Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genetic associations with cortical thickness and surface area and their distinct developmental trajectories

Chunshui Yu

chunshuiyu@tmu.edu.cn

Tianjin Medical University General Hospital https://orcid.org/0000-0001-5648-5199

Nana Liu

Tianjin Medical University General Hospital

Jiayuan Xu

Tianjin Key Lab of Functional Imaging & Tianjin Institute of Radiology https://orcid.org/0000-0001-9473-1047

Jie Tang

Laboratory of Functional Imaging, Tianjin Medical University

Jilian Fu

Tianjin Medical University General Hospital

Sijia Wang

Tianjin Medical University General Hospital https://orcid.org/0000-0001-8949-303X

Yuan Ji

Tianjin Medical University General Hospital https://orcid.org/0000-0003-2874-5904

Hui Xue

Tianjin Medical University General Hospital

Nannan Zhang

Tianjin Medical University General Hospital

Qiang Xu

Tianjin Medical University General Hospital https://orcid.org/0000-0003-4744-5616

Lining Guo

Tianjin Medical University General Hospital

Hao Ding

School of Medical Imaging and School of Medical Technology, Tianjin Medical University

Libo Su

Tianjin Medical University

Ping Wang

School of Medical Imaging and School of Medical Technology, Tianjin Medical University

Huaigui Liu
Tianjin Medical University General Hospital
Feng Liu
Tianjin Medical University General Hospital https://orcid.org/0000-0002-3570-4222
Meng Liang
Tianjin Medical University https://orcid.org/0000-0003-0916-520X
Wen Qin
Tianjin Medical University General Hospital https://orcid.org/0000-0002-9121-8296

Biological Sciences - Article

Keywords:

Posted Date: May 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8949457/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Abstract

In the human cerebral cortex, surface area follows a more prolonged developmental trajectory than cortical thickness; however, the genetic mechanisms underlying this difference remain unclear. We conducted genome-wide association meta-analyses for 64 cortical traits in 100,628 participants and identified 213 loci for cortical thickness and 417 loci for surface area (45 and 79 new loci), mapping to 184 and 431 genes (82 overlaps), respectively. Although thickness- and area-related genes exhibited similar functional enrichments, their cell type-specific expression curves during development showed distinct associations with growth-rate curves, with thickness growth showing a greater proportion of correlations with gene expression in inhibitory neurons. Even the shared genes influenced thickness growth and area expansion through distinct cell types and temporal lags. These findings indicate that the differing developmental trajectories of cortical thickness and surface area may arise from distinct cell type-specific gene expression and temporal dynamics.

Introduction

The expansion and elaboration of the cerebral cortex are hallmarks of human evolution, underpinning higher-level cognitive functions[1]. The enlargement of the human cerebral cortex is primarily driven by a dramatic expansion of surface area (SA), accompanied by a moderate increase in cortical thickness (CT). For instance, humans demonstrate an approximately 1000-fold increase in SA and a twofold increase in CT relative to mice[1,2]. Additionally, humans exhibit a more prolonged cortical developmental trajectory than other species, resulting in a larger number of neurons, a higher proportion of glial cells, and more complex neuronal morphologies and synaptic connections[3]. Furthermore, in the human cerebral cortex, SA expansion continues for a much longer period than CT growth. For example, according to brain charts of mean CT and total SA derived from 123,984 magnetic resonance imaging (MRI) scans spanning the age range from mid-gestation to 100 postnatal years[4], SA expansion peaked at approximately 11.0 years, while CT growth peaked around 1.7 years. However, the mechanisms underlying the distinct developmental timelines of CT and SA remain poorly understood.

Twin and family-based studies indicate that image-derived phenotypes (IDPs) for both CT and SA are highly heritable[5]. Consequently, genome-wide association studies (GWASs) have been conducted on these IDPs, identifying hundreds of loci associated with CT and SA[6–16]. Although these findings advance our understanding of the genetic architecture underlying inter-individual differences in CT and SA, current sample sizes remain insufficient to detect all genetic variants accounting for their heritability, calling for larger-scale GWASs. Additionally, the cerebral cortex is key for cognitive function and vulnerable to neuropsychiatric disorders[17,18]; therefore, linking CT- and SA-related genetic variants to neurocognition and neuropsychiatric diseases is valuable and may shed light on their genetic and neural mechanisms.

Using single-nucleus RNA sequencing (snRNA-seq) data from the human cerebral cortex across developmental periods, previous studies have identified numerous genes whose cell type-specific

expression is related to cortical development[19–21]. However, these studies cannot distinguish genes associated with CT growth from those associated with SA expansion, a limitation that may be addressed by integrating GWAS findings for CT-IDPs and SA-IDPs. Because both CT and SA growth rates are regulated by gene expression, investigating the associations of cell type-specific gene expression curves with CT and SA growth-rate curves within the same timeframe may reveal the genetic mechanisms driving CT and SA development at the cellular level.

In this study, we conducted the largest GWAS meta-analyses to date on 68 cortical IDPs in 100,628 participants to identify genetic loci associated with CT and SA, and mapped these loci to genes. We then examined associations between cell type-specific gene expression curves and CT and SA growth-rate curves from fetal to adult stages to elucidate genetic mechanisms underlying CT and SA development.

Results

Genetic associations with CT and SA

We assessed cortical morphology using 68 bilateral averaged cortical IDPs, including the CT and SA of 33 distinct cortical regions (Fig. 1a) from the Desikan-Killiany atlas[22], and the mean CT and total SA of the entire cerebral cortex. In the discovery phase, we performed GWAS meta-analyses on cortical IDPs in 80,935 individuals of European ancestry (EUR) by integrating data from the UK Biobank (UKB) cohort (n = 52,278), the Enhancing Neuroimaging Genetics through Meta-Analysis (ENIGMA) consortium (n = 23,909), and the Adolescent Brain Cognitive Development (ABCD) study (n = 4,748). The intercepts of Linkage Disequilibrium (LD) score regression (LDSC) for all GWASs were close to 1 (Table S1), providing evidence for no population stratification. We estimated single nucleotide polymorphism (SNP)-based heritability for each cortical IDP, with h^2 ranging from 0.08 to 0.36 (Table S2). These GWASs identified 2,028 significant variant-trait associations ($P < 7.35 \times 10^{-10}$, Bonferroni correction for 68 IDPs; Fig. S1 and Table S3), with 1,606 associations for SA (Fig. 1b) and 422 for CT (Fig. 1c). These associations were merged into 1,508 trait-associated loci (1,154 for SA and 354 for CT; Fig. 1d). Of the 354 CT-associated loci, only 75 (21.19%) overlapped with SA-associated loci for the same cortical region (Fig. 1d), indicating that CT and SA of the same cortical region are primarily controlled by different genetic loci. After merging association signals (loci) for SA-IDPs and CT-IDPs, respectively, we identified 787 independent association signals (417 loci) for SA and 300 association signals (213 loci) for CT, including 137 overlapping loci. Of these, 79 SA- and 45 CT-related loci have not previously been reported at $P < 7.35 \times 10^{-10}$. In the replication phase, we validated the discovered variant-trait associations in 19,693 mixed-ancestry participants, including 8,333 non-White British participants from UKB, 4,329 multi-ancestry participants from ABCD, and 7,031 East Asian participants from the Chinese Imaging Genetics (CHIMGEN) study. Among the 2,022 discovered associations available for replication, we found high concordance ($r = 0.88$; Fig. S2) of effect sizes between the discovery and replication samples. We defined replication by considering both lead and high-LD variants ($r^2 > 0.6$), and validated 1,306 (64.59%) variant-trait associations at $P < 0.05$ with consistent effect direction. Consequently, the EUR-GWAS

summary statistics for the 68 cortical IDPs, derived from the discovery sample, were used in all subsequent analyses.

We utilized C-GWAS[23] to conduct multivariate GWAS ($P < 5 \times 10^{-8}$) for the 34 CT-IDPs and the 34 SA-IDPs, respectively. We found 1,989 genetic associations (772 loci) for SA and 835 genetic associations (498 loci) for CT, including 410 overlapping loci (Table S4). Of these, 280 SA-related and 202 CT-related loci were not found in univariate GWASs, comprising 106 SA-related and 45 CT-related loci (Fig. 1e) that have not been reported previously ($P < 5 \times 10^{-8}$).

We also investigated genetic correlations ($P_c < 0.05$, Bonferroni corrected) within and between CT-IDPs and SA-IDPs. Among the genetic correlations between CT and SA within the same region, we identified 16 negative and 6 positive genetic correlations (Fig. S3). The extensive negative genetic correlations aligned with a prior study[7] and were primarily present in late-maturing association cortices (e.g., frontal and parietal regions), whereas positive genetic correlations were mainly observed in early-maturing somatosensory and visual areas[24]. These findings indicate that genetic variants associated with SA expansion may contribute to cortical thinning in association cortices, but to cortical thickening in sensory cortices. For within-measure genetic correlations between cortical regions (Fig. S4), we found positive genetic correlations for CT or SA between physically adjacent regions, suggesting that SA expansion or CT increases in these regions share the same genetic substrates.

Variant prioritization and annotation

After excluding loci in the major histocompatibility complex (MHC) region, we applied the Sum of Single Effects (SuSiE) model[25] to perform statistical fine mapping for 1,501 trait-associated loci (1,151 for SA and 350 for CT) from EUR-GWASs. For each locus, we estimated the posterior inclusion probability (PIP) of each variant within the locus and calculated its 95% credible set of causal variants. We found 1,203 credible causal variants (PIP > 0.8) for SA and 366 for CT (Fig. 2a), including 81 shared by both CT and SA. For instance, rs2279829 in the 3'-UTR of *ZIC4* was a causal variant (all PIPs > 0.8) for seven SA-IDPs, and *ZIC4* plays a key role in neural development[26]. A missense mutation (rs13107325) in *SLC39A8* was a causal variant for eight CT-IDPs (all PIPs > 0.8). *SLC39A8* encodes a membrane transporter critical for manganese homeostasis and glycosylation; its deficiency in humans is linked to abnormal brain development[27]. The rs2696531 in the 3'-UTR of *ARL17B* was a causal variant for four SA-IDPs and three CT-IDPs. *ARL17B* is a risk gene for neurodegenerative diseases[28].

We used FUMA[29] to perform functional annotation of causal variants within 95% credible sets for CT- and SA-associated loci, respectively. ANNOVAR[30] was utilized to categorize variants by genic position, while combined annotation-dependent depletion (CADD) score[31] was applied to assess their pathogenicity. We identified 128 missense variants in 96 genes for SA and 46 missense variants in 41 genes for CT, including 12 shared missense variants in 12 unique genes (Table S5). Using PIP > 0.8 to filter causal missense variants with high confidence, we identified 14 credible causal missense variants

for SA and six for CT, comprising three shared variants (Fig. 2a). The three shared missense variants included rs1801133 (PIP_{max} = 0.98 for SA, PIP_{max} = 0.99 for CT, CADD = 25.5) in *MTHFR*, rs10283100 (PIP_{max} = 1 for SA, PIP_{max} = 0.99 for CT, CADD = 22.4) in *ENPP2*, and rs35891966 (PIP = 0.99 both SA and CT, CADD = 28.1) in *NAV2*. All three genes are involved in brain development: *MTHFR* in neural tube development[32], *ENPP2* in the localization and adhesion of neuronal progenitors[33], and *NAV2* in neurite outgrowth and axonal elongation[34]. The rs2066827 was identified as a SA-only missense variant (PIP_{max} = 1, CADD = 18.05) in *CDKN1B*, which plays a role in neural progenitor proliferation, cell cycle progression, and neuron migration[35–37]. The rs2286471 was identified as a CT-only missense variant (PIP_{max} = 1, CADD = 18.05) in *NPW*, a key regulator of the hypothalamic-pituitary-adrenal (HPA) axis[38], with a profound influence on cortical development[39].

