

**Improving the Robustness and Reproducibility of Functional Connectomics-
based Biomarkers in Neurodevelopmental and Psychiatric Conditions**



Jivesh Ramduny

Thesis submitted to Trinity College Dublin, the University of Dublin
for the degree of

Doctor in Philosophy

2024

Declaration

I, **Jivesh Ramduny**, declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

A handwritten signature in black ink, appearing to read 'Jivesh Ramduny', written diagonally.

Jivesh Ramduny

August 31, 2023

Summary

Functional magnetic resonance imaging (fMRI)-based functional connectivity offers remarkable potential to non-invasively map individual differences in brain functional organisation to individual differences in behavioural phenotypes. Applied to neurodevelopmental and psychiatric conditions, this potential has renewed the search for neuroimaging-based biomarkers that have diagnostic and prognostic validity. Reproducible functional connectivity-based measures are key to achieving this goal. The overarching goal of this thesis is to improve the reproducibility of conventional and advanced functional connectomics-based measures to enhance their sensitivity to phenotypic differences of clinical relevance in neurodevelopmental and psychiatric conditions. I achieve this through four *empirical* studies using Open Science datasets as well as in-house data. All analysis code used in this thesis are publicly accessible via GitHub repositories: <https://github.com/JRam02>.

Using the functional connectome fingerprinting approach, **Study 1** demonstrates that the test-retest fingerprint matching algorithm can serve as a simple and intuitive measure to compare, benchmark, and adjudicate between analysis pipelines. I show that this approach enables the identification of data pre- and post-processing parameters that optimise the detection of individual differences in the functional connectome. I also argue that fingerprint accuracy offers a number of advantages over the conventional approach to comparing pipelines — the computation of test-retest reliability using the intraclass correlation coefficient (ICC). Across five datasets spanning childhood and adulthood derived from the Consortium for Replicability and Reproducibility (CoRR; $N = 264$, 7-43 years, mean age $\pm$ SD = 19.4 $\pm$ 5.6 years), I show that identification accuracies of up to 94% are achieved for pipelines that employ: (I) global signal regression; (II) finer-grained brain parcellations; (III) cortical regions, excluding subcortical and cerebellar structures; (IV) medial frontal and frontoparietal networks; (V) a focus on discriminative edges; and (VI) longer scan duration. Together, these factors form components of two pipelines that are optimised for reproducing the individual functional connectome. I suggest that a pipeline defined in this way should better enable the detection of robust and reproducible brain-behaviour relationships, particularly in developmental samples.

In **Study 2**, I address the trade-off between strict data quality requirements (particularly regarding head motion) and the large sample sizes required for robust and reproducible brain-behaviour associations. I propose the implementation of a machine learning technique referred to as “bagging” or bootstrap aggregation at the timeseries level for estimating individual functional connectomes, enabling the salvage of “high-motion” individuals that would otherwise be discarded in these analyses. Using resting-state data from CoRR and Healthy Brain Network (HBN; $N = 423$, 6-20 years, mean age $\pm$ SD = 10.7 $\pm$ 3.0 years), I show that deriving the functional connectome from bootstrap samples of low-motion timepoints can successfully differentiate an

individual from a group and increases sample inclusivity. I further demonstrate that bagging yields robust univariate brain-behaviour associations (i.e., relationships between functional connectivity and age) even when as few as 60 “clean” timepoints are retained (16% of the full timeseries; sampled *with* replacement over 500 iterations): the brain-behaviour associations obtained using bagging are comparable with those obtained using a standard (full timeseries) approach. The utility of bagging lies in its ability to rescue high-motion individuals by maximising participant inclusivity by ~4-30% of the high-motion individuals from CoRR and HBN; its application may reduce the tension between data quality vs. sample size requirements, mitigating resource wastage and the unethical practice of discarding large numbers of participants to ensure inclusive and reproducible brain-behaviour associations.

Study 3 builds on the work of **Studies 1 and 2** by investigating whether the factors that maximise the detection of individual differences in the functional connectome, together with head motion strategies that maximise participant inclusivity and reduce motion artefacts also boost the predictive power of functional connectomes for behavioural phenotypes, using connectome-based predictive modelling (CPM). Using the HBN ($N = 540$, 6-21 years, mean age $\pm$ SD = 10.7 $\pm$ 3.0 years) and the components and strategies identified in **Studies 1 and 2**, I sought to strengthen connectome-based predictions and improve their generalisability across three domains: age, general psychopathology [CBCL], and intellectual ability [FSIQ]. I found that the data pre- and post-processing factors previously identified for improving the identifiability of individual functional connectomes — (I) dimensionality of parcellation schemes; (II) focus on discriminatory functional networks such as the medial frontal and frontoparietal networks; and (III) head motion mitigation strategies including scrubbing and bagging — neither improved nor disimproved the CPM-based predictions. Generalisability to an independent sample of non-Autistic individuals obtained from the Autism Brain Imaging Data Exchange (ABIDE; $N = 439$, 6-57 years, $M = 16.2\pm 6.7$ years) was also generally unaffected, although finer parcellation schemes and bagging improved generalisability for FSIQ. Therefore, while the factors and strategies that I examined do not necessarily improve validity (i.e., strength of brain-behaviour relationships), the evidence suggests that they do strengthen reproducibility (i.e., generalisability of brain-behaviour relationships).

In **Study 4**, I take a different route to reproducibility: ecological validity. While resting-state fMRI conditions have been widely applied to study brain functions and dysfunctions across development, naturalistic paradigms may offer a more ecologically valid index of cognitive, emotional, and social processes. I propose a naturalistic analysis framework that employs a short emotionally evocative movie called “The Present” (<https://www.jacob-frey.com>) to explore how functional brain responses are modulated by depressive symptomatology and emotion regulation strategies at multiple spatial and temporal scales in a sample of female adolescents experiencing subclinical depressive symptoms ($N = 50$, 13-20 years, mean age $\pm$ SD = 17.0 $\pm$ 1.9 years). In this

study, I examined the (I) shared brain responses using intersubject correlation, (II) “static” and dynamic functional connectivity patterns, and (III) signal complexity of the naturalistic fMRI signals using the Hurst Exponent (H), as well as their phenotypic relevance for depressive symptom severity and adaptive/maladaptive emotion regulation strategies. The findings suggest that depressive symptomatology and emotion regulation strategies shape the processing of affect-laden events in real-life scenarios.

Taken together, the findings of these four thesis projects highlight opportunities for improving the reproducibility of functional connectivity-based measures in developing populations, but also reveal several limitations that remain to be addressed. I discuss these findings in the context of several broader topics of concern within the neuroimaging field: sample size, the phenotype challenge, trade-offs of the resting-state fMRI approach, methodological variability, rate of developmental changes, and the quest for multimodal MRI approaches. I also consider several future avenues towards increased reproducibility, including meta-matching frameworks, multivariate techniques, multiverse pipelines, and sex differences. I conclude that a reproducibility lens remains central to the collective effort to discover functional connectivity-based biomarkers of clinical relevance, and ultimately, to improve outcomes for those with neurodevelopmental and psychiatric conditions.

Acknowledgment

The research conducted during this PhD was supported by the Trinity College Dublin Provost's PhD Awards (2018-2019) from 01/09/2019 to 31/08/2023.

First and foremost, I wish to express my sincere gratitude to Dr. Clare Kelly for her mentorship and guidance during the course of the PhD programme which allowed me to embark upon an exciting journey of learning and discovery which lies at the intersection of methodological innovation and brain development. Clare has offered me her continued support despite the global pandemic to facilitate my academic journey and personal growth. She trusted me as a Graduate Teaching Assistant by allowing me to lead Python-based lab sessions for the Year 1 Psychology students to develop scientific research skills and perform statistical analyses in a reproducible manner.

I am privileged to have collaborated with Dr. Tamara Vanderwal from the University of British Columbia, Canada, Prof. Hugh Garavan from the University of Vermont, USA, and Dr. Robert Whelan from Trinity College Dublin, Ireland. The collaborations have offered me unique perspectives about my research by allowing me to think about different methods in addressing the research questions, hypotheses, and novelty in the findings.

The work carried out in this thesis would not have been possible without the moral support of my colleagues and friends within and outside Trinity College Dublin: Sharon Chi Tak Lee, Colin Simon, Yihe Weng, Chiara Caldinelli, Nadja Enz, Camilla Roselli, Michael Connaughton, Mélanie Garcia, Andrew Breen, Daniel McLoone, Abhimanyu Bhardwaj, Tom Farnan, Celine Fox, Kelly Donegan, Steph Suddell, Anna Hanlon, Eike Buabang, Parnian Rafei, Francesca Sibilía, Alannah Smyth, Julia Paterson, Vanessa Teckentrup, Anna Marie Rosická, Luisa Franzen, Maike Splittgerber, Gabriel Byczynski, Feifan Chen, Katherine Adcock, Colum Ó Sé, Elva Arulchelvan, Kathy Ruddy, Áine Dineen, Anna Truzzi, Graham King, Clíona O'Doherty, Lorijn Zaadnoordijk, Oisín Joyce, Sarah McComish, David Lee, Aisling Leavy, Aimée Freeburn, Hugh Delaney, Emanuele Plini, Eoin O'Neill, MaryAnn Ciosk, Tevfik Syonmez, Carrie Mackintosh, Jack Smyth, Elon Mitchell, Nora Rex, Felix Betz, Danny Brosnan, Mihail Dimitrov, Matthew Bowdler, Michelangelo Domina, Michael Minotakis, Shaun Warrington, Jose Manzano Patron, Elinor Thompson, Sarina Iwabuchi, William Cottam, Matteo Bastiani, Stefan Pszczolkowski Parraguez, Andrea Torchi, and Sofiya Sergeyeva. They were instrumental in helping me maintain a work-life balance and allowing me to thrive throughout the hectic schedule of the PhD.

I wish to thank the administrative staff at the Trinity College Institute of Neuroscience for promoting a conducive, cordial, and inclusive research environment during my PhD. They have always supported my queries with regards to my studentship, academic registry, and funding.

I am eternally grateful to my family for their sacrifice and support throughout my academic journey which started in Mauritius followed by Singapore, UK, and now, Ireland.

List of Manuscripts in Preparation, Preprints & Publications

Ramduny, J., Vanderwal, T., & Kelly, C. Using functional connectome fingerprinting as a testbed for strategies to improve reproducibility.

Ramduny, J., Vanderwal, T., & Kelly, C. Towards data salvage in high-movement cohorts: bagging yields robust and reproducible brain-behaviour relationships.

Ramduny, J., Weng, Y., Whelan, R., Vanderwal, T., & Kelly, C. What methodological factors can improve the reproducibility of connectome-based predictive modelling in developing youths?

Ramduny, J., MacSweeney, N., Petkova, E., Kellaghan, E., Lahert, N., Gallagher, L., Vanderwal, T., & Kelly, C. Naturalistic emotional processing in adolescent females with depressive symptoms.

Ramduny, J., Garcia, M., & Kelly, C. (2023). Establishing a reproducible and sustainable analysis workflow. In R. Whelan & H. Lemaitre (Eds.). *Methods for Analysing Large Neuroimaging Datasets*. Springer Nature. doi: [10.31219/osf.io/RCXG8](https://doi.org/10.31219/osf.io/RCXG8).

List of Invited Talks

Ramduny, J. (2023). Using functional connectome fingerprint to investigate the robustness of connectome-based phenotype predictions in developing youths. Invited Talk for the Development and Lifespan Session at the Computational Psychiatry Conference, Dublin, Ireland.

Ramduny, J. (2023). Improving connectome-based fingerprinting for predicting robust and reliable brain-behaviour relationships. Invited presentation at the meeting of the Brain Connectivity and Cognition Lab, UCLA, Los Angeles, California, USA.

Ramduny, J. (2022). Improving the Robustness and Reproducibility of Functional Connectomics-based Analyses in Neurodevelopmental and Psychiatric Conditions. Invited presentation at the 9th meeting of the Predictive Brain Health Modelling Group, Trinity College Dublin, Dublin, Ireland.

Whelan, R., Lemaitre, H., **Ramduny, J.**, Madhyastha, T., Hoffstaedter, F., Esteban, O. et al. (2022). An Introduction to Methods for Analyzing Large Neuroimaging Datasets. Presentation as part of an Educational Symposium at the 28th annual meeting of the Organisation for Human Brain Mapping, Glasgow, Scotland.

Ramduny, J. (2022). Getting Started with Reproducible Neuroimaging Analyses. Invited presentation conducted for the degree of MSc Neuroscience, King's College London, London, UK.

Ramduny, J. (2021). Improving the reproducibility of functional connectomics-based biomarkers using resting-state and naturalistic fMRI. Invited presentation at the meeting of the Modinos Lab, King's College London, held virtually due to COVID-19.

Ramduny, J. (2020). Improving the sensitivity and reproducibility of functional connectomics-based biomarkers in neurodevelopmental conditions. Invited presentation at the annual meeting of The Psychological Society of Ireland EGG, held virtually due to COVID-19.

Ramduny, J. (2020). Why should we care about the reproducibility crisis in research? Presentation at the 1st meeting of the ReproducibiliTea Journal Club, Dublin, Ireland.

List of Conference Posters

Ramduny, J., Weng, Y., Whelan, R., Vanderwal, T., & Kelly, C. (2023). Does maximising connectome fingerprint identifiability improve connectome-based phenotype prediction? Poster presented at the 10th annual meeting of the Flux Society, Santa Rosa, California, USA. <https://github.com/JRam02/FLUX2023POSTER TCD.pdf>.

Ramduny, J., Weng, Y., Whelan, R., Vanderwal, T., & Kelly, C. (2023). Testing the use of fingerprinting-informed pipelines to enhance brain-behaviour predictions. Poster presented at the 29th annual meeting of the Organisation for Human Brain Mapping, Montreal, Canada. <https://github.com/JRam02/OHBM2023POSTER TCD.pdf>

Ramduny, J., Weng, Y., Whelan, R., Vanderwal, T., & Kelly, C. (2023). Delineating the functional brain as fingerprints to predict robust brain-behaviour relationships in developmental youths. Poster presented at the 78th annual meeting of the Society of Biological Psychiatry, San Diego, California, USA. <https://github.com/JRam02/SOPB2023POSTER TCD.pdf>

Ramduny, J., Weng, Y., Whelan, R., Vanderwal, T., & Kelly, C. (2023). Does maximising connectome fingerprint identifiability improve connectome-based phenotype prediction? Poster presented at the 1st meeting of the School of Psychology Symposium, Trinity Biomedical Sciences Institute, Dublin, Ireland. <https://github.com/JRam02/SOP2023POSTER TCD.pdf>.

Ramduny, J., Vanderwal, T., & Kelly, C. (2022). Towards data salvage in high-movement cohorts: bagging yields brain-behaviour relationships. Poster presented at the 14th annual meeting of the Neuroscience Ireland Young Investigator Symposium, Dublin, Ireland. <https://github.com/JRam02/NSI2022POSTER TCD.pdf>.

Ramduny, J., Vanderwal, T., & Kelly, C. (2022). Towards data salvage in high-movement cohorts: bagging yields robust brain-behaviour relationships. Poster presented at the 28th annual meeting of the Organisation for Human Brain Mapping, Glasgow, United Kingdom. <https://github.com/JRam02/OHBM2022POSTER TCD.pdf>.

Ramduny, J., & Kelly, C. (2021). Shifting towards movie-watching fMRI to investigate emotional processing in adolescent depression. Poster presented at the 27th annual meeting of the Organisation for Human Brain Mapping, held virtually due to COVID-19. <https://github.com/JRam02/OHBM2021POSTER TCD.pdf>.

Ramduny, J., & Kelly, C. (2021). Naturalistic emotional processing in adolescent depression. Poster presented at the 76th annual meeting of the Society of Biological Psychiatry, held virtually due to COVID-19. <https://github.com/JRam02/SOBPPoster.png>.

Ramduny, J., & Kelly, C. (2021). Shifting towards a naturalistic paradigm to investigate emotional processing in adolescent depression. Poster presented at the 26th biennial meeting of the British Neuroscience Association, held virtually due to COVID-19. <https://github.com/JRam02/BNAPoster.pdf>.

Ramduny, J., & Kelly, C. (2020). Optimising the individual functional connectome for neurodevelopmental and psychiatric disorders. Poster presented at the 12th annual meeting of the Neuroscience Ireland Young Investigator Symposium, held virtually due to COVID-19. <https://github.com/JRam02/NSIPoster.pdf>.

Ramduny, J., & Kelly, C. (2020). Refining the fingerprint: Optimising connectome fingerprinting for neurodevelopmental applications. Poster presented at the 26th annual meeting of the Organisation for Human Brain Mapping, held virtually due to COVID-19. https://github.com/JRam02/OHBM2020POSTER_TCD.pdf.

Ethics Statement

The study procedures in Chapters 2, 4, and 6, which used Open Science datasets such as the Consortium for Replicability and Reproducibility [CoRR], Healthy Brain Network [HBN], and Autistic Brain Imaging Data Exchange [ABIDE], were approved by the School of Psychology Research Ethics Committee, Trinity College Dublin, on September 12, 2017 (see Appendix A for the Ethics Approval Letter). All neuroimaging and phenotypic data were collected and openly shared in accordance with ethical approval from the relevant Ethics Committee or Institutional Review Board, as documented on the project websites: http://fcon_1000.projects.nitrc.org/indi/CoRR/html/samples.html; http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/; and <http://preprocessed-connectomes-project.org/abide/>.

The study procedures reported in Chapter 8, which examines data collected from a community-based sample of female adolescents experiencing subclinical depressive symptoms, were approved by the School of Psychology Research Ethics Committee, Trinity College Dublin, the Research Ethics Committee at Linn Dara Child & Adolescent Mental Health Service, and Saint James's Hospital-Tallaght University Hospital Joint Research Ethics Committee on October 28, 2015 (see Appendix B for the Ethics Approval Letter).

Data Sharing Statement

The neuroimaging data used in this thesis are primarily Open Science consortia-level datasets. The data examined in Chapter 8 were gathered from a community-based sample collected locally at Trinity College Institute of Neuroscience and Trinity College Dublin. The characteristics of each dataset (contributing sites, demographics, imaging parameters, and other relevant information) are described in more detail in the relevant chapters.

Analyses for Chapter 2 were conducted using five datasets from the Consortium for Replicability and Reproducibility [CoRR] which comprise the Beijing Normal University [BNU], New York University [NYUado, NYUadu], Southwest University [SWU], and University of Pittsburgh School of Medicine [UPSM]. The neuroimaging data are publicly available from the CoRR platform: http://fcon_1000.projects.nitrc.org/indi/CoRR/html/samples.html. I have shared the mean timeseries of the preprocessed resting-state fMRI data derived from the five datasets, which can be used to reproduce the analyses, via GitHub: https://github.com/JRam02/conn_fingerprintr_neurodev.

Chapter 4 used neuroimaging and phenotypic data from the Cornell Brain Imaging Center [CBIC] site in Releases 4-9 derived from Healthy Brain Network [HBN]: http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/. The data are publicly available to download via the LORIS and COINS platforms. I have shared the mean timeseries of the preprocessed resting-state fMRI data obtained from the CBIC site in Releases 4-9, which were used in the analyses, via GitHub: <https://github.com/JRam02/bagging>.

Chapter 6 used the Healthy Brain Network [HBN] data from the Cornell Brain Imaging Center [CBIC] site in Releases 4-10: http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/. The data are publicly available to download via the LORIS and COINS platforms. I have shared the mean timeseries of the preprocessed resting-state fMRI data obtained from the CBIC site in Releases 4-10 via GitHub: <https://github.com/JRam02/cpm>. Preprocessed neuroimaging and phenotypic data for non-Autistic participants including in the Autistic Brain Imaging Data Exchange [ABIDE] were also included to test generalisability: <http://preprocessed-connectomes-project.org/abide/>. To ensure reproducibility, I have shared the timeseries of the preprocessed resting-state fMRI data from the ABIDE dataset via GitHub: <https://github.com/JRam02/cpm>.

Analyses for Chapter 8 used neuroimaging and phenotypic data from a community-based sample of female adolescents experiencing subclinical depressive symptoms, recruited, assessed, and scanned at Trinity College Institute of Neuroscience in collaboration with Louise Gallagher

(Department of Psychiatry at the School of Medicine, Trinity College Dublin) and Ewa Goncerz. I have shared the mean timeseries of the preprocessed movie-watching fMRI data, which can be used to reproduce the analyses, via GitHub: <https://github.com/JRam02/naturalistic-analysis>.

Code Availability Statement

The structural and functional MRI data preprocessing for the empirical projects reported in Chapters 2, 4, 6, and 8 were performed using in-house neuroimaging scripts that used commands from AFNI (<https://afni.nimh.nih.gov>), FSL (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL>), and ANTs (<https://github.com/ANTsX/ANTs>). Structural and functional data preprocessing for Chapter 8 was performed using fMRIPrep version 1.5.0 (<https://fmriprep.org/en/stable/>; Esteban et al., 2019) in addition to in-house scripts.

Across the four research projects, the publicly available whole-brain Shen ($k = 268$) parcellation scheme (<https://identifiers.org/neurovault.image:395091>; Shen et al., 2013) was employed to extract mean fMRI timeseries. Several additional parcellation schemes were also used in Chapters 2 and 6, and they are described in the relevant empirical projects.

Analysis code was written in Python 3.7 and executed on either Google Colaboratory or Jupyter Notebook using Anaconda version 1.10.0. All code produced for this thesis is publicly accessible via GitHub repositories: <https://github.com/JRam02>.

Honours and Awards

2023	Society of Biological Psychiatry Award Predoctoral Travel Fellowship Award Amount: \$2,000
2022	Trinity College Dublin Postgraduate Teaching Award Nominated and shortlisted for Best Graduate Teaching Assistant for the academic year 2021/22
2022	Trinity College Dublin Trust Travel Grant OHBM Educational Symposium and Poster Presentation Amount: €200
2021	OHBM Open Science SIG Fellowship Teaching fellow for the OHBM 2021 Hackathon Amount: €300
2020	Neuroscience Ireland Young Investigator Symposium Best Poster Presentation
2019	Trinity College Dublin Provost Award Non-EU PhD Studentship awarded from 01/09/2019 to 31/08/2023 Amount: €130,000

List of Figures

Figure 1.1	(A) Functional brain organisation described by Smith et al. (2009). Visual network (Maps 1 ₂₀ – 3 ₂₀); Default Mode network (Map 4 ₂₀); Cerebellar network (Map 5 ₂₀); Sensorimotor network (Map 6 ₂₀); Auditory network (Map 7 ₂₀); Executive Control network (Map 8 ₂₀); and Frontoparietal network (Maps 9 ₂₀ – 10 ₂₀). Adapted from Smith et al., 2009. (B) Similar functional brain organisation has been described by the Yeo-7Network definition. Visual network; Somatomotor network; Dorsal Attention network; Ventral Attention network; Limbic network; Frontoparietal network; and Default Mode network. The seven colour codes correspond to the respective functional networks from the left hemisphere. Note: There is also a Yeo-17Network definition which further extends the Yeo-7Network scheme. Adapted from Yeo et al., 2011.....	39
Figure 1.2	Growth in the number of published functional connectivity-based studies in neurotypical and clinical populations since the first study published by Biswal and et al. (1995). Adapted from Zhang et al. (2021).....	41
Figure 1.3	The term reproducibility describes the properties of analytical reproducibility (i.e., same data, same methods), replicability (i.e., new data, same methods), and robustness to analytical flexibility (i.e., same data, new methods). Adapted from Botvinik-Nezer and Wager (2022).....	44
Figure 1.4	Barriers to reproducible findings across disciplines. Adapted from Munafò et al., 2017 and Chambers et al., 2014.....	45
Figure 1.5	Identification procedure as described by Finn et al. (2015). An individual is identified on the basis of their functional connectivity matrices derived between two imaging sessions (i.e., Database and Target Matrix). This matching procedure uses Pearson’s correlation coefficient such that an individual is successfully identified (ID = 1) if the correlation is maximal between the two functional connectivity matrices of that participant, otherwise that individual is misidentified (ID = 0). Adapted from Kauffmann et al., 2017.....	57
Figure 2.1	Analysis framework to describe an optimal pipeline for maximising connectome-based identification. <u>Data pre- and post-processing parameters</u> : A. Applying GSR as a denoising strategy. B. Describing the functional connectome with four publicly available parcellation schemes such as Shen 268, Glasser 360, MIST 444, and Schaefer 1000-17Networks. C. Describing cortical, subcortical, and cerebellar regions from the Harvard-Oxford and Buckner atlases. D. Computing the functional connectome using multi-resolution parcellation schemes with 8-10 levels of dimensionality. E. Examining the Finn Networks and Yeo Networks. F. Using the top 1-75% of most discriminative edges. G. Examining scan duration — from 0.5 min to the full acquisition time of each dataset. H. Examining the impact of sample size (N) by pooling data across the five datasets and resampling such that $N \in \{20, 40, 60, 80, 100, 120, 140, 160, 180, 200, 220, 240, 264\}$	72
Figure 2.2	(A) ID accuracy derived from BID in addition to the degree of confusion as indexed by RR for the database-target matrix configuration. (B) ID accuracy derived from BID in addition to the degree of confusion as indexed by RR for	

target matrix-database configuration. (C) ID accuracy and TRT reliability derived from the respective fingerprint approach using BID and the ICC(2,1) model. The Pearson's correlation coefficient (r) between ID accuracy and ICC is 0.82. (D). ID accuracy derived from BID with and without applying GSR as a denoising strategy. The five independent datasets are as follows: BNU ($N = 60$), NYUado ($N = 24$), NYUadu ($N = 31$), SWU ($N = 82$), and UPSM ($N = 67$).....74

Figure 2.3 (A) ID accuracy obtained from BID across the five independent datasets using four publicly available parcellation schemes including Shen 268, Glasser 360, MIST 444, and Schaefer 1000-17Networks. (B) Relationship between parcel resolution and ID accuracy measured by BID in addition to the degree of confusion as indexed by RR using 8 and 10 levels of parcel dimensionality derived from the Schaefer 17Networks and MIST parcellation schemes across the five independent datasets. (C) ID accuracy derived from BID based on the spatial coverage of the functional connectome covering cortical, subcortical, and cerebellar structures across the five independent datasets. Cor: Cortical; Cor+Sub: Cortical and Subcortical; Cor+Cer: Cortical and Cerebellum; Sub: Subcortical; Sub+Cer: Subcortical and Cerebellum; and Cer: Cerebellum.....79

Figure 2.4 (A) ID accuracy derived from BID for the Finn Networks in addition to the combined medial frontal and frontoparietal networks [FPNs] and whole-brain connectome [WBN] across the five datasets. The Finn Networks are as follows: (1) medial frontal [MFN]; (2) frontoparietal [FPN]; (3) default mode [DMN]; (4) subcortical [SUB]; (5) motor [MOT]; (6) visual I [VISI]; (7) visual II [VISII]; and (8) visual association [VISA]. (B) ID accuracy derived from BID for the Yeo Networks across the five datasets. The Yeo Networks are as follows: (1) ventral attention [VAN]; (2) dorsal attention [DAN]; (3) frontoparietal [FPN]; (4) default mode [DMN]; (5) somatomotor [SOM]; (6) limbic [LIM]; and (7) visual [VIS].....80

Figure 2.5 Relationship between number of discriminative edges and ID accuracy derived from BID in addition to the degree of confusion as indexed by RR across the five independent datasets. The discriminative edges are selected based on their DP values falling between the 25th and 99th percentiles of the whole-brain connectome, i.e., the top 1-75% of edges in the whole-brain connectome were retained to compute the ID accuracy of each dataset. The black box represents the ID accuracy derived from BID when all the edges in the whole-brain connectome were used for each dataset.....82

Figure 2.6 (A) Relationship between scan duration and ID accuracy measured by BID in addition to the degree of confusion as indexed by RR across the five independent datasets. The scan duration ranges from 0.5 min to the full acquisition time of each dataset. (B) Relationship between amount of resting-state fMRI data and ID accuracy measured by BID in addition to the degree of confusion as indexed by RR. The resting-state fMRI scans were aggregated across the five datasets ($N = 264$) for each imaging session, and then subsets of individuals were randomly selected across 100 iterations such that $N \in \{20, 40, 60, 80, 100, 120, 140, 160, 180, 200, 220, 240, 264\}$83

Figure 2.7 ID accuracy and TRT reliability of the functional connectome measured by BID and ICC, respectively, using the standard and optimal fingerprint-informed

pipelines. Dots and error bars represent the mean and SD of the BID accuracies and ICC values. Standard: GSR + Shen 268 parcellation + full scan length + full sample size. Pipeline 1: GSR + Schaefer 1000-17Network parcellation + full scan length + full sample size. Pipeline 2: GSR + frontoparietal networks derived from Finn networks + full scan length + full sample size. * $P < 0.05$, two-tailed paired sample t-test. For ID accuracy, the effect sizes measured by Cohen's d are: 1.5 (Pipeline 1) and 0.8 (Pipeline 2). For TRT reliability, the effect sizes measured by Cohen's d are: 1.0 (Pipeline 1) and 1.6 (Pipeline 2).....84

Figure 3.1 From Marek et al. (2022): Effect sizes and sampling variability of univariate brain-wide associations (BWAS) from the ABCD Study. (a) Effect sizes of brain-wide associations derived from cortical thickness and cognitive ability as well as general psychopathology at the vertex, ROI, and network levels as a function of sample size. (b) Effect sizes of brain-wide associations derived from resting-state functional connectivity and cognitive ability as well as general psychopathology at the vertex, ROI, and network levels as a function of sample size. (c) Largest brain-wide associations of cortical thickness with cognitive ability and general psychopathology spatially displayed at the ROI-level. (d) Largest brain-wide associations of resting-state functional connectivity with cognitive ability and general psychopathology spatially displayed at the ROI-level. (e) Sampling variability resamples per sample size from $N = 25$ to $N = 3,604$ for the largest brain-wide associations of cortical thickness with cognitive ability and general psychopathology. (f) Sampling variability resamples per sample size from $N = 25$ to $N = 3,928$ for the largest brain-wide associations of resting-state functional connectivity with cognitive ability and general psychopathology. (g) Examples of two $N = 25$ subsamples, where inaccurate default mode network (DMN) associations were observed for cortical thickness with cognitive ability and general psychopathology. (h) Examples of two $N = 25$ subsamples, where inaccurate default mode network (DMN) associations were observed for resting-state functional connectivity with cognitive ability and general psychopathology.....97

Figure 4.1 Schematic describing the scrubbing and bagging procedures. The Shen 268 parcellation scheme is applied to extract the fMRI timeseries of all the participants in the CoRR and HBN datasets. When scrubbing is performed, a motion-corrupted timepoint is identified in the fMRI timeseries if its rmsFD > 0.20 mm. After censoring that timepoint, its preceding and two succeeding timepoints are also removed in the timeseries. Bagging is performed using the scrubbed timeseries by (1) selecting the least motion-corrupted timepoints (ranked by their rmsFD values) that match a predefined threshold denoted by minTP to prevent participants from having varying number of timepoints which can be sampled from, and (2) bootstrapping samples of a given size TP from the motion-limited timepoints *with* replacement across 500 iterations. For each participant, the mean bagged functional connectivity matrix is computed by averaging the 500 functional connectivity matrices.....111

Figure 4.2 Applying bagging and one-sample bootstrap (minTP = 60) to perform connectome-based fingerprint identification derived from CoRR and HBN. Left Panel: Fingerprint accuracy measured by BID as a function of TP ranging from 60 to 194 by excluding “high-motion” participants with a rmsFD > 0.20 mm. Right Panel: Fingerprint accuracy measured by BID as a function of TP ranging from 60 to 375 by excluding “high-motion” participants with a rmsFD > 0.20

mm. Dotted black lines represent the BID accuracy obtained from the standard approach using the full timeseries of each dataset. Solid navy lines denote the BID accuracies obtained from bagging. Solid teal lines correspond to the BID accuracies obtained from one-sample bootstrap.....114

Figure 4.3 Correlations between functional connectivity and age obtained from REST1 using the standard approach as a function of sample size. Solid teal line shows the mean correlations derived from 500 bootstrap samples for a given sample size. Teal shading denotes the minimum and maximum correlations across 500 bootstrap samples for a given sample size. Black dotted line represents the lower and upper bounds of the 95% CIs. After applying a $\text{rmsFD} \leq 0.20$ mm to exclude the “high-movement” participants, the sample size was reduced to 255.....115

Figure 4.4 (A) Correlations between functional connectivity and age obtained from REST1 using the standard approach as a function of sample size. The sample size was set to 239 to facilitate the comparison of the standard approach with bagging. (B) Correlations between functional connectivity and age obtained from REST1 as a function of sample size using bagging during which participants have least amount of contaminated timepoints that match the minTP threshold (e.g., 200) and number of bagged timepoints, i.e., TP (e.g., 20, 60, 100). (C) Comparison of the 95% CIs between the standard approach and bagging with the area under curve (AUC). Dotted coloured line represents the 95% CIs derived from the standard approach. Solid coloured line represents the 95% CIs derived from bagging. (D) Correlations between functional connectivity and age obtained from REST1 as a function of sample size using one-sample bootstrap.....116

Figure 4.5 Correlations between functional connectivity and age obtained from REST1 as a function of sample size using bagging during which participants have minimally bootstrap timepoints that match the minTP threshold (e.g., 60) and number of bagged timepoints, i.e., TP (e.g., 20). Solid purple line shows the mean correlations, purple shading denotes the minimum and maximum correlations, and black dotted line represents the lower and upper bounds of the 95% CIs. The AUC is displayed to compare the 95% CIs between the standard approach and bagging. Dotted blue line represents the 95% CIs derived from the standard approach. Solid purple line represents the 95% CIs derived from bagging. The correlations between functional connectivity and age obtained from REST1 as a function of sample size is also shown using one-sample bootstrap.....117

Figure 4.6 (A) Influence of minTP using bagging on the width of the 95% CIs by keeping TP constant at 20 such that $\text{minTP} \in \{60, 100, 200\}$. (B) Influence of TP using bagging on the width of the 95% CIs by keeping minTP constant at 60 such that $\text{TP} \in \{20, 60, 100\}$118

Figure 4.7 Applying bagging and one-sample bootstrap ($\text{minTP} = 60$) to perform connectome-based fingerprint identification derived from CoRR and HBN. Left Panel: Fingerprint accuracy measured by BID as a function of TP ranging from 60 to 194 by including the salvaged “high-motion” participants. Right Panel: Fingerprint accuracy measured by BID as a function of TP ranging from 60 to 375 by including the salvaged “high-motion” participants. Dotted black lines represent the BID accuracy obtained from the standard approach using the full

timeseries of each dataset. Solid navy lines denote the BID accuracies obtained from bagging. Solid teal lines correspond to the BID accuracies obtained from one-sample bootstrap.....119

Figure 4.8 Salvaging “high-movement” participants to compute univariate brain-behaviour associations with bagging without applying the initial head motion threshold of $\text{rmsFD} \leq 0.20$ mm, i.e., the timeseries of all the participants were scrubbed with a $\text{rmsFD} \leq 0.2$ mm. By applying bagging with $\text{minTP} = 60$ and $\text{TP} = 20$, the sample size was increased to 379. The correlations between functional connectivity and age obtained from REST1 as a function of sample size are shown using bagging and one-sample bootstrap. The AUC is displayed to compare the 95% CIs between the standard approach and bagging in addition to the standard approach and one-sample bootstrap. Solid pink line shows the mean correlations, pink shading denotes the minimum and maximum correlations, and black dotted line represents the lower and upper bounds of the 95% CIs.....119

Figure 4.9 Impact of applying a strict head motion threshold on brain-behaviour associations using bagging ($\text{minTP} = 60$ and $\text{TP} = 20$). (A) After applying a $\text{rmsFD} \leq 0.08$ mm to exclude “high-movement” participants, the sample size was reduced to 77. Correlations between functional connectivity and age obtained from REST1 as a function of sample size are shown using bagging across 500 and one-sample bootstrap. (B) After scrubbing the timeseries with a $\text{rmsFD} \leq 0.08$ mm of all the participants, the sample size was increased to 92. Correlations between functional connectivity and age obtained from REST1 as a function of sample size are shown using bagging across 500 and one-sample bootstrap. The AUC is displayed to compare the 95% CIs derived from bagging across 500 and one-sample bootstrap. Solid green line shows the mean correlations, green shading denotes the minimum and maximum correlations, and black dotted line represents the lower and upper bounds of the 95% CIs.....121

Figure 4.10 Reproducibility of the standard approach in deriving the associations between functional connectivity and age from an additional 9 edges in the connectome. Correlations between functional connectivity and age obtained from REST1 as a function of sample size are shown using the standard approach across the 9 edges. The AUCs are displayed to compare the 95% CIs derived from the standard approach across the 9 edges. Solid coloured lines show the mean correlations, coloured shadings denote the minimum and maximum correlations, and black dotted lines represent the lower and upper bounds of the 95% CIs. After applying a $\text{rmsFD} \leq 0.20$ mm to exclude the “high-movement” participants, the sample size was reduced to 255.....122

Figure 4.11 Reproducibility of bagging ($\text{minTP} = 60$ and $\text{TP} = 20$) including the salvaged “high-movement” participants in deriving the associations between functional connectivity and age from an additional 9 edges in the connectome. Correlations between functional connectivity and age obtained from REST1 as a function of sample size are shown using bagging across the 9 edges. The AUCs are displayed to compare the 95% CIs derived from bagging across the 9 edges. Solid coloured lines show the mean correlations, coloured shadings denote the minimum and maximum correlations, and black dotted lines represent the lower and upper bounds of the 95% CIs. After rescuing the “high-movement” participants, the sample size increased to 379.....123

Figure 5.1	Connectome-based predictive modelling (CPM) framework which involves feature selection, feature summarisation, model building, and assessment of the prediction significance. Adapted from Shen et al., 2017.....	130
Figure 6.1	Phenotypic distributions across the complete HBN dataset. Histograms showing the distributions of age [in years], general psychopathology [CBCL], and intellectual ability [FSIQ] in developing youths.....	144
Figure 6.2	Data pre-/post-processing factors and head motion mitigation strategies for connectome-based fingerprint identification. Data pre-/post-processing factors include the application of four publicly available <u>parcellation schemes</u> such as the Glasser 360, Glasser 426, MIST 444, and Schaefer 500-17Networks atlases in addition to using the <u>frontoparietal networks</u> derived from a 10-network definition to derive the REST1 and REST2 functional connectivity matrices of the participants. Head motion mitigation strategies include <u>scrubbing</u> and <u>bagging</u> to compute the REST1 and REST2 functional connectivity matrices of the participants on the basis of their least motion-corrupted timepoints in the fMRI timeseries.....	150
Figure 6.3	Data pre-/post-processing factors (i.e., frontoparietal networks, parcellation schemes) for connectome-based fingerprint identification. A. Connectome-based fingerprint accuracy derived from the frontoparietal networks using a 10-network definition which involves the combination of the medial frontal and frontoparietal networks. B. Connectome-based fingerprint accuracy derived from multiple parcellation schemes that differ in terms of their parcel dimensionality (i.e., 268 → 500 parcels), space (i.e., volume, surface), and spatial coverage (i.e., cortical, whole-brain).....	153
Figure 6.4	Head motion mitigation strategies (i.e., scrubbing, bagging) for connectome-based fingerprint identification. A. Connectome-based fingerprint accuracy derived from scrubbing where a motion-corrupted timepoint (T) is identified and censored if its rmsFD > 0.20 mm followed by a preceding (T – 1) and two succeeding timepoints (T + 1, T + 2). B. Connectome-based fingerprint accuracy derived from bagging with and without the inclusion of “high-motion” participants, where the least 60 or 200 motion-corrupted timepoints are selected from each participant’s fMRI timeseries out of which 20 or 100 timepoints are respectively bootstrapped across 500 iterations.....	154
Figure 6.5	Neuroanatomy of the CPM-based predictive networks for age, CBCL, and FSIQ. <u>First Panel</u> : Relationships between predicted and observed phenotypes derived from the respective POS and NEG network strengths. <u>Second Panel</u> : Chord plots illustrating the contributions of the predictive POS and NEG edges in different functional networks using a 10-network definition. Note: To facilitate the post-hoc visualisations, the POS and NEG edges in predicting age were thresholded with a degree of 20. <u>Third Panel</u> : Glass brains illustrating the connectivity patterns across the whole brain for the POS and NEG network strengths. <u>Fourth Panel</u> : Connectivity matrices summarising the connectivity patterns within and between different functional networks based on their POS and NEG network strengths. The shades represent the degree of edge contributions such that darker shades indicate greater contribution of the edges in a functional network. MF: Medial Frontal; FP: Frontoparietal; DMN: Default	

Mode; Mot: Motor; VI: Visual I; VII: Visual II; VAs: Visual association; SAL: Salience; SC: Subcortical; CBL: Cerebellum.....156

Figure 6.6	Generalisability of baseline CPM-based predictions of age and FSIQ in ABIDE. <u>First Panel</u> : Relationships between predicted and observed age derived from the respective POS, NEG, and Combined network strengths. <u>Second Panel</u> : Relationships between predicted and observed FSIQ derived from the respective POS, NEG, and Combined network strengths. predicted and observed phenotypes. r_p = positive network strength, r_n = negative network strength, r_c = combined network strength.....160
Figure 7.1	A bottom-up representation of the layers that are involved in narratives. The features of a narration can reflect audio signals, words, sentences, interactions, characters, events, and story-telling. Adapted from Willems et al., 2020.....170
Figure 7.2	Naturalistic stimuli derived from three different movies including “Inscapes” (Vanderwal et al., 2015), “The Present” (Alexander et al., 2017), and “Ocean’s Eleven”. Top Panel: Inscapes (http://headspacestudios.org/inscapes) is a 7-min non-verbal, non-social series of slowly evolving abstract shapes with a piano score sound directed by Tamara Vanderwal. Middle Panel: The Present (https://www.jacob-frey.com) is a 4.18-min animated movie directed by Jacob Frey that portrays an amputated youth being gifted a disabled dog by his mother. Bottom Panel: Ocean’s Eleven is a 116-min Hollywood movie with an ensemble cast that covers a heist comedy plot. The movie was directed by Steven Soderbergh and released in 2001.....172
Figure 7.3	Illustration of ISC to quantify shared brain synchrony derived from the fMRI timeseries of a voxel x from a given participant A with the timeseries of voxel x in participant B. Adapted from Nastase et al., 2019.....174
Figure 8.1	Phenotypic distributions. Histograms showing the distributions of the depressive symptom severity derived from Depression subscale of the Beck Youth Inventory (BYI; the Depression Inventory, BYI-D), adaptive cognitive emotion regulation (Adaptive CERQ) scores and maladaptive cognitive emotion regulation (Maladaptive CERQ) scores. The skewed distributions of the phenotypic measures indicate that they are not normally distributed (Shapiro-Wilk test, $P < 0.05$).....189
Figure 8.2	Movie stills obtained from The Present (https://www.jacob-frey.com).....191
Figure 8.3	Movie-watching analysis framework at multiple spatial and temporal scales. First, whole-brain full-length intersubject correlation (ISC) was performed by calculating the pairwise correlations between participant A and the remaining participants B, C, and so on such that $A \neq B, C$. Second, dynamic intersubject correlation (dISC) was carried out using a tapered cosine sliding window with a width of $W = 20$ TRs and step size of $W = 1$ TR (tapered cosine fraction = 0.95) to capture time-varying patterns of brain synchrony throughout the movie. Temporal ICA was then applied to extract functional networks that capture dynamic patterns of synchronised brain activities over the course of the movie. Third, functional connectivity (FC) was performed between <i>a priori</i> ROIs that

form part of an emotion regulation network derived from NeuroSynth. Fourth, dynamic functional connectivity (dFC) was derived using the same tapered cosine sliding window procedure as previously described to compute the FC-phenotypic correlations over the course of the movie. Lastly, the Hurst Exponent (H) was employed to index the self-similarity of the BOLD fMRI signal at the network level using the Yeo-7Network parcellation scheme and ROI level using the predefined ROIs that are involved in the emotion regulation network. Subsequently, the relationships between each of these measures and phenotypic scores (i.e., BYI-D, adaptive CERQ, maladaptive CERQ) were performed.....196

Figure 8.4 Intersubject synchronisation measured by full-length ISC. (A) Full-length ISC during the natural viewing of an emotionally evocative movie. (B) Full-length ISC during high-emotion segment relative to low-emotion segment of the movie. The ISC maps are corrected for multiple comparisons thresholded at $P < 0.05$ FDR.....198

Figure 8.5 Relationships between intersubject synchronisation and phenotypic measures during emotional movie viewing. (A) Relationships between full-length ISC and self-reported depressive symptom severity (BYI-D; Top Panel), adaptive emotion regulation (Adaptive CERQ; Middle Panel) and maladaptive emotion regulation (Maladaptive CERQ; Bottom Panel). The relationships are displayed at an uncorrected threshold of $P < 0.01$. (B) Histograms illustrate the permutation test null distributions and observed mean ρ for relationships between the phenotypic measures and global ISC averaged across the whole brain (i.e., 268 parcels).....199

Figure 8.6 Temporal ICA shows several functional networks that capture dynamic intersubject synchronisation. (A) The percentage of explained variance in each IC is displayed in descending order than ranges from 4% to 9%. (B) The functional networks that correspond to each IC during naturalistic emotional viewing. The Z-statistic maps of the ICs are thresholded at $Z > 2$200

Figure 8.7 Timecourse of the functional networks derived during emotional movie watching. The x-axis represents the tapered sliding windows such that each window has a width of $W = 20$ TRs with a step size of $T = 1$ TR. The y-axis illustrates the normalised response during each sliding window as measured by dISC. Solid black curves represent the timecourse of the functional networks. Dotted black line denote the mean dISC across the functional networks. Purple shades indicate the functional networks that exhibit idiosyncratic responses during high-emotion segment of the movie. Orange shades correspond to the functional networks that exhibit high synchrony during high-emotion segment of the movie. Some scenes from the high-emotion segment that relate to the timecourse of the functional networks are also displayed in the top panel. A Mann-Whitney U test was performed to quantify the increases and decreases in brain synchrony within each functional network throughout the movie.....202

Figure 8.8 Edge-level correlations between dFC and phenotypic measures that persisted across at least 5 consecutive sliding window during the emotionally evocative movie. (A) Edges in consecutive windows that showed a significant relationship between dFC and BYI-D score. (B) Edges in consecutive windows that showed

a significant relationship between dFC and adaptive CERQ score. (C) Edges in consecutive windows that showed a significant relationship between dFC and maladaptive CERQ score. Red line denotes the positive association between dFC and phenotypic measures derived from Spearman's ρ . Blue line represents the negative association between dFC and phenotypic measures derived from Spearman's ρ . The number of consecutive edges that captured the significant relationships between dFC and phenotypic measures is also displayed under each edge. The significance of these relationships is thresholded at $P < 0.01$204

Figure 8.9 Relationships between H and phenotypic measures at the network and ROI level. (A) Distributions of H derived from the functional networks based on the Yeo-7Network parcellation scheme. (B) Distributions of H derived from the ROIs that form part of the emotional regulation network. (C) Top Panel: There is a significant positive association between H of the VAN and depressive symptom severity. There is a significant positive association between H of the vlPFC and depressive symptom severity. Bottom Panel: There is a significant negative relationship between H of the left vlPFC and adaptive emotional regulation strategies. The H of the left vlPFC is positively related to maladaptive emotion regulation strategies but this relationship did not survive statistical significance. All the associations between H and phenotypic measures are corrected for multiple comparisons at the network and ROI level using an FDR threshold set at $P < 0.05$. Amyg: amygdala; OFC: orbitofrontal cortex, vlPFC: ventrolateral prefrontal cortex; dlPMC: dorsolateral premotor cortex; dlPFC: dorsolateral prefrontal cortex; aPFC: anterior prefrontal cortex; rACC: rostral anterior cingulate cortex; aINS: anterior insula. VAN: ventral attention network; DAN: dorsal attention network; FPN: frontoparietal network; DMN: default mode network; SMN: somatomotor network; LIM: limbic; VIS: visual; SUB: subcortical.....206

Figure 9.1 Overview of the functional connectivity-based frameworks used in this thesis: functional connectome fingerprinting, bagging, connectome-based predictive modelling, and the application of a naturalistic (e.g., movie-watching) paradigm.....216

Figure 9.2 Inter-rater reliability of diagnostic scales from the DSM-5 measured by the Kappa statistic. The Kappa statistic can be interpreted as follows: ≤ 0 indicates no agreement and 0.01-0.20 as none to minimal agreement, 0.21-0.40 as fair agreement, 0.41-0.60 as moderate agreement, 0.61-0.80 as substantial agreement, and 0.91-1 as almost perfect agreement. Adapted from Freedman et al., 2013.....223

Figure 9.3 Subtle manipulations of the functional connectome by introducing “enhancement attacks” in the CPM to predict a chosen behavioural phenotype such as IQ obtained from the HCP. Model performance is quantified by the Pearson's correlation coefficient (r) and cross-validated R^2 denoted by q^2 between observed and predicted IQ. MF: medial frontal; FP: frontoparietal; DMN: default mode; MOT: motor; VI: visual I; VII: visual II; VAs: visual associations; SAL: salience; SC: subcortical; CBL: cerebellum. Adapted from Rosenblatt et al., 2023.....229

Figure 9.4	Discriminability, TRT Reliability (ICC), and Fingerprinting (FP) of the functional connectome. Left Panel: First scenario where FP reveals a low ID accuracy of 30%, ICC shows a “good” reliability of 60%, and Discriminability is high with a value of 84%. Middle Panel: Second scenario where both Discriminability and fingerprint ID accuracy are optimal (100%), but the ICC shows a “good” reliability of 62%. Right Panel: Third scenario where no individual has been successfully identified, i.e., fingerprint ID accuracy = 0%, but the TRT reliability (ICC = 57%) and Discriminability (77%) are fair/moderate. Adapted from Milham et al., 2021.....	230
Figure 9.5	Factors capturing the rate of developmental changes such demographics, genetics, environment, and imaging modalities. Adapted from Cao et al., 2017.....	232
Figure A.1	Bagging framework which applies a bootstrap sampling strategy to create an ensemble model.....	289

List of Tables

Table 2.1	Demographic information and fMRI parameters for the five datasets obtained from CoRR that were used in this study. All sites contributed two resting-state scans per participant.....	68
Table 2.2	Gains in ICC and ID accuracy derived from BID across the five datasets obtained from the standard and optimal fingerprint-informed pipelines. Standard: GSR + Shen 268 parcellation + full scan length + full sample size. *Pipeline 1: GSR + Schaefer 1000-17Network parcellation + full scan length + full sample size. †Pipeline 2: GSR + frontoparietal networks derived from Finn Networks + full scan length + full sample size.....	78
Table 6.1	Demographic characteristics and resting-state fMRI acquisition parameters of the HBN that were used in this study. CBIC = Cornell Brain Imaging Center. RMS FD = root-mean-square framewise displacement. Note: The CBCL scores derived from the HBN reflect externalising behaviour that have been T-standardised.....	143
Table 6.2	Properties of the parcellation schemes that were employed to derive the REST1 and REST2 functional connectivity matrices for the connectome-based identification procedure. The parcellation schemes differ in parcel dimensionality (i.e., 268 → 500 parcels), space (i.e., volume, surface), and spatial coverage (i.e., cortex, whole-brain). *Baseline connectome-based fingerprint identification was performed using the whole-brain Shen 268 parcellation scheme.....	146
Table 6.3	Baseline connectome-based predictive modelling of age, CBCL, and FSIQ. The three phenotypes were predicted using the positive [POS], negative [NEG], and combined [Combined] network strengths. Model performance assessed the linear relationships between predicted and observed phenotypes using Pearson's correlation coefficient (r), explained variance (R^2), and mean absolute error (MAE). *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$	155
Table 6.4	CPM configurations derived from data pre-/post-processing factors and head motion mitigation strategies for predicting age, CBCL, and FSIQ. The data pre-/post-processing factors involve using <u>predictive networks</u> and several <u>parcellation schemes</u> that differ in terms of their parcel dimensionality, space, and spatial coverage. The head motion mitigation strategies include <u>scrubbing</u> and <u>bagging</u> . Here, model performance is indicated by the Pearson's correlation coefficient (r) between predicted and observed phenotypes. r_p = positive network strength, r_n = negative network strength, r_c = combined network strength. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$	158
Table 6.5	CPM configurations derived from data pre-/post-processing factors and head motion mitigation strategies for predicting age, CBCL, and FSIQ. The data pre-/post-processing factors involve using <u>predictive networks</u> and several <u>parcellation schemes</u> that differ in terms of their parcel dimensionality, space, and spatial coverage. The head motion mitigation strategies include <u>scrubbing</u> and <u>bagging</u> . Here, the Steiger's Z tests are presented for each CPM	

configuration and network strength to compare the baseline HBN-derived predictions of age, CBCL, and FSIQ with those obtained from the data pre-/post-processing factors and head motion mitigation strategies. Note: Steiger's Z test was not performed to compare the baseline CPM performance with that derived from bagging as the sample sizes are not equal: the assumption of the test relies on computing the statistical significance of the difference between two dependent correlations with the same sample size. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$159

Table 6.6	CPM configurations derived from data pre-/post-processing factors and head motion mitigation strategies for predicting age and FSIQ in ABIDE. The data pre-/post-processing factors involve using <u>predictive networks</u> and several <u>parcellation schemes</u> that differ in terms of their parcel dimensionality, space, and spatial coverage. The head motion mitigation strategies include <u>scrubbing</u> and <u>bagging</u> . Here, model performance is indicated by the Pearson's correlation coefficient (r) between predicted and observed phenotypes. The Steiger's Z tests are also computed for each CPM configuration and network strength to compare the baseline ABIDE-derived predictions of age and FSIQ with those obtained from the data pre-/post-processing factors and head motion mitigation strategies. Note: Steiger's Z test was not performed to compare the baseline CPM performance with that derived from bagging as the sample sizes are not equal: the assumption of the test relies on computing the statistical significance of the difference between two dependent correlations with the same sample size. r_p = positive network strength, r_n = negative network strength, r_c = combined network strength. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$161
Table 8.1	Demographic Characteristics. CERQ = Cognitive Emotion Regulation Questionnaire (*Maladaptive Strategy; ‡Adaptive Strategy). BYI-D = Depressive Inventory from Beck Youth Inventory. FD = Framewise Displacement.....189
Table 8.2	Relationships between H and phenotypic measures at the network level after correcting for multiple comparisons using an FDR threshold set at $P < 0.05$207
Table 8.3	Relationships between H and phenotypic measures at the ROI level after correcting for multiple comparisons using an FDR threshold set at $P < 0.05$207
Table A.1	Recommendations for describing an optimal pipeline to reliably measure the functional connectome at the individual level. The recommendations follow the underlying factors that influence fingerprint accuracy at the data pre- and post-processing stages as well as during the data acquisition phase.....287

List of Abbreviations

ABCD	Adolescent Brain Cognitive and Development
ABIDE	Autism Brain Imaging Data Exchange
ADHD	Attention-deficit-hyperactivity disorder
ASD	Autism spectrum disorder
AUC	Area Under Curve
BIDS	Brain Imaging Data Structure
BOLD	Blood-oxygenation-level-dependent
BWAS	Brain-wide association studies
BYI-D	Depressive Inventory of Beck Youth Inventory
CCA	Canonical Correlation Analysis
CERQ	Cognitive Emotion Regulation Questionnaire
CI	Confidence Interval
CoRR	Consortium for Replicability and Reproducibility
CPM	Connectome-based Predictive Modelling
CV	Cross Validation
E:I	Excitation-Inhibition
FAIR	Findability, Accessibility, Interoperability, Reusability
FCP	Functional Connectome Project
fMRI	Functional magnetic resonance imaging
FD	Framewise Displacement
GSR	Global Signal Regression
H	Hurst Exponent
HARKing	Hypothesis After Results are Known
HBN	Healthy Brain Network
HCP	Human Connectome Project
ICA	Independent Component Analysis
ICC	Intraclass Correlation Coefficient
ISC	Intersubject Correlation
MAE	Mean Absolute Error
MDD	Major depressive disorder
mPFC	Medial prefrontal cortex
NHST	Null Hypothesis Significance Testing
NIH	National Institutes of Health
NIMH	National Institute of Mental Health

PCA	Principal Component Analysis
RP	Realignment Parameter
RDoC	Research Domain Criteria
SNR	Signal-to-noise ratio
SVR	Support Vector Regression
TRT	Test-Retest
VAN	Ventral attention network
vlPFC	Ventrolateral prefrontal cortex
vmPFC	Ventromedial prefrontal cortex

Contents

Summary	iii
Acknowledgment.....	vi
List of Manuscripts in Preparation, Preprints & Publications	viii
List of Invited Talks.....	ix
List of Conference Posters	x
Ethics Statement	xii
Data Sharing Statement	xiii
Code Availability Statement	xv
Honours and Awards.....	xvi
List of Figures	xvii
List of Tables	xxvii
List of Abbreviations	xxix
Chapter 1	36
1.1 Introduction.....	38
1.2 Development of Functional Networks Across the Lifespan.....	40
1.3 Clinical Relevance of Functional Connectivity in Neurodevelopmental and Psychiatric Conditions	40
1.4 Robustness and Reproducibility of the Functional Connectome.....	43
1.5 Sample Size, Poor Statistical Power, Methodological Flexibility	46
1.6 Measuring Reproducibility: Test-Retest Reliability	48
1.6.1 Assessing the Reliability of the Functional Connectome During Adulthood	50
1.6.2 Assessing the Reliability of the Functional Connectome Across Childhood and Adolescence.....	52
1.6.3 Factors Driving TRT Reliability of the Functional Connectome	53
1.6.4 Limitations of ICC in Evaluating the Reliability of the Functional Connectome	55
1.7 Emergence of Functional Connectome Fingerprinting.....	56
1.7.1 Individual Differences in the Functional Connectome derived from Predictive Networks.....	57
1.7.2 Age and Gender Related Dependencies on the Individual Functional Connectome	58
1.8 Objectives and Hypotheses	60
Chapter 2	63
2.1 Introduction.....	66
2.2 Materials & Methods	68

2.2.1	Datasets.....	68
2.2.2	Preprocessing.....	68
2.2.3	Data Pre- and Post-Processing Parameters.....	69
2.2.4	Functional Connectivity Matrices — Fingerprints.....	71
2.2.5	Identification Procedure	71
2.2.6	TRT Reliability.....	71
2.2.7	Comparing Analysis Pipelines Using ID Accuracy vs. ICC	72
2.3	Results.....	73
2.3.1	Identifiability of the individual functional connectome	73
2.3.2	TRT reliability of the functional connectome is consistently poor	73
2.3.3	Applying GSR as a denoising strategy improves identification accuracy	73
2.3.4	Finer cortical parcellations yield stronger identification accuracy.....	74
2.3.5	Identification accuracy across brain networks	75
2.3.6	Greater number of discriminative edges yields stronger identification accuracy	76
2.3.7	Longer scan duration facilitates stronger identification accuracy	76
2.3.8	Increasing sample size does not necessarily improve identification accuracy	77
2.3.9	Selecting a fingerprint-optimised pipeline	77
2.3.10	Applying functional connectome fingerprinting informed pipelines to improve ICC	78
2.4	Discussion.....	84
2.5	Conclusions.....	89
Chapter 3		91
3.1	The Big Data Revolution	93
3.2	Considerations of Sample Size and Amount of Data for Reproducible Brain-Behaviour Relationships.....	96
3.3	Impact of Head Motion on Functional Connectivity in Youths and Adults.....	99
Chapter 4		102
4.1	Introduction.....	105
4.2	Materials & Methods	107
4.2.1	Participants	107
4.2.2	HBN MRI Data Acquisition.....	108
4.2.3	MRI Preprocessing Pipeline	108
4.2.4	Head Motion.....	108
4.2.5	Parcellation Scheme	109
4.2.6	Functional Connectivity: Standard Approach	109
4.2.7	Functional Connectivity: Scrubbing.....	109
4.2.8	Functional Connectivity: Bagging Approach.....	110

4.2.9	Functional Connectome Fingerprinting.....	111
4.2.10	Brain-behaviour Relationships: Standard and Bagging Approaches	112
4.2.11	Salvaging high-movement participants for connectome-based fingerprinting and brain-behaviour relationships	112
4.3	Results.....	113
4.3.1	Applying bagging to perform connectome-based fingerprint identification	113
4.3.2	Standard brain-behaviour relationships as a function of sample size.....	114
4.3.3	Bagged brain-behaviour relationships as a function of sample size.....	115
4.3.4	Robust brain-behaviour associations with minimally bootstrap timepoints.....	117
4.3.5	Influence of minTP and TP on bagged brain-behaviour relationships	117
4.3.6	Applying bagging to perform connectome-based fingerprint identification with “high-movement” individuals.....	118
4.3.7	Salvaging “high-movement” participants for brain-behaviour associations	119
4.3.8	Reproducibility of bagging across head motion thresholds and edges in the functional connectome.....	120
4.4	Discussion.....	123
4.5	Conclusions.....	127
Chapter 5		128
5.1	Principles of Connectome-based Predictive Modelling.....	130
5.2	Predicting Behavioural Phenotypes with Individual Functional Connectome	131
5.3	Methodological Considerations for Robust Prediction of Individual Behaviours.....	133
5.4	Towards Reproducible CPM-based Predictions	137
Chapter 6		138
6.1	Introduction.....	141
6.2	Materials & Methods	142
6.2.1	Participants	142
6.2.2	Phenotypic assessment of general psychopathology and intellectual ability	143
6.2.3	MRI Data Acquisition	144
6.2.4	MRI Preprocessing Pipeline	144
6.2.5	Functional Connectivity Matrices	145
6.2.6	Connectome-based Fingerprinting	145
6.2.7	Connectome-based Fingerprinting Pipeline	145
6.2.8	Connectome-based Predictive Modelling.....	147
6.2.9	External Validation of the CPM-based Predictions.....	149
6.3	Results.....	150
6.3.1	Connectome-based identifiability of developing youths	150

6.3.2	Factors influencing connectome-based fingerprint identification in developing youths	151
6.3.3	Connectome-based predictions of age, CBCL, and FSIQ in developing youths	154
6.3.4	Neuroanatomy of the CPM-based predictive networks.....	155
6.3.5	Impact of methodological configurations on CPM-based predictions	156
6.3.6	Generalisability of connectome-based predictions in developing populations ..	159
6.4	Discussion	162
6.5	Conclusions.....	166
Chapter 7		168
7.1	Emergence of Naturalistic fMRI Paradigms.....	170
7.2	Understanding Brain Synchronisation, Functional Connectivity, and Scale-free Dynamics of the fMRI Signal Using Naturalistic Paradigms	172
7.2.1	Functional Brain Synchronisation	173
7.2.2	Naturalistic Functional Connectome	176
7.2.3	Scale-free Properties of the fMRI BOLD Signal	178
Chapter 8		182
8.1	Introduction.....	185
8.2	Methods & Materials	188
8.2.1	Participants	188
8.2.2	Self-Reported Depressive Symptomatology	189
8.2.3	Self-Reported Emotion Regulation Strategies.....	190
8.2.4	Naturalistic Stimulus	190
8.2.5	MRI Data Acquisition	191
8.2.6	MRI Preprocessing Pipeline	191
8.2.7	Intersubject Correlation Analyses	191
8.2.8	Functional Connectivity Analyses.....	194
8.2.9	Hurst Exponent Analyses	195
8.3	Results.....	197
8.3.1	Naturalistic emotional viewing reveals whole-brain intersubject synchronisation..	197
8.3.2	Brain synchrony is related to depressive symptom severity and emotion regulation strategies during the emotionally evocative movie	198
8.3.3	Temporal ICA captures dynamic intersubject synchronisation in several functional networks during naturalistic emotional viewing	199
8.3.4	High emotion content evokes idiosyncratic responses during movie-watching	201

8.3.5	Moment-to-moment functional connectivity shows edge-level correlations with self-reported depressive symptomatology and emotion regulation strategies during emotional movie-watching	203
8.3.6	Self-similarity signal structure of the ventral attention network and ventrolateral prefrontal cortex is associated with depressive symptom severity and adaptive emotional regulation strategies during naturalistic emotional viewing.....	205
8.4	Discussion	207
8.4.1	Shared brain responses in adolescents experiencing depressive symptoms during free emotional viewing	208
8.4.2	Temporal dynamics of brain synchrony in functionally defined networks	209
8.4.3	Dynamic functional connectivity in emotional processing regions during naturalistic movie stimuli	210
8.4.4	Self-similarity signal structure of the ventral attention network and ventrolateral prefrontal cortex in adolescents experiencing subclinical depressive symptoms during movie-watching.....	211
8.4.5	Limitations and future directions.....	211
8.5	Conclusions.....	212
Chapter 9		214
9.1	Contributions of Empirical Findings	216
9.2	Consortium-level versus Typical Sample Sizes.....	218
9.3	Reliable Measurement of Phenotypes.....	221
9.4	Trade-offs of Resting-State fMRI.....	224
9.5	Methodological Variability in fMRI Pipelines	226
9.6	Rate of Developmental Changes.....	231
9.7	A Quest for Multimodal MRI Applications.....	233
9.8	Limitations	234
9.9	Future Directions	236
9.10	Conclusions.....	239
References		242
Appendices		284

Chapter 1

Introduction

Summary

Chapter 1 provides an overview of the background and motivations for this thesis. I start by describing how resting-state fMRI can be used to study functional brain organisation across development, lifespan, and clinical conditions, and explain why the approach has become central to the search for brain-based biomarkers of neurodevelopmental and psychiatric conditions. Next, I discuss the importance of reproducibility (encompassing analytic reproducibility, replicability, and robustness to analytical variability) to that search, and some of the obstacles to identifying reproducible functional connectivity-based brain-behaviour relationships. I review the literature on test-retest reliability of the functional connectome in the context of neurodevelopmental and psychiatric conditions, and identify factors affecting reliability over time. Next, I review the literature on functional connectome fingerprinting, and outline how this approach offers an alternative way to assess the reproducibility of functional connectivity estimates. Finally, the primary objectives and hypotheses of this thesis are elaborated.

1.1 Introduction

Over the past two decades, task-free, or “resting-state” fMRI has risen to prominence as a non-invasive technique for the investigation of the macroscopic organisation of the human brain (Smith et al., 2009; Damoiseaux et al., 2006; Biswal et al., 1995). A plethora of fMRI studies has shown that spontaneous or intrinsic low frequency activity is temporally correlated within distinct functional networks when participants are at rest (Smith et al., 2013; Biswal et al., 2010; Biswal et al., 1995) — these correlations are referred to as “functional connectivity”. The networks that exhibit functional connectivity at rest have been shown to recapitulate patterns of activity observed during task performance across a wide range of functional domains, including cognitive (Petersen et al., 2015; Bressler et al., 2010), attention (Vossel et al., 2014; Fox et al., 2006; Corbetta et al., 2002), motor (Biswal et al., 1995), language (Li et al., 2020; Bernal et al., 2015; Tomasi et al., 2012), and memory (Vincent et al., 2006) functions.

