

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 (<i>n</i>) 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 <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☐	☒ 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. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ 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 <i>d</i> , Pearson's <i>r</i>), 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	Open-source neuroimaging datasets from two cohorts of the Human Connectome Project (HCP), the developing HCP (dHCP) and HCPDevelopment (HCPD), were used. A separate dataset from the Nathan Kline Institute - Rockland Sample (NKI-RS) was used for replication purposes. dHCP: MR images were acquired on a 3T Philips Achieva scanner equipped with a dedicated neonatal imaging system and a neonatal 32 channel phased array head coil at the Evelina Neonatal Imaging Centre, St Thomas' Hospital, London. HCPD: Participants are scanned on Siemens 3T Prisma (whole-body scanner with 80mT/m gradient coil, slew rate of 200T/m/s; Siemens 32-channel head coil; software version: E11C) using matched protocols across 4 sites (Harvard University, University of California – Los Angeles, University of Minnesota, and Washington University in St. Louis) in the United States. NKI-RS: Scanned using a 3.0T Siemens TIM Trio located at the Nathan Kline Institute, with a resting state sequence of a multiband imaging protocol (T= 645 ms/3 mm isotropic voxels). For more details on data collection, please refer to the following references: - dHCP: Edwards, A. David, et al. "The developing human connectome project neonatal data release." <i>Frontiers in neuroscience</i> 16 (2022): 886772. - HCPD: Somerville, L. H. et al. The Lifespan Human Connectome Project in Development: A large-scale study of brain connectivity development in 5–21 year olds. <i>Neuroimage</i> 183, 456–468 (2018). - NKI-RS: Nooner, K. B. et al. The NKI-Rockland Sample: A Model for Accelerating the Pace of Discovery Science in Psychiatry. <i>Front. Neurosci.</i> 6, 152 (2012).
Data analysis	The codes used in this study are publicly available. Preprocessing codes for the dHCP data are available at https://github.com/BioMedia/dhcp-structural-pipeline (structural data) and https://git.fmrib.ox.ac.uk/seanf/dhcp-neonatal-fmri-pipeline (functional data). Scripts to align dHCP native surfaces to template space can be found at https://github.com/ecr05/dHCP_template_alignment. The NKI-RS dataset was

preprocessed using the ciftify toolbox from <https://github.com/edickie/ciftify>. Connectopic mapping of resting-state functional data was performed using <https://github.com/koenhaak/congrads>. The 'procrustes' function in MATLAB was used for alignment between individual and group level gradients. Statistical testing of age effects was implemented using surface-based linear models in a Matlab toolbox, SurfStat (<http://www.math.mcgill.ca/keith/surfstat/>; <https://brainstat.readthedocs.io/en/master/>). The 'silhouette' function and 'inpolygon' function in MATLAB was used to compute silhouette scores and network overlap, respectively. The 'fitlme' function in MATLAB was used for mixed effects analysis while the 'fitrlinear' function in MATLAB was used for predictive modeling. Regional microarray expression data for the transcriptomic analysis was processed with the abagen toolbox (version 0.1.3; <https://github.com/rmarkello/abagen>), and for the spatial autocorrelation-preserving permutation tests, surrogate maps were generated by spatially permuting parcellated brain regions using the BrainSMASH python toolbox (<https://brainsmash.readthedocs.io/en/latest/>). Gene specificity was tested using the Gene Annotation by Macroscale Brain-Imaging Association Matlab toolbox (<https://github.com/dutchconnectomelab/GAMBA-MATLAB>). Gene enrichment analysis utilized ShinyGO 0.77 (<http://bioinformatics.sdstate.edu/go/>) and a cell-type specific expression analysis developmental expression tool (<http://doughertylab.wustl.edu/csea-tool-2/>; version 1.1). Alternative tools including Enrichr (<https://maayanlab.cloud/Enrichr/>; using version available after updated on June 8, 2023) and g:Profiler (<https://biit.cs.ut.ee/gprofiler/gost>; version e111_eg58_p18_30541362) were used for testing robustness. Generative network modeling was implemented by adopting the analysis code made available at <https://github.com/StuartJO/GenerativeNetworkModel>. For visualization of NEOMAPs including parcellation boundaries the matlab codes available at <https://github.com/StuartJO/plotSurfaceROIBoundary>. The following toolboxes were used for analyses of simulated data: Brain connectivity toolbox (<https://sites.google.com/site/bctnet/>) and Brainspace toolbox (<https://brainspace.readthedocs.io/en/latest/>). All code generated for analyses and for creating figures are available at: https://github.com/COMBINE-SKKU/NatNeurosci_2024_SWPark.