For 1,501 trait-associated loci, we used OPERA[40] to jointly analyze GWAS signals and multi-omics quantitative trait loci (xQTLs) of the human cerebral cortex to identify molecular phenotypes that share genetic signals with cortical IDPs. We included QTLs of DNA methylation (meQTLs, n = 1,160)[41], histone acetylation (haQTLs, n = 561)[42], chromatin accessibility (caQTLs, n = 272)[43], gene expression (eQTLs, n = 2,865)[44], and protein (pQTLs, n = 1,277)[45]. For each GWAS signal for cortical IDPs, we calculated the joint posterior probability of association (PPA) for every combination of molecular phenotypes. For pleiotropic associations with PPA > 0.9, we performed a multi-trait HEIDI test to exclude the LD-related false positive findings ($P_{HEIDI} \leq 0.01$). We found that 631 GWAS signals were shared with at least one molecular phenotype (PPA > 0.9 and $P_{HEIDI} > 0.01$), of which 27 were shared with four or more molecular phenotypes (Table S6). For instance, a GWAS signal (lead SNP: rs17055142 at chr8; gene: *STMN4*) for mean CT was associated with all five molecular phenotypes (joint PPA = 0.99, P_{HEIDI} = 0.12; Fig. 2b), and another GWAS signal (lead SNP: rs7550758 at chr1; gene: *ATP13A2*) for the SA of the pars triangularis in the inferior frontal cortex was associated with four molecular phenotypes (joint PPA = 0.99, P_{HEIDI} = 0.13; Fig. 2c). *STMN4* plays a key role in microtubule depolymerization and neuron projection development[46], while *ATP13A2* maintains neuronal integrity[47] and has been linked to an early onset form of Parkinson's disease (PD)[48]. Several GWAS signals from both CT and SA traits converged on molecular phenotypes of *KANSL1*, a gene involved in the regulation of synaptic structure[49].

We used the FLAMES[50] framework for gene prioritization by integrating SNP-to-gene evidence and convergence-based evidence into a single prediction for each fine-mapped GWAS signal. FLAMES identified 431 unique genes for SA-IDPs and 184 for CT-IDPs, with 82 overlapping genes (Table S7). We used WebGestalt[51] to perform statistical over-representation analysis to identify the biological processes enriched by these genes. We identified 403 significant pathways for SA and 53 for CT (P_c < 0.05, Bonferroni-corrected; Table S8), including 52 overlapping pathways (Fig. 2d). The shared pathways primarily comprised neurodevelopmental processes, such as neurogenesis ($P_c = 3.09 \times 10^{-25}$ for SA and $P_c = 6.52 \times 10^{-9}$ for CT), neuron differentiation ($P_c = 3.34 \times 10^{-21}$ for SA and $P_c = 3.71 \times 10^{-6}$ for CT), and neuron development ($P_c = 8.14 \times 10^{-13}$ for SA and $P_c = 1.23 \times 10^{-3}$ for CT); signaling

pathways, such as the Wnt signaling pathway (P_c = 2.51 × 10^{-13} for SA and P_c = 2.55 × 10^{-4} for CT); neuronal projection development, including neuron projection morphogenesis (P_c = 2.93 × 10^{-14} for SA and P_c = 2.89 × 10^{-6} for CT) and axonogenesis (P_c = 5.81 × 10^{-9} for SA and P_c = 7.73 × 10^{-3} for CT); and cell migration (P_c = 1.10 × 10^{-21} for SA and P_c = 1.58 × 10^{-6} for CT). The high proportion (98.1%) of overlap between CT-related and SA-related pathways indicates that enrichment analysis cannot distinguish the genetic mechanisms underlying CT and SA.

Cell type-specific gene expression and CT and SA growth-rates

We investigated genetic mechanisms underlying the distinct developmental trajectories of CT and SA by integrating cell type-specific expression trajectories of the prioritized CT-related and SA-related genes across developmental periods with lifetime growth-rate curves of CT and SA. We focused on 102 CT-related, 349 SA-related, and 82 CT-SA shared genes that were prioritized from CT-GWASs and SA-GWASs, respectively. We subsequently obtained their cell type-specific gene expression trajectories across six developmental periods (fetal, neonatal, infancy, childhood, adolescence, and adult) in 14 major cell clusters from the human prefrontal cortex[19]. The 14 major clusters were derived from 17 cell clusters (Fig. 3a) after integrating three categories of immature neurons (PN-dev, CGE-dev, and MGE-dev) into their respective mature neuronal categories for gene expression curve fitting. The lifetime growth-rate curves of mean CT and total SA were derived from brain charts for the human lifespan[4].

For each cell type, we generated cell type-specific time-expression curve of each gene from the 22nd week of gestation to 40 years[19]. Developmental ages were arcsinh-transformed to expand the scale of early developmental windows, thereby capturing the accelerated pace of early-life developmental dynamics. Gene expression values were extracted at 100 uniformly distributed time points across this transformed temporal axis, and corresponding growth-rate values for CT and SA were subsequently quantified at the same time points[4]. The first time point lacked a growth-rate value since derivatives require prior data points, leaving 99 values for correlation analysis. We then calculated cross-correlations between cell type-specific gene expression values (x) and growth-rate values (y) across the 99 time points. For each lag k, the coefficient r_{xy} (k) was defined as normalized cross-correlation coefficients. We defined the optimal lag as the value of k that produced the largest absolute correlation. Cross-correlations were considered significant if the maximal coefficient satisfied $|r_{xy}(k)| > 0.8$. We then validated the main findings by extracting cell type-specific gene expression values for the same 99 time points from an independent snRNA-seq dataset[20].

In the discovery snRNA-seq dataset[19], we found 1,380 cross-correlations between cell type-specific expression of 304 SA-related genes and SA growth-rate, and 301 cross-correlations between cell type-specific expression of 84 CT-related genes and CT growth-rate (Table S9). We categorized these correlations by cell types (Fig. 3b) and calculated the relative proportion of correlations for each cell type (Fig. 3c). We also calculated the relative proportions of correlations for three major categories

(excitatory neurons, inhibitory neurons, and glial cells, Fig. 3c). Among the 1,380 SA correlations, the proportion were 36.74% for inhibitory neurons, 33.41% for excitatory neurons, and 29.86% for glial cells (Fig. 3c). *NCAM1* expression in excitatory neurons (L2/3-CUX2) showed the strongest positive correlation with SA growth-rate ($r_{xy\,(0)}$ = 0.996; Fig. 3d). Among the 301 CT correlations (Fig. 3e), the proportion was much higher in inhibitory neurons (47.84%) than in excitatory neurons (27.24%) or glial cells (24.92%, Fig. 3f). For example, *SYBU* expression in PV-SCUBE3 inhibitory neurons exhibited a negative correlation with CT growth-rate ($r_{xy\,(0)}$ = -0.97; Fig. 3g). Given the distinct cell type definitions in the replication dataset[20], we focused our validation on the proportion of CT- and SA-related cross-correlations. We observed similar SA and CT correlation distributions across the three cell categories (Fig. S5). The proportions of SA correlations were 44.82% for inhibitory neurons, 36.35% for excitatory neurons, and 18.83% for glial cells, and the proportions of CT correlations were 52.84% for inhibitory neurons, 28.22% for excitatory neurons, and 18.94% for glial cells. The differences in the proportions of SA and CT correlations across major cell categories may contribute to the developmental timeline differences between SA expansion and CT growth.

Among the 82 CT-SA shared genes, we identified 362 cross-correlations for SA and 284 cross-correlations for CT (Fig. 4a), with only 72 overlapping correlations in the discovery dataset. In the validation dataset, we replicated this pattern by identifying 368 cross-correlations for SA and 529 cross-correlations for CT (Fig. 4b), with only 51 overlapping correlations. These findings indicate that most of these CT-SA shared genes identified by GWASs act within distinct cell types. For instance, *NAV2* was identified as a CT-SA shared gene. Its expression in excitatory neurons correlated with SA growth-rate ($r_{xy(-5)}$ = 0.94; Fig. 4c), while its expression in inhibitory neurons correlated with CT growth-rate ($r_{xy(0)}$ = -0.91; Fig. 4d). Among the 72 CT-SA shared cross-correlations identified in the discovery dataset (Fig. 4a) and the 51 CT-SA shared cross-correlations identified in the validation dataset (Fig. 4b), most (all in discovery and 46 in validation) showed zero-lag relative to CT growth-rate, but all showed negative lags relative to SA growth-rate. For instance, *DACT1* expression in ID2 neurons preceded SA growth-rate by 9-time units ($r_{xy(-9)}$ = 0.81; Fig. 4e), while it was correlated with CT growth-rate with zero-lag ($r_{xy(0)}$ = 0.87; Fig. 4f). These results suggest that even the CT-SA shared gene expression within the same cell type may influence CT growth and SA expansion with distinct temporal lags.

Genetic correlation and colocalization between cortical and other traits

For the 68 cortical traits, we computed their genetic correlations with three cognitive traits and eight neuropsychiatric disorders. We found 37 significant genetic correlations ($P < 6.68 \times 10^{-5}$, Bonferroni corrected), including 18 correlations for SA and 19 for CT traits (Fig. 5a-b and Table S11). The total SA showed positive genetic correlation with all cognitive traits: years of schooling (r_g = 0.24, $P = 5.87 \times 10^{-87}$), intelligence (r_g = 0.25, $P = 2.11 \times 10^{-67}$), and cognitive performance (r_g = 0.26, $P = 1.11 \times 10^{-81}$). However, total SA showed mixed correlations with neuropsychiatric disorders: negative correlations

with attention-deficit/hyperactivity disorder (ADHD, r_g = -0.21, $P = 3.18 \times 10^{-24}$) and depression (r_g = -0.09, $P = 2.63 \times 10^{-7}$) and a positive correlation with PD (r_g = 0.20, $P = 2.21 \times 10^{-5}$). The SA of the pars orbitalis, inferior parietal lobule, insula, entorhinal cortex, and superior temporal gyrus showed positive correlations with cognitive traits, while the SA of the precuneus, isthmus cingulate, lingual gyrus, and pericalcarine cortex exhibited negative correlations with cognitive traits (Fig. 5a). The CT of the rostral anterior cingulate, superior and inferior parietal, and lateral occipital cortices showed negative genetic correlations with cognitive traits, with the strongest correlation between the inferior parietal CT and years of schooling (r_g = -0.14, $P = 4.45 \times 10^{-16}$) (Fig. 5b-c). Conversely, the precentral and superior temporal CT showed positive genetic correlations with cognitive traits. As for neuropsychiatric disorders, the rostral anterior cingulate (r_g = 0.13, $P = 2.17 \times 10^{-6}$) and precentral (r_g = -0.11 $P = 1.05 \times 10^{-5}$) cortices exhibited genetic correlations with ADHD. The medial orbitofrontal cortex (r_g = -0.07, $P = 3.32 \times 10^{-6}$) and pars orbitalis (r_g = -0.09, $P = 3.97 \times 10^{-5}$) exhibited genetic correlations with schizophrenia. The superior parietal CT showed genetic associations with depression (r_g = 0.08, $P = 7.65 \times 10^{-6}$). These findings indicated that SA and CT exhibited complex genetic correlation patterns with cognitive functions and neuropsychiatric disorders.