These “resting-state” or intrinsic connectivity networks are highly reproducible across large population samples (e.g., Yeo et al., 2011; Smith et al., 2009; Damoiseaux et al., 2006). Further research has also detected analogous networks across non-human species (Gorges et al., 2017; Pan et al., 2015; Lu et al., 2012; Hutchison et al., 2012; Margulies et al., 2009; Vincent et al., 2007), suggesting that they reflect a fundamental organisational principle of the brain. Across human studies, a set of canonical networks has been identified (Figure 1.1) that typically includes: (I) the *default mode* network, which includes ventral medial prefrontal cortex, dorsal medial prefrontal cortex, posterior cingulate cortex and precuneus, and lateral parietal cortex, and is thought to integrate current perception and cognition, guided by memory representations (Murphy et al., 2019; Margulies et al., 2016; Raichle et al., 2001); (II) the *frontoparietal* network, comprising lateral prefrontal cortex and posterior parietal cortex, which supports cognitive control abilities (Cole et al., 2013; Zanto et al., 2013; Dosenbach et al., 2008); (III) the *dorsal attention* network, which consists of the intraparietal sulcus and frontal eye fields, and is involved in top-down goal-directed processing (Rohr et al., 2016; Corbetta et al., 2002); (IV) the *ventral attention* network, which includes the temporoparietal junction and ventral frontal cortex, and which is active during the presentation of unexpected stimuli that are behaviourally relevant (Vossel et al., 2014; Fox et al., 2006); (V) the *sensorimotor* network, which consists of somatosensory (i.e., postcentral gyrus), motor (i.e., precentral gyrus), and supplementary motor areas, which supports motor and somatosensory functioning and learning (Vahdat et al., 2011; Albert et al., 2009); (VI) the *limbic* network, comprising limbic cortex, hippocampal formation, amygdala, septal area, and hypothalamus, which guides motivational, emotional, and olfactory processes in addition to long-term memory formation (Rajmohan et al., 2007); and (VII) the

visual network, including the primary visual cortex and portions of the temporal and occipital lobes, which supports visual processing and perception (Smith et al., 2009; Wilms et al., 2013).

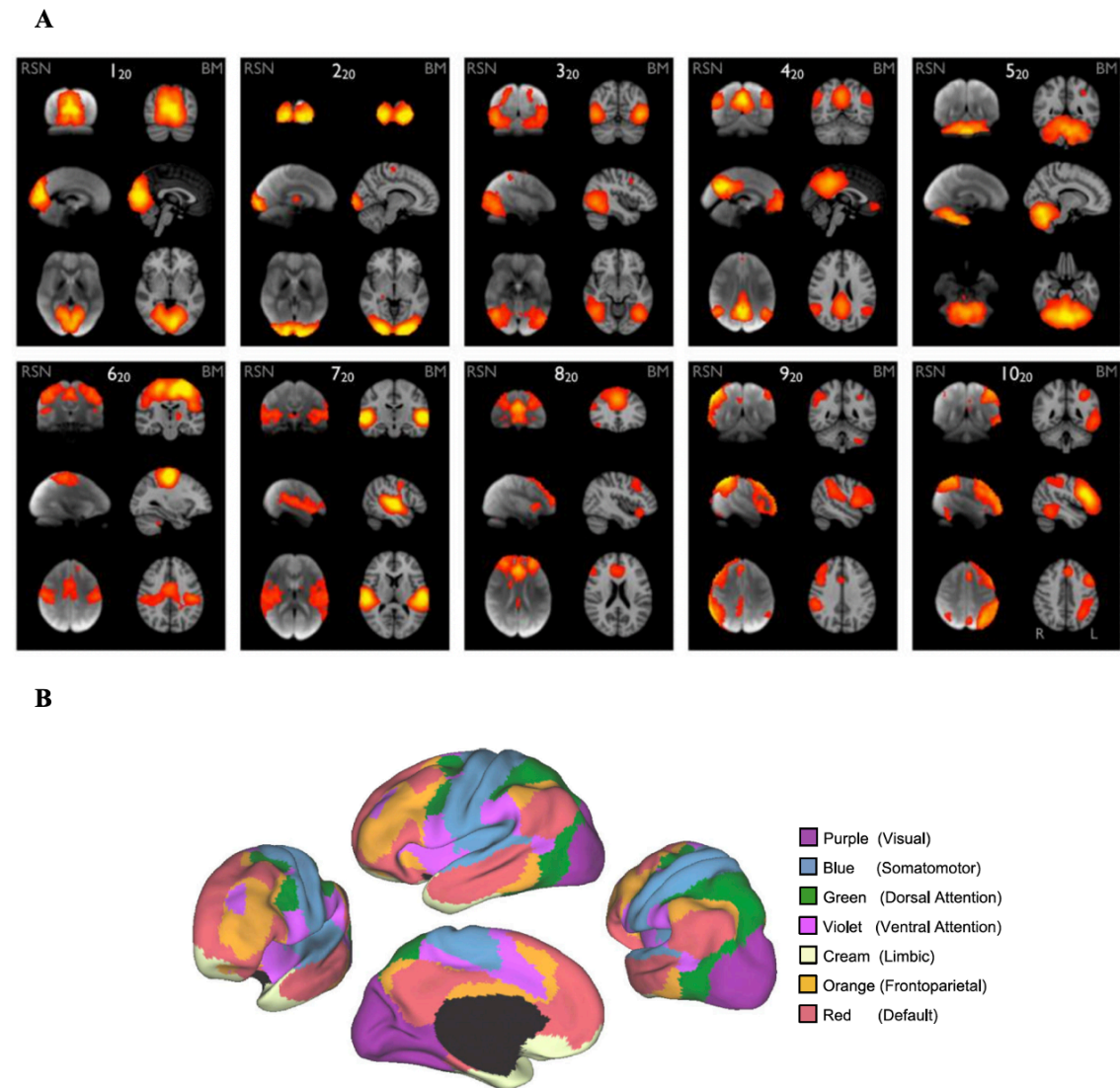


Figure 1.1. (A) Functional brain organisation described by Smith et al. (2009). Visual network (Maps 1₂₀ – 3₂₀); Default Mode network (Map 4₂₀); Cerebellar network (Map 5₂₀); Sensorimotor network (Map 6₂₀); Auditory network (Map 7₂₀); Executive Control network (Map 8₂₀); and Frontoparietal network (Maps 9₂₀ – 10₂₀). Adapted from Smith et al., 2009. (B) Similar functional brain organisation has been described by the Yeo-7Network definition. Visual network; Somatomotor network; Dorsal Attention network; Ventral Attention network; Limbic network; Frontoparietal network; and Default Mode network. The seven colour codes correspond to the respective functional networks from the left hemisphere. Note: There is also a Yeo-17Network definition which further extends the Yeo-7Network scheme. Adapted from Yeo et al., 2011.

Studies conducted over the past two decades show that functional connectivity offers a valid and robust approach to mapping the functional organisation of the human brain — the functional connectome (Fox et al., 2015; Constable et al., 2013; Van Dijk et al., 2010; Sporns et al., 2005; Raichle et al., 2001; Biswal et al., 1995). Moreover, a variety of methods aimed at identifying

individual differences in the functional connectome have provided fundamental insights into how brain functional organisation evolves across the development, and how that organisation is disrupted or altered in neurodevelopmental and psychiatric conditions (Zhang et al., 2021; Saggar and Uddin, 2019; Castellanos et al., 2013; Fox and Greicius, 2010).

1.2 Development of Functional Networks Across the Lifespan

Although the canonical functional networks can be detected across the lifespan, from prenatal life through early childhood to late adulthood, the functional connectome is substantially reshaped throughout development (Sanders et al., 2023; Bruchhage et al., 2020; Chong et al., 2019; Zuo et al., 2017; Thornburgh et al., 2017; Thomason et al., 2017; Thomason et al., 2015; Thomason et al., 2014; Tomasi and Volkow, 2011; Andrews-Hanna et al., 2007). With increasing age, neighbouring brain regions tend to segregate from one another (i.e., decrease in functional connectivity strength) and at the same time, integrate into more distributed “adult-like” functional networks through strengthening connectivity with distant regions (Fair et al., 2009). Across the lifespan, an inverted U-shape profile is observed, such that increasing integration from childhood to adulthood is followed by a “plateau”, followed by a return towards more localised network topology in later life (Zhang et al., 2015; Di Martino et al., 2014a). There are also regional differences: children tend to exhibit stronger subcortical-cortical connections, especially in the limbic and paralimbic networks, and weaker cortico-cortical connectivity compared to adults (Supekar et al., 2009). The default mode network seems particularly sensitive to developmental processes; while it is sparsely connected in childhood, its connectivity gradually increases into adulthood (Fair et al., 2009), followed by age-related reductions in the default mode functional connectivity — not only in normal ageing but also in neurodegenerative disorders such as Alzheimer’s Disease (Grieder et al., 2018; Finotelli et al., 2018; Mevel et al., 2011; Jones et al., 2011; Greicius et al., 2004). Studying the functional connectome in developmental populations is fundamental to understand typical and atypical brain functions in addition to understanding the emergence of neurodevelopmental and psychiatric conditions.

1.3 Clinical Relevance of Functional Connectivity in Neurodevelopmental and Psychiatric Conditions

Wide availability combined with low participant demands has made the functional connectivity approach a feasible, accessible, and popular approach for the investigation of the functional integrity of the human brain across a wide range of clinical conditions — neurodevelopmental and psychiatric conditions in particular. Disruptions in functional connectivity — patterns of *dysconnectivity* — have been well-studied in conditions such as autism spectrum disorder (ASD)

(Hong et al., 2020; Nomi and Uddin, 2015; Uddin et al., 2013; Nielsen et al., 2013; Gotts et al., 2012; Di Martino et al., 2011; Monk et al., 2009; Kennedy and Courchesne, 2008), attention deficit hyperactivity disorder (ADHD) (Zhang et al., 2020; Castellanos and Aoki, 2016; Li et al., 2014; Fair et al., 2012; Uddin et al., 2008), aphasia (Klingbeil et al., 2019; Baldassarre et al., 2019; Marcotte et al., 2013; Turken et al., 2011), dyslexia (Bailey et al., 2018; Schurz et al., 2015; Finn et al., 2014), schizophrenia (Li et al., 2019; Dong et al., 2018; Fornito et al., 2015; Damaraju et al., 2014; Lynall et al., 2010; Whitfield-Gabrieli et al., 2009; Zhou et al., 2007; Garrity et al., 2007), and depression (Yan et al., 2019; Drysdale et al., 2017; Mulders et al., 2015; Kaiser et al., 2015; Veer et al., 2010; Greicius et al., 2007), amongst others. Relative to conventional task-based fMRI approaches, the ability to collect data from participants from a wider spectrum of ages and abilities, together with the greater ease of pooling samples for increased statistical power, gave rise to considerable hope that functional connectivity would rapidly enable the identification of sensitive brain-based markers for neurodevelopmental and psychiatric conditions (Castellanos et al., 2013; Craddock et al., 2013b). This led to rapid growth in the number of published functional connectivity-based studies, particularly in the last decade (Figure 1.2; Zhang et al., 2021). Despite considerable research efforts, however, no robust clinically relevant biomarkers have yet been identified for any condition.

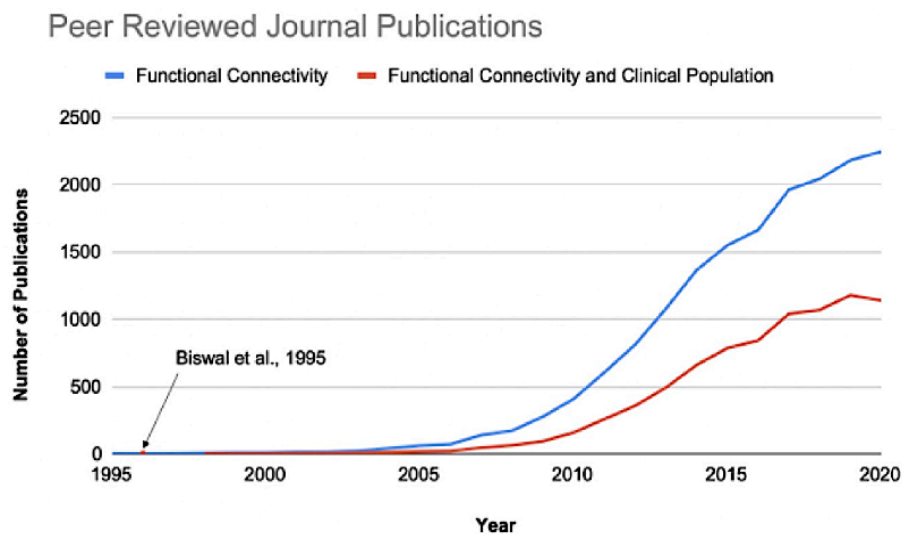


Figure 1.2. Growth in the number of published functional connectivity-based studies in neurotypical and clinical populations since the first study published by Biswal and et al. (1995). Adapted from Zhang et al. (2021).

There are a number of challenges to the identification of brain-based biomarkers. One is that there can be considerable variability across studies, sites, countries, and cultures in terms of factors such as disease severity, demographics, socioeconomic status, comorbidities, and/or medication status (He et al., 2020). These heterogeneities may be responsible for the wide range of alterations in functional connectivity that have been reported in ASD studies, for example (Dichter et al.,

2012; Ecker et al., 2010), which concluded that ASD is characterised by hyperconnectivity (Supekar et al., 2013; Di Martino et al., 2011), hypoconnectivity (von dem Hagen et al., 2013; Weng et al., 2010; Kennedy et al., 2008), or both (Monk et al., 2009).

Beyond variability in how they are measured, neurodevelopmental disorders, and psychiatric conditions more generally, are themselves highly heterogeneous in nature — marked variability at both the phenotypic and biological levels makes the identification of biomarkers particularly challenging. The diagnostic categories that we, researchers, currently rely on, do not necessarily reflect biological validity for a single symptom or condition. The National Institute of Mental Health [NIMH] has thus proposed the Research Domain Criteria (RDoC) framework to develop a diagnostic system for precision medicine — to better understand and classify the biological and psychosocial bases of mental disorders (Insel, 2014; Cuthbert and Insel, 2013; Insel et al., 2010). One response has been to identify *biotypes* (i.e., distinct subgroups within a population that exhibit specific characteristics or features such as brain functions) for neurodevelopmental and psychiatric conditions. Such efforts have been enabled by the collation of large multi-site resting-state fMRI datasets containing imaging and phenotypic data. In ASD, for example, studies employing large-scale datasets (e.g., Autism Brain Imaging Data Exchange [ABIDE]) have identified biotypes using clustering approaches such as *k*-means, such that each subgroup (cluster) exhibits distinct functional connectivity patterns (Hong et al., 2020). Easson et al. (2019) applied this approach to identify two biotypes expressing atypical functional connectivity patterns (i.e., hypoconnectivity, hyperconnectivity) within and/or between the default mode, cingulo-opercular, and frontoparietal networks. While shared dysfunctions in these networks, which support higher-order social and cognitive functioning, may be related to the core symptoms of ASD, subtype-specific network atypicalities may be differentially mapped to patterns of symptoms and symptom severity (Hong et al., 2020). Similarly, distinct functional connectivity profiles have been linked to clinical subgroups in ADHD, with different biotypes related to different behavioural profiles in terms of response variability and low arousal as well as executive and information processing performances (Fair et al., 2012). Distinct profiles of atypical functional connectivity have also been linked with specific depressive symptom profiles, such that frontoamygdala hypoconnectivity is most severe in the biotypes that are characterised by increased anxiety, anterior cingulate and orbitofrontal hypoconnectivity is observed in biotypes associated with elevated fatigue, and frontostriatal and thalamic hyperconnectivity is pronounced in biotypes associated with elevated anhedonia and psychomotor retardation (Drysdale et al., 2017).

While there continues to be great excitement about the potential of functional connectivity to uncover biomarkers, and the potential for large datasets to provide sufficient statistical power to

identify biotypes, concerns about reproducibility have emerged (Schmaal et al., 2020; Hong et al., 2020; Xia et al., 2019; King et al., 2019). Two recent examples include a failure to replicate the stratification of depressive biotypes (Dinga et al., 2019) and a lack of reproducibility in ASD-related atypicalities across data collection sites (He et al., 2020). These concerns highlight a second major challenge to the search for sensitive biomarkers for neurodevelopmental and psychiatric conditions — the need for reproducible, replicable, and robust functional connectivity-based measures.

1.4 Robustness and Reproducibility of the Functional Connectome

Looking at the fMRI literature over the past 20 years, functional connectivity has been very successfully applied to measure the functional brain organisation across development, lifespan, and a wide range of clinical conditions. In our search for functional connectivity-based biomarkers, the neuroimaging community has been accelerating their efforts to achieve *robust* and *reproducible* brain-behaviour relationships. The term robustness refers to the ability of a result to withstand variations in experimental conditions, populations, or methodologies while still producing consistent and reliable outcomes (Nosek et al., 2021). Achieving robust measures is important as it increases our confidence to ensure they remain valid and reliable over time. On the other hand, reproducibility denotes the ability to obtain similar results of a prior study using closely matched procedures that were applied in the original research (Nosek et al., 2021; Goodman et al., 2016). Obtaining reproducible measures is fundamental to independently ensure that they are valid with respect to the original experiments and methodologies. As Botvinik-Nezer and Wager (2022) outlined, reproducibility describes three properties (Figure 1.3): analytical reproducibility, replicability, and robustness to analytical variability. Analytical reproducibility refers to the ability of reproducing the same results using the data and methods described in the original study. Replicability refers to the ability of replicating similar results using the methods described in the original study but applied to new data. Robustness to analytical variability refers to the ability of reproducing similar findings using the same data described in the original study but applied to new methods. Throughout this thesis, I will use their definitions to refer to the reproducibility of functional connectivity measures and/or brain-behaviour relationships.

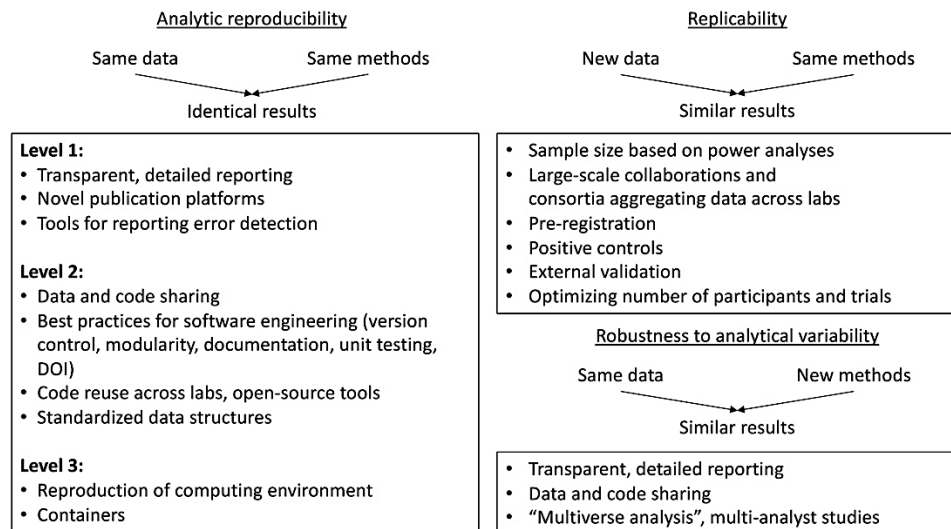


Figure 1.3. The term reproducibility describes the properties of analytical reproducibility (i.e., same data, same methods), replicability (i.e., new data, same methods), and robustness to analytical flexibility (i.e., same data, new methods). Adapted from Botvinik-Nezer and Wager (2022).

In general, achieving reproducible measures is not straightforward as there is a “reproducibility crisis” not only in neuroimaging, but in the broader context of science. Munafò et al. (2017) summarise the main causes of irreproducibility in research as follows (Figure 1.4; Munafò et al., 2017; Chambers et al., 2014): (I) failure to control for bias — no randomised control, blinding; (II) low statistical power — underpowered studies to correctly reject the null hypothesis; (III) poor quality control — presence of outliers or noisy observations in measurements; (IV) *P*-hacking — performing more statistical tests to report just the significance ones ($P < 0.05$ uncorrected); (V) HARKing [Hypothesis After Results are Known] — collecting more data after discovering statistically significant outcomes or to push “near significant” results over the boundary; and (VI) publication bias — publishing only positive findings, i.e., findings that achieve statistical significance ($P < 0.05$) and discarding null outcomes, an issue known as the “file drawer problem”.

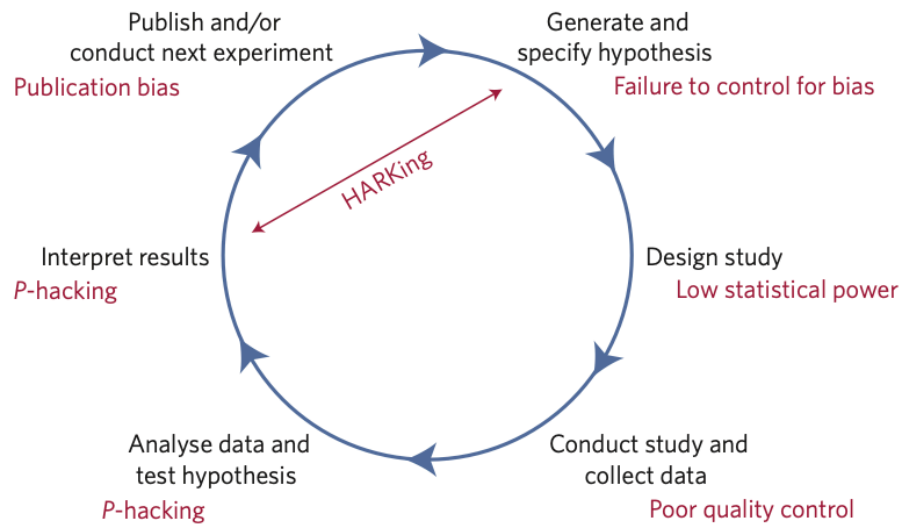


Figure 1.4. Barriers to reproducible findings across disciplines. Adapted from Munafò et al., 2017 and Chambers et al., 2014.

It is important to remember that irreproducible findings are not just disappointing from a scientific perspective — they contribute to a substantial waste of effort by researchers who design research, seek funding, and seek ethical approval, as well as the individuals who participate in research, sometimes for periods spanning days, months, or even years. In addition, there is an underlying opportunity cost, since other potential brain markers could have been pursued instead. Ultimately, reproducibility is a responsibility we owe to our participants — who should be at the heart of our research questions. This is why we are compelled to improve the reproducibility of functional connectivity-based measures — to facilitate biomarker discovery for early diagnosis, prognosis, and therapeutic intervention in clinical populations.

It is important to note that the search for brain-based markers will also depend on our ability to detect reproducible imaging-derived features from other modalities such as diffusion MRI. Diffusion MRI captures microstructural properties of the white matter tissue as well as characterises changes in structural connectivity across several behavioural domains (Ramduny et al., 2022; Brosnan et al., 2022). The advantage of diffusion MRI lies in its ability to probe white matter integrity and complexity (e.g., myelination) *in vivo* using water molecules that diffuse within the tissue with displacement at the micron scale. The application of diffusion MRI in children and adolescents is becoming increasingly popular to study age-related alterations in white matter tracts both cross-sectionally and longitudinally (Tamnes et al., 2018). Further, this modality has been applied to examine white matter developmental changes related to cognitive functions in early life (e.g., infants). For example, Short et al. (2013) showed that diffusion MRI measures in several white matter tracts were associated with visuospatial working memory performance in infants — the brain regions that are involved in these white matter tracts have

been thought to play an important role in the cognitive functioning of older children and adults. A multimodal MRI approach offers a potentially fruitful route to achieving reproducible brain-behaviour relationships during development, particularly since it provides an opportunity to examine cross-modal reproducibility. I return to this in the General Discussion in **Chapter 9**.

In the context of biomarkers, the National Institutes of Health [NIH] Biomarkers Definitions Working Group defines them as “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention” (Califf, 2018; Castellanos et al., 2013; Strimbu and Tavel, 2010; Atkinson et al., 2001). To identify functional connectivity-based measures as biomarkers, they must capture clinically relevant information (e.g., individual differences in the brain that relate to behaviour) and express valid characteristics (e.g., a measure’s effectiveness or utility in capturing the “true” estimates). Hence, focusing our efforts in detecting reproducible brain-behaviour relationships will improve the likelihood of discovering functional connectivity-based biomarkers in challenging populations such as those suffering from neurodevelopmental and psychiatric conditions.

Why are reproducible functional connectivity-based biomarkers difficult to attain? The barriers to reproducibility summarised by Munafò et al. (2017) are common across scientific disciplines, but there are some factors that have proven particularly challenging for researchers using the functional connectomics approach with clinical groups. First, as discussed above, the phenotypic heterogeneity inherent to neurodevelopmental and psychiatric conditions poses a particular challenge. While this heterogeneity is an important consideration, addressing it is beyond the scope of research conducted in this thesis; I return to this factor in the General Discussion (see **Chapter 9**). Other considerations are related to the functional connectivity measurements themselves. Understanding how we can assess the reproducibility, replicability, and robustness of functional connectivity measures, and how we can improve existing and develop new reproducible methods is what motivates this thesis.

1.5 Sample Size, Poor Statistical Power, Methodological Flexibility

As described above, achieving clinically useful functional connectivity-based biomarkers remains a challenge. Over the years, the reproducibility of many findings in the fMRI literature has been called in question as a result of the growing recognition of issues associated with small sample size, poor statistical power, methodological flexibility, and “researcher degrees of freedom”. Of all the sources of irreproducibility, sample size has received considerable attention. It is understandable that sample size and power constitute key challenges to fMRI studies, given

the expense involved. Nevertheless, low sample size has several detrimental effects, including an increased probability of false negative (type II error) outcomes, inflated effect sizes for effects that are detected, and exaggerated false reporting in the literature (Szucs and Ioannidis, 2020). While power calculations for fMRI studies were traditionally uncommon and challenging to perform (Mumford, 2012), there is now an expectation that studies be adequately powered. In addition to increasing the likelihood of false positives and false negatives, smaller sample sizes can lead to noisy or inaccurate measurements due to sampling variability. This means that the estimates obtained from smaller samples are likely to deviate from the true values of the population, which increases the difficulty of making accurate predictions or drawing valid conclusions. I return to the debate of sample size in **Chapters 3 and 9**.

One of the key advantages of the resting-state fMRI approach over traditional task-based approaches is the ease of pooling data from multiple sites — retrospectively (without prior coordination) as well as through prospective collaborations. Since the 2010s, there has been a shift towards data sharing and pooling of resting-state fMRI data, which has helped to address the challenge of small samples and low statistical power. The first such effort was the 1000 Functional Connectomes Project [FCP] (Biswal et al., 2010), which established an openly available dataset involving resting-state data collected from healthy adults at 20 sites around the world. Subsequently, efforts have focused on several clinical cohorts, including the ADHD-200 (Milham et al., 2012) and ABIDE (Di Martino et al., 2013); coordinated multimodal data collection (Human Connectome Project [HCP]; Van Essen et al., 2013); multiple scan sessions (Consortium for Replicability and Reproducibility [CoRR]; Zuo et al., 2014a); and deeply phenotyped, large community samples (UKBiobank (Miller et al., 2016); Healthy Brain Network [HBN] (Alexander et al., 2017); and Adolescent Brain Cognitive Development [ABCD] Study (Casey et al., 2018)). The ability to aggregate independent datasets have favoured the possibility to increase the sample sizes of fMRI studies by recruiting significantly more participants, thereby increasing the statistical power of the effects of interest, and variance between disease states and control conditions (Nielsen et al., 2013). The creation of pooled datasets has enabled the investigation of factors that influence reproducibility for functional connectivity studies. Two well-studied factors, for example, are the resting conditions (e.g., eyes opened, eyes closed, fixation) and acquisition parameters (e.g., temporal resolution, data acquisition sequence, scanner manufacturer) (Pomponio et al., 2020; Tagliazucchi and Laufs, 2014), which, if not controlled for, are significant sources of variability or noise in functional connectivity estimates.

Despite the promise of these large-scale datasets, reproducible functional connectivity-based biomarkers have remained out of reach. A key component of reproducibility is *reliability* (Castellanos et al., 2013) — its ability to yield similar results under a number of plausible

scenarios (Mellinger and Hanson, 2020; Heale and Twycross, 2015). Because functional connectivity can be influenced by measurement errors which are either systematic or random, measuring reliability provides an important index of how consistent the measure is (Castellanos et al., 2013). For a measure to be reproducible, it must be reliable — stable over time. In fact, for Botvinik-Nezer and Wager (2022), reliability — being replicable in independent studies across different datasets, populations, and/or sample sizes — is one of the three key components of reproducibility. Next, I will consider how reproducibility is quantified through the lens of reliability.

1.6 Measuring Reproducibility: Test-Retest Reliability

As discussed by Noble et al. (2019), reliability and validity — the correspondence of a test measure (e.g., functional connectivity) with a target measure (e.g., behaviour) — are highly complementary. Indeed, the reliability of a measure establishes the upper limit on its validity, where validity is measured in terms of its correlation with a target measure. In other words, while a reliable measure may exhibit *high* correlation with a target measure (e.g., fluid intelligence), and therefore *high* validity, an unreliable measure will likely exhibit *low* correlation with the target measure (e.g., fluid intelligence) and *low* validity. Accordingly, qualifying the reliability of a measurement can help build evidence about its validity — a desirable characteristic for brain-behaviour relationships. Indeed, in the absence of a known and measurable physiological basis of functional connectivity — a “ground truth” of validity — many researchers have used reliability as a proxy index of validity. It is important to note, however, that reliability alone cannot ensure validity — a measure can have very high reliability (e.g., due to the impact of a scanner artefact) but no validity. It is therefore important to measure both reliability *and* validity (Noble et al., 2019).

Across disciplines, reliability is used to assess the consistency of questionnaires, surveys, protocols, and clinical measures. There are several forms of reliability which include: internal consistency reliability (i.e., consistency of items within a measurement); inter-rater reliability (i.e., consistency of ratings or observations made by different “raters” or observers); inter-method reliability (i.e., consistency between two versions of a measurement); and test-retest (TRT) reliability (i.e., consistency of a measurement administered to the same group of individuals over time). In the context of fMRI, the TRT reliability has been adapted to quantify the consistency of functional connectivity estimates (Shehzad et al., 2009). To index the TRT reliability of functional connectivity measures, the intraclass correlation coefficient (ICC) is typically computed, which estimates the statistical stability of two measurements (Shrout et al., 1979).

To derive the ICC for an fMRI-based measure such as functional connectivity, a minimum of two fMRI scans is required for each participant. In theory, a sample of N participants ($i = 1, 2, 3 \dots N$) is selected randomly from a population P , and the functional connectivity at the edge-level, E is measured k times ($j = 1, 2 \dots k$) across repeated runs or sessions. Hence, E_{ij} denotes the edge-level measurement j for a given participant i which can be represented as a matrix with N rows and k columns. The participants are usually referred to as “targets” and the measurements are referred to as conditions or “raters”. The ICC of a given sample can thus be computed from the $N \times k$ matrix.

In principle, ICC is formulated as a ratio of variances (σ) based on the source of error or unwanted variances:

$$\text{ICC} = \frac{\text{variance of interest}}{\text{total variance}} = \frac{\text{variance of interest}}{\text{variance of interest} + e} \quad (2.1)$$

There are a number of ways to compute the ICC. In the case of its application to functional connectivity measures, the specific equation as described by Shehzad et al. (2009) is:

$$\text{ICC} = \frac{\text{MSBS} - \text{MSWS}}{\text{MSBS} + (k - 1) \text{MSWS}} \quad (2.2)$$

where MSBS = mean square between subjects, MSWS = mean square within subjects, and MSE = mean square error.

As such, ICC can also be described as the ratio of within-subject variance to between-subject variance (Liljequist et al., 2019; Koo and Li, 2016; Shehzad et al., 2009; Shrout et al., 1979).

In a scenario where the error source, denoted by e (e.g., fMRI runs, sessions), is at least equal to or greater than the variance of interest, the reliability of the functional connectivity is likely to be poor ($\text{ICC} < 0.4$). As ICC is estimated on the basis of how e is specified, it can take the following three forms: ICC(1,1) represents “agreement” in exact values when e is not specified; ICC(2,1) refers to as “absolute agreement” in exact values during which e is specified and modelled as random effects; and ICC(3,1) corresponds to “consistent agreement” in exact values during which e is specified and modelled as fixed effects. As an example, the ICC(1,1) and ICC(2,1) models can be mathematically expressed as follows:

$$\text{ICC}(1, 1) = \frac{\sigma_\lambda^2}{\sigma_\lambda^2 + \sigma_e^2} = \frac{\text{MSBS} - \text{MSWS}}{\text{MSBS} + (k - 1) \text{MSWS}} \quad (2.3)$$

$$\text{ICC}(2, 1) = \frac{\sigma_{\lambda}^2}{\sigma_{\lambda}^2 + \sigma_{\pi}^2 + \sigma_{\varepsilon}^2} = \frac{\text{MSBS} - \text{MSE}}{\text{MSBS} + (k-1) \text{MSE} + \frac{k}{N} (\text{MSBM} - \text{MSE})} \quad (2.4)$$

where λ = object of measurement such as subjects, π = specified error source, ε = unspecified error source, $\theta_{\pi}^2 = \frac{\sum_j (\pi_j - \bar{\pi})^2}{k-1}$, and MSBM = mean square between measurements (e.g., fMRI scans). Of note, the reliability of the functional connectivity estimates can be generalised by averaging across k measurements and be further expressed as ICC(1, k), ICC(2, k), and ICC(3, k) models (Noble et al., 2019; Liljequist et al., 2019; Koo and Li, 2016).

ICC values generally range from 0 to 1 such that an ICC = 1 indicates perfect reliability whereas an ICC = 0 denotes complete unreliability. Values are typically divided into four levels: poor (ICC < 0.4), fair (0.4 < ICC < 0.59), good (0.6 < ICC < 0.74), and excellent (ICC ≥ 0.75) (Cicchetti and Sparrow, 1981). It is important to note that the use of these categories can differ across studies. For example, an ICC = 0.4 can be taken to mean “poor” reliability in one study (Noble et al., 2019) but denote “fair” reliability in another study (Zuo and Xing, 2014b).

An important question that arises is how best to compute and summarise ICC for a given application. ICC can be computed in a variety of ways — for example, at a voxelwise level for a given region of interest (ROI)’s whole-brain pattern of functional connectivity, or on an edge-level basis, i.e., for a given connection between a pair of ROIs. Voxelwise maps contain ~200,000 ICC values (depending on voxelwise resolution) quantifying the TRT reliability of a single ROI; these individual ICC values can then be averaged across the whole-brain, or within specific functional networks to obtain a summary ICC value. Computing ICC edgewise can reduce complexity and dimensionality (e.g., 268 x 268 ROIs for the Shen parcellation scheme), but the question of how best to summarise the metric (e.g., across all edges? Within networks?) still arises. How best to handle negative ICC values is also unclear — it is possible to have a negative ICC for an edge in the functional connectome if within-subject variance exceeds between-subject variance. Such ICC values are typically set to 0 to indicate complete unreliability (Noble et al., 2019; Noble et al., 2017; Bartko, 1976). However, the ICC’s reliance on a ratio of within- to between-subject variance merits further consideration — I return to this below.

1.6.1 Assessing the Reliability of the Functional Connectome During Adulthood

The majority of TRT reliability studies have focused on adult populations, and have shown a range of reliability estimates, from poor to excellent TRT reliability (Guo et al., 2012; Zuo et al., 2010a; Shehzad et al., 2009). Using within-session (< 1 hour), between-session (> 5 months), and

multi-scan session (3 scans) scans, Shehzad et al. (2009) showed that the TRT reliability of functional connectivity derived from seed-based and voxelwise measures ranged from poor to excellent. Importantly, within-session ICCs were found to be higher, on average, compared to between-session ICCs. These TRT reliability estimates, however, depend on three factors (Shehzad et al., 2009): (1) *statistical significance* — significant correlations exhibit higher TRT reliability relative to non-significant correlations; *valence* — significantly positive correlations show higher TRT than significantly negative correlations; and (3) *network membership* — subregions that form part of the default mode network are more reliably correlated with each other compared to regions within a task-induced network.

During adulthood, ICC has been computed to evaluate the TRT reliability of several functional connectivity measures. The application of independent component analysis (ICA; a data-driven approach which identifies maximally independent components that represent neuronal activities and components which purely reflect physiological artefacts) has shown fair-to-good reliability (ICCs: 0.45-0.65) of canonical functional networks (e.g., frontoparietal, default mode, and sensorimotor networks) within (< 45 min) and between (5-16 months) scanning sessions (Zuo et al., 2010a). Across studies, the highest reliability estimates are obtained for the frontoparietal networks that support attentional control abilities as well as the default mode network (Zuo et al., 2010a), which may be related to the fact that the highest *inter*-individual variability in functional connectivity has been reported for association areas, including those that make up the frontoparietal and default mode networks (Noble et al., 2019; Mueller et al., 2013). These observations also suggest that cortical networks exhibit greater TRT reliability than subcortical networks, potentially due to stronger and positive functional connections within the cortex (Noble et al., 2019). Another study with adults (55-85 years) showed that between-network connections tend to be more vulnerable to ageing effects, relative to within-network connections (Song et al., 2012). The proportion of significantly reliable correlations between-network connections was found to decrease in ageing adults, but the opposite pattern was observed for within-network connections, especially in the frontoparietal and sensorimotor networks (Song et al., 2012). The same study further demonstrated that younger adults (18-35 years) achieved higher proportion of reliable and positive correlations compared to ageing adults, leading to better TRT reliability of functional connectivity (Song et al., 2012). It is important to note that developmental change in functional connectivity is not linear — non-linear age-related changes (i.e., quadratic, cubic) in resting-state functional connectivity likely reflect underlying structural maturational processes including synaptic pruning and myelination (Zuo et al., 2010b). Differential rates of brain development or maturation could introduce large *intra*- and *inter*-individual variability in functional connectivity-based measures, especially in early life and around puberty — which can then affect the ICC estimates. As mentioned above, poorer TRT reliability also tends to be noted

for negative correlations (“anticorrelations”), for example, between task-based networks and the default mode network (Shehzad et al., 2009). This finding agrees with prior reports which showed that anticorrelated patterns of functional connectivity exhibit lower stability than positively correlated voxels (Tian et al., 2007). In a recent review, Noble et al. (2019) concluded that ICC shows varying reliability estimates in general, potentially as a result of greater variance in the estimates of both within- and between-subject variation.

Early work examining the TRT reliability of functional connectivity measures focused on data collected from adult participants. Since neurodevelopmental and psychiatric conditions emerge during childhood and adolescence, there is a need to understand the stability of functional connectivity-based measures during these sensitive developmental periods during which the brain undergoes significant maturational change. To date, only a small number of efforts have assessed the TRT reliability for functional connectivity metrics either across development or in neurodevelopmental populations.

1.6.2 Assessing the Reliability of the Functional Connectome Across Childhood and Adolescence

Several resting-state fMRI studies have observed that spontaneous fluctuations in the fMRI signal are more variable during childhood and adolescence compared to adulthood — which may be expected, since these periods are marked by large maturational changes and cognitive development (Marek et al., 2019; Thomason et al., 2011; Thomason et al., 2005). In terms of TRT reliability, Thomason et al. (2011) computed ICC for ICA-based measures of functional connectivity, and showed that typically developing children and adolescents aged from 9 to 15 years mostly exhibited fair-to-excellent reliability in functional networks that correspond to cognitive and sensorimotor areas (Thomason et al., 2011). Similar TRT reliability has been observed longitudinally over a span of 2-3 years although higher consistency of functional connectivity in these networks was revealed within sessions (Thomason et al., 2011). This suggests that patterns of functional connectivity are stable even across periods during which the brain continues to mature. TRT reliability studies in clinically developing populations are even more rare. One study revealed fair-to-excellent reliability (ICCs: 0.40-0.80) of functional connectivity using measures of regional variability, local connectivity, and default mode connectivity in typically developing and ADHD children (Somandepalli et al., 2015). The majority of the functional connectivity estimates showed stronger reliability in the typically developing cohort compared to the ADHD group. The lower ICCs in ADHD children could be explained by the high between-subject and *intra*-subject variances which were observed in these children (Somandepalli et al., 2015). Within both groups, the highest ICCs were observed in

highly reliable functional networks involving the frontoparietal, default mode, and attention networks whereas the least ICCs were found in the visual network (Somandepalli et al., 2015). Other studies have painted a less positive picture of TRT reliability in developing populations, but for task-based, rather than resting-state functional connectivity measures (Heilicher et al., 2022; Kennedy et al., 2021; Herting et al., 2017).

The lack of studies in youths likely reflects some of the challenges to quantifying TRT reliability. One of the most obvious is the requirement to have two scans, ideally separated in time (i.e., between-session, rather than within-session), which can be challenging to obtain in developmental populations. In addition, since participants are, by definition, still *developing*, TRT reliability measures for developmental populations are likely to be negatively impacted by the normal maturational changes in the brain that may occur in the time between scans, as well as by factors that reduce reliability. Disambiguating these influences is likely to be challenging. This highlights a need to consider factors that we can control, so that we can avoid suboptimal experimental designs and analyses. The ICC exhibits other shortcomings too, further motivating the consideration of alternatives.

1.6.3 Factors Driving TRT Reliability of the Functional Connectome

A review of studies, including Noble et al. (2019), suggest that the following key factors affect the reliability of functional connectivity estimates. Importantly, these factors have been used to identify optimal functional connectivity analysis pipelines across brain states, populations, and datasets:

1. Resting-state Conditions: TRT reliability of functional connectivity appears to be higher for data collected during fixation conditions, relative to eyes-opened (but not fixated) or eyes-closed; the eyes-closed condition exhibits the least reliability in the default mode, attention, and motor networks (Patriat et al., 2013). The default mode network also shows higher TRT reliability during eyes-opened, relative to eyes-closed (Zuo et al., 2015). The explanatory factor may be wakefulness; fixation and eyes-opened conditions may better enable participants to stay awake during scanning — individuals tend to experience drowsiness or sleepiness when they are at rest (Tagliazucchi and Laufs, 2014), particularly with eyes closed. Further, the TRT reliability of functional connectivity is improved when sleep effects are removed from the fMRI signal (Wang et al., 2017a).
2. Naturalistic Paradigms: The majority of TRT reliability studies have employed resting-state fMRI conditions across several populations. With increasing efforts in maximising data quality, particularly in developing cohorts, naturalistic paradigms such as movies have been

proposed (Vanderwal et al., 2019a; Greene et al., 2018b; Dosenbach et al., 2017; Vanderwal et al., 2015). Movies have been found to improve the TRT reliability of functional connectivity compared to resting-state conditions in young adults (21-31 years) (Wang et al., 2017b). Wang et al. (2017b) showed that movies increased the ICCs from fair to excellent compared to resting-state conditions. Although these improvements are spatially located across the whole-brain, the greatest increase in the TRT reliability estimates have been observed in the frontoparietal, default mode, and limbic networks (Wang et al., 2017b).

3. Spatial Distribution of Functional Connectivity: The most reliable functional networks include the default mode, frontoparietal, and attention networks (Noble et al., 2019; Noble et al., 2017). However, significantly lower TRT reliability of functional connectivity-based measures has been observed when subcortical connections are considered, relative to cortical connections (Shah et al., 2016). For example, poor-to-good ICCs have been measured for the bilateral nucleus accumbens, globus pallidus, and amygdala in addition to deep gray matter nuclei (Shah et al., 2016). Therefore, TRT reliability estimates may suffer from the inclusion of subcortical regions, potentially due to their smaller volumes, close proximity to physiological sources, and low signal-to-noise ratio (SNR) due to signal dropout (Noble et al., 2017).
4. Scan Duration: The TRT reliability of functional connectivity-based measures appears to be dependent on scan length, such that scan durations between 6 min to 16 min increase TRT reliability (Birn et al., 2013). One study has posited that even longer scans are required: Noble et al. (2017b) showed that over 36 min is required to achieve high TRT reliability, which may not be ideal for developing or clinical populations. On the other hand, the average correlation strengths within and between functional networks such as the frontoparietal and default mode networks can be improved with shorter scan lengths of ~5 min, with further increases in imaging time (up to 41 min) only marginally improve the TRT reliability of functional connectivity-based estimates (Van Dijk et al., 2010). This finding is in line with previous evidence showing that the organisation of functional networks may be moderately differentiable at the individual level with such short acquisition length (Shehzad et al., 2009). Similarly, while improved reliability of functional networks can be achieved with increasing scan duration, their spatial distributions may attain stability within 5 min of imaging time (Mueller et al., 2015).
5. Denoising Strategies: Denoising strategies minimise the contribution of non-neural signals in functional connectivity-based measures. We would therefore expect that denoising would increase the TRT reliability of functional connectivity estimates. However, the application of

denoising strategies has yielded mixed results in the fMRI literature. Global signal regression (GSR), which is still a controversial confound regression method, achieved improved TRT reliability estimates for one study (Yan et al., 2013b) but revealed detrimental effects for others (Wanger et al., 2022; Tozzi et al., 2020; Varikuti et al., 2016; Shirer et al., 2015). Similarly, ICA techniques such as ICA-AROMA have shown both improvement and reduction in the TRT reliability of functional connectivity measures (Parkes et al., 2018; Pruim et al., 2015; Beckmann et al., 2005). When a temporal censoring procedure such as scrubbing (i.e., detection and removal of motion-corrupted timepoints in the fMRI timeseries) was further applied, the TRT reliability of some network-based functional connectivity measures increased (Du et al., 2015; Power et al., 2014). In another study, scrubbing up to 50% of the fMRI timepoints had minimal effect on the TRT reliability of functional connectivity whereas truncating the timepoints had more drastic effect on the reliability estimates (Shah et al., 2016). It is challenging to identify a single denoising technique that performs consistently well across all quality control metrics (Parkes et al., 2018; Ciric et al., 2017; Varikuti et al., 2016). Hence, it is likely that the relatively low TRT reliability in resting-state fMRI studies reflects noise signals that remain in the preprocessed data (Shirer et al., 2015).

The literature suggests that there is a broad spectrum of factors that affects or are likely to affect TRT reliability, due to their differential influences on both *intra*- and *inter*-individual variability. Studies that minimise these factors should benefit from increased measurement reliability. While ICC has proven very useful for assessing the TRT reliability of functional connectivity measures, it has some shortcomings that limit its utility in functional connectivity applications, particularly in developing populations. I suggest that we should also consider looking beyond the ICC as a primary measure of TRT reliability.

1.6.4 Limitations of ICC in Evaluating the Reliability of the Functional Connectome

As described in **Section 1.6**, it is unclear how to best compute and summarise the ICC for functional connectivity measures — for example, whether this should be done at the voxel, network, or edge level. Moreover, it is a challenge to achieve excellent TRT reliability (i.e., ICC ≥ 0.75). From Equation (2.2), it would be essential to observe low *intra*-individual and high *inter*-individual variability (Zuo and Xing, 2014b; Shehzad et al., 2009). Conceptually, this is because low *intra*-individual variability reflects high temporal stability, while high *inter*-individual variability indicates greater discrimination across participants (Zuo and Xing, 2014b). However, we know that functional connectivity patterns are remarkably consistent across participants, potentially leading to relatively low *inter*-individual variability, particularly for some functional networks. Indeed, Noble et al. (2019) showed that *inter*-individual variability explains 80% of

the variance, indicating that the ICC for a given functional connectivity measure is dominated by between-subject differences.

Dissecting the variance components used to compute the ICC may offer one solution, but Noble et al. (2019) caution that this requires careful understanding of how within-subject and between-subject variability can be affected across different population samples. Even in adult populations, within-subject variance is difficult to interpret as individuals can undergo changes in brain structure and function over short (e.g., < 1 day), long (e.g., days, months), and developmental periods (e.g., > 1 year; Noble et al., 2019) as a result of myriad factors including learning, health, and work or activity patterns. As noted above, disentangling these influences from those that reflect random time-related variability in the functional connectivity metrics of interest is likely to be challenging; treating them as random effects may not be adequate to derive the “true” ICC (Noble et al., 2019). The challenge of disentangling these influences underscores a need to identify alternative metrics for assessing the stability of functional connectivity estimates.

1.7 Emergence of Functional Connectome Fingerprinting

What are the potential alternatives to the ICC as a measure of reproducibility of functional connectivity metrics? In a seminal study, Finn et al. (2015) showed that the functional connectivity profile of an individual is unique and stable, and can serve as a “*fingerprint*” to identify that individual over time. Here, and in **Chapter 2**, I suggest that the identifiability of the functional connectome fingerprints may serve as a more intuitive, simple, and less computationally intensive way to assess reproducibility, which can complement measures of validity in our search for better brain-based biomarkers.

The process of identifying an individual based on their functional connectivity profiles is known as functional connectome fingerprinting (Finn et al., 2015). In order to identify an individual using their functional fingerprint, a minimum of two fMRI scans (e.g., REST-REST, REST-TASK, TASK-TASK) is needed, either from the same scan session or between scan sessions that may span across multiple days, months, and even years (Finn et al., 2015; Finn et al., 2017). For a given dataset comprising N participants with Scan 1 and Scan 2, the process involves the creation of a “database” $\mathbf{D} = \{X_i, i = 1, 2, 3, \dots, N\}$, where X_i refers to the vectorised functional connectivity matrix derived from the non-zero ROI values from Scan 1 and the subscript i represents the i th participant in $\mathbf{D}$. Iteratively, the functional connectivity matrix X_j from a given individual j from Scan 2 is selected and this matrix is treated as the “target matrix”, $\mathbf{T} = \{X_j, j = 1, 2, 3, \dots, N\}$. The target matrix X_j is then compared with each of the functional connectivity

matrices in $\mathbf{D}$ and a Pearson's correlation coefficient (r) is computed for each pairing. For a given individual j , the correlation coefficients (r) are then ordered and $\mathbf{ID} = \text{argmax}(\{r_1, r_2, r_3 \dots r_N\})$. A participant is considered correctly identified ($\mathbf{ID} = 1$) if the identity (X_j) of the highest-ranking correlation coefficient (r) matches the true identity X_i of the individual in $\mathbf{D}$, otherwise the $\mathbf{ID} = 0$ (Figure 1.5). The ID accuracy is therefore a simple and intuitive measure that captures the reproducibility of an individual's functional connectome with a single number.

The ID accuracy of an entire dataset is computed as:

$$\frac{\sum \text{Number of correctly identified participants}}{N} \quad (2.5)$$

where N denotes the total number of participants in a given dataset. For a given dataset, the ID accuracy is averaged by exchanging the roles of the database-target matrix ($\mathbf{D} \leftrightarrow \mathbf{T}$). The higher the ID accuracy of a dataset, the better the ability to differentiate an individual from a group based on their functional connectivity profile — that is, the better the ability to pick a person out from a crowd on the basis of their functional connectome fingerprint.

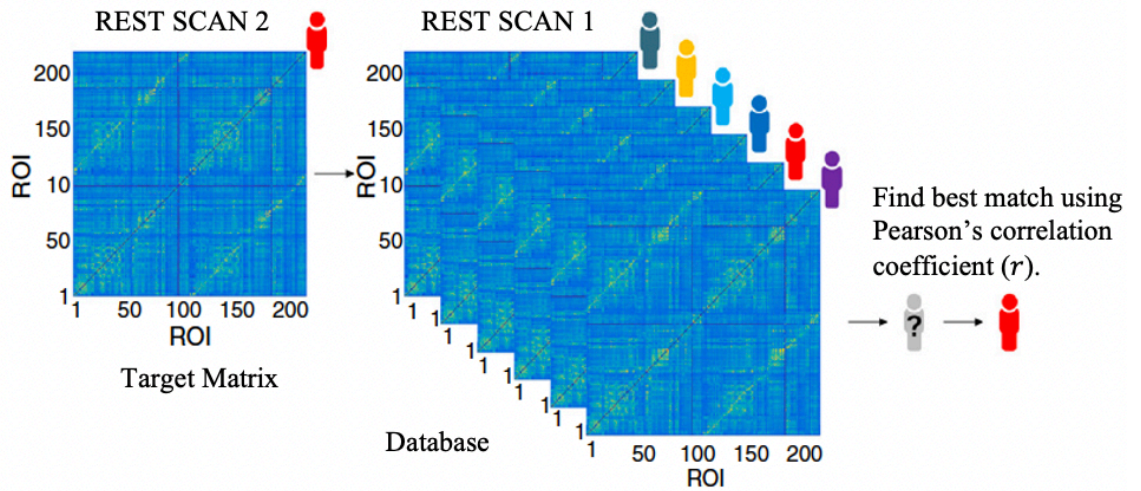


Figure 1.5. Identification procedure as described by Finn et al. (2015). An individual is identified on the basis of their functional connectivity matrices derived between two imaging sessions (i.e., Database and Target Matrix). This matching procedure uses Pearson's correlation coefficient such that an individual is successfully identified ($\mathbf{ID} = 1$) if the correlation is maximal between the two functional connectivity matrices of that participant, otherwise that individual is misidentified ($\mathbf{ID} = 0$). Adapted from Kauffmann et al., 2017.

1.7.1 Individual Differences in the Functional Connectome derived from Predictive Networks

High ID accuracy for the fingerprinting approach has been observed using resting-state (i.e., REST-REST: 98-99%), task-based (i.e., TASK-TASK: 64-93%), and cross-state (i.e., REST-TASK: 85-92%) paradigms. This pattern of variability in ID accuracy suggests that it is more challenging to identify unique functional connectivity profiles across different task conditions, and between task and rest conditions (Finn et al., 2015). Several studies have shown that the ability to identify an individual from a group is stronger when participants perform a task, across a broad range of domains (Cai et al., 2021; Venkatesh et al., 2020; Finn et al., 2017; Finn et al., 2015). This has been attributed to the fact that functional connectivity patterns tend to be more idiosyncratic at rest, since the unconstrained nature of the scanning environment allows participants to “mind-wander,” leading to different patterns of cognition, arousal, and other processes. Although task-based fMRI can produce relatively high fingerprint ID accuracy, participant identification accuracy depends on factors such as task novelty, difficulty, and load. It is also important to note that naturalistic fMRI paradigms such as movies can produce stronger individual differences in functional connectivity patterns compared to resting-state and task-based conditions (Vanderwal et al., 2021). It remains a challenge to identify a single movie that performs consistently well in identifying individuals across a diverse range of populations, since each movie aims to evoke different processes with their underlying features (e.g., stimulus, length, narration). An important lesson from these findings, however, is that although changes in brain states modulate the functional connectivity patterns of an individual, a given individual’s intrinsic functional architecture is unique from a group of participants, regardless of how the brain is engaged (Finn et al., 2015).

There also appears to be some spatial differentiation — the highest ID accuracies have been found in the medial frontal and frontoparietal networks, two functional networks which share high reproducibility in neurotypical and developing populations (Marek et al., 2019; Finn et al., 2015; Mueller et al., 2013). The combination of these two networks (referred to as the “frontoparietal” networks) further shows improved ID accuracy relative to using either network alone (Finn et al., 2015). This is consistent with the observation that frontoparietal networks tend to be the most relevant in linking individual differences of functional connectivity to individual differences in behaviour (e.g., fluid intelligence) (Finn et al., 2015). It is likely that factors that affect TRT reliability, i.e., the ICC (see **Section 1.6.3**), will also influence the fingerprint ID accuracy.

1.7.2 Age and Gender Related Dependencies on the Individual Functional Connectome

The ability to successfully identify an individual is determined by that individual’s connectome distinctiveness, i.e., the degree to which an individual connectome differentiates that person from a group (Kaufmann et al., 2017). Recent work has shown that greater connectome distinctiveness

is associated with increasing age during the critical period that spans puberty (8-22 years, peaking at 14.5 ± 1.8 years). Kaufmann et al. (2017) found that developmental youths revealed higher ID accuracy in functional networks such as the medial frontal, frontoparietal, default mode, and motor networks; these functional networks appear to be highly predictive of an individual's functional connectivity profile during childhood and adolescence (Kaufmann et al., 2017). However, the distinctiveness of the functional connectome was decreased or delayed with increasing clinical scores related to depression, attention deficit disorder, and prodromal symptoms of schizophrenia (Kauffman et al., 2017). High ID accuracy appears to be relatively stable during development; individuals can be successfully identified (ID accuracy: 75-86%) on the basis of their functional connectome fingerprints across spans of several years (1-2 years), even extending to 3 years during adolescence and adulthood (Horien et al., 2019). These observations have been derived using the most predictive networks, i.e., the medial frontal and frontoparietal networks, which agree with the original fingerprinting findings derived from Finn et al. (2015). This suggests that long-term individual differences in the functional connectome are preserved over long timespans, even during developmental periods. Similar observations have been found when measuring the sensitivity (i.e., likelihood of detecting a true positive ID) and specificity (i.e., likelihood of detecting a true negative ID) of the ID accuracy in youths and adulthood (12-30 years) on short-term (same-day scan) and long-term (scans 1.5 years apart) (Jalbrzikowski et al., 2019). Importantly, there appears to be no significant differences in ID accuracy between youths and adulthood over short- and long-term timespans, suggesting that the individual functional connectome is stable in youths, despite underlying maturational change (Jalbrzikowski et al., 2019; Horien et al., 2019). Significant improvements have been consistently reported in ID accuracy when the most predictive connections from the frontoparietal, default mode, and dorsal attention networks are used (Jalbrzikowski et al., 2019).

Moreover, there are significant gender differences in ID accuracy, such that male youths demonstrate delayed connectome distinctiveness (i.e., lower ID accuracy), relative to female youths (i.e., higher ID accuracy) (Kaufmann et al., 2017). This finding, however, appears inconsistent with the findings of a previous study, which found that female adults from the HCP dataset exhibit greater between-subject correlation relative to male adults (Finn et al., 2017). That is, across brain states derived from resting-state and task-based conditions, female adults from one session are more similar (less distinctive) to other female adults in another session, whereas male adults tend to be less similar to other male adults from one session to the other. If females are more similar to each other, then intuitively, they should be harder to identify from their functional connectome fingerprints, relative to males. The impact of sex differences on the connectome distinctiveness remains unclear and needs to be further investigated. I return to the topic of sex differences in the General Discussion (see **Chapter 9**).

One of the main challenges to developmental studies using resting-state fMRI data and functional connectome measures is that children tend to move more during scanning, reducing data quality and leading to data loss. In addition, when studying developmental changes, differential movement between youth and adult groups constitute a significant confound (Power et al., 2015; Van Dijk et al., 2012; Satterthwaite et al., 2012; Power et al., 2012). One solution is to use naturalistic paradigms such as movies, rather than “rest” — this has been shown to significantly increase participant compliance and reduce motion (Vanderwal et al., 2019; Greene et al., 2018b; Dosenbach et al., 2017; Vanderwal et al., 2015). This is because movie-watching can increase the engagement of participants as the storyline develops (Vanderwal et al., 2019; Vanderwal et al., 2015). Another advantage of using movies is that they tend to exhibit higher TRT reliability than resting-state conditions (Wang et al., 2017b; see **Section 1.6.3**). I will return to the benefits of using naturalistic fMRI paradigms in **Chapter 7**. While the majority of the fingerprinting analyses have focused on conventional resting-state and task-based conditions, there has been a growing effort to examine the distinctiveness and stability of the individual functional connectome using movies (Vanderwal et al., 2021; Vanderwal et al., 2017; Finn et al., 2017). In contrast to findings using resting-state alone (Jalbrzikowski et al., 2019; Kaufmann et al., 2017; Horien et al., 2019; reviewed above), fingerprint ID accuracies across rest and movie-watching conditions in a large developmental psychiatric cohort ($N = 210$, 6-21 years) were lower (18-44%) than previously reported in adults (80-94%; Vanderwal et al., 2021; Vanderwal et al., 2017). It is possible that several key methodological factors may have contributed to this conflicting finding. Identifying these factors is an important step towards understanding how we can optimise the detection of individual differences in the functional connectome, especially in neurodevelopmental and psychiatric populations, such as ASD, ADHD, or depression.

1.8 Objectives and Hypotheses

This thesis aims to identify ways in which we can improve the reproducibility of functional connectivity measures in neurodevelopmental and psychiatric conditions, so that these measures can be used in the search for brain-based biomarkers. In light of my review of the literature on TRT reliability and connectome-based fingerprinting, above, I suggest that fingerprint ID accuracy could serve as a simple, intuitive, and less computationally intensive reproducibility metric than the ICC. Further, I suggest that the adoption and deployment of fingerprint ID accuracy for this purpose will help advance the field towards more reproducible functional connectivity measures, and ultimately, better biomarkers. In this thesis, I use the process of functional connectome fingerprinting as a testbed for identifying functional connectivity analysis methods that maximise reproducibility. My expectation is that improving the *reproducibility* of functional connectivity measures should also improve their *validity* — and translate to better

brain-behaviour relationships. This is because improving reproducibility can improve validity in its own right, but also because better fingerprint ID accuracy suggests better differentiation of an individual's functional connectome, which should lead to better differentiation of their behaviour — based on the assumption that the unique functional connectivity features which distinguish them are behaviourally relevant, at least for the behavioural phenotypes measured in this thesis.

In line with the objectives of this thesis, I hypothesised that (1) connectome-based fingerprinting can be used as an alternative test (to ICC) of reproducibility of functional connectivity measures; (2) methodological factors that increase fingerprint ID accuracy can be considered to increase the reproducibility of functional connectivity estimates; (3) methodological factors that increase reproducibility (quantified as fingerprint ID accuracy) will also increase validity, in terms of stable and reproducible brain-behaviour relationships for functional connectivity measures. The roadmap for the research conducted in this thesis is as follows:

- **Chapter 2** demonstrates the use of connectome-based identification as a testbed for benchmarking, comparing, and adjudicating between methods for deriving functional connectivity estimates in developing and adult populations. It examines factors that affect reproducibility, quantified in terms of fingerprint ID accuracy, and describes a fingerprint-informed pipeline for functional connectivity analyses that optimises the detection of individual differences in the functional connectome.
- **Chapters 3 and 4** apply the fingerprint approach to test the reproducibility of a novel machine learning approach known as “bagging” for rescuing “high-movement” participants for functional connectivity analyses. Across several samples, including developmental and neurodevelopmental groups, I show that reproducible brain-behaviour relationships can be computed using a subsample of the least motion-corrupted fMRI data. I suggest that this can help researchers strike a balance between data quality requirements and sample size in brain-behaviour association studies.
- **Chapters 5 and 6** apply the factors that improve connectome-based identification and head motion strategies that minimise the exclusion of “high-motion” individuals to investigate the robustness of connectome-based predictions of behavioural phenotypes in developing youths — assessing whether the methods developed in **Chapters 2 and 4** also boost the prediction and generalisability of brain-behaviour relationships.
- **Chapters 7 and 8** adopts a different approach to boosting reproducibility by examining a naturalistic paradigm — an emotionally evocative movie — intended to boost individual

differences in emotional responses in a community-based sample of female adolescents experiencing subclinical depressive symptoms. This work aims to develop a naturalistic fMRI framework that captures emotional processing at multiple spatial and temporal scales.

Chapter 2

Using functional connectome fingerprinting as a testbed for strategies to improve reproducibility

Paper (In Preparation)

Ramduny, J., Vanderwal, T., & Kelly, C. Using functional connectome fingerprinting as a testbed for strategies to improve reproducibility.

Authors

Jivesh Ramduny^{1,3}, Tamara Vanderwal^{4,5}, Clare Kelly^{1,2,3}

¹School of Psychology, Trinity College Dublin, Dublin, Ireland

²Department of Psychiatry, School of Medicine, Trinity College Dublin, Dublin, Ireland

³Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland

⁴Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada

⁵BC Children's Hospital Research Institute, Vancouver, BC, Canada

Authorship Contribution Statement

Jivesh Ramduny: Conceptualisation, Methodology, Formal analysis, Investigation, Writing — original draft, review & editing. **Tamara Vanderwal:** Conceptualisation, Investigation, Writing — original draft, review & editing. **Clare Kelly:** Conceptualisation, Funding, Methodology, Formal analysis, Investigation, Writing — original draft, review & editing, Supervision, Project administration.

Abstract

Resting-state fMRI has provided ways to capture individual differences in the human functional connectome, and much of the field is striving to optimise methods that relate these differences to behaviour. “Fingerprinting” is the use of a test-retest matching algorithm that asks whether an individual’s connectome is more correlated with another of their own connectomes relative to connectomes from other participants. Here, I test the use of this straightforward matching algorithm as a testbed for reproducible functional connectomes and to serve as a sensitive index for comparing, benchmarking, and adjudicating functional connectivity analysis pipelines. To achieve this, I identified several data pre- and post-processing parameters that form part of two optimal fingerprint-informed pipelines for improving the reproducibility of the individual functional connectome. I hypothesised that whichever parameters yielded the best identification accuracies would also yield the best intraclass correlations across scans. Using five datasets from the Consortium for Replicability and Reproducibility, I found that higher identification accuracies were achieved when I employed: (I) global signal regression; (II) finer brain parcellations; (III) cortical regions compared to subcortical and cerebellar structures; (IV) medial frontal and frontoparietal networks relative to the whole-brain; (V) greater number of discriminative edges; (VI) longer scan duration; and (VII) lower sample size. I then estimated the TRT reliability of the functional connectome using a conventional method, i.e., ICC. In line with recent work, I observed that ICC was consistently “poor” across the five datasets even with the application of the optimal pipelines. The connectome-based fingerprint pipelines may serve as a measure of comparison, benchmarking, and adjudication of functional connectivity-based analysis pipelines and novel methodologies, as a means to enhance their reproducibility, especially in heterogeneous populations such as those suffering from neurodevelopmental and psychiatric conditions.

2.1 Introduction

As outlined in **Chapter 1**, the identification of clinically useful brain-based biomarkers requires that our measures of the brain (here, functional connectivity) and of behaviour (e.g., clinically important symptoms or behaviours) are reproducible, defined by Botvinik-Nezer and Wager (2022) as analytical reproducibility, replicability, and robustness to analytical variability. The goal of this thesis is to address the former — the reproducibility of functional connectivity metrics.

One challenge to assessing functional connectivity metrics is the lack of a “ground truth” against which to evaluate them and which could serve as a basis for comparing methods and pipelines for computing functional connectivity (Zuo et al., 2019). For example, while functional connectivity shows good correspondence with measures of anatomical or structural connectivity (e.g., as revealed by retrograde tracer studies in non-human animals, or diffusion-based connectivity studies in humans) (Litwińczuk et al., 2022; Margulies et al., 2009), it is clear that functional connectivity metrics reflect more than is captured by these anatomical measures, so correspondence with such measures alone cannot be used to evaluate functional connectivity metrics or methods for their computation. In the absence of a ground truth, many studies have focused on TRT reliability as an outcome measure; if a particular functional connectivity metric or analysis step or pipeline is advantageous, it should exhibit improved TRT reliability. The most consistently used index to evaluate TRT reliability of functional connectivity-based measures is the ICC (Noble et al., 2019; Shehzad et al., 2009; Shrout et al., 1979; see **Chapter 1 (Section 1.6)** for detailed overview on the ICC). While the ICC has been informative, it has a number of limitations. First, it is typically computed either as a voxelwise or an edge-level measure and exhibits significant variation both within and between networks (e.g., Shehzad et al., 2009); when an average value is computed across the brain, it is often quite low, suggesting that the TRT reliability for functional connectivity-based measures is poor-to-fair, at best (see Noble et al., 2019 for a review). This may, in part, reflect the fact that ICC is computed as the ratio of between-subject variance to within-subject variance (Shehzad et al., 2009). Since between-subject variance tends to account for majority of the variability in the ICC (Noble et al., 2019), achieving *high* between-subject and *low* within-subject variances for all functional connectivity measures remains a challenge (Tozzi et al., 2020; Chen et al., 2018). For these reasons, the ICC might not provide the optimal testbed for arbitrating between analysis pipelines.

