http://smart.servier.com/) licensed under CC BY 3.0.

https://doi.org/10.1038/s41562-023-01722-6



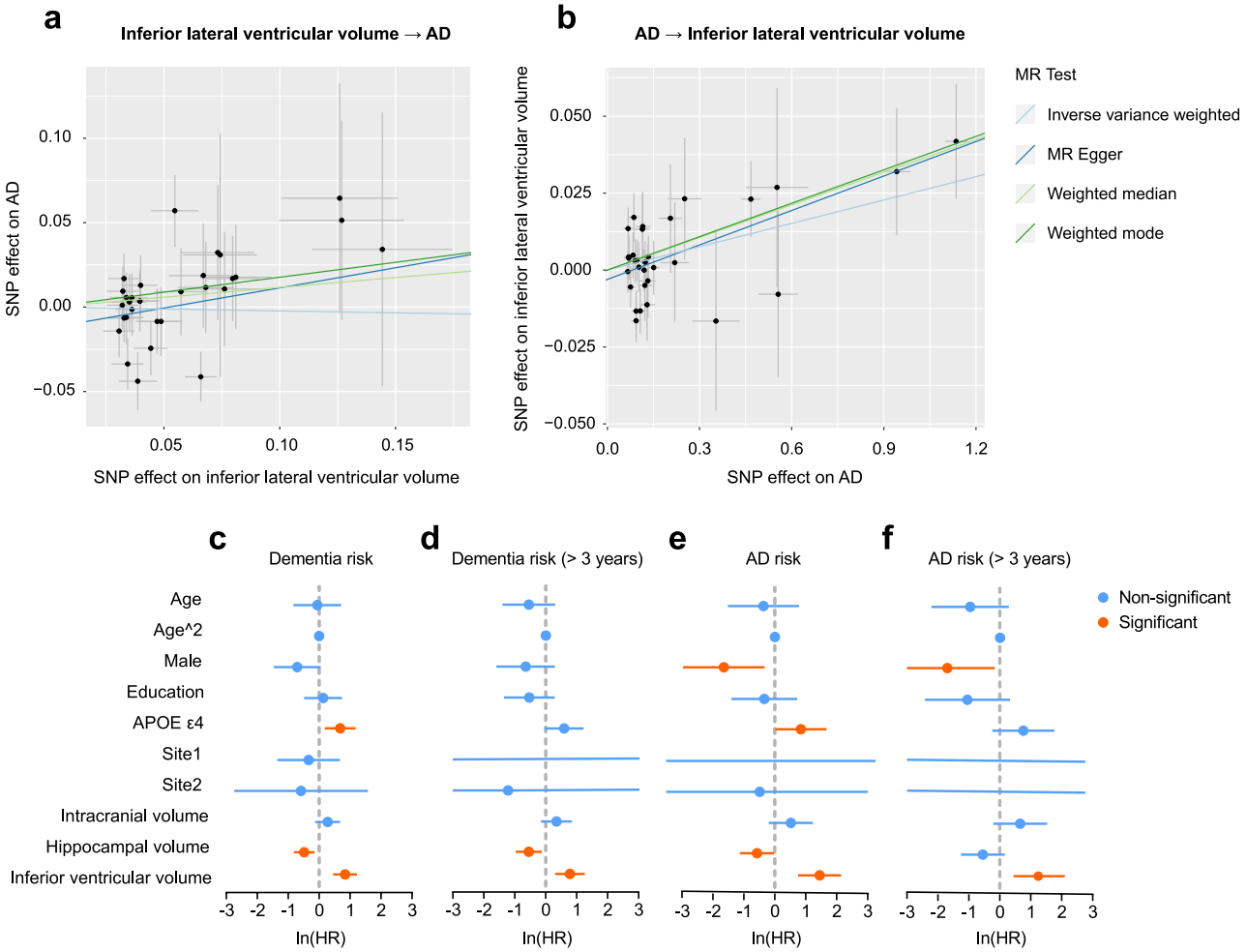
Extended Data Fig. 2 | Top hits in FUMA gene set enrichment analysis. Bar plots show the top enriched GO terms (a), pathways (b), and phenotypes in the GWAS Catalog (c) through the summary statistics of multivariate ventricular GWAS using hypergeometric tests in FUMA. The length of the bars indicates the -log10 P-value of the enrichment analysis. The color of bars shows the number of genes overlapping between cerebral ventricles and other traits. This figure displays raw P-values without FDR correction and full results could be found in Supplementary Table 17.

https://doi.org/10.1038/s41562-023-01722-6



Extended Data Fig. 3 | Associations between age and volumetric metrics in cohorts with different age groups. Individual-level data from ABCD, IMAGEN, HCP, and UKB were used to construct the density plots with no covariates included. This plot only includes data from cross-sectional genetic analyses (N = 38,441).