For each trait-associated locus (1,151 for SA and 350 for CT), we performed multi-trait colocalization with three cognitive traits and eight neuropsychiatric disorders. We identified 60 colocalizations with a posterior probability of full colocalization (PPFC) ≥ 0.8 (Fig. 5d and Table S12), including 41 for SA and 19 for CT. Among these colocalizations, 12 SA-related and 12 CT-related traits were colocalized with two or more cognitive and neuropsychiatric phenotypes. SA traits were colocalized with all cognitive traits (years of schooling, intelligence, and cognitive performance) and with five neuropsychiatric disorders (depression, ADHD, bipolar disorder, schizophrenia, and autism spectrum disorder (ASD)). For example, total SA colocalized with cognitive performance and intelligence (PPFC = 0.99), with the highest posterior probability at rs11079849 in *IGF2BP1* (Fig. 5e); paracentral SA colocalized with schizophrenia (PPFC = 0.99), with the highest posterior probability at a missense variant (rs11692435) in *ACTR1B*; and total SA colocalized with ADHD, cognitive performance, intelligence, and years of schooling (PPFC = 0.95) at a shared locus in *ARHGAP39*. CT traits showed colocalizations with all cognitive traits and schizophrenia. For instance, the mean CT colocalized with years of schooling (PPFC = 0.98), exhibiting the highest posterior probability at a missense variant (rs10901333) in *LAMC3*; and CT in the caudal middle frontal cortex colocalized with schizophrenia, cognitive performance, intelligence, and years of schooling (PPFC = 0.99), with the highest posterior probability at a missense variant (rs13107325) in *SLC39A8* (Fig. 5f).

Discussion

In this study, we conducted the largest-to-date GWAS meta-analyses on 68 SA and CT traits in 100,628 participants, identifying 417 SA-associated and 213 CT-associated loci. Compared with previous studies[6–16], we found 79 and 45 new loci associated with SA and CT traits, respectively, which advance

our understanding of the genetic architecture of these two morphological measures of the human cerebral cortex.

We also conducted several post-GWAS analyses to enhance understanding of the genetic mechanisms underlying SA and CT development. For instance, fine mapping prioritized 1,203 causal variants for SA and 366 for CT, including 128 missense variants in 96 genes for SA and 46 missense variants in 41 genes for CT. Some of these genes (e.g., *MTHFR*, *ENPP2*, *NAV2*, and *CDKN1B*) have been linked to brain development[35–37]. These findings provide more candidate variants and genes for functional experiments to investigate the genetic control of cortical development. We also examined pleiotropic associations of SA-and CT-associated loci with DNA methylation, histone acetylation, chromatin accessibility, gene expression, and protein expression, annotating 631 loci with at least one molecular phenotype, 27 of which were shared across four or more phenotypes. The results may inform future studies on the molecular pathways by which genetic variation affects cortical morphology. We prioritized 431 genes for SA and 184 for CT, with 82 shared genes, which can enhance biological interpretation of GWAS findings. We revealed that CT-related genes exhibited substantial overlap (52/53, 98.1%) with SA-related genes in functional enrichment for biological processes. These results indicate that conventional enrichment analysis cannot distinguish biological processes specific to CT growth and SA expansion. We investigated functional significance of cortical GWAS findings, and found genetic correlations and colocalizations between cortical traits and cognitive and neuropsychiatric phenotypes, including several multi-trait colocalizations with both categories of phenotypes. These findings may inform genetic and neural mechanisms underlying cognitive impairment in neuropsychiatric disorders.

The most important contribution of this study to the field of genetic studies of the human cerebral cortex is the identification of potential genetic mechanisms underlying SA expansion and CT growth during development. For the first time, we calculated the cross-correlations to connect CT and SA growth-rate trajectories[4] to cell type-specific expression curves[19,20] of the CT- and SA-related genes across developmental periods. We identified 1,380 and 301 correlations between cell type-specific expression of 304 SA-related genes and SA growth-rate, and between cell type-specific expression of 84 CT-related genes and CT growth-rate, respectively. These findings indicate that distinct genes may affect CT and SA growth-rates through their cell type-specific expression. SA and CT differences in developmental trajectories may also be attributed to the cell-type differences between CT and SA cross-correlations. For instance, we found that a greater proportion of CT correlations were observed for inhibitory neurons compared with SA correlations (discovery: 47.84% vs 36.74%; validation: 52.84% vs 44.82%). Additionally, most of the 82 CT-SA shared genes correlated with CT and SA growth-rates by affecting gene expression in distinct cell types. For example, *NAV2* accelerated SA expansion through excitatory neurons, while suppressing CT growth via inhibitory neurons. For CT-SA shared gene expression within the same cell type, they influenced CT growth and SA expansion with distinct temporal lags. These results may provide insights into the genetic and cellular mechanisms governing the distinct developmental trajectories between SA and CT.

Our findings expand the genetic architecture of cortical morphology by identifying novel genetic loci, implicating key causal variants, genes, and pathways, and revealing genetic and cellular mechanisms underlying human cortical development. The genetic correlations and colocalizations of CT- and SA-associated loci with cognitive functions and neuropsychiatric disorders highlight the functional implications of our findings.

Declarations

Acknowledgements

We are grateful to ENIGMA for providing GWAS summary statistics of cortical thickness and surface area. This work was funded by the National Natural Science Foundation of China (Grant No. 82430063 and 82030053 to C.Y., 82402218 to N.L., 82472052 and 81971599 to W.Q., 82371924 to J.X. and 82202093 to J.T., the Tianjin Municipal Health Commission Science and Technology Project to N.L. (Grant No. TJWJ2025QN006), the Tianjin Medical University Young Scholar Program to N.L., the Tianjin Young Talents in Science and Technology to J.X. (Grant No. QN20230336), the National Key Project of “Inter-governmental International Scientific and Technological Innovation Cooperation” to J.X. (Grant No. 2023YFE0199700), the Natural Science Foundation of Tianjin to J.X. (25JCZDJC00640), the Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences to J.X. (Grant No. 2024-JKCS-18), the Tianjin Science and Technology Commission Major Special Project in Public Health Science and Technology to J.X. (Grant No. 24ZXGQSY00050) and the Tianjin Medical University “Clinical Talent Training 123 Climbing Plan” to J.X.

Author contributions

Conceptualization: Chunshui Yu, Nana Liu

Methodology: Nana Liu, Wen Qin,, Chunshui Yu

Investigation: Nana Liu

Visualization: Nana Liu

Funding acquisition: Chunshui Yu, Wen Qin, Nana Liu, Jiayuan Xu, Jie Tang

Project administration: Chunshui Yu, Wen Qin, Meng Liang, Jiayuan Xu, Jie Tang

Supervision: Chunshui Yu, Wen Qin, Meng Liang

Writing – original draft: Nana Liu

Writing – review & editing: Chunshui Yu, Nana Liu

All authors critically reviewed the manuscript. Nana Liu, Jiayuan Xu, Jie Tang, Jilian Fu, Sijia Wang, Yuan Ji, Hui Xue, Nannan Zhang, Qiang Xu, Lining Guo, Hao Ding, Huaigui Liu, Feng Liu, Meng Liang, Wen Qin and Chunshui Yuacquired the data.

Competing interests

The authors have declared that no competing interests exist.

Data availability

The GWAS summary statistics for cortical thickness and surface area used in this work are publicly available from the ENIGMA consortium (http://enigma.usc.edu/research/download-enigma-gwas-results/). Raw genotype and neuroimaging data for the UKB and the ABCD study were accessed under authorized applications (application no. 75556 for UKBB; application no. 17607 for ABCD). The meta-analyses summary statistics for the 68 cortical traits will be made publicly available upon publication following peer review. Two human cortex single-nucleus RNA-seq datasets are publicly accessible (https://brain.listerlab.org/ and https://cells.ucsc.edu/?ds=pre-postnatal-cortex).

Code availability

The software used in this study is publicly available. The sMRI process and cortical parcellation were performed using FreeSurfer (https://surfer.nmr.mgh.harvard.edu) and harmonization was conducted using Combat harmonization (v.1.0.1) (https://github.com/Jfortin1/ComBatHarmonization). Genetic data processing involved PLINK (v.2.0) (http://zzz.bwh.harvard.edu/plink), SHAPEIT2 (v.2.r904) (https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html), IMPUTE2 (v.2.3.2) (http://mathgen.stats.ox.ac.uk/impute/impute_v2.html) and UCSC LiftOver tools (Jan.2022 release) (https://genome.ucsc.edu/cgi-bin/hgLiftOver). The R package ukbtools (v.0.11.3) (https://kenhanscombe.github.io/ukbtools) was used to estimate the relatedness of UKB participants. GWASs were performed using BGENIE (v.1.3) (https://jmarchini.org/bgenie), GCTA (v1.94.1) (https://yanglab.westlake.edu.cn/software/gcta/). Multivariate GWASs were performed using C-GWAS (v.0.9.3) (https://github.com/Fun-Gene/CGWAS). Meta-analyses were conducted using METAL (v.2011-03-25) (https://csg.sph.umich.edu/abecasis/Metal/). The post-GWAS analyses mainly involved LDSC (v.1.0.1) (https://github.com/bulik/ldsc), SuSiE (v.0.12.41) (https://github.com/stephenslab/susieR), FUMA (v.1.8.1) (https://fuma.ctglab.nl/), FLAMES (v.1.1.2) (https://github.com/Marijn-Schipper/FLAMES), WebGestalt (v.2024) (https://www.webgestalt.org/), OPERA (https://github.com/wuyangf7/OPERA), HyPrColoc (v.0.0.2) (https://jrs95.github.io/hyprcoloc/articles/hyprcoloc.html) and HDL (v.1.4.1) (https://github.com/zhenin/HDL). Analysis for RNA-seq datasets were used the published pipeline (https://zenodo.org/records/7113422).

Methods

Study populations

The study populations for genome-wide association studies (GWASs) on image-derived phenotypes (IDPs) of cortical thickness (CT) and surface area (SA) from brain magnetic resonance imaging (MRI) data were obtained from four independent datasets: the UK biobank (UKB) study[52], the Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) consortium[53], the Chinese Imaging Genetics (CHIMGEN) study[54], and the Adolescent Brain Cognitive Development (ABCD) study[55]. A pioneer study conducted GWAS meta-analyses on 70 bilateral averaged brain IDPs (35 for CT and 35 for SA) in 33,992 European ancestry (EUR) participants (23,909 from ENIGMA and 10,083 from UKB)[7]. These 70 cortical IDPs included total SA and mean CT for the cerebral cortex, along with CT and SA from 34 regions parcellated based on the Desikan-Killiany atlas using FreeSurfer[56]. From that study[7], we included the GWAS summary data of these IDPs, which were conducted exclusively on ENIGMA participants. For each ENIGMA site, genotype data was filtered by excluding variants with an imputation info score ≤ 0.5 and minor allele frequency (MAF) ≤ 0.005. These GWASs were conducted using an additive model, adjusting for age, age^2, sex, sex × age, sex × age^2, ancestry (the first four multidimensional scaling components), diagnostic status, and dummy variables for scanner. Subsequently, meta-analyses were conducted using METAL[57]. For the other three datasets, we tried to align with the ENIGMA-GWASs. As the CT and SA of the temporal pole have been deemed unreliable IDPs in the UKB study[6,58], we performed our GWASs solely on the remaining 68 cortical IDPs.