Here, I test a straightforward alternative to the ICC for comparing analyses and pipelines — connectome-based fingerprint identification. This process, first proposed by Finn et al. (2015), and its ability to successfully identify (ID) an individual based on their functional connectivity

profiles (connectomes) have been described in **Chapter 1 (Section 1.7)**. I suggest that connectome fingerprint ID offers a simple, intuitive, and less computationally demanding alternative to the ICC for the assessment of analysis pipelines and novel methodologies. The same logic applies — if a particular analysis step or pipeline is advantageous, it should improve fingerprint ID.

As reviewed in the previous Chapter, several papers have already shown that ID accuracy is influenced by a number of processing choices. For example, Finn et al. (2015) reported lower ID accuracy when a cortical parcellation scheme was used to derive the functional connectivity profiles, relative to a whole-brain atlas. This may reflect the impact of differential parcel resolutions or spatial coverage (i.e., cortical vs. subcortical regions). Indeed, fingerprint ID accuracy exhibits variation across functional networks, with the highest ID accuracies being found in the medial frontal and frontoparietal networks (Cai et al., 2021; Vanderwal et al., 2019; Jalbrzikowski et al., 2019; Horien et al., 2019; Byrge and Kennedy, 2019; Vanderwal et al., 2017; Finn et al., 2015). Somewhat predictably, restricting analyses to a small subset of the most distinct connectome edges (e.g., $< 1\%$) also improves ID accuracy, producing near-ceiling accuracy (Byrge and Kennedy, 2019). Other factors known to affect ID accuracy include scan duration (Vanderwal et al., 2019; Vanderwal et al., 2017; Finn et al., 2015) and participant age (Horien et al., 2019; Vanderwal et al., 2017; Kaufmann et al., 2017; Finn et al., 2015), with the youngest participants (1.5-9 months) showing lowest ID, reflecting poor connectome distinctiveness, particularly between-sessions (Dufford et al., 2021). Recent work also indicates that denoising strategies (e.g., GSR, temporal censoring, ICA-AROMA) improve ID accuracy in early childhood (Graff et al., 2022). Importantly, and contrary to the benefits commonly inherent to increasing sample size, increasing sample size has been shown to *decrease* ID accuracy, making it more difficult to identify an individual from a group (Li et al., 2021; Waller et al., 2017).

The present study has twin aims. First, I aim to show how fingerprint accuracy can serve as a robust, easily computed, intuitive, and sensitive index for comparing or benchmarking pipeline parameters. To achieve this aim, I expand on recent efforts to maximise functional connectome identification accuracy by examining the contributions of data pre- and post-processing factors, as well as acquisition parameters, using five independent datasets spanning childhood to adulthood, obtained from the Consortium for Replicability and Reproducibility (Zuo et al., 2014a). By identifying several factors that maximise ID accuracy across these independent datasets, I describe components of a pipeline that is optimised for the reliable measurement of functional connectivity at the individual level. I suggest that the use of fingerprint accuracy to identify workflows that maximise the identifiability of individual connectomes may help improve

the identification of robust and reproducible brain-behaviour relationships for biomarker discovery.

2.2 Materials & Methods

2.2.1 Datasets

Five datasets from the Consortium for Replicability and Reproducibility (CoRR; http://fcon_1000.projects.nitrc.org/indi/CoRR/html/samples.html) were used: Beijing Normal University (BNU), New York University (NYUado, NYUadu), Southwest University (SWU), and University of Pittsburgh School of Medicine (UPSM). Ethical approval was obtained from the relevant committee at each imaging site. Each dataset included resting-state fMRI scans obtained from two imaging sessions on either the same day or several months or years apart. The demographics and fMRI acquisition parameters are provided in Table 2.1.

Dataset	NYUadu	NYUado	UPSM	BNU	SWU
Sample Size (N)	31	24	67	60	82
Age Range (years)	18-43	7-13	10-19	19-23	18-25
Gender (Male/Female)	16/15	14/10	34/33	32/28	33/49
Scan Duration (min)	06:00	06:00	05:06	08:06	08:00
Time between Scans	Days	Days	1-4 years	3 months	1 year
Scanner Manufacturer	Siemens	Siemens	Siemens	Siemens	Siemens
Scanner Model	Allegra	Allegra	TrioTim	TrioTim	TrioTim
Field Strength	3.0T	3.0T	3.0T	3.0T	3.0T
TR (ms)	2000	2000	1500	2000	2000

Table 2.1. Demographic information and fMRI parameters for the five datasets obtained from CoRR that were used in this study. All sites contributed two resting-state scans per participant.

2.2.2 Preprocessing

All data were preprocessed using the same basic pipeline (Yan et al., 2013a; Yan et al., 2013b; Craddock et al., 2013c). Structural and functional data were preprocessed using AFNI (<http://afni.nimh.nih.gov/afni/>), FSL (<http://fsl.fmrib.ox.ac.uk/>), and ANTs (<http://stnava.github.io/ANTs>). The preprocessing pipeline included: (1) volume-based motion correction; (2) grand mean scaling; (3) linear and quadratic detrending; (4) nuisance signal regression on 24 motion parameters (i.e., 3 translational, 3 rotational, their first derivatives, and the squares of each of these terms) and 12 nuisance regressors (i.e., white matter, cerebrospinal fluid, global signal (Note that GSR was one factor investigated, so for some analyses, global signal was excluded from the nuisance regressors), their squares, and the squares of their first derivatives); (5) spatial

smoothing using a Gaussian kernel at 6mm FWHM; and (6) temporal filtering ($0.009 \leq f \leq 0.1$ Hz). Functional to session-averaged anatomical co-registration was performed using boundary-based registration in FSL (Greve and Fischl, 2009). Diffeometric normalisation of individual anatomical to the MNI152 standard space was carried out using ANTs; the same non-linear transformation was applied to the fMRI data. Participants were excluded if either scan exhibited a root-mean-square framewise displacement (rmsFD) > 0.20 mm (Jenkinson et al., 2002).

2.2.3 Data Pre- and Post-Processing Parameters

To demonstrate how fingerprint ID accuracy can be used to test pipeline parameters, and to compare the utility of ID accuracy vs. ICC for this purpose, I examined how several data acquisition, pre-processing, and post-processing parameters affected ID accuracy and ICC (Figure 2.1), specifically: (1) GSR; (2) parcellation schemes (i.e., scheme, spatial coverage, resolution); (3) network organisation; (4) use of discriminative edges; (5) scan duration; and (6) sample size.

2.2.3.1 GSR

Although debated, GSR is a highly effective denoising strategy (Parkes et al., 2018; Satterthwaite et al., 2017; Ciric et al., 2017). It remains the most effective strategy for reducing associations between motion and functional connectivity-based measures (Power et al., 2017; Power et al., 2014; Yan et al., 2013a; Yan et al., 2013b) in addition to increasing the biological plausibility of patterns of functional connectivity (Weissenbacher et al., 2009; Fox et al., 2009). I examined the impact of including vs. excluding the global signal from the 36-parameter nuisance signal regression step during data preprocessing.

2.2.3.2 Parcellation Schemes (Selection, Spatial Coverage & Resolution)

Parcellation schemes differ in aspects such as their spatial coverage (e.g., cortex, subcortex), space (e.g., volume, surface), and resolution (e.g., 10-1000). I parcellated the fMRI data using four publicly available brain atlases including Shen 268 (Shen et al., 2013), Glasser 360 (Glasser et al., 2016), MIST 444 (Urchs et al., 2017), and Schaefer 1000-17Networks (Schaefer et al., 2018), to examine whether these different commonly applied schemes are associated with differences in ID accuracy. I examined the influence of spatial brain coverage on ID accuracy by comparing cortical-only parcellations to parcellations that also included subcortical and cerebellar regions. Finally, I assessed the relationship between parcel resolution and ID accuracy by examining two parcellation schemes with 8 (Schaefer 17Networks, 100-1000) and 10 (MIST, 7-1095) levels of dimensionality.

2.2.3.3 Network Organisation

Several studies have demonstrated that restricting the fingerprint to specific functional networks can increase ID accuracy (Jalbrzikowski et al., 2019; Horien et al., 2019; Finn et al., 2015). To assess differences between functional networks in terms of ID accuracy, I examined 8 functional “Finn” networks: (1) medial frontal; (2) frontoparietal; (3) default mode; (4) subcortical and cerebellum; (5) motor; (6) visual I; (7) visual II; and (8) visual association (Finn et al., 2015), and 7 “Yeo” networks: (1) ventral attention; (2) dorsal attention; (3) frontoparietal; (4) default mode; (5) somatomotor; (6) limbic; and (7) visual (Yeo et al., 2011).

2.2.3.4 Discriminative Edges

Fingerprint identification is influenced by the degree to which different edges contribute to the identification of an individual (Vanderwal et al., 2019; Horien et al., 2019; Byrge and Kennedy, 2019; Finn et al., 2015). For example, Vanderwal et al. (2019) showed that increasing the number of discriminative edges improves ID accuracy. I employed the discriminative power (DP) measure to quantify the usefulness of an edge in successfully identifying an individual from a group. The DP of a given edge reflects the empirical probability of an individual being related to I relative to another individual based on the same edge. It is computed as the product of the edge values between-sessions from an individual and compared with the product of the edge values between-sessions from other individuals in a dataset (see Horien et al., 2019, Finn et al., 2015 for details). Edges with a high DP tend to yield similar values within individuals than between individuals, i.e., within-subject similarity is greater than between-subject similarity. To compute ID accuracy using the most discriminative edges across all individuals in the five datasets, I selected the edges whose DP values fell between the 25th and 99th percentiles of the functional connectome, i.e., I retained the top 1-75% of edges in the whole-brain connectome to compute ID accuracy for each dataset.

2.2.3.5 Scan Duration

Converging evidence has revealed that the TRT reliability of functional connectivity-based measures depends on scan duration (Noble et al., 2019; Elliott et al., 2019; Noble et al., 2017; O’Connor et al., 2017; Mueller et al., 2015; Birn et al., 2013; Anderson et al., 2011; Van Dijk et al., 2010; Shehzad et al., 2009) and that longer scan duration improves ID accuracy (Vanderwal et al., 2021; Vanderwal et al., 2017; Airan et al., 2016; Finn et al., 2015). I assessed the relationship between scan duration and ID accuracy by increasing the scan lengths from 0.5 min to the full acquisition time of each dataset (scan parameters for each site are described in Table 2.1). Of note, the range of the scan lengths is limited based on the relatively short scan durations of the five datasets (e.g., the maximum scan length is approximately 8 min derived from the BNU and SWU datasets).

2.2.3.6 Sample Size

While sample size typically improves the TRT reliability of functional connectivity-based measures (Cho et al., 2020; Elliott et al., 2019), identification accuracy tends to decrease with increasing sample sizes (Li et al., 2021; Waller et al., 2017). To assess the impact of sample size on ID accuracy, and to avoid site/sample effects, I pooled the five datasets ($N = 264$), then randomly selected subsets of individuals such that $N \in \{20, 40, 60, 80, 100, 120, 140, 160, 180, 200, 220, 240, 264\}$ *without* replacement and computed ID accuracy over 100 of each N iterations.

2.2.4 Functional Connectivity Matrices — Fingerprints

For every individual and each scan, the mean timeseries of each ROI within a given parcellation scheme was extracted and the Pearson’s correlation coefficient (r) between all possible ROI pairs was computed, yielding two functional connectivity matrices (fingerprints) per individual that were subsequently used in the ID procedure.

2.2.5 Identification Procedure

To identify an individual based on their functional connectivity fingerprint, I followed the identification procedure as previously described in Finn et al. (2015) and **Chapter 1 (Section 1.7)**. The mean binary identification (BID) accuracy for each dataset and each combination of analysis factors was computed as the number of correctly identified participants divided by N . To reach a final ID accuracy per dataset, I recomputed mean BID by exchanging the roles of the database-target matrix ($\mathbf{D} \leftrightarrow \mathbf{T}$), and averaging across the two mean BIDs.

In addition to BID, I computed the relative rank (RR), which is a continuous measure ranging from 0 to 1 that quantifies the degree of “confusion” for inaccurately identified individuals. RR was calculated as the j th position of the rank ordered correlation that matches the predicted identity X_j in $\mathbf{T}$ with X_i in $\mathbf{D}$ divided by N . For example, if the ranked ordered correlation of participant 1 in the SWU dataset ($N = 82$) is at the first position, then $RR = 1/82 = 0.012$ whereas if the ranked ordered correlation for the same participant is at the last position, then $RR = 82/82 = 1$. The fewer individuals inaccurately ranked above their true identity, the lower the degree of confusion and lower the RR ($RR \rightarrow 0$).

2.2.6 TRT Reliability

As described in **Chapter 1 (Section 1.6)**, I computed the TRT reliability, as indexed by the ICC, for each edge and each dataset. I applied the ICC(2,1) model to refer to “absolute agreement” for which the sources of errors are known (e.g., multiple sessions) and modelled as random effects

(Noble et al., 2019). For each dataset, the mean ICC was obtained by averaging the edgewise ICC values in addition to setting the negative values to 0 to indicate complete unreliability (Noble et al., 2019; Noble et al., 2017; Bartko, 1976).

2.2.7 Comparing Analysis Pipelines Using ID Accuracy vs. ICC

For each dataset, the identification accuracy was computed with each of the data acquisition, pre-processing, and post-processing parameters, and then compared with a “standard” functional connectivity pipeline as described in **Sections 2.2.2 and 2.2.4**. For each dataset, the ICC was further computed for several configurations as described in **Sections 2.2.3**. Given that I assessed how fingerprint ID accuracy can be applied to test analysis pipeline parameters, I did not perform permutation testing to assess the statistical significance of the ID accuracies themselves across the five datasets.

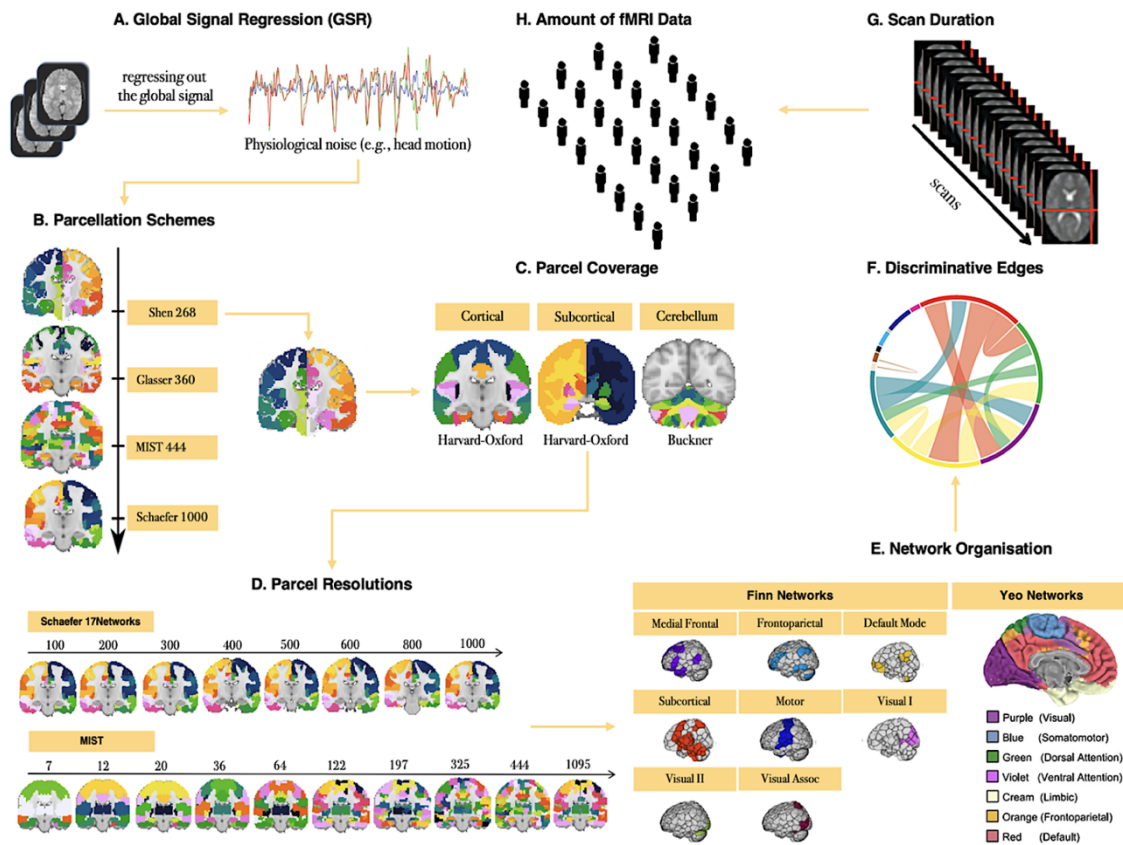


Figure 2.1. Analysis framework to describe an optimal pipeline for maximising connectome-based identification. Data pre- and post-processing parameters: A. Applying GSR as a denoising strategy. B. Describing the functional connectome with four publicly available parcellation schemes such as Shen 268, Glasser 360, MIST 444, and Schaefer 1000-17Networks. C. Describing cortical, subcortical, and cerebellar regions from the Harvard-Oxford and Buckner atlases. D. Computing the functional connectome using multi-resolution parcellation schemes with 8-10 levels of dimensionality. E. Examining the Finn Networks and Yeo Networks. F. Using the top 1-75% of most discriminative edges. G. Examining scan duration — from 0.5 min to the full acquisition time of each dataset. H. Examining the impact of sample size (N) by pooling data

across the five datasets and resampling such that $N \in \{20, 40, 60, 80, 100, 120, 140, 160, 180, 200, 220, 240, 264\}$.

2.3 Results

2.3.1 Identifiability of the individual functional connectome

Using whole-brain functional connectomes collected from two scans per participant derived from CoRR, I initially applied a standard functional connectivity pipeline which involves minimal data pre-processing (including GSR), post-processing (i.e., whole-brain Shen 268 parcellation), and acquisition (i.e., full scan duration and sample size) parameters to compute the identification accuracy across the five datasets. I first computed binary identification (BID) such that the BID accuracy is 1 if an individual is successfully identified and 0 if they are not, using all five datasets. BID accuracies ranged from 42% to 77% which were averaged across the two database-target matrix configurations (Figure 2.2A and Figure 2.2B). On average for both directions ($\mathbf{D} \rightarrow \mathbf{T}$ and $\mathbf{T} \rightarrow \mathbf{D}$), the NYU developmental dataset (NYUado; $N = 24$) achieved the strongest identification accuracy (BID = 77%) whereas the UPSM adolescent group ($N = 67$) had the lowest (BID = 42%). I also employed relative rank (RR) to measure the degree of confusion for misidentifying an individual from a group such that the lower the RR, the lower the confusion for incorrectly identifying that individual. RR showed that the degree of confusion varied from 6% to 13% for the first direction ($\mathbf{D} \rightarrow \mathbf{T}$) (Figure 2.2A), and 4% to 16% for the second direction ($\mathbf{T} \rightarrow \mathbf{D}$) (Figure 2.2B). On average, the adult dataset (SWU) exhibited the lowest degree of confusion for incorrectly identifying an individual (RR = 5%) between-sessions whereas the UPSM adolescent dataset exhibited the highest degree of confusion (RR = 15%).

2.3.2 TRT reliability of the functional connectome is consistently poor

I then evaluated the TRT reliability of the functional connectome by applying the ICC(2,1) model using the standard functional connectivity pipeline across the five datasets. I computed the average ICC across 268 x 268 edges for the Shen ROIs. I found that the mean TRT reliability was consistently “poor” (i.e., $\text{ICC} < 0.4$) across the independent datasets (Figure 2.2C). The highest mean TRT reliability was observed for the adult dataset (SWU; $\text{ICC} = 0.29$) whereas the lowest mean TRT reliability was found for the adolescent group (UPSM; $\text{ICC} = 0.15$).

2.3.3 Applying GSR as a denoising strategy improves identification accuracy

While investigating the impact of data pre-processing factors on identification accuracy across the five datasets, I assessed the contribution of GSR by computing identification accuracy following preprocessing *with* and *without* GSR (Figure 2.2D). For the purpose of this comparison, all other noise regressors were held constant (i.e., the global signal was the only regressor

removed from the 36-parameter nuisance signal regression). The range of the BID accuracies was higher *with* GSR (42-77%) than *without* GSR (35-69%); GSR significantly increased BID accuracies across all five datasets (paired sample t-test; $t(4) = 6.14$, $P = 0.004$).

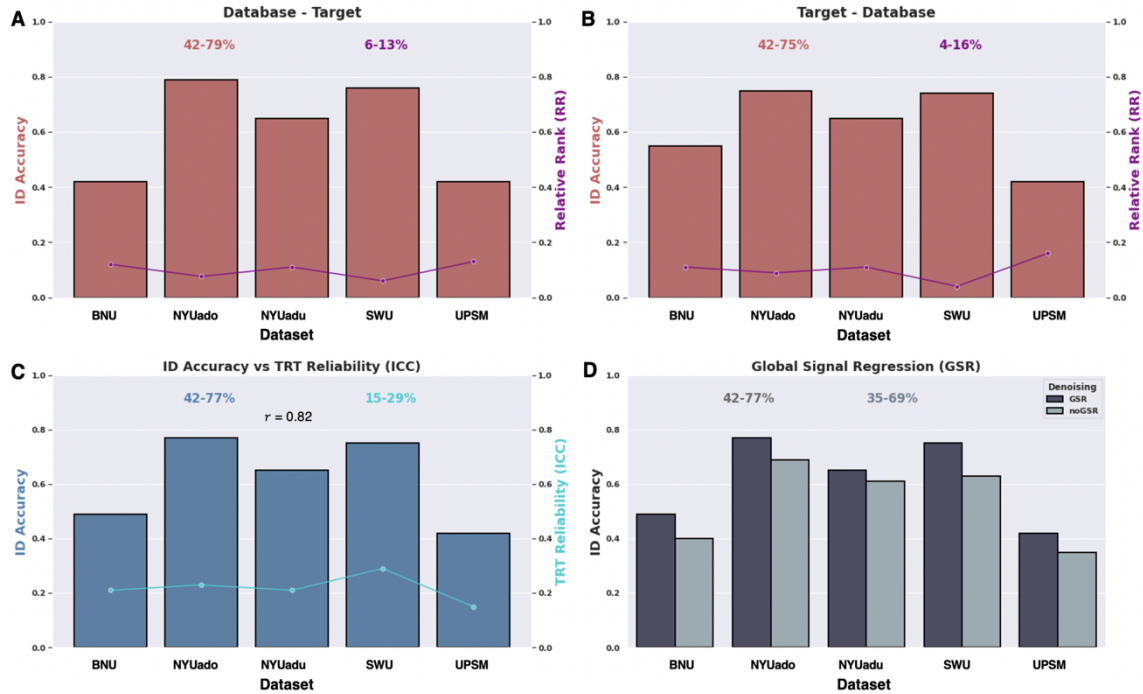


Figure 2.2. (A) ID accuracy derived from BID in addition to the degree of confusion as indexed by RR for the database-target matrix configuration. (B) ID accuracy derived from BID in addition to the degree of confusion as indexed by RR for target matrix-database configuration. (C) ID accuracy and TRT reliability derived from the respective fingerprint approach using BID and the ICC(2,1) model. The Pearson's correlation coefficient (r) between ID accuracy and ICC is 0.82. (D). ID accuracy derived from BID with and without applying GSR as a denoising strategy. The five independent datasets are as follows: BNU ($N = 60$), NYUado ($N = 24$), NYUadu ($N = 31$), SWU ($N = 82$), and UPSM ($N = 67$).

2.3.4 Finer cortical parcellations yield stronger identification accuracy

To understand how different parcellation features influence the identification accuracy, I computed functional connectivity using four parcellation schemes with different spatial resolutions and spatial coverage. I found that the BID accuracies differed significantly across the parcellation schemes (repeated measure one-way ANOVA test; $F(3,12) = 31.1$, $P < 0.001$). There was a tendency to achieve greater BID accuracies across the five datasets when the functional connectome was derived from finer parcellation schemes, i.e., parcellations with higher resolutions. While the lowest BID accuracies were obtained with the Shen 268 parcellation (42-77%), the highest BID accuracies were found with the Schaefer 1000-17Network parcellation (68-94%) (Figure 2.3A). Both Glasser 360 and MIST 444 parcellation schemes produced intermediate BID accuracies ranging from 55% to 86% across the five datasets (Figure 2.3A).

Because the parcellation schemes differ in terms of their resolutions, I applied the multi-resolution Schaefer 17Networks (i.e., 100-1000) and MIST (i.e., 7-1095) parcellations to examine the relationship between parcel resolution and identification accuracy. There was a non-linear relationship between parcel resolution and BID accuracy that was evident across the five datasets (Figure 2.3B). BID showed that there were large gains in ID accuracy as parcel resolution increased from 7 to 300, with smaller but still significant gains thereafter (Figure 2.3B). Similarly, there was a non-linear relationship between parcel resolution and RR across the five datasets such that there was a rapid decline in RR from 7 to 300 parcels, and then plateaued until reaching the finest resolutions (Figure 2.3B).

I also investigated how the inclusion of the cortex and subcortex including the cerebellum influences identification accuracy. Fingerprint ID accuracies differed significantly across different combinations of the spatial coverage (repeated measure one-way ANOVA test; $F(5,20) = 54.4$, $P < 0.001$). The highest ID accuracies were derived when the functional connectome included cortical regions (41-77%), with the addition of subcortical and cerebellar regions making little-to-no difference to ID accuracy. Fingerprints computed on the basis of subcortical (17-49%) and/or cerebellar (6-36%) structures alone were associated with much lower BID across the five datasets (Figure 2.3C). It should be noted, however, that there are differences in the macroscale division of the cortex, subcortex, and cerebellum — the number of brain regions within the cortex is usually much greater than that within subcortical and cerebellar structures in most parcellation schemes, which likely impacts the stability of the fingerprints restricted to these compartments.

2.3.5 Identification accuracy across brain networks

To examine variation in fingerprint identifiability across functional networks, I compared BID across eight “Finn” networks and seven “Yeo” networks. For the Finn networks, BID was highest when the fingerprints were restricted to the medial frontal (except for the UPSM dataset) and frontoparietal networks, relative to the other functional networks (Figure 2.4A). In contrast, fingerprinting using the visual networks (i.e., Visual I, Visual II, Visual Association) produced the lowest BID accuracies (Figure 2.4A). Given previous research showing that combining medial frontal and frontoparietal networks improved BID accuracies (Horien et al., 2019; Finn et al., 2015), I examined the impact of combining these functional networks into the “frontoparietal networks”. Restricting the fingerprint to these frontoparietal networks was associated with significantly greater identification accuracy compared to the whole-brain functional connectome (paired sample t-test; $t(4) = 2.83$, $P = 0.047$). When I applied the Yeo networks, BID revealed that the fingerprints restricted to the frontoparietal (except for the NYUadu dataset; 57-87%) and default mode networks outperformed the other functional networks across the five datasets

(Figure 2.4B). Across the five datasets, fingerprinting with the limbic network exhibited the lowest BID accuracies (21%-67%) followed by fingerprinting with the visual network (25%-69%) (Figure 2.4B). Similarly, there are network-level differences in how these functional networks are defined — brain networks tend to have more cortical regions compared to subcortical and/or cerebellar areas, thus providing access to a greater number of cortical edges for fingerprint identification.

2.3.6 Greater number of discriminative edges yields stronger identification accuracy

To examine whether restricting the fingerprint to discriminative edges improves identification accuracy, I restricted the fingerprint to those edges whose DP values fell between the 25th and 99th percentiles of the functional connectome for each dataset (i.e., selecting the top 1-75% of all edges, ordered by discriminability). I observed a non-linear relationship between the number of discriminative edges and ID accuracy across the five datasets (Figure 2.5). When the fingerprint was restricted to the top 1% of DP edges, BID accuracies (21-64%) were significantly lower compared to using all the edges from the whole-brain connectome (paired sample t-test; $t(4) = 7.21$, $P = 0.002$). When the fingerprint was computed with the top 10% of DP edges, BID accuracies (37-77%) were significantly higher compared to the top 1% of DP edges (paired sample t-test; $t(4) = 8.22$, $P = 0.001$). On the other hand, when the fingerprint was computed with the top 75% of DP edges, BID accuracies slightly improved (43-77%) compared to the whole-brain connectome. However, this improvement was not significantly different from the whole-brain connectome (paired sample t-test; $t(4) = 1.84$, $P = 0.14$). Similar observation was noted when comparing the BID accuracies from the top 10% of DP edges to the whole-brain connectome (paired sample t-test; $t(4) = -0.14$, $P = 0.90$). The non-linear relationship between number of discriminative edges and ID accuracy is striking — there is a large jump in BID accuracies between 1% and 10%, but little-to-no improvement after that, suggesting that good BID accuracy can be achieved with as little as the top 10% of the most discriminative edges.

2.3.7 Longer scan duration facilitates stronger identification accuracy

I evaluated the effort of acquiring increasing scan length to boost identification accuracy by varying the scan duration from 0.5 min to the full acquisition length of each dataset. There was a non-linear relationship between scan duration and identification accuracy across the five datasets (Figure 2.6A), such that BID was very negatively impacted by very short fMRI scans (e.g., 0.5 min — ranging from 5% to 21% across the five datasets). The BID accuracies were optimal (42-77%) when the full acquisition length of each dataset was used to compute the functional connectivity matrices of each individual. RR revealed a similar pattern such that the degree of confusion for unsuccessfully identifying an individual was highest (23-41%) with short

acquisition time (e.g., 0.5 min) but minimal (5-15%) when the entire scan length was used to derive the functional connectivity profiles of the individuals across the five datasets.

2.3.8 Increasing sample size does not necessarily improve identification accuracy

I also examined the effect of sample size (N) on identification accuracy by creating random subgroups of individuals $N \in \{20, 40, 60, 80, 100, 120, 140, 160, 180, 200, 220, 240, 264\}$ across the five datasets. For each N , 100 samples were selected, and the average BID was computed across samples. There was a non-linear relationship between sample size and BID accuracy such that the ability to identify an individual based on their functional connectivity profiles decreased non-linearly as the sample size increased (Figure 2.6B), with the steepest decrease observed as sample size increased from 20 to 100, and a more gradual decrease in BID as sample size increased further to 264. The overall change across sample sizes was relatively small, however — BID ranged from 70% when $N = 20$ to 53% when $N = 264$. As a mathematical reality, increasing sample sizes also increases the potential for “confusion” (high correlation) amongst individuals — it increases the pool of potential matches, thus decreasing identification accuracy. Interestingly, the aggregated/subsampling approach produced higher BIDs for sample sizes equivalent to those of the UPSM, BNU, and NYUadu sites, relative to the BIDs for those sites alone, suggesting some site or sample effects that were overcome by the aggregating and subsampling approach. Further, and contrary to observations for the preceding manipulations, for which the trend for RR was inverted relative to the trend for BID, RR *decreased* slightly but non-linearly from 12% to 7% as sample size increased (Figure 2.6B).

2.3.9 Selecting a fingerprint-optimised pipeline

On the basis of the results presented above, I identified two optimal fingerprint-informed pipelines featuring those parameters associated with the *highest* BID accuracy (Table 2.2; Figure 2.7). The following two pipelines have been described given that predictive networks such as the frontoparietal networks and finer cortical parcellation schemes such as the Schaefer 1000-17Network scheme, separately, encapsulate the most distinct features of an individual’s functional connectome:

- Pipeline 1: minimal pre-processing, GSR, Schaefer 1000-17Network parcellation, full scan length, and full sample size;
- Pipeline 2: minimal pre-processing, GSR, frontoparietal networks (medial frontal + frontoparietal) derived from Finn networks, full scan length, and full sample size.

As expected, relative to the standard pipeline (minimal pre-processing, GSR, whole-brain Shen 268 parcellation, full scan length, and full sample size), fingerprint ID accuracy was significantly

higher for both Pipeline 1 (paired sample t-test; $t(4) = 6.11$, $P = 0.004$) and Pipeline 2 (paired sample t-test; $t(4) = 2.83$, $P = 0.047$). Notably, some datasets (e.g., NYUado; SWU) exhibited BID of 90% or above following processing with Pipeline 1.

2.3.10 Applying functional connectome fingerprinting informed pipelines to improve ICC

To further assess the utility of these two fingerprint-informed pipelines, I computed TRT reliability as indexed by the ICC for each dataset following the application of these pipelines to compute functional connectivity (Table 2.2; Figure 2.7). The ICC increased significantly with Pipeline 1 (paired sample t-test; $t(4) = 7.22$, $P = 0.002$) and Pipeline 2 (paired sample t-test; $t(4) = 7.22$, $P = 0.002$), although it is worth noting that the ICCs across the five datasets would still be considered “poor” according to the Cicchetti and Sparrow (1981) definition.

Pipeline	NYUadu	NYUado	UPSM	BNU	SWU
◊Standard ICC	0.21	0.23	0.15	0.21	0.29
*Pipeline 1 ICC	0.25	0.28	0.19	0.29	0.35
‡Pipeline 2 ICC	0.29	0.33	0.21	0.35	0.40
◊Standard BID	0.65	0.77	0.42	0.49	0.75
*Pipeline 1 BID	0.78	0.94	0.68	0.79	0.90
‡Pipeline 2 BID	0.81	0.86	0.44	0.77	0.83

Table 2.2. Gains in ICC and ID accuracy derived from BID across the five datasets obtained from the standard and optimal fingerprint-informed pipelines. Standard: GSR + Shen 268 parcellation + full scan length + full sample size. *Pipeline 1: GSR + Schaefer 1000-17Network parcellation + full scan length + full sample size. ‡Pipeline 2: GSR + frontoparietal networks derived from Finn Networks + full scan length + full sample size.

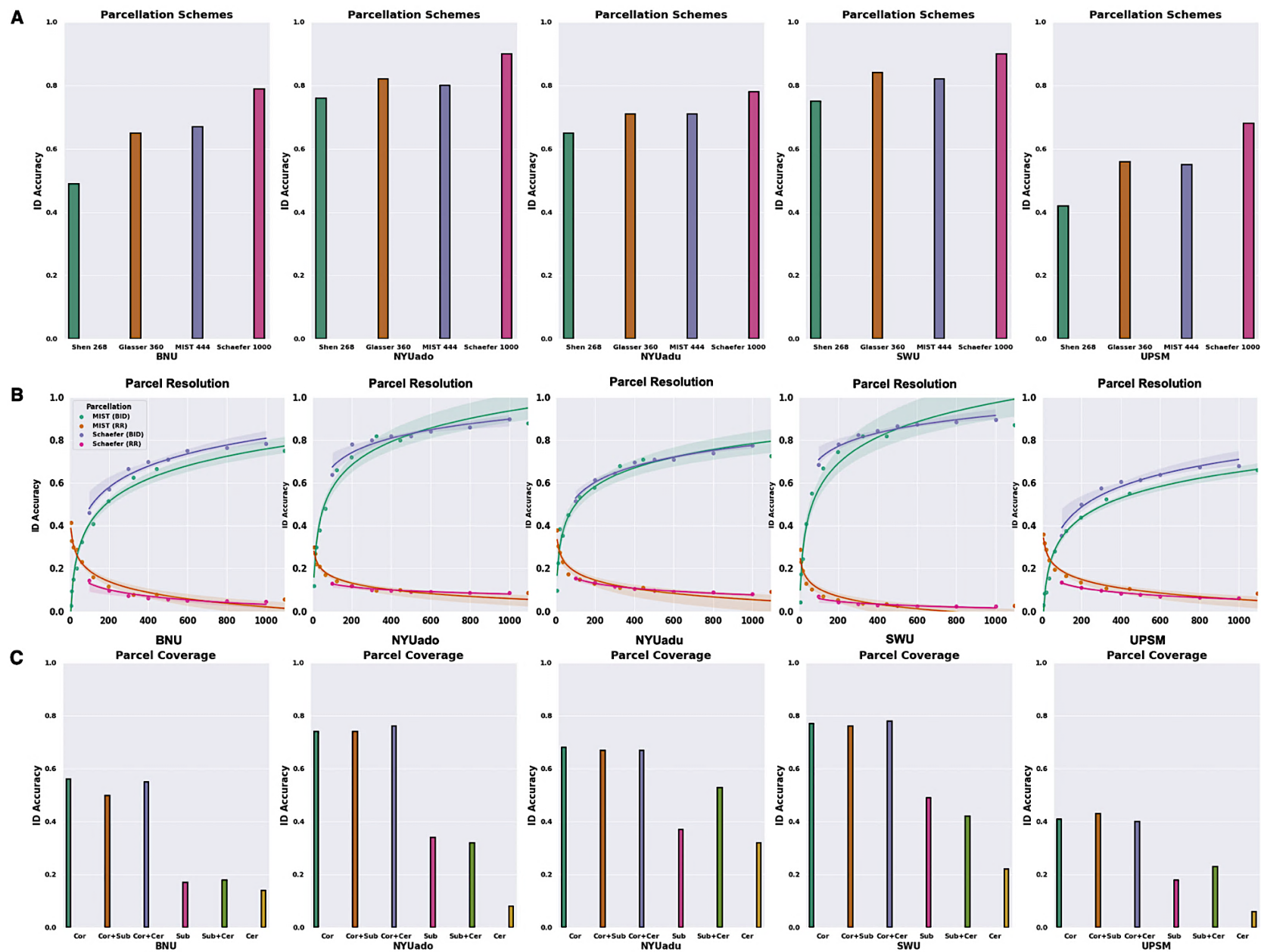


Figure 2.3. (A) ID accuracy obtained from BID across the five independent datasets using four publicly available parcellation schemes including Shen 268, Glasser 360, MIST 444, and Schaefer 1000-17Networks. (B) Relationship between parcel resolution and ID accuracy measured by BID in addition to the degree of confusion as indexed by RR using 8 and 10 levels of parcel dimensionality derived from the Schaefer 17Networks and MIST parcellation schemes across the five independent datasets. (C) ID accuracy derived from BID based on the spatial coverage of the functional connectome covering cortical, subcortical, and cerebellar structures across the five independent datasets. Cor: Cortical; Cor+Sub: Cortical and Subcortical; Cor+Cer: Cortical and Cerebellum; Sub: Subcortical; Sub+Cer: Subcortical and Cerebellum; and Cer: Cerebellum.

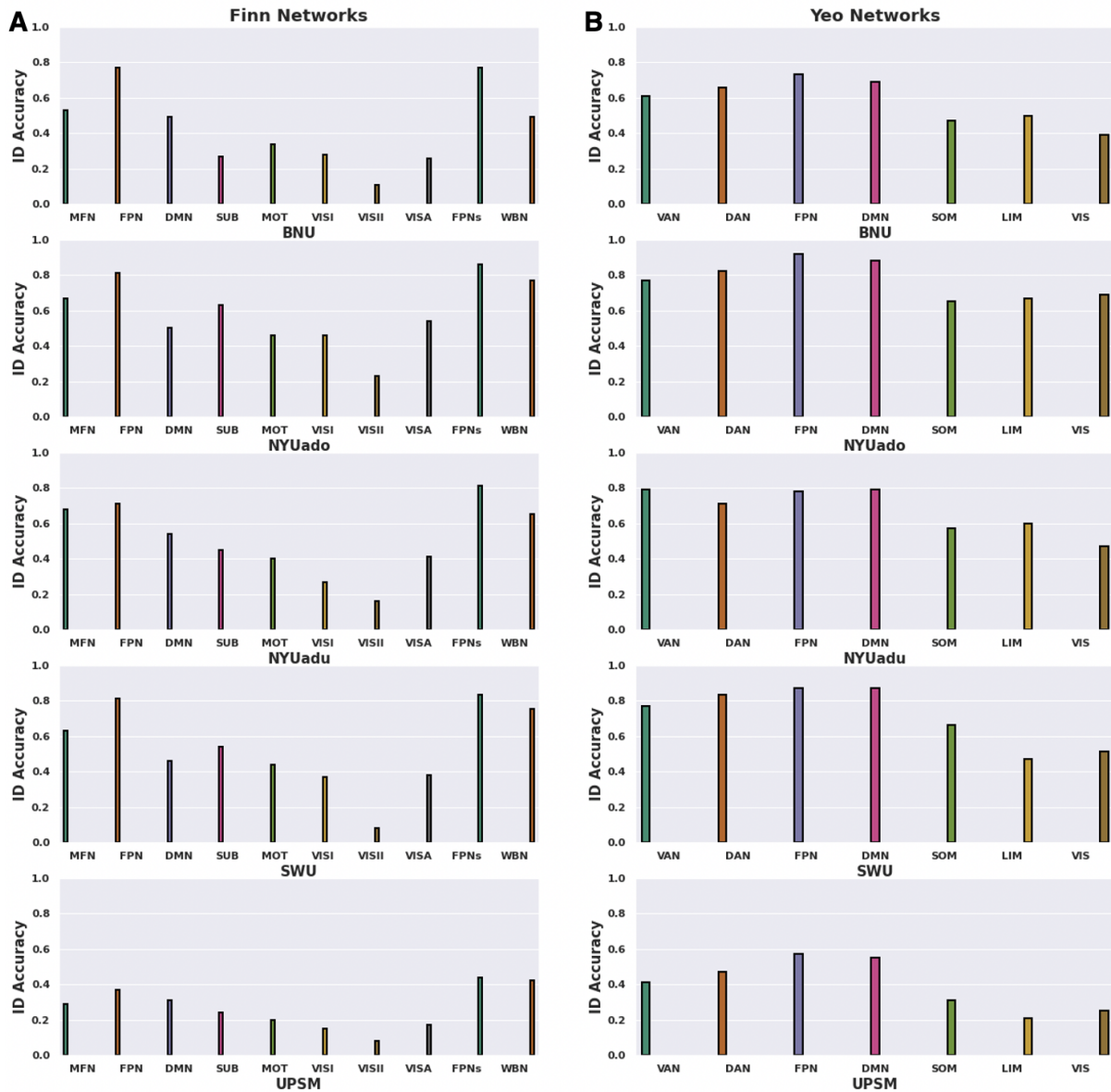


Figure 2.4. (A) ID accuracy derived from BID for the Finn Networks in addition to the combined medial frontal and frontoparietal networks [FPNs] and whole-brain connectome [WBN] across the five datasets. The Finn Networks are as follows: (1) medial frontal [MFN]; (2) frontoparietal [FPN]; (3) default mode [DMN]; (4) subcortical [SUB]; (5) motor [MOT]; (6) visual I [VISI]; (7) visual II [VISII]; and (8) visual association [VISA]. (B) ID accuracy derived from BID for the Yeo Networks across the five datasets. The Yeo Networks are as follows: (1) ventral attention [VAN]; (2) dorsal attention [DAN];

(3) frontoparietal [FPN]; (4) default mode [DMN]; (5) somatomotor [SOM]; (6) limbic [LIM]; and (7) visual [VIS].

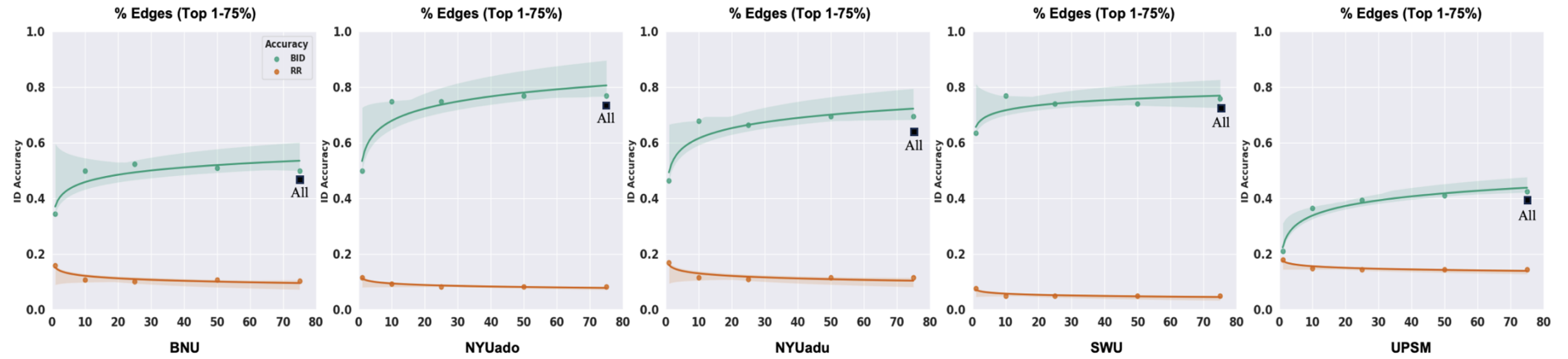


Figure 2.5. Relationship between number of discriminative edges and ID accuracy derived from BID in addition to the degree of confusion as indexed by RR across the five independent datasets. The discriminative edges are selected based on their DP values falling between the 25th and 99th percentiles of the whole-brain connectome, i.e., the top 1-75% of edges in the whole-brain connectome were retained to compute the ID accuracy of each dataset. The black box represents the ID accuracy derived from BID when all the edges in the whole-brain connectome were used for each dataset.

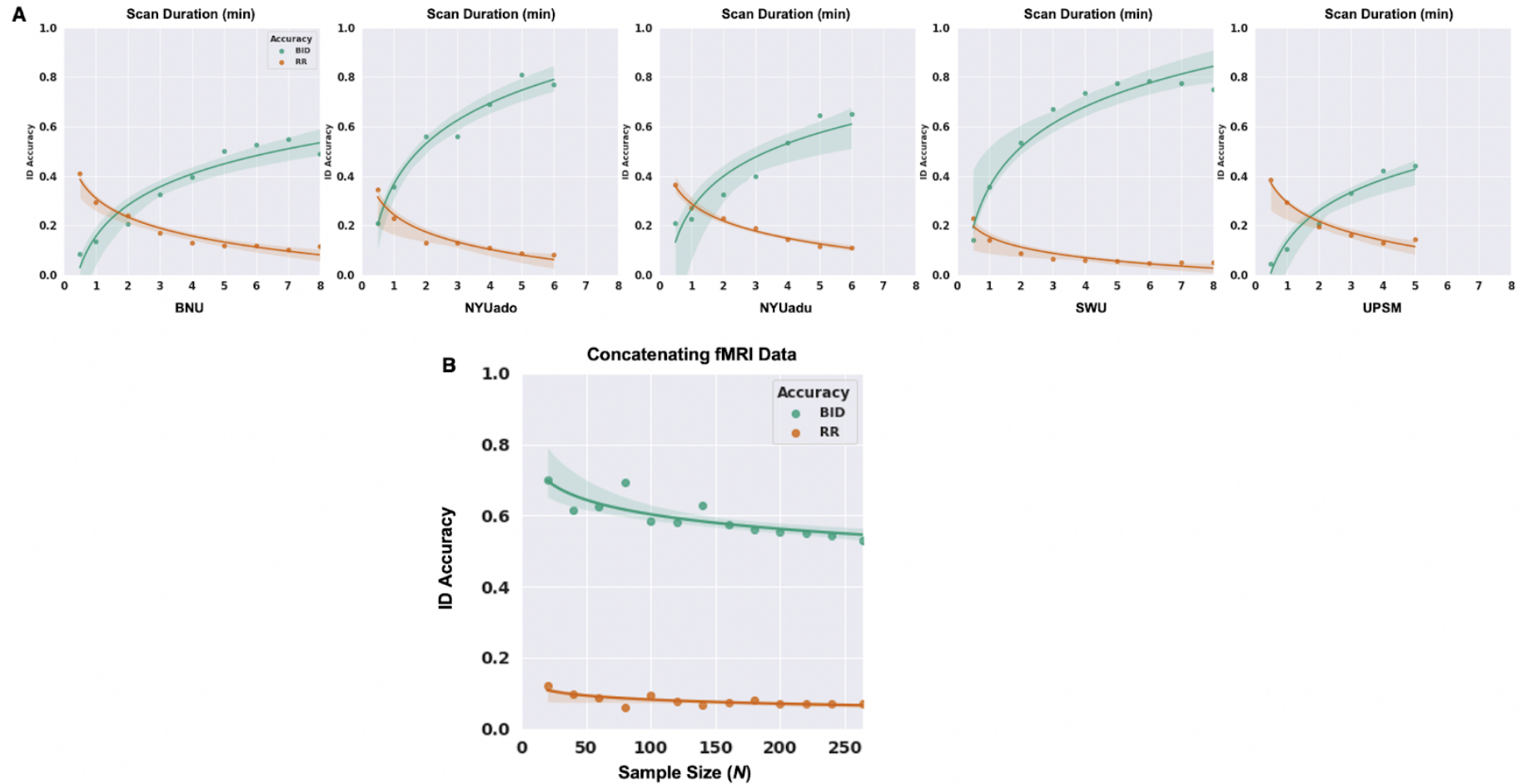


Figure 2.6. (A) Relationship between scan duration and ID accuracy measured by BID in addition to the degree of confusion as indexed by RR across the five independent datasets. The scan duration ranges from 0.5 min to the full acquisition time of each dataset. (B) Relationship between amount of resting-state fMRI data and ID accuracy measured by BID in addition to the degree of confusion as indexed by RR. The resting-state fMRI scans were aggregated across the five datasets ($N = 264$) for each imaging session, and then subsets of individuals were randomly selected across 100 iterations such that $N \in \{20, 40, 60, 80, 100, 120, 140, 160, 180, 200, 220, 240, 264\}$.

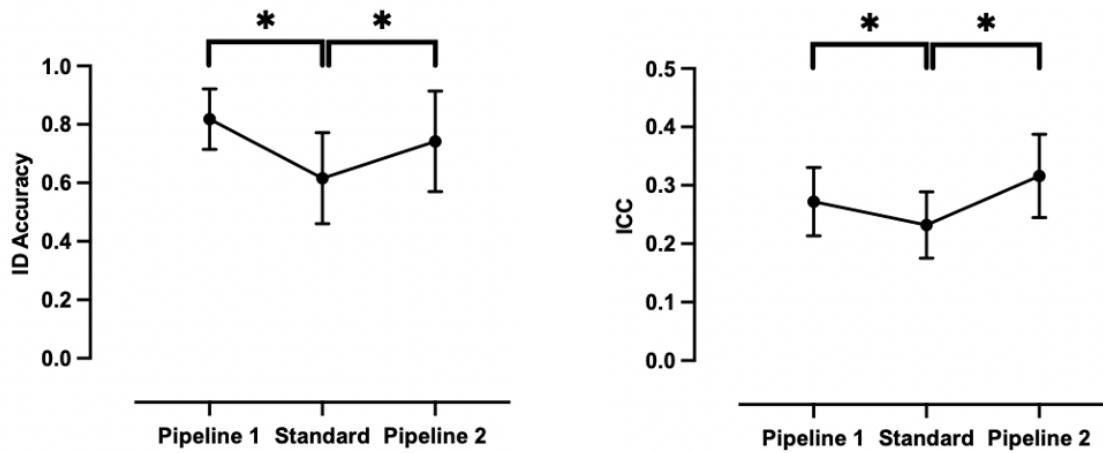


Figure 2.7. ID accuracy and TRT reliability of the functional connectome measured by BID and ICC, respectively, using the standard and optimal fingerprint-informed pipelines. Dots and error bars represent the mean and SD of the BID accuracies and ICC values. Standard: GSR + Shen 268 parcellation + full scan length + full sample size. Pipeline 1: GSR + Schaefer 1000-17Network parcellation + full scan length + full sample size. Pipeline 2: GSR + frontoparietal networks derived from Finn networks + full scan length + full sample size. $*P < 0.05$, two-tailed paired sample t-test. For ID accuracy, the effect sizes measured by Cohen’s d are: 1.5 (Pipeline 1) and 0.8 (Pipeline 2). For TRT reliability, the effect sizes measured by Cohen’s d are: 1.0 (Pipeline 1) and 1.6 (Pipeline 2).

2.4 Discussion

The present study aimed to demonstrate how connectome-based fingerprint identification accuracy could be used as a simple, intuitive, and sensitive index to adjudicate between functional connectivity-based analysis parameters. Across five independent Open Science datasets spanning childhood to adulthood, I found that functional connectome fingerprint identification accuracy ranged from 42% to 77%. I then tested the effect of a range of analysis parameters on ID accuracy and found that GSR, fine-grained (i.e., ~1000 ROIs) brain parcellations encompassing cortical regions, increasing number of discriminative edges, and longer scan duration improved identification accuracy — for some datasets to $\geq 90\%$. Interestingly, pooling fMRI data across datasets to increase sample size decreased, rather than increased, ID accuracy. Importantly, the effects of many of these manipulations on ID accuracy were quantitatively and qualitatively unambiguous, allowing for easy selection of the “best performing” parameters. These “best performing” parameters were then combined into two analysis pipelines that optimise the detection of individual differences in the functional connectome.

Following the investigation of the impact of a set of data acquisition, pre-, and post-processing parameters on fingerprint accuracy, I described two analysis pipelines as follows: (1) Pipeline 1 — GSR, cortical-based parcellation scheme with 1000 parcels, full scan length, and sample size, and (2) Pipeline 2 — GSR, frontoparietal networks (medial frontoparietal and frontoparietal

networks derived from the Finn networks), full scan length, and sample size. The two fingerprint-optimal pipelines were described as the first applied a finer cortical-based parcellation scheme whereas the second employed the most predictive functional networks, i.e., the frontoparietal networks. While Pipeline 1 boosted fingerprint accuracies to 68%-94%, Pipeline 2 increased the accuracies to 44%-86%. The utility of these pipelines was also supported by the fact that they significantly increased TRT reliability (indexed by the ICC), relative to the “standard” pipeline — although it is notable that the ICC values remained “poor”, according to the traditional categorisation (i.e., $ICC < 0.4$). This echoes the TRT reliability literature which suggests ICC for functional connectivity metrics is generally poor (mean: 0.29; 95% CI: 0.23-0.36; Noble et al., 2019). While a direct quantitative comparison of ICC and fingerprint BID is not possible (since these metrics are computed differently and cannot be converted to e.g., z-scores), we can compare them qualitatively given that fingerprint accuracy tracks the ICC values in the five datasets. For example, an $ICC = 0.40$ implies that the TRT reliability of a given functional connectivity measure is fair/moderate (see **Section 1.6**), but is above the upper meta-analytic estimate for the ICC across studies (Noble et al., 2019). This means that 40% of the total variance in the measurement of interest is attributed to agreement between the “raters” while the remaining 60% of the variance corresponds to individual variation or measurement errors. In contrast, an $ID = 40\%$ would be considered a low level of self-self correlation within a sample. However, it is clear from Figure 2.2C that the ICC and BID accuracy track each other quite well, suggesting good correspondence. What these comparisons suggest is that, while the ICC values that are typical for the field are considered to reflect “poor” reliability, fingerprint identification accuracy suggests a more optimistic perspective on functional connectome stability. To take one example, BID accuracy for the NYUado dataset processed with Pipeline 1 is 94% ($ICC = 0.28$), indicating that, on average, 23 of 24 participants are correctly identified on the basis of their functional connectome. Similarly, the BID accuracy derived from Pipeline 1 for the SWU dataset is 90% ($ICC = 0.35$), which successfully identified, on average, 74 of 82 participants. As described in **Chapter 1 (Section 1.7.2)**, my findings show that connectome-based identification provides a metric of reproducibility of an individual’s functional connectome which is stable on both short-term (days) and long-term (years) from childhood to adulthood.

Although I found that GSR was beneficial in terms of its impact on fingerprint ID accuracy, I note that it remains the subject of dispute and there is no clear consensus on its application in publicly available pre-processing pipelines (Li et al., 2021; see **Chapter 1 (Section 1.6.3)**). This disagreement has been traditionally rooted in evidence that showed that GSR introduces negative correlations between brain regions (Saad et al., 2012; Weissenbacher et al., 2009; Murphy et al., 2009), but often countered by its efficacy at mitigating the effects of motion and other sources of noise (Graff et al., 2022; Yan et al., 2013b). The debate is ongoing with recent work showing that

the global signal could reflect neural information of behavioural relevance: for example, in a sample of 1094 adults, the topography of the global signal has been shown to capture a positive-negative axis of life outcomes and psychological functions linking to an increased component loading of the frontoparietal network (Li et al., 2019). Across the lifespan, the topographical patterns of the global signal have revealed spatially specific linear and quadratic associations with age ranging from 6 to 86 years (Nomi et al., 2022). Given my observations of a significant and positive effect of GSR on fingerprint accuracy, in combination with strong evidence that GSR reduces relationships between functional connectivity and head motion, particularly in age groups and vulnerable populations that exhibit heightened in-scanner movement, I recommend its inclusion in processing pipelines.

Describing the functional brain with high-resolution parcellation schemes produced the highest identification accuracies. This likely reflects the fact that increasing the parcel resolution preserves finer properties of the functional connectome from large-scale functional networks to localised brain areas, thus improving the sensitivity and specificity of functional connectivity analyses (Bellec et al., 2015). The choice of selecting a parcel resolution to investigate individual differences in the functional connectome typically depends on the desired spatial resolution of the effects being investigated. I found that cortical regions produced the highest fingerprint accuracy, which is consistent with observations that their functional connectivities are highly reliable compared to that of subcortical and cerebellar structures (Noble et al., 2019; Noble et al., 2017; Shah et al., 2016), which may be driven by factors such as poor SNR due to signal dropout and susceptibility to physiological noise (Noble et al., 2017). It is important to note that there may not be a unique parcellation scheme that captures individual differences in the functional connectome across all brain states (Salehi et al., 2020). Instead, there may exist a set of brain parcellations that is condition dependent which provides better parcel definition to detect such individual properties across different states (Salehi et al., 2020); investigating the impact of such state-specific parcellations on fingerprint ID would be an important future direction. Fingerprints restricted to the medial frontal and frontoparietal networks also yielded the highest ID accuracies compared to the whole-brain connectome. This is consistent with previous observations as elaborated in **Chapter 1 (Section 1.7.1)** that these networks are highly reproducible in both neurotypical and developing populations (Marek et al., 2019; Jalbrzikowski et al., 2019; Horien et al., 2019; Finn et al., 2015). The frontoparietal network exhibits higher inter-subject variability in its functional connectivity relative to other functional networks including subcortical, limbic, and visual networks (Mueller et al., 2013), suggesting that this functional network can better discriminate an individual from a group. Restricting the functional connectome fingerprint with the frontoparietal networks (Pipeline 2) would be suitable for some functional connectivity applications but not necessarily for all measures, pipelines, and procedures. Therefore, describing

the individual functional connectome with a second fingerprint-informed pipeline that includes a fine-grained parcellation scheme (Schaefer 1000-17Networks; Pipeline 2) offers another optimal strategy to test the reproducibility of functional connectivity applications. Importantly, the higher accuracy obtained for cortical regions and functional networks comprising cortical regions, relative to the subcortex and/or cerebellum may be largely explained by differences in the number of nodes in these different compartments and networks. For example, Venkatesh et al. (2020) showed that some functional networks (e.g., limbic) may achieve poorer fingerprint accuracies due to their small size (i.e., fewer brain regions) compared to others (e.g., frontoparietal networks) with larger sizes (i.e., more brain regions) — though this is not universally the case.

As others have shown (Vanderwal et al., 2021; Vanderwal et al., 2017; Airan et al., 2016; Finn et al., 2015), identification accuracies further increased non-linearly with longer scan duration. Scan duration has long been identified as a factor that affects the reliability of functional connectivity measures (Noble et al., 2019; Elliott et al., 2019; Noble et al., 2017; O'Connor et al., 2017; Mueller et al., 2015; Birn et al., 2013; Van Dijk et al., 2010; Shehzad et al., 2009; see **Chapter 1 (Section 1.6.3)**), along with resting-state conditions such as fixation, eyes-opened, and eyes-closed (Wang et al., 2017a; Zuo et al., 2015; Tagliazucchi and Laufs, 2014; Patriat et al., 2013). While initial reports suggested that acquisition times of ~5 min (typical for resting-state fMRI studies) are sufficient to yield fair-to-excellent reliability (Van Dijk et al., 2010; Shehzad et al., 2009), more recent studies have suggested that much longer duration scans are required to yield the most reliable (e.g., > 36 min; Noble et al., 2017) and reproducible (e.g., 15-25 min; Anderson et al., 2011) metrics. Researchers have employed several strategies to increase the amount of fMRI data available per participant, including longer scan duration (Noble et al., 2019; Noble et al., 2017; Tomasi et al., 2016; Shah et al., 2016), repeated scans within-/between-sessions (Cho et al., 2020; Elliott et al., 2019; O'Connor et al., 2017; Noble et al., 2017), and higher sampling rates (Birn et al., 2013). Acquiring and combining more scans for each individual across different brain states could further improve the fingerprint accuracy as within-subject variability may be better captured than at rest (Finn et al., 2017). This echoes recent evidence which suggests that having more fMRI data per participant increases the ability to capture unique properties of the functional connectome using resting-state and naturalistic conditions (Vanderwal et al., 2021). It is, however, challenging to identify an ideal amount of fMRI data that is needed per participant to optimise the reliability of functional connectivity — longer scan lengths undermine the practical utility of fMRI, especially in developing and clinical populations with lower in-scanner tolerance or wakefulness.

In agreement with previous work (Vanderwal et al., 2021; Vanderwal et al., 2017), identification patterns were improved when a greater number of discriminative edges was used to compute a

given individual's functional connectivity profile. This finding showed that as few as 10% of the most discriminative edges were needed to yield good fingerprint accuracies as there was no significant difference in identification accuracies for the top 10% or 75% of the most discriminative edges relative to the full whole-brain connectome. This suggests that computational complexity and resource demands could be used by downsizing the functional connectome to the top 10% of discriminative edges. This also suggests that there are some edges that do not contribute to the identifiability of an individual's functional connectome, thus positing that the entire functional connectome may not be required to compute functional connectome fingerprints.

I also pooled fMRI data across the five datasets to investigate the impact of sample size, but this did not improve the fingerprint accuracies. That is, larger sample sizes increase the difficulty of differentiating an individual from a group. Similar observations have been reported which showed that fingerprint accuracy decreases non-linearly when the sample size derived from the HCP increases from typical (e.g., $N = 100$) to large (e.g., $N = 1000$) (Li et al., 2021; Waller et al., 2017), which reflects that greater potential for “confusion” (erroneous high correlations/matches) due to the larger number of potential “lures.” Another potential limitation of this sampling strategy is that I aggregated the fMRI data across a wide range of age groups (i.e., 7-43 years) which may not be ideal to detect individual differences in the functional connectome. There could also be a site effect that requires further investigation, particular in a larger dataset acquired from several more sites. This may be a challenge as a downside of the connectome-based identification procedure is that a minimum of two imaging scans is required which is not always feasible during cross-sectional study designs. Future work may investigate the utility of split-half fingerprinting as a proxy method (King et al., 2023; Wang et al., 2021). Along the same lines, employing a larger multisite dataset could allow the investigation of inter-scan intervals on fingerprint accuracy. The CoRR datasets used in this study make the comparison very difficult due to the wide range of inter-scan intervals from days to years.

Although I aimed to identify a pipeline that would optimise functional connectome fingerprint ID, and while high identification accuracies were attained for some samples, the accuracy did not reach 100% for any sample. This demonstrates that there are some individuals who remain incorrectly identified even with the fingerprint-informed pipelines. This suggests that there is scope to identify further factors that might influence fingerprint identification accuracy. On the other hand, it may also suggest that there are individuals who remain misidentified, even in the context of overall high ID accuracy. I attempted to quantify the degree of misidentification for a given individual from a group using the RR metric which was as low as 2% when the optimal

pipelines were employed. The RR metric did not provide additional discriminative information beyond BID, however.

Can we expect resting-state functional connectome fingerprint ID accuracy to be perfect? There is evidence for 100% identification accuracy, achieved using naturalistic paradigms (Vanderwal et al., 2017). The answer may lie in the cognitive states of participants being scanned (Finn and Rosenberg, 2021; Finn et al., 2017); scanning conditions that evoke more specific cognitive states tend to be associated with higher fingerprint accuracy relative to rest (Vanderwal et al., 2017; Finn et al., 2017; see **Chapter 1 (Section 1.7.2)**). It is possible that connectome-based identification accuracy at rest may reflect the influence of arousal, thought, and even physiological processes (Finn et al., 2017). Another important consideration is the influence of anatomical variability on individual differences in functional connectivity estimates that have been characterised as a mixture of morphological and functional variability (Langs et al., 2015). Converging evidence from human and primate studies have shown that there is a significant relationship between structural and functional connectivity such that structural connectivity may predict *inter*-individual variability in cognition (Sarwar et al., 2020; Sporns, 2013). This relationship may, however, be heterogeneous as the structure-function coupling could be stronger for some networks but weaker for others (Gu et al., 2021; Suárez et al., 2020; Baum et al., 2019). Future work could investigate the contributions of anatomical variability (e.g., morphology) in measuring individual differences in functional connectome identifiability. Yet another avenue could focus on investigating whether unique or similar structure-function coupling across individuals relates to better or worse connectome-based identification.

2.5 Conclusions

In this study, I showed that functional connectome fingerprint accuracy can be used as a simple, intuitive, and sensitive metric for examining the effects of different functional connectivity analysis parameters. By examining a set of data acquisition, pre-, and post-processing parameters, I identified a number of recommended steps that form two pipelines optimised for reproducible and stable functional connectivity at the individual level (see **Appendix C**). The pipelines comprise the following strategies: (I) applying GSR as a denoising strategy; (II) fine-grained brain parcellations; (III) cortical regions relative to subcortical and cerebellar structures; (IV) predictive functional networks such as medial frontal and frontoparietal networks; (V) greater number of discriminative edges; and (VI) increasing scan duration. In contrast, increasing the sample size by pooling resting-state fMRI data across different populations do not necessarily yield higher fingerprint accuracies. I suggest that the inclusion of fingerprint accuracy in analysis workflows can help researchers identify the most reproducible pipelines and as a result, minimise

the “researcher degrees of freedom” to achieve reproducible brain-behaviour relationships for biomarker discovery.

So far, **Chapter 2** has employed functional connectome fingerprinting in five datasets that comprised low-motion individuals spanning childhood and adulthood. Retaining only participants exhibiting the lowest levels of head motion is common practice in functional connectivity studies, given that head motion has been identified as a major confound, particularly in developmental populations. In **Chapter 3**, I will describe the efforts of the neuroimaging community to improve both sample sizes and data quality, as routes to increase reproducibility. In **Chapter 4**, I investigate whether the “high-motion” participants who are usually discarded in brain-behaviour association studies can be salvaged using a novel sampling approach. To test the utility of this sampling approach, I will use the same functional connectome fingerprinting procedure described in this Chapter.

