nature portfolio

Corresponding author(s): Shuyu Li
Yong He

Last updated by author(s): Mar 11, 2026

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: MRI acquisition: Siemens Prisma 3T scanner with standard MPRAGE sequence

Data analysis	Preprocessing: - HCP pipeline v4.3.0-rc.2 (https://github.com/Washington-University/HCPpipelines); - BrainVISA Morphologist 5.1.0 (https://brainvisa.info) Statistical analysis: - R v4.3.1 (https://www.r-project.org) • mgcv package: GAMMs • ppcor package: Partial correlations - Python 3.9.13 • scikit-learn+: LASSO regression • abagen v0.1.3: Gene expression mapping (https://github.com/rmarkello/abagen) • BrainSmash v0.11.0: Spatial permutation testing (https://github.com/murraylab/brainsmash) Gene enrichment: - Allen Human Brain Atlas (AHBA): https://human.brain-map.org - Metascape: https://metascape.org (GO + KEGG pathway analysis) Custom scripts: https://github.com/yijinshan1/sulcus2025

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

Data

Policy information about availability of data

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

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

The raw structural MRI data are not publicly available due to institutional restrictions regarding participant privacy and protection of minors, but can be requested from the corresponding authors (S.L. at shuyuli@bnu.edu.cn or Y.H. at yong.he@bnu.edu.cn) upon reasonable academic request.
All processed sulcal morphometric metrics and cognitive scores used for analyses are publicly available at: https://github.com/yijinshan1/sulcus2025

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Sex information was collected from parent/guardian reports at enrollment. Sex was included as a fixed effect in all GAMMs. Results showed modest sex differences primarily in raw sulcal metrics (males: larger SA, CT, depth; females: greater SW), but most differences disappeared after TIV adjustment, supporting allometric scaling rather than divergent neurodevelopmental programs. A total of 214 models showed significant sex effects across sulcal features. Final sample: 145 females, 167 males (312 total participants) WM dataset: 212 females, 190 males (402 scans) ANT dataset: 213 females, 198 males (411 scans)
Reporting on race, ethnicity, or other socially relevant groupings	All participants were recruited from primary schools in Beijing, China, representing a geographically homogeneous Han Chinese population. Race/ethnicity was not explicitly collected as a variable due to population homogeneity. Inclusion was based on standardized cognitive ability tests administered in Chinese. No stratification or analysis by race/ethnicity was performed. Generalizability to other populations should be evaluated in future multi-ethnic cohorts.
Population characteristics	312 typically developing children (145 females, 167 males) Age range: 6.1-13.9 years Recruitment: Primary schools in Beijing, China Inclusion criteria: Based on performance on standardized cognitive ability test administered in Chinese Exclusion criteria: History of neurological disorders, mental disorders, cognitive abnormalities, head injuries, or physical illnesses
Recruitment	Participants were recruited from primary schools in Beijing as part of the Children School Functions and Brain Development Project (CBD, Beijing Cohort). This is a community-based sample and no specific selection bias is expected beyond geographic location (Beijing area).
Ethics oversight	Written informed consent was obtained from each child's parent or guardian. The study protocol was approved by the Ethics Committee of Beijing Normal University. All procedures were conducted in accordance with ethical standards.