UKB

UKB is a prospective study involving over 500,000 individuals aged 40 to 69 years at recruitment from the United Kingdom[52]. It gathered numerous biological phenotypes, including brain IDPs from more than 60,000 participants[59]. The study was approved by the North West Multi-centre Research Ethics Committee, and written informed consent was obtained from each participant. We accessed the data under application NO. 75556. We utilized the imputed genomic data (version 3) from 487,409 individuals and retained 486,361 individuals after excluding participants according to the UKB pipeline[60]. The excluded participants comprised individuals with discrepancies between reported and chromosome-X determined sexes; those with sex chromosome aneuploidy; those displaying excessive heterozygosity or elevated missing rates; and those without a kinship inference.

All brain MRI data from UKB participants were processed using a standard pipeline[6], resulting in various IDPs. We used the 136 CT and SA IDPs that were symmetrically distributed across the left and right cerebral hemispheres. Of the 61,850 participants with both IDP and genetic data, we excluded 31 participants due to poor image quality, 173 participants with outliers (exceeding six times the median absolute deviation from the median) in any of cortical IDPs, and 1,035 related participants, retaining 60,611 unrelated individuals (White British: n = 52,278; non-White British: n = 8,333). We then computed

the 68 bilateral averaged cortical IDPs to align with those used in the ENIGMA-GWASs. Its plausibility was confirmed by the strong genetic correlations (mean = 0.83) of these IDPs from the bilateral homologous regions in the White British sample (Table S13).

The White British dataset served as the discovery cohort, whereas the non-White British dataset functioned as the replication cohort. Of the non-White British dataset, 6,220 individuals (75.49%) reported White ancestry, while the remainder included 679 Asian or Asian British, 417 Black or Black British, 344 from other ethnic groups, 316 with mixed ancestry, 182 with Chinese ancestry, and 175 with unknown ancestries. For each dataset, we implemented the following procedures: ComBat harmonization was applied to these IDPs to remove between-scanner variation while preserving biological variability[61]; normal score transformation was then applied to the harmonized data to enhance normality; and variants were filtered by only including those with minor allele frequency (MAF) > 0.005 and info > 0.5. The demographic and IDP data of included participants are presented in Table S14.

ABCD

The ABCD study collected a longitudinal dataset comprising nearly 12,000 participants aged 9-10 years at their baseline assessment from 21 sites[55]. Centralized institutional review board approval was obtained from the University of California San Diego, and each site also obtained local institutional review board approval. Parents or caregivers provided written informed consent, and children provided written assent. We accessed the data under application (ID 17607). After variant- and sample-level quality control procedures[62], this study provided imputed genotype data for 11,666 participants. After excluding those with sex mismatches, a total of 11,449 participants were retained. Of these, 11,319 participants also had data on 136 cortical IDPs. According to the inclusion criteria of ABCD, we excluded 142 participants due to poor image quality, 412 because of cortical parcellation failure, 56 outliers, and 1,632 with genetic relatedness, resulting in a total of 9,077 individuals with these 68 cortical IDPs.

From the raw genotyped data of 9,077 ABCD participants, we excluded variants with a call rate < 95%, MAF < 0.001, or located within genomic regions exhibiting long-range Linkage Disequilibrium (LD)[63,64], including the MHC region. We merged the remaining variants with 1000 Genomes data, and performed LD pruning using PLINK[65] to identify independent variants with a window of 1000 variants, a step size of 80 variants, and an $r^2 < 0.1$. Subsequently, we used PLINK to perform Principal Component Analysis (PCA) with default parameters, calculating the top 20 principal components (PCs) based on the 100,022 independent variants from the 9,077 participants. We utilized UMAP to identify clusters in 1000 Genomes data using the first four PCs. We identified seven major populations, including the non-Finnish Europeans, Finnish Europeans, Africans, Americans, East Asians, South Asian, and Bengali. Using the first four PCs from the ABCD dataset, we projected individuals onto the seven clusters, identifying broadly homogeneous populations (Fig. S6). We defined the 4,748 non-Finnish Europeans as the discovery dataset (ABCD EUR), whereas the remaining 4,329 participants were categorized as the replication dataset (ABCD multi-ancestry cohorts), comprising 374 Finnish Europeans, 1,788 Africans,

1,980 Americans, 114 East Asians, 68 South Asians, and 5 Bengalis. The liftover tool[66] in the UCSC Genome Browser was used to map variants to build 37 of the human reference genome. We performed ComBat harmonization, normal score transformation, and variant filtration (MAF > 0.005 and info > 0.5) for the discovery and replication datasets, respectively. The demographic and IDP data of the included participants are presented in Table S14.

CHIMGEN

The CHIMGEN study has collected both genomic and neuroimaging data from 7,306 healthy Chinese Han participants aged 18 to 30 years[54]. Participants were recruited from 32 centers, with MRI data acquired using 30 scanners. Informed consent was obtained from all participants, and centralized ethics approval was granted by Tianjin Medical University General Hospital, with additional local ethical approvals from each center.

Among the 7,306 participants, 7,195 were genotyped using the Illumina ASA-750K (Asian Screening Array). The genotype data were aligned to the human reference genome (GRCh37/hg19). In the sample-level quality control, we excluded participants with a missing rate > 3%, identity by descent (IBD) > 0.1875, PCA deviations from the Asian population, mismatches between reported and chromosome X determined sexes, or excess heterozygosity. In the variant-level quality control, we excluded variants with a call rate < 95%, MAF < 0.001, or those deviating from Hardy-Weinberg equilibrium (HWE; $P < 1 \times 10^{-6}$). After quality control, 7,163 individuals and 549,309 variants were retained for imputation. The variants were pre-phased using SHAPEIT2[67] and imputed by IMPUTE2[68], with the merged reference panel from the 1000 Genomes and SG10K[69] projects[70]. Among the 7,163 participants with qualified genetic data, we excluded 104 participants lacking qualified structural MRI data, 1 due to parcellation failure, and 27 outliers, resulting in 7,031 participants with these 68 cortical IDPs. We then performed ComBat harmonization, normal score transformation, and variant filtration (MAF > 0.005 and info > 0.5) for the dataset (Table S14).

GWAS of cortical IDPs

For the UKB White British, ABCD EUR, and CHIMGEN Han cohorts, we conducted GWAS using BGENIE v1.3[60,71] with an additive model to test the linear associations between genetic variants and cortical IDPs. For the UKB non-white British and ABCD multi-ancestry cohorts, we utilized mixed linear model[72] in GCTA[73] to perform GWASs, which prevents false-positive associations and enhances statistical power by accounting for population structure and relatedness. All GWASs were controlled for age, age^2, sex, sex × age and sex × age^2, PCs (40 for UKB, 10 for CHIMGEN and ABCD).

GWAS meta-analysis

In the discovery phase, EUR-GWAS meta-analyses of cortical IDPs were performed using summary statistics from three datasets: UKB White British, ABCD EUR, and ENIGMA. After aligning the genotype data from all datasets to the human reference genome (GRCh37/hg19), we excluded variants from ABCD EUR and ENIGMA that had allele frequency differences exceeding 0.5 or mismatched alleles when compared to UKB White British. We also removed variants with a total sample size < 10,000. We applied inverse variance weighted fixed-effects meta-analysis in METAL[57] to integrate GWAS summary statistics of cortical IDPs from 52,278 UKB White British participants and 4,748 ABCD EUR participants. We then used sample size weighted meta-analyses in METAL to combine GWAS summary statistics from 23,909 ENIGMA participants. In the replication phase, we also used sample size weighted meta-analyses to integrate GWAS summary data from CHIMGEN, UKB non-white British, and ABCD multi-ancestry populations.

C-GWAS

We used C-GWAS[23] to perform a multi-trait GWAS for 34 CT-IDPs and 34 SA-IDPs, respectively. C-GWAS integrates GWAS summary statistics from multiple correlated traits, enabling the effective detection of multi-trait effects across complex scenarios. As recommended[23], we defined statistical significance as $P < 5 \times 10^{-8}$.

Defining LD reference, genetic associations, and loci

We utilized the imputed genotypic data from 52,278 UKB White British participants to construct an EUR-LD reference. To identify independent lead variants associated with each IDP, we applied the PLINK clumping algorithm to select LD-independent variants ($r^2 < 0.1$ within 3-Mb windows, clump $P_1 = 7.35 \times 10^{-10}$) as lead variants. The method also identified all variants in LD ($r^2 > 0.1$ within 3-Mb windows, clump $P_2 = 5 \times 10^{-8}$) with each lead variant, forming a clump. Then, we extended a 250-kb window on both sides of each clump and merged overlapping clumps into a single locus. The association between each independent variant and the trait was defined as a variant-trait association, while all non-overlapping loci linked to the trait were defined as trait-associated loci. We performed PLINK clumping (LD $r^2 > 0.1$) for all variants associated with CT-IDPs and SA-IDPs, resulting in independent association signals for CT and SA, respectively. After merging overlapping loci, the remaining loci were defined as independent CT- and SA-associated loci. For CT-IDPs or SA-IDPs, a locus was considered novel if all lead variants within the locus were located more than 500 kb away from and not in LD ($r^2 < 0.1$) with any previously reported lead variants associated with CT-IDPs or SA-IDPs. Previously reported lead variants were extracted from GWASs on CT-IDPs and SA-IDPs[6-16], as well as from the GWAS Catalog[74].

Heritability and GWAS assessment

For each of the five sets of GWASs (UKB White British, ABCD EUR, ENIGMA, meta-analysis of UKB White British and ABCD EUR, and the final meta-analysis) in the discovery phase, we used LD score regression (LDSC)[75] with LD scores derived from the EUR reference panel of the 1000 Genomes phase 3 to estimate the LDSC intercept for each GWAS. The intercept can distinguish polygenicity from confounding biases[76]. We estimated the SNP-based heritability of each IDP based on the final meta-analysis. We conducted these analyses after excluding the MHC region (chr6:25-35 Mb) due to its extreme LD.

Statistical fine mapping

Statistical fine mapping was performed using the Sum of the Single Effects framework (SuSiE) model[25]. We performed fine-mapping for trait-associated loci based on EUR-GWASs and EUR-LD reference after excluding the MHC region. We executed SuSiE using z-scores as inputs and allowed a maximum of ten signals per locus. For each signal, we calculated its 95% credible set, representing the minimum number of variants whose posterior inclusion probability (PIP) for the signal summed to ≥ 0.95.