Chapter 3

Salvaging high-motion participants to maximise sample size for inclusive and reproducible brain-behaviour relationships

Summary

Chapter 3 discusses the “big data” era in neuroimaging — from data-sharing and the Open Science movement to the establishment of large multi-site collaborative consortia. Reflecting on the active scholarly debate of sample sizes within functional neuroimaging, I elaborate on the efforts of researchers to identify reproducible brain-behaviour associations with both typical sample sized and large consortia-level datasets. Given that developing youths exhibit heightened head motion during scanning, I further discuss the impact of head motion in estimating functional connectivity in addition to describing several commonly applied head motion corrections strategies. As “high-movement” individuals tend to be excluded in typical brain-behaviour association studies, I examine whether some useable, low-motion data these individuals have, can be salvaged — to address the trade-off between the requirement that data are high quality and free from motion artefacts, and the large sample sizes required for reproducible brain-behaviour associations.

3.1 The Big Data Revolution

Over the past two decades, resting-state functional connectivity mapping has gained mainstream acceptance as a powerful non-invasive technique for the investigation of macroscopic human brain organisation, and for the elucidation of associations between variation in this functional architecture and variation in behavioural phenotypes (Shen et al., 2017; Finn et al., 2015; Biswal et al., 1995). While early brain-behaviour studies generated considerable excitement about functional connectivity's potential to reveal biomarkers for a range of neurodevelopmental, psychiatric, and neurological conditions (Castellanos et al., 2013), the robustness and reproducibility of many findings have recently been called into question (Marek et al., 2022; Szucs and Ioannidis, 2020; Cremers et al., 2017; Button et al., 2013; Thirion et al., 2007; see **Chapter 1 (Sections 1.4-1.5)**). Although accepted practice for the field at the time, subsequent analyses have found that overly liberal statistical approaches to thresholding, combined with poor statistical power, may have increased the likelihood of false-positive outcomes (i.e., spurious findings) (Button et al., 2016; Button et al., 2013). Even when a low-powered study reveals a “true” effect size, there is a high likelihood that the magnitude of that effect is inflated (Marek et al., 2022; Anderson et al., 2017; Button, 2016; Simmons et al., 2011; Ioannidis, 2005). Recognising this reality, the resting-state neuroimaging community has shifted towards an emphasis on data sharing and both retrospective and prospective pooling of neuroimaging data from multiple sites, creating consortia-level datasets (Casey et al., 2018; Alexander et al., 2017; Miller et al., 2016; Zuo et al., 2014a; Thompson et al., 2014; Van Essen et al., 2013; Di Martino et al., 2013; Milham et al., 2012; Petersen et al., 2010; Biswal et al., 2010). The ease of pooling resting-state data has enabled the creation of datasets containing thousands and even tens of thousands of participants — a feat that is impossible for any individual laboratory, due to resource constraints. These collaborative efforts have established several consortia-level resting-state datasets including the ABCD Study ($N \approx 12,000$), ABIDE ($N \approx 1,112$), ADHD-200 ($N \approx 947$), ADNI ($N \approx 1,821$), CoRR ($N \approx 1,639$, 33 datasets), HCP ($N \approx 1,200$), HBN ($N \approx 10,000$), Enhanced NKI-RS ($N \approx 1,589$), OASIS ($N \approx 3,059$), and UKBiobank ($N \approx 60,000$) among others (see Madan, 2021 for a review on large-scale neuroimaging datasets). Many of these datasets are openly available (under Open Science CC-BY licences; e.g., ABIDE, ADHD-200, CoRR) and can be used by anyone, anywhere, while others are available under Data Use agreements (e.g., ABCD Study, UKBiobank). All are widely used by the neuroimaging community.

Open Science data sharing has provided an unprecedented opportunity to accelerate scientific discovery, improve reproducible practices, and to “democratise” neuroimaging by dramatically reducing the resources required to do research (Markiewicz et al., 2021; Milham et al., 2018; Poldrack and Gorgolewski, 2014; Poline et al., 2012). Given the increasing cost of acquiring an

MRI scan (e.g., US\$500-US\$1000/HR), combined with the fact that many researchers do not have access to an MRI, particularly in the Global South, the Open Science big data movement has provided researchers with unprecedented opportunities to access large-scale datasets to ask new questions and devise new hypotheses. For example, Milham et al. (2018) showed that if we take the sample sizes of 913 studies published using the ABIDE dataset, the cost savings associated with data collection, phenotyping, and imaging would range from US\$893 million to US\$1.7 billion. The proliferation of Open-Source processing pipelines or frameworks such as fMRIPrep and C-PAC, and the wide adoption of Open Science practices, such as code sharing (e.g., GitHub), has further boosted this democratising movement.

The advent of consortia-level datasets, particularly those comprising retrospective data collected without pre-agreed and standardised imaging protocols, generated a need to standardise neuroimaging data. Two outcomes are community adoption of the FAIR Principles (Wilkinson et al., 2016) and BIDS format (Gorgolewski et al., 2016). The FAIR Principles describe the characteristics of data and digital objects to support sharing, exploration, reuse, and deposition by third parties (Wilkinson et al., 2016) — namely, that they should be findable, accessible, interoperable, and reusable (<https://www.go-fair.org/fair-principles/>). The BIDS format is a standard for the organisation of neuroimaging datasets (e.g., file naming conventions, direct structure, provenance records) that follows FAIR principles and facilitates both data reuse and automated processing by Open-Source data analysis pipelines (Gorgolewski et al., 2016). For example, OpenNeuro (<https://openneuro.org>) shares hundreds of BIDS compliant resting-state, task-based, and naturalistic fMRI datasets, collected from more than 21,900 participants. Increasingly, Open-Source pipelines such as fMRIPrep and C-PAC expect BIDS format data, further reinforcing its adoption across the community.

There has also been a collective effort to deposit statistical maps, brain parcellations, and atlases from published studies so that they are accessible and available for reuse by the wider scientific community. One such platform is Neurovault (<https://neurovault.org>), which is a centralised repository for unthresholded statistical maps that researchers can reuse in their own analyses (Gorgolewski et al., 2015). For example, if one wishes to use a whole-brain parcellation as a mask, they can find the Shen 268 (<https://identifiers.org/neurovault.image:395091>) or multi-resolution MIST 7-444 (<https://identifiers.org/neurovault.collection:6227>) parcellation schemes available on Neurovault. Another effort is NeuroSynth (<https://neurosynth.org>; Yarkoni et al., 2011), which is a meta-analytic database that synthesises large-scale fMRI studies to provide term-based spatial maps. For example, if one is interested in the functional networks that play a role in emotion regulation, they can search the term “emotion regulation” in NeuroSynth, which yields an automated meta-analytic map of 247 studies, from which they can identify

coordinates of peak meta-analytic activation probability: <https://neurosynth.org/analyses/terms/emotion%20regulation/>. Such maps can be important starting points for analyses — for example, by providing a basis for ROI definition.

Despite the strengths of Open Science and consortia-level data, we should note that not all research questions can be answered by such datasets, even though we have tens of thousands of participants at our disposal. For example, researchers interested in the neural bases of disease-specific processes or rare clinical conditions will not find those well represented in openly available datasets. Even for neurotypical participants, or for conditions that are relatively well represented in consortia-level datasets (e.g., ASD, ADHD), factors such as site-to-site variation in imaging protocols, lack of harmonised phenotypic and clinical assessments, and condition heterogeneity (e.g., severity, comorbidity, medication status, treatment response) create challenges that may not be sufficiently addressed by large sample sizes. Finally, although efforts such as the ABCD Study and UKBiobank are leading the way in longitudinal data collection with deep phenotyping, the majority of available studies to date use cross-sectional paradigms with limited or uncoordinated phenotyping, inhibiting the precision with which we can measure developmental processes and brain-behaviour relationships.

As discussed in **Chapter 1**, for functional connectivity metrics to serve as biomarkers for clinical conditions means that methods for the detection of brain-behaviour relationships in a given cohort must be reproducible. Researchers have long noted that a central concern in this endeavour is sample size — we must have adequate sample sizes at our disposal to achieve enough statistical power for estimating the “true” effect sizes of brain-behaviour associations (Yarkoni, 2009; Kelly et al., 2012). Power reflects the rate of true positives and is inversely related to the rate of false negatives, such that the lower the statistical power of a study, the lower the likelihood of detecting true effects above chance. Large-scale studies typically have higher statistical power, so they can help detect smaller effects, whereas studies with small sample sizes can detect only large effects, which are often subject to inflation (Button, 2016). Therefore, smaller sample sizes in fMRI studies tend to reduce the likelihood of finding a true brain-behaviour association (if one truly exists) and increase the likelihood of observing a false-positive association, i.e., inflated brain-behaviour relationships. This phenomenon then leads to an inflation of positive effect sizes that are typically observed in the fMRI literature (Poldrack et al., 2017; Yarkoni, 2009). It is important to remember that underpowered studies are not only prevalent in neuroimaging but are also present beyond the realm of neuroscience (Szucs and Ioannidis, 2017; Button et al., 2013; Ioannidis, 2005). Nonetheless, this motivates the need for the neuroimaging community to work collaboratively to establish sufficiently powered datasets for functional connectivity-based biomarker discovery.

3.2 Considerations of Sample Size and Amount of Data for Reproducible Brain-Behaviour Relationships

The importance of sample size was recently underscored by Marek et al. (2022), who showed that very large sample sizes ($N \gtrapprox 2,000$) are required to achieve reproducible univariate and multivariate estimates of brain-wide phenotype associations (BWAS) — statistical associations between structural and functional brain markers, such as functional connectivity, and phenotypes such as cognitive ability and general psychopathology. Using the large, longitudinal, richly phenotyped ABCD Study ($N = 11,572$; 9-10 years), the authors revealed the largest 1% univariate BWAS across all possible associations are weaker ($0.06 < |r| < 0.16$) than those typically reported in low-powered BWAS-based studies and are not reproducible at typical sample sizes ($N = 25$ -100) (Figure 3.1). Subsequently, they showed that the small BWAS effect sizes derived from resting-state functional connectivity were consistent in two other consortia-level datasets (i.e., $N = 900$; ABCD Study: $|r| > 0.11$; HCP: $|r| > 0.12$, UKBiobank: $|r| > 0.09$). The authors further demonstrated that typical sample sizes inflate estimates of and effect sizes of brain-behaviour associations, due to sampling variability. Sampling variability, i.e., random variation in a brain-behaviour association across subsamples of the population, can generate stronger associations with typical samples sizes by chance. Marek et al. (2022) demonstrated that the 99% confidence interval (CI) of the univariate BWAS for $N = 25$ was $r \pm 0.52$ which was drastically reduced at $r \pm 0.07$ for much larger sample sizes at $N = 1,964$. The statistical power ($1 - \beta$, where β is the Type II error, i.e., false negative rate) remains low even for considerably large sample sizes ($N \approx 1000$). This is because Type II errors and inflation rate of the BWAS are still high for very large N . Moreover, Marek et al. (2022) also showed that the statistical thresholding of BWAS (i.e., correction for multiple comparisons) can limit the ability to replicate univariate correlation effects; conservative thresholding in low-powered studies is likely to capture largely inflated correlations unless N grows to very large ($N = 9,500$) to achieve 80% power for the detection of the strongest brain-wide associations. On the other hand, multivariate BWAS has been shown to achieve larger effects relative to univariate BWAS with large sample sizes ($N \gtrapprox 2,000$) although there is still an inflation rate that persists (Marek et al., 2022). An important feature of this study is that the authors ensured the univariate and multivariate brain-wide associations were not affected by head motion by applying a strict motion correction strategy and threshold to retain only the “low-movement” participants, resulting in the exclusion of a substantially large number of participants — an issue which I will return to in **Section 3.3**.

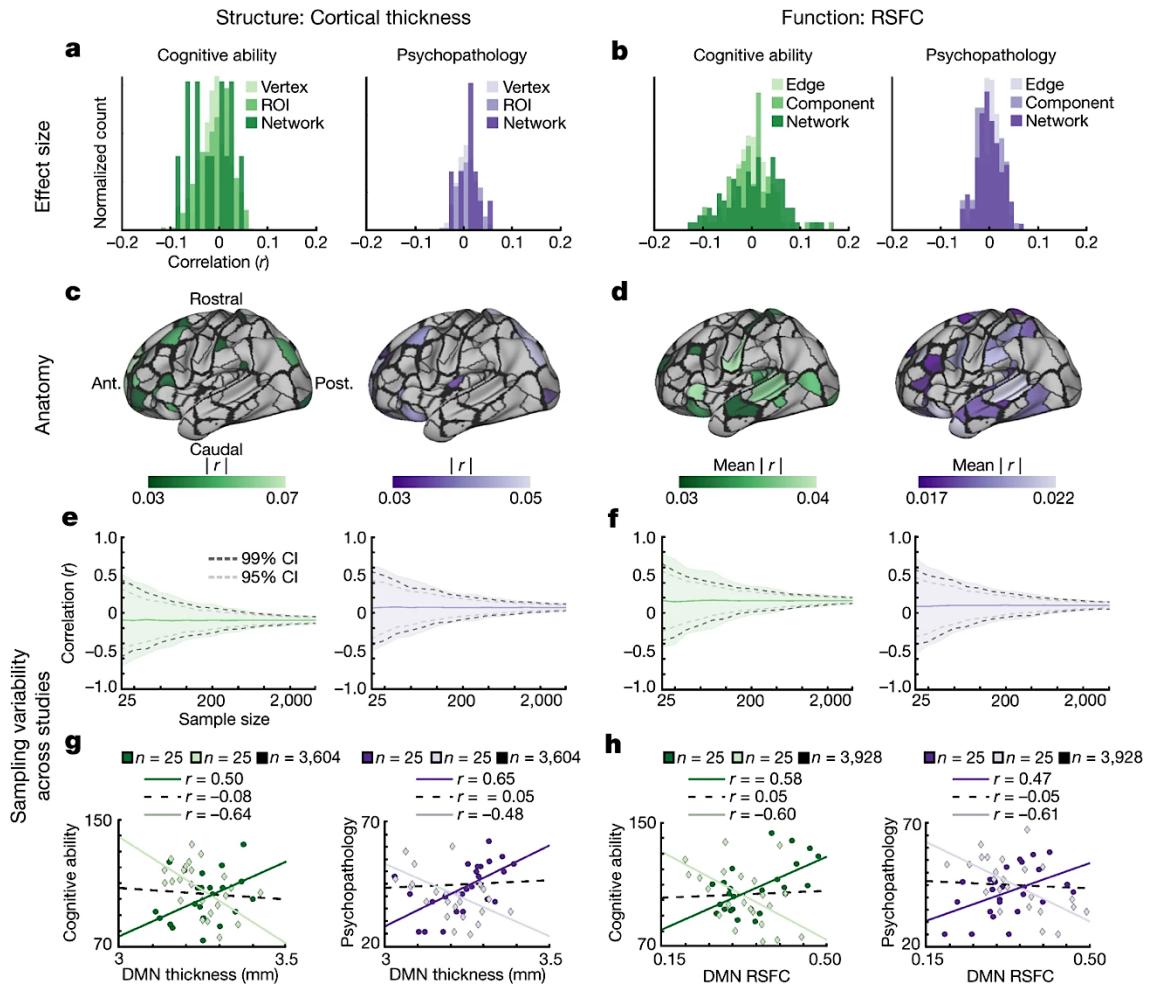


Figure 3.1. From Marek et al. (2022): Effect sizes and sampling variability of univariate brain-wide associations (BWAS) from the ABCD Study. (a) Effect sizes of brain-wide associations derived from cortical thickness and cognitive ability as well as general psychopathology at the vertex, ROI, and network levels as a function of sample size. (b) Effect sizes of brain-wide associations derived from resting-state functional connectivity and cognitive ability as well as general psychopathology at the vertex, ROI, and network levels as a function of sample size. (c) Largest brain-wide associations of cortical thickness with cognitive ability and general psychopathology spatially displayed at the ROI-level. (d) Largest brain-wide associations of resting-state functional connectivity with cognitive ability and general psychopathology spatially displayed at the ROI-level. (e) Sampling variability resamples per sample size from $N = 25$ to $N = 3,604$ for the largest brain-wide associations of cortical thickness with cognitive ability and general psychopathology. (f) Sampling variability resamples per sample size from $N = 25$ to $N = 3,928$ for the largest brain-wide associations of resting-state functional connectivity with cognitive ability and general psychopathology. (g) Examples of two $N = 25$ subsamples, where inaccurate default mode network (DMN) associations were observed for cortical thickness with cognitive ability and general psychopathology. (h) Examples of two $N = 25$ subsamples, where inaccurate default mode network (DMN) associations were observed for resting-state functional connectivity with cognitive ability and general psychopathology.

By suggesting that thousands of participants are needed to identify robust and reproducible BWAS, Marek et al. (2022)'s paper prompted a scholarly debate (Nature Neuroscience, 2022). If true, this would put the ability to study brain-behaviour relationships out of the hands of all but a small number of researchers with access to large enough datasets and sufficient computational

resources. In response, Rosenberg and Finn (2022) suggest there are other routes toward robust brain-behaviour relationships that do not require participants in the thousands, including the use of appropriate study designs (e.g., rest vs. task vs. movies), refined methodological approaches (e.g., univariate versus multivariate), and Open Science practices (e.g., pre-registration, model sharing and validation, replication). For example, certain brain states such as those induced by task performance or by watching a movie are more predictive of an individual's behaviour compared to resting-state (i.e., unconstrained) conditions (Greene et al., 2023; Finn and Bandettini, 2021b; Gao et al., 2019; Greene et al., 2018a). Rosenberg and Finn (2022) cite research showing that multivariate methods can also improve the robustness of brain-behaviour associations compared to univariate techniques, because individual functional connectivity features tend to have lower effect sizes relative to (weighted) brain patterns in the functional connectome (e.g., Marek et al., 2022). In terms of Open Science practices, pre-registering a study helps to reduce the “researcher degrees of freedom” by defining the hypotheses and analysis framework prior to conducting the research. Similarly, researchers should attend carefully to pre- and post-processing factors that could influence the statistical power of brain-wide associations, such as spatial normalisation, brain parcellations, and automated processing pipelines (DeYoung et al., 2022; Bandettini et al., 2022). As described in **Chapter 1 (Section 1.5)**, the number of options available and the lack of consensus across researchers can lead to thousands of possible combinations of pre-/post-processing data analysis steps. For example, Carp et al. (2012) compared the output of 6,912 task-based fMRI analysis pipelines and found a lack of correspondence in the activation maps related to strength, localisation, and thresholding. Poldrack et al. (2017) further identified that the choice of data pre-/post-processing steps available to researchers creates a total of 69,120 possible workflows, potentially introducing significant variability into estimates, including estimates of functional connectivity. If we take the example of brain parcellations, use of different parcellation schemes can significantly impact the strength of brain-behaviour associations, such as relationships between functional connectivity and age or cognitive ability (Bryce et al., 2021).

Other researchers have also contributed to the discussion stimulated by Marek et al. (2022). Another route toward increased reproducibility in functional connectivity studies, and fMRI more generally, is to carefully control for physiological processes, including arousal states and breathing (Bandettini et al., 2022). Yet another avenue is to focus on smaller samples but using studies designed to maximise SNR and effect sizes for brain-behaviour associations (Gratton et al., 2022). For example, within-subject designs can be strengthened using experimental manipulations that are known to be sensitive for a chosen behavioural measure, while between-subject designs can be applied to maximise the variations between brain functions and behavioural measures, e.g., in developing or clinical samples that exhibit significant changes in

brain function. Moreover, reducing the variability in brain functions and behavioural measures that relate to “noise” in the fMRI signal could also serve to increase the effect sizes of brain-behaviour relationships with typical sample sizes. Such studies could also employ “precision mapping” data — individual-specific brain network organisation derived from highly sampled fMRI data (Gordon et al., 2017; Poldrack et al., 2015) or individualised brain parcellations (Kong et al., 2021; Salehi et al., 2020; discussed in **Chapter 2 (Section 2.4)**) — to boost signal by minimising the “noise” associated with morphological and functional variability in functional connectivity estimates (Sarwar et al., 2020; Lang et al., 2015; Sporns, 2013). One conclusion we can draw from the discussion sparked by the Marek et al. (2022) paper is that sample size is not the only route to robust and reproducible brain-behaviour associations: methodological choices aimed at improving the reproducibility of functional connectivity estimates themselves may also achieve this end. In recent years, one factor that negatively influences functional connectivity estimates has received considerable attention: head motion.

3.3 Impact of Head Motion on Functional Connectivity in Youths and Adults

Head motion has been a long-standing concern in fMRI research as it can significantly inflate functional connectivity estimates across a diverse range of populations (Frew et al., 2022; Power et al., 2014; Yan et al., 2013a; Yan et al., 2013b; Van Dijk et al., 2012; Power et al., 2012; Kelly et al., 2012). This is particularly true in youth samples due to the presence of heightened in-scanner motion (Satterthwaite et al., 2012); head motion typically decreases with increasing age (Frew et al., 2022; Vanderwal et al., 2019; Greene et al., 2018b; Satterthwaite et al., 2012). The canonical trajectory of developmental change in functional connectivity from childhood to adulthood is increased distant connectivity (e.g., integration) and reduced local connectivity (i.e., segregation) with increasing age (Dosenbach et al., 2010; Fair et al., 2009; Fair et al., 2007). Unfortunately, head motion exhibits the inverse pattern: more severe head motion is associated with weaker long-distance functional connectivity and increased local connectivity (Zeng et al., 2014; Yan et al., 2013a; Yan et al., 2013b; Van Dijk et al., 2012) — thus, the typical pattern of developmental change in functional connectivity may simply be the spurious product of age-related decreases in the severity of head motion (Power et al., 2012). Head motion is therefore a significant confound that must be both minimised and, because it can never be fully removed, controlled for in analyses of functional connectivity data.

In addition to directly influencing functional connectivity estimates, head motion can constitute a phenotypic index in itself. For example, head motion is associated with poorer inhibition and cognitive flexibility in older adults (Hausman et al., 2022), and with impulsivity in youths and adults, further confounding studies of clinical cohorts, such as ADHD (Kong et al., 2014). Head

motion exhibits significant relationships with behavioural phenotypes across a broad range of domains such as cognition, emotion, executive function, and personality (Siegel et al., 2016), extending to demographic (e.g., education) and physiological (e.g., blood pressure, weight, BMI) measures (Siegel et al., 2016). The presence of in-scanner motion may also reflect neurobiological correlates of functional brain organisation mainly in the default mode network as evidenced by diminished distant functional connectivity within the network (Zheng et al., 2014).

One approach to minimising biases associated with head motion and potentially spurious relationships with behavioural phenotypes is to apply motion correction strategies during pre- and post-processing. All such strategies begin with the quantification of motion. One of the most commonly used metrics to quantify in-scanner motion is framewise displacement (FD; Power et al., 2012; Jenkinson et al., 2002). During the functional preprocessing stage, each fMRI volume is rigidly realigned to a base volume, which produces a set of realignment parameters (RPs), indexing how much a given volume must be rotated (along the x, y, and z axis) or translated (pitch, roll, yaw) to align with the base volume. FD is computed as the sum of the absolute values of the differentiated RP estimates for every volume (i.e., the difference between the current and the preceding volume) and a participant's overall head motion is usually computed as mean FD (Power et al., 2012) or root-mean-square [rms] FD (Jenkinson et al., 2002).

While multiple strategies are available to mitigate the influence of head motion during preprocessing (e.g., GSR, temporal censoring, ICA, tissue-specific signal regression, principal component analysis [PCA]), none succeed in removing the influence of motion completely (Parkes et al., 2018; Ciric et al., 2017; Yan et al., 2013a; Yan et al., 2013b). An important first step in any analysis to reduce the impact of motion, therefore, is to exclude participants that have a mean FD or rmsFD above a predefined threshold. However, there is no clear consensus on the ideal head motion threshold that one should apply. Given the concerns discussed above, about motion biases on functional connectivity measures, particularly in the context of development or group differences in motion severity, some studies have applied very strict motion thresholds (e.g., mean FD/rmsFD < 0.10 mm) to retain only those participants with minimal movement (Marek et al., 2022; Byrge and Kennedy, 2018; Zheng et al., 2014). Although this is an understandable approach, a conservative head motion threshold can result in having to exclude large numbers of participants from a sample. This trade-off is exemplified in the landmark paper of Marek et al. (2022), which employed a stringent FD threshold (filtered FD < 0.08 mm) that resulted in only ~34% of the total sample size ($N = 3,928/11,572$) being retained for analyses. A requirement to discard nearly two-thirds of a sample could certainly be deemed a considerable waste of resources, and potentially unethical, from the perspective of the participants who dedicated considerable time and effort to participating in the study, as well as the scientists who

dedicated their time and effort to collecting the data. Moreover, in the context of developmental and clinical cohorts, the participants who tend to be excluded for motion are usually those who are the youngest (Vanderwal et al., 2015) or who exhibit more severe symptoms/difficulties (e.g., ADHD, multiple sclerosis) (Wylie et al., 2014; Durston et al., 2003). Salvaging their data would maximise inclusivity, enabling a wider spectrum of behaviour to be investigated. This could be particularly beneficial for the investigation of brain-behaviour relationships, which might otherwise suffer from range restriction (i.e., due to being limited only to the less severe or unimpaired range of the behavioural spectrum). While motion corrupts the timepoints during which it occurs as well as their surrounding timepoints (Power et al., 2015; Power et al., 2014; Power et al., 2013), even high-motion participants may still have some useable (low-motion) timepoints within a given scan. Rather than discarding all participants who exceed a mean FD or rmsFD threshold, could some of their data be salvaged? Chapter 4 aims to answer this question.

Chapter 4

Towards data salvage in high-movement cohorts: bagging yields robust and reproducible brain-behaviour relationships

Paper (In Preparation)

Ramduny, J., Vanderwal, T., & Kelly, C. Towards data salvage in high-movement cohorts: bagging yields robust and reproducible brain-behaviour relationships.

Authors

Jivesh Ramduny^{1,3}, Tamara Vanderwal^{4,5}, Clare Kelly^{1,2,3}

¹School of Psychology, Trinity College Dublin, Dublin, Ireland

²Department of Psychiatry, School of Medicine, Trinity College Dublin, Dublin, Ireland

³Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland

⁴Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada

⁵BC Children's Hospital Research Institute, Vancouver, BC, Canada

Authorship Contribution Statement

Jivesh Ramduny: Conceptualisation, Methodology, Formal analysis, Investigation, Writing — original draft, review & editing. **Tamara Vanderwal:** Conceptualisation, Investigation, Writing — original draft, review & editing. **Clare Kelly:** Conceptualisation, Funding, Methodology, Formal analysis, Investigation, Writing — original draft, review & editing, Supervision, Project administration.

Abstract

The conflicting demands of achieving larger sample sizes and good data quality have pushed the neuroimaging community towards acquiring thousands of participants for adequate statistical power and at the same time, have compelled researchers to exclude large number of participants to minimise the presence of motion artefacts in typical brain-behaviour association studies as these individuals are often regarded as “high-motion”. In the present study, I proposed a resampling strategy referred to as “bagging” or bootstrap aggregation at the level of timeseries to identify an individual’s functional connectome from resting-state fMRI data obtained from CoRR and HBN. First, I showed that functional connectomes computed on the basis of bagged low-motion data yield good ID accuracy. Second, I demonstrated that bagging produces robust brain-behaviour associations as a function of sample size with as few as 60 “clean” timepoints. These associations were comparable with a standard approach employing the full fMRI timeseries. Lastly, I showed that the utility of bagging lies in maximising participant inclusivity by salvaging high-motion participants that would otherwise be discarded. While the findings from bagging were reproducible across different head motion thresholds and edges in the functional connectome, I recommend the examination of brain-behaviour relationships with other behavioural phenotypes and consortia datasets that involve youth populations (e.g., ABIDE, ABCD Study) to assess the generalisability of bagging for typical brain-behaviour association studies.

4.1 Introduction

As discussed in **Chapter 3**, researchers who examine brain-behaviour relationships need to meet sample size demands, to obtain adequate statistical power for reproducible associations, as well as data quality demands, to mitigate the confounding influence of motion-related and other artefacts. While the sample size requirement is often satisfied by the use of large consortia datasets, good data quality is usually achieved by retaining only low-motion data, i.e., discarding participants who move excessively. In some cohorts, the latter can lead to large numbers of participants being excluded.

When discarding participants on the assumption that their data are “unusable,” we reduce the inclusivity of our dataset and limit the diversity of the samples we have at our disposal for brain-behaviour analyses. Participants who are excluded in such analyses are those who are the youngest and/or experience severe clinical conditions (Vanderwal et al., 2015; Wylie et al., 2014; Durston et al., 2003; see **Chapter 3 (Section 3.3)**). This limits our study of a particular symptom or behavioural dimension, since such exclusions likely truncate its range. In the context of brain-behaviour analyses, this impedes reproducibility — imposing a range restriction violates the assumptions of correlation by reducing the representativeness of the sample, relative to the population (Goodwin and Leech, 2006). Restricting the range of symptoms and/or behaviours also reduces the effect sizes of brain-behaviour relationships by decreasing the variability of the data (Gratton et al., 2022).

Instead of discarding participants whose motion exceeds a given threshold, here, I propose a means to salvage these individuals’ data so that they can be retained in analyses, thus boosting sample size and diversity for the examination of brain-behaviour associations. Specifically, I implement “bagging”, also known as bootstrap aggregation, at the timeseries level (Opitz and Maclin, 1999; Breiman, 1996); in terms of its implementation, bagging means resampling *with* replacement to create multiple random subsets of the training data. These subsamples can then be used to generate several instances of a machine learning model, and their predictions are aggregated to yield more robust performance estimates (see Appendix D for more details about the implementation of bagging). The main advantage of this bootstrap sampling technique is that it can minimise variability and bias in a noisy high-dimensional dataset (Varoquaux et al., 2017; Dudoit and Fridlyand, 2003). That is, creating multiple subsamples of the training data and aggregating their performances reduce model underfitting (i.e., low variance and high bias) and/or overfitting (i.e., high variance and low bias) in addition to improving model generalisability in unseen datasets (Varoquaux et al., 2017; Dudoit and Fridlyand, 2003). Bagging can be applied at the level of samples by computing bootstrap subsets of the training data or at the level of

timeseries themselves by sampling random timepoints — both across a number of iterations (e.g., ≥ 100).

An approach similar to bagging, jack-knife resampling, was first applied to fMRI data in a proof-of-concept study to estimate CIs for correlation coefficients indexing task-evoked activation during a finger-tapping paradigm (Biswal et al., 2001). Specifically, 85 timepoints were repeatedly resampled (*without* replacement) from a timeseries of 90 timepoints acquired during 4.5 cycles of 20 seconds alternating finger-tapping/rest. For each of the 1000 jack-knife samples, the box-car reference waveform was also resampled in a matched fashion and the correlation between the reordered fMRI timeseries and waveform was computed. This process produces an approximation of the theoretical population from the observed data (sample), enabling the estimation of reliability and CIs for the parameter of interest (here, correlation with the reference waveform). Biswal et al. (2001) then generated maps of task-evoked activation thresholded on the basis of the estimated distributions, rather than a simple fixed correlation coefficient threshold. Biswal et al. (2001) argued that this approach produced more robust, sensitive activation maps because it better accounts for variation in both signal (response) and noise across voxel. They also suggested that this approach reduces the chance of false positives, thus increasing reproducibility.

Bagging can also be applied at the level of participants to create bootstrap resampled groups. For example, bagging was used by Richiardi et al. (2011) to create bootstrapped sets of training data (participant-level functional connectivity matrices), which were used to train classifiers to decode brain states (e.g., resting-state, movie-watching) from functional connectivity patterns. Averaging predictions across bootstrapped samples improved classification accuracy. Participant-level resampling of training data and aggregation of model parameters and features also improve the accuracy and generalisability of brain-behaviour models for a range of phenotypic measures (e.g., fluid intelligence, sustained attention, cognitive flexibility, processing speed) using connectome-based modelling as well as other prediction frameworks (O'Connor et al., 2021; Wei et al., 2020).

Combining both approaches, Bellec et al. (2010) and Nikolaidis et al. (2020) used bagging *within* individual participants' timeseries and *across* participants to improve the reproducibility of functional brain parcellations derived from relatively short acquisition scans (e.g., 6 min). Specifically, Nikolaidis et al. (2020) applied bagging to create multiple bootstrap resampled timeseries (using circular block resampling) for each participant. Clustering was applied to these resampled timeseries and cluster solutions were transformed to adjacency matrices, which were then averaged for each participant. Bagging was then applied again at the participant level — individual adjacency matrices were averaged within many resampled groups, and clustering was

applied to these group-mean adjacency matrices to obtain a final cluster solution. This approach increased the reliability (assessed using the ICC), and reproducibility of functional connectivity-based parcellations across a number of acquisition (e.g., scan lengths, scanner, imaging sites) and processing (e.g., clustering parameters and algorithms) parameters. Moreover, reproducibility for the bagging approach applied to 6-min scans was greater than that for the standard parcellation approach applied to scans that were double the duration. This suggests that bagging may compensate for the effects of short scan durations (see **Chapters 1 and 2** for a discussion of the association between scan duration and reproducibility).

In the present study, my focus is on bagging at the level of individual timeseries; this approach is conceptually similar to that used by Biswal et al. (2001), Bellec et al. (2010), and Nikolaidis et al. (2020). A key difference is that, rather than randomly resampling timepoints from an individual timeseries, I first sorted each individual's timepoints according to motion (rmsFD; see **Chapter 3 (Section 3.3)**). I then retained those timepoints with the least motion (requiring all timepoints to exhibit motion below a given threshold) and performed bootstrap resampling (*with replacement*) within this subset to create multiple bootstrap resampled timeseries for each participant. I computed functional connectivity for each sample and averaged across samples to derive an average functional connectivity matrix for each participant. In **Chapter 2**, I showed that fingerprint ID accuracy can be used to assess the impact of methodological factors on the reproducibility of functional connectivity estimates. Here, I followed the same logic to assess the reproducibility of this bootstrap resampling strategy by examining the effect of bagging on fingerprint ID accuracy. I complemented this test of reproducibility with an examination of validity — following Marek et al. (2022), I also examined the impact of bagging on estimates and CIs for brain-behaviour relationships as a function of sample size. For both analyses, the results obtained using bagged functional connectivity matrices were compared to those obtained using standard, full timeseries. I hypothesised that bagging would yield reproducible functional connectivity matrices, and robust brain-behaviour relationships. I suggest that the bagging method offers a promising avenue to rescue data from high-motion participants who would otherwise be excluded, thus enabling researchers to maximise sample size, and compute more inclusive and reproducible brain-behaviour relationships.

4.2 Materials & Methods

4.2.1 Participants

Consortium for Reproducibility and Replicability [CoRR]: From the five CoRR datasets that were employed in **Chapter 2**, the UPSM ($N = 92$) dataset was used in the present study. Each participant had at least one anatomical and two-resting fMRI scans (see **Chapter 2 (Section**

2.2.1) for the demographic information and acquisition parameters of the dataset). The purpose of including this CoRR dataset was to assess the impact of bagging on connectome-based fingerprint identification, following the same logic in **Chapter 2**.

Healthy Brain Network [HBN]: Data were obtained from the HBN (http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/), which is a community-based help-seeking sample of children and adolescents from the Greater New York area (Alexander et al., 2017). Neuroimaging and phenotypic data were obtained from the Cornell Brain Imaging Center (CBIC) site in Releases 4-9 comprising 682 participants. Approval was obtained from the Chesapeake Institutional Review Board; written consent was provided by participants who were at least 18 years old, while written consent was obtained from legal guardians and from the participants who were under 18 years. For each participant, at least one anatomical and two resting-state fMRI scans were obtained from the same imaging session (i.e., within-session scans). Quality control was performed on all the scans to check for potential scan artefacts in addition to ensuring that each participant had two runs of resting-state fMRI data as well as their phenotypic data to reflect their age and sex. Note that motion was addressed as part of the analyses; no initial head motion threshold was applied. A total of 423 participants (153 females; age range = 6-20 years; mean age $\pm$ SD = 10.7 $\pm$ 3.0 years) remained in the cohort for subsequent analyses.

4.2.2 HBN MRI Data Acquisition

Structural T1-weighted and functional scans were acquired at the CBIC using a Siemens 3T Prisma scanner with a 32-channel head coil. A 3D T1-weighted MPRAGE volume with spatial resolution 0.8 x 0.8 x 0.8 mm³ was obtained for each participant (224 sagittal slices, TR = 2500 ms, TE = 3.15 ms, flip angle = 8°). The two runs of resting-state fMRI data were acquired with echo-planar imaging sequences (60 slices, voxel size = 2.4 x 2.4 x 2.4 mm³, multiband factor = 6, TR = 800 ms, TE = 30 ms, flip angle = 31°). The participants underwent two 5 min functional runs while they fixated a cross on the screen and both runs consisted of 375 volumes.

4.2.3 MRI Preprocessing Pipeline

The structural T1-weighted and functional data derived from CoRR and HBN were preprocessed using a previously defined pipeline (Yan et al., 2013a; Yan et al., 2013b; Craddock et al., 2013c; see **Chapter 2 (Section 2.2.2)**) which includes AFNI (<https://afni.nimh.nih.gov/>), FSL (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL/>), and ANTs (<https://github.com/ANTsX/ANTs>).

4.2.4 Head Motion

To identify high-motion participants in the CoRR and HBN datasets, a head motion threshold of rmsFD > 0.20 mm was applied (Satterthwaite et al., 2013; Jenkinson et al., 2002). This predefined

threshold is consistent with the one applied in **Chapter 2**. After applying a $\text{rmsFD} > 0.20$ mm, a total of 67 participants remained in the UPSM dataset which corresponds to $\sim 73\%$ of the original sample size. For the HBN, a total of 255 participants were retained in the cohort which represents $\sim 60\%$ of the original sample size.

4.2.5 Parcellation Scheme

I used a whole-brain parcellation scheme to compute the functional connectivity matrices of each participant in the CORR and HBN datasets (<https://neurovault.org/images/395091/>; Shen et al., 2013). The Shen 268 parcellation scheme describes the cortical, subcortical, and cerebellar regions with a reasonable number of parcels. Although the properties of the individual functional connectome are better captured with finer cortical brain atlases (e.g., Schaefer 1000-17Networks; **see Chapter 2**), their applications make them highly computationally intensive to construct functional connectivity matrices for each participant. Similarly, functional networks such as the medial frontal and frontoparietal networks are highly predictive of an individual's functional connectome (**see Chapter 2**). The aims of the present study, however, focus on the whole-brain functional connectome rather than on connectomes derived from specific functional networks. When the Shen 268 parcellation scheme was applied, a total of $[(268 \times 268) - 268] / 2 = 35,778$ unique edges were obtained in the whole-brain functional connectome.

4.2.6 Functional Connectivity: Standard Approach

Using the Shen 268 parcellation scheme, the mean timeseries of each ROI was extracted prior to computing the Pearson's correlation coefficient (r) between all possible ROI pairs to construct a 268×268 functional connectivity matrix for each participant. Hence, the correlation values represent the connectivity strengths (edges) between two ROIs (nodes). This procedure was repeated for the two functional runs (REST1, REST2) such that an individual had two functional connectivity matrices which reflect their connectivity profiles. I eliminated some edges in the functional connectivity matrices for each participant due to the lack of spatial coverage (i.e., zero values) across the whole brain. I also removed duplicate edges by considering only the upper triangular part of the functional connectivity matrices. For CoRR, there were 30,381 unique edges that remained in REST1 and REST2. For HBN, there were 25,200 distinct edges that remained in REST1 and REST2.

4.2.7 Functional Connectivity: Scrubbing

To identify motion-corrupted timepoints in the fMRI timeseries, a volume censoring strategy commonly referred to as “scrubbing” was applied for the CoRR and HBN datasets (Figure 4.1; Power et al., 2015; Power et al., 2014; Power et al., 2013). For each participant, a contaminated timepoint T is identified in their fMRI timeseries if its $\text{rmsFD} > 0.20$ mm. Once T has been

identified, the preceding timepoint ($T - 1$) and two succeeding timepoints ($T + 1$, $T + 2$) are censored in the fMRI timeseries. By removing a total of four timepoints for each T in the fMRI timeseries, the residual motion that persists in the neighbouring timepoints of T is minimised.

4.2.8 Functional Connectivity: Bagging Approach

Using the scrubbed timeseries, I employed the bagging approach (Figure 4.1), where I selected samples of the least motion-corrupted timepoints denoted by minTP. I ranked the scrubbed timeseries by their least rmsFD values and selected the top timepoints based on the minTP threshold. This is an important step that prevents participants from having varying numbers of timepoints in their scrubbed fMRI timeseries. I then took 500 bootstrapped samples, *with* replacement, of TP timepoints. For each participant, a mean bagged functional connectivity matrix was then obtained by averaging the functional connectivity matrices across the 500 samples. I also evaluated the importance of bagging by comparing results to those obtained using only 1 iteration (relative to 500) — a procedure I refer to as “one-sample bootstrap”.

The approach to resampling that was taken here is more similar to that of Biswal et al. (2001), who subsampled task-based fMRI timeseries *without* replacement, than that of Bellec et al. (2010) and Nikolaidis et al. (2020), who used the circular block bootstrap. The circular block bootstrap is considered preferable because it better allows for temporal autocorrelation in the timeseries, which is known to contribute to functional connectivity (Arbabshirani et al., 2014) and fingerprint ID accuracy (Shinn et al., 2023). Here, due to the prioritisation of low-motion timepoints, and the exclusion of the preceding and two succeeding timepoints after a motion-contaminated timepoint, it was not possible to apply circular block bootstrap while still retaining sufficient timepoints for the calculation of functional connectivity. I may expect some impact of disrupted temporal autocorrelation, although the literature is mixed on whether this would be a positive (e.g., Arbabshirani et al., 2014) or negative (e.g., Shinn et al., 2023) outcome.

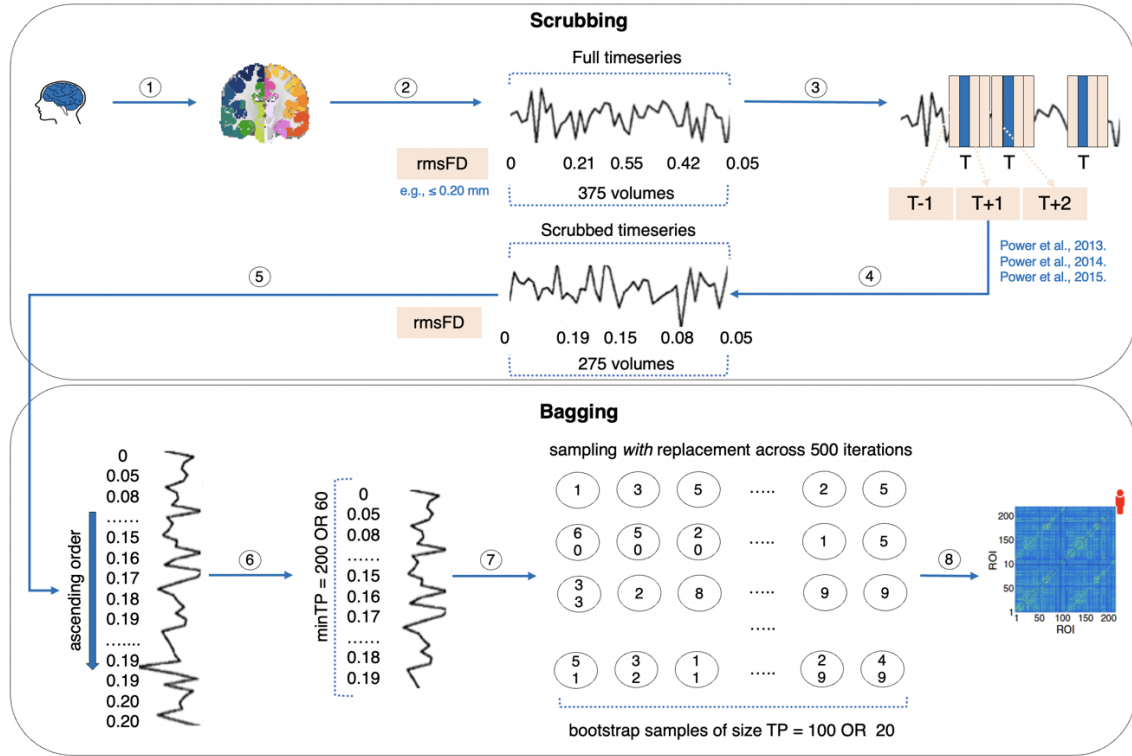


Figure 4.1. Schematic describing the scrubbing and bagging procedures. The Shen 268 parcellation scheme is applied to extract the fMRI timeseries of all the participants in the CoRR and HBN datasets. When scrubbing is performed, a motion-corrupted timepoint is identified in the fMRI timeseries if its rmsFD > 0.20 mm. After censoring that timepoint, its preceding and two succeeding timepoints are also removed in the timeseries. Bagging is performed using the scrubbed timeseries by (1) selecting the least motion-corrupted timepoints (ranked by their rmsFD values) that match a predefined threshold denoted by minTP to prevent participants from having varying number of timepoints which can be sampled from, and (2) bootstrapping samples of a given size TP from the motion-limited timepoints *with* replacement across 500 iterations. For each participant, the mean bagged functional connectivity matrix is computed by averaging the 500 functional connectivity matrices.

4.2.9 Functional Connectome Fingerprinting

I tested the abilities of the standard and bagging approaches to identify an individual based on their functional connectivity profiles from REST1 and REST2 as previously described by Finn et al. (2015), and in **Chapters 1 (Section 1.7) and 2**. The identification procedure was performed by creating the “database” **D** which stored either the standard or bagged functional connectivity matrices of each individual from REST1. Iteratively, the standard or bagged functional connectivity matrix from a given individual from REST2 was selected, and this functional connectivity matrix was treated as the “target matrix” **T**. Similar binary identification (BID) was performed on the standard and bagged functional connectivity matrices by exchanging the **D** $\leftrightarrow$ **T** configurations to compute the mean BID accuracy for the CoRR and HBN datasets. The mean BID accuracies were also computed for the one-sample bootstrap. The mean BID accuracies were then plotted as a function of timepoint (TP) ranging from 60 to the full timeseries of each dataset.

4.2.10 Brain-behaviour Relationships: Standard and Bagging Approaches

To detect univariate brain-behaviour relationships, I performed partial Spearman's Rank correlation (R_s) between functional connectivity and age at the edge level using the full timeseries (i.e., unscrubbed — 375 volumes) from the HBN while treating sex and head motion (indexed by rmsFD) as covariates. The edge that shared the strongest brain-behaviour relationship (highest R_s) was selected after applying a False Discovery Rate (FDR) to correct for multiple comparisons (across 25,200 edges). I identified the strongest edge and its ROI-ROI pair to reduce the high computational costs for the subsequent analyses and I refer to this procedure as the *standard approach*. Following the logic of Marek et al. (2022), I then established the robustness of the brain-behaviour relationships using the standard approach across different sample sizes. Each sample size was bootstrapped *without* replacement over 500 iterations, where I randomly selected samples of the low-motion participants at 13 intervals: $N \in \{20, 40, 60, 80, 100, 120, 140, 160, 180, 200, 220, 240, 255\}$. For each sample size, the mean brain-behaviour R_s and 95% CIs were plotted as a function of N .

I repeated the brain-behaviour analyses by calculating R_s between functional connectivity and age using the scrubbed timeseries from the HBN while treating sex and head motion (indexed by rmsFD) as covariates for the same edge that shared the highest correlation strength with the standard approach. The robustness of the brain-behaviour R_s was computed on the basis of the bagged functional connectivity matrices across different sample sizes which follows similar sampling procedures as the standard approach. The mean brain-behaviour R_s and 95% CIs were plotted as a function of N . I assessed the performance of the standard and bagging approaches by computing their areas under the curve (AUC) such that a lower AUC indicates tighter 95% CIs. To compare the difference in AUC between the two approaches, I computed the rate of change in AUC as denoted by ΔAUC . I also derived the ΔAUC to compare the performance of bagging and one-sample bootstrap.

4.2.11 Salvaging high-movement participants for connectome-based fingerprinting and brain-behaviour relationships

Given that connectome-based fingerprinting and brain-behaviour analyses often exclude a large number of “high-movement” participants with a predefined rmsFD threshold, I examined the possibility of computing fingerprint ID accuracies and brain-behaviour R_s using the bagged functional connectivity matrices derived from the scrubbed timeseries of these individuals *without* applying the initial rmsFD threshold set at 0.20 mm. Connectome-based fingerprinting was first performed as described in **Section 4.2.9** with the bagged functional connectivity matrices from REST1 and REST2 with all the participants in the CoRR and HBN datasets

including the high-motion individuals who had at least minTP uncorrupted timepoints available. The mean BID accuracies were plotted as a function of timepoint (TP) ranging from 60 to the full timeseries of each dataset. Subsequently, the univariate brain-behaviour analyses were repeated with bagging by including the high-motion participants who had least motion-corrupted timepoints that matched the minTP threshold and who would have otherwise been discarded. The robustness of the brain-behaviour R_s was computed on the basis of the bagged functional connectivity matrices across varying sample sizes derived from all the participants in the sample who had enough useable timepoints. The mean brain-behaviour R_s and 95% CIs were also plotted as a function of N .

Finally, I assessed the reproducibility of bagging in two ways. First, I assessed the ability of bagging to salvage the high-movement participants who would otherwise be discarded for brain-behaviour associations by applying a stringent rmsFD threshold set at 0.08 mm as previously reported by Marek et al. (2022). Second, I investigated the impact of bagging on brain-behaviour associations for the first 10 edges selected on the basis of their correlation strengths (R_s) after applying an FDR threshold to correct for multiple comparisons (across 25,200 edges). Given that the edges share a behavioural relevance with age that may differ in their magnitude and direction, I assessed whether bagging can produce robust brain-behaviour relationships regardless of the valence of specific edges in the functional connectome. I computed the AUC for the bagging and standard approaches followed by the Δ AUC and one-sample bootstrap comparison. The mean brain-behaviour R_s and 95% CIs derived from the bagging and standard approaches were also plotted as a function of N .

4.3 Results

4.3.1 Applying bagging to perform connectome-based fingerprint identification

To test whether bagging generates reproducible functional connectivity matrices, I derived the fingerprint ID accuracies from CoRR ($N = 67$) and HBN ($N = 255$) after applying a rmsFD ≤ 0.20 mm threshold to exclude participants with excess motion. Given that fingerprint ID accuracy is usually computed with the full timeseries, and tends to increase with longer timeseries (Vanderwal et al., 2021; Venkatesh et al., 2020; Vanderwal et al., 2017; Finn et al., 2015; see **Chapter 2**), I expectedly found that the BID accuracies were optimal with the standard approach (Figure 4.2). For CoRR, the BID accuracy was 42% and for HBN, the BID accuracy was 78%. When I applied bagging with the scrubbed timeseries with minTP = 60 and varied TP from 60 to the full number of timepoints of each dataset, the BID accuracy for CoRR and HBN achieved good accuracy, albeit lower than that for the full timeseries (33% for CoRR and 64% for HBN;

Figure 4.2). This means that using bootstrap samples of only 60 timepoints largely preserved BID accuracy. There was a clear benefit of computing the functional connectome on the basis of bootstrapped data —when only one sample (rather than 500) was taken, BID accuracies were lower, at least for smaller TP (TP = 60, 80; Figure 4.2).

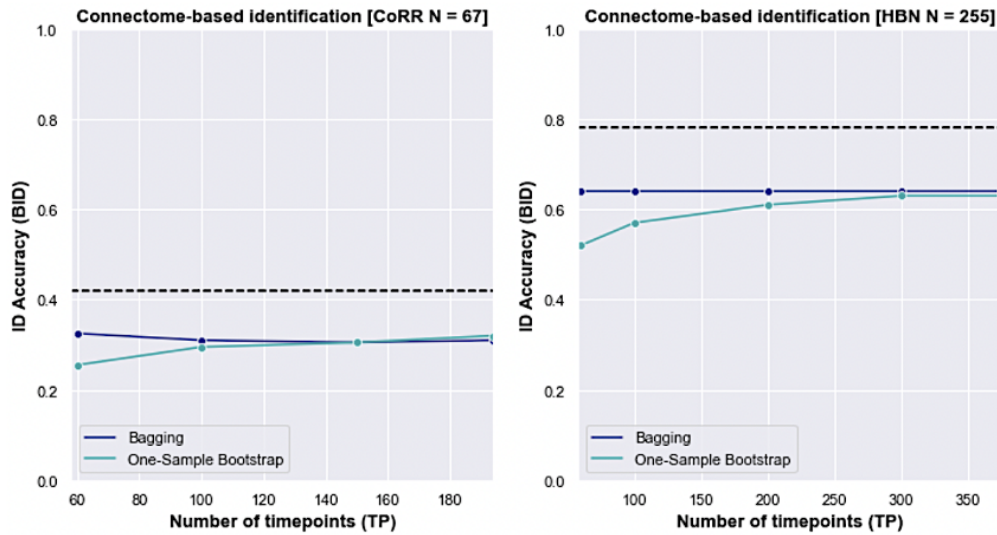


Figure 4.2. Applying bagging and one-sample bootstrap (minTP = 60) to perform connectome-based fingerprint identification derived from CoRR and HBN. Left Panel: Fingerprint accuracy measured by BID as a function of TP ranging from 60 to 194 by excluding “high-motion participants” with a rmsFD > 0.20 mm. Right Panel: Fingerprint accuracy measured by BID as a function of TP ranging from 60 to 375 by excluding “high-motion” participants with a rmsFD > 0.20 mm. Dotted black lines represent the BID accuracy obtained from the standard approach using the full timeseries of each dataset. Solid navy lines denote the BID accuracies obtained from bagging. Solid teal lines correspond to the BID accuracies obtained from one-sample bootstrap.

4.3.2 Standard brain-behaviour relationships as a function of sample size

I computed univariate brain-behaviour associations by performing partial Spearman’s Rank correlation (R_s) between functional connectivity and age while treating head motion (indexed by rmsFD) and sex as covariates. I first employed the standard approach, which uses the unscrubbed timeseries (i.e., 375 volumes) from REST1 derived from the HBN ($N = 255$). The brain-behaviour relationships between functional connectivity and age were obtained using the strongest edge (highest R_s) from the whole-brain functional connectome after applying an FDR threshold to correct for multiple comparisons (across 25,200 edges).

Plotting the brain-behaviour associations between functional connectivity and age as a function of N , revealed an expected tightening of the 95% CIs as the sample size increased from small ($N = 20$) to large ($N = 255$) (Figure 4.3). $|\bar{R}_s|$ was 0.35 across N s.

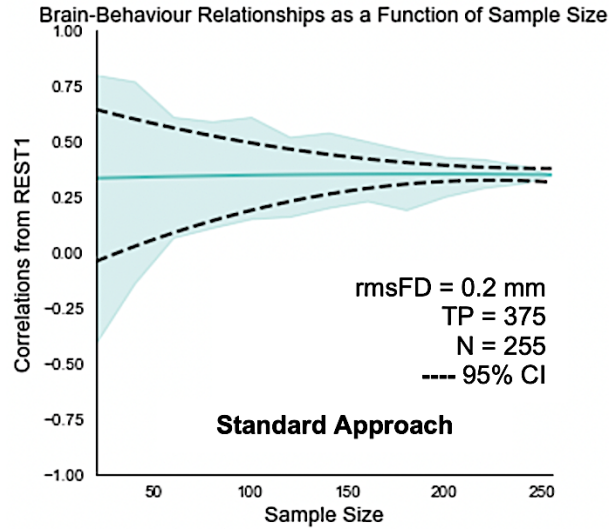


Figure 4.3. Correlations between functional connectivity and age obtained from REST1 using the standard approach as a function of sample size. Solid teal line shows the mean correlations derived from 500 bootstrap samples for a given sample size. Teal shading denotes the minimum and maximum correlations across 500 bootstrap samples for a given sample size. Black dotted line represents the lower and upper bounds of the 95% CIs. After applying a $\text{rmsFD} \leq 0.20$ mm to exclude the “high-movement” participants, the sample size was reduced to 255.

4.3.3 Bagged brain-behaviour relationships as a function of sample size

I repeated the analysis of brain-behaviour relationships as a function of sample size using the bagging approach (applied to REST1, HBN dataset). The brain-behaviour relationships between functional connectivity and age were computed using the strongest edge (highest R_s) from the whole-brain functional connectome that was obtained from the standard approach. In the first instance, I set $\text{minTP} = 200$, i.e., I selected 200 least motion-corrupted timepoints from the scrubbed timeseries based on their rmsFD values. As a result, the sample size was reduced to $N = 239$. To compare the performance of bagging with the standard approach, I repeated the brain-behaviour analyses using the standard approach for $N = 239$. Using the standard approach, I observed similar tightening of the 95% CIs as the sample size increased from typical ($N = 20$) to large ($N = 239$), and $|\bar{R}_s|$ was 0.37 (Figure 4.4A). Bagging using $\text{TP} \in \{20, 60, 100\}$ achieved robust correlations between functional connectivity and age which were comparable to those obtained using the standard approach, across the three TP thresholds ($|\bar{R}_s| = 0.35$ for each TP) (Figure 4.4B).

These observations were confirmed by the AUC analysis (Figure 4.4C). The ΔAUC , which quantifies the amount of change in AUC between the standard and bagging approaches, was small, ranging from 0.9% to 2.7% across the three TP thresholds ($\Delta\text{AUC}_{\text{TP}20} = 2.7\%$; $\Delta\text{AUC}_{\text{TP}60} = 0.9\%$; $\Delta\text{AUC}_{\text{TP}100} = 1.0\%$). In contrast, ΔAUC for one-sample bootstrap was associated with significant widening of the 95% CIs ($\Delta\text{AUC}_{\text{TP}20} = 29.1\%$; $\Delta\text{AUC}_{\text{TP}60} = 14.5\%$; $\Delta\text{AUC}_{\text{TP}100} = 12.6\%$) (Figure 4.4D).

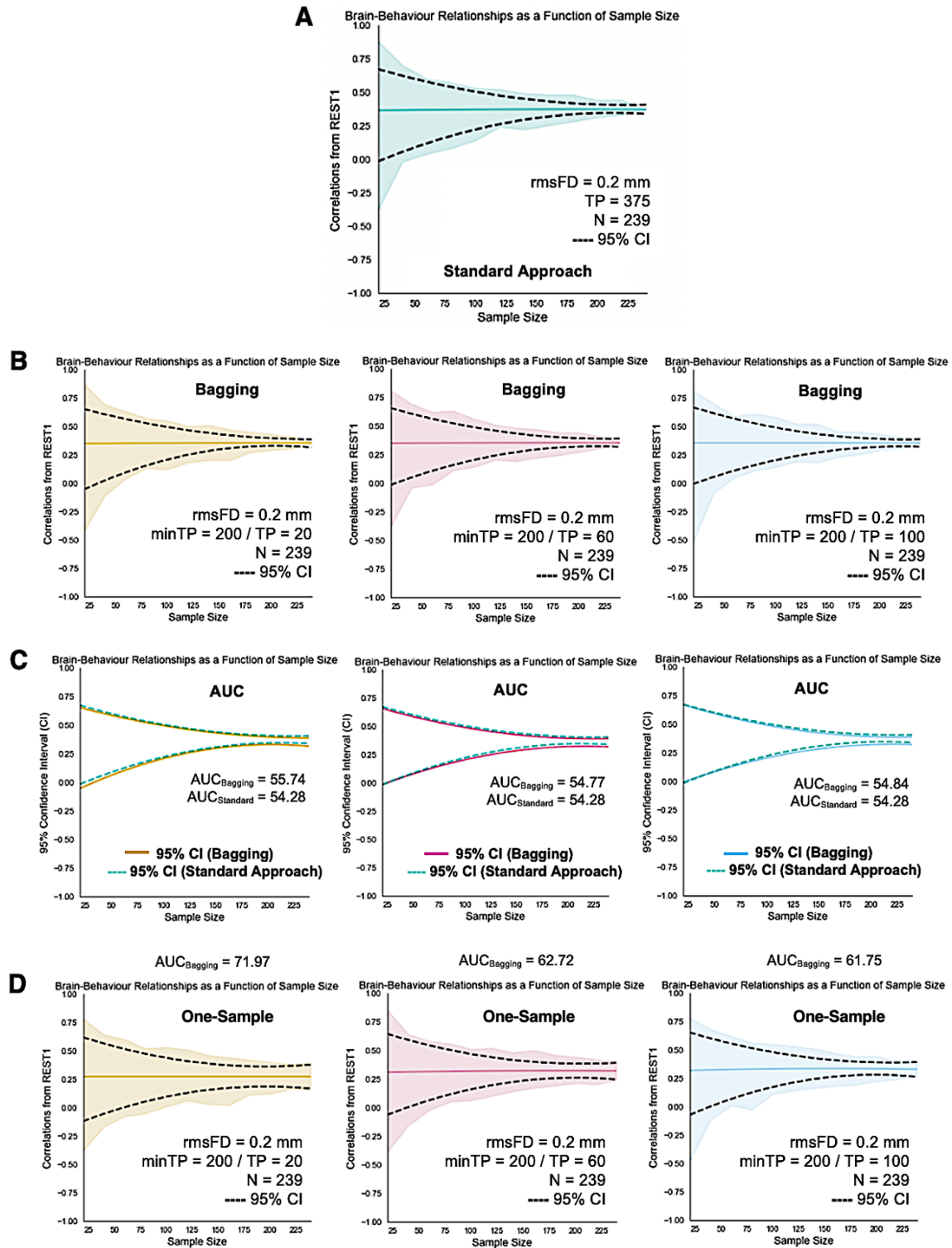


Figure 4.4. (A) Correlations between functional connectivity and age obtained from REST1 using the standard approach as a function of sample size. The sample size was set to 239 to facilitate the comparison of the standard approach with bagging. (B) Correlations between functional connectivity and age obtained from REST1 as a function of sample size using bagging during which participants have least amount of contaminated timepoints that match the minTP threshold (e.g., 200) and number of bagged timepoints, i.e., TP (e.g., 20, 60, 100). (C) Comparison of the 95% CIs between the standard approach and bagging with the area under curve (AUC). Dotted coloured line represents the 95% CIs derived from the standard approach. Solid coloured line represents the 95% CIs derived from bagging. (D) Correlations between functional

connectivity and age obtained from REST1 as a function of sample size using one-sample bootstrap.

4.3.4 Robust brain-behaviour associations with minimally bootstrap timepoints

To investigate the impact of the number of timepoints available for analysis, I reduced minTP from 200 to 60. Setting TP to 20 increased the sample size to $N = 255$. Even with as few as 20 timepoints bootstrap sampled from 60 total timepoints, I observed robust brain-behaviour associations (Figure 4.5). Standard and bagging approaches yielded comparable brain-behaviour associations as demonstrated by their AUCs ($\Delta AUC_{TP20} = 2.9\%$). Increased variability in the brain-behaviour correlations was also observed with one-sample bootstrap relative to bagging ($\Delta AUC_{TP20} = 18.3\%$) (Figure 4.5).

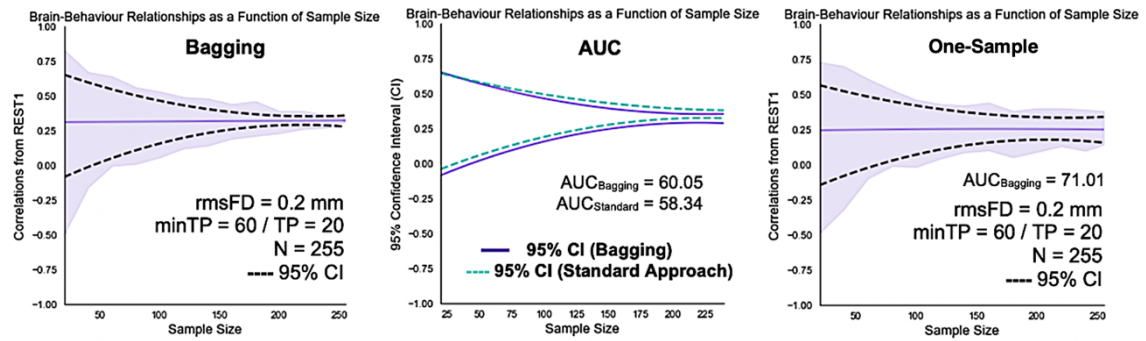


Figure 4.5. Correlations between functional connectivity and age obtained from REST1 as a function of sample size using bagging during which participants have minimally bootstrap timepoints that match the minTP threshold (e.g., 60) and number of bagged timepoints, i.e., TP (e.g., 20). Solid purple line shows the mean correlations, purple shading denotes the minimum and maximum correlations, and black dotted line represents the lower and upper bounds of the 95% CIs. The AUC is displayed to compare the 95% CIs between the standard approach and bagging. Dotted blue line represents the 95% CIs derived from the standard approach. Solid purple line represents the 95% CIs derived from bagging. The correlations between functional connectivity and age obtained from REST1 as a function of sample size is also shown using one-sample bootstrap.

4.3.5 Influence of minTP and TP on bagged brain-behaviour relationships

Figure 4.6 illustrates the influence of minTP ($\text{minTP} \in \{60, 100, 200\}$) and TP ($\text{TP} \in \{20, 60, 100\}$) directly. Varying the minTP threshold revealed a non-linear relationship between the widths of the 95% CIs and sample size (Figure 4.6A). Across the three minTP thresholds, the widths of the 95% CIs were considerably large ($\text{width}_{\text{minTP}60} = 0.86$; $\text{width}_{\text{minTP}100} = 0.83$; $\text{width}_{\text{minTP}200} = 0.80$) at typical sample sizes ($N = 20$) but narrowed non-linearly ($\text{width} = 0.01$ for each minTP) with increasing number of participants ($N = 255$). When I varied the TP thresholds, I detected a similar non-linear relationship between the widths of the 95% CIs and sample size (Figure 4.6B). Across the three TP thresholds, the widths of the 95% CIs were large ($\text{width}_{\text{TP}20} =$

0.86; $\text{width}_{\text{TP}60} = 0.77$; $\text{width}_{\text{TP}100} = 0.80$) at small sample sizes ($N = 20$) but non-linearly tightened (width = 0.01 for each TP) as the sample size became large ($N = 255$).