https://doi.org/10.1038/s41562-023-01722-6



Extended Data Fig. 4 | The volume of the inferior lateral ventricles is a consequence of prodromal AD and could predict AD risk. a-b, Bidirectional MR shows high genetic liability to AD is potentially causal to large inferior lateral ventricles, but not vice versa. The dot refers to each variant's mean value of association estimate with exposure and outcome, while the cross symbol represents SE. The slope of the line corresponds to the total estimated causal effect. The number of instruments is 31 for inferior lateral ventricular volume (a) and 31 for AD (b), respectively. c-f, Multivariate COX analyses indicate the volume of the inferior lateral ventricles could predict the risk of indicant dementia (c-d) and AD (e-f). Analyses were performed without restriction (c and e, sample size = 15,198) and restricted in participants with at least three years of follow-up (d and f, sample size = 4,930). Compared to volumes of the hippocampus, the volume of the inferior lateral ventricles has even stronger significance and larger effect size.

nature portfolio

Corresponding author(s): Jin-Tai Yu, Wei Cheng

Last updated by author(s): Sep 6, 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 used to collect data.
Data analysis	R version 4.0.3 were used for data cleansing and visualization. PLINK version 2.0 (https://www.cog-genomics.org/plink/) for genetic control and univariate GWAS; MOSTest on matlabR2018b(https://github.com/precimed/mostest) for multivariate GWAS; SAIGE-GENE+ (https://saigegit.github.io/SAIGE-doc/) for WES analysis; GCTA version 1.93.2 (http://cnsgenomics.com/software/gcta/) and R for heritability estimation; IMPUTE version 2 (https://mathgen.stats.ox.ac.uk/impute/impute_v2.html) and Michigan Imputation Server (https://imputationserver.sph.umich.edu/) for imputation; FUMA version 1.5.4 (https://fuma.ctglab.nl/) for annotation of genetic locus; MAGMA version 1.08 (https://ctg.cncr.nl/software/magma/, also implemented in FUMA) for gene analysis; METAL version 2011-03-25 (http://www.sph.umich.edu/csg/abecasis/Metal/) for meta analysis of GWAS summary statistics; Enrichr (https://maayanlab.cloud/Enrichr/) and FUMA version 1.5.4 (https://fuma.ctglab.nl/) for gene set enrichment; LDSC version 1.0.1 (https://github.com/bulik/ldsc/) for genetic correlation 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

Our GWAS summary statistics for cerebral ventricles can be found at https://doi.org/10.6084/m9.figshare.21529491. GWAS results are also available on the FUMA website (https://fuma.ctglab.nl/browse/; ID 619–625). The individual-level data used in the present study were obtained from UKB (https://www.ukbiobank.ac.uk/), ABCD (https://abcdstudy.org/), IMAGEN (http://imagen-project.org/), HCP (http://www.humanconnectome.org/), and ADNI (adni.loni.usc.edu). The summary-level data for the CHARGE consortium were obtained from dbGaP (phs000930.v9.p1). GWAS Catalog resource can be found in https://www.ebi.ac.uk/gwas/. GWAS atlas can be found in https://atlas.ctglab.nl/PheWAS/. DSigDB database can be found in http://dsigdb.tanlab.org/DSigDBv1.0/. Agora can be found in https://agora.adknowledgeportal.org/. The Human Protein Atlas can be found in https://www.proteinatlas.org/.

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, IMAGEN, HCP, and ADNI was determined based on self-reporting data via questionnaire, and all included participants gave written informed consent for sharing of individual-level data.
Population characteristics	This study included individual-level data from ABCD (age range = 9-11 years, 48% female), IMAGEN (age range = 13-20, 51% female), HCP (age range = 22-36, 54% female), UKB (age range = 45-82, 54% female), and summary-level data from the CHARGE consortium (age range = 20-98, 56% female). Baseline descriptions can be found in Supplementary Table 1.
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. In HCP study, the 1200 participants were to be recruited as families of twins and non-twin siblings to enable heritability and imaging genetic analyses of individual variability in brain connectivity.
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. 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. The HCP was reviewed and approved by the Institutional Ethics Committee of Washington University in St. Louis, Missouri.

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	The CHARGE cohort has similar demographics to UKB, and its summary statistics were applied for replication analysis. Most (83%) independent significant variants were replicated in CHARGE after Bonferroni correction.
Randomization	Age, age2, sex, intracranial volume, imaging center, and the first ten 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	No behavioral performance was assessed, only structural MRI.

Acquisition

Imaging type(s)	T1-weighted structural imaging
Field strength	3T
Sequence & imaging parameters	Varying (description in Supplementary Table 41)
Area of acquisition	Whole brain, including cerebral ventricles

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software	Varying (description in each dataset description) but mainly based on Freesurfer recon-all
Normalization	Varying (description in each dataset description) but mainly based on Freesurfer recon-all
Normalization template	Varying (description in each dataset description) but mainly based on Freesurfer recon-all
Noise and artifact removal	Varying (description in each dataset description) but mainly based on Freesurfer recon-all
Volume censoring	Varying (description in each dataset description) but mainly based on Freesurfer recon-all

Statistical modeling & inference

Model type and settings	Linear regression under the addictive model
Effect(s) tested	Associations of SNPs across the whole genome with MRI-based ventricular traits

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

Anatomical location(s) subcortical volumetric segmentation (aseg)

Statistic type for inference (See Eklund et al. 2016) Region-based measures

Correction The paper includes various different analyses and the multiple testing approach for each is described explicitly in the methods, either Bonferroni or FDR.

Models & analysis

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