Functional annotation

FUMA[29] was used to perform functional annotations for the variants extracted from the 95% credible sets. FUMA is an online platform that annotates and prioritizes genetic variants by integrating comprehensive annotation data sources. Variants were annotated using ANNOVAR[30] and combined annotation-dependent depletion (CADD) scores[31]. ANNOVAR annotated each SNP according to genic position, while CADD prioritized pathogenic variants using scores > 12.37[77]. Gene prioritization was performed using the framework[50]. FLAMES leverages machine learning predictions based on biological data and integrates them with convergence evidence to interpret fine-mapped GWAS signals. The resulting genes were included in a statistical over-representation analysis (P_c < 0.05, Bonferroni corrected, overlap gene number > 10) using the WebGestalt[51] online tool, based on Gene Ontology (GO) terms of biological processes.

Multiple omics analysis

The Omics PlEiotRopic Association (OPERA)[40] can integrate GWAS summary data of studied traits (cortical IDPs) with multi-omics molecular quantitative trait loci (xQTLs) data to identify shared causal variants between molecular and studied traits, providing insights into the molecular mechanisms associated with studied traits. OPERA is an extension of Summary-based Mendelian Randomization (SMR)[78] that can effectively control the false discovery rate while maintaining high detection power for association patterns. Using the estimated prior probabilities across the genome, OPERA computes a marginal posterior probability of association (PPA) for each trait pair and a joint PPA for combinations of multiple traits. PPA > 0.9 indicates pleiotropic associations between each studied trait and one or more

molecular traits through shared causal variants[40]. To exclude false associations arising from LD in SNPs from xQTLs and GWAS or those from different xQTLs, a multi-trait HEIDI test was performed on associations with PPA > 0.9, applying P_{HEIDI} > 0.01 to identify true pleiotropic associations[40]. xQTLs datasets comprised five molecular traits from the human brain tissues, including cis-eQTL data from cortical samples of 2,443 individuals[44]; mQTLs from a meta-analysis of 1,160 brain samples[41]; caQTLs from 272 adult prefrontal samples[43]; pQTLs from 1,277 cortical proteomes[45]; and haQTLs from 561 individuals in the ROSMAP study[42].

Correlation between cell type-specific gene expression and cortical growth rate

For each CT- or SA-related gene prioritized using FLAMES, we extracted its cell type-specific expression data from two snRNA-seq datasets of the human cerebral cortex. One dataset for discovery was obtained from 27 donors, including more than 150,000 snRNA-seq profiles across cell types and developmental stages[19]. We included 17 major cell clusters, comprising five (L2/3-CUX2, L4-RORB, L5/6-THEMIS, L5/6-TLE4, and PN dev) for excitatory neurons, eight (VIP, ID2, LAMP5-NOS1, GGE dev, SST, PV, PV-SCUBE3, and MGE dev) for inhibitory neurons, and four (astrocytes, microglia, oligodendrocytes, and oligodendrocyte precursor cells) for glial cells. The data were obtained from five developmental stages: fetal (22nd gestation week to birth), neonatal (first month after birth), infancy (from 2nd month to 1 year), childhood (1-10 years), adolescence (10-20 years), and adult (20 years). Another dataset for replication was obtained from 106 donors with over 700,000 nuclei[20], of which 79 donors and 554,624 snRNA-seq profiles were retained after excluding donors overlapping with dataset 1. The data were sampled from multiple cortical regions across an age range spanning from 14 postconceptional weeks to 54 years. The major cell types included: excitatory neurons (progenitors, L2-3, L4, L5, L5-6-IT, L6), inhibitory neurons (progenitors, VIP, CALB2, CCK, NOS, RELN, SV2C, PV, PV-MP, SST, SST-RELN), and glial cells (astrocytes, oligodendrocytes, OPCs and microglia). Detailed preprocessing of the two snRNA-seq datasets has been described previously[19,20].

To characterize cell type-specific gene expression curves during development, we created pseudo-bulk counts by aggregating nuclei (minimum 10) grouped by batches, cell types, and developmental stages. The data were filtered with edgeR[79] (filterByExpr function) to remove low-expression genes, normalized by the trimmed mean of M-values (TMM)[80], and scaled to $\log_2$CPM (counts per million). Given that the replication dataset had undergone normalization and log-transformation, we aggregated snRNA-seq data into pseudo-bulk samples Based on the mean of log-transformed expression values for each gene within specific cell types and developmental stages. We used an inverse hyperbolic sine (arcsinh) transformation for age to linearize the developmental time. For each gene, we performed gene expression trend analysis using generalized additive models (GAMs) in each cell type. The pyGAM Python package was used to model continuous expression dynamics from pseudo-bulked data, with developmental trends captured by computing fitted gene expression values at arcsinh-transformed age

spanning the entire developmental timeline. From the arcsinh-transformed age values, 100 evenly spaced time points were extracted. Their corresponding gene expression values at these time points were then retrieved, and the time points were converted back to the actual ages. We extracted the CT and SA growth rates at the corresponding time points, which were calculated based on over 100,000 MRI scans[4]. The first time point lacked a growth rate since derivatives require previous data points, leaving 99 values for cross-correlation analysis. Based on the trajectories of cell-type gene expression and cortical growth rates, we computed cross-correlations across the 99 time points. We calculated cross-correlations for all possible time lags using SciPy in python. In the replication stage, gene expression corresponding to the same time points of the discovery dataset were extracted for cross-correlations.

Colocalization analysis

We utilized HyPrColoc (Hypothesis Prioritisation for multi-trait Colocalization)[81] to identify colocalization between cortical traits and cognitive traits and neuropsychiatric disorders. The method enables multi-trait colocalization and can yield reliable results even in scenario where there is sample overlap between traits. HyPrColoc generates the posterior probability of full colocalization (PPFC), which represents the likelihood that all traits share a common causal variant. HyPrColoc also indicates the potential causal variant for colocalization, accounting for the largest proportion of PPFC. We used PPFC ≥ 0.8 as evidence of colocalization. We included EUR-GWASs for three cognitive traits (years of schooling[82], cognitive performance[82], and intelligence[83]) and eight disorders (attention-deficit/hyperactivity disorder[84], autism spectrum disorder[85], obsessive compulsive symptoms[86], depression[87], bipolar disorder[88], schizophrenia[89], Alzheimer's disease[90], and Parkinson's disease[91]).

Genetic correlation

Genetic correlations were computed within 68 cortical IDPs, and between cortical IDPs and cognitive/neuropsychiatric disorders using the high-definition likelihood (HDL)[92] method, based on EUR-GWAS summary statistics. HDL can enhance the precision of genetic correlation estimation and reduce their variance. We used the reference panel derived from imputed UKB genotype data in EUR populations that was available at the HDL website, restricting to HapMap SNPs. The MHC region was removed from these analyses.

References

1. Rakic P (2009) Evolution of the neocortex: a perspective from developmental biology. Nat Rev Neurosci 10:724–735. 10.1038/nrn2719
2. Rakic P (1988) Specification of cerebral cortical areas. Science 241:170–176. 10.1126/science.3291116

3. Lindhout FW, Krienen FM, Pollard KS, Lancaster MA (2024) A molecular and cellular perspective on human brain evolution and tempo. Nature 630:596–608. 10.1038/s41586-024-07521-x
4. Bethlehem RAI et al (2022) Brain charts for the human lifespan. Nature 604:525–533. 10.1038/s41586-022-04554-y
5. Panizzon MS et al (2009) Distinct genetic influences on cortical surface area and cortical thickness. Cereb Cortex 19:2728–2735. 10.1093/cercor/bhp026
6. Smith SM et al (2021) An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. Nat Neurosci 24:737–745. 10.1038/s41593-021-00826-4
7. Grasby KL et al (2020) The genetic architecture of the human cerebral cortex. Science 367. 10.1126/science.aay6690
8. Makowski C et al (2022) Discovery of genomic loci of the human cerebral cortex using genetically informed brain atlases. Science 375:522–528. 10.1126/science.abe8457
9. Warrier V et al (2023) Genetic insights into human cortical organization and development through genome-wide analyses of 2,347 neuroimaging phenotypes. Nat Genet 55:1483–1493. 10.1038/s41588-023-01475-y
10. van der Meer D et al (2021) The genetic architecture of human cortical folding. Sci Adv 7:eabj9446. 10.1126/sciadv.abj9446
11. van der Meer D et al (2020) Understanding the genetic determinants of the brain with MOSTest. Nat Commun 11:3512. 10.1038/s41467-020-17368-1
12. Hofer E et al (2020) Genetic correlations and genome-wide associations of cortical structure in general population samples of 22,824 adults. Nat Commun 11:4796. 10.1038/s41467-020-18367-y
13. Shin J et al (2020) Global and Regional Development of the Human Cerebral Cortex: Molecular Architecture and Occupational Aptitudes. Cereb Cortex 30:4121–4139. 10.1093/cercor/bhaa035
14. Makowski C et al (2023) Larger cerebral cortex is genetically correlated with greater frontal area and dorsal thickness. Proc Natl Acad Sci U S A 120:e2214834120. 10.1073/pnas.2214834120
15. Zhukovsky P et al (2024) Genetic influences on brain and cognitive health and their interactions with cardiovascular conditions and depression. Nat Commun 15:5207. 10.1038/s41467-024-49430-7
16. Fu J et al (2024) Cross-ancestry genome-wide association studies of brain imaging phenotypes. Nat Genet 56:1110–1120. 10.1038/s41588-024-01766-y
17. Matsumoto J et al (2023) Cerebral cortical structural alteration patterns across four major psychiatric disorders in 5549 individuals. Mol Psychiatry 28:4915–4923. 10.1038/s41380-023-02224-7
18. Koechlin E, Basso G, Pietrini P, Panzer S, Grafman J (1999) The role of the anterior prefrontal cortex in human cognition. Nature 399:148–151. 10.1038/20178
19. Herring CA et al (2022) Human prefrontal cortex gene regulatory dynamics from gestation to adulthood at single-cell resolution. *Cell* 185, 4428–4447 e4428. 10.1016/j.cell.2022.09.039