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 data used in this study are publicly available. The neuroimaging data from dHCP is available at <http://www.developingconnectome.org/data-release/second-data-release/> and preprocessed with available, recommended pipelines. Surface and volumetric neonatal atlas used were downloaded from <https://web.gin.g-node.org/BioMedIA/dhcp-volumetric-atlas-groupwise> and <http://brain-development.org/brain-atlases/atlases-from-the-dhcp-project/cortical-surface-atlas-bozek/>, respectively. The HCPD data can be downloaded from the NDA website at <https://nda.nih.gov/general-query.html?q=query=featured-datasets:HCP%20Aging%20and%20Development>. The minimally preprocessed structural and functional data from the public release HCP pipelines v3.22 (<https://github.com/Washington-University/HCPpipelines>) that were released as a part of the Lifespan HCP-Development Release 2.0 were used. The NKI-RS dataset can be downloaded from the International Neuroimaging Data-Sharing Initiative database (https://fcon_1000.projects.nitrc.org/indi/enhanced/sharing_neuro.html). The sleep/wake comparison dataset is available from the NDAR database (<https://nda.nih.gov/>; Collection # 2072). Regional microarray expression data were provided by the Allen Human Brain Atlas (<https://human.brain-map.org/static/download>) and BrainSpan (<https://www.brainspan.org/static/download.html>). We downloaded the Yeo-Krienen 7 Network Atlas and the Schaefer atlas from https://github.com/ThomasYeoLab/CBIG/tree/master/stable_projects/brain_parcellation.

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

This study reports on self-reported participant sex (as a biological attribute). For dHCP, 107 males and 88 females, for HCPD, 417 males and 186 females and for NKI-RS, 33 males and 37 females were included. Study findings apply to both male and female sexes. Sex was considered as a potential covariate in all statistical analyses.

Reporting on race, ethnicity, or other socially relevant groupings

The dHCP dataset does not provide information on race, ethnicity, or other socially relevant groupings. The HCPD dataset does report on race/ethnicity, however we did not use this information as a part of our analyses. We did not use any information related to race or ethnicity for the NKI-RS dataset.

Population characteristics

dHCP: All anatomical scans were reviewed by a neuroradiologist who provided a radiology score. Scans with a score of 3 (incidental findings with unlikely clinical significance but possible analysis significance) or 5 (incidental findings with possible/likely significance for both clinical and imaging analysis) were excluded from further analyses. After quality control and preprocessing, the final sample consisted of 195 healthy term infants (107 male, mean [SD] gestational age at birth=39.9 [1.26] weeks, age at scan=41.0 [1.67] weeks) and 60 pre-term infants (42 male, gestational age at birth=32.6 [3.12] weeks, at scan=35.5 [2.78] weeks).
HCPD: After quality control and preprocessing, the final sample consisted of 603 healthy participants (417 male, mean [SD] age=14.8 [3.88] years). Participants were excluded for insufficient English, premature birth, MRI contraindications, or a lifetime history of serious medical conditions, endocrine disorders, head injury, psychiatric/neurodevelopmental disorders.
NKI-RS: To ensure replicability, we utilized the cross-sectional and longitudinal pediatric data from an independent open-source dataset, the Nathan Kline Institute / Rockland Sample (NKI-RS; n=71 (34 male); age range 6 - 19 years; mean [SD] age=10.9±2.9). Initially, we selected participants without any prior diagnosis and who had at least two timepoints of fMRI data.

Recruitment

Various recruitment methods were employed to gather participants, including targeting related siblings for the HCP data, which were not directly related to the objectives of the current study. However, our findings are not likely to be affected by the recruitment approach used.

Ethics oversight

All participants provided written informed consent statements before participation in the study. The dHCP data were reviewed and approved by United Kingdom Health Research Authority (Research Ethics Committee reference number: 14/LO/1169). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. The HCP data were acquired using protocols approved by the Washington University institutional review board. For the NKI-RS, Institutional Review Board Approval was obtained at the Nathan Kline Institute (Phase I #226781 and Phase II #239708) and at Montclair State University (Phase I #000983A and Phase II #000983B). All data articles of the open-source datasets used in this study mention that written consent was obtained from all participants or a parent/legal guardian in the case of those under 18 years of age. Information regarding the participant compensation is currently not available on the websites and/or data papers.