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	Sample size (N=312, 490 scans) was determined by available high-quality longitudinal data from the Children School Functions and Brain Development Project (CBD, Beijing Cohort). This sample size provides adequate statistical power for developmental modeling using GAMMs to detect non-linear age-related trajectories while accounting for within-subject correlations across repeated measurements (up to 3 timepoints per participant). For brain-behavior analyses, working memory analysis included 402 scans and ANT analysis included 411 scans. Based on prior developmental neuroimaging studies reporting modest effect sizes ($\Delta R^2 \approx 0.01$-0.10), our sample provides sufficient power (>80%) to detect small-to-moderate effects after FDR correction.
Data exclusions	From an initial dataset of 360 participants, comprehensive quality control procedures excluded participants with: (1) cognitive abnormalities, (2) history of neurological disorders, mental disorders, head injuries, or physical illnesses, resulting in 312 participants for final analysis. For sulcal morphometry: outliers were removed using Median Absolute Deviation (MAD) method with ±3 threshold. Sulci with automated extraction success rates <75% were excluded, resulting in 53 sulci retained from initial 56 (excluded: F.I.P.r.int.2_right, S.Pa.t._left, S.F.orbitaire._right). For cognitive analyses: WM dataset included 402 scans (212 females, 190 males); ANT dataset included 411 scans (213 females, 198 males).
Replication	Developmental findings were validated through multiple approaches: 1. GAMM results required significance in both smooth term Wald test AND likelihood ratio test ($p<0.05$, FDR-corrected) 2. Brain-behavior associations validated through: LASSO regression (80/20 train-test split), Partial Least Squares correlation (1000 permutations), and univariate partial correlations 3. Gene enrichment validated using 1000 spatially-constrained null maps (BrainSmash) to control for spatial autocorrelation 4. Cross-validation: 5-fold CV for LASSO with independent test set validation
Randomization	Not applicable - this is an observational developmental cohort study. All children were recruited from primary schools in Beijing. To ensure the robustness of the LASSO regression in predicting cognitive gains, we employed a nested 5-fold cross-validation on the training set (80%) to optimize the regularization parameter \lambda. The final model performance was evaluated on a completely independent test set (20%). To assess whether the sparse LASSO solution outperformed a diffuse global model, we compared its predictive power against a Ridge regression and a "Full" linear regression model. Gene expression associations were validated using 1000 spatially-constrained null maps generated by BrainSmash to account for spatial autocorrelation in cortical data. This ensures that identified gene-morphometry correlations are not simply products of spatial proximity.
Blinding	Not applicable - this is a computational neuroimaging study with objective, automated measurements. Sulcal identification and morphometric quantification were performed using automated pipelines (BrainVISA Morphologist with SPAM model). All morphometric features were extracted algorithmically without manual intervention beyond quality control.

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
☒	☐ Plants

Methods

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

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.

Magnetic resonance imaging

Experimental design

Design type

This part is not included in the experiment

Design specifications

This part is not included in the experiment

Behavioral performance measures

The participants' cognitive performance was assessed via the classic numerical N-back working memory (WM) task and a child-friendly version of the Attention Network Test (ANT) 93.
In the WM task, participants completed 12 blocks of tasks under three different workload conditions: 0-back, 1-back, and 2-back. In the 0-back condition, participants were instructed to identify whether the current digit was the number 1. In the 1-back and 2-back conditions, participants had to determine whether the current digit matched the one presented one or two steps earlier in the sequence, respectively4. The final WM dataset included 402 scans (212 females, 190 males, see Supplementary Fig. 1). The metric rx2-back was computed by regressing 2-back hits on 0-back hits.
Additionally, ANT performance was measured across 411 scans (213 females, 198 males, see Supplementary Fig. 1). Network efficiency scores were calculated using orthogonal subtractions of mean response times (RT) across conditions. Specifically: The Alerting effect was calculated as RT_{no\mathrm{-}cue}-RT_{double\mathrm{-}cue}, representing the benefit of a temporal warning cue. The Orienting effect was calculated as RT_{center\mathrm{-}cue}-RT_{spatial \mathrm{-}cue}, representing the benefit of a spatial orienting cue. The Executive Control (Conflict) effect was calculated as RT_{incongruent}-RT_{congruent}, representing the performance cost induced by flanker conflict. Note that for Alerting and Orienting, higher scores indicate greater efficiency (better performance), whereas for Executive Control, a higher score indicates a larger conflict cost (worse performance).