20. Velmeshev D et al (2023) Single-cell analysis of prenatal and postnatal human cortical development. Science 382:eadf0834. 10.1126/science.adf0834
21. Jeffries AM et al (2025) Single-cell transcriptomic and genomic changes in the ageing human brain. Nature 646:657–666. 10.1038/s41586-025-09435-8
22. Desikan RS et al (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 31:968–980. 10.1016/j.neuroimage.2006.01.021
23. Xiong Z et al (2022) Combining genome-wide association studies highlight novel loci involved in human facial variation. Nat Commun 13:7832. 10.1038/s41467-022-35328-9
24. Sydnor VJ et al (2021) Neurodevelopment of the association cortices: Patterns, mechanisms, and implications for psychopathology. Neuron 109:2820–2846. 10.1016/j.neuron.2021.06.016
25. Zou Y, Carbonetto P, Wang G, Stephens M (2022) Fine-mapping from summary data with the Sum of Single Effects model. PLoS Genet 18:e1010299. 10.1371/journal.pgen.1010299
26. Grinberg I et al (2004) Heterozygous deletion of the linked genes ZIC1 and ZIC4 is involved in Dandy-Walker malformation. Nat Genet 36:1053–1055. 10.1038/ng1420
27. Park JH et al (2015) SLC39A8 Deficiency: A Disorder of Manganese Transport and Glycosylation. Am J Hum Genet 97:894–903. 10.1016/j.ajhg.2015.11.003
28. Allen M et al (2016) Gene expression, methylation and neuropathology correlations at progressive supranuclear palsy risk loci. Acta Neuropathol 132:197–211. 10.1007/s00401-016-1576-7
29. Watanabe K, Taskesen E, van Bochoven A, Posthuma D (2017) Functional mapping and annotation of genetic associations with FUMA. Nat Commun 8:1826. 10.1038/s41467-017-01261-5
30. Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38:e164. 10.1093/nar/gkq603
31. Kircher M et al (2014) A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 46:310–315. 10.1038/ng.2892
32. Ueland PM, Hustad S, Schneede J, Refsum H, Vollset SE (2001) Biological and clinical implications of the MTHFR C677T polymorphism. Trends Pharmacol Sci 22:195–201. 10.1016/s0165-6147(00)01675-8
33. Greenman R et al (2015) Non-cell autonomous and non-catalytic activities of ATX in the developing brain. Front NeuroSci 9. 10.3389/fnins.2015.00053
34. Muley PD et al (2008) The atRA-responsive gene neuron navigator 2 functions in neurite outgrowth and axonal elongation. Dev Neurobiol 68:1441–1453. 10.1002/dneu.20670
35. Pruitt SC, Freeland A, Rusiniak ME, Kunnev D, Cady GK (2013) Cdkn1b overexpression in adult mice alters the balance between genome and tissue ageing. Nat Commun 4:2626. 10.1038/ncomms3626
36. Nguyen L, Besson A, Roberts JM, Guillemot F (2006) Coupling cell cycle exit, neuronal differentiation and migration in cortical neurogenesis. Cell Cycle 5:2314–2318. 10.4161/cc.5.20.3381

37. Goto T, Mitsuhashi T, Takahashi T (2004) Altered patterns of neuron production in the p27 knockout mouse. Dev Neurosci 26:208–217. 10.1159/000082138
38. Yan F et al (2020) FoxO3a suppresses neuropeptide W expression in neuronal cells and in rat hypothalamus and its implication in hypothalamic-pituitary-adrenal (HPA) axis. Int J Biol Sci 16:2775–2787. 10.7150/ijbs.45619
39. Fernandez V, Borrell V (2024) Keep calm and make neurons: The effects of glucocorticoids on human cortical neurogenesis. Neuron 112:1373–1375. 10.1016/j.neuron.2024.04.004
40. Wu Y et al (2023) Joint analysis of GWAS and multi-omics QTL summary statistics reveals a large fraction of GWAS signals shared with molecular phenotypes. Cell Genom 3:100344. 10.1016/j.xgen.2023.100344
41. Qi T et al (2018) Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. Nat Commun 9:2282. 10.1038/s41467-018-04558-1
42. Ng B et al (2017) An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. Nat Neurosci 20:1418–1426. 10.1038/nn.4632
43. Bryois J et al (2018) Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. Nat Commun 9:3121. 10.1038/s41467-018-05379-y
44. Qi T et al (2022) Genetic control of RNA splicing and its distinct role in complex trait variation. Nat Genet 54:1355–1363. 10.1038/s41588-022-01154-4
45. Wingo AP et al (2023) Sex differences in brain protein expression and disease. Nat Med 29:2224–2232. 10.1038/s41591-023-02509-y
46. Stelzer G et al (2016) The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr Protoc Bioinf 54(30):31–31. 10.1002/cpbi.5
47. Ramonet D et al (2012) PARK9-associated ATP13A2 localizes to intracellular acidic vesicles and regulates cation homeostasis and neuronal integrity. Hum Mol Genet 21:1725–1743. 10.1093/hmg/ddr606
48. Fujii T et al (2023) Parkinson's disease-associated ATP13A2/PARK9 functions as a lysosomal H(+),K(+)-ATPase. Nat Commun 14:2174. 10.1038/s41467-023-37815-z
49. Li T et al (2022) Kansl1 haploinsufficiency impairs autophagosome-lysosome fusion and links autophagic dysfunction with Koolen-de Vries syndrome in mice. Nat Commun 13:931. 10.1038/s41467-022-28613-0
50. Schipper M et al (2025) Prioritizing effector genes at trait-associated loci using multimodal evidence. Nat Genet 57:323–333. 10.1038/s41588-025-02084-7
51. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B (2019) WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res 47:W199–W205. 10.1093/nar/gkz401
52. Sudlow C et al (2015) 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. 10.1371/journal.pmed.1001779

53. Thompson PM et al (2020) ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. Transl Psychiatry 10:100. 10.1038/s41398-020-0705-1
54. Xu Q et al (2020) CHIMGEN: a Chinese imaging genetics cohort to enhance cross-ethnic and cross-geographic brain research. Mol Psychiatry 25:517–529. 10.1038/s41380-019-0627-6
55. Casey BJ et al (2018) The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. Dev Cogn Neurosci 32:43–54. 10.1016/j.dcn.2018.03.001
56. Fischl B, FreeSurfer (2012) NeuroImage 62:774–781. 10.1016/j.neuroimage.2012.01.021
57. Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26:2190–2191. 10.1093/bioinformatics/btq340
58. Alfaro-Almagro F et al (2018) Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. NeuroImage 166:400–424. 10.1016/j.neuroimage.2017.10.034
59. Miller KL et al (2016) Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Nat Neurosci 19:1523–1536. 10.1038/nn.4393
60. Bycroft C et al (2018) The UK Biobank resource with deep phenotyping and genomic data. Nature 562:203–209. 10.1038/s41586-018-0579-z
61. Fortin JP et al (2017) Harmonization of multi-site diffusion tensor imaging data. NeuroImage 161:149–170. 10.1016/j.neuroimage.2017.08.047
62. Fan CC, Loughnan R, Wilson S, Hewitt JK, Group AG (2023) W. Genotype Data and Derived Genetic Instruments of Adolescent Brain Cognitive Development Study((R)) for Better Understanding of Human Brain Development. Behav Genet 53:159–168. 10.1007/s10519-023-10143-0
63. Price AL et al (2008) Long-range LD can confound genome scans in admixed populations. *Am J Hum Genet* 83, 132–135; author reply 135–139. 10.1016/j.ajhg.2008.06.005
64. Anderson CA et al (2010) Data quality control in genetic case-control association studies. Nat Protoc 5:1564–1573. 10.1038/nprot.2010.116
65. Purcell S et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575. 10.1086/519795
66. Hinrichs AS (2006) The UCSC Genome Browser Database: update 2006. Nucleic Acids Res 34:D590–D598. 10.1093/nar/gkj144
67. Delaneau O, Zagury JF, Marchini J (2013) Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods 10:5–6. 10.1038/nmeth.2307
68. Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 5:e1000529. 10.1371/journal.pgen.1000529
69. Wu D et al (2019) Large-Scale Whole-Genome Sequencing of Three Diverse Asian Populations in Singapore. Cell 179:736–749e715. 10.1016/j.cell.2019.09.019

70. Liu N et al (2023) Cross-ancestry genome-wide association meta-analyses of hippocampal and subfield volumes. Nat Genet 55:1126–1137. 10.1038/s41588-023-01425-8
71. Elliott LT et al (2018) Genome-wide association studies of brain imaging phenotypes in UK Biobank. Nature 562:210–216. 10.1038/s41586-018-0571-7
72. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL (2014) Advantages and pitfalls in the application of mixed-model association methods. Nat Genet 46:100–106. 10.1038/ng.2876
73. Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88:76–82
74. Buniello A et al (2019) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res 47:D1005–D1012. 10.1093/nar/gky1120
75. Bulik-Sullivan B et al (2015) An atlas of genetic correlations across human diseases and traits. Nat Genet 47:1236–1241. 10.1038/ng.3406
76. Bulik-Sullivan BK et al (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 47:291–295. 10.1038/ng.3211
77. Amendola LM et al (2015) Actionable exomic incidental findings in 6503 participants: challenges of variant classification. Genome Res 25:305–315. 10.1101/gr.183483.114
78. Zhu Z et al (2016) Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet 48:481–487. 10.1038/ng.3538
79. Chen Y, Chen L, Lun ATL, Baldoni PL, Smyth GK (2025) edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. Nucleic Acids Res 53. 10.1093/nar/gkaf018
80. Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 11, R25. 10.1186/gb-2010-11-3-r25
81. Foley CN et al (2021) A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. Nat Commun 12:764. 10.1038/s41467-020-20885-8
82. Lee JJ et al (2018) Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet 50:1112–1121. 10.1038/s41588-018-0147-3
83. Savage JE et al (2018) Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nat Genet 50:912–919. 10.1038/s41588-018-0152-6
84. Demontis D et al (2023) Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. Nat Genet 55:198–208. 10.1038/s41588-022-01285-8
85. Grove J et al (2019) Identification of common genetic risk variants for autism spectrum disorder. Nat Genet 51:431–444. 10.1038/s41588-019-0344-8

86. Strom NI et al (2024) Genome-Wide Association Study of Obsessive-Compulsive Symptoms including 33,943 individuals from the general population. Mol Psychiatry 29:2714–2723. 10.1038/s41380-024-02489-6
87. Als TD et al (2023) Depression pathophysiology, risk prediction of recurrence and comorbid psychiatric disorders using genome-wide analyses. Nat Med 29:1832–1844. 10.1038/s41591-023-02352-1
88. Mullins N et al (2021) Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nat Genet 53:817–829. 10.1038/s41588-021-00857-4
89. Trubetskoy V et al (2022) Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature 604:502–508. 10.1038/s41586-022-04434-5
90. Kunkle BW et al (2019) Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. Nat Genet 51:414–430. 10.1038/s41588-019-0358-2
91. Nalls MA et al (2019) Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol 18:1091–1102. 10.1016/S1474-4422(19)30320-5
92. Ning Z, Pawitan Y, Shen X (2020) High-definition likelihood inference of genetic correlations across human complex traits. Nat Genet 52:859–864. 10.1038/s41588-020-0653-y