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](https://www.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	We included all available data from the relevant datasets, after applying several exclusion criteria explained below.
Data exclusions	The dHCP second release dataset consists of a developmental sample of n=558 (224 female). Among individuals with both structural and functional MRI scans (n=461) individuals with a neuro-radiological score of 3 (incidental findings with unlikely clinical significance but possible analysis significance) or 5 (incidental findings with possible/likely significance for both clinical and imaging analysis) were excluded (n=374). The final sample consisted of a total of 255 infants (106 females), after the MRI scans finished preprocessing without errors, survived motion-related quality control (mean framewise displacement [FD] < 0.4) as well as showed strong resemblance between the individual thalamic gradient and group-level gradient. The HCPD dataset consists of a total of n=652. Likewise, among individuals with both structural and functional MRI scans, those that survived the preprocessing pipeline and motion-related quality control of Mean FD < 0.4 (n=615), as well as showed strong spatial correlation to the group-level thalamic gradient were included resulting in a final sample of n=603 (327 females). NKI-RS: NKI-RS: To ensure replicability, we utilized the cross-sectional and longitudinal pediatric data from an independent open-source dataset, the Nathan Kline Institute / Rockland Sample (NKI-RS; n=71 (34 male); age range 6 - 19 years; mean [SD] age=10.9±2.9). We selected participants without any prior diagnosis and who had at least two time-points of all imaging data modalities.
Replication	The results were reproduced over several relevant biological and image-processing factors, including 1) an independent dataset (i.e., NKI-RS) utilizing both cross-sectional and longitudinal samples, 2) global signal regression, 3) sleep-wake state dependency, 4) different gene enrichment analysis tools.
Randomization	As this study did not entail distinct experimental groups or conditions, randomization was not performed.
Blinding	Each participant in the study underwent an identical study protocol, and there was no allocation to distinct experimental groups. Thus blinding was unnecessary.

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	NA
Novel plant genotypes	NA
Authentication	NA

Magnetic resonance imaging

Experimental design

Design type	Resting-state functional MRI
Design specifications	NA
Behavioral performance measures	NA

Acquisition

Imaging type(s)	Resting-state functional MRI
Field strength	3 Tesla
Sequence & imaging parameters	dHCP: High-resolution T2w and T1w structural images were acquired with 0.8 mm isotropic spatial resolution and thickness of 1.6mm, overlapped by 0.8 mm (T2w: TE/TR=156 ms/12 s; T1w: TE/TR/TI=8.7/4795/1740 ms). Resting-state fMRI data were acquired using multiband accelerated echo-planar imaging for approximately 15 minutes (TE/TR=38ms/392ms, flip angle=34°, spatial resolution=2.15 mm isotropic, 2300 volumes). Single-band reference scans were also collected with bandwidth matched readout and additional spin-echo EPI acquisitions with both AP/PA phaseencoding directions. HCPD: T1-weighted images were acquired using a 3D multi-echo MPAGE sequence (TR=2500, TI=1000 ms, TE=1.8/3.6/5.4/7.2 ms, flip angle=8°, FOV=256×240×166 mm, 320×300 matrix, 208 slices, 0.8 mm isotropic voxels). Resting-state fMRI data were acquired with a multiband gradient-recalled (GRE) EPI sequence (TR=800 ms, TE=37 ms, flip angle=52°, FOV=208 mm, 104×90 matrix, 72 oblique axial slices, 2 mm isotropic voxels, multiband acceleration factor of 8, 1912 volumes). NKI-RS: A 3.0-T Siemens MAGNETOM Trio-Tim scanner was used to collect resting state fMRI data using the following imaging parameters: TR =645 ms, TE =30 ms, 40 slices, flip angle = 60° and voxel size = 3.0 × 3.0 × 3.0 mm.
Area of acquisition	whole brain
Diffusion MRI	☐ Used ☒ Not used

Preprocessing

Preprocessing software	We refer the reader to the original articles (and related reporting summaries) for a detailed description: dHCP MRI preprocessing - Makropoulos et al., 2018 (NIMG) & Fitzgibbon et al., 2020 (NIMG); and HCPD MRI preprocessing – Somerville et al., 2018 (NIMG). A summary can be found in the article methods section.
Normalization	We refer the reader to the original articles (and related reporting summaries) for a detailed description: dHCP MRI preprocessing - Makropoulos et al., 2018 (NIMG) & Fitzgibbon et al., 2020 (NIMG); and HCPD MRI preprocessing – Somerville et al., 2018 (NIMG). A summary can be found in the article methods section.
Normalization template	We refer the reader to the original articles (and related reporting summaries) for a detailed description: dHCP MRI preprocessing - Makropoulos et al., 2018 (NIMG) & Fitzgibbon et al., 2020 (NIMG); and HCPD MRI preprocessing – Somerville et al., 2018 (NIMG). A summary can be found in the article methods section.
Noise and artifact removal	We excluded participants with mean framewise displacement (FD) > 0.4 mm, which enables a trade-off between sample size and data quality.
Volume censoring	The first 10 volumes were discarded to achieve steady-state magnetization of the fMRI signal

Statistical modeling & inference

Model type and settings

Neuroimaging analysis: We estimated thalamocortical connectopic maps, which represent dominant modes of functional connectivity change within the thalamus, based on the voxel-wise connectivity between each thalamic voxel and the neocortex, and its neocortical projection maps. This procedure is described in Haak et al., 2018 (NIMG).