Acquisition

Imaging type(s)

structural

Field strength

3.0T

Sequence & imaging parameters

Scanner: 3T Siemens Prisma scanners at Peking University, Beijing, China
Sequence: T1-weighted MPRAGE sequence
Parameters:
- TR = 2530 ms
- TE = 2.98 ms
- TI = 1100 ms
- Flip angle = 7°
- FOV = 256 × 224 mm²
- Number of slices = 192
- Slice thickness = 1 mm
- Bandwidth = 240 Hz/Px
- Scan time = 5 min 58 s

Area of acquisition

Whole brain

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software

HCP pipeline v4.3.0-rc.,BrainVISA Morphologist 5.1.0

Normalization

Nonlinearly normalized to the Chinese Pediatric Atlas (CHN-PD) space

Normalization template	Chinese Pediatric Atlas (CHN-PD)
Noise and artifact removal	A bandpass filter (0.01 Hz<f< 0.08 Hz) was applied; then, additional motion noise in rs-fMRI was removed using the ICA-AROMA procedure. The shared variance between the global signal and time series was also removed.
Volume censoring	This part is not included in the experiment

Statistical modeling & inference

Model type and settings	Mass univariate analysis:Generalized Additive Mixed Models (GAMMs) with thin-plate regression splines (k=3 basis functions) to capture non-linear developmental trajectories. Mixed-effects structure: fixed effects (age smooth term + sex), random effects (subject-specific intercepts to account for repeated measures). Estimation: Restricted Maximum Likelihood (REML). No drift or auto-correlation structures were explicitly modeled. Multivariate analysis: Partial Least Squares (PLS) correlation to identify global modes of covariance between sulcal morphometry and cognitive performance. Predictive modeling: LASSO (Least Absolute Shrinkage and Selection Operator) regression with L1 penalty for sparse feature selection. Training: 80% data with 5-fold cross-validation. Testing: 20% independent hold-out set.
Effect(s) tested	GAMM developmental analysis: - Primary effect: Non-linear age-related changes in sulcal morphometry (6 features × 53 sulci) - Covariate: Sex (fixed effect) - Nested model comparison: Full model (age + sex) vs. Null model (sex only) Brain-behavior associations: - Longitudinal change-to-change coupling: Δ morphometry predicting Δ cognitive performance - Covariates controlled: baseline age, sex, inter-scan interval - Cognitive outcomes: Working memory (rx2-back), ANT sub-networks (Alerting, Orienting, Executive Control) No factorial designs or task-condition contrasts were used.

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

Statistic type for inference (See Eklund et al. 2016)	no voxel-wise or vertex-wise analysis
Correction	Multiple comparisons correction:False Discovery Rate (FDR) using Benjamini-Hochberg procedure applied separately to: 1. GAMM smooth term p-values (age effects) 2. GAMM likelihood ratio test p-values (model comparison) 3. GAMM sex fixed effect p-values 4. Univariate brain-behavior correlations Threshold: pFDR < 0.05 Gene enrichment: Spatial autocorrelation correction using BrainSmash with 1000 spatially-constrained null permutations to control for spatial proximity effects. LASSO predictive models:Permutation-based p-values (10,000 permutations) with FDR correction across multiple cognitive domains.

Models & analysis

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

Multivariate modeling and predictive analysis	LASSO Regression: - Independent variables: Longitudinal changes in 6 sulcal morphometry features (ΔSA, ΔmaxD, ΔmeanD, ΔSL, ΔCT, ΔSW) across 53 regions per hemisphere - Feature extraction: TIV-adjusted residuals for SA, maxD, meanD, SL; raw values for CT and SW - Dimension reduction: L1 regularization penalty driving irrelevant coefficients to zero - Model: Minimize penalized residual sum of squares with λ optimized via 5-fold CV (grid search: 50 logarithmic values from 10^{-5} to 10^{0}) - Training/evaluation: 80/20 split with standardization (Z-score) based on training set statistics to prevent data leakage - Metrics: Root Mean Squared Error (RMSE_cv) for model selection; Spearman correlation (r) for test set performance PLS Correlation:

- Variables: X (sulcal morphometry matrix), Y (cognitive scores)
- Extraction: First latent variable (LV1) capturing maximum covariance
- Validation: 1000 permutations for significance testing; bootstrap resampling (1000 iterations) for feature stability with FDR correction