Figures

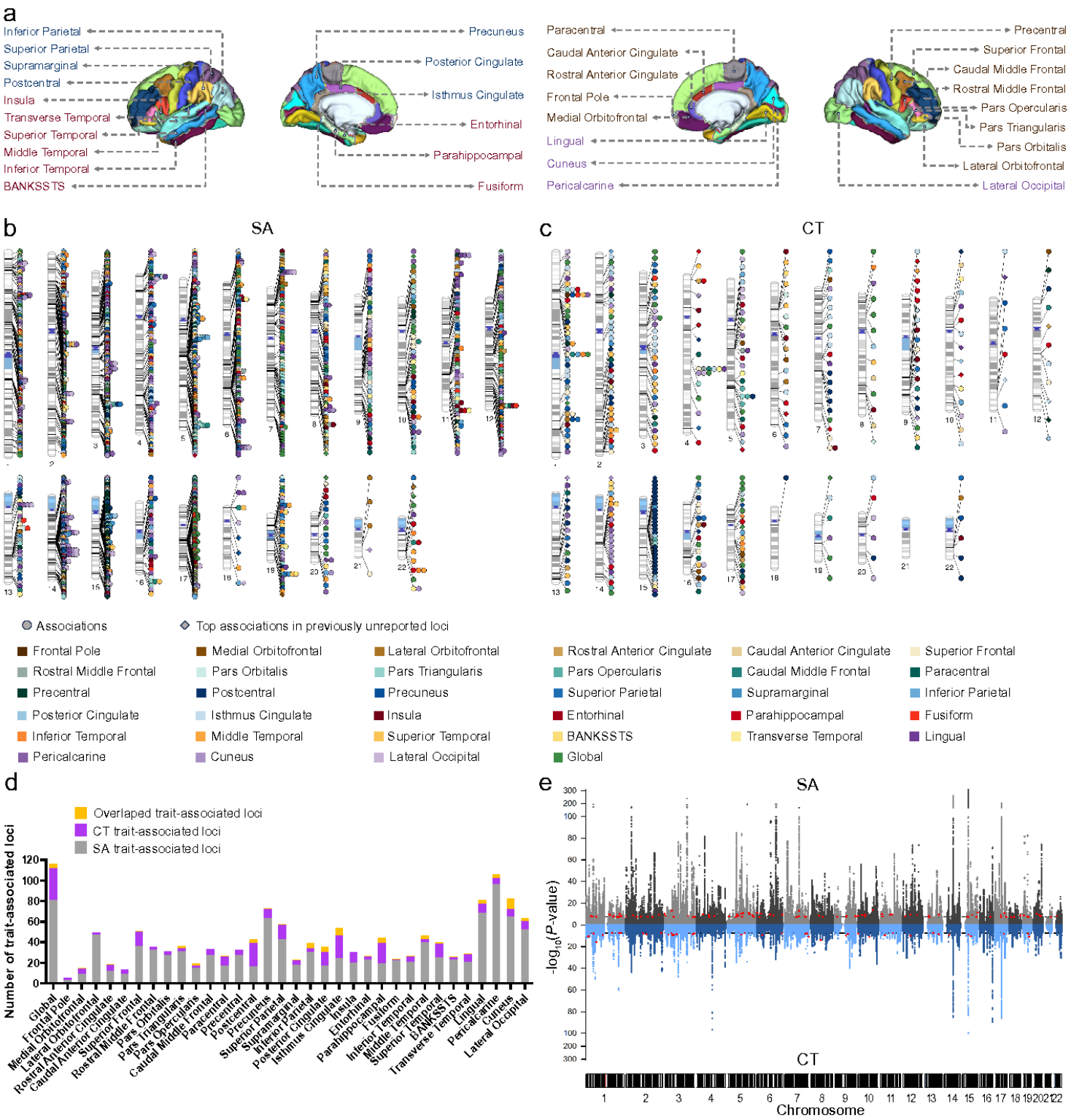


Figure 1

GWAS meta-analyses of CT and SA. a, The spatial distribution of the 33 cortical regions defined by the Desikan-Killiany atlas (excluding the temporal pole). **b-c**, Ideograms show genomic locations of lead variants of the 1,606 variant-trait associations for SA (**b**) and 422 associations for CT (**c**) (two-sided $P < 7.35 \times 10^{-10}$, Bonferroni corrected). **d,** Bar plot displays the number of trait-associated loci for CT, SA, or both. **e,** Manhattan plots show the -$\log_{10}$ (*P*-value) of each variant from multi-trait GWAS for SA (upper

panel) and CT (lower panel). The x-axis indicates genomic locations grouped by chromosome, and black lines mark the genome-wide significance threshold of 5×10^{-8}. Red dots represent lead variants in new loci identified by multi-trait GWAS.

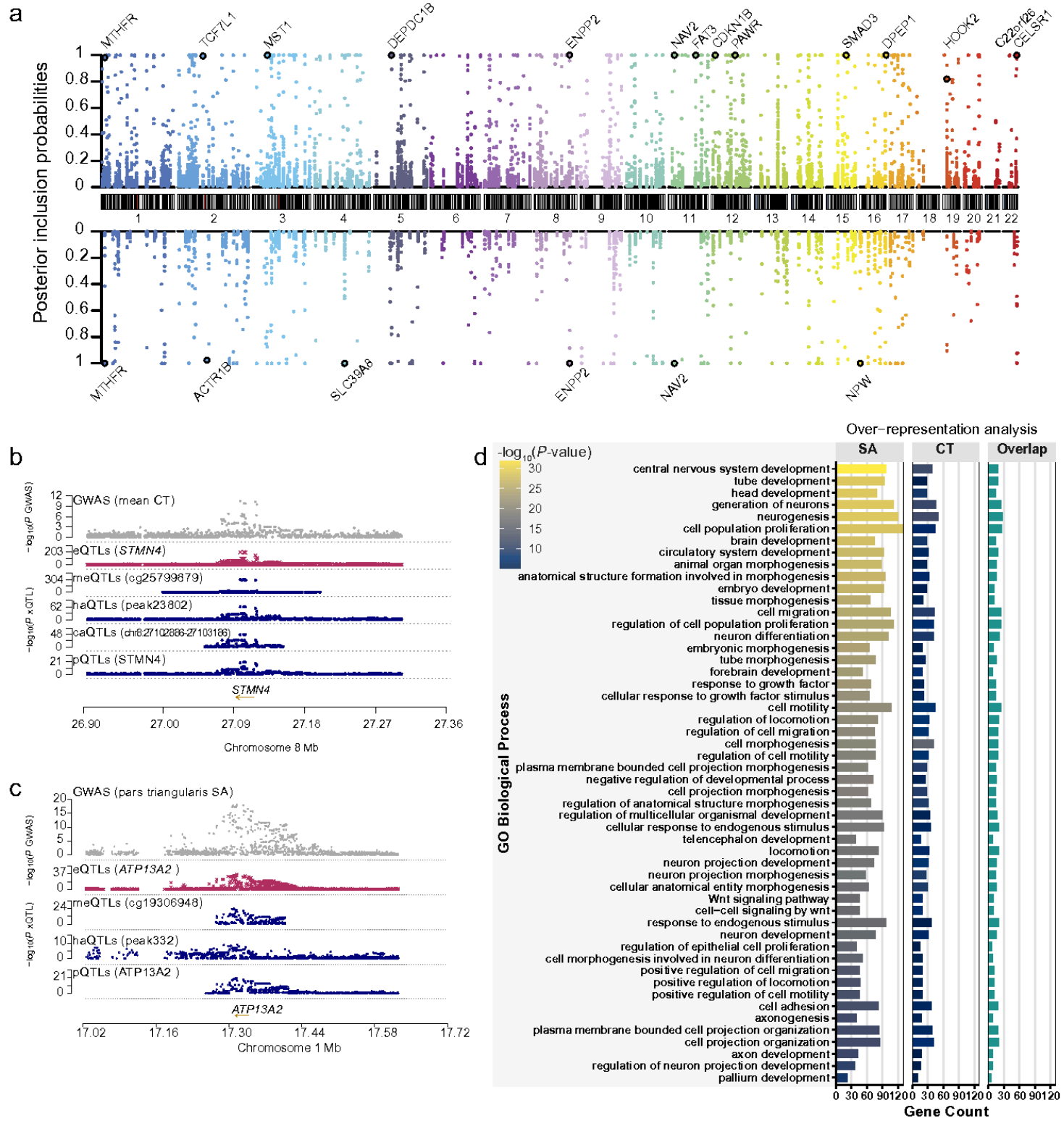


Figure 2

Fine mapping and functional annotation of genetic variants associated with CT and SA. a, Genomic distribution of posterior inclusion probabilities for fine mapped variants (95% credible sets) of SA (upper panel) and CT (lower panel). Genes corresponding to missense variants with PIP > 0.8 are highlighted. **b-c,** Pleiotropic associations between molecular phenotypes and cortical traits at the loci containing the *STMN4* (**b**) and *ATP13A2* (**c**) genes. The top track shows the -$\log_{10}$(*P*-values) of genetic associations with cortical traits. The other tracks show -$\log_{10}$(*P*-values) of genetic associations with molecular phenotypes derived from their corresponding multi-omics datasets. **d,** The bar plots display the enriched gene ontology (GO) terms (P_c < 0.05, Bonferroni-corrected) shared by genes associated with SA and CT, as well as the number of overlapping genes per term. The x-axis represents the number of genes for each term, while the y-axis lists the terms. Bar colors indicate the significance level for each term in the enrichment analysis.

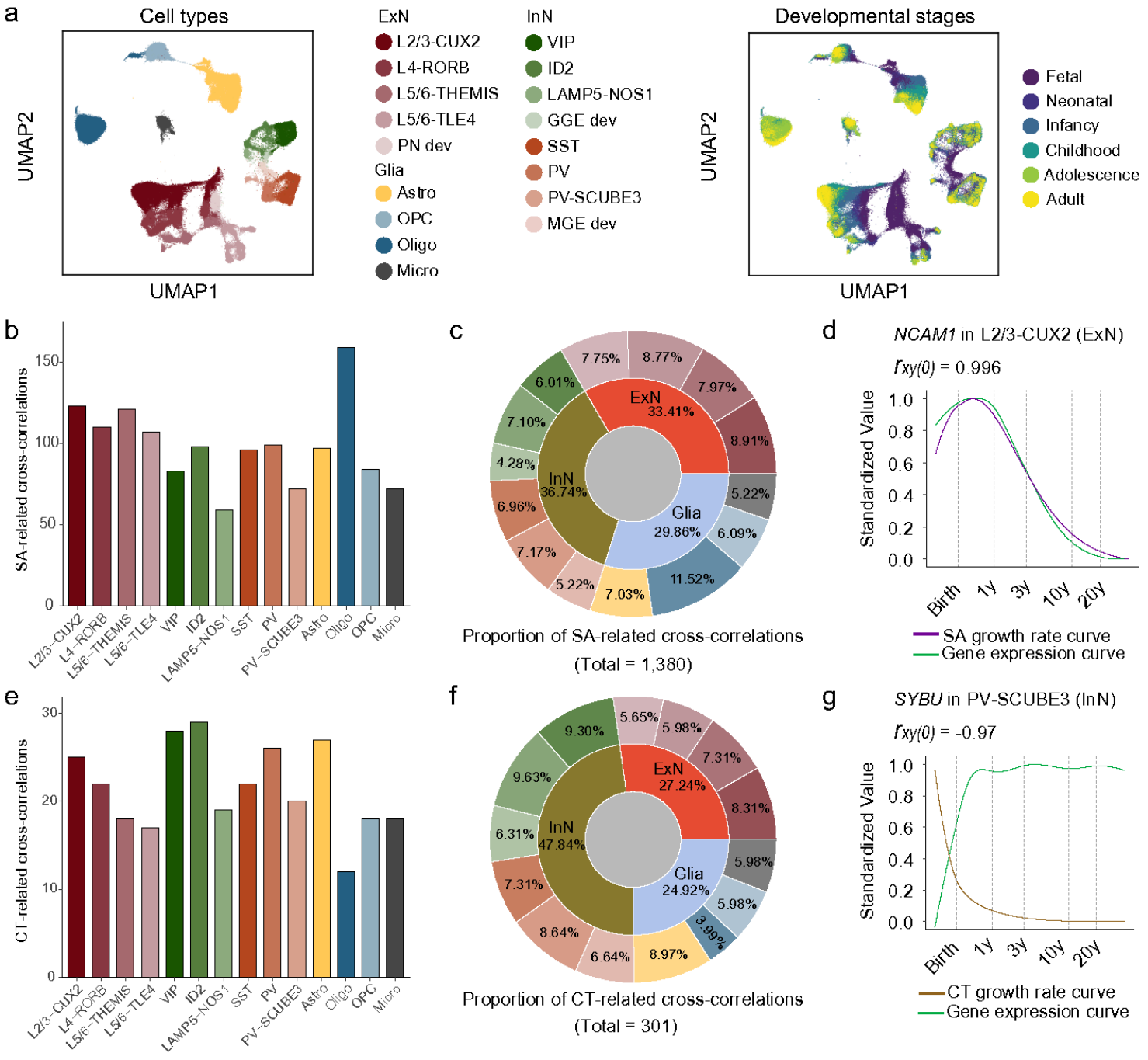


Figure 3

Cross-correlations between cell type-specific gene expression curves and SA and CT growth-rate curves during human cortical development. a, UMAP (uniform manifold approximation and projection) visualization of cortical cells, with cluster labels indicating annotated cell types (left panel) and developmental stages (right panel). **b-c,** 1,380 cell type-gene pairs associated with SA growth rate. Bar plot (**b**) shows the counts of SA cross-correlations ($|r_{xy}(k)| > 0.8$) for 14 cell types in the discovery dataset. Sunburst chart (**c**) showing the proportion of SA cross-correlations across major cell types (inner ring) and their subtypes (outer ring). **d,** Example developmental expression trajectories of *ROBO2* showing correlation with SA growth rate in L2/3-CUX2 excitatory neurons. **e-f,** 301 cell type-gene pairs