Transcriptomic analysis: Partial least squares (PLS) analysis was used to investigate the correlation with gene expression patterns. An ensemble of randomized null phenotypes controlling for spatial autocorrelation (SBP-spatial models) were used to verify the statistical significance of these correlations. The surrogate maps of spatially permuting parcellated brain regions while conserving for spatial autocorrelation were generated using the BrainSMASH python toolbox (<https://brainsmash.readthedocs.io/en/latest/>). We further assessed the specificity of our genes of interest using the most stringent null-brain-gene model that generates null distributions based on random genes that are over-expressed in brain tissues ($n=2,711$ selected by $q < 0.05$, FDR correction, one-sided two-sample t-test comparing brain tissues and other body tissues) via the Gene Annotation by Macroscale Brain-imaging Association Matlab toolbox. The significant gene sets were analyzed for biological processes and cell-type specificity using ShinyGO and cell-type specific expression analysis (CSEA), respectively. For the former, significant results are reported after Benjamin-Hochberg FDR correction ($q < 0.05$). For the latter, statistical significance of overlap between our candidate genes and cell-specific genes for each developmental stage was tested with Fisher's exact test with Benjamin-Hochberg multiple testing correction for the number of cell types assayed. Genes enriched in a particular cell type were identified by specificity index p-value (pSI), where lower pSI values indicate more specific genes.

Computational modeling: Generative network modeling based on the cost-topology trade-off model was used to investigate the effects of thalamocortical, genetic, spatial, and topological constraints by synthetically creating functional connectome topology. Model performance was evaluated by comparing the synthetic networks derived from the models to empirically extracted networks in terms of degree, clustering, betweenness, and edge length distributions. For each measure, the distributions were compared using the Kolmogorov-Smirnov (KS) statistic, a non-parametric test that quantifies the distance between two distributions, where lower values imply greater similarity. Detailed descriptions for generative network modeling are in Oldham et al., 2022.

Effect(s) tested

Neuroimaging analysis: These maps were statistically tested for significant aging effects at FDR-corrected $p < 0.05$, controlling for sex and mean FD, using surface-based linear models implemented in SurfStat.

Transcriptomic analysis: PLS analysis with an ensemble of randomized null phenotypes controlling for spatial autocorrelation (SBP-spatial models) were used to verify the statistical significance of correlations with gene expression patterns.

Computational modeling: Model performance was evaluated by comparing the synthetic networks derived from the models to empirically extracted networks in terms of degree, clustering, betweenness, and edge length distributions. For each measure, the distributions were compared using the Kolmogorov-Smirnov statistic.

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

Statistic type for inference

Neuroimaging analysis: Voxelwise, massive univariate analysis with an FDR-corrected $p < 0.05$ (See [Eklund et al. 2016](#))

Correction

Neuroimaging analysis: False discovery rate correction was used

Transcriptomic analysis: spin tests and Benjamin-Hochberg FDR correction was used

Computational modeling: Kolmogorov-Smirnov statistic and a conservative 3-step multi-resolution optimization procedure that sampled 10,000 combinations of different parameters was used to obtain the best-fitting parameters.

Models & analysis

n/a | Involved in the study

- ☐ ☒ Functional and/or effective connectivity
- ☐ ☒ Graph analysis
- ☐ ☒ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Neuroimaging analysis: Thalamocortical connectopic maps were calculated by conducting dimensionality reduction on a similarity matrix which was derived from a Pearson correlation matrix between the functional connectivity of thalamus and neocortex.

Graph analysis

Computational modeling: To estimate how well the synthetic networks made from thalamocortical connectivity models reflect the functional segregation between cortical networks as seen in the neuroimaging data analyses, the affinity matrices derived from the synthetic networks showing best performance were analyzed in terms of modularity network measure, which was derived using the Brain Connectivity Toolbox.

Multivariate modeling and predictive analysis

Two distinct analyses were performed on the longitudinal dataset. The first analysis involved a mixed effects model, where we employed the 'fitlme' function in MATLAB to execute the mixed effects analysis. In the model, age was treated as a fixed effect and participant as a random effect to model the segregation indices of both salience-external and salience-internal networks. Secondly, we aimed to estimate the influence of thalamocortical connectivity development on network differentiation of the neocortex during later developmental stages. For this, we applied a supervised learning framework using the 'fitrlinear' function in MATLAB, with a 5-fold cross-validation, employing least squares and lasso regularization. Specifically, we investigated whether the difference of the thalamocortical NEOMAPs between follow-up and baseline data can predict the segregation index between internal and external networks within the 'cortico-cortical' gradients. The mean average error and median correlation coefficient were used for evaluating model fit.