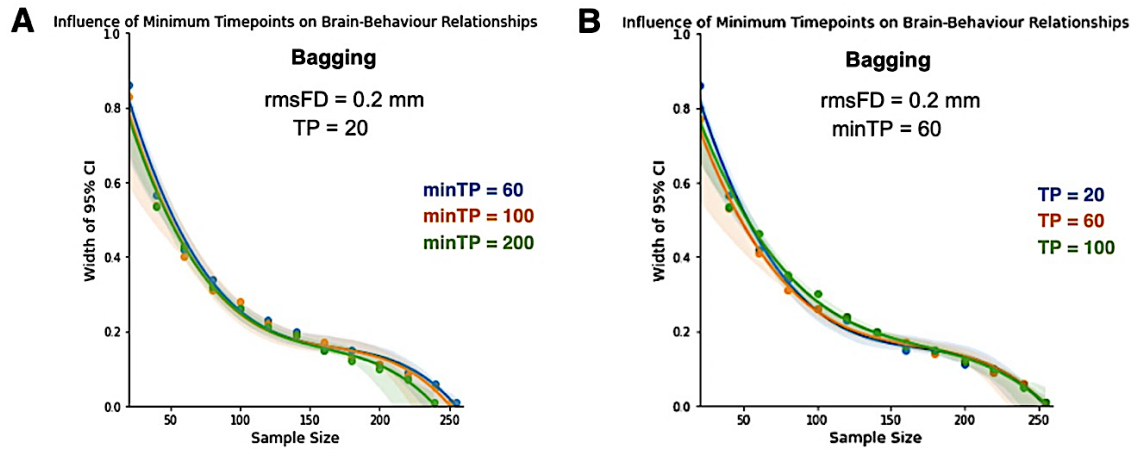


Figure 4.6 (A) Influence of minTP using bagging on the width of the 95% CIs by keeping TP constant at 20 such that $\text{minTP} \in \{60, 100, 200\}$. (B) Influence of TP using bagging on the width of the 95% CIs by keeping minTP constant at 60 such that $\text{TP} \in \{20, 60, 100\}$.

4.3.6 Applying bagging to perform connectome-based fingerprint identification with “high-movement” individuals

Until now, I employed bagging to perform all the analyses after excluding participants in the CoRR and HBN datasets with a $\text{rmsFD} > 0.20$ mm, resulting in the retention of $\sim 73\%$ of the CoRR ($N = 67/92$) and $\sim 60\%$ of the HBN ($N = 255/423$) participants. In this analysis, I examined the impact of using bagging to retain as many participants as possible. I set $\text{minTP} = 60$ and $\text{TP} = 20$, which helped retain $\sim 88\%$ of the CoRR ($N = 81/92$) and $\sim 90\%$ of the HBN ($N = 379/423$) participants. BID accuracies were optimal with the standard (full timeseries) approach for both CoRR (33%) and HBN (71%) (Figure 4.7). Applying bagging to the scrubbed timeseries with $\text{minTP} = 60$ and varied TP from 60 to the full timeseries length of each dataset, reduced BID accuracy to 28% for CoRR and 54% for HBN. The one-sample bootstrap exhibited greater variability across TP (Figure 4.7).

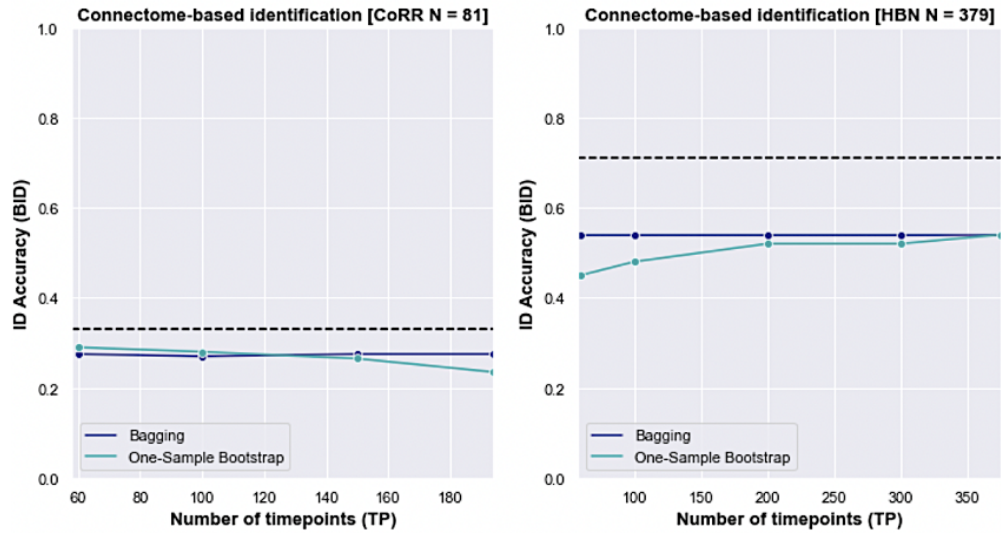


Figure 4.7. Applying bagging and one-sample bootstrap (minTP = 60) to perform connectome-based fingerprint identification derived from CoRR and HBN. Left Panel: Fingerprint accuracy measured by BID as a function of TP ranging from 60 to 194 by including the salvaged “high-motion” participants. Right Panel: Fingerprint accuracy measured by BID as a function of TP ranging from 60 to 375 by including the salvaged “high-motion” participants. Dotted black lines represent the BID accuracy obtained from the standard approach using the full timeseries of each dataset. Solid navy lines denote the BID accuracies obtained from bagging. Solid teal lines correspond to the BID accuracies obtained from one-sample bootstrap.

4.3.7 Salvaging “high-movement” participants for brain-behaviour associations

Including high-motion participants ($N = 379$) to compute the brain-behaviour associations with bagging (minTP = 60 and TP = 20) yielded $|\bar{R}_s|$ of 0.24 and the expected narrowing of the 95% CIs with increasing sample sizes (Figure 4.8). As before, brain-behaviour correlations showed increased variability when one-sample bootstrap was performed, relative to bagging ($\Delta\text{AUC}_{\text{TP}20} = 22.5\%$) (Figure 4.8).

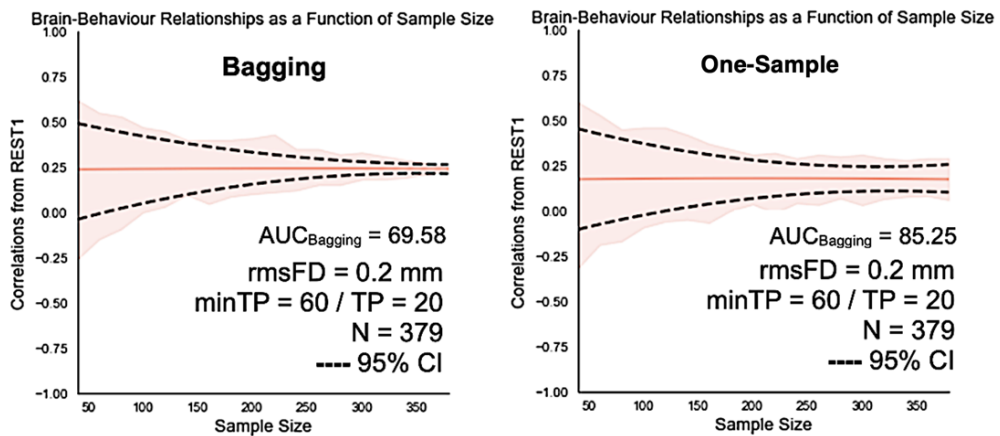


Figure 4.8. Salvaging “high-motion” participants to compute univariate brain-behaviour associations with bagging without applying the initial head motion threshold of $\text{rmsFD} \leq 0.20$ mm, i.e., the timeseries of all the participants were scrubbed with a $\text{rmsFD} \leq 0.2$ mm. By applying bagging with minTP = 60 and TP = 20, the sample size was increased to 379. The correlations between functional connectivity and age obtained from REST1 as a function of sample size are

shown using bagging and one-sample bootstrap. The AUC is displayed to compare the 95% CIs between the standard approach and bagging in addition to the standard approach and one-sample bootstrap. Solid pink line shows the mean correlations, pink shading denotes the minimum and maximum correlations, and black dotted line represents the lower and upper bounds of the 95% CIs.

4.3.8 Reproducibility of bagging across head motion thresholds and edges in the functional connectome

Up to this point, a motion threshold of $\text{rmsFD} \leq 0.20$ mm was applied. Next, I evaluated the impact of a strict head motion threshold, i.e., $\text{rmsFD} \leq 0.08$ mm as recently utilised by Marek et al. (2022) to identify the high-movement participants. First, I applied this threshold to the HBN dataset prior to computing the univariate correlations with bagging ($\text{minTP} = 60$ and $\text{TP} = 20$), which resulted in the retention of only $\sim 18\%$ of the participants ($N = 77/423$). $|\bar{R}_s|$ was 0.22, and 95% CIs tightened as before (Figure 4.9A). Greater variability in the brain-behaviour associations was observed when one-sample bootstrap was performed relative to bagging ($\Delta\text{AUC}_{\text{TP}20} = 38.3\%$) (Figure 4.9A). Second, I retained participants with at least $\text{minTP} = 60$ and $\text{TP} = 20$, which resulted in $\sim 22\%$ of the participants ($N = 92/423$) for the brain-behaviour analyses. $|\bar{R}_s|$ was 0.25 (Figure 4.9B). However, it is important to note that this strict rmsFD threshold retains fewer participants which can inflate the effect sizes of brain-behaviour relationships due to smaller N s.

So far, I computed R_s using the standard and bagging approaches derived from the edge that shares the strongest correlation between functional connectivity and age. To test the reproducibility of bagging ($\text{minTP} = 60$ and $\text{TP} = 20$) across different edges in the whole-brain connectome, I selected an additional 9 edges to compute the brain-behaviour associations as a function of N . I first employed the standard approach and found that the AUCs varied from 58.34 to 62.31 (mean $\text{AUC} \pm \text{SD}$ 60.19 ± 1.38 ; Figure 4.10). When comparing the performance of the standard approach for these 9 edges ($0.30 \leq R_s \leq 0.32$) to performance for the strongest edge ($R_s = 0.35$), ΔAUC was relatively small, ranging from 0% to 7.2%. Bagging applied to data that included the salvaged high-motion participants ($N = 379$) yielded AUCs ranging from 77.79 to 82.99 (mean $\text{AUC} \pm \text{SD}$ 80.34 ± 1.79 ; Figure 4.11), and ΔAUC ranging from 11.2% to 19.2%.

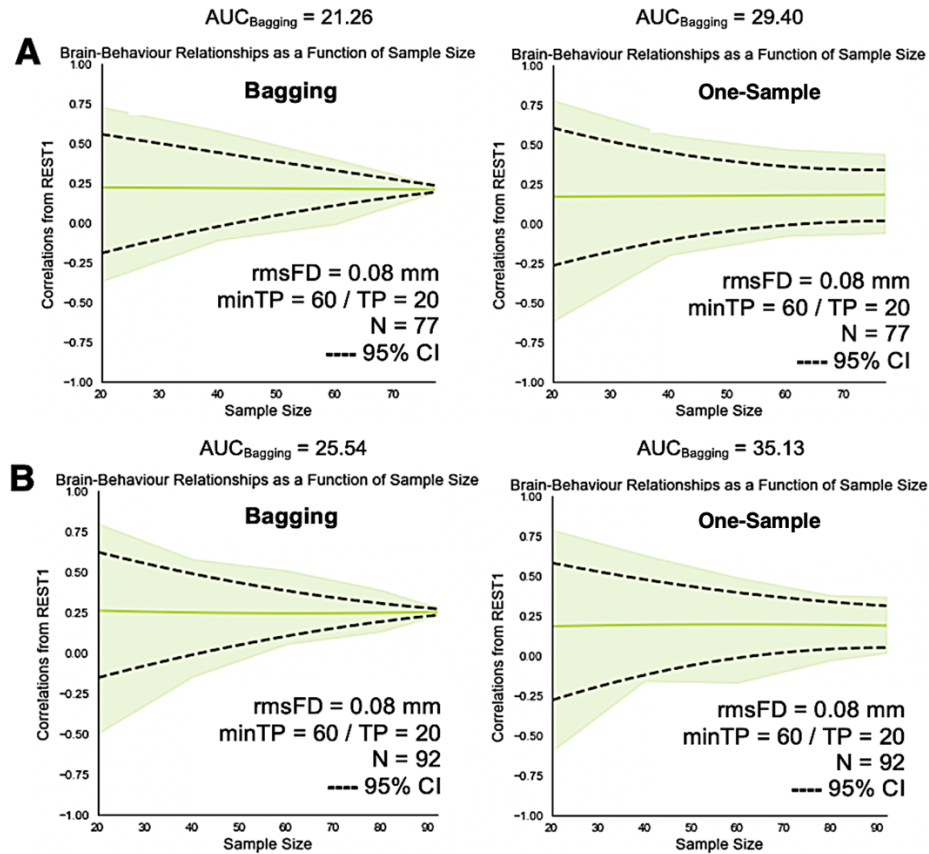


Figure 4.9. Impact of applying a strict head motion threshold on brain-behaviour associations using bagging ($minTP = 60$ and $TP = 20$). (A) After applying a $rmsFD \leq 0.08 \text{ mm}$ to exclude “high-movement” participants, the sample size was reduced to 77. Correlations between functional connectivity and age obtained from REST1 as a function of sample size are shown using bagging across 500 and one-sample bootstrap. (B) After scrubbing the timeseries with a $rmsFD \leq 0.08 \text{ mm}$ of all the participants, the sample size was increased to 92. Correlations between functional connectivity and age obtained from REST1 as a function of sample size are shown using bagging across 500 and one-sample bootstrap. The AUC is displayed to compare the 95% CIs derived from bagging across 500 and one-sample bootstrap. Solid green line shows the mean correlations, green shading denotes the minimum and maximum correlations, and black dotted line represents the lower and upper bounds of the 95% CIs.

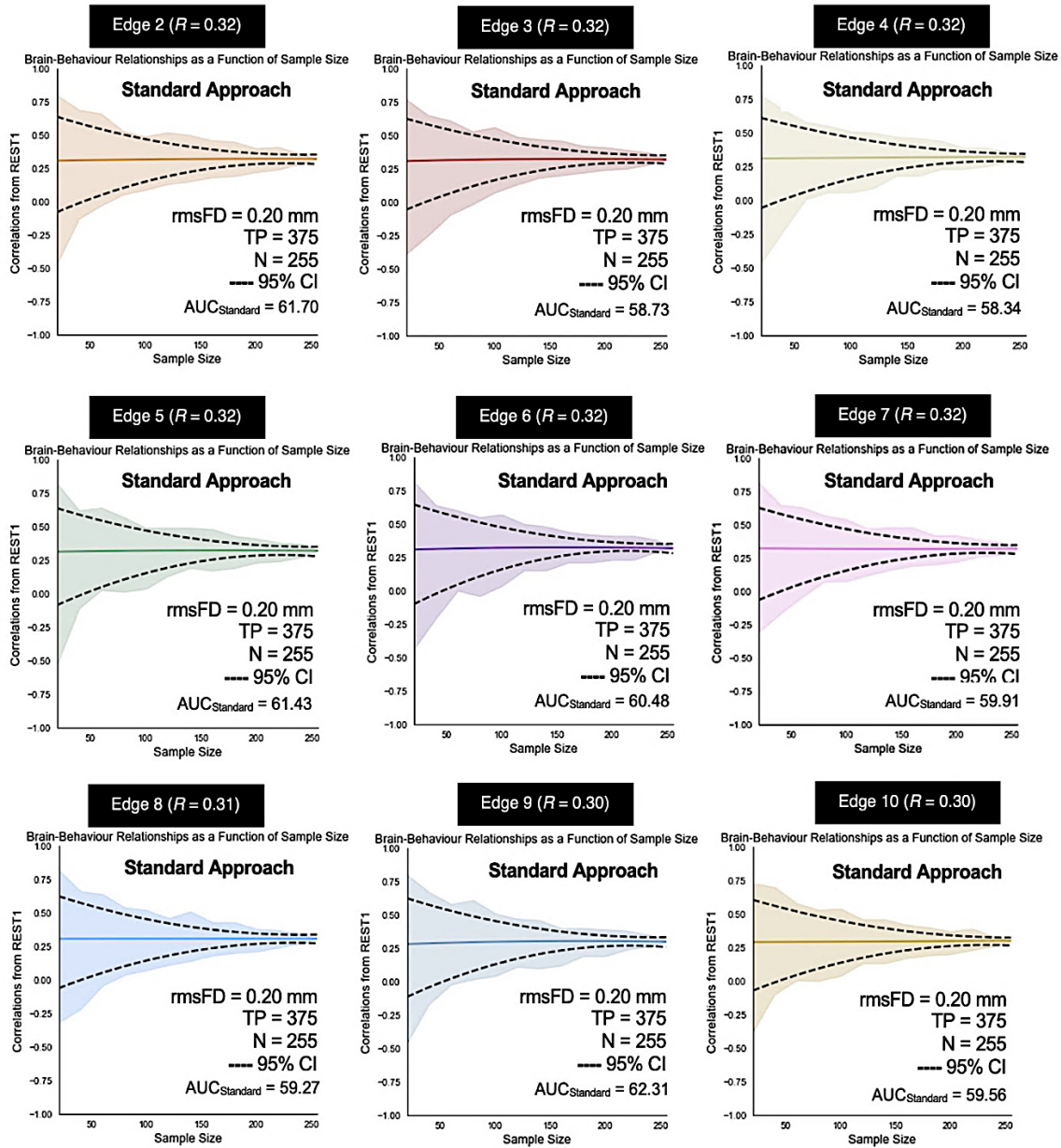


Figure 4.10. Reproducibility of the standard approach in deriving the associations between functional connectivity and age from an additional 9 edges in the connectome. Correlations between functional connectivity and age obtained from REST1 as a function of sample size are shown using the standard approach across the 9 edges. The AUCs are displayed to compare the 95% CIs derived from the standard approach across the 9 edges. Solid coloured lines show the mean correlations, coloured shadings denote the minimum and maximum correlations, and black dotted lines represent the lower and upper bounds of the 95% CIs. After applying a rmsFD $\leq$ 0.20 mm to exclude the “high-movement” participants, the sample size was reduced to 255.

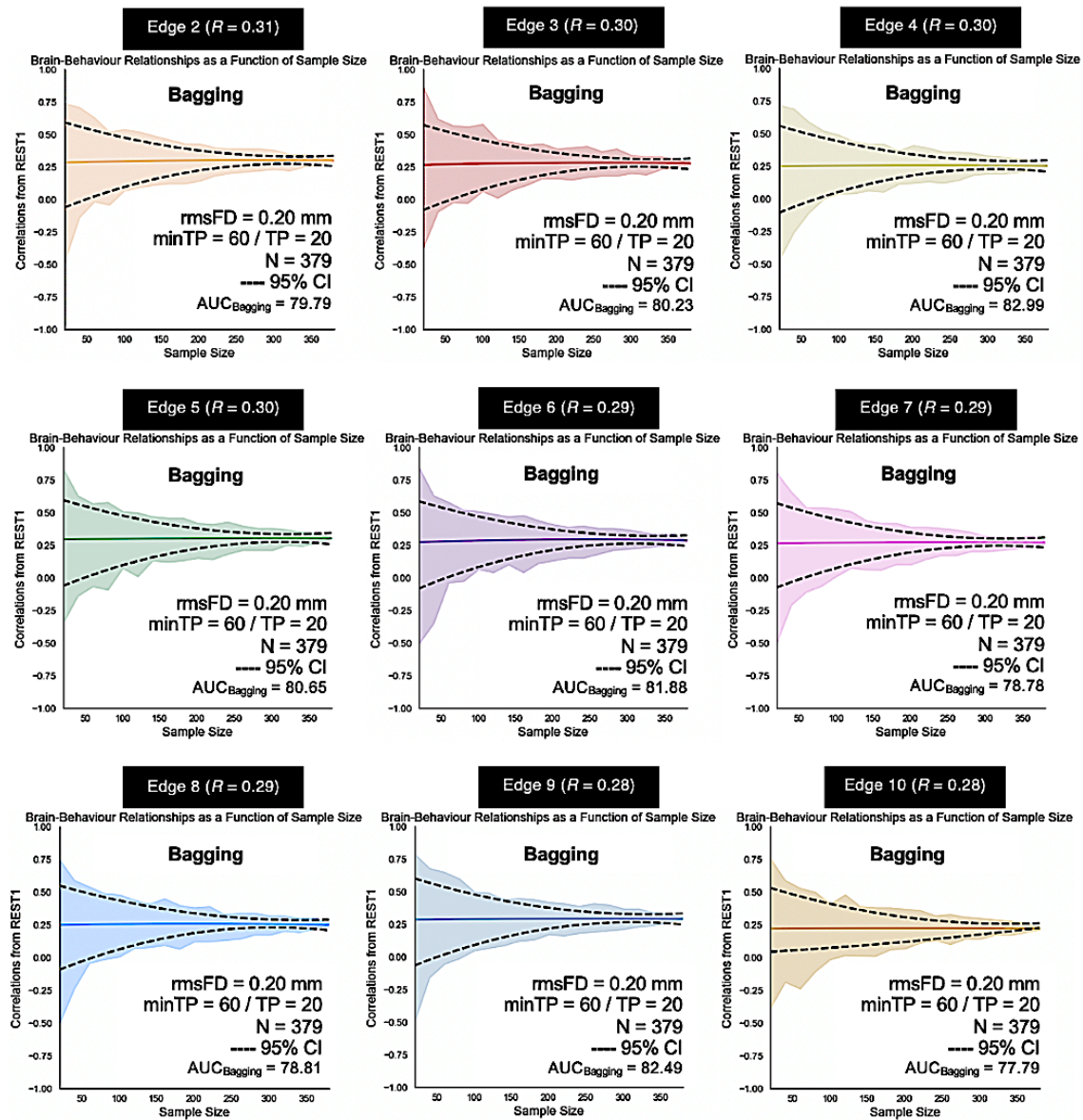


Figure 4.11. Reproducibility of bagging (minTP = 60 and TP = 20) including the salvaged “high-movement” participants in deriving the associations between functional connectivity and age from an additional 9 edges in the connectome. Correlations between functional connectivity and age obtained from REST1 as a function of sample size are shown using bagging across the 9 edges. The AUCs are displayed to compare the 95% CIs derived from bagging across the 9 edges. Solid coloured lines show the mean correlations, coloured shadings denote the minimum and maximum correlations, and black dotted lines represent the lower and upper bounds of the 95% CIs. After rescuing the “high-movement” participants, the sample size increased to 379.

4.4 Discussion

In modern neuroimaging research, the quest for reproducible brain-behaviour relationships has underscored a need for large sample sizes that can provide adequate statistical power and rigorous head motion thresholds that can improve data quality. These imperatives can come into conflict: applying strict head motion thresholds reduces sample size, sometimes substantially, particularly in developmental cohorts. To address this conflict, the present study assessed the impact of a

resampling technique known as bagging on fingerprint accuracy and brain-behaviour associations in a sample of developing youths. First, I demonstrated that bagging produces reproducible functional connectivity matrices that successfully differentiate an individual from a group — *with* and *without* the inclusion of the high-motion individuals' bootstrap motion-limited timepoints. This was achieved in samples of developing youths from both typical- N (CoRR) and large- N (HBN) datasets. Second, I showed that functional connectomes computed using bagging retain validity, yielding robust brain-behaviour correlations that vary as a function of sample size, even when computed using as few as 60 timepoints (i.e., 16% of the original HBN fMRI timeseries). Third, I showed that bagging improves inclusivity by salvaging high-motion participants — analyses including these individuals yielded brain-behaviour association patterns that were comparable to analyses that excluded them. Last, I demonstrated that bagging generates reproducible brain-behaviour associations across different head motion thresholds and edges in the whole-brain functional connectome.

To test the utility of bagging, connectome-based identification was performed using aggregated functional connectomes computed using bootstrapped timeseries samples in two developmental datasets. Although I and others (Vanderwal et al., 2021; Venkatesh et al., 2020; Vanderwal et al., 2017; Finn et al., 2015) have previously shown that the ability to distinguish someone from their functional connectivity fingerprints increases in association with scan duration (see **Chapter 2**), here, I demonstrated that bagging can successfully identify developing youths from bootstrap motion-limited timepoints — using as few as 60 timepoints, which represents scan lengths < 2 min. Although there was a small decrement in ID accuracy, using even minimal timeseries ($\text{minTP} = 60 / \text{TP} = 60$) preserved fingerprint ID accuracy by more than 75% *with* and *without* the salvaged high-motion participants across typical (CoRR) and consortia (HBN) datasets. This should be weighed against the possible advantages of inclusivity and larger N s which typically decrease fingerprint ID accuracy (Li et al., 2021; Waller et al., 2017; see **Chapter 2**). While there is considerable scope for further analyses to probe the opportunities and limitations of bagging, I suggest that it may be applicable as a reproducible technique to increase N and minimise motion-related data loss, particularly in challenging youth populations for which the connectome distinctiveness is low (Dufford et al., 2021; see **Chapter 1 (Sections 1.8)**).

When bagging was applied *with* and *without* the inclusion of high-motion participants, robust brain-behaviour associations were observed as the sample size grew from small to large. This is to be expected due to sampling variability; using two metrics, the CI and AUC, I showed that variability reduces as expected as the number of participants being sampled increases. This means that the brain-behaviour correlations are inflated with typical (e.g., $N = 20$) compared to larger (e.g., $N = 255$) sample sizes as smaller studies are more prone to exhibit random variations in

these associations in addition to overestimating effect sizes that limit their reproducibility (Marek et al., 2022; see **Chapters 1 and 3** for more details about the impact of sampling variability in brain-behaviour association studies). Although the magnitude of the impact quantified by the change in the area under the CI curve (ΔAUC) varied across edges, bagging had a relatively small impact, relative to the standard approach, suggesting that it does not negatively impact validity. This also demonstrates that the salvaged high-motion developing youths have useable timepoints that are meaningful for brain-behaviour associations; their inclusion may also increase both validity and reproducibility by mitigating the range restriction that occurs when they are excluded. This presents a promising avenue for future research. While several head motion mitigation strategies have been proposed (see **Chapter 3 (Section 3.3)** for a review), most have focused on excluding individuals, rather than maximising inclusion. One recent study by Smith et al. (2022) proposed a framework to salvage first grade high-motion children ($N = 101$, 6-8 years) by employing two motion correction procedures (i.e., scrubbing and ICA-FIX) and two quality control metrics (e.g., functional connectivity and temporal SNR). By varying the mean FD thresholds, the authors showed that a mean FD = 0.30 mm would be sufficient to remove most motion artefacts by retaining ~83% of the first grade children (Smith et al., 2022). The main advantage of bagging is that it minimises the risk of discarding large amount of high-motion participants. Excluding tens or thousands of these high-motion participants leads to a waste of participant time and effort as well as research resources, considering the time, cost, and effort it took for data collection and pre-/post-processing the imaging data in addition to the infrastructure needed to store the raw/preprocessed data (e.g., Marek et al., 2022; see **Chapter 3**). Moreover, the lack of sample inclusivity and diversity in a dataset means that the estimated brain-behaviour relationships may not be reproducible in other studies. Bagging may enable researchers to strike a better balance between data quality and data inclusivity.

Importantly, bagging yielded robust brain-behaviour associations across different edges in the functional connectome, suggesting that it generalises across connections and networks. Nonetheless, future work should evaluate the generalisability and reproducibility of bagging in greater depth, using a larger number of edges, behavioural phenotypes, and datasets. Age was chosen as the behaviour of interest to perform brain-behaviour associations as its relationship with functional connectivity is highly robust and has been well-studied (Grayson et al., 2017; Geerlings et al., 2015a; Song et al., 2014; Di Martino et al., 2014a; Menon, 2013; Uddin et al., 2011; Park and Reuter-Lorenz, 2009; Fair et al., 2009; Fair et al., 2007). Future studies should examine the impact of bagging on brain-behaviour associations for cognitive ability and general psychopathology — which have recently been shown to generate reproducible BWAS at the expense of very large sample sizes (Marek et al., 2022). Moreover, achieving generalisable brain-behaviour associations with other datasets remains important, especially in neurodevelopmental

cohorts (e.g., ABIDE). Since children with psychiatric diagnoses generally are most likely to be excluded due to motion, applying bagging to maximise inclusivity may help address reproducibility concerns (discussed in **Chapter 1**). I return to the generalisability of bagging in **Chapter 6**.

Bagging further demonstrated to yield reproducible brain-behaviour associations across different edges in the functional connectome. This suggests that bagging does not depend on specific connections across the whole brain. Nonetheless, future work should evaluate the generalisability of bagging using other behavioural phenotypes. Age was chosen as the behaviour of interest to perform brain-behaviour associations as its relationship with functional connectivity has been well-studied in neurodevelopment (Grayson et al., 2017; Geerlings et al., 2015a; Song et al., 2014; Di Martino et al., 2014a; Menon, 2013; Uddin et al., 2011; Park and Reuter-Lorenz, 2009; Fair et al., 2009; Fair et al., 2007). For example, bagging could be applied to study the brain-behaviour associations with measures that index cognitive ability and general psychopathology — which have recently been shown to generate reproducible BWAS though at the expense of very large sample sizes (Marek et al., 2022). Moreover, achieving generalisable brain-behaviour associations with other datasets remains important, especially in challenging youth populations such as ABIDE and ABCD Study. Given that these populations are those who exhibit the most motion and are also most likely to be excluded, externally validating the performance of bagging in rescuing these high-motion individuals would generalise its potential across other (typical and/or consortia) datasets. I return to the generalisability of bagging in **Chapter 6**.

Although bagging has been shown to successfully produce robust and reproducible brain-behaviour associations in developing youths, the current study has several limitations. First, bagging requires the fMRI timeseries to be scrubbed using a head motion threshold, i.e., a proportion of volumes has to be removed from the total number of volumes acquired after being confounded by head motion. If a study acquires a small number of volumes (e.g., 100), then scrubbing may still exclude a large number of participants due to insufficient amount of timepoints that are least confounded by head movement, even if the aim is to retain as few as 60 timepoints. Second, bagging was tested using scans collected both within and between imaging sessions. Comparing fingerprint ID accuracies across different datasets is difficult due to such temporal factors as well as different sample sizes and number of timepoints. Lastly, bagging was associated with a decrement in fingerprint ID accuracy relative to the standard approach. One potential explanation is that this reflects disruption of temporal autocorrelation, which would negatively impact functional connectivity strengths (see **Section 4.2.8**). There is considerable scope for future work to address these limitations and to further probe the utility, reproducibility, and validity of the bagging approach.

4.5 Conclusions

Timeseries bagging can generate reproducible functional connectomes that exhibit robust univariate brain-behaviour relationships in developing youths. By maximising participant inclusivity in a given dataset, this strategy affords a promising opportunity to salvage high-motion participants who would otherwise be discarded in typical brain-behaviour association studies. Bagging may reduce the tension between data quality and sample size requirements in neuroimaging, and in doing so, mitigate against resource wastage and the unethical practice of discarding large number of participants who have given their time, effort, and trust to the research process. While the results of this study may be considered to offer preliminary support for the approach, future work should probe its utility, reproducibility, and validity in greater depth and across samples and phenotypes. The next chapters (**Chapters 5 and 6**) begin this process by assessing whether bagging and the fingerprint-informed pipelines identified in **Chapter 2** can improve the strength and generalisability of behavioural predictions derived from connectome-based predictive modelling.

Chapter 5

Detecting reproducible brain-behaviour relationships: the connectome-based predictive modelling framework

Summary

Chapter 5 describes the connectome-based predictive modelling (CPM) framework, which builds predictive linear models from functional connectivity patterns to predict behavioural phenotypes at the individual level. I discuss the application of CPM for predicting a wide range of behavioural phenotypes that are also generalisable across independent datasets, and identify methodological factors that affect the robustness of CPM-derived predictions. Finally, I build on the *empirical* research conducted in **Chapters 2 and 4** by considering whether the factors that maximise the detection of individual differences in the functional connectome and head motion strategies that minimise the exclusion of “high-motion” individuals might also boost the predictive power of functional connectomes for behavioural phenotypes using CPM.

5.1 Principles of Connectome-based Predictive Modelling

Predictive modelling, applied within the realm of neuroimaging, ultimately aims to identify brain signatures of mental health conditions that can predict clinically relevant outcomes (Tejavibulya et al., 2022; Woo et al., 2017). An increasingly popular data-driven technique for predictive modelling of functional connectivity-behaviour relationships is known as CPM (Shen et al., 2017; Rosenberg et al., 2015; Finn et al., 2015). CPM extracts statistically significant features from functional connectivity matrices (*functional connectome signatures*) that share the highest correlation strengths with a behavioural measure, to construct linear predictive models. The CPM protocol comprises the following key steps (Figure 5.1): (1) feature selection; (2) feature summarisation; (3) model building; and (4) assessment of the prediction significance. The main advantage of employing CPM is that it has an in-built cross validation (CV) procedure such as leave-one-out [LOOCV] or k -fold CV to test the generalisability of model predictions in unseen participants (see **Appendix E** for details on how CPM identifies functional connectome signatures and performs CV using either LOOCV or k -fold).

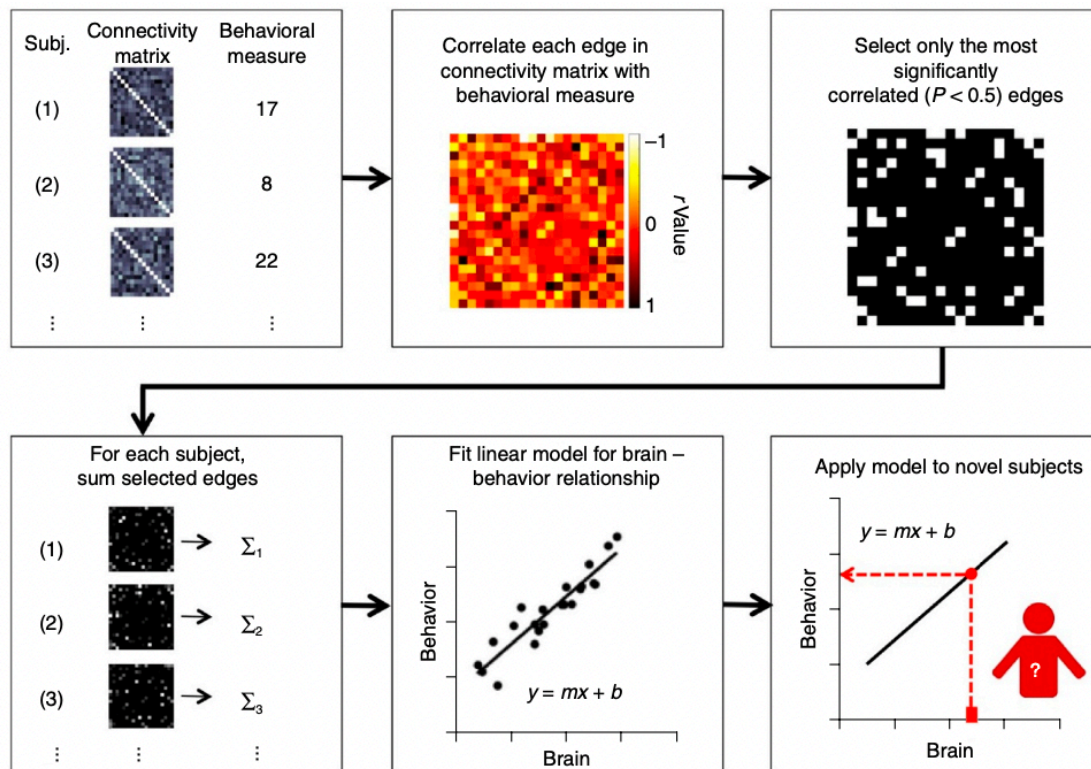


Figure 5.1. Connectome-based predictive modelling (CPM) framework which involves feature selection, feature summarisation, model building, and assessment of the prediction significance. Adapted from Shen et al., 2017.

CPM has gained increasing popularity since its inception to predict a broad range of behavioural domains using both resting-state and task-based fMRI paradigms spanning childhood to adulthood including developing and clinical populations.

5.2 Predicting Behavioural Phenotypes with Individual Functional Connectome

Using resting-state fMRI conditions, Finn et al. (2015) demonstrated that individual patterns of functional connectivity were strongly related to fluid intelligence when CPM included positive and negative edges, separately, derived from the whole-brain and frontoparietal connectomes. Such brain-behaviour relationships have been reproduced in independent studies (Cai et al., 2021; Lin et al., 2020), with high degrees of variance also explained by resting-state functional connectivity in the cingulo-opercular, default mode, and visual networks (Dubois et al., 2018b). Beyond fluid intelligence, resting-state functional connectivity has been shown to predict reliable patterns of personality traits obtained from the non-uniform contributions of the positive and negative edges of the whole brain such as openness to experience (Dubois et al., 2018a), neuroticism and extraversion (Hsu et al., 2018), narcissism (Feng et al., 2018), temperament (Jiang et al., 2018), and loneliness (Feng et al., 2019) — across different preprocessing pipelines, predictive models, and populations. CPM has also been recently applied to predict individual anxiety traits in neurotypical participants which highlights the importance of whole-brain functional connectivity, especially the limbic, prefrontal, and temporal areas, in predicting the anxiety scores (Wang et al., 2021). “Lesioning” of these brain regions, i.e., a non-invasive computational manipulation of the brain regions (Feng et al., 2018), produced a drastic reduction in the predictive power of the individual scores, which was externally validated in an independent dataset (Wang et al., 2021).

While a substantial amount of CPM research has used resting-state fMRI paradigms, deriving individual functional connectomes from task-based fMRI conditions may improve the prediction of behavioural phenotypes. Predictive models derived from task-based functional connectivity have been shown to significantly outperform those derived from rest for prediction of fluid intelligence (Greene et al., 2018a), reading comprehension skills (Jiang et al., 2020), and extending executive domains (e.g., language, inhibition, flexibility; Mantwill et al., 2022). This suggests that certain (task-evoked) brain states are more predictive of some behaviours than resting-state conditions (Greene et al., 2023). Task-based CPM has also been employed across clinical groups (e.g., schizophrenia, bipolar, ADHD) to characterise brain networks that correspond to different memory processes (Barron et al., 2018). For example, transdiagnostic individual differences in memory function were predicted by functional connectivity within networks involving default mode, limbic, subcortical, and cerebellar structures (Barron et al.,

2018). Another significant body of work has focused on predicting attentional abilities using both task-based and resting-state functional connectivity in neurotypical and neurodevelopmental populations (Chamberlain and Rosenberg, 2022; Rosenberg et al., 2020; Rosenberg et al., 2018; Rosenberg et al., 2016; Rosenberg et al., 2015). In one instance, a CPM was trained using task-based data to predict sustained attention and inhibition; its ability to predict behaviour generalised to resting-state data and to a sample of participants with ADHD (characterised by difficulties with sustained attention, hyperactivity, and impulsivity; Rosenberg et al., 2015).

Recent work has further derived CPM-based models of sustained attention that differentiate individuals with better or worse sustained attention (Rosenberg et al., 2016) and are stable over time in neurodiverse youths, including those suffering from ADHD, anxiety, autism, bipolar, and broader autism phenotype conditions (Horien et al., 2023). Sustained attention varies over both short and long timescales; functional connectivity patterns within the “low-attention” and “high-attention” networks that predict individual differences in sustained attention were found to also predict fluctuations in attention over periods ranging from minutes to days (Rosenberg et al., 2020). This CPM model was also sensitive to pharmacological intervention with methylphenidate, a psychostimulant commonly used for the treatment of ADHD. Neurotypical adults who received methylphenidate exhibited patterns of functional connectivity that are associated with better sustained attention and decreased functional connectivity in networks associated with poorer sustained attention (Rosenberg et al., 2016). The converse pattern was observed for the anaesthetic agent, propofol, compared to an awake state; participants in deep sedation (propofol condition) showed lower high-attention and higher low-attention network strengths (Chamberlain and Rosenberg, 2022). Further reinforcing their potential for the identification of clinically meaningful brain-based markers, CPM-based models have been shown to predict treatment response over time; a CPM was used to characterise positive and negative cocaine abstinence brain networks that successfully predicted cognitive and executive functioning during a 12-week treatment period in addition to a 6-month follow-up (Yip et al., 2019). Recent efforts have extended the ability of CPM to predict individual differences in behaviour using structural connectomes obtained from diffusion MRI data (Yoo et al., 2022; Feng et al., 2022; Zhang et al., 2019). For example, fractional anisotropy (FA)-derived white matter fiber tracts successfully predicted trait anxiety levels in young and older adults (Yoo et al., 2022). Finally, a negative network that forms part of the amygdala-prefrontal circuits was negatively related to heightened anxiety in young adults (20-30 years), while a distributed positive network consisting of the frontotemporal and frontolimbic circuits was positively associated with individual trait of anxiety levels across younger and older adults (20-80 years) (Yoo et al., 2022).

Functional connectome signatures (i.e., functional connectivity patterns predicting a behavioural phenotype at the individual level) have successfully been used for the prediction of *individual* behavioural phenotypes, prompting the question of whether the factors that improve functional connectome identifiability might also benefit CPM-based predictions. In other words, can boosting the uniqueness of a given individual’s functional connectome boost connectome-based predictions of behavioural phenotypes? This link is suggested based on the fact that improving the reproducibility of brain-related measures (here, functional connectivity) can enhance validity of those measures, as revealed by brain-behaviour relationships (see **Chapters 2 and 4**). The studies reviewed in **Section 5.2** show that CPM identifies functional connectome signatures that have specificity to clinically relevant cognitive functions and outcomes, and which are generalisable across samples, participant cohorts, and time. CPM therefore offers considerable promise for the search of reproducible brain-based markers in neurodevelopmental and psychiatric conditions. It is important, therefore, to investigate the methodological factors that have the ability to influence CPM, so that we can optimise it for reproducibility.

5.3 Methodological Considerations for Robust Prediction of Individual Behaviours

In **Chapter 2**, I described a processing pipeline that optimised connectome-based identification by maximising fingerprint ID accuracy — i.e., maximising an individual’s identifiability based on their functional connectome. In **Chapter 4**, I showed that bagging can help salvage “high-motion” individuals for deriving robust and reproducible brain-behaviour associations. I now identify and summarise the following set of factors that may contribute to the robustness and reproducibility of behavioural phenotypic predictions obtained from CPM — a subset of which will be investigated *empirically* in **Chapter 6**:

1. Sample Size: A large sample is required so that the dataset can be divided into training and testing sets to enable CV to be performed (see **Section 5.1 and Appendix E**). This is necessary to avoid overfitting the model and foster generalisability to novel datasets (Davatzikos, 2022; Varoquaux et al., 2017; Whelan and Garavan, 2014). Predictive models using small sample sizes ($N \approx 100$) tend to yield inflated predictions (Yeung et al., 2022; Varoquaux et al., 2018), leading to poor reproducibility in independent datasets (Dhamala et al., 2022; Botvinik-Nezer and Wager, 2022). This is because small samples are susceptible to sampling variability that produce stronger associations by chance (Marek et al., 2022; Szucs and Ioannidis, 2020; Schönbrodt and Perugini, 2013; see **Chapter 4**), while large sample sizes decrease the sampling variability and produce stable brain-behaviour relationships at a rate of $\sqrt{N}$ (Marek et al., 2022).

2. **Behavioural Phenotypes:** Many behavioural phenotypes, such as clinical diagnoses, are categorical in nature. Creating models that predict these phenotypes is a challenging task, due to the fact that many of these phenotypic categories are highly *heterogeneous* in nature, yet many behavioural signs and symptoms show significant overlap amongst diagnostic categories (Marquand et al., 2016; Insel et al., 2010). Nearly 50% of individuals who meet the criteria for one disorder are likely to meet the criteria for a co-occurring condition (Tejavibulya et al., 2022; Astle et al., 2022; Forbes et al., 2016; Insel et al., 2010). One approach to this challenge is to identify *subtypes* of neurodevelopmental and psychiatric disorders such as ASD (Qi et al., 2020; Easson et al., 2019), ADHD (Fair et al., 2012), and depression (Drysdale et al., 2017; Dhamala et al., 2022). Normative modelling has also gained popularity to parse the heterogeneity in these conditions and model the biological variation within the samples (Marquand et al., 2016). Another is to focus on behavioural dimensions that are relevant across diagnoses, and rather than treating categories as isolated disorders, to identify shared variances arising from a common latent construct (Astle et al., 2022; Forbes et al., 2016). The RDoC defined by the NIMH were introduced to foster this notion of treating mental health conditions as transdiagnostic or dimensional (Insel, 2014; Cuthbert and Insel, 2013; Insel et al., 2010; see **Chapter 1 (Section 1.3)**). A good example of the successful implementation of this approach is the work on predictive models of sustained attention, as reviewed in **Section 5.2** (Rosenberg et al., 2020; Rosenberg et al., 2018; Rosenberg et al., 2016; Rosenberg et al., 2015). Such approaches benefit from samples that include individuals who represent different points on a behavioural or biological spectrum of relevance — from healthy, through subclinical, to clinical severity (Tejavibulya et al., 2022).
3. **Head Motion:** Predictive models are vulnerable to factors that introduce biased relationships between potential confounds and target variables (Rao et al., 2017). For example, physiological artefacts in the form of head motion, respiration, and cardiac movements could be correlated with disease symptom severity such that individuals with heightened head motion exhibit worse symptoms (see **Chapter 3**). Failure to control for confounds can severely impair data quality and model interpretability (Yeung et al., 2022; Tejavibulya et al., 2022); analyses using CPM should carefully mitigate the effects of head motion using correction strategies such as GSR and scrubbing (see **Chapters 1 and 3**). Although scrubbing can yield small improvements in differentiating an individual's functional connectome (Pham et al., 2023), its impact on CPM remains to be investigated. One study showed that task-based CPM better predicts cognitive abilities such as reading comprehension skills with longer fMRI timecourses as a function of $\frac{1}{\sqrt{\text{imaging time}}}$ (Jiang et al., 2020). Since scrubbing decreases the

length of timeseries (see **Chapter 4**), it may negatively impact CPM — a factor which I consider in **Chapter 6**.

4. Task-Induced versus Resting-State Connectomes: CPM has traditionally employed functional connectivity matrices derived from resting-state conditions to predict a range of behavioural phenotypes (see **Section 5.2**). Emerging evidence has strongly suggested that task-induced paradigms improve the detection of individual differences in the functional connectome and their underlying brain-behaviour predictions (Jiang et al., 2020; Greene et al., 2018a; Finn et al., 2017; Gratton et al., 2016). While some task conditions have been shown to exhibit higher predictive power compared to others, combining different task-based connectomes in a single predictive model yield better predictions than using individual conditions in isolation (Jiang et al., 2020; Gao et al., 2019; Greene et al., 2018a). However, this observation does not necessarily mean that integrating more task-based conditions leads to better predictive performances (Jiang et al., 2020). This is because task conditions that are more predictive by themselves contribute more edges to the predictive model compared to the less predictive tasks (Gao et al., 2019). On the other hand, CPM derived from task conditions such as reading comprehension can generalise across several other fMRI conditions (e.g., rest, emotion, working memory) to predict language related scores (Jiang et al., 2020). This shows that the CPMs which are built using activation patterns that are relevant to cognitive domains are generally predictive of other domains.
5. Parcellation Schemes: In order to predict a behavioural phenotype from individual patterns of functional connectivity, the connectivity matrices are typically derived from a parcellation scheme (Shen et al., 2017). While a range of brain parcellations has been applied for this purpose, the Shen ($k = 268$) parcellation remains the most popular as it is derived from resting-state fMRI data that covers the whole brain with a reasonable number of regions (Yeung et al., 2022; Shen et al., 2013). Over the years, brain parcellations have been derived from anatomical, functional, or multimodal techniques to define nodes that form part of either the cortex or whole-brain (i.e., cortex and subcortex including the cerebellum) (Yeung et al., 2022). While a number of brain parcellations have been shown to yield consistent predictions of cognitive measures (Mantwill et al., 2022), they can also produce significantly different results (Bryce et al., 2019; Dubois et al., 2018a). As described in **Chapters 2**, parcellation schemes differ in parcel resolution (e.g., $7 \rightarrow 1095$ parcels) — higher-resolution parcellations capture finer properties of the functional connectome at the individual level. However, there is a trade-off: brain parcellations with increasing dimensionality do not necessarily improve behavioural predictions (Dadi et al., 2019; Greene et al., 2018a), and can make predictive modelling highly computationally intensive, variable, and less interpretable (Tejavibulya et

al., 2022; Dhamala et al., 2022). Although the majority of parcellation schemes have been defined at the group-level, the emergence of individualised brain parcellations has allowed the detection of specific biologically plausible features that could increase the strength of functional connectivity estimates, thus improving model predictions (Kong et al., 2021; Salehi et al., 2020).

6. **Model Building Strategies:** Unsurprisingly, CPM is affected by methodological choices related to functional connectivity definition, feature selection, and predictive algorithm parameterisation (Dhamala et al., 2022; Dadi et al., 2019; Shen et al., 2017). Functional connectivity is commonly estimated using Pearson’s correlation coefficient, but other methods include: *geodesic distance* — non-Euclidean geometry of functional connectivity matrices (Abbas et al., 2021; Venkatesh et al., 2020); *accordance* — in-phase synchronisation of two fMRI signals (Yoo et al., 2018; Meskaldji et al., 2016); and *discordance* — out-of-phase anticorrelation of two fMRI signals (Yoo et al., 2018; Meskaldji et al., 2016). Similarly, other forms of regression models have been employed in the CPM. Partial Least Squares (PLS) regression is a predictive technique that combines the features of PCA and multiple linear regression to obtain a linear relationship between the predictor and response variables. PLS regression is beneficial when the predictor variables (edges) are significantly larger than the number of observations (participants) such that fewer components are required to explain the covariance between the predictor and response variables. Resting-state and task-based CPMs derived from PLS regression have been shown to predict similar, if not, better sustained attention abilities than the traditional linear regression models (Yoo et al., 2018). Canonical Correlation Analyses (CCA), which is a technique that compares a set of fMRI variables (e.g., spatial maps, network matrices) to a set of behavioural variables based on the assumption that they are maximally correlated, has further been adapted to combine multiple task-induced functional connectomes into a single CPM — to maximise the correlation between the multidimensional patterns of functional connectivity and behavioural measures (e.g., fluid intelligence) (Gao et al., 2019). Other regression models have also been incorporated into a CPM to derive a significant number of edges from several task-based connectomes to predict fluid intelligence through means of ℓ_1 [LASSO] and ℓ_2 [Ridge] regularisation (Gao et al., 2019). Specifically, ridge regression which reduces model overfitting, complexity, and multicollinearity, has been shown to improve the prediction of fluid intelligence compared to other regression models (Gao et al., 2019).
7. **Discriminatory Networks and Edges:** As described in **Chapters 1 (Section 1.8) and 2**, medial frontal and frontoparietal networks tend to exhibit the highest discriminatory power for identifying an individual based on their functional connectivity matrices. Could restricting the

functional connectivity matrices to the most discriminatory functional networks and edges improve predictive performances? By improving the unique functional connectivity features of an individual, they may boost the detection of individual differences in behavioural phenotypes. The evidence to date provides mixed support for this notion. Predictive models consisting of discriminatory edges that consistently relate to behaviour successfully predict sustained attention in neurodiverse samples (Horien et al., 2023). However, building a CPM with discriminative networks and edges to predict a range of psychometric measures such as fluid intelligence, language comprehension, and grip strength has been unsuccessful (Mantwill et al., 2022). It remains to be seen whether highly discriminatory connectivity features show better predictive power for behavioural phenotypes.

5.4 Towards Reproducible CPM-based Predictions

To formulate the motivations for **Chapter 6**, I reinforce the *empirical* findings of **Chapters 2 and 4**: (1) connectome-based fingerprint identification serves as a testbed for reproducible functional connectivity estimates; and (2) bagging enables the retention of “high-motion” individuals. In **Chapter 6**, I first aimed to assess whether improving the identifiability of functional connectome fingerprints also boosts CPM-based predictions by applying the fingerprint-informed pipeline identified in **Chapter 2**. Second, I aimed to examine how head motion mitigation strategies (i.e., scrubbing, bagging), which maximise sample inclusivity in a dataset by salvaging “high-motion” participants as described in **Chapter 4**, influence the CPM-based predictions. In both cases, I also investigated the impact on the generalisability of CPM-based predictions to an independent dataset.

Chapter 6

What methodological factors can improve the reproducibility of connectome-based predictive modelling in developing youths?

Paper (In Preparation)

Ramduny, J., Weng, Y., Whelan, R., Vanderwal, T., & Kelly, C. What methodological factors can improve the reproducibility of connectome-based predictive modelling in developing youths?

Authors

Jivesh Ramduny^{1,3}, Yihe Weng^{1,3}, Robert Whelan^{1,2,3,4}, Tamara Vanderwal^{5,6}, Clare Kelly^{1,2,3}

¹School of Psychology, Trinity College Dublin, Dublin, Ireland

²Department of Psychiatry, School of Medicine, Trinity College Dublin, Dublin, Ireland

³Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland

⁴Global Brain Health Institute, Trinity College Dublin, Dublin, Ireland

⁵Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada

⁶BC Children's Hospital Research Institute, Vancouver, BC, Canada

Authorship Contribution Statement

Jivesh Ramduny: Conceptualisation, Methodology, Formal analysis, Investigation, Writing — original draft, review & editing. **Yihe Weng:** Methodology, Writing — original draft, review & editing. **Robert Whelan:** Methodology, Writing — original draft, review & editing. **Tamara Vanderwal:** Conceptualisation, Investigation, Writing — original draft, review & editing. **Clare Kelly:** Conceptualisation, Funding, Methodology, Formal analysis, Investigation, Writing — original draft, review & editing, Supervision, Project administration.

Abstract

This study extends the work of **Chapters 2 and 4**, which focused on improving the differentiation of an individual's functional connectome, by assessing whether boosting the reproducibility of the functional connectome translates to better brain-behaviour predictions in developing youths. Using the connectome-based predictive modelling (CPM) approach and the HBN dataset, I examined whether the data pre-/post-processing factors and head motion strategies previously identified can improve the predictions of age, general psychopathology, and intellectual ability. Connectome-based fingerprints were computed using the most discriminatory functional networks, multi-resolution parcellation schemes, and head motion mitigation strategies such as scrubbing and bagging. CPM was then performed using a 10-fold CV to predict three phenotypes: age, FSIQ, and externalising behaviours (per the CBCL). The generalisability of these predictions was quantified using an independent sample of non-Autistic individuals from the ABIDE dataset. Baseline CPM successfully predicted the three phenotypes. Unexpectedly, most data configurations had minimal impact on the CPM-based predictions. The CPM-based predictions of age and intellectual ability were generalisable to the ABIDE datasets, with finer brain parcellations and bagging achieving better generalisability. Taken together, these findings paint a mixed picture — while several factors and strategies do not necessarily improve validity (i.e., strengths of brain-behaviour relationships), they do have the ability to strengthen reproducibility (i.e., generalisability of brain-behaviour relationships).

6.1 Introduction

Resting-state fMRI has provided an exciting opportunity to develop functional connectivity measures that reveal associations with a wide range of behavioural phenotypes as a means to accelerate the search for functional connectivity-based biomarkers. To achieve biomarkers that are clinically viable, functional connectivity measures must be reproducible. In my previous work, I focused on improving functional connectivity estimates as a route to boost the reproducibility of functional connectivity-based biomarkers using functional connectome fingerprinting (Finn et al., 2015). I demonstrated that the connectome-based identification procedure can serve as a testbed for methods to increase the reproducibility of functional connectivity estimates, and used this approach to describe two fingerprint-informed pipelines to improve the differentiation of the individual functional connectome spanning childhood to adulthood including developing populations (see **Chapter 2**). By improving the ability to identify an individual from their functional connectivity profiles, the unique properties of their functional connectome are enhanced — I suggest this could translate to better detection of individual differences in their behaviour. To test this hypothesis, I examined whether improving the identifiability of the functional connectome boosts brain-behaviour predictions in developing youths.

Individual differences in functional connectivity, which are unique and stable like a “fingerprint”, have been shown to predict individual differences in a broad spectrum of phenotypes using CPM (Shen et al., 2017; Rosenberg et al., 2015; Finn et al., 2015; see **Chapter 5** for details about the conceptualisation of CPM). As described in **Chapter 5 (Section 5.2)**, a notable body of work employing CPM has applied resting-state and task-based fMRI to successfully predict individual phenotypes across several domains such as sustained attention (Horien et al., 2023; Chamberlain and Rosenberg, 2022; Rosenberg et al., 2020; Rosenberg et al., 2018; Rosenberg et al., 2016; Rosenberg et al., 2015), fluid intelligence (Cai et al., 2021; Lin et al., 2020; Dubois et al., 2018b; Finn et al., 2015), personality traits (Jiang et al., 2018; Hsu et al., 2018; Feng et al., 2019; Feng et al., 2018; Dubois et al., 2018a), executive functions (Mantwill et al., 2022; Greene et al., 2018a), social skills (Lake et al., 2019), cognitive abilities (Ju et al., 2023; Boyle et al., 2023; Jiang et al., 2020; Beaty et al., 2018; Barron et al., 2018), and more recently, suicide risk (Gao et al., 2023).

In **Chapter 2**, I identified a fingerprint-informed pipeline comprising a set of data pre-/post-processing factors that maximised connectome-based identification in five independent datasets ($N = 264$, 7-43 years). For example, I found that defining the individual functional connectome using the most discriminatory brain networks (i.e., medial frontal and frontoparietal networks),

fine-grained parcellation schemes ($7 \rightarrow 1095$ parcels), and cortical regions compared to subcortical and cerebellar structures increased the identifiability of the functional connectome up to 94% and minimised the degree of “confusion” for misidentifying an individual up to 2%. In **Chapter 3**, I addressed the impact of head motion and the fact that the main approach to head motion mitigation can lead to a large number of participants being discarded in order to maintain good data quality, which can have the unintended knock-on consequence of reducing statistical power as a result of reduced sample size. To address this, and to strike a better balance between data quality and sample size, I previously proposed the implementation of “bagging” as a means to retain the “high-motion” individuals who would otherwise be discarded (Power et al., 2015; Power et al., 2014; Power et al., 2013; Dudoit and Fridlyand, 2003; Opitz and Maclin, 1999; Breiman, 1996; see **Chapter 4**). I showed that bagging can enable the salvage of data from “high-motion” developing youths ($N = 423$, 6-21 years), thus maximising sample size for deriving inclusive and reproducible brain-behaviour associations (see **Chapter 4**). However, how bagging affects connectome-based phenotypic predictions has not yet been investigated.

The goal of the present study is to examine whether optimising connectome-based fingerprint identification for the reliable measurement of individual differences in functional connectivity also improves the sensitivity of connectome-based phenotype predictions in developing youths. To achieve this, I first build on the work of **Chapter 2** to assess whether the same data pre-/post-processing factors that maximised connectome-based identification also improve CPM-based predictions of three phenotypes — i.e., age, general psychopathology, and intellectual ability — in developing youths derived from the HBN (Alexander et al., 2017). Second, I examine whether the bagging approach tested in **Chapter 4** affects CPM-based predictions. In each case, I also assess the impact of these factors using a gold-standard test — the generalisability of the CPM-based predictions of the phenotypes to an independent sample (i.e., non-Autistic individuals from ABIDE; Di Martino et al., 2014a).

6.2 Materials & Methods

6.2.1 Participants

Data were obtained from the HBN (http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/), which is a community-based help-seeking sample of children and adolescents from the Greater New York area (Alexander et al., 2017). Neuroimaging and phenotypic data were obtained from the Cornell Brain Imaging Center (CBIC) site in Releases 4-10. Approval was obtained from the Chesapeake Institutional Review Board; written consent was provided by the participants who were at least 18 years old, while written consent was obtained from legal guardians and from the participants who were under 18 years. For each participant, at least one

anatomical and two resting-state fMRI scans were obtained from the same imaging session (i.e., within-session scans). Quality control was performed on all the scans to check for potential scan artefacts in addition to ensuring that each participant had two runs of resting-state fMRI data as well as their phenotypic data. A total of 540 participants (198 females, age range = 6-21 years; mean age $\pm$ SD = 10.7 $\pm$ 3.0 years) were available for this study. The demographic information and resting-state fMRI parameters of the HBN are provided in Table 6.1.

Demographic & fMRI Characteristics	AGE	CBCL	FSIQ
Sample Size (N)	540	516	479
Range	6-21	33-88	47-147
Gender (Female/Male)	198/342	187/329	178/301
Scanning Site	CBIC	CBIC	CBIC
Scan Duration (min)	5	5	5
Time between Scans	Within-Session	Within-Session	Within-Session
Scanner Manufacturer	Siemens	Siemens	Siemens
Scanner Model	Prisma	Prisma	Prisma
Field Strength	3.0T	3.0T	3.0T
TR (ms)	800	800	800
Volumes	375	375	375
REST1 RMS FD (mm)	0.03-2.09	0.03-2.09	0.03-2.09
REST2 RMS FD (mm)	0.03-2.89	0.03-2.89	0.04-2.89

Table 6.1. Demographic characteristics and resting-state fMRI acquisition parameters of the HBN that were used in this study. CBIC = Cornell Brain Imaging Center. RMS FD = root-mean-square framewise displacement. Note: The CBCL scores derived from the HBN reflect externalising behaviour that have been T-standardised.

6.2.2 Phenotypic assessment of general psychopathology and intellectual ability

6.2.2.1 General Psychopathology

The Child Behaviour Checklist (CBCL) was administered to assess general psychopathology linked to behavioural and emotional problems in children and adolescence (Achenbach and Rescoral, 2001). The CBCL is a 113-item parent report measure comprising eight subscales assessing a range of internalising and externalising behaviours. For each participant, subscale scores are T-standardised, and a total score, an internalising score, and an externalising score are computed. The higher the CBCL T-score, the greater the risk of experiencing emotional and behavioural problems. While externalising behaviours tend to reflect behaviours related to conditions like ADHD and oppositionality (e.g., impulsivity, aggression, disruption), internalising behaviours capture mood-related behaviours (e.g., anxiety, withdrawal, dysphoria)

(Achenbach and Edelbrock, 1978). Because the HBN is a community sample of psychiatric help-seeking individuals and because these individuals are very young, externalising behaviours are likely to be well represented in this sample. For this study, I focused on the externalising subscale to derive the CBCL T-scores. The distribution of the CBCL T-scores is illustrated in Figure 6.1.

6.2.2.2 Intellectual Ability

A full-scale intelligence quotient (FSIQ) score was obtained using the Wechsler Intelligence Scale for Children (WISC-V), administered for participants aged between 6 and 17 years old (Wechsler, 2003). For participants who are aged 18 years and older, the Wechsler Adult Intelligence Scale (WAIS-IV) was administered (Wechsler, 2008). For both WISC-V and WAIS-IV, the higher the FSIQ score, the greater the intellectual ability of the participants. The distribution of the FSIQ scores is illustrated in Figure 6.1.

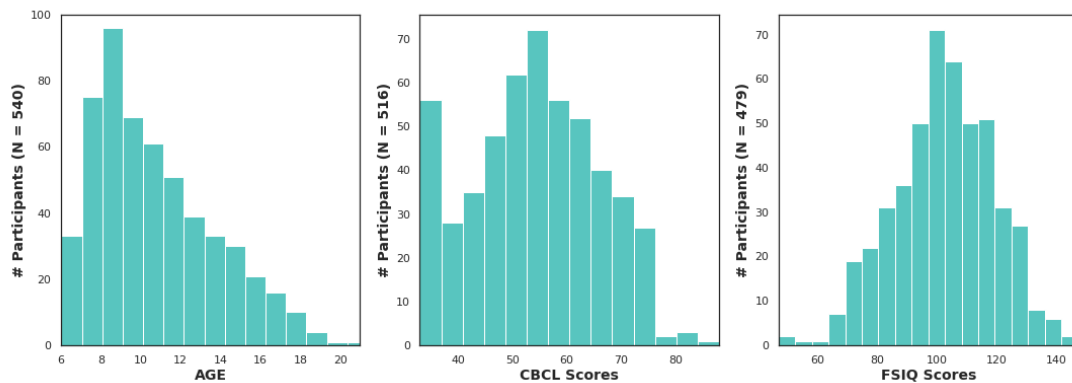


Figure 6.1. Phenotypic distributions across the complete HBN dataset. Histograms showing the distributions of age [in years], general psychopathology [CBCL], and intellectual ability [FSIQ] in developing youths.

6.2.3 MRI Data Acquisition

Structural T1-weighted and functional MRI scans were acquired at the CBIC using a Siemens 3T Prisma scanner with a 32-channel head coil. A 3D T1-weighted MPRAGE volume with spatial resolution $0.8 \times 0.8 \times 0.8 \text{ mm}^3$ was obtained for each participant (224 sagittal slices, TR = 2500 ms, TE = 3.15 ms, flip angle = 8°). The two runs of resting-state fMRI data were acquired with echo-planar imaging sequences (60 slices, voxel size = $2.4 \times 2.4 \times 2.4 \text{ mm}^3$, multiband factor = 6, TR = 800 ms, TE = 30 ms, flip angle = 31°). Each participant underwent two 5 min functional MRI scans while they fixated a cross on the screen; both runs consisted of 375 volumes.

6.2.4 MRI Preprocessing Pipeline

The structural T1-weighted and functional data derived from the HBN were preprocessed using a previously defined pipeline (Yan et al., 2013a; Yan et al., 2013b; Craddock et al., 2013c; see Chapters 2 (Section 2.2.2) and 4 (Section 4.2.3)) which includes AFNI (<https://afni.nimh>).

[nih.gov](https://www.nih.gov)), FSL (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL>), and ANTs (<https://github.com/ANTsX/ANTs>). Participants exhibiting mean rmsFD > 0.20 mm (Jenkinson et al., 2002) in either fMRI scan were initially excluded. A total of 306 participants (~57% of the original sample size) remained in the cohort for the subsequent analyses. There was a significant relationship between mean rmsFD and age ($r = -0.28$, $P < 0.001$) but not for CBCL ($r = 0.02$, $P = 0.78$) and FSIQ ($r = -0.11$, $P = 0.08$).

6.2.5 Functional Connectivity Matrices

A whole-brain parcellation was used to compute the functional connectivity matrices of each participant within the same imaging session (<https://identifiers.org/neurovault.image:395091>; Shen et al., 2013). For each participant, the mean fMRI timeseries of each ROI was extracted prior to computing the Pearson's correlation coefficient (r) between all possible ROI pairs to construct a 268 x 268 functional connectivity matrix. This procedure was repeated for the two functional runs such that each participant had two functional connectivity matrices as denoted by REST1 and REST2. Some edges in the functional connectivity matrices were eliminated for each participant to account for the lack of spatial coverage (i.e., zero values) across the whole brain. I also removed duplicate edges by considering only the upper triangular part of the functional connectivity matrices. From the $[(268 \times 268) - 268] / 2 = 35,778$ edges in the whole-brain functional connectome, there were 25,200 distinct edges that remained in the REST1 and REST2 functional connectivity matrices.

6.2.6 Connectome-based Fingerprinting

The fingerprint identification procedure was performed as previously described (Finn et al., 2015; see **Chapters 1 (Section 1.8), 2, and 4**). Baseline identification was carried out with the 268 x 268 REST1 and REST2 functional connectivity matrices for each participant. The mean baseline binary identification (BID) accuracy of the HBN was calculated by exchanging the roles of the database-target matrix ($\mathbf{D} \leftrightarrow \mathbf{T}$) configurations and averaging the two BID accuracies.

6.2.7 Connectome-based Fingerprinting Pipeline

To maximise the connectome-based identification as a means to improve the reproducibility of functional connectivity estimates, two data pre-/post-processing factors and two head motion mitigation strategies were applied to derive the REST1 and REST2 functional connectivity matrices as described below (see Figure 6.2 for details):

1. Parcellation Schemes: Parcellation schemes differ in terms of their dimensionality, space, and spatial coverage. Given that finer brain parcellations tend to yield the highest identification

accuracy in developing youths (see **Chapter 2**), the resting-state fMRI data were parcellated with four publicly available brain atlases including Glasser 360 (Glasser et al., 2016), Glasser Extended 426 (Huang et al., 2021), MIST 444 (Urchs et al., 2017), and Schaefer 500-17Networks (Schaefer et al., 2018). To reduce the increasing computational resources of applying CPM with greater parcel resolutions, the maximum resolution was restricted to the Schaefer 500-17Networks (e.g., 124,750 edges) as opposed to the Schaefer 1000-17Networks (499,500 edges) described in **Chapter 2**. The properties of the four parcellation schemes are elaborated in Table 6.2.

Schemes	Dimensionality	Edges	Space	Coverage
*Shen	268	35,778	Volume	Whole-brain
Glasser	360	64,620	Surface	Cortical
Glasser Extended	426	90,525	Volume	Whole-brain
MIST	444	98,346	Volume	Whole-brain
Schaefer 17Networks	500	124,750	Surface	Cortical

Table 6.2. Properties of the parcellation schemes that were employed to derive the REST1 and REST2 functional connectivity matrices for the connectome-based identification procedure. The parcellation schemes differ in parcel dimensionality (i.e., 268 → 500 parcels), space (i.e., volume, surface), and spatial coverage (i.e., cortex, whole-brain). *Baseline connectome-based fingerprint identification was performed using the whole-brain Shen 268 parcellation scheme.

2. **Frontoparietal Networks:** Functional networks have consistently shown to yield the highest identification accuracy with the combination of the medial frontal and frontoparietal networks in neurotypical and developing populations compared to the whole-brain connectome (Vanderwal et al., 2021; Cai et al., 2021; Jalbrzikowski et al., 2019; Horien et al., 2019; Byrge and Kennedy, 2019; Vanderwal et al., 2017; Finn et al., 2015; **see Chapters 1 (Section 1.8) and 2**). The frontoparietal networks were identified using a 10-network definition (Horien et al., 2023; Horien et al., 2019; Noble et al., 2017; Finn et al., 2015) to construct a 63 x 63 REST1 and REST2 functional connectivity matrix for each participant.
3. **Scrubbing:** Head motion strategies such as scrubbing have been shown to minimise motion artefacts in the fMRI timeseries as a means to maximise data quality (Power et al., 2015; Power et al., 2014; Power et al., 2013; **see Chapters 3 (Section 3.3) and 4**). I performed scrubbing as previously described by censoring motion-corrupted timepoints (T) if their rmsFD > 0.20 mm in addition to their neighbouring timepoints (T - 1, T + 1, T + 2). The

scrubbed timeseries were then used to derive the REST1 and REST2 functional connectivity matrices for each participant. To maximise participant inclusion for performing connectome-based identification, a rmsFD threshold set at 0.20 mm was applied as opposed to more strict thresholds (e.g., rmsFD > 0.08 mm) which increase the risk of excluding a larger number of participants in the analyses (Marek et al., 2022; see **Chapter 3 (Section 3.3)** for a discussion).

4. **Bagging:** Given that bagging has been shown to have the ability to successfully identify an individual's functional connectome based on their scrubbed timeseries and maximise sample inclusivity by salvaging “high-motion” participants (see **Chapter 4**), I applied bagging to compute the REST1 and REST2 functional connectivity matrices of the participants. Bagging was first performed on the “low-motion” participants ($N = 306$) after applying the initial rmsFD threshold, set at 0.20 mm. For each participant, and each scan (REST1, REST2), their scrubbed timepoints were ranked by rmsFD. From the $\text{minTP} \in \{60, 200\}$ timepoints exhibiting the least motion, 500 bootstrapped timeseries volumes of size $\text{TP} \in \{20, 100\}$ were randomly selected *with* replacement. The two parameters were tuned to be consistent with the values applied in **Chapter 4**. At each iteration, functional connectivity was computed between all pairs of ROIs, creating 500 bootstrapped REST1 and REST2 functional connectivity matrices. For each participant, these matrices were then averaged, creating a single mean REST1 and REST2 functional connectivity matrix. The bagging procedure was repeated with the original sample size ($N = 540$) *without* discarding participants whose rmsFD > 0.20 mm, excluding only those participants who did not have minTP timepoints remaining after scrubbing.

6.2.8 Connectome-based Predictive Modelling

CPM was carried out as previously described (Green et al., 2018a; Shen et al., 2017; Rosenberg et al., 2015; Finn et al., 2015; see **Chapter 5**) to predict age, CBCL, and FSIQ using a k -fold CV, where $k = 10$. The three phenotypes were chosen as they differ in their measurements: age (*true biological* measure), CBCL (*self-reported, subjective* measure), and FSIQ (*task performance* measure). The strength of each edge from the whole-brain connectome in the $k - 1$ samples was correlated with each of the three phenotypes across 100 iterations. Edges that shared the strongest *positive* correlations with each of the phenotypes and were 95% relevant in all of the $k - 1$ samples were assigned to the positive [POS] network whereas those that exhibited the strongest *negative* correlations with the phenotypes were assigned to the negative [NEG] network. The significance of the correlations between each edge's strength and the phenotypes was thresholded at $P < 0.01$. The POS and NEG networks of each participant in the training set were derived as follows:

$$\text{POS} = \sum_{i,j} \text{FC}_{i,j} m_{i,j}^+ \quad (6.1)$$

$$\text{NEG} = \sum_{i,j} \text{FC}_{i,j} m_{i,j}^- \quad (6.2)$$

where FC is the vectorised connectivity matrix of a participant obtained from REST1, and m^+ and m^- are the binary matrices for a given pair of edges i and j that survived the significance threshold for the respective POS and NEG networks (Greene et al., 2018a). In addition, a Combined network can be derived by computing the difference between the POS and NEG network strengths. Linear regression models were built to assess the relationship between the three network strengths and each of the phenotypes in the $k - 1$ samples. When age was predicted, the regression models covaried for sex and rmsFD derived from REST1. When CBCL and FSIQ were predicted, the regression models included age, sex, and rmsFD as covariates. The three network strengths were then calculated for the (unseen) k samples to predict the phenotypes and this procedure was repeated k times to allow each k sample to serve as the testing set once. Permutation testing was performed to compute the significance of the model predictions by randomly shuffling the participant labels across 1000 iterations, thus deriving the number of times the permuted (null) predictions were greater than or equal to the true (unpermuted) predictions:

$$P = \frac{\#\{r_{\text{null}} > r_{\text{true}}\}}{1000} \quad (6.3)$$

The significance of the model predictions was further corrected for multiple comparisons using FDR. Model performances were quantified as the relationships between the predicted and observed phenotypes using Pearson's correlation coefficient (r), explained variance (R^2), and mean absolute error (MAE). CPM was repeated to test whether the data pre-/post-processing factors and head motion mitigation strategies would improve the baseline relationships between the predicted and observed phenotypes. Finally, a Steiger's Z test was performed to assess whether the CPM-based predictions differed significantly from baseline for each of the factors examined (Horien et al., 2023; Beaty et al., 2018; Steiger, 1980). Of note, a Steiger's Z test cannot be performed to compare the baseline CPM performances with those derived from bagging as the sample sizes are not equal: the assumption of the test relies on computing the statistical significance of the difference between two dependent correlations with the same sample size.

6.2.9 External Validation of the CPM-based Predictions

To test generalisability, the POS, NEG, and Combined networks obtained from the HBN CPM-based predictions of age and FSIQ were applied to an independent dataset (Horien et al., 2023; Greene et al., 2018a; Rosenberg et al., 2016). Given that the CBCL T-scores were not available for ABIDE, the generalisability focused on predicting age and FSIQ. The network strengths were first used to derive the model coefficients in the independent sample. The network masks and model coefficients were then used to predict age and FSIQ. The independent dataset included non-Autistic individuals from ABIDE (84 females, $N = 439$, age range = 6-57 years, mean age $\pm$ SD = 16.2 $\pm$ 6.7 years). The preprocessed resting-state fMRI data were obtained from C-PAC which included GSR (<http://preprocessed-connectomes-project.org/abide/>; Craddock et al., 2013a). After applying an initial head motion threshold of rmsFD < 0.20 mm, a total of 407 participants remained in the cohort for the CPM analyses. Model performances were quantified for the relationships between the predicted and observed phenotypes using Pearson's correlation coefficient (r), explained variance (R^2), and mean absolute error (MAE). A Steiger's Z test was recomputed to assess whether the CPM-based predictions differed significantly from baseline for each of the factors (excluding bagging) examined.

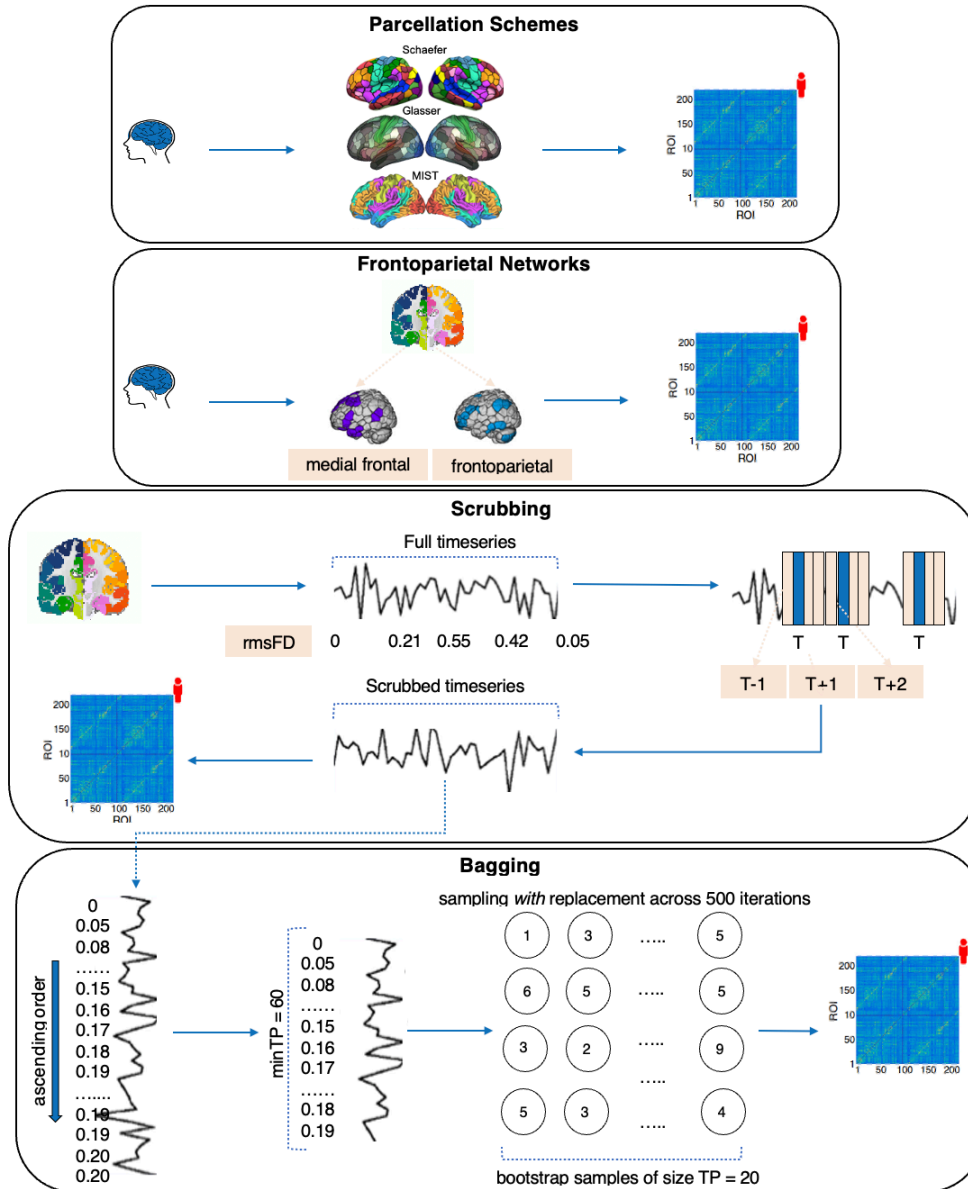


Figure 6.2. Data pre-/post-processing factors and head motion mitigation strategies for connectome-based fingerprint identification. Data pre-/post-processing factors include the application of four publicly available parcellation schemes such as the Glasser 360, Glasser 426, MIST 444, and Schaefer 500-17Networks atlases in addition to using the frontoparietal networks derived from a 10-network definition to derive the REST1 and REST2 functional connectivity matrices of the participants. Head motion mitigation strategies include scrubbing and bagging to compute the REST1 and REST2 functional connectivity matrices of the participants on the basis of their least motion-corrupted timepoints in the fMRI timeseries.

6.3 Results

6.3.1 Connectome-based identifiability of developing youths

Connectome-based identification was performed for the HBN by applying a standard preprocessing pipeline including GSR followed by the Shen 268 parcellation scheme which consisted of 25,200 distinct edges in the whole-brain connectome. Prior to deriving the baseline

connectome-based identification, an initial head motion threshold of $\text{rmsFD} > 0.20$ mm was employed to identify the “high-movement” participants in the cohort. Using REST1 and REST2 scans acquired within the same imaging session, binary identification was carried out such that the BID accuracy is 1 if the participants are successfully identified and 0 if they are misidentified with other individuals in the cohort. For the **D** $\rightarrow$ **T** configuration, the BID accuracy was 76% and for the **T** $\rightarrow$ **D** configuration, the BID accuracy increased to 80%. On average, the developing youths in the HBN ($N = 306$, age range = 6-21 years, mean age $\pm$ SD = 11.6 $\pm$ 3.1 years) had an identification accuracy of 78% based on their unique whole-brain functional connectome profiles.

6.3.2 Factors influencing connectome-based fingerprint identification in developing youths

Data Pre-/Post-Processing Factor #1: Frontoparietal Networks

I first computed functional connectivity for the medial frontal and frontoparietal networks using a 10-network definition to construct a 63 x 63 REST1 and REST2 functional connectivity matrix for each participant instead of using the whole-brain connectome. Restricting the fingerprint to these networks alone, accounting for $\sim 25\%$ of the whole-brain connectome, decreased the BID accuracy to 71% (Figure 6.3A). Contrary to the findings in **Chapter 2**, the functional networks that typically share the highest discriminatory power exhibited a drop of 9.0% in their ability to identify developing youths based on their distinct functional connectivity patterns compared to the baseline whole-brain connectome identification.

Data Pre-/Post-Processing Factor #2: Parcellation Schemes

I then examined how different parcellation schemes that differ in their parcel dimensionality, space (e.g., volume vs. surface), and spatial coverage influence the identifiability of the developing youths based on their functional connectomes. Similar to the findings in **Chapter 2**, there was a tendency to observe greater BID accuracies in the HBN with finer parcellation schemes (i.e., 268 $\rightarrow$ 500 parcels), especially with those covering cortical regions only, rather than whole-brain. When the functional connectome was defined using the whole-brain Shen 268 parcellation scheme, identification accuracy was 78% but improved to 87% when the cortical Schaefer 500-17Networks scheme was employed (Figure 6.3B). That is, increasing parcel dimensionality ~ 2 times resulted in a gain of 11.5% in the ability to differentiate developing youths from the cohort based on their individual functional connectomes.

Head Motion Mitigation Strategy #1: Scrubbing

As described in **Chapter 4**, a temporal censoring strategy commonly known as scrubbing was applied to identify and remove each timepoint (**T**) including one preceding (**T** - 1) and two succeeding (**T** + 1, **T** + 2) timepoints that are corrupted by head motion (i.e., $\text{rmsFD} > 0.20$ mm) prior to computing the REST1 and REST2 functional connectivity matrices of the participants.

After scrubbing, BID accuracy decreased to 77% (Figure 6.4A) — a marginal drop of 1.3% compared to when the full timeseries of the participants were used to construct the REST1 and REST2 functional connectivity matrices.

Head Motion Mitigation Strategy #2: Bagging

Bagging was performed using the scrubbed timeseries as described in **Chapter 4**. First, I set $\text{minTP} = 200$ and $\text{TP} = 100$, i.e., I selected the 200 least motion-contaminated timepoints from each individual's scrubbed timeseries for REST1 and REST2, then bagged 100 timepoints, computed functional connectivity matrices for each of 500 iterations, and averaged across iterations to produce a mean functional connectivity matrix for REST1 and REST2 for each participant. For this combination of parameters, which resulted in the retention of just over 50% of the full timeseries, BID accuracy decreased to 69% (Figure 6.4B). In the second instance, I set $\text{minTP} = 60$ and $\text{TP} = 20$, i.e., I identified the 60 least motion-contaminated timepoints from the scrubbed timeseries, from which only 20 timepoints were bagged across 500 iterations. This combination of parameters resulted in the retention of just 16% of the full timeseries, and BID accuracy further decreased to 58% (Figure 6.4B). The decrease in fingerprint ID accuracy is consistent with the results obtained in **Chapter 4** when bagging was performed with the $\text{minTP} = 200/\text{TP} = 100$ and $\text{minTP} = 60/\text{TP} = 20$ thresholds.

The bagging procedure was repeated to salvage “high-motion” participants in the HBN cohort, i.e., those initially excluded using a head motion threshold of $\text{rmsFD} \leq 0.20$ mm. Participants from the original $N = 540$ sample were retained provided that they had enough timepoints remained after scrubbing to meet the minTP thresholds. I retained ~60% ($N = 320/540$) and ~88% ($N = 474/540$) of the original full sample when $\text{minTP} = 200/\text{TP} = 100$ and $\text{minTP} = 60/\text{TP} = 20$ thresholds were applied, corresponding to a gain of 3% and 31% in sample size, respectively. BID accuracies were 68% and 47% (Figure 6.4B). While bagging enables the rescue of “high-motion” participants, in this sample, it resulted in relatively low fingerprint ID accuracy which was consistent with the findings obtained in **Chapter 4**.

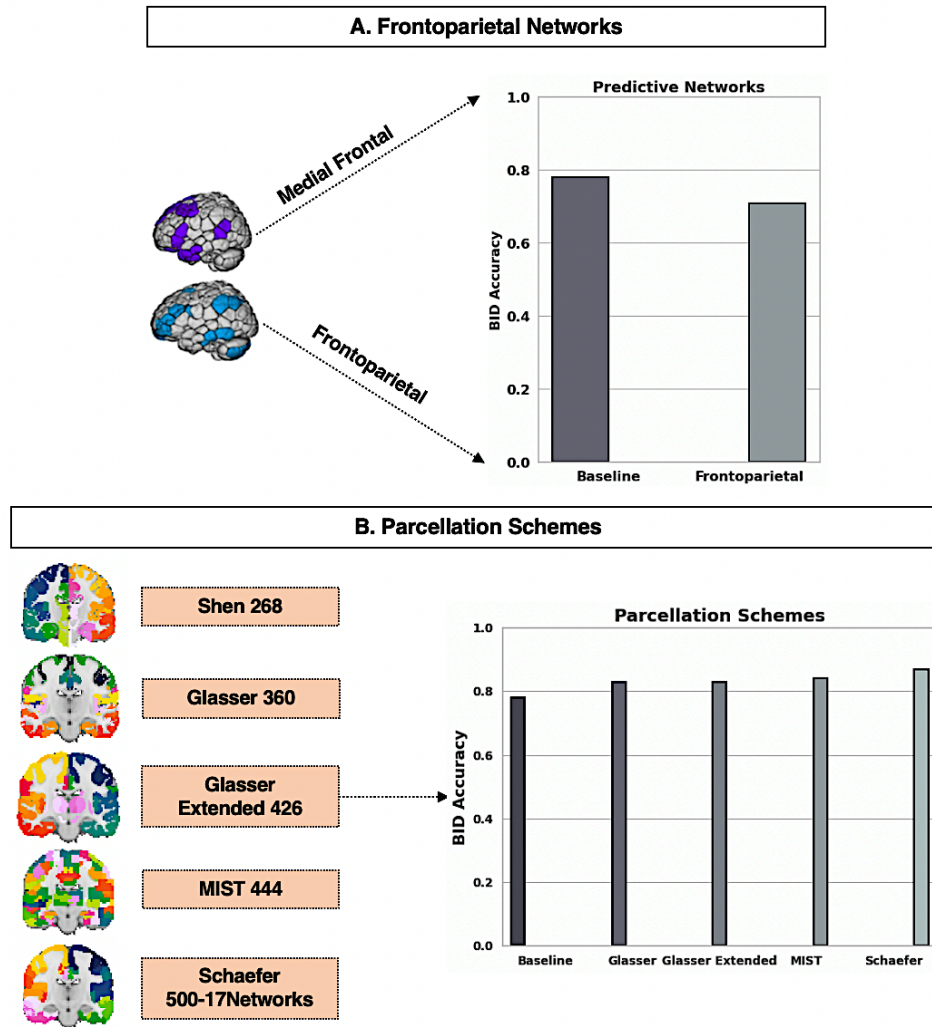


Figure 6.3. Data pre-/post-processing factors (i.e., frontoparietal networks, parcellation schemes) for connectome-based fingerprint identification. A. Connectome-based fingerprint accuracy derived from the frontoparietal networks using a 10-network definition which involves the combination of the medial frontal and frontoparietal networks. B. Connectome-based fingerprint accuracy derived from multiple parcellation schemes that differ in terms of their parcel dimensionality (i.e., 268 → 500 parcels), space (i.e., volume, surface), and spatial coverage (i.e., cortical, whole-brain).

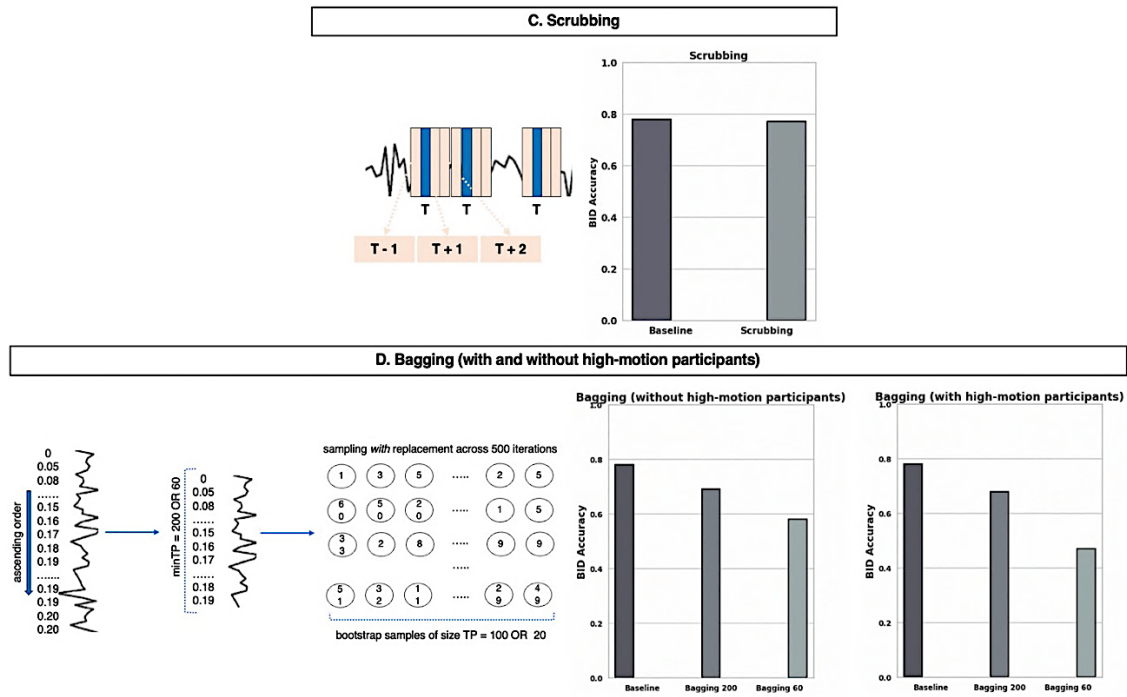


Figure 6.4. Head motion mitigation strategies (i.e., scrubbing, bagging) for connectome-based fingerprint identification. A. Connectome-based fingerprint accuracy derived from scrubbing where a motion-corrupted timepoint (T) is identified and censored if its rmsFD > 0.20 mm followed by a preceding (T - 1) and two succeeding timepoints (T + 1, T + 2). B. Connectome-based fingerprint accuracy derived from bagging with and without the inclusion of “high-motion” participants, where the least 60 or 200 motion-corrupted timepoints are selected from each participant’s fMRI timeseries out of which 20 or 100 timepoints are respectively bootstrapped across 500 iterations.

6.3.3 Connectome-based predictions of age, CBCL, and FSIQ in developing youths

CPM was performed to predict three phenotypes (i.e., age, CBCL, FSIQ) in developing youths using a k -fold CV, where $k = 10$. CPM was adjusted to control for sex and rmsFD to predict age in addition to treating age as an additional covariate for predicting CBCL and FSIQ. The baseline CPM predicted age using 5.3% (1,330/25,200) of the edges derived from the POS network and 3.1% (774/25,200) of the edges from the NEG network (thresholded at $P < 0.01$ and consistent in 95% of all instances of 10-fold CV). The relationship between predicted and observed age was strongest when the edges were derived from the Combined network [$r = 0.63$, $P < 0.001$; $R^2 = 0.39$; MAE = 2.00] (Table 6.3). The POS network predicted CBCL with the baseline CPM [$r = 0.16$, $P = 0.019$; $R^2 = 0.03$; MAE = 10.4] using 0.3% (76/25,200) of the thresholded edges (Table 6.3), while FSIQ was also predicted by the POS network [$r = 0.16$, $P = 0.027$; $R^2 = 0.03$; MAE = 14.0] using 0.7% (178/25,200) of the thresholded edges (Table 6.3). Of note, there was a significant relationship between rmsFD and the three predicted phenotypes that varied according to their network strengths: age (POS: $r = -0.31$, $P < 0.001$; NEG: $r = -0.34$, $P < 0.001$; Combined: $r = -0.34$, $P < 0.001$), CBCL (POS: $r = 0.07$, $P = 0.23$; NEG: $r = -0.17$, $P = 0.004$; Combined: $r = -0.34$, $P < 0.001$), FSIQ (POS: $r = 0.07$, $P = 0.23$; NEG: $r = -0.17$, $P = 0.004$; Combined: $r = -0.34$, $P < 0.001$).

= -0.03, $P = 0.67$), and FSIQ (POS: $r = 0.21$, $P < 0.001$; NEG: $r = 0.19$, $P = 0.002$; Combined: $r = 0.22$, $P < 0.001$).

Baseline Connectome-based Predictive Modelling									
	POS Network Strength			NEG Network Strength			Combined Network Strength		
	r	R^2	MAE	r	R^2	MAE	r	R^2	MAE
AGE	0.61***	0.37	2.04	0.57***	0.32	2.08	0.63***	0.39	2.00
CBCL	0.16*	0.03	10.4	0.0002	0.001	11.1	0.11	0.01	10.6
FSIQ	0.16*	0.03	14.0	0.02	0.001	14.8	0.12	0.02	14.2

Table 6.3. Baseline connectome-based predictive modelling of age, CBCL, and FSIQ. The three phenotypes were predicted using the positive [POS], negative [NEG], and combined [Combined] network strengths. Model performance assessed the linear relationships between predicted and observed phenotypes using Pearson's correlation coefficient (r), explained variance (R^2), and mean absolute error (MAE). *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$.

6.3.4 Neuroanatomy of the CPM-based predictive networks

A 10-network template was used (Horien et al., 2023; Horien et al., 2019; Greene et al., 2018a; Noble et al., 2017; Finn et al., 2015) to examine and visualise network membership for the three models. Edges that were predictive of age in the POS and NEG networks were distributed across several functional networks spanning the whole brain (Figure 6.5). The greatest proportion of predictive edges in the POS network was found in the salience, motor, and subcortical networks, followed by moderate contributions of the frontoparietal, default mode, and visual networks. For the NEG network, the majority of the predictive edges involved the salience, medial frontal, frontoparietal, motor, and visual networks, whereas the connections derived from the default mode, subcortical, and cerebellar networks were least important. For CBCL, the POS network involved edges concentrated in the subcortical, motor, default mode, medial frontal, and visual networks (Figure 6.5), while the model for FSIQ featured edges in the subcortical and motor networks, followed by minimal involvement of the cerebellar and salience networks (Figure 6.5).

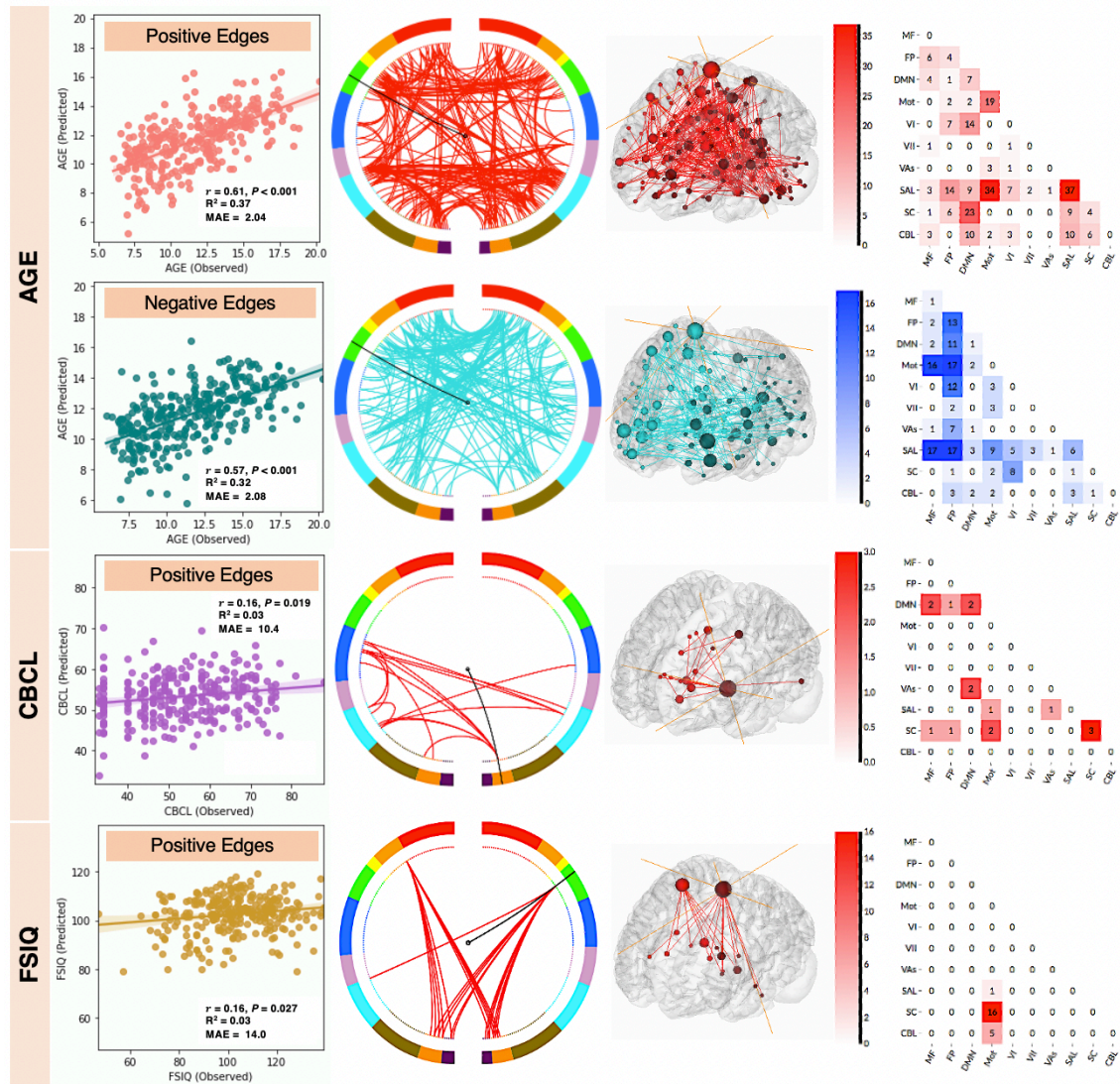


Figure 6.5. Neuroanatomy of the CPM-based predictive networks for age, CBCL, and FSIQ. First Panel: Relationships between predicted and observed phenotypes derived from the respective POS and NEG network strengths. Second Panel: Chord plots illustrating the contributions of the predictive POS and NEG edges in different functional networks using a 10-network definition. Note: To facilitate the post-hoc visualisations, the POS and NEG edges in predicting age were thresholded with a degree of 20. Third Panel: Glass brains illustrating the connectivity patterns across the whole brain for the POS and NEG network strengths. Fourth Panel: Connectivity matrices summarising the connectivity patterns within and between different functional networks based on their POS and NEG network strengths. The shades represent the degree of edge contributions such that darker shades indicate greater contribution of the edges in a functional network. MF: medial frontal; FP: frontoparietal; DMN: default mode; Mot: motor; VI: visual I; VII: visual II; VA: visual association; SAL: salience; SC: subcortical; CBL: cerebellum.

6.3.5 Impact of methodological configurations on CPM-based predictions

After deriving the baseline connectome-based predictions of age, CBCL, and FSIQ, CPM was rerun using two data pre-/post-processing factors (i.e., frontoparietal networks and parcellation schemes) and two head motion mitigation strategies (i.e., scrubbing and bagging) to assess whether these alternative configurations improve predictive power. Regardless of configuration,

the relationships between predicted and observed age were largely maintained (Table 6.4), with the exception being the age model constructed using the frontoparietal networks, which showed a significantly weaker relationship between predicted and observed age, relative to the baseline model across the three network strengths [$z_p = 5.00, P < 0.001$; $z_n = 5.25, P < 0.001$; $z_c = 5.06, P < 0.001$]. Further, only finer brain parcellations such as Glasser 360 [$z_c = 5.06, P = 0.028$] and MIST 444 [$z_n = -2.29, P = 0.022$] significantly altered the relationship between predicted and observed age (Table 6.5). Note: the Steiger's Z test could not be applied to test the significance of the difference between baseline and bagged CPM-based predictions.

For the CBCL models, scrubbing was the only manipulation that preserved the relationship between predicted and observed behaviour [$r_p = 0.18, P = 0.013$] (Table 6.4); this relationship was not statistically different from the baseline prediction [$z_p = -1.01, P = 0.32$] (Table 6.5). For the FSIQ models, finer whole-brain parcellation schemes, scrubbing, and bagging all maintained the relationships between predicted and observed FSIQ, with some additional significant relationships for Combined networks observed (Table 6.4). Only the finer MIST 444 parcellation scheme [$z_c = -2.06, P = 0.039$] and scrubbing [$z_c = -2.69, P = 0.007$] significantly altered the relationship between predicted and observed FSIQ compared to the baseline predictions (Table 6.5).

Configurations of Connectome-based Predictive Modelling					
METHOD	BID	BID PATTERN	AGE	CBCL	FSIQ
Shen 268	0.78	Baseline	$r_p = 0.61^{***}$ $r_n = 0.57^{***}$ $r_c = 0.63^{***}$	$r_p = 0.16^*$ $r_n = 0.0002$ $r_c = 0.11$	$r_p = 0.16^*$ $r_n = 0.02$ $r_c = 0.12$
Predictive Networks					
Frontoparietal	0.71	↓	$r_p = 0.40^{***}$ $r_n = 0.35^{***}$ $r_c = 0.46^{***}$	$r_p = 0.03$ $r_n = 0.006$ $r_c = 0.03$	$r_p = 0.03$ $r_n = -0.003$ $r_c = 0.009$
Parcellation Schemes					
Glasser 360	0.83	↑	$r_p = 0.56^{***}$ $r_n = 0.53^{***}$ $r_c = 0.58^{***}$	$r_p = 0.12$ $r_n = 0.05$ $r_c = 0.10$	$r_p = 0.10$ $r_n = 0.08$ $r_c = 0.10$
Glasser 426	0.83	↑	$r_p = 0.61^{***}$ $r_n = 0.53^{***}$ $r_c = 0.60^{***}$	$r_p = 0.14$ $r_n = 0.03$ $r_c = 0.10$	$r_p = 0.17^*$ $r_n = 0.10$ $r_c = 0.15^*$
MIST 444	0.84	↑	$r_p = 0.61^{***}$ $r_n = 0.60^{***}$ $r_c = 0.62^{***}$	$r_p = 0.07$ $r_n = 0.005$ $r_c = 0.04$	$r_p = 0.18^*$ $r_n = 0.13$ $r_c = 0.17^*$
Schaefer 500	0.87	↑	$r_p = 0.56^{***}$ $r_n = 0.58^{***}$ $r_c = 0.60^{***}$	$r_p = 0.09$ $r_n = -0.004$ $r_c = 0.06$	$r_p = 0.06$ $r_n = 0.09$ $r_c = 0.08$
Head Motion Mitigation Strategies					
Scrubbing	0.77	↓	$r_p = 0.60^{***}$ $r_n = 0.58^{***}$ $r_c = 0.63^{***}$	$r_p = 0.18^*$ $r_n = -0.06$ $r_c = 0.10$	$r_p = 0.18^*$ $r_n = 0.06$ $r_c = 0.16^*$
Bagging (minTP = 200 & TP = 100)	0.69	↓	$r_p = 0.62^{***}$ $r_n = 0.53^{***}$ $r_c = 0.62^{***}$	$r_p = -0.08$ $r_n = -0.12$ $r_c = -0.12$	$r_p = 0.21^{**}$ $r_n = 0.04$ $r_c = 0.18^*$
Bagging (minTP = 60 & TP = 20)	0.58	↓	$r_p = 0.56^{***}$ $r_n = 0.54^{***}$ $r_c = 0.58^{***}$	$r_p = 0.12$ $r_n = 0.01$ $r_c = 0.08$	$r_p = 0.20^{**}$ $r_n = -0.03$ $r_c = 0.15^*$
Bagging (minTP = 200 & TP = 100) w/ High-Motion	0.68	↓	$r_p = 0.60^{***}$ $r_n = 0.50^{***}$ $r_c = 0.58^{***}$	$r_p = -0.04$ $r_n = 0.03$ $r_c = -0.006$	$r_p = 0.16^*$ $r_n = 0.06$ $r_c = 0.14^*$
Bagging (minTP = 60 & TP = 20) w/ High-Motion	0.47	↓	$r_p = 0.54^{***}$ $r_n = 0.49^{***}$ $r_c = 0.54^{***}$	$r_p = 0.06$ $r_n = 0.09$ $r_c = 0.09$	$r_p = 0.09$ $r_n = -0.01$ $r_c = 0.06$

Table 6.4. CPM configurations derived from data pre-/post-processing factors and head motion mitigation strategies for predicting age, CBCL, and FSIQ. The data pre-/post-processing factors involve using predictive networks and several parcellation schemes that differ in terms of their parcel dimensionality, space, and spatial coverage. The head motion mitigation strategies include scrubbing and bagging. Here, model performance is indicated by the Pearson's correlation coefficient (r) between predicted and observed phenotypes. r_p = positive network strength, r_n = negative network strength, r_c = combined network strength. $***P < 0.001$, $**P < 0.01$, $*P < 0.05$.

METHOD	Steiger's Z AGE	Steiger's Z CBCL	Steiger's Z FSIQ
Shen 268	Baseline Baseline Baseline	Baseline Baseline Baseline	Baseline Baseline Baseline
Predictive Networks			
Frontoparietal	$z_p = 5.00^{***}$ $z_n = 5.25^{***}$ $z_c = 5.06^{***}$	$z_p = 2.29^*$ $z_n = -0.09$ $z_c = 1.38$	$z_p = 1.69$ $z_n = 0.34$ $z_c = 1.56$
Parcellation Schemes			
Glasser 360	$z_p = 1.79$ $z_n = 1.72$ $z_c = 2.19^*$	$z_p = 0.90$ $z_n = -0.86$ $z_c = 0.22$	$z_p = 1.06$ $z_n = -1.33$ $z_c = 0.42$
Glasser 426	$z_p = 0$ $z_n = 1.89$ $z_c = 1.50$	$z_p = 0.49$ $z_n = -0.59$ $z_c = 0.25$	$z_p = -0.34$ $z_n = -2.01^*$ $z_c = -1.01$
MIST 444	$z_p = 0$ $z_n = -2.29^*$ $z_c = 0.79$	$z_p = 2.67^{**}$ $z_n = -0.12$ $z_c = 2.20^*$	$z_p = -0.74$ $z_n = -3.83^{***}$ $z_c = -2.06^*$
Schaefer 500	$z_p = 1.94$ $z_n = -0.46$ $z_c = 1.44$	$z_p = 1.79$ $z_n = 0.08$ $z_c = 1.24$	$z_p = 1.88$ $z_n = -1.68$ $z_c = 0.91$
Head Motion Mitigation Strategies			
Scrubbing	$z_p = 0.90$ $z_n = -0.68$ $z_c = 0$	$z_p = -1.01$ $z_n = 2.31^*$ $z_c = 0.46$	$z_p = -1.35$ $z_n = -1.74$ $z_c = -2.69^{**}$
Bagging (minTP = 200 & TP = 100)	— — —	— — —	— — —
Bagging (minTP = 60 & TP = 20)	— — —	— — —	— — —
Bagging (minTP = 200 & TP = 100) with High-Motion	— — —	— — —	— — —
Bagging (minTP = 60 & TP = 20) w/ High-Motion	— — —	— — —	— — —

Table 6.5. CPM configurations derived from data pre-/post-processing factors and head motion mitigation strategies for predicting age, CBCL, and FSIQ. The data pre-/post-processing factors involve using predictive networks and several parcellation schemes that differ in terms of their parcel dimensionality, space, and spatial coverage. The head motion mitigation strategies include scrubbing and bagging. Here, the Steiger's Z tests are presented for each CPM configuration and network strength to compare the baseline HBN-derived predictions of age, CBCL, and FSIQ with those obtained from the data pre-/post-processing factors and head motion mitigation strategies. Note: Steiger's Z test was not performed to compare the baseline CPM performance with that derived from bagging as the sample sizes are not equal: the assumption of the test relies on computing the statistical significance of the difference between two dependent correlations with the same sample size. $***P < 0.001$, $**P < 0.01$, $*P < 0.05$.

6.3.6 Generalisability of connectome-based predictions in developing populations

To externally validate the connectome-based predictions derived from the HBN, an independent sample of non-Autistic individuals were obtained from ABIDE to predict age and FSIQ (CBCL

scores were not available for ABIDE). Baseline model predictions of age generalised for the POS [$r = 0.27$, $P < 0.001$], NEG [$r = 0.43$, $P < 0.001$], and Combined [$r = 0.35$, $P < 0.001$] networks (Figure 6.6). The baseline CPM-based predictions of age were significantly different from those derived from the frontoparietal networks and finer parcellation schemes (Steiger's Z test, $P < 0.05$; Table 6.6). On the other hand, the baseline predictions of FSIQ did not generalise across the three network strengths (Figure 6.6). However, frontoparietal networks and finer parcellation schemes were beneficial in predicting FSIQ, which were significantly different from the baseline predictions (Steiger's Z test, $P < 0.05$; Table 6.6). Bagging was the only head motion strategy that facilitated the prediction of FSIQ across the POS [$r = 0.09$, $P = 0.04$], NEG [$r = 0.11$, $P = 0.01$], and Combined networks [$r = 0.12$, $P = 0.004$] (Table 6.6). The complete impact of the data pre-/post-processing factors and head motion mitigation strategies on the CPM-based predictions of age and FSIQ are further presented in Table 6.6.

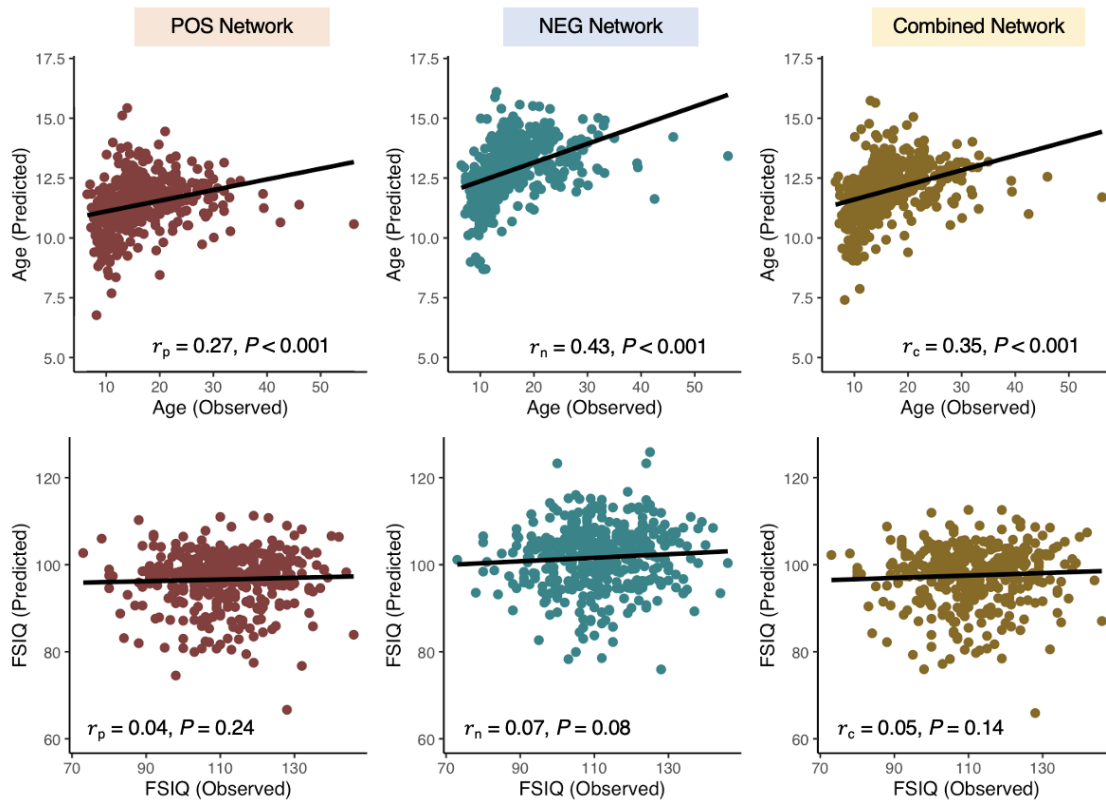


Figure 6.6. Generalisability of baseline CPM-based predictions of age and FSIQ in ABIDE. First Panel: Relationships between predicted and observed age derived from the respective POS, NEG, and Combined network strengths. Second Panel: Relationships between predicted and observed FSIQ derived from the respective POS, NEG, and Combined network strengths. Here, model performance is indicated by the Pearson's correlation coefficient (r) between predicted and observed phenotypes. r_p = positive network strength, r_n = negative network strength, r_c = combined network strength.

Configurations of Connectome-based Predictive Modelling				
METHOD	AGE	FSIQ	Steiger's Z AGE	Steiger's Z FSIQ
Shen 268	$r_p = 0.27^{***}$ $r_n = 0.43^{***}$ $r_c = 0.35^{***}$	$r_p = 0.04$ $r_n = 0.07$ $r_c = 0.05$	Baseline Baseline Baseline	Baseline Baseline Baseline
Predictive Networks				
Frontoparietal	$r_p = 0.14^{**}$ $r_n = 0.13^{**}$ $r_c = 0.19^{***}$	$r_p = 0.06$ $r_n = 0.08^*$ $r_c = 0.09^*$	$z_p = 2.57^*$ $z_n = 5.73^{***}$ $z_c = 3.52^{***}$	$z_p = -0.31$ $z_n = -0.20$ $z_c = -0.73$
Parcellation Schemes				
Glasser 360	$r_p = 0.19^{***}$ $r_n = 0.32^{***}$ $r_c = 0.27^{***}$	$r_p = 0.09^*$ $r_n = 0.04$ $r_c = 0.07$	$z_p = 2.62^{**}$ $z_n = 4.39^{***}$ $z_c = 3.10^{**}$	$z_p = -1.13$ $z_n = 0.84$ $z_c = -0.50$
Glasser 426	$r_p = 0.25^{***}$ $r_n = 0.40^{***}$ $r_c = 0.35^{***}$	$r_p = 0.06$ $r_n = 0.06$ $r_c = 0.07$	$z_p = 0.74$ $z_n = 1.66$ $z_c = 0$	$z_p = -0.82$ $z_n = 0.37$ $z_c = -0.95$
MIST 444	$r_p = 0.28^{***}$ $r_n = 0.39^{***}$ $r_c = 0.34^{***}$	$r_p = 0.07$ $r_n = 0.10^*$ $r_c = 0.09^*$	$z_p = -0.56$ $z_n = 2.36^*$ $z_c = 0.62$	$z_p = -1.51$ $z_n = -1.14$ $z_c = -2.02^*$
Schaefer 500	$r_p = 0.18^{***}$ $r_n = 0.32^{***}$ $r_c = 0.27^{***}$	$r_p = 0.10^*$ $r_n = 0.11^{**}$ $r_c = 0.12^{**}$	$z_p = 2.81^{**}$ $z_n = 4.39^{***}$ $z_c = 3.00^{**}$	$z_p = -1.45$ $z_n = -1.35$ $z_c = -2.13^*$
Head Motion Mitigation Strategies				
Scrubbing	$r_p = 0.26^{***}$ $r_n = 0.42^{***}$ $r_c = 0.34^{***}$	$r_p = 0.03$ $r_n = 0.06$ $r_c = 0.04$	$z_p = 0.74$ $z_n = 0.70$ $z_c = 0.76$	$z_p = 0.82$ $z_n = 0.64$ $z_c = 1.01$
Bagging (minTP = 200 & TP = 100)	$r_p = 0.41^{***}$ $r_n = 0.41^{***}$ $r_c = 0.43^{***}$	$r_p = 0.09$ $r_n = 0.01$ $r_c = 0.08$	— — —	— — —
Bagging (minTP = 60 & TP = 20)	$r_p = 0.19^{***}$ $r_n = 0.39^{***}$ $r_c = 0.31^{**}$	$r_p = 0.06$ $r_n = 0.05$ $r_c = 0.07$	— — —	— — —
Bagging (minTP = 200 & TP = 100) with High-Motion	$r_p = 0.38^{***}$ $r_n = 0.45^{***}$ $r_c = 0.44^{***}$	$r_p = 0.06$ $r_n = 0.10$ $r_c = 0.09$	— — —	— — —
Bagging (minTP = 60 & TP = 20) w/ High-Motion	$r_p = 0.21^{***}$ $r_n = 0.35^{***}$ $r_c = 0.31^{***}$	$r_p = 0.09^*$ $r_n = 0.11^*$ $r_c = 0.12^{**}$	— — —	— — —

Table 6.6. CPM configurations derived from data pre-/post-processing factors and head motion mitigation strategies for predicting age and FSIQ in ABIDE. The data pre-/post-processing factors involve using predictive networks and several parcellation schemes that differ in terms of their parcel dimensionality, space, and spatial coverage. The head motion mitigation strategies include scrubbing and bagging. Here, model performance is indicated by the Pearson's correlation coefficient (r) between predicted and observed phenotypes. The Steiger's Z tests are also computed for each CPM configuration and network strength to compare the baseline ABIDE-derived predictions of age and FSIQ with those obtained from the data pre-/post-processing factors and head motion mitigation strategies. Note: Steiger's Z test was not performed to compare the baseline CPM performance with that derived from bagging as the sample sizes are not equal: the assumption of the test relies on computing the statistical significance of the difference between two dependent correlations with the same sample size. r_p = positive network strength, r_n = negative network strength, r_c = combined network strength. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$.

6.4 Discussion

The present study builds on my prior work — which sought to improve the measurement of individual differences in the functional connectome — and assesses whether the strategies I identified in **Chapters 2 and 4** also enhance the robustness and reproducibility of brain-behaviour relationships, prerequisites for biomarker discovery. Applying the same logic of **Chapters 2 and 4**, I first tested the reproducibility of my findings in an independent, developing, and help-seeking sample — that is, I examined whether the previously identified data pre-/post-processing factors and head motion mitigation strategies improved fingerprint-based identification accuracy. Connectome-based identification for the baseline pipeline was 78%; while finer parcellation schemes improved the ability to identify an individual from their functional connectomes up to 87%, consistent with **Chapter 2**, restricting the connectome to the frontoparietal networks did not boost ID accuracy, nor did bagging. Next, I investigated whether the data pre-/post-processing factors and head motion mitigation strategies improved connectome-based phenotypic predictions of age, general psychopathology, and intellectual ability. Finally, I assessed whether the same factors could improve the generalisability of the CPM-based predictions, using an independent sample. CPM-based predictions were generally preserved, but not improved, using the pre-/post-processing factors and head motion mitigation strategies that I examined. Phenotype predictions were also generalisable to an independent sample, with some factors (i.e., finer grained parcellation and bagging) significantly improving generalisability. These findings thus paint a mixed picture — strategies for increasing connectome-based fingerprint identification accuracy identified in neurotypical adults and youths in **Chapter 2** did not fully reproduce in an independent sample of help-seeking youths, nor did they translate to stronger connectome-based phenotypic predictions. Several strategies did improve generalisability, however — suggesting that while these strategies do not necessarily improve validity (strength of brain-behaviour relationships), they do strengthen reproducibility (generalisability of brain-behaviour relationships).

Consistent with my observations in **Chapter 2**, connectome-based fingerprint identification accuracy increased with parcellation schemes that encapsulate finer properties of the functional connectome. To limit the bottleneck for increasing the computational resources of CPM, the highest parcel dimensionality was captured with 500 parcels instead of 1000 parcels that were applied in **Chapter 2**. Model-based predictions of age were generally unaffected by parcellation schemes, while there were some increases in sensitivity for predicting intellectual ability. These predictions were also generalisable, with increased generalisability observed for intellectual ability for the highest parcel dimensionality (i.e., 500 parcels). The former finding is consistent with previous work, which found that CPM-based phenotypic predictions do not differ

significantly across different brain parcellations (Mantwill et al., 2022; Dadi et al., 2019; Greene et al., 2018a). This suggests that the process of summarising meaningful edges (sometimes as many as 10^2 or 10^3) that share an association with a chosen phenotype into a singular value is robust irrespective of the parcellation scheme being applied to define the functional connectome. My observations on improved generalisability are novel and important, however, particularly given the developmental samples examined. Using high-dimensionality parcellations can make the CPM highly computationally intensive, variable, and potentially less interpretable (Tejavibulya et al., 2022; Dhamala et al., 2022) — however, such costs may be acceptable in light of improved reproducibility of brain-behaviour relationships (i.e., generalisability). Given that there are no one-parcellation-fits-all datasets, individualised parcellations may be one route to improve the CPM performance without necessarily increasing dimensionality (Kong et al., 2021; Salehi et al., 2020).

In contrast to my observations for higher dimensionality parcellations, there was a decrease in connectome-based fingerprint identification when the functional connectome was restricted to the frontoparietal networks. This finding is not consistent with mounting evidence of the frontoparietal networks being the discriminatory functional networks that maximise participant identification (Horien et al., 2023; Venkatesh et al., 2020; Horien et al., 2019; Finn et al., 2015) nor with my findings in **Chapter 2**. Further, restricting the CPM to frontoparietal networks alone significantly dampened the connectome-based phenotypic predictions. This might be expected — the full (baseline) predictive model showed that several other functional networks, including the salience, default mode, visual, motor, subcortical, and cerebellar networks were important for predictions of the three phenotypes. If we restrict the connectome only to the frontoparietal networks, it might continue to show some relevance for behavioural phenotypes such as age and intellectual ability, but predictive power will be lost due to the exclusion of other regions. Hence, a narrow focus on the most stable and reproducible functional networks does not necessarily generate the most useful or meaningful features for predicting individual phenotypes (Finn and Rosenberg, 2021; Noble et al., 2017). These observations are consistent with Finn and Rosenberg (2021) who argued that an individual’s functional connectome can be highly differentiable from a group (i.e., *high* reproducibility) but exhibit low utility for phenotypic predictions (i.e., *high* reliability and *low* validity). Instead, we require a balance between selecting stable and reproducible brain networks that boost the identifiability of the functional connectome but also, choosing a behavioural phenotype that is relevant to these networks.

The motion strategies I investigated also yielded a mixed set of findings. Scrubbing had minimal impact on the fingerprint identification accuracy — which is a positive outcome, since connectome-based identification tends to increase with longer scan duration (Vanderwal et al.,

2021; Venkatesh et al., 2020; Vanderwal et al., 2017; Finn et al., 2015; see **Chapter 2**). The application of scrubbing suggests that participants who have motion-limited timeseries can still be included in the fingerprint identification procedure to produce reproducible functional connectivity estimates. A recent study has shown that scrubbing yields small improvements in the fingerprint matching of participants with network-based functional connectomes (Pham et al., 2023). While scrubbing had a significant impact on the prediction of intellectual ability, it did not affect the generalisation of connectome-based phenotypes in developing youths. On the other hand, while bagging reduced the identifiability of the functional connectome, it had relatively minimal effects on the strength of predictions for age and intellectual ability, and boosted generalisability for these phenotypes. That is, bagging allowed the retention of the “high-motion” developing youths who were included in the CPM for the successful predictions of age and FSIQ scores. Given that connectome-based identification decreases with larger sample sizes (see **Chapter 2**), the inclusion of these “high-movement” individuals with bagging (minTP = 200, TP = 100) did not drastically affect the fingerprinting procedure. The primary advantage of bagging, however, is that these developing youths did not undermine the connectome-based phenotypic prediction of age and intellectual ability. The applications of finer brain parcellations and bagging suggest that the generalisability tests of age and intellectual ability in ABIDE are relatively strong. Given that the training HBN sample includes help-seeking developmental participants and the samples are therefore enriched for psychiatric symptoms, validating these predictions in the ABIDE non-Autistic “comparison” sample means that the predictions for age and intellectual ability are generalisable to neurotypical individuals who are non-symptomatic for psychiatric or neurodevelopmental conditions. In contrast, scrubbing and bagging were not beneficial for improving the connectome-based prediction of general psychopathology in this developmental sample. This is likely due to the few contributions of edges that are meaningful or useful from the whole-brain connectome for predicting the CBCL scores. This also suggests that the CBCL may not be a robust behavioural measure to index externalising behaviours in developing youths, at least not as robust as age and FSIQ. It is important to note that self-reported trait scores such as CBCL are more difficult to predict than objective (e.g., age) and/or task performance (e.g., FSIQ) measurements. This may be because they are less biologically valid and/or represented more heterogeneously, since they often constitute composite scores reflecting several types of behaviours.

Using different strategies to manipulate the individual functional connectome for testing the robustness and generalisability of CPM-based predictions demonstrate that several of them had minimal impact for predicting the three chosen phenotypes. There is a possibility that CPM imposes a ceiling for predicting certain phenotypes, beyond which little to no improvement can be observed. If we take the example of fluid intelligence (one of the most studied phenotypes in

the CPM literature), its out-of-sample generalisability with resting-state conditions is $r \approx 0.20$ (Greene et al., 2018a). This is an important example as the prediction of fluid intelligence has mostly been derived from young adults from the HCP to train the CPM; these studies show that r can range from as low as 0.02 to as high as 0.50 depending on the pre-/post-processing pipeline being applied (Cai et al., 2021; Lin et al., 2020; Greene et al., 2018a; Dubois et al., 2018a; Finn et al., 2015). In the case of intellectual ability studied here, its successful prediction from the HBN cohort showed that r ranged from 0.14 to 0.18 across different CPM configurations. In terms of generalisability, intellectual ability was successfully validated in the ABIDE sample with r ranging from 0.08 to 0.12. Therefore, we cannot assume that these phenotypic predictions will always be robust and generalisable because of the variations in CPM-based predictions. Assessing the robustness and generalisability of CPM-based predictions should be prioritised as manipulating the individual functional connectome to enhance predictive performances can be harmful without proper justifications. Recently, Rosenblatt et al. (2023) showed that introducing subtle manipulations in the functional connectome can significantly inflate the CPM-derived predictive performances (e.g., relationships between observed and predicted fluid intelligence) that are not easily detectable by visual inspection. I return to the prioritisation of the robustness of CPM-based predictions in **Chapter 9**.

This study has several limitations that outline opportunities for future investigations. One caveat is that the scan duration was relatively short, and the two scans were collected within- rather than between-sessions. Although an acquisition length of 5 min produced a relatively high baseline connectome-based fingerprint identification followed by successful phenotypic predictions in developing youths, longer scan lengths (e.g., 200-1200 timepoints) can increase the predictions of connectome-based phenotypes derived from resting-state and task-based conditions (Jiang et al., 2020; see **Chapter 5 (Section 5.3)**). Unfortunately, such scanning durations present a challenge in practice as developing youths, particularly those with psychiatric symptoms, often suffer from low in-scanner tolerance due to heightened head motion, lower compliance, and drowsiness (Zou et al., 2015; Tagliazucchi et al., 2014; Patriat et al., 2013; see **Chapter 3 (Section 3.3)**). Nonetheless, concatenating fMRI scans across sessions and runs could be a plausible strategy to increase the amount of fMRI data required per participant (Cho et al., 2020; Elliott et al., 2019; O'Connor et al., 2017; Noble et al., 2017). A future avenue could also investigate the impact of within- vs. between-session functional connectomes on the chosen phenotypic predictions to understand whether acquiring imaging data at shorter/longer intervals make the phenotypic predictions less/more sensitive to data manipulations. A second shortcoming is that I was limited in my ability to test the generalisability of the phenotypic predictions — the analyses were restricted to age (one of the most robust and well-studied phenotypes in the fMRI literature) and intellectual ability. Given my interest in identifying biomarkers for

neurodevelopmental and psychiatric conditions, assessing the robustness and generalisability of relevant phenotypes should be a priority for future work. Such an example is the study of sustained attention which has consistently been predicted and validated in neurodevelopmental populations (Horien et al., 2023; Chamberlain and Rosenberg, 2022; Rosenberg et al., 2020; Rosenberg et al., 2018; Rosenberg et al., 2016; Rosenberg et al., 2015; see **Chapter 5 (Section 5.2)**). Future work could focus on deeply phenotyped and longitudinal datasets such as the ABCD Study ($N \approx 12,000$; Casey et al., 2018). Finally, here, CPM was applied using resting-state fMRI data to predict three phenotypes. Task-based paradigms have been shown to enhance the detection of individual differences in the functional connectome as well as brain-behaviour predictions (Jiang et al., 2020; Greene et al., 2018a; Finn et al., 2017; see **Chapter 5 (Sections 5.2-5.3)**). Some task-induced conditions have been shown to yield higher predictive power compared to others whereas aggregating different task connectomes in a single CPM can also improve model predictions than using individual conditions (Jiang et al., 2020; Gao et al., 2019; Greene et al., 2018a). Certain tasks characterise specific brain states that are more relevant to reveal trait-like individual differences across the whole brain, thus contributing to more meaningful edges in the CPM (Jiang et al., 2020; Greene et al., 2018a). Future work can further investigate whether these trait-like properties can be enhanced across different task paradigms and data processing strategies to boost the predictions of relevant phenotypes.

6.5 Conclusions

My work points out that improving the reproducibility of the individual functional connectome (i.e., *higher* fingerprint accuracy) preserves the brain-behaviour relationships but does not necessarily improve its validity (i.e., strength of brain-behaviour relationships). However, the generalisability of brain-behaviour relationships can be improved under certain configurations such as higher-dimensional brain parcellations and head motion mitigation strategies such as bagging in neurotypical populations. Therefore, the ability to capture finer properties of the functional connectome and maximising participant inclusivity is beneficial for generalisable phenotypic predictions. Given the variations that are present in-sample and out-of-sample CPM-based predictions, manipulating the functional connectome to improve the robustness and generalisability of these predictions should be properly justified. Future work should encourage the study of longer scan lengths either using concatenation strategies and/or increasing the acquisition time, within- vs. between-session scans, reliable phenotypes, and/or task-based fMRI paradigms.

The empirical research conducted in this thesis has so far investigated the reproducibility of functional connectivity with an examination of its validity for brain-behaviour relationships,

particularly in developing populations. While I achieved this with methodological innovations (e.g., functional connectome fingerprinting, bagging, CPM) using resting-state fMRI conditions, ecologically valid paradigms such as movies have the ability to address some of the limitations of both resting-state and task-based conditions (reviewed in **Chapters 1-6**). I will now take a different approach to improving reproducibility in **Chapter 7** by describing how movies can be applied to elicit neural responses as well as examining their behavioural utility in neurodevelopmental and psychiatric conditions. Subsequently, I will examine a naturalistic fMRI framework in **Chapter 8** that aims to capture individual differences in emotional responses at multiple spatial and temporal scales in female adolescents experiencing depressive symptoms.

Chapter 7

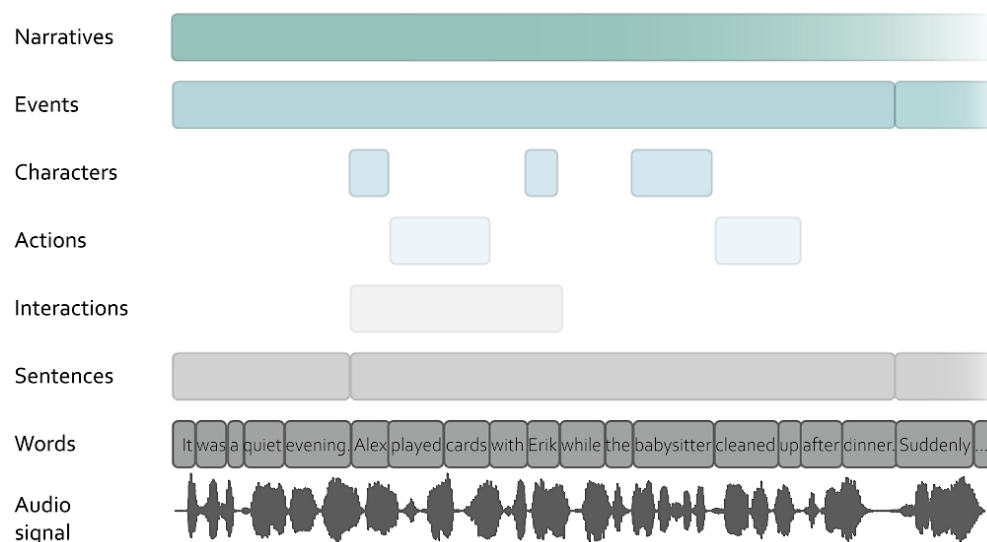
Towards ecologically valid naturalistic paradigms to study functional brain responses at multiple spatial and temporal scales

Summary

Chapter 7 takes a different approach to reproducibility, focusing on naturalistic fMRI as an ecologically valid, diverse, rich, and engaging approach to study neural responses and their relationships with behavioural phenotypes. First, I discuss the advantages of using naturalistic fMRI paradigms, relative to resting-state and task-based conditions. Second, I describe the analytic frameworks that index functional brain responses and their phenotypic relevance at multiple spatial and temporal scales — specifically, intersubject correlation, functional connectivity, and the Hurst Exponent.

7.1 Emergence of Naturalistic fMRI Paradigms

Recent years have seen growing efforts to develop study paradigms that improve in-scanner compliance and data quality. These efforts have largely focused on naturalistic fMRI paradigms — study designs that elicit neural responses to real-world experiences and situations, typically delivered in movie form — leading to the development and application of dynamic, feature-rich, and engaging audiovisual stimuli (Finn et al., 2022; Willems et al., 2020; Vanderwal et al., 2019; Nguyen et al., 2019; Dosenbach et al., 2017; Alexander et al., 2017; Maguire, 2012; Spiers and Maguire, 2007; Bartels and Zeki, 2005; Hasson et al., 2004). Naturalistic stimuli are considered ecologically valid¹ paradigms, since they better replicate the dynamic world we operate in than traditional task-based paradigms (Willems et al., 2020). In traditional experimental designs, real-world scenarios are decomposed into subcomponents that reflect highly controlled experimental manipulations (Nastase et al., 2020). While controlled experimental designs can provide information that correspond to ecological brain functions, no single paradigm can be representative and generalisable across all domains because the subcomponents are usually interconnected with each other (Willems et al., 2020; Nastase et al., 2020). In contrast, ecologically valid paradigms (e.g., narratives) are better suited to study the variations that naturally occur in these stimuli, and tend to be more generalisable than traditional experimental designs (Willems et al., 2020). These paradigms further enable the separation of individual embedded features of interest during naturalistic stimulation (Figure 7.1) — a procedure that is constrained for traditional paradigms comprising strictly controlled stimuli.



¹The term “ecological validity” is described in Nastase et al. (2020) which reflects Egon Brunswik’s definition that describes the utility of a perceptual cue with respect to an ecologically-relevant state of the environment (Brunswik, 1949; Brunswik, 1947).

Figure 7.1. A bottom-up representation of the layers that are involved in narratives. The features of a narration can reflect audio signals, words, sentences, interactions, characters, events, and story-telling. Adapted from Willems et al., 2020.

Naturalistic stimuli often feature extracts from Hollywood films, shorter animated movies, or abstract scenes (Figure 7.2). Given that it is difficult to accommodate an entire movie (e.g., *Ocean's Eleven* lasts for 116 minutes) within a study design, researchers tend to employ edited versions of a film by selecting relevant story-telling segments to study the desired constructs, thus allowing for well-controlled narratives. Children and adolescents tend to engage more with the narratives of the stimuli (Vanderwal et al., 2019; Vanderwal et al., 2015), even when the storyline involves complex scenes that capture higher-order processes such as memory, attention, and emotion. When participants are scanned at rest, they can be bored, “mind-wander”, and/or fall asleep unintentionally, especially with longer scan duration, leading to less reliable fMRI signals (see **Chapter 1 (Section 1.6.3)**). On the other hand, when participants perform a task to evoke a specific process, the need for practice and training, task difficulty, repetition, and cognitive load pose challenges for younger participants and those with neurodevelopmental, neurological, or psychiatric conditions. If we take the examples of ADHD and depressed participants, difficulties with attention or anxiety could interfere with task acquisition and performance. Naturalistic fMRI conditions could, therefore, circumvent the limitations of both resting-state and task-based paradigms, and facilitate the search for fMRI-based biomarkers in neurodevelopmental and psychiatric conditions (Finn, 2021a; Eickhoff et al., 2020; Sonkusare et al., 2019).

As elaborated in **Chapter 3 (Section 3.3)**, youth populations tend to naturally move more than adults, which often limits the amount of useable data that we have at our disposal to perform brain-behaviour analyses. Some large-scale datasets such as the HBN and HCP now provide movie-watching scans, thus allowing researchers to examine different naturalistic stimuli prior to formulating their hypotheses or research questions. Hypotheses developed for one naturalistic stimulus can be validated with other ecological stimuli (Willems et al., 2020). Researchers can also take advantage of the recently established Open-Source naturalistic-based dataset known as the Naturalistic Neuroimaging Database (NDDb; $N = 86$; 18-58 years; <https://openneuro.org/datasets/ds002837/>), which includes 10 full-length movies released between 1985 and 2016, covering 10 different genres such as action, drama, sci-fi, historical, and documentary, among others (Aliko et al., 2020). The “Narratives” collection is another Open-Source large-scale dataset that aggregates 27 spoken stories ($N = 345$; ~4.6 hours; ~43,000 words) to benchmark models of naturalistic language and narrative comprehension in a reproducible manner (<https://openneuro.org/datasets/ds002345/>; Nastase et al., 2021). Naturalistic stimuli not only provide a rich repertoire of real-word contexts, but also allow researchers to focus on different story-telling segments by annotating features of a movie such as faces, eye and bodily movements, sound, brightness, and

so forth (Finn and Bandettini, 2021b; Meer et al., 2020). Therefore, the richness of naturalistic stimuli allows researchers to compare and evaluate multiple hypotheses at the same time. Given the potential of movies to elicit neural processes at different spatial and temporal scales, several analytic frameworks have been applied — investigating the synchronisation of brain activity across participants, connectivity within and between participants, and the temporal complexity of the fMRI signals.



Figure 7.2. Naturalistic stimuli derived from three different movies including “Inscapes” (Vanderwal et al., 2015), “The Present” (Alexander et al., 2017), and “Ocean’s Eleven”. Top Panel: Inscapes (<http://headspacestudios.org/inscapes>) is a 7-min non-verbal, non-social series of slowly evolving abstract shapes with a piano score sound directed by Tamara Vanderwal. Middle Panel: The Present (<https://www.jacob-frey.com>) is a 4.18-min animated movie directed by Jacob Frey that portrays an amputated youth being gifted a disabled dog by his mother. Bottom Panel: Ocean’s Eleven is a 116-min Hollywood movie with an ensemble cast that covers a heist comedy plot. The movie was directed by Steven Soderbergh and released in 2001.

7.2 Understanding Brain Synchronisation, Functional Connectivity, and Scale-free Dynamics of the fMRI Signal Using Naturalistic Paradigms

There are several established and emerging methodological approaches to the detection of reproducible brain-behaviour associations within naturalistic paradigms. As for task-based and resting-state data, functional brain responses during naturalistic paradigms can be studied at multiple spatial and temporal scales by examining: (1) shared brain responses across individuals using intersubject correlation; (2) functional connectome derived from “stationary” and moment-

to-moment functional connectivity; and (3) scale-free dynamics of the fMRI signal indexed by the Hurst Exponent.

7.2.1 Functional Brain Synchronisation

One of the most common approaches to the analysis of naturalistic fMRI data is the intersubject correlation (ISC) framework, which examines shared brain responses across individuals (Hasson et al., 2010; Hasson et al., 2004). Amongst a set of individuals who are attending to the same stimulus (e.g., movies, audio clips), ISC indexes the extent to which their activity within specific brain regions and networks supporting information processing fluctuates synchronously in a time-locked manner (Hasson et al., 2010; Hasson et al., 2004). Specifically, when a participant A responds to a stimulus such as a movie at time t , the brain activity, denoted by $x_A(t)$, in voxel or ROI x , reflects a combination of three signals (Figure 7.3): $c(t)$ — stimulus-locked component that is consistent across participants; $id_A(t)$ — idiosyncratic stimulus-evoked component that is relevant to participant A such as memories and emotions; and $\epsilon_A(t)$ — noise component that is unrelated to the stimulus such as “mind-wandering”, head motion, or respiration. The fMRI signal in voxel or ROI x of participant A can then be derived as:

$$x_A(t) = \alpha_A c(t) + \beta_A id_A(t) + \epsilon_A(t) \quad (7.1)$$

If the activity in voxel or ROI x is correlated with activity in voxel or ROI x in participants B, C, D, and so forth, then the correlations of $c(t) \rightarrow 1$, and the other components including $id_A(t)$ and $\epsilon_A(t) \rightarrow 0$. Therefore, ISC depends on the relative contributions of the three signals. While ISC is a continuous index ranging from 0 to 1, an $ISC = 1$ denotes perfect synchronisation such that a region encodes information about the stimulus that is consistent across all the participants, whereas an $ISC = 0$ indicates no synchronisation (completely idiosyncratic), where a region encodes almost to no information about the stimulus that is consistent across participants. There are two primary reasons why ISC has been traditionally used to study shared brain responses in naturalistic fMRI conditions: (1) it can be used to filter individual-specific signals to reveal voxels or ROIs that are time-locked and consistent across participants; and (2) it is a model-free approach that quantifies brain responses in a natural environment without requiring *a priori* information about the timecourse of the stimulus unlike general linear models that require information such as stimulus features (e.g., faces, scenes, bodily movements, emotions) prior to defining the models (Finn et al., 2020; Nastase et al., 2019; Nummenmaa et al., 2018; Hasson et al., 2010; Jääskeläinen et al., 2008; Hasson et al., 2004).

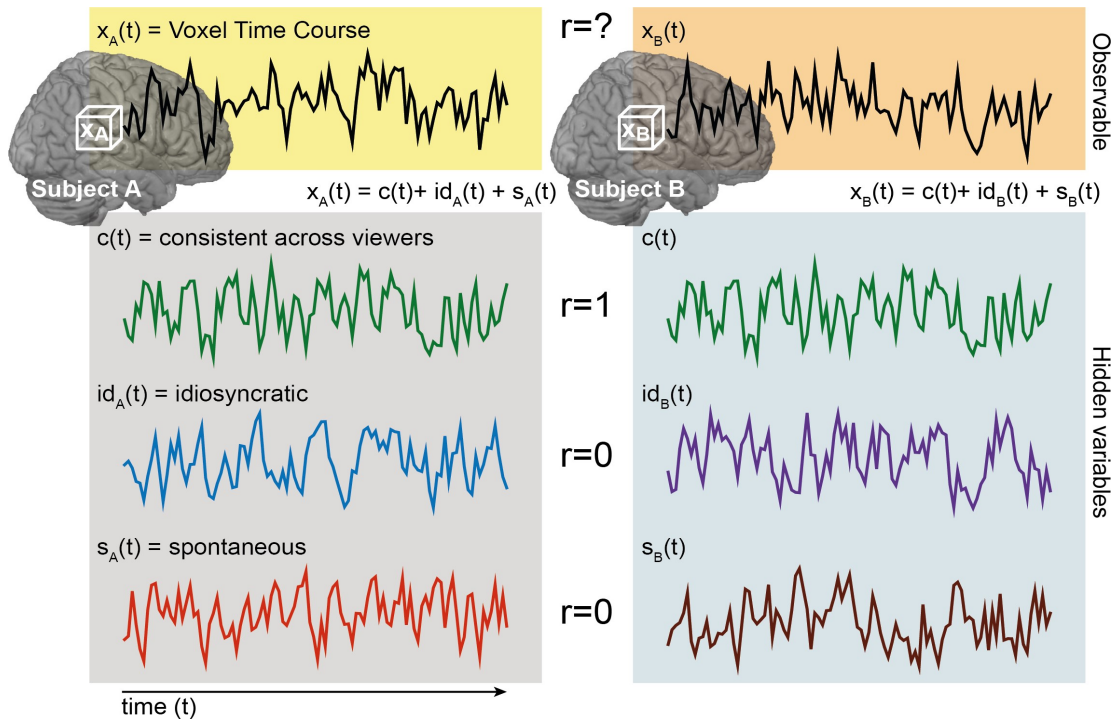


Figure 7.3. Illustration of ISC to quantify shared brain synchrony derived from the fMRI timeseries of a voxel x from a given participant A with the timeseries of voxel x in participant B. Adapted from Nastase et al., 2019.

ISC was first proposed to study the functional brain organisation under naturalistic fMRI conditions using a 30-min audiovisual feature film called “The Good, the Bad and the Ugly” released in 1966 that portrays the distinct use of violence, tension, and gunfights (Hasson et al., 2004). The individuals showed highly synchronised cortical brain activity during movie-watching that correlated with emotional arousal scenes in addition to selective components (e.g., faces, buildings) of the movie. These findings opened the door to an exciting opportunity to study the synchronisation or similarity of brain responses across individuals using ecologically valid, rich, and engaging stimuli in younger typical and atypical populations. For example, ASD adults showed idiosyncratic cortical brain responses that are distributed across primary sensory cortices and higher-order association areas during natural viewing of a Western feature film (Hasson et al., 2009). While these individuals revealed variable ISC patterns that were weaker relative to the non-ASD participants during the course of the movie, the greater variability amongst ASD participants was suggested to reflect sensory overload (Hasson et al., 2009). Further ISC studies revealed that both ASD children and adults exhibited idiosyncratic responses in several brain networks (e.g., default mode, limbic, medial frontal) compared to neurotypical individuals, consistent with differences in information processing underlying social cognition and comprehension in natural environments (Lyons et al., 2020; Nummenmaa et al., 2018; Byrge et al., 2015; Salami et al., 2013).

In early childhood (4-11 years), ISC has been applied to examine synchronous neural activity during free watching of the educational television show, “Sesame Street”, such that children with more, relative to less mature cognition and verbal abilities, exhibited stronger ISCs with adults (Cantlon et al., 2013). These findings suggest that ecologically valid paradigms can enhance our understanding of how children process information in a real-world environment. Extending movie-watching paradigms, brain synchrony has been captured in cortical and subcortical auditory networks that reflect working memory and attention functions as well as medial frontal regions that are involved in motor abilities during natural music listening (Abrahams et al., 2013). This suggests that brain regions outside the auditory system are time-locked to the features of the musical stimulus (e.g., structure, rhythms).

The data-driven ISC approach has also been employed to study the synchronisation of emotional brain responses using short clips from Hollywood feature films such as “When Harry Met Sally” and “The Godfather” to evoke positive, negative, or neutral emotions during adulthood (25-49 years) (see Saarimäki, 2021 for a review; Nummenmaa et al., 2012). As emotional valence decreases from positive to negative, the shared brain responses increase in regions that contribute to emotion processing in addition to areas that correspond to the default mode functions (Nummenmaa et al., 2012). On the other hand, emotional arousal tends to be positively associated with ISC in the somatosensory, attention, and visual networks, suggesting that emotional valence and arousal have distinct patterns of synchronised brain activity. Relevant to the goals of **Chapter 8**, ISC has been applied to examine both brain synchrony and idiosyncrasy in major depressive disorder (MDD) during adolescence and adulthood. Individuals suffering from melancholic MDD (22-63 years) exhibited diminished synchrony in sensory association areas and higher-order regions during naturalistic viewing of positive and negative emotionally evocative movies in addition to a significant reduction in the ventromedial prefrontal cortex (vmPFC) — a region that plays an important role in emotion regulation — during negative movie watching (Guo et al., 2015). The melancholic MDD individuals further showed weaker ISC in most functional networks, including the dorsal attention and salience networks, whose timecourses peaked during the most emotional “disgusting” scene. This suggests that melancholia is associated with disrupted brain responses to negative emotional stimuli (Guo et al., 2015). More recently, a study investigated the spatiotemporal patterns of brain synchrony and their relationships with emotional processing in adolescent depression (7-21 years) (Gruskin et al., 2020). Using an emotionally evocative movie known as “Despicable Me” from the HBN, Gruskin et al. (2020) showed that the synchrony of brain activity across depressed adolescents exhibited a negative relationship with depressive symptomatology, with more severe symptoms associated with weaker ISC. However, adolescents who exhibited similar depressive symptom profiles also exhibited more similar brain responses during the emotional movie viewing. The same study characterised the

temporal dynamics of brain synchronisation and suggested that the relationship between brain synchrony and depressive symptomatology may be modulated by moment-to-moment responses to the emotional valence of the movie amongst individuals suffering from the most severe symptoms (Gruskin et al., 2020).

Applied to other clinical conditions, ISC has also been used to understand how paranoia modulates neural responses during natural audio story-telling in a sample of adults (19-35 years) with subclinical paranoia symptoms (Finn et al., 2018). While widespread brain synchrony was captured in higher-order regions across these participants, the high-paranoia participants exhibited more shared neural responses compared to the low-paranoia participants in the left temporal pole and right medial prefrontal cortex (mPFC) — regions that are associated with processes such as theory of mind, mentalising, mental states, and intentions (Finn et al., 2018). Finally, the application of ISC has further been extended to study dynamic or moment-to-moment brain responses across individuals using short animated movies (Gruskin et al., 2020; Di and Biswal, 2020). Using a short Pixar movie called “Partly Cloudy”, highly consistent patterns of dynamic connectivity were identified across individuals in the visual, supramarginal gyrus, and default mode related areas, regions that are implicated in theory of mind, emotional processes, and understanding the representation of the world — the narratives of the movie (Di and Biswal, 2020).

7.2.2 Naturalistic Functional Connectome

Moving beyond brain synchronisation across participants, standard functional connectivity analyses can be applied to naturalistic stimuli to understand the interactions of functional networks supporting sensory, memory, emotional, and cognitive processes (Song et al., 2021; Chen et al., 2017; Baldassano et al., 2017; Simony et al., 2016) and to examine the impact of different stimulus conditions on fingerprint identifiability (Vanderwal et al., 2021; Vanderwal et al., 2017). Movie-watching stimuli have been shown to improve CPM-based brain-behaviour predictions for cognition and emotion, relative to resting-state conditions (Finn and Bandettini, 2021b). While these authors revealed that movies portraying high social content are better predictors of brain-behaviour relationships, the improvements can also be attributed to certain underlying features; for example, movies that presented more human, verbal, and social interactions positively improved the brain-behaviour predictions whereas movies that portrayed natural scenes and landscapes negatively impacted the predictions. Naturalistic movie stimuli have also been shown to yield consistently high *inter*-individual variability in functional connectivity patterns in the frontoparietal networks across conditions (i.e., rest, movie) followed by moderate variability in the heteromodal cortex, and lowest variability in the sensorimotor

cortices (Vanderwal et al., 2017). This explains why naturalistic stimulation boosts the identifiability of the functional connectome, even achieving perfect fingerprint ID accuracy (Vanderwal et al., 2017).

Natural audio stimuli such as music and audiobooks have been used to examine the density and topology of functional connectivity networks (Brauchli et al., 2020). Several networks showed diminished functional connectivity in the vmPFC, default mode regions, and temporal areas, suggesting that the degree of connectivity is mediated by acoustic pitch, musical processing, and auditory processing (Brauchli et al., 2020). Naturalistic fMRI paradigms have also been applied in developmental populations (Di and Biswal, 2022a; Vanderwal et al., 2019; Vanderwal et al., 2017; Emerson et al., 2015). For example, 6-year old children have demonstrated an uncoupling, i.e., decreased functional connectivity, between the visual and dorsal attention networks followed by less-developed “adult-like” brain organisation with the interactions of the dorsal attention, frontoparietal, and default mode networks during movie-watching (Emerson et al., 2015). This suggests that the functional brain in early childhood undergo dynamic reorganisation that reflect maturation processes and brain plasticity.

While the majority of the naturalistic functional connectomes have been studied in a “static” or stationary manner, more recent efforts have characterised dynamic or moment-to-moment neural responses over the course of a movie (Yang et al., 2023; Di et al., 2022b; Meer et al., 2020). Naturalistic movie stimuli provide an exciting platform to characterise brain state dynamics that co-occur across individuals during viewing of specific scenes that are linked to several behavioural profiles such as emotion processing, sensorimotor functions, and inhibition (Meer et al., 2020). Transitions between brain states tend to be synchronised across participants over the course of the movie, in accordance with narratives or story-telling segments. Moreover, dynamic functional connectivity can provide temporal insights that cannot be captured by “static” connectivity-based frameworks. For example, Di et al. (2022b) showed that individuals exhibit widespread moment-to-moment responses that are consistent across participants during movie-watching. However, these dynamic patterns may be dependent on the context of the movie as different movies produce distinct dynamic functional connectivity patterns. This critical information cannot be captured by “static” functional connectivity (Di et al., 2022b) — the authors further showed that stationary neural responses did not differ across movies.

In an innovative examination of dynamic functional connectivity, Yang et al. (2023) employed a natural language processing framework to a full 2-hour Hollywood movie known as “Forrest Gump” which was released in 1994 to further characterise brain state dynamics and the dwell time (i.e., time spent in a given brain state). While dwell times appear to differ across individuals

and region/network combinations, the default mode network yielded the longest time across brain states and participants which was significantly different compared to the visual network. Although variability in dwell times could reflect subject-dependent cognitive ability to process the narratives of the movie, there is also the possibility that brain state dynamics differ between higher-order (e.g., default mode and hippocampus) and lower-order (visual and hippocampus) networks (Yang et al., 2023). Yet another potential advantage of dynamic, relative to static functional connectivity is greater sensitivity to age and gender effects, as shown in a sample spanning childhood to early adulthood (5-22 years) (Di et al., 2023). This study showed that dynamic functional connectivity could capture non-linear age effects and gender differences in regional brain activity and moment-to-moment functional connectivity. However, the authors of that study noted that stationary and dynamic functional connectivity may capture different characteristics of a chosen behavioural phenotype — for example, only stationary functional connectivity predicted FSIQ in the developmental cohort (Di et al., 2023).

7.2.3 Scale-free Properties of the fMRI BOLD Signal

Spontaneous fluctuations in the fMRI signal exhibit properties of scale-free dynamics in time and space that reflect the temporal complexity or self-similarity structure of the fMRI timeseries. While the term temporal complexity refers to whether fMRI signals are temporally redundant and predictable based on the degree of correlation, self-similarity refers to the geometrical similarity between elements of an event regardless of a scale (Omidvarnia et al., 2022; Expert et al., 2011; Bullmore et al., 2009; Eke et al., 2000). For example, a geometric fractal structure can be partitioned into smaller windows and each window can be scaled down by a given factor. When the scaling maps the original structure onto the partitioned structure, it becomes self-similar. In the frequency domain, the scale-free dynamics is related to a power law of the power spectrum in the form of $1/f$ like: $P \propto f^\beta$, where P is the power, f denotes the frequency, and β is the power law exponent for which $f \rightarrow 0$ and $\beta > 0$ (Campbell and Weber, 2022; Ciuciu et al., 2012; Expert et al., 2011; He, 2011; Maxim et al., 2005). One of the most popular measures to characterise the scale-free properties of the fMRI timeseries is the Hurst Exponent (H ; Eke et al., 2000; Hurst, 1951). H is a continuous measure ranging from 0 to 1 such that an H which is closer to 0.5 indicates more random behaviour whereas an H which is nearer to 1 indicates more persistent fluctuations (Eke et al., 2000). In other words, an $H < 0.5$ indicates that the timeseries has a “short-term memory” and fast return to the mean whereas an $H > 0.5$ represents “long-term memory” and slower return to the mean. In a scenario where $H = 0.5$, the timeseries demonstrates Gaussian random white noise (e.g., Brownian motion), where past scale-free dynamics of a structure do not influence future dynamics of the same structure, and therefore, is not predictable. H further indexes whether a timecourse exhibits positively autocorrelated [$0.5 < H < 1$] or negatively

autocorrelated [$0 < H < 0.5$] behaviours. In theory, there is a linear relationship between β and H such that $\beta = 2H - 1$ for some scale-free processes such as the fractal Brownian motion. Hence, the higher the H of a structure, the more correlated the fMRI timeseries becomes, reflecting $1/f$ processes and interpreting as being more self-similar. Note: In the fMRI literature, several definitions have been proposed to explain the behaviours that a timeseries reflects as indexed by H including self-similarity, autocorrelation, long-term memory, persistent fluctuations, and temporal complexity. To maintain consistent terminology in this thesis, the terms self-similarity and temporal complexity have been used to describe the scale-free properties of the fMRI timeseries.

Recently, H has been employed to index excitation-inhibition (E:I) balance driven by glutamatergic and GABAergic signalling. E:I balance is thought to be disrupted in neuropsychiatric disorders, reflecting changes in glutamatergic and GABAergic signalling, leading to weaker SNR in neural circuits and impaired information processing (Yao et al., 2022; Trakoshis et al., 2020; Sohal and Rubenstein, 2019; Fogaça and Duman, 2019; Selten et al., 2018; Gao et al., 2017; Rubenstein and Merzenich, 2003). In the frequency domain, E:I can be captured from the power law of the power spectrum as excitatory glutamate and inhibitory GABA transmissions exhibit differential temporal profiles in neuronal signals. While increasing E:I ratio leads to a “flatter” $1/f$ slope, decreasing E:I ratio produces a more negative $1/f$ slope (Gao et al., 2017). It is understood that reduced H indicates increased excitation whereas greater H corresponds to increased inhibition (Trakoshis et al., 2020), and E:I balance has been broadly implicated in information transfer, neural computation, and homeostasis (Gao et al., 2017).

H has been examined under resting-state and task-based conditions, and more recently, naturalistic stimulation, in typical and atypical populations. Resting-state and task-based fMRI studies have shown that the temporal complexity of the fMRI signal is highest in brain regions that form part of several functional networks including the default mode, frontoparietal, and visual networks. Although fractal properties are present across the whole brain, the temporal complexity of the fMRI signals are non-uniform, i.e., H varies in gray matter, white matter, and cerebrospinal fluid tissues (Omidvarnia et al., 2022). An important observation that follows is that the self-similarity structure of fMRI signals derived from task-induced conditions has been shown to decrease with task difficulty and novelty (Churchill et al., 2016; Ciuciu et al., 2014; Ciuciu et al., 2012; He, 2011; Barnes et al., 2009). He (2011) showed that H was reduced during a button pressing task compared to resting-state conditions. Upon completing a task, H may return to its initial temporal state, suggesting that any potential deviations could be indicative of physiological and functional changes (Campbell and Weber, 2022a). Moreover, the scale-free properties of fMRI signals derived from task-based paradigms that require more cognitive effort

experience longer recovery times to their pre-task Hs, after returning from a higher frequency dominant state (Barnes et al., 2009). Such phenomena could reflect the processing of online information — a system that processes online information in an activated state and maintains the memory to prepare for future events during a baseline state (Campbell and Weber, 2022a; He, 2011). Recently, a movie-watching paradigm has been shown to exhibit higher temporal complexity at rest; higher fractal dynamics have been observed in the dorsal attention, somatomotor, and visual networks, whereas lower fractal dynamics have been found in the frontoparietal and default mode networks (Campbell et al., 2022b). Although there is only one study, to our knowledge, which has studied the temporal complexity of the naturalistic fMRI signals, evidence shows that the TRT reliability of H observed from resting-state and naturalistic conditions is good ($0.60 < ICC < 0.74$) (Campbell et al., 2022b). The presence of scale-free dynamics during naturalistic processing, therefore, suggests plausible (re) configurations of brain states as the brain integrates sensory information with the narrative unfolding over the course of the movie (Campbell et al., 2022b).

The $1/f$ -like dynamics of the fMRI signals have been studied initially in participants suffering from early onset of Alzheimer's Disease (69-85 years) such that abnormal temporal complexity has been found in brain regions that are implicated in the disease (Maxim et al., 2005). Promisingly, H may also be used to identify patients with mild cognitive impairment vs. neurotypical individuals, with a classification accuracy of 83.1% (Long et al., 2018). Across the lifespan, normal ageing has been associated with both losses and gains of signal complexity across the whole brain; reduced Hs have been revealed mainly in the frontal and parietal regions whereas greater Hs have been found in the insular, limbic, and temporal areas (Dong et al., 2018; Wink et al., 2006).

The scale-free dynamics of the default mode network have been shown to relate to personality traits such as extraversion, with greater H corresponding to lower extraversion scores during adulthood (18-29 years; Lei et al., 2013). Another line of investigation has revealed that depressive symptom severity is associated with reduced self-similarity signal structure of the right frontoparietal and default mode networks, with opposite association patterns in the left frontoparietal, ventromedial prefrontal, and salience networks (Jiang et al., 2017; Wei et al., 2015; Wei et al., 2013). Extending this finding, the temporal complexity of the default mode network has been shown to mediate the association between adaptive emotion regulation (e.g., cognitive appraisal) and depressive symptomatology in young adults (21.1 ± 1.58 years) (Gao et al., 2018). Highly impulsive adults (19-28 years) exhibited alterations in the signal complexity of the fronto-hippocampal pathways that are sensitive for learning, reward, and memory processing (Neufang and Akhrif, 2020). For example, deviations in H of the anterior cingulate cortex are associated

with response inhibition in the right hippocampus as well as reward processing in the dorsolateral prefrontal cortex and middle frontal gyrus (Neufang and Akhrif, 2020).

H has proven to be informative in developmental populations too. Children suffering from mild traumatic brain injury (13.4 ± 2.3 years) demonstrated a reduction in the temporal complexity of the fMRI signal of the basal ganglia, hippocampus, and cerebellar regions — supporting memory and cognitive dysfunctions that relate to neurophysiological changes that occur in the hippocampus (Dona et al., 2017b). Reduced signal complexity has also been observed in ASD children (12.7 ± 2.4 years), with the lowest fractal dimensions found in the left amygdala and left nucleus accumbens (Dona et al., 2017a). Further, using H as an index of the synaptic E:I ratio, Trakoshis et al. (2020) showed that ASD females expressed higher H, which was associated with a stronger ability to socially “camouflage” as a means to behaviourally compensate for social and communicative difficulties. Taken together, H is an emerging index in naturalistic fMRI that may reflect fundamental neuronal functioning such as E:I balance, thus showing promise as a biomarker for a wide range of clinical conditions.

Throughout this thesis, I have tested a number of methodological innovations aimed at improving the reproducibility of brain-behaviour associations in developing populations. In the next Chapter, I will take a different approach — I will examine whether ecologically valid paradigms such as movies can help address some of the limitations of resting-state and task-based conditions as elaborated in **Sections 7.1 and 7.2**. **Chapter 8** describes a set of analyses of movie-watching data aimed at examining whether information processing during an emotionally evocative naturalistic stimulus is modulated by depressive symptomatology and emotion regulation strategies in a community-based sample of female adolescents experiencing subclinical depressive symptoms.

Chapter 8

Naturalistic emotional processing in adolescent females with depressive symptoms

Paper (In Preparation)

Ramduny, J., MacSweeney, N., Petkova, E., Kellaghan, E., Lahert, N., Gallagher, L., Vanderwal, T., & Kelly, C. Naturalistic emotional processing in adolescent females with depressive symptoms.

Authors

Jivesh Ramduny^{1,3}, Niamh MacSweeney⁴, Eva Petkova^{5,6}, Elisabeth Kellaghan¹, Nessa Lahert¹, Louise Gallagher^{7,8}, Tamara Vanderwal^{9,10}, Clare Kelly^{1,2,3}

¹School of Psychology, Trinity College Dublin, Dublin, Ireland

²Department of Psychiatry, School of Medicine, Trinity College Dublin, Dublin, Ireland

³Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland

⁴Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, United Kingdom

⁵Department of Child and Adolescent Psychiatry, New York University, New York, NY, USA

⁶Department of Population Health, New York University, New York, NY, USA

⁷Child and Youth Mental Health Collaborative, University of Toronto, Toronto, ON, Canada

⁸Division of Child and Youth Mental Health, University of Toronto, Toronto, ON, Canada

⁹Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada

¹⁰BC Children's Hospital Research Institute, Vancouver, BC, Canada

Authorship Contribution Statement

Jivesh Ramduny: Conceptualisation, Methodology, Formal analysis, Investigation, Writing — original draft, review & editing. **Niamh McSweeney:** Data Collection, Methodology. **Eva Petkova:** Conceptualisation, Methodology. **Elisabeth Kellaghan:** Data Collection, Methodology. **Nessa Lahert:** Data Collection, Methodology. **Louise Gallagher:** Conceptualisation, Methodology. **Tamara Vanderwal:** Conceptualisation, Investigation, Writing — original draft, review & editing. **Clare Kelly:** Conceptualisation, Funding, Methodology, Formal analysis, Investigation, Writing — original draft, review & editing, Supervision, Project administration.

Abstract

Adolescence is a sensitive developmental period during which the brain undergoes substantial maturational changes. Heightened brain plasticity during this critical period may increase vulnerabilities to influences that precipitate later mental health difficulties such as anxiety and depression, particularly amongst females. The movie-watching fMRI paradigm has gained popularity, due to the opportunities it offers to interrogate shared and idiosyncratic neural responses as the storyline develops. In the present study, I used an emotionally evocative movie to study brain function and brain-behaviour relationships at multiple spatial and temporal scales in a community-based sample of 50 female adolescents experiencing subclinical depressive symptoms. First, intersubject correlation analyses revealed patterns of brain synchrony that differed between high- and low-emotion segments of the movie, and which were globally modulated by depressive symptom severity and use of adaptive emotion regulation strategies. Second, analyses of dynamic ISC revealed changes in intersubject synchronisation over the course of the movie. Third, an exploratory analysis of dynamic functional connectivity between edges of an emotion regulation network demonstrated time-varying brain-behaviour relationships. Finally, the self-similarity (fractal) structure of the fMRI signal was associated with depressive symptomatology and adaptive emotion regulation at the network and ROI level. While my findings require replication in an independent, well-powered sample, they suggest that depressive symptomatology and emotion regulation strategies may shape how female adolescents experiencing subclinical depressive symptoms process emotional information during affect-laden events that reflect real-life scenarios.

8.1 Introduction

As outlined in **Chapter 1 (Section 1.8)**, the last aim of this thesis is to take an alternative approach to reproducibility: by examining whether a naturalistic fMRI paradigm — an emotionally evocative movie — can enhance individual differences in emotional responses in a sample of female adolescents experiencing subclinical depressive symptoms.

Adolescence is a sensitive developmental period during which substantial maturational change and enhanced brain plasticity appear to increase vulnerability to influences that can precipitate mental health difficulties such as depression, particularly amongst females (Sisk and Gee, 2022; Saluja et al., 2004; Kessler et al., 1993). During this critical phase, adolescents have a heightened risk of experiencing intense negative emotions in their daily life (Young et al., 2019; Zimmerman and Iwanski, 2014; Silk et al., 2003), increasing demands for emotion regulation. Emotion regulation is defined as the processes through which an individual modulates their emotions to change the magnitude and/or type of their emotional experiences (Schäfer et al., 2016; Gross, 1998; Thompson, 1994). Individual differences in emotion regulation abilities or strategies may be linked with risks for depressive symptoms (Schäfer et al., 2016; Ahmed et al., 2015; Zimmerman and Iwanski, 2014; Larsen et al., 2012); better appreciating such individual differences may improve our understanding of risk and resilience, and suggest targets for intervention.

An emotion regulation strategy can be categorised as either *adaptive* or *maladaptive*; more frequent engagement of adaptive emotion regulation strategies is associated with fewer depressive symptoms, while more frequent engagement of maladaptive emotion regulation strategies is associated with increased depressive symptomatology (Young et al., 2019; Domaradzka and Fajkowska, 2018; Schäfer et al., 2016; Aldao et al., 2010). Studies aimed at characterising the neural underpinnings of emotion regulation strategies have shown that adaptive emotion regulation to reappraise sad feelings and negative stimuli (e.g., cognitive appraisal) recruits prefrontal areas that form part of an emotional regulation network, and depressed individuals exhibit perturbations in these circuits (Keller et al., 2022; Ebneabbasi et al., 2021; Erk et al., 2010; Johnstone et al., 2007; Beauregard et al., 2006; Etkin et al., 2015). Further, depressed individuals tend to engage more with maladaptive emotion regulation strategies (e.g., self-blame, rumination, suppression) and exhibit aberrant corticolimbic activity during social isolation (Jankowski et al., 2018), loss error processing (Willinger et al., 2022), and negative emotional memories (Visted et al., 2018; Joormann and Gotlib, 2010). Here, I was interested in extending these investigations by examining the neural correlates of emotion regulation under ecologically rich and valid conditions that echo real-world scenarios (see **Chapter 7**).

This aim aligns with a recent shift in fMRI research, towards naturalistic paradigms such as movie-watching. As outlined in **Chapter 7 (Section 7.2)**, using naturalistic movie stimuli during fMRI elicits time-locked responses that are shared across participants and are thought to provide ecologically valid indices of social, emotional, and cognitive processing (Eickhoff et al., 2020; Vanderwal et al., 2019; Vanderwal et al., 2015; Hasson et al., 2010; Hasson et al., 2004). Conventionally, naturalistic fMRI paradigms have permitted the examination of shared neural responses across participants, in the form of ISC (Natase et al., 2019; Nummenmaa et al., 2018; Hasson et al., 2010; Jääskeläinen et al., 2008; Hasson et al., 2004; see **Chapter 7, Section 7.2.1**). This data-driven approach has previously been employed to study the synchronisation of emotional brain responses during a movie paradigm in neurotypical adults (Nummenmaa et al., 2012), adults suffering from melancholic depression (Guo et al., 2015), and depressed adolescents (Gruskin et al., 2020). For example, Gruskin et al. (2020) showed that the shared brain responses across depressed adolescents revealed a negative association with depressive symptomatology — the youths exhibiting more severe depressive symptoms showed weaker brain synchronisation. Recent efforts have also characterised the temporal dynamics of brain synchronisation in adolescent depression. The relationship between brain synchrony and depressive symptom severity has been shown to be modulated by the dynamic responses of the youths who experience the more severe symptoms to the emotional valence of the naturalistic movie stimulus (Gruskin et al., 2020). Here, I build on these investigations by also examining the relationship between brain synchrony and emotion regulation strategies.

In addition to ISC, movie-watching fMRI data offer an interesting opportunity to examine both static and dynamic functional connectivity patterns (Vanderwal et al., 2021; Song et al., 2021; Finn and Bandettini, 2021b; Wang et al., 2017; Vanderwal et al., 2017; Finn et al., 2017; Chen et al., 2017; Baldassano et al., 2017; Simony et al., 2016; Geerlings et al., 2015b; see **Chapter 7 (Section 7.2.2)**). Studies examining functional connectivity during movie-watching stimuli that elicit emotion regulation have revealed associations with functional connectivity within the salience (Saarimäki, 2021; Jääskeläinen et al., 2021), frontoparietal (Berkman and Lieberman, 2006), and limbic (Jääskeläinen et al., 2021) networks in addition to other emotion-specific brain areas consisting of the ventrolateral prefrontal cortex and amygdala (Torre and Lieberman, 2018; Kohno et al., 2015). These regions have also been shown to exhibit altered functional connectivity patterns in depression during the regulation of positive and negative emotions (Kaiser et al., 2015; Siegle et al., 2007). To date, these examinations have focused on static functional connectivity — whether depressive symptoms are associated with altered functional connectivity dynamics during emotional processing remains an important open question.

Measures of ISC and functional connectivity during naturalistic fMRI provide information about the degree to which stimulus-evoked activity is synchronised across large-scale networks within and between individuals. There is increasing interest in examining the characteristics of this activity itself, however, since this can provide insights into the integrity of the stimulus-evoked processing underlying ISC and functional connectivity. Spontaneous and stimulus-linked brain activities can be characterised in terms of their scale-free dynamics which reflect the self-similarity or temporal complexity of the BOLD fMRI signal (Omidvarnia et al., 2022; Expert et al., 2011; Bullmore et al., 2009; Eke et al., 2000). One such index is the Hurst Exponent (H), which has been applied to characterise fMRI signal complexity with depressive symptom severity and adaptive emotion regulation such as cognitive appraisal (see **Chapter 7 (Section 7.2.3)**). While one recent study investigated the scale-free dynamics of the fMRI signal using naturalistic movie stimuli during adulthood (Campbell et al., 2022), H during movie-watching fMRI has not yet been studied during development or in association with depressive symptoms, which could reveal new information about signal dynamics that underlie vulnerabilities to depressive symptomatology and emotion regulation strategies under ecologically valid conditions.

Together, these lines of investigation demonstrate that naturalistic movie-watching data offer a powerful means for the examination of brain functions supporting emotional processing and emotion regulation at multiple spatial and temporal scales. Here, I employ this approach to examine the relationships between depressive symptom severity and emotion regulation strategies and intersubject synchrony, individual functional connectivity, and brain activity dynamics during an emotionally evocative movie in a community-based sample of 50 female adolescents reporting subclinical depressive symptoms. As many as two-thirds of those who experience depressive symptoms during adolescence receive a clinical diagnosis of MDD by their early 30s (Klein et al., 2009). Investigating the neural correlates of individual differences in emotion regulation strategies within this at-risk group may improve our understanding of risk and resilience factors, and suggest targets for intervention aimed at averting the development of the full clinical phenotypes.

The aims of the present study are three-fold: to (1) investigate whole-brain static and dynamic ISC during naturalistic emotional viewing, and explore the relationships of each with self-reported depressive symptom severity and emotion regulation strategies; (2) examine static and dynamic functional connectivity within an *a priori* emotion regulation network; and (3) characterise the underlying stimulus-evoked brain activity using the H, and explore its associations with the phenotypic measures at the network and ROI level. First, I hypothesised that patterns of ISC would be modulated by depressive symptomatology and emotion regulation strategies and these relationships would vary over the course of the movie. Second, I expected

that edges connecting key nodes of the emotion regulation network would also exhibit both static and dynamic relationships with the phenotypic measures. Third, I hypothesised a loss of signal complexity in the functional networks and regions that form part of the emotion regulation network. Finally, I hypothesised that the self-similarity signal structure of these networks and regions would share an association with depressive symptom severity and emotion regulation strategies.

8.2 Methods & Materials

8.2.1 Participants

Sixty female adolescents, approximately half of whom reported significant but subclinical depressive symptoms, were recruited from the community and a community-based Child and Adolescent Mental Health Services clinic in Dublin, Ireland (Figure 8.1). The School of Psychology Research Ethics Committee (Trinity College Dublin), Research Ethics Committee at Linn Dara Child & Adolescent Mental Health Service, and the Saint James's Hospital/Tallaght University Hospital Joint Research Ethics Committee approved the study procedures. Written consent was provided by participants who were at least 18 years old, and written assent and parental consent were obtained for those under 18 years of age.

Adolescents were not included if they had been diagnosed with any comorbidities related to autism, psychosis, or other pervasive developmental disorders. They were also excluded if they had received psychopharmacological treatment in the preceding 6 months or had a Full-Scale IQ < 80. High-motion participants who had a mean FD > 0.30 mm during the movie-watching fMRI scan were excluded from the subsequent analyses (Vanderwal et al., 2021; Power et al., 2012). This mean FD threshold differs from the one (i.e., rmsFD > 0.20 mm) applied in **Chapters 2, 4, and 6** to maximise participant inclusivity. A total of 50 adolescents (age range = 13-20 years, mean age $\pm$ SD = 17.0 $\pm$ 1.9 years) remained in this study. The demographic characteristics of the youth sample are presented in Table 8.1.

Variable	Sample (<i>N</i> = 50)	Mean (SD)	Range
Age (years)	—	17.0 (1.9)	13-20
13-17	25	—	—
18-20	25	—	—
Gender (Female)	50	—	—
Maladaptive CERQ*	—	29.2 (8.4)	15-49
Self-blame*	—	9.68 (3.6)	4-17
Rumination*	—	12.2 (3.9)	5-20
Catastrophizing*	—	7.32 (2.4)	4-12
Adaptive CERQ‡	—	11.8 (4.4)	4-20
Positive Reappraisal‡	—	11.8 (4.4)	4-20
BYI-D	—	17.4 (10.4)	2-45
Mean FD	—	0.14 (0.05)	0.08-0.30

Table 8.1. Demographic Characteristics. CERQ = Cognitive Emotion Regulation Questionnaire (*Maladaptive Strategy; ‡Adaptive Strategy). BYI-D = Depressive Inventory from Beck Youth Inventory. FD = Framewise Displacement.

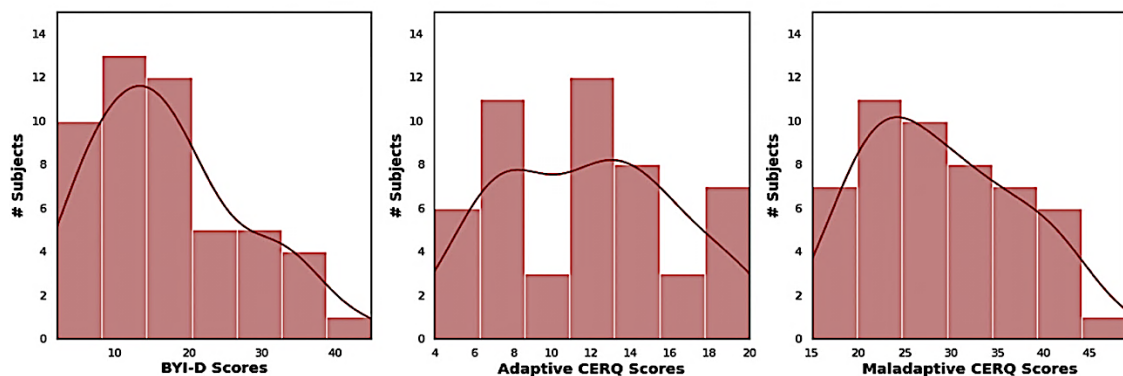


Figure 8.1. Phenotypic distributions. Histograms showing the distributions of the depressive symptom severity derived from Depression subscale of the Beck Youth Inventory (BYI; the Depression Inventory, BYI-D), adaptive cognitive emotion regulation (Adaptive CERQ) scores and maladaptive cognitive emotion regulation (Maladaptive CERQ) scores. The skewed distributions of the phenotypic measures indicate that they are not normally distributed (Shapiro-Wilk test, $P < 0.05$).

8.2.2 Self-Reported Depressive Symptomatology

The Depression subscale of the Beck Youth Inventory (BYI; the Depression Inventory, BYI-D) was administered to measure self-reported depressive symptom severity (Beck et al., 2005). The BYI-D subscale contains 20 items which refer to feelings, thoughts, and behaviours related to depressive symptoms experienced in the preceding 2 weeks. The BYI-D score is calculated by summing the items from the inventory, each of which is answered using a 4-point scale ranging from 0 [“never”] to 3 [“always”]. The higher the BYI-D score, the greater the severity of depressive symptoms. Hence, the severity of the depressive symptoms that a female youth may

experience can be categorised as average [$\text{BYI-D} < 17$], mildly elevated [$17 \leq \text{BYI-D} \leq 21$], moderately elevated [$22 \leq \text{BYI-D} \leq 31$], and extremely elevated [$\text{BYI-D} \geq 32$].

8.2.3 Self-Reported Emotion Regulation Strategies

The Cognitive Emotion Regulation Questionnaire (CERQ) was administered to the participants to assess their emotional responses following stressful or negative events in their daily lives (Garnefski et al., 2001). The CERQ is a multidimensional questionnaire that characterises nine coping strategies. Each subscale consists of 4 items (36 items in total). For each item, participants indicated the extent to which they used the coping strategies after experiencing a negative event on a 5-point Likert scale ranging from 1 [“almost never”] to 5 [“almost always”]. A CERQ subscale score is then generated for each participant by summing up the individual scores from the 4 items ranging from 4 to 20; the higher the subscale score, the more frequently that specific cognitive coping strategy is used. Of the nine coping strategies, four are identified as maladaptive strategies (i.e., self-blame, other-blame, rumination, catastrophising) and the remaining five are categorised as adaptive (i.e., acceptance, positive reappraisal, refocus on planning, putting into perspective, positive refocusing) (Domaradzka and Fajkowska, 2018; Aldao et al., 2010). Given the negative associations between adaptive strategies such as positive reappraisal and psychopathology (Garnefski and Kraaij, 2007; Garnefski et al., 2004; Garnefski and Kraaij, 2003), and positive associations between maladaptive strategies such as self-blame, rumination, and catastrophising and psychopathology (Garnefski and Kraaij, 2007; Garnefski et al., 2004; Garnefski and Kraaij, 2003), I derived global adaptive and maladaptive CERQ scores for each participant by summing up the items from the respective subscales.

8.2.4 Naturalistic Stimulus

The participants underwent an fMRI scan during which they watched a short movie called “The Present” (<https://www.jacob-frey.com>). The Present is an animated film which describes a poignant story featuring a 12-year-old boy who struggles with emotion regulation and self-image in the context of having only one leg. The narrative arc of the film captures changes in the youth’s emotions and emotion regulation around the difficult topic of being differently abled. The film also features a broad range of themes such as frustration, screen-time, disgust, self-image, humour, empathy, and attachment (Figure 8.2). This naturalistic stimulus has recently been used to evoke emotional responses in a psychiatrically enriched developmental sample derived from the HBN (Alexander et al., 2017). The movie lasted 4 minutes 18 seconds in total, with the main story taking place during the first 3 minutes 23 seconds, followed by the ending credits (55 seconds). Since the most significant emotional moment of the short movie takes place during the final 87 seconds, I analysed the entire timeseries to account for the haemodynamic delay in the BOLD fMRI signal, and enduring emotional processing and regulation.

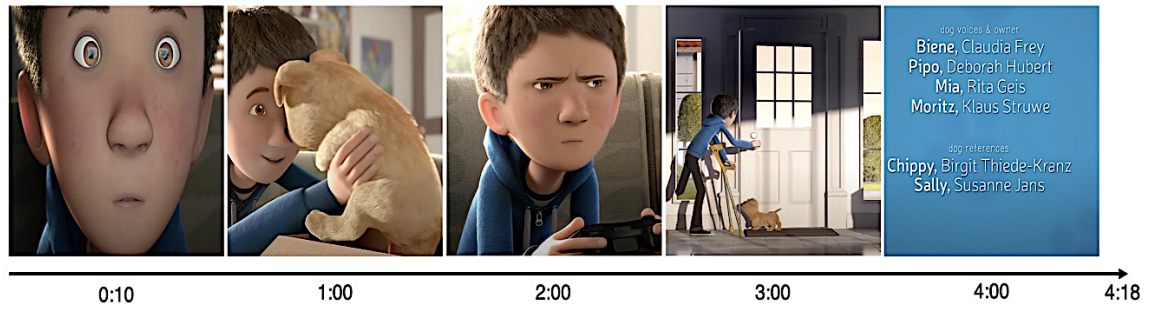


Figure 8.2. Movie stills obtained from *The Present* (<https://www.jacob-frey.com>).

8.2.5 MRI Data Acquisition

Structural and functional scans were acquired using a Philips Achieva 3T scanner from the Trinity College Institute of Neuroscience which is located at Trinity College Dublin. A T1-weighted MPRAGE volume with a spatial resolution of $0.9 \times 0.9 \times 0.9 \text{ mm}^3$ (FOV = 230 mm, TR = 8.4 ms, TE = 3.8 ms, flip angle = 8°) was obtained for each participant. The participants underwent resting-state and movie-watching fMRI scans although the resting-state data were not used in the present study. The movie-watching fMRI data were acquired with 132 echo-planar imaging (EPI) volumes (37 slices, TR = 2000 ms, TE = 30 ms, voxel size = $3.2 \times 3 \times 3 \text{ mm}^3$).

8.2.6 MRI Preprocessing Pipeline

The structural and functional scans were preprocessed using the fMRIPrep version 5.1.0 pipeline (<https://fmriprep.org/en/stable/>; Esteban et al., 2019), and in-house scripts using FSL (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL>), and AFNI (<https://afni.nimh.nih.gov>) — which is similar to the pipeline applied in **Chapters 2, 4, and 6**. The structural data underwent bias field correction, skull stripping, tissue segmentation (i.e., gray matter, white matter, cerebrospinal fluid), and spatial normalisation of the anatomical data to the MNI152 standard space using non-linear transformation. Preprocessing of the functional data comprised volume-based motion correction and functional to anatomical co-registration using boundary-based registration (Greve and Fischl, 2009). Confound regression was performed with 24 motion parameters (i.e., 3 translational and 3 rotational parameters, their first derivatives, and the squares of each of these terms) and 12 nuisance signal regressors (i.e., cerebrospinal fluid, white matter, global signal, their squares, and the squares of their first derivatives). Spatial smoothing was applied to the functional data with a Gaussian kernel at 6 mm FWHM followed by temporal filtering ($0.01 \leq f \leq 0.1 \text{ Hz}$).

8.2.7 Intersubject Correlation Analyses

I first performed whole-brain full-length and dynamic ISC analyses to quantify the shared brain responses across participants during naturalistic viewing of the emotionally evocative movie

(Figure 8.3; see **Chapter 7 (Section 7.2.1)**). I then assessed the relationships between ISC and each of the phenotypic measures (i.e., BYI-D, adaptive CERQ, maladaptive CERQ) to determine whether the degree of brain synchrony is related to depressive symptom severity and emotion regulation strategies across participants.

8.2.7.1 Intersubject Correlation

I used a whole-brain parcellation (<https://identifiers.org/neurovault.image:395091>; Shen et al., 2013) to extract the mean timeseries across 268 parcels for each participant. Pairwise correlations were computed between the timecourses of all participant pairs using Pearson's correlation coefficient (r) to generate a 50 x 50 ISC matrix for each parcel [$r_{AB} = r(x_A, x_B)$, $r_{AC} = r(x_A, x_C)$, ... where $A \neq B, C$]. I derived the mean of the lower triangle of the ISC matrices to summarise the degree of brain synchrony across the participants. Hypothesis testing was then performed to obtain the significance of the ISC at each parcel using a non-parametric approach as proposed by Chen et al. (2016). To obtain a null distribution, a subject-wise bootstrapping method was used to randomly sample participants *with* replacement to generate a new ISC matrix, and this procedure was repeated 5,000 times. The statistical significance of ISC was adjusted for multiple comparisons (across 268 parcels) by applying an FDR threshold corrected at $P < 0.05$.

8.2.7.2 Difference in ISC between low and high emotion segments of the movie

Given that the emotional content evolves over the course of the movie, two epochs of equal length were identified which contained the high and low emotion contents. The high-emotion segment started at 2 minutes 51 seconds and was immediately preceded by the low-emotion segment, from 1 minute 24 seconds. Each epoch comprised 45 volumes. It is important to note that the segmentation on the basis of emotional content was not performed using independent raters — a step that should be taken in further research. Accordingly, it is possible that that emotion ratings are similar or even stronger in other parts of the movie. I computed the pairwise correlation between all participants' timecourses for each parcel using the low-emotion and high-emotion segments as previously described (see **Section 8.2.7.1**). For each parcel, I subtracted the ISC_{low} matrix from the ISC_{high} matrix to obtain the difference in ISC. The statistical significance of the difference in ISC was assessed using a non-parametric approach (see **Section 8.2.7.1**) and FDR corrected at $P < 0.05$.

8.2.7.3 Relationships between ISC and phenotypic measures

To examine relationships between ISC and the phenotypic measures, I computed the correlations (Spearman's ρ) between full-length ISC and each of the phenotypic measures. ISC was computed as described in **Section 8.2.7.1** by computing the pairwise correlations between the timecourses of a given participant and the remaining participants using Pearson's correlation coefficient (r)

for each parcel. Hypothesis testing was achieved by computing the significance of the Spearman's ρ using a non-parametric approach (see **Section 8.2.7.1**). To obtain a null distribution, a permutation method was applied which randomly shuffled one of the vectors for every parcel to compute the Spearman's ρ between ISC and each of the phenotypic measures, and this procedure was repeated 5,000 times. Statistical significance was adjusted for multiple comparisons (across 268 parcels) using an FDR corrected threshold at $P < 0.05$. I also assessed the consistency of these relationships across the whole brain by obtaining the observed mean ρ which represents the correlation coefficients between ISC and each of the phenotypic measures averaged across 268 parcels. I further derived the mean ρ from the null distribution of each permutation test and performed two-tailed significance testing as previously described by Gruskin et al. (2020).

8.2.7.4 Dynamic Intersubject Correlation (dISC)

To assess changes in ISC during the movie, I quantified moment-to-moment changes in intersubject synchronisation by computing ISC for each parcel using a tapered cosine sliding window approach. A tapered cosine window prevents all the timepoints in a window from having equal weights, especially those that are at the boundary and extreme of the window (Gruskin et al., 2020; Preti et al., 2017). Here, dynamic intersubject correlation (dISC) was calculated by setting the width of the tapered window to $W = 20$ TRs with a step size of $W = 1$ TR (tapered cosine fraction = 0.95), thus producing 112 ISC maps. Parameter selection was performed by referencing previous studies which reported that such W offers a good trade-off between detecting reliable dISCs with long enough W and capturing transient dISCs with short enough W (Gruskin et al., 2020; Di and Biswal, 2020; Trost et al., 2015; Nummenmaa et al., 2012). To extract functional networks that capture dynamic intersubject synchrony during movie-watching, I performed ICA using FSL's MELODIC (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/MELODIC>; Smith et al., 2004). The ICs correspond to the spatial maps that are statistically independent and non-Gaussian. I concatenated the 112 ISC maps temporally to create a single 4D image and then allowed ICA to automatically decompose the maps into ICs. The ICs underwent visual inspection to ensure that they did not relate to physiological noise and their Z-statistic maps were thresholded at $Z > 2$. I further assessed the relationships between dISC and each of the phenotypic measures for each IC by extracting the mean dISC across each of the ICs for each participant, followed by the same procedures for testing the brain-behaviour relationships as described earlier (see **Section 8.2.7.3**). This approach allowed me to evaluate whether patterns of temporal dynamics of brain synchrony covary with depressive symptom severity and/or emotion regulation strategies. The statistical significance of the Spearman's ρ was adjusted for multiple comparisons based on the number of ICs using an FDR threshold at $P < 0.05$.

8.2.8 Functional Connectivity Analyses

After performing the whole-brain ISC analyses, I proceeded to perform a series of functional connectivity analyses which focused on an *a priori* emotion regulation network (Figure 8.3). I assessed both static and dynamic functional connectivity patterns derived from the emotion regulation network, and examined their relationships with the phenotypic measures.

8.2.8.1 Functional Connectivity

I defined a set of *a priori* ROIs for emotion regulation using NeuroSynth (Yarkoni et al., 2011). I queried the term “emotion regulation”, which resulted in an automated meta-analytic map for 247 studies, spanning 10 ROIs (<https://neurosynth.org/analyses/terms/emotion%20regulation/>). These ROIs include the following: bilateral amygdala [R_Amyg; L_Amyg], bilateral dorsolateral prefrontal cortex [R_dIPFC; L_dIPFC], left dorsal premotor cortex [L_dPMC], left ventrolateral prefrontal cortex [L_vIPFC], left rostral anterior cingulate cortex [L_rACC], left anterior insula [L_aINS], right orbitofrontal cortex [R_OFC], and right anterior prefrontal cortex [R_aPFC]. For each participant, I extracted the mean timeseries of each ROI and computed the Pearson’s correlation coefficient (r) between all possible ROI pairs to construct a 10 x 10 functional connectivity matrix. I then vectorised the 50 functional connectivity matrices by selecting their respective lower triangular part to produce a 50 x 45 functional connectivity matrix. Using these functional connectivity matrices, I performed edge-level correlations using Spearman’s ρ to assess the relationships between functional connectivity involving the ROIs and each of the phenotypic measures. The statistical significance of the relationships was derived using the same non-parametric approach as previously described (see **Section 8.2.7.3**). A permutation method was employed to randomly shuffle one of the vectors for every edge to calculate the Spearman’s ρ between functional connectivity and each of the phenotypic measures, and this procedure was repeated 5,000 times. To control for multiple comparisons (across 45 edges), the significance of the Spearman’s ρ was FDR corrected at $P < 0.05$.

8.2.8.2 Dynamic Functional Connectivity (dFC)

I computed dFC for the same 10 ROIs using the same $W = 20$ TRs tapered cosine sliding window approach as described earlier (see **Section 8.2.7.4**). I performed edge-level correlations for each window using Spearman’s ρ to assess the associations between functional connectivity and each of the phenotypic measures. The statistical significance of the associations was obtained using a non-parametric procedure as previously described (see **Section 8.2.7.4**). For every W , a permutation method was used to randomly shuffle one of the vectors to calculate the Spearman’s ρ between dFC and each of the phenotypic measures, and this procedure was repeated 5,000 times. Given that the sliding windows overlap with each other (i.e., they are not independent), the significance of the Spearman’s ρ was corrected for multiple comparisons across 45 edges using

an FDR threshold corrected at $P < 0.01$. To determine the connections that facilitate significant associations with the phenotypic measures over the course of the movie, I retained the edges that consistently revealed such relationships in at least 5 consecutive windows.

8.2.9 Hurst Exponent Analyses

Finally, I was interested in examining the self-similarity structure of the movie-watching fMRI signals given that these signals may exhibit abnormal autocorrelated patterns as a result of depressive symptom severity (Jing et al., 2017; Wei et al., 2015; Wei et al., 2013). I employed H to assess the self-similarity of the BOLD fMRI timeseries at the network and ROI level (Figure 8.3; see **Chapter 7 (Section 7.2.3)** for more details about the properties of H).

8.2.9.1 Network-based H

H was first computed using the `hurst` package in Python (<https://github.com/Mottl/hurst>) at the network level using the network definition derived from the Yeo-7 Network scheme (Yeo et al., 2011) which includes the following: (1) ventral attention; (2) dorsal attention; (3) frontoparietal; (4) default mode; (5) somatomotor; (6) limbic; (7) visual; and (8) extending the subcortical network consisting of the bilateral hippocampus and amygdala. To derive H for each functional network, I averaged the respective timeseries to obtain a 50×132 matrix. I computed Spearman's ρ between H of each network and each of the phenotypic measures while covarying for age (Gao et al., 2018; Lei et al., 2013). To further control for multiple comparisons (across 8 networks), the significance of the Spearman's ρ was FDR corrected at $P < 0.05$ for each of the phenotypic measures.

8.2.9.2 ROI-based H

H was also computed at the ROI level using the same `hurst` package in Python with the *a priori* regions obtained from NeuroSynth as defined in **Section 8.2.8.1**. I computed the Spearman's ρ between H of each ROI and each of the phenotypic measures while treating age as a covariate (Gao et al., 2018; Lei et al., 2013). To control for multiple comparisons (across 10 ROIs), the significance of the Spearman's ρ was FDR corrected at $P < 0.05$ for each of the phenotypic measures.

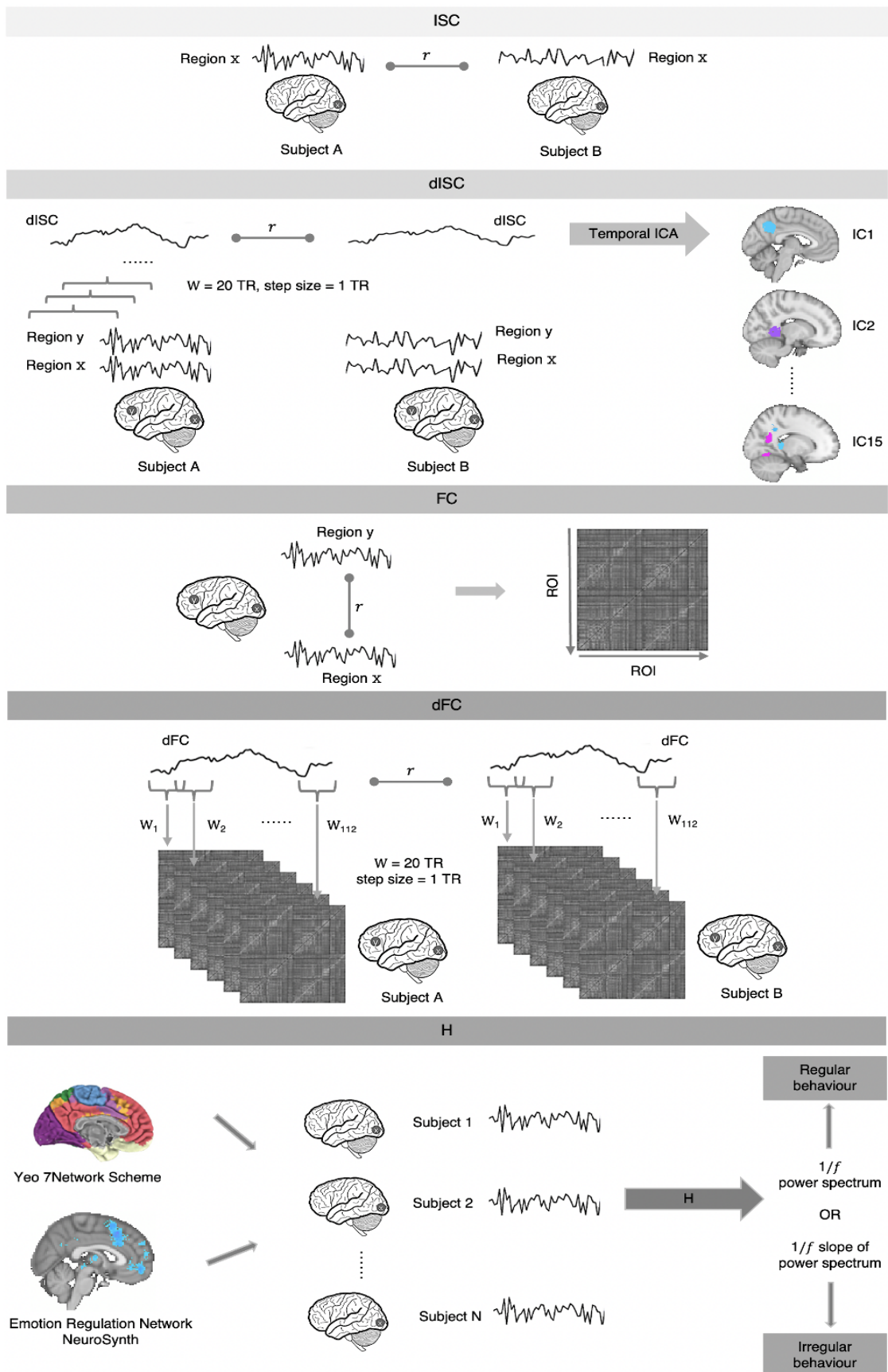


Figure 8.3. Movie-watching analysis framework at multiple spatial and temporal scales. First, whole-brain full-length intersubject correlation (ISC) was performed by calculating the

pairwise correlations between participant A and the remaining participants B, C, and so on such that $A \neq B, C$. Second, dynamic intersubject correlation (dISC) was carried out using a tapered cosine sliding window with a width of $W = 20$ TRs and step size of $W = 1$ TR (tapered cosine fraction = 0.95) to capture time-varying patterns of brain synchrony throughout the movie. Temporal ICA was then applied to extract functional networks that capture dynamic patterns of synchronised brain activities over the course of the movie. Third, functional connectivity (FC) was performed between *a priori* ROIs that form part of an emotion regulation network derived from NeuroSynth. Fourth, dynamic functional connectivity (dFC) was derived using the same tapered cosine sliding window procedure as previously described to compute the FC-phenotypic correlations over the course of the movie. Lastly, the Hurst Exponent (H) was employed to index the self-similarity of the BOLD fMRI signal at the network level using the Yeo-7Network parcellation scheme and ROI level using the predefined ROIs that are involved in the emotion regulation network. Subsequently, the relationships between each of these measures and phenotypic scores (i.e., BYI-D, adaptive CERQ, maladaptive CERQ) were performed.

8.3 Results

8.3.1 Naturalistic emotional viewing reveals whole-brain intersubject synchronisation

I first characterised the degree of brain synchrony across participants by performing whole-brain full-length ISC. Natural viewing of “The Present” elicited substantial whole-brain intersubject synchronisation in 176/268 parcels (Figure 8.4A; $P < 0.05$ FDR). Whole-brain ISCs were similar to patterns observed in previous work with strong ISCs in the visual and somatosensory regions (Vanderwal et al., 2019; Vanderwal et al., 2015). Relatively weaker ISCs were observed in regions involved in social and emotional processing. Weak-to-moderate ISCs were also noted in the prefrontal and association cortices extending to the cerebellum.

Comparing ISC during high-emotion and low-emotion contents of the movie revealed significant differences for 39/268 parcels (Figure 8.4B; $P < 0.05$ FDR). ISCs were significantly weaker during the high- relative to low-emotion segment in the visual-related areas, primary somatosensory cortex, cerebellar structures, and default network regions, but stronger in the primary and supplementary motor areas in addition to brain regions that support language processing and face perception ($P < 0.05$ FDR).

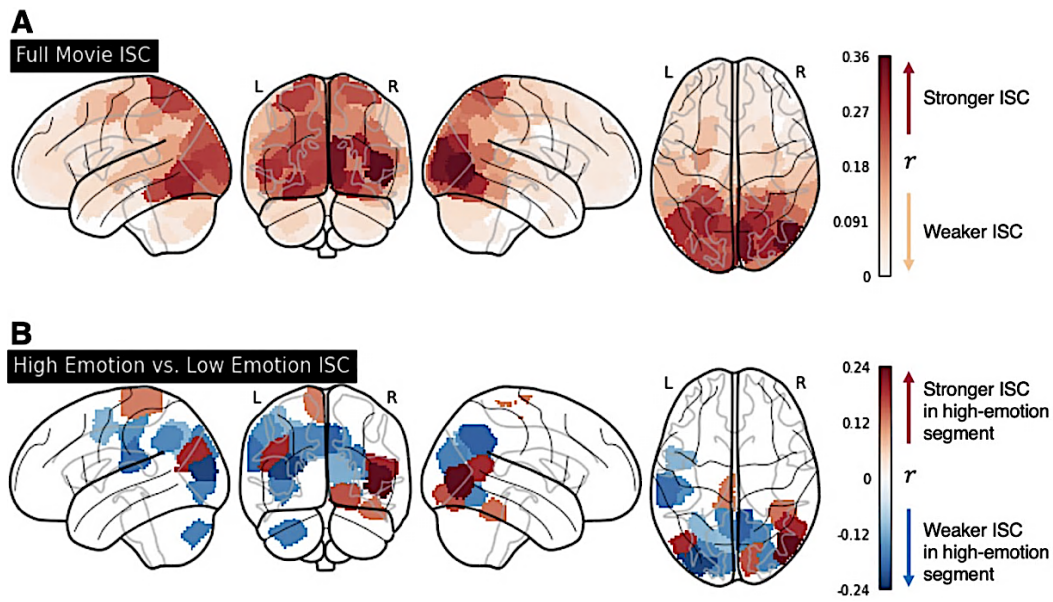


Figure 8.4. Intersubject synchronisation measured by full-length ISC. (A) Full-length ISC during the natural viewing of an emotionally evocative movie. (B) Full-length ISC during high-emotion segment relative to low-emotion segment of the movie. The ISC maps are corrected for multiple comparisons thresholded at $P < 0.05$ FDR.

8.3.2 Brain synchrony is related to depressive symptom severity and emotion regulation strategies during the emotionally evocative movie

There were no significant associations between ISC and self-reported depressive symptomatology and emotion regulation strategies that survived the FDR threshold at $P < 0.05$ (across 268 parcels). At an uncorrected threshold ($P < 0.01$), greater depressive symptom severity was related to weaker ISC in 1/268 parcel, i.e., left anterior prefrontal cortex ($\rho = -0.43$, $P < 0.01$ uncorrected; Figure 8.5A). Stronger adaptive CERQ scores were related to increased ISC in 2/268 parcels (Figure 8.5A) — in the right anterior cingulate cortex ($\rho = 0.36$, $P < 0.01$ uncorrected) and right striatum ($\rho = 0.40$, $P < 0.01$ uncorrected). In contrast, stronger adaptive CERQ scores were related to weaker ISC in 3/268 parcels (Figure 8.5A) — in right posterior prefrontal cortex ($\rho = -0.37$, $P < 0.01$ uncorrected), left frontal pole ($\rho = -0.38$, $P < 0.01$ uncorrected), and left hippocampus ($\rho = -0.49$, $P < 0.01$ uncorrected). There was a negative association between ISC and maladaptive CERQ scores in 2/268 parcels, located in the right cerebellum ($\rho = -0.37$, $P < 0.01$ uncorrected; $\rho = -0.40$, $P < 0.01$ uncorrected) (Figure 8.5A).

While regional relationships between ISC and self-reported depressive symptomatology and emotion regulation strategies were weak, when I examined a global measure of ISC, averaged across all 268 parcels, I observed a significant (following permutation testing) negative relationship between depressive symptom severity and ISC, such that more severe symptoms were associated with weaker intersubject synchrony ($\rho = -0.04$, $P < 0.001$) (Figure 8.5B). An inverse relationship was also observed for adaptive emotion regulation strategies ($\rho = -0.06$, $P <$

0.001; Figure 8.5B), but the association between mean ISC and maladaptive emotion regulation strategies was not significant ($\rho = -0.006$, $P = 0.25$; Figure 8.5B).

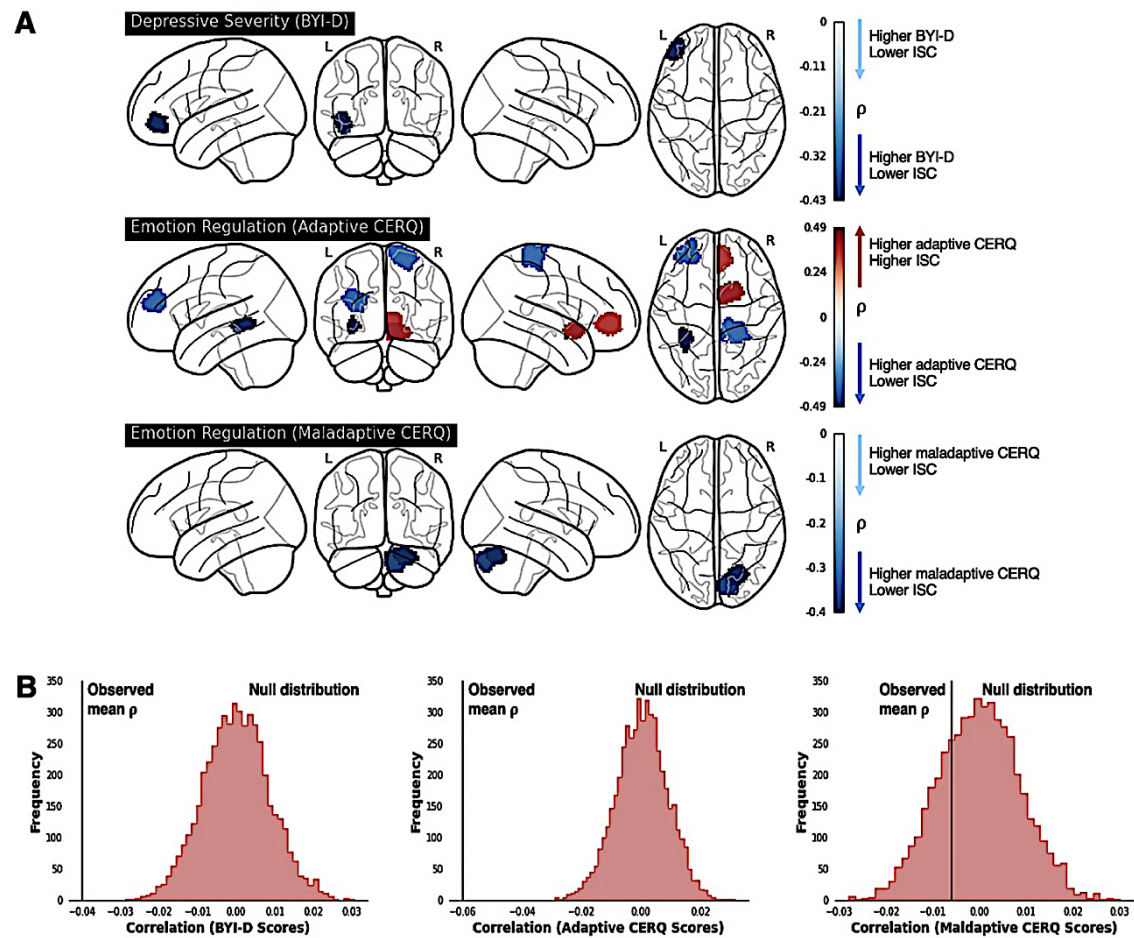


Figure 8.5. Relationships between intersubject synchronisation and phenotypic measures during emotional movie viewing. (A) Relationships between full-length ISC and self-reported depressive symptom severity (BYI-D; Top Panel), adaptive emotion regulation (Adaptive CERQ; Middle Panel) and maladaptive emotion regulation (Maladaptive CERQ; Bottom Panel). The relationships are displayed at an uncorrected threshold of $P < 0.01$. (B) Histograms illustrate the permutation test null distributions and observed mean ρ for relationships between the phenotypic measures and global ISC averaged across the whole brain (i.e., 268 parcels).

8.3.3 Temporal ICA captures dynamic intersubject synchronisation in several functional networks during naturalistic emotional viewing

To explore dynamic patterns of brain synchrony, I first computed ISC for each ROI using a tapered cosine sliding window approach, with 112 temporal windows spanning the length of the movie. Applying ICA to the resultant ISC maps revealed 15 networks exhibiting coordinated changes in intersubject synchronisation over time with an explained variance ranging from 4% to 9% (Figure 8.6A). These functional networks were largely located posteriorly and comprised regions of the default mode, visual, sensorimotor, and cerebellar networks (Figure 8.6B), sometimes in combination with the limbic regions including the hippocampus and thalamus.

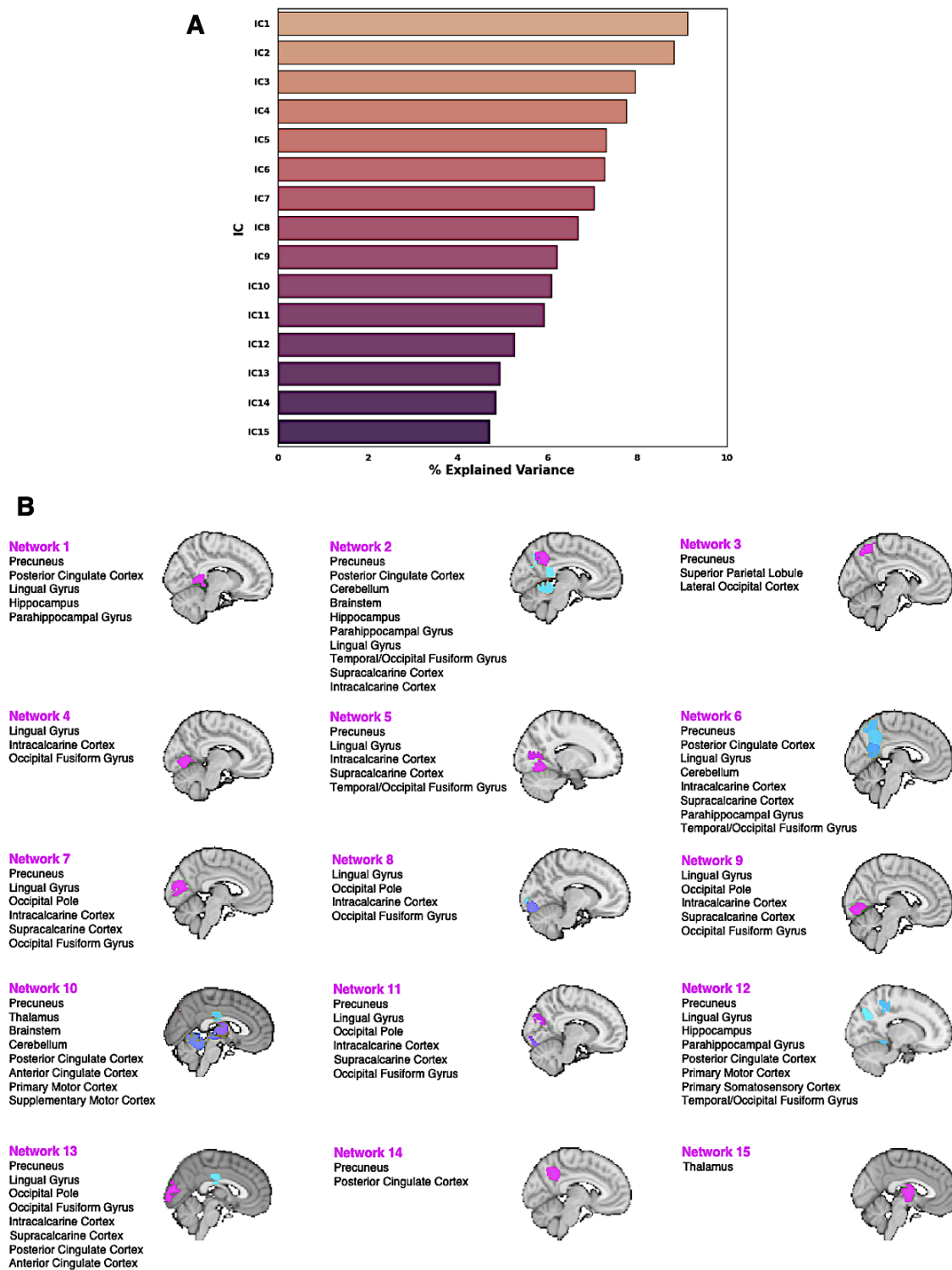


Figure 8.6. Temporal ICA shows several functional networks that capture dynamic intersubject synchronisation. (A) The percentage of explained variance in each IC is displayed in descending order than ranges from 4% to 9%. (B) The functional networks that correspond to each IC during naturalistic emotional viewing. The Z-statistic maps of the ICs are thresholded at $Z > 2$.

8.3.4 High emotion content evokes idiosyncratic responses during movie-watching

To assess changes in the degree of synchronisation within each of the functional networks described above throughout the movie, I plotted the ISC timecourses of the networks (i.e., across the 112 temporal windows; Figure 8.7). The high-emotion segment of the movie (previously defined from $TR \geq 67$, comprising 45 windows) includes the scenes where the young amputee empathises with a three-legged dog and decides to engage and play with the dog — illustrated in the top panel of Figure 8.7. Compared to the rest of the movie (67 windows), 4/15 functional networks (networks 1, 2, 10, 14) exhibited relatively decreased ISC during the high-emotion segment (Mann-Whitney U test; $P < 0.05$ FDR; Figure 8.7), suggesting that individuals expressed more idiosyncratic neural responses during this time. In contrast, 3 networks (networks 11, 13, 15), including the visual and default mode regions, became more synchronised during the high-emotion segment, compared to the rest of the movie (Mann-Whitney U test; $P < 0.05$ FDR; Figure 8.7).

I further investigated whether the timecourse of the functional networks covaried with the phenotypic measures (i.e., BYI-D, adaptive CERQ, maladaptive CERQ). After applying an FDR threshold to account for multiple comparisons across the 15 functional networks, there were no associations between dISC and the phenotypic measures that survived statistical significance at $P < 0.05$.

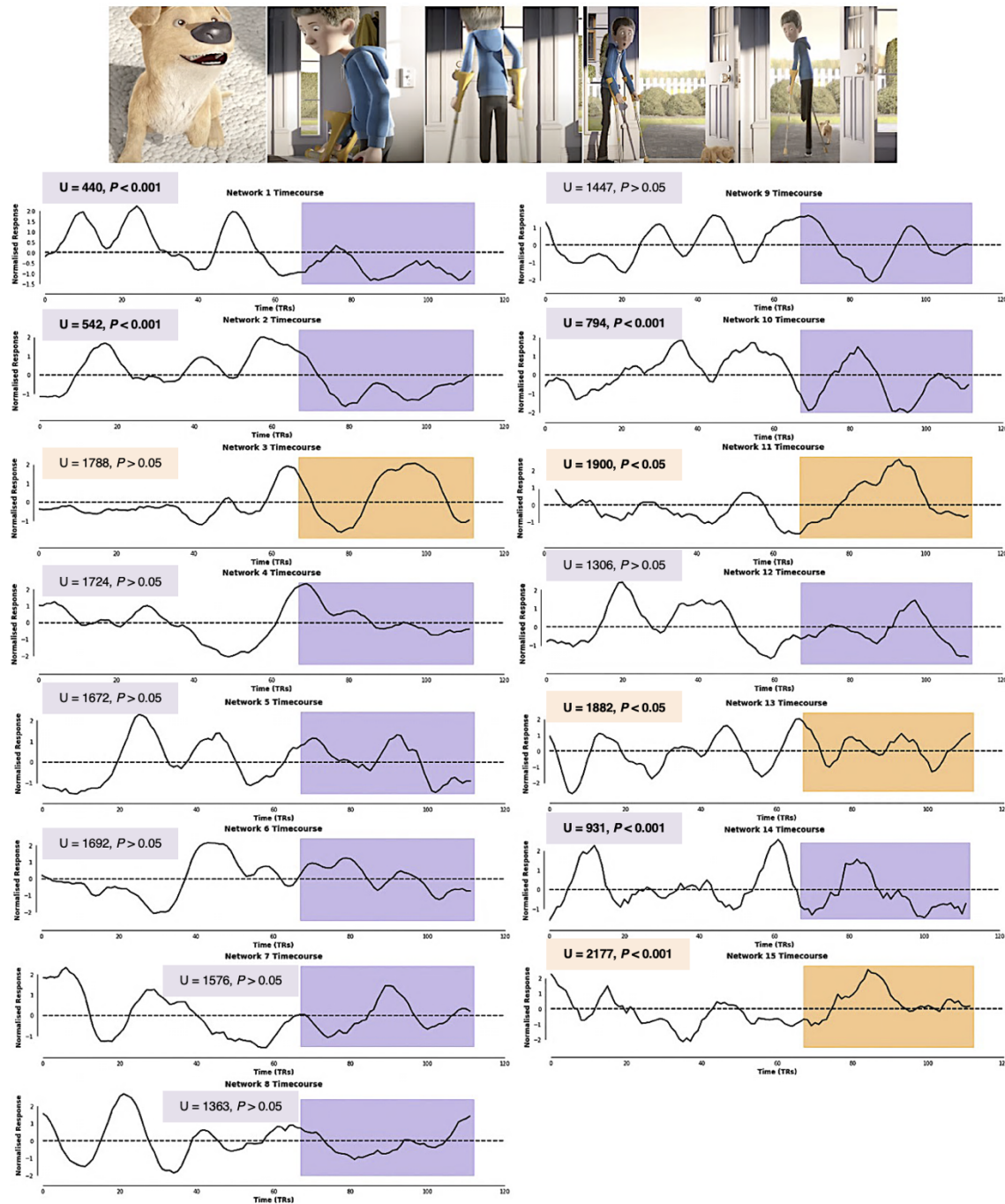


Figure 8.7. Timecourse of the functional networks derived during emotional movie watching. The x-axis represents the tapered sliding windows such that each window has a width of $W = 20\text{TRs}$ with a step size of $T = 1\text{TR}$. The y-axis illustrates the normalised response during each sliding window as measured by dISC. Solid black curves represent the timecourse of the functional networks. Dotted black line denote the mean dISC across the functional networks. Purple shades indicate the functional networks that exhibit idiosyncratic responses during high-emotion segment of the movie. Orange shades correspond to the functional networks that exhibit high synchrony during high-emotion segment of the movie. Some scenes from the high-emotion segment that relate to the timecourse of the functional networks are also displayed in the top panel. A Mann-Whitney U test was performed to quantify the increases and decreases in brain synchrony within each functional network throughout the movie.

8.3.5 Moment-to-moment functional connectivity shows edge-level correlations with self-reported depressive symptomatology and emotion regulation strategies during emotional movie-watching

Using the *a priori* ROIs derived from a NeuroSynth-defined emotion regulation network, I first computed the edge-level associations between static functional connectivity and each of the phenotypic measures. No edge-level relationships were observed, nor were any relationships detected for the global mean static functional connectivity across the 10 ROIs.

In a novel and exploratory approach to examining how brain-behaviour relationships unfold over the course of a movie, I identified edges that exhibited a relationship with a given phenotypic measure that persisted across at least 5 consecutive sliding windows. This analysis suggested the presence of relationships between dFC and BYI-D scores that changed over the course of the movie. There were 5/8 edges that showed a positive relationship between dFC and BYI-D scores — two of these relationships were observed during the low-emotion segment of the movie whereas the remaining were found during the high-emotion segment (Figure 8.8A): (1) L_vlPFC – L_rACC ($\rho = 0.40$, $P < 0.01$); (2) L_vlPFC – L_aINS ($\rho = 0.41$, $P < 0.01$); (3) L_Amyg – L_aINS ($\rho = 0.39$, $P < 0.01$); (4) L_dIPFC – R_aPFC ($\rho = 0.40$, $P < 0.01$); and (5) L_dPMC – L_dIPFC ($\rho = 0.40$, $P < 0.01$). There were 3/8 edges that exhibited a negative association between dFC and BYI-D scores during both low- and high-emotion segments of the movie (Figure 8.8A): (1) L_Amyg – R_OFC ($\rho = -0.43$, $P < 0.01$); (2) R_OFC – L_dPMC ($\rho = -0.41$, $P < 0.01$); and (3) R_Amyg – L_dPMC ($\rho = -0.40$, $P < 0.01$).

Interestingly, an edge connecting the L_dPMC and L_Amyg yielded a positive relationship between dFC and adaptive CERQ scores ($\rho = 0.41$, $P < 0.01$) at the beginning of the high-emotion segment of the movie (Figure 8.8B). Lastly, a total of 3 edges showed an association between dFC and maladaptive CERQ scores, where the first two edges were found during the low-emotion segment and the last edge was captured during the high-emotion segment (Figure 8.8C): (1) L_Amyg – R_OFC ($\rho = -0.43$, $P < 0.01$); (2) L_dPMC – R_dIPFC ($\rho = -0.38$, $P < 0.01$); and (3) L_dIPFC – R_aPFC ($\rho = 0.39$, $P < 0.01$).

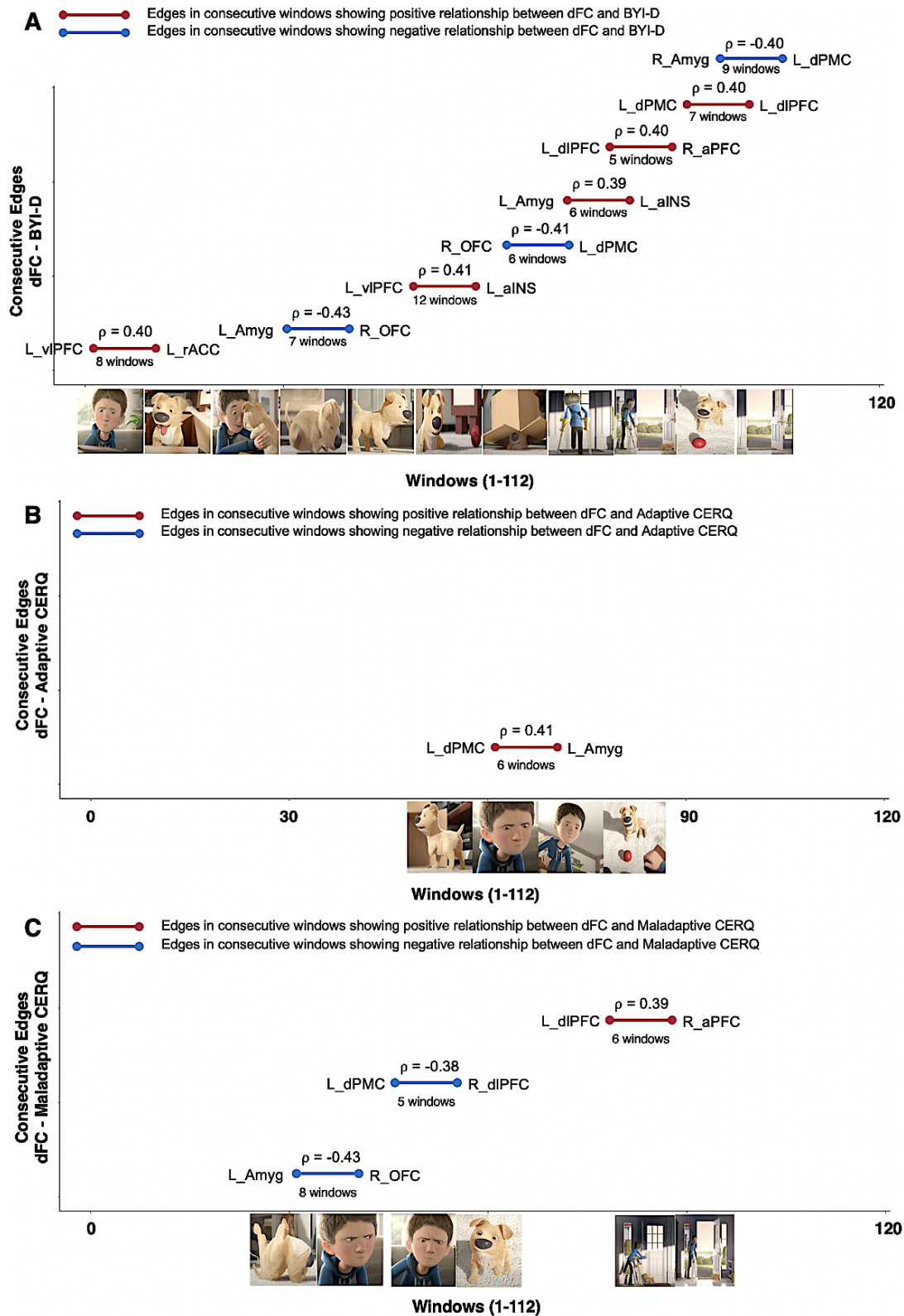


Figure 8.8. Edge-level correlations between dFC and phenotypic measures that persisted across at least 5 consecutive sliding window during the emotionally evocative movie. (A) Edges in consecutive windows that showed a significant relationship between dFC and BYI-D score. (B) Edges in consecutive windows that showed a significant relationship between dFC and adaptive CERQ score. (C) Edges in consecutive windows that showed a significant relationship between dFC and maladaptive CERQ score. Red line denotes the positive association between dFC and phenotypic measures derived from Spearman's ρ . Blue line represents the negative

association between dFC and phenotypic measures derived from Spearman's ρ . The number of consecutive edges that captured the significant relationships between dFC and phenotypic measures is also displayed under each edge. The significance of these relationships is thresholded at $P < 0.01$ for each window.

8.3.6 Self-similarity signal structure of the ventral attention network and ventrolateral prefrontal cortex is associated with depressive symptom severity and adaptive emotional regulation strategies during naturalistic emotional viewing

Finally, I computed H and examined its relationships with the phenotypic measures at the network and ROI level (Figure 8.9; Table 8.2; Table 8.3). A significant positive association between depressive symptom severity and H within the VAN was observed ($\rho = 0.43$, $P = 0.015$ FDR; Figure 8.9C; Table 8.2). Within the set of 10 *a priori* emotion regulation ROIs, the left vlPFC exhibited a significant positive relationship between depressive symptom severity and H ($\rho = 0.40$, $P = 0.039$ FDR) and a negative relationship between adaptive emotion regulation and H ($\rho = -0.40$, $P = 0.044$ FDR; Figure 8.9C; Table 8.3). Although maladaptive emotion regulation was positively related to the H of the left vlPFC (Figure 8.9C; Table 8.3), this relationship did not survive statistical significance after correcting for multiple comparisons ($\rho = 0.38$, $P = 0.075$ FDR).

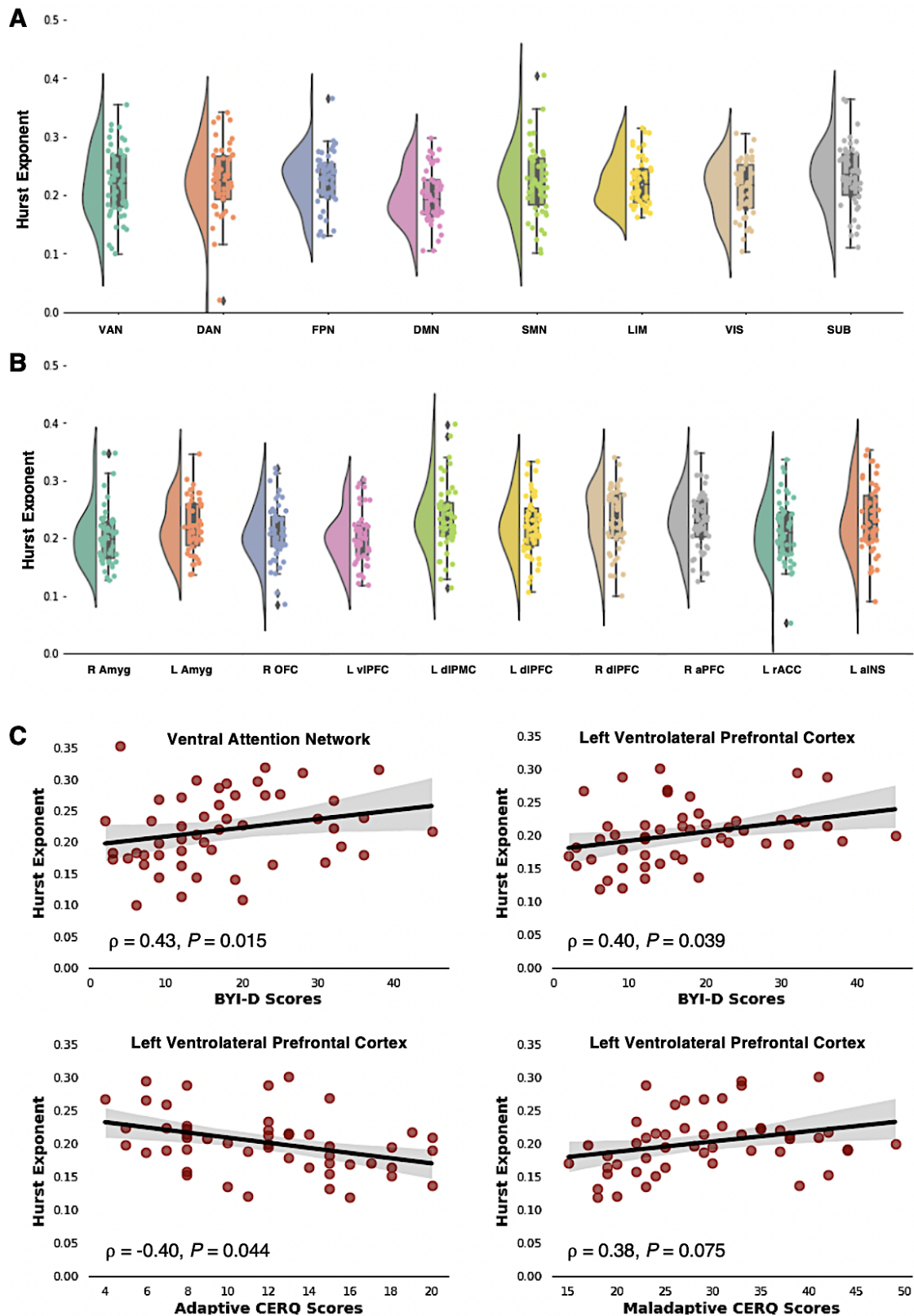


Figure 8.9. Relationships between H and phenotypic measures at the network and ROI level. (A) Distributions of H derived from the functional networks based on the Yeo-7Network parcellation scheme. (B) Distributions of H derived from the ROIs that form part of the emotional regulation network. (C) Top Panel: There is a significant positive association between H of the VAN and depressive symptom severity. There is a significant positive association between H of the vlPFC and depressive symptom severity. Bottom Panel: There is a significant negative relationship between H of the left vlPFC and adaptive emotional regulation strategies. The H of

the left vIPFC is positively related to maladaptive emotion regulation strategies but this relationship did not survive statistical significance. All the associations between H and phenotypic measures are corrected for multiple comparisons at the network and ROI level using an FDR threshold set at $P < 0.05$. Amyg: amygdala; OFC: orbitofrontal cortex, vIPFC: ventrolateral prefrontal cortex; dIPMC: dorsolateral premotor cortex; dIPFC: dorsolateral prefrontal cortex; aPFC: anterior prefrontal cortex; rACC: rostral anterior cingulate cortex; aINS: anterior insula. VAN: ventral attention network; DAN: dorsal attention network; FPN: frontoparietal network; DMN: default mode network; SMN: somatomotor network; LIM: limbic; VIS: visual; SUB: subcortical.

Networks	ρ_{BYI-D}	<i>P</i> value	$\rho_{adapCERQ}$	<i>P</i> value	$\rho_{maladapCERQ}$	<i>P</i> value
Ventral Attention	0.43	0.015	-0.33	0.18	0.38	0.057
Dorsal Attention	-0.22	0.43	-0.14	0.64	-0.16	0.62
Frontoparietal	0.032	0.98	0.11	0.70	-0.066	0.73
Default Mode	0.013	0.98	-0.054	0.80	0.058	0.73
Somatomotor	0.16	0.56	-0.20	0.59	-0.050	0.73
Limbic	-0.0037	0.98	0.067	0.80	-0.13	0.62
Visual	0.089	0.86	0.037	0.80	0.13	0.62
Subcortical	-0.21	0.43	-0.18	0.59	-0.26	0.30

Table 8.2. Relationships between H and phenotypic measures at the network level after correcting for multiple comparisons using an FDR threshold set at $P < 0.05$.

ROIs	ρ_{BYI-D}	<i>P</i> value	$\rho_{adapCERQ}$	<i>P</i> value	$\rho_{maladapCERQ}$	<i>P</i> value
L_Amyg	0.098	0.83	-0.037	0.92	-0.011	0.94
R_Amyg	-0.0035	0.98	0.024	0.92	-0.060	0.93
R_OFC	-0.023	0.98	-0.16	0.83	-0.045	0.93
L_vIPFC	0.40	0.039	-0.40	0.044	0.38	0.075
L_dIPMC	0.062	0.96	-0.14	0.83	-0.079	0.93
L_dIPFC	0.10	0.83	-0.071	0.92	0.12	0.93
R_dIPFC	0.14	0.83	-0.22	0.70	0.059	0.93
R_aPFC	-0.0043	0.98	0.089	0.92	0.029	0.93
L_rACC	0.22	0.70	0.015	0.92	0.14	0.93
L_aINS	-0.18	0.73	-0.049	0.92	0.070	0.93

Table 8.3. Relationships between H and phenotypic measures at the ROI level after correcting for multiple comparisons using an FDR threshold set at $P < 0.05$.

8.4 Discussion

Movie-watching fMRI has become an increasingly popular approach, due to higher in-scanner compliance and the ecologically valid nature of the stimuli. In the present study, I employed an emotionally evocative movie to investigate brain functions at multiple spatial and temporal scales,

and associations with key measures of depressive phenotypes in a community-based sample of 50 female adolescents experiencing subclinical depressive symptoms. First, ISC analyses revealed regionally specific patterns of brain synchrony that exhibited differential modulation during high- relative to low-emotion segments of the movie. Although these localised shared neural responses did not exhibit any phenotypic relevance after correcting for multiple comparisons, there were significant associations between global mean intersubject synchrony and both depressive symptom severity and adaptive emotion regulation. Second, novel analyses of dynamic ISC revealed changes in intersubject synchronisation within and between high- and low-emotion segments of the movie. Third, while static functional connectivity showed no associations with behaviour, an exploratory analysis of dynamic functional connectivity between edges of an emotion regulation network demonstrated time-varying brain-behaviour relationships. Finally, the self-similarity (fractal) structure of the VAN, and specifically, of the left vlPFC, was positively related to depressive symptom severity and negatively associated with adaptive emotion regulation.

8.4.1 Shared brain responses in adolescents experiencing depressive symptoms during free emotional viewing

The strongest degree of brain synchrony was evident in the temporal and occipital cortices during the emotional movie. These regions have been widely implicated in processing different aspects of a movie such as stimulus features, emotional information, and story-telling narratives. I observed stronger synchronisation in visual processing areas that are associated with face recognition and perception (Weiner and Zilles, 2016; Haxby et al., 2000; Tovée, 1995), audio-visual emotional processing (Pourtois et al., 2005), selective visual attention (Driver and Spence, 2000), encoding of environmental scenes (Aminoff et al., 2013), and visuospatial orientation and processing (Seghier, 2012; Husain and Nachev, 2007; Karnath, 1997; McFadzean et al., 1994). In contrast, weaker intersubject correlations were revealed in brain regions that are involved in emotional processing which echoed a previous finding of reduced ISC in melancholic depression (Guo et al., 2015). This distributed pattern of ISC may relate to the emotional valence of the movie, given that there were significant differences in brain synchrony in the high- relative to the low-emotion segments. Despite the emotional context of the movie, the majority of regions exhibiting significant ISC were located posteriorly in the brain — I did not observe significant ISC in the vmPFC, for example. This aligns with previous work suggesting that emotional processing circuits may process information related to an individual's interpretations or meaning of the stimulus, thus yielding idiosyncratic responses (and therefore weaker ISC) across participants (Chang et al., 2021). For example, the vmPFC, which is critical for emotion regulation, has been shown to exhibit weak cross-participant similarity compared to exogeneous regions (e.g., visual cortex) during movie-watching (Chang et al., 2021).

Examination of brain-behaviour relationships revealed two global effects — both depressive symptom severity and adaptive emotion regulation strategies were positively associated with the global mean of ISC across the whole brain. Contrary to expectations, however, the strength of ISC within the functional networks did not exhibit any phenotypic relevance for depressive symptoms or emotion regulation, after correcting for multiple comparisons. This differs from previous studies that have found that individuals exhibiting greater depressive symptomatology expressed weaker intersubject correlations in the anterior prefrontal cortex (Kaya and McCabe, 2019; Pandya et al., 2012). My expectations were also informed by structural MRI studies which posited that the maturation of the prefrontal cortex is linked to the development of adaptive emotion regulation during adolescence (Pizzagalli and Roberts, 2021; Kaya and McCabe, 2019). For example, greater cortical thinning of the prefrontal regions has been shown to predict more frequent use of cognitive reappraisal in female adolescents (Vijayakumar et al., 2014). Taken together, female adolescents exhibiting subclinical depressive symptoms revealed that their shared brain responses are likely to be modulated by a global contribution of brain regions that are linked to depression and adaptive emotion regulation. This finding requires further investigation in an independent subclinical sample.

8.4.2 Temporal dynamics of brain synchrony in functionally defined networks

As discussed in **Section 8.2.7.4**, emerging approaches characterising dynamic ISC offer opportunities to examine the temporal dynamics of brain synchrony, potentially revealing new insights into shared neural responses during specific moments or windows of events over the course of a movie (Chang et al., 2021; Meer et al., 2020; Gruskin et al., 2020; Di and Biswal, 2020; Guo et al., 2015). Temporal ICA revealed several functional networks that exhibited coherent moment-to-moment changes in ISC, many of which recapitulate patterns of neural responses during rest and task performances (Allen et al., 2011; Smith et al., 2009; Emerson et al., 2015). The recruitment and modulation of these higher-order networks suggest that they are engaged to support cognitive and emotional processing as the movie unfolds, with some networks showing differential modulation within and between high- and low-emotion segments. Further, the temporal patterns of several functional networks showed idiosyncratic responses during the high-emotion segment which agrees with weaker cross-participant similarity in individuals suffering from melancholic depression (Guo et al., 2015). However, I observed some network components that revealed greater normalised responses in the same segment, thus supporting a previous report that emotionally evocative events increase the brain synchrony in the frontolimbic network derived from adolescent depression (Gruskin et al., 2020). These temporal patterns were, therefore, consistent with previous work on modes of variations in functional connectivity and brain synchrony — some networks may dynamically reconfigure the narrative comprehension

during movie-watching whereas others may have a role in minimising the dwelling time during negative stimuli (Demirtaş et al., 2019; Simony et al., 2016; Guo et al., 2015).

8.4.3 Dynamic functional connectivity in emotional processing regions during naturalistic movie stimuli

To examine whether moment-to-moment changes in functional connectivity over the course of the emotionally evocative movie have phenotypic relevance, I performed an exploratory dynamic functional connectivity analysis. Although preliminary and in need of further refinement, the analysis suggested that the functional connections between regions involved in emotional processing exhibited phenotypic associations that evolved over time, and across low- and high-emotion segments of the movie. Several edges showed positive or negative relationships with depressive symptom severity over the course of the movie, suggesting that changing connection strengths between emotional processing regions may reflect emotional processing that is modulated by depressive symptom severity. Converging evidence from the literature has shown that depressed individuals exhibit aberrant functional connectivity patterns in these regions which may be associated with anhedonia (Light et al., 2011), suicidal thoughts (Chen et al., 2021; Wei et al., 2018), anxiety (He et al., 2019), poor social skills, and positive relations with others (He et al., 2022; Segrin et al., 2009). A functional connection between the dorsal premotor cortex and amygdala shared a significant relevance with adaptive emotion regulation strategies during the start of the high-emotion segment. Greater use of such regulation styles may strengthen the premotor-amygdala functional connectivity, thus allowing depressive individuals to modulate their internal sequence of actions when they experience affect-laden events (Ohbayashi et al., 2016; Solopchuk et al., 2016). During the low-emotion segment, there were two prefrontal-limbic connections that revealed an inverse association with maladaptive regulation strategies, which aligns with evidence that engagement of maladaptive emotion regulation styles is associated with altered interactions between the orbitofrontal cortex and amygdala (Fiuzat et al., 2017), and between the dorsal premotor and dorsolateral prefrontal cortices (Bakhit et al., 2020). These observations must be caveated by the exploratory nature of the analysis, however. Due to the sliding window approach, edges are temporally dependent and overlap across the sliding windows. A modified correction for multiple comparisons should be applied, but this may limit the practical utility of the approach. Here, in line with the exploratory nature of the study, I retained only the edges whose significant brain-behaviour relationships persisted across at least five sliding windows. Future work should refine this approach and examine its reproducibility.

8.4.4 Self-similarity signal structure of the ventral attention network and ventrolateral prefrontal cortex in adolescents experiencing subclinical depressive symptoms during movie-watching

Spontaneous fluctuations in the BOLD fMRI signal demonstrate properties of scale-free dynamics, which can be characterised in terms of temporal complexity or self-similarity structure (Dong et al., 2018; Churchill et al., 2016; Ciuciu et al., 2012; He, 2011; Maxim et al., 2005; see **Chapter 7, Section 7.2.3**). In this sample of adolescents, the self-similarity signal structure of the VAN was positively associated with depressive symptom severity. The VAN comprises the bilateral vLPFC and temporoparietal junction (TPJ), and is thought to orient the attention towards unexpected salient stimuli that are behaviourally relevant (Vossel et al., 2014; Corbetta et al., 2008; Corbetta and Shulman, 2002). This observation is consistent with reports of increased within-network effective functional connectivity between vLPFC and TPJ in both neurotypical and depressed adolescents (Liu et al., 2019; Sylvester et al., 2013). At the ROI level, the self-similarity signal structure of the left vLPFC was also significantly related to depressive symptomatology and negatively related to the use of adaptive emotion regulation strategies, such that individuals who report using adaptive emotion regulation less frequently show more regular behaviour in the signal of the left vLPFC. In addition to being part of the VAN, activity in the left vLPFC has been associated with cognitive appraisal to help reduce the emotional impact of an event or scene after changing the interpretation of a negative stimulus (McRae et al., 2010; Goldin et al., 2008). In addition, the left vLPFC has been shown to down-regulate the responses in emotion-specific regions including the amygdala and insula during negative movie-watching (McRae et al., 2010; Goldin et al., 2008). Although these findings require replication in an independent sample, they suggest that dysregulation of emotional information processing in the vLPFC in adolescents experiencing mood symptoms represents an important avenue for further research.

8.4.5 Limitations and future directions

Four major limitations warrant consideration. The primary shortcoming of the present study, which focused on a unique dataset collected in-house, is its sample size. As has been discussed throughout this thesis, larger sample sizes are recommended to achieve reproducible functional connectomes that are valid for brain-behaviour relationships. Future work focusing on naturalistic emotional processing in adolescent depression could benefit from publicly available developmental datasets such as the HBN (Alexander et al., 2017) and ABCD Study (Casey et al., 2018). Indeed, these datasets could be used to examine the validity and reproducibility of the exploratory analytic approaches and findings described above. The second challenge of the present study lies in the heterogeneity of psychiatric disorders such as depression (Buch and Liston, 2021; Mihalik et al., 2019; Xia et al., 2018; Drysdale et al., 2017; Zhang et al., 2016).

Since the adolescents in the sample exhibited subclinical symptoms, it means that some individuals exhibited virtually no symptoms, while others exhibited few or moderate symptoms. The lack of clinical severity in this community-based sample is likely to have negatively impacted my ability to detect brain-behaviour relationships. Third, the duration of the movie may also have played an important role. Given that *The Present* is only 4 minutes 18 seconds including the ending credits and the high emotion segment is predominantly towards the end of the movie, the lack of significant relationships may be attributed to the short movie length. As discussed in the previous Chapters, existing evidence suggests that cross-participant similarity derived from movie-watching fMRI may be improved with increasing scan duration (Vanderwal et al., 2021; Gruskin et al., 2020; Vanderwal et al., 2019; Guo et al., 2015). Future work could also apply the bagging approach as proposed in **Chapter 4** to examine whether motion-limited naturalistic fMRI timepoints could boost brain synchrony, functional connectivity, and self-similarity structure of the fMRI signals as well as their phenotypic relevance with depressive symptomatology and emotion regulation strategies. Lastly, movie annotations such as luminance and human faces have been shown to predict trait-like phenotypes that are related to cognition (Finn and Bandettini, 2021b). Naturalistic stimuli have further demonstrated to evoke reliable brain states that are associated with physiological indices such as heart rate and pupil diameter (Meer et al. 2020). Meer et al. (2020) showed that these brain states share significant associations with several stimulus features (e.g., face expressions, scene valence/switching, dialogues) that reflect behavioural profiles of anxiety and inhibition. While I did not extract stimulus features, future work might benefit from such annotations, to complement and strengthen dynamic analyses of ISC, functional connectivity, and brain-behaviour relationships. Validating stimulus features by gathering emotion ratings from an independent set of raters will be an important future step, particularly since the low- and high-emotion segments as currently defined are confounded with time (i.e., first and second halves of the movie). As a result, the comparison of the low- and high-emotion segments could be confounded by processes and factors that covary with time, such as stimulus-level (e.g., narrative arcs) and subject-level factors (e.g., arousal, attention). Independently provided emotion ratings would provide a more robust means to define the emotional blocks across the movie, especially because the first half of the movie also captures emotional moments related to frustration, anger, surliness, and so forth.

8.5 Conclusions

Naturalistic paradigms offer opportunities for a multiplicity of analyses that provide a window on the neural correlates of shared and idiosyncratic information processing and how these evolve in an ecologically valid and dynamic context. The analyses reveal dynamic patterns of within- and between-individual brain synchrony that may share associations with depressive symptom

severity and emotion regulation strategies. Emerging and exploratory methodologies can further elucidate the temporal complexity of the attention and prefrontal networks that are thought to play an important role in emotional processing. While future research should address the limitations of the present study (e.g., sample size, subclinical sample, movie length) and directly examine the reproducibility and validity of the analyses, my findings suggest that naturalistic paradigms offer a fruitful means to assess how depressive symptomatology and emotion regulation strategies shape how adolescents experiencing subclinical depressive symptoms process emotional information.

Chapter 9

General Discussion

Summary

Chapter 9 elaborates on the contributions of the *empirical* findings of this thesis. I consider these findings in the context of several broader topics of concern within the neuroimaging field: sample sizes, the phenotype challenge, trade-offs of the resting-state fMRI approach, methodological variability, rates of developmental changes, and the quest for multimodal MRI approaches. In addition to presenting the limitations of the *empirical* research carried out in this thesis, I consider several avenues towards increased reproducibility including meta-matching frameworks, multivariate techniques, multiverse pipelines, and sex differences. Finally, I conclude that a reproducibility lens remains central to the collective effort to discover functional connectivity-based biomarkers of clinical relevance, and ultimately, to improve outcomes for those with neurodevelopmental and psychiatric conditions.

9.1 Contributions of Empirical Findings

The overarching goal of this thesis is to improve the reproducibility of functional connectomics-based measures in neurodevelopmental and psychiatric conditions as a means to enhance their sensitivity to phenotypic differences of clinical relevance (i.e., validity criteria for brain-based measures). As outlined in **Chapter 1**, this thesis relies on the definition of reproducibility as proposed by Botvinik-Nezer and Wager (2022): encompassing analytical reproducibility, replicability, and robustness to analytical variability. The contributions of the four *empirical* research are presented below, and their underlying analytical and conceptual frameworks are illustrated in Figure 9.1.

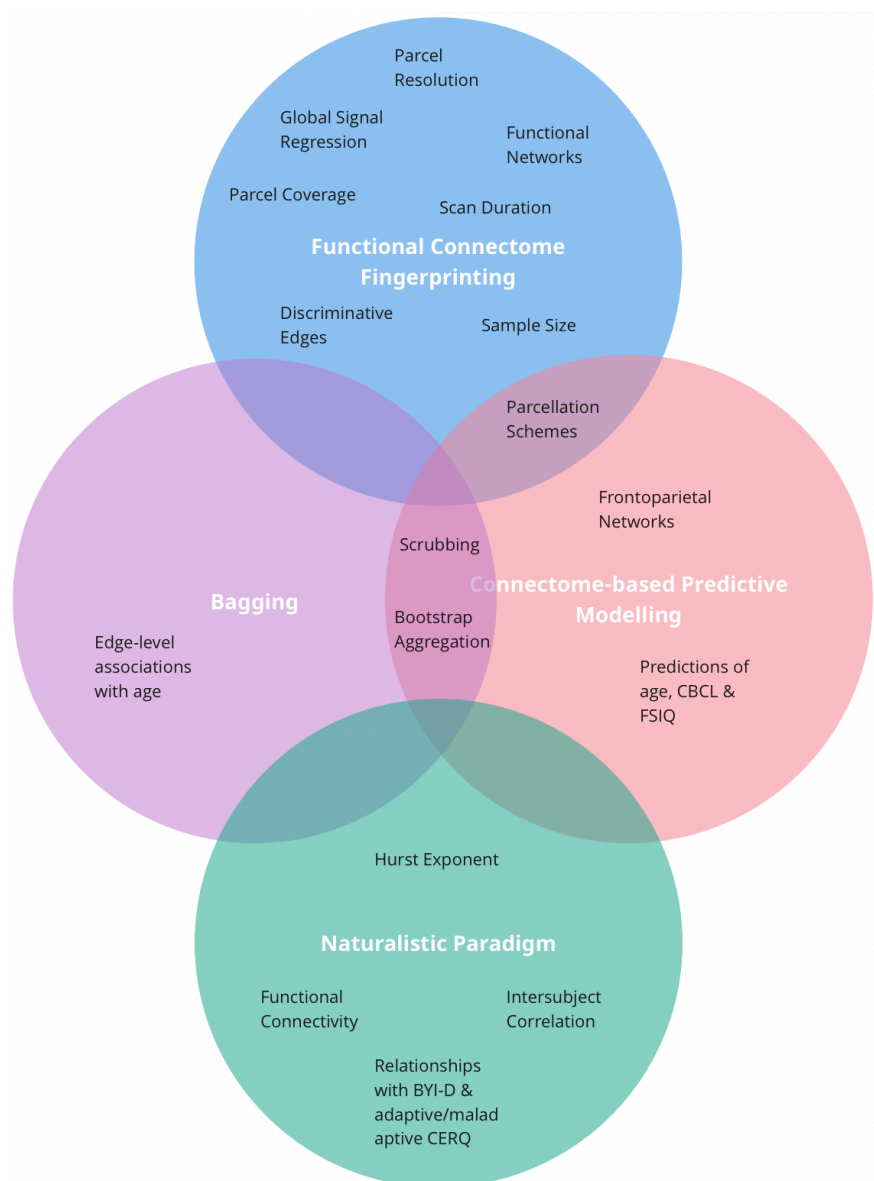


Figure 9.1. Overview of the functional connectivity-based frameworks used in this thesis: functional connectome fingerprinting, bagging, connectome-based predictive modelling, and the application of a naturalistic movie-watching paradigm.

First, I showed that functional connectome fingerprinting can be used as an index of reproducibility for functional connectivity-based measures, and offers advantages over the traditional TRT reliability index, the ICC. By investigating the impact of several methodological factors on fingerprint ID accuracy, I showed that the fingerprint paradigm can serve as a testbed for comparing the impact of different functional connectivity-based pipeline parameters. Using this approach, I identified several factors that increased ID accuracy across five independent datasets. Together, these components form two pipelines that are optimised for the reliable measurement of functional connectivity at the individual level.

Second, I grappled with the challenge to reproducibility that is posed by two countervailing imperatives: the requirement to minimise the impact of head motion and the requirement to maximise sample size. To address this challenge, I proposed bagging or bootstrap aggregation at the level of timeseries. I showed that functional connectivity matrices computed from bootstrap samples of motion-limited timepoints successfully act as reproducible “fingerprints”. I further demonstrated that the bootstrap approach retains validity: univariate brain-behaviour associations (i.e., relationships between functional connectivity and age) are largely preserved for bagged functional connectomes computed from as few as 60 “clean” timepoints. Bagging also maximised sample inclusivity by enabling the retention of up to 30% of the “high-motion” participants that would normally be excluded. This study suggests that we can achieve both high data quality standards as well as inclusive and reproducible functional connectomes without discarding large proportions of our data, even in developing populations.

Third, I encountered an unexpected result: optimising the uniqueness of the individual functional connectome did not necessarily improve connectome-based prediction of behavioural phenotypes in developing youths. While it seems intuitive to infer that factors enabling superior connectome fingerprint differentiation would improve brain-behaviour predictions, I showed that several pre- and post-processing factors, such as maximally discriminatory networks, multi-resolution parcellation schemes, and head motion mitigation strategies (i.e., scrubbing and bagging) did not improve (or disimprove) individual predictions of age, general psychopathology, and intellectual ability. An important observation was that bagged connectomes did not disimprove predictions, further supporting the bagging approach for the rescue of high-motion data. I further showed that certain factors, including finer parcellation schemes and bagging, improved the generalisability of age and intellectual ability predications to an independent sample. These observations suggest that while validity of CPM-based predictions (i.e., strength of brain-behaviour relationships) may be near ceiling, there is scope to boost reproducibility (i.e., generalisation to independent samples).

Lastly, I explored how naturalistic paradigms such as movie-watching offer rich opportunities for the elucidation of how functional brain responses are modulated by phenotypic dimensions of interest, such as depressive symptom severity. I proposed a movie-watching framework that examines the following: (I) shared full-length and moment-to-moment responses using intersubject correlation; (II) “static” and moment-to-moment functional connectivity derived from an *a priori* emotion regulation network; (III) and temporal complexity of the fMRI signal derived from the emotion regulation network and its underlying ROIs using the H. While subject to a number of shortcomings, particularly sample size, my final study provided preliminary illustrations of how a short emotionally evocative movie can serve as an ecologically valid index of emotional processing and regulation captured at multiple spatial and temporal scales in female adolescents experiencing subclinical depressive symptoms.

In this general discussion, I reflect on the main outcomes of this thesis and how they relate to gaps that remain to be addressed so that we can attain reproducible functional connectivity-based biomarkers with clinical viability: (1) sample size; (2) phenotypic measurements; (3) resting-state fMRI conditions; (4) methodological variations in fMRI pipelines; (5) rate of developmental changes; and (6) multimodal MRI applications.

9.2 Consortium-level versus Typical Sample Sizes

To improve the reproducibility of functional connectivity measures, sufficiently powered studies are required to detect the true effect sizes in question, which may be smaller than previously appreciated (Marek et al., 2022; see **Chapter 3 (Section 3.2)** for a detailed discussion). As elaborated in **Chapter 3**, Marek et al. (2022) revealed that thousands of participants ($N \gtrsim 2,000$) are required to achieve reproducible univariate and multivariate BWAS. There has been a scholarly discussion within the neuroimaging community about whether achieving brain-behaviour associations that are robust and reproducible do indeed require thousands of participants. My research supports a point made in these discussions (e.g., Rosenberg and Finn, 2022; Nature Neuroscience, 2022; Gratton et al., 2022; DeYoung et al., 2022; Bandettini et al., 2022; see **Chapter 3, Section 3.2**), that attention to methodological choices other than sample size can boost the robustness and reproducibility of functional connectivity measures. First, I showed that fingerprint ID accuracy can serve as a testbed for functional connectome reproducibility, and can be improved based on a set of data pre-/post-processing factors. On the other hand, I demonstrated that increasing the sample size (from $N = 20$ to $N = 264$) did not help improve the identifiability of an individual’s functional connectome; higher N was found to decrease the fingerprint ID accuracy, thus increasing the challenge of differentiating an individual’s functional connectivity profile (see **Chapter 2**). This shows that there is scope for

further methodological improvements to boost the fingerprint ID accuracy in the context of larger samples, particularly with the ABCD Study ($N \approx 12,000$). Second, I showed the utility of bagging for maximising the sample size requirements by improving participant inclusivity without decrement to the reproducibility of brain-behaviour relationships. As elaborated in **Chapters 3 (Section 3.3) and 4**, it is common practice to discard large number of “high-movement” participants in typical brain-behaviour association studies. Instead, I showed that participants who are usually considered both “low-motion” and “high-motion” can yield robust and reproducible brain-behaviour relationships, even for $N < 500$ — on the basis of the participants’ timepoints that are least corrupted by head motion. An important consideration is that consortia-level datasets do not always capture inclusive and diverse populations, especially those from underrepresented and minority groups (Ricard et al., 2023; Dhamala et al., 2023). For example, Ricard et al. (2023) argued that the lack of inclusivity in neuroimaging data collection is evident in consortia datasets such as the UKBiobank and HCP which respectively include $\sim 95\%$ and 76% white participants. If we are going to acquire and process thousands of participants’ data only to discard a substantial amount prior to the analyses, then this contradicts the motivation to acquire large sample sizes that promote inclusivity and diversity in the first place.

It is clear that sufficiently powered studies with adequate statistical power are required to estimate the true effect sizes of brain-behaviour associations. This is predominantly due to the sampling variability that manifests at low sample sizes to produce inflated brain-behaviour correlations. I found that the 95% CIs for brain-behaviour relationships tightened with increasing sample sizes, showing that the effect sizes converge to their “true” estimates when the study is sufficiently powered, aligned with the findings of Marek et al. (2022). However, the stern focus on sample size is not enough to observe reproducible relationships in both BWAS and non-BWAS studies because large sample sizes do not always translate to reliable phenotypes (Gell et al., 2023; Nikolaidis et al., 2022). Hence, collaborative efforts to acquire thousands of participants (e.g., UKBiobank $N \approx 60,000$) will only address half of the equation for optimising the detection of univariate and multivariate brain-behaviour associations. Recently, Li et al. (2023) showed that certain phenotypes such as age only require 300 individuals to generate replicable brain-behaviour associations while others including mental health and lifestyle measures need consortia-level sample sizes at $N = 1,500$ - $3,900$. For example, the authors showed that a sample size of $N = 300$ can generate fair ICCs ($ICC = 0.50$) whereas a much larger sample size of $N = 3,500$ would be required to replicable alcohol consumption with similar TRT reliability from the UKBiobank. Surprisingly, even IQ, which is one of the most consistently used phenotypes in deriving brain-behaviour relationships would need thousands of participants to achieve the largest effect size for excellent regional replicability ($ICC = 0.75$). Here, regional replicability indexes the replication probability between independent samples for individual brain-behaviour

associations using a resampling procedure (Li et al., 2023). Similarly, I showed in **Chapter 6** that CPM can yield robust connectome-based predictions of age and intellectual ability in developing youths that are generalisable in an independent cohort of neurotypical individuals based on a set of methodological factors for $N < 500$. This work reinforces the message that datasets with N s numbering in the thousands are not the only route to robust and reproducible brain-behaviour associations.

The active debate on whether reproducible brain-behaviour associations require thousands of participants has prompted several methodological suggestions to reduce the sample size requirement. These methodological innovations reveal new opportunities to extend the work carried out in thesis by deriving brain-behaviour associations without the urgency to acquire sample sizes in the thousands. For example, Spisak et al. (2023) showed that applying a cross validation (CV) method to derive in-sample effect sizes using the HCP yielded strong effects for multivariate BWAS from moderate sample sizes ($N = 75$ -500), and eliminated the inflation rate in the discovery sample, reported by Marek et al. (2022). While estimating the CV-based effect sizes, multivariate BWAS attained 80% in-sample statistical power and out-of-sample replication probability for $N < 500$ when phenotypes such as age, cognitive ability, and fluid intelligence were predicted. It is important to note, however, that CV does not always guarantee the detection of true effect sizes and replication rate as CV on its own can introduce model overfitting, bias, and complexity (see **Chapter 5 and Appendix E**). Studies that employ CV can still suffer from insufficient power and increased variability in effect size estimation (Spisak et al., 2023). Similar observations have been reported when CV is employed to estimate the effect sizes of univariate BWAS derived from voxel-based morphometry (Goltermann et al., 2023). While CV shows its utility in generating sufficiently large effect sizes for sample sizes in the hundreds ($N > 400$), smaller samples could inflate the effect sizes of univariate voxel-based associations. Nonetheless, the application of CV in univariate BWAS could serve as a solution to evaluate the generalisability of the effect sizes in typical sample sizes compared to large-scale datasets regardless of the modality (Goltermann et al., 2023). A further clarification about the application of CV is that multivariate BWAS cannot be fully replicated in smaller samples from independent consortia even with the largest effect sizes. Tervo-Clemmens et al. (2023) showed that only 31% of the multivariate BWAS effects are replicable in the ABCD Study at $N = 1000$ followed by 14% at $N = 500$, and 0% at $N = 75$. When sample sizes tend to be typical, the rate of false negatives reaches 55%. If we take the largest multivariate BWAS effect size, the ability to replicate such effect is possible at $N = 400$ at the expense of a 40% reduction in the out-of-sample prediction from the HCP to the ABCD Study (Tervo-Clemmens et al., 2023). Another approach lies in detecting regional univariate and multivariate BWAS which seems to achieve 80% power with sample sizes in the order of hundreds derived from the HCP (Cecchetti and Handjaras, 2022).

Expectedly, multivariate brain-behaviour associations revealed stronger effect sizes compared to univariate BWAS; multivariate BWAS achieved statistical power greater than 80% when predicting cognitive abilities at $N = 350$ whereas reliable associations between brain features (e.g., functional connectivity, cortical thickness) and psychopathology required larger sample sizes ($N \approx 1,500$) to achieve sufficient power.

While it is not possible to propose the ideal sample size that is required for reproducible brain-behaviour associations, i.e., there is no one-size-fits-all studies, neuroimaging-based researchers should not be discouraged from working with typical sample sizes — the median sample size in fMRI studies is $N \approx 25$ (Szucs and Ioannidis, 2020). Although the neuroimaging field is moving towards consortia datasets arising from the Open-Source and data sharing movements, I believe that studies employing typical sample sizes are equally as important as those relying on large-scale datasets. Individual laboratories usually employ their own datasets of typical sample sizes as they are not able to acquire thousands of participants due to the lack of infrastructure, funding, and ethical justification. Typical sample-size studies have promising potentials to truly advance the search for functional connectivity-based biomarkers in challenging populations by enabling targeted investigations that are generally not possible in larger studies, such as: (I) allowing precise phenotyping of behaviours; (II) developing repeated measurements of imaging data cross-sectionally and longitudinally; (III) lesioning of specific brain regions and functional networks; (IV) inducing brain states under resting-state, task-based, naturalistic, and pharmacological conditions; (V) studying rare clinical conditions and/or subtypes in neurodevelopmental and psychiatric conditions; and (VI) establishing rich, diverse, age-specific datasets to investigate brain-behaviour associations.

9.3 Reliable Measurement of Phenotypes

While the neuroimaging community is pushing towards “big data” (see **Chapter 3 (Section 3.1)**), the reliability of both imaging data and behavioural phenotypes has garnered particular attention over the years. As I mentioned in **Chapter 1**, the ability to observe reproducible brain-behaviour relationships for biomarker discovery requires us to not only focus on improving the reproducibility of the functional connectome but also, requires us to optimise for reliable phenotypes. While the split-half reliability of resting-state functional connectivity generally fluctuates across consortia ranging from 0.39 [UKBiobank] to 0.79 [HCP] with the ABCD Study achieving moderate estimate (Marek et al., 2022), the TRT reliability of behavioural phenotypes typically investigated in fMRI studies ranges from poor to excellent, and may be hindered by the presence of measurement noise (Gell et al., 2023). Attenuating phenotypic reliability produces a drastic reduction in the brain-behaviour predictions as a result of increasing measurement noise

irrespective of methodological variations in brain parcellations and predictive models (Gell et al., 2023). In **Chapter 6**, I showed that CPM successfully predicted three phenotypes including age, general psychopathology, and intellectual ability in developing youths with baseline model predictive performances ranging from 0.16 to 0.61. I further demonstrated that optimising the individual functional connectome did not necessarily improve or disimprove the brain-behaviour predictions. While the chosen phenotypes were sufficiently reliable to be predicted, it is possible that phenotypes in other domains may not be reliable enough. This is because behavioural phenotypes often suffer from heterogeneity (e.g., subtypes, comorbidities, prevalence), ill-informed diagnostic criteria (e.g., categorical, subclinical, dimensional), biases (e.g., sampling bias, method bias), and low resolution (see **Chapter 5 (Section 5.3)**).

Moreover, I observed that the CPM-based prediction of general psychopathology as indexed by the CBCL is a challenge, since it appears to be less sensitive to methodological considerations — at least for the factors I examined in **Chapter 6**. If we take a look at the inter-rater reliability of diagnostic scales in neurodevelopmental and psychiatric conditions based on the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013), their Kappa statistics — a measure of inter-rater reliability — show great variability, ranging from unacceptable agreement to very good agreement (Figure 9.2; Freedman et al., 2013; Cohen, 1960). For example, MDD diagnoses in adult and youth populations exhibit a Kappa statistic of 0.28 which refers to questionable agreement (Figure 9.2) — a limitation I pointed out in **Chapter 8**. It is, however, possible to improve the TRT reliability of phenotypic measures by acquiring several repeated measurements (e.g., ≥ 4) — this can increase the reliability of CBCL and NIH Toolbox from poor-to-good ($0.40 \leq \text{ICC} \leq 0.70$) to good-to-excellent ($0.66 \leq \text{ICC} \leq 0.91$) (Nikolaidis et al., 2022). Even with the acquisition of two phenotypic measurements for $N = 1000$, the mean square error (MSE) in predicting brain-behaviour associations decreases compared to one phenotypic measurement for much greater sample sizes ($N = 10,000$ - $100,000$) (Nikolaidis et al., 2022). Other routes to improving the reliability of behavioural phenotypes such as psychopathology are dimensional measurement, deep phenotyping, and hierarchical modelling (Tiego et al., 2022). While developing new phenotypes in an effort to improve univariate and multivariate brain-behaviour relationships, age can be used as a phenotype of reference as I have done in **Chapters 4 and 6** given that age is the most robust and reliable phenotype ($\text{ICC} \approx 1.0$) in the literature, though obviously differs from behavioural and clinical measures in that it is a direct biological index. Although it is unlikely that new phenotypic measures will achieve perfect reliability due to biological variations that manifest within and between individuals in addition to a potential reliability plateau, establishing more novel sensitive and reliable phenotypes should not be discounted.