associated with CT growth rate. Bar plot (**e**) shows the counts of CT cross-correlations ($|r_{xy}(k)| > 0.8$) for 14 cell types in the discovery dataset. Sunburst chart (**f**) showing the proportion of SA cross-correlations across major cell types (inner ring) and their subtypes (outer ring). **g,** Example developmental expression trajectories of *SYBU* showing correlation with CT growth rate in PV-SCUBE3 inhibitory neurons.

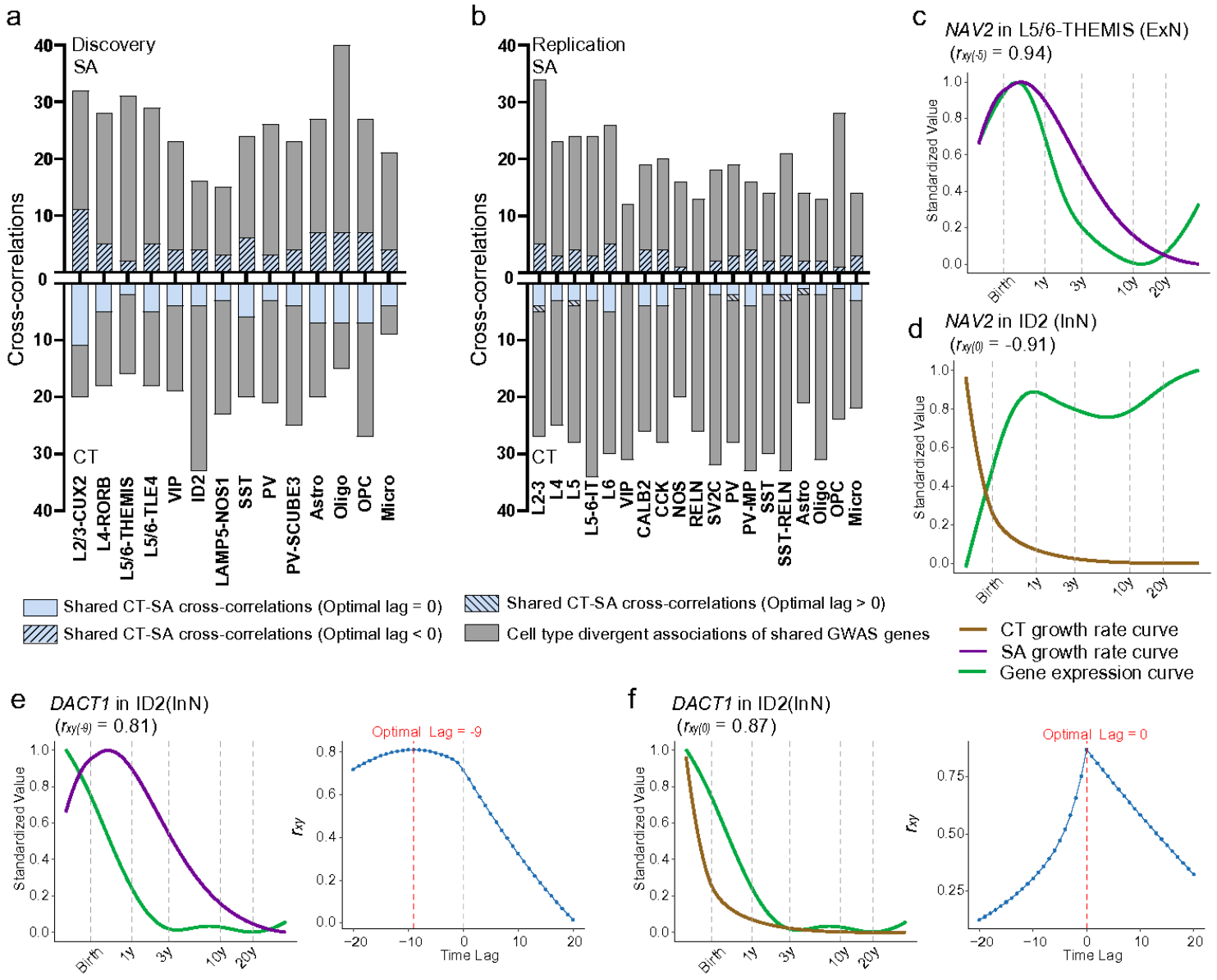


Figure 4

Distinct cross-correlations of cell type-specific expression curves of CT-SA shared genes with CT and SA growth-rate curves. a-b, Bar plot shows the number (y-axis) of significant cross-correlations ($|r_{xy}(k)| > 0.8$) of CT-SA shared gene expression curves with SA (upper) and CT (lower) growth-rate curves in discovery (**a**) and replication (**b**) datasets, stratified by cell types (x-axis). **c-d,** An example showing that SA (**c**) and CT (**d**) growth-rates are influenced by the same gene expression in different cell types. **e-f,** An example showing that SA (**e**) and CT (**f**) growth-rates are influenced by the same gene expression in the same cell type with distinct temporal lags. In **e-f**, the left panel shows the cell type-specific gene

expression curve and the SA or CT growth-rate curve from fetal to adult stages, while the right panel shows cross-correlations (20-time units) between the two curves. Red dotted lines mark peak correlations and corresponding lags.

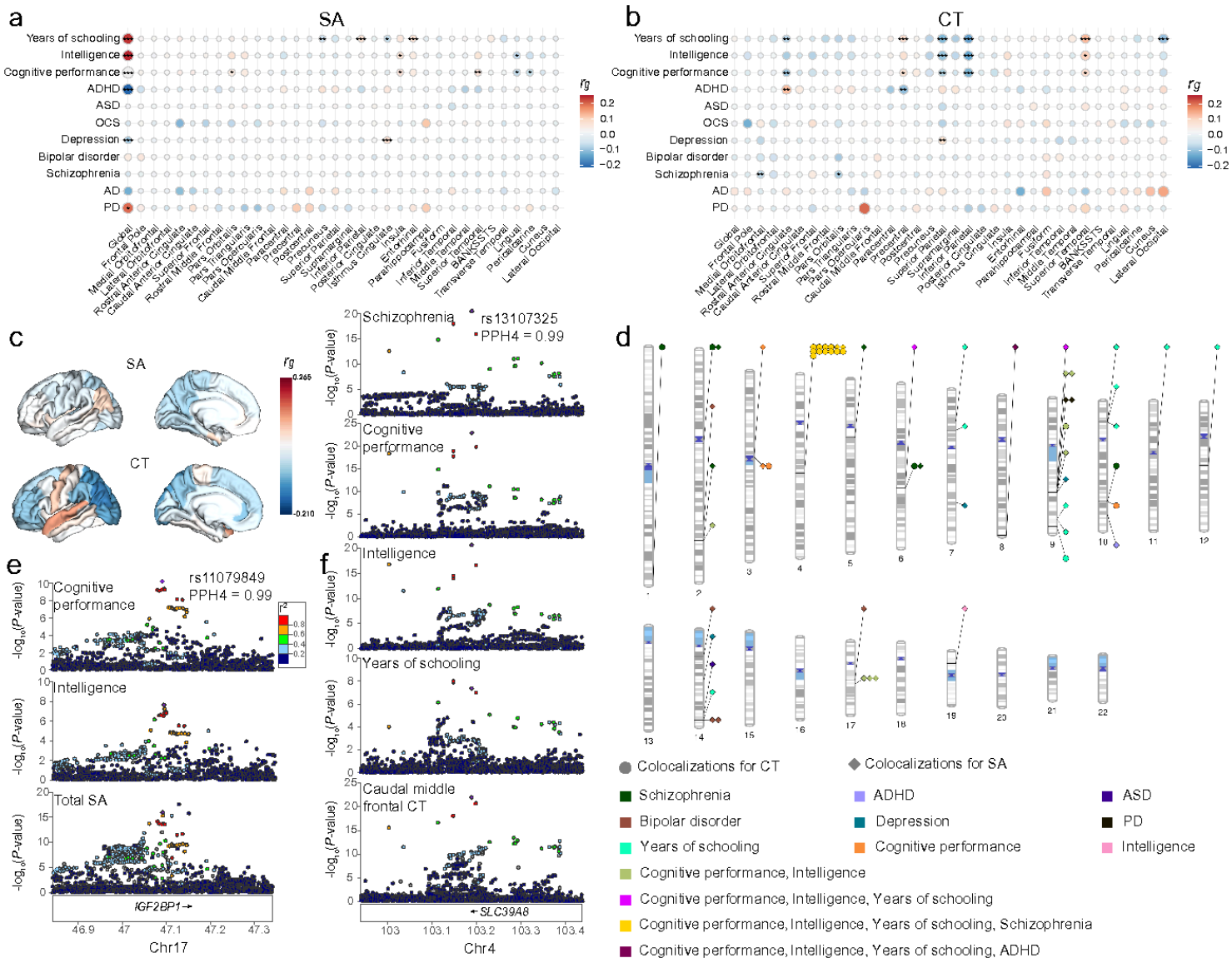


Figure 5

Genetic correlations and colocalizations between cortical traits and brain-related traits. a-b, Heatmaps show genetic correlations of SA (**a**) and CT (**b**) traits with three cognitive traits and eight neuropsychiatric disorders. Circle color indicates the sign and size of genetic correlation, while its size reflects the absolute r_g value. Bonferroni-corrected significance: *P_c < 0.05, **P_c < 0.01, ***P_c < 0.001. **c,** Brain chart shows genetic correlations of regional SA and CT with years of schooling. **d,** Ideograms illustrate colocalizations between cortical traits and brain-related traits. **e-f,** Regional plots display two examples of colocalizations of SA (**e**) and CT (**f**) with brain-related traits.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTables.xlsx
- SupplementaryFigures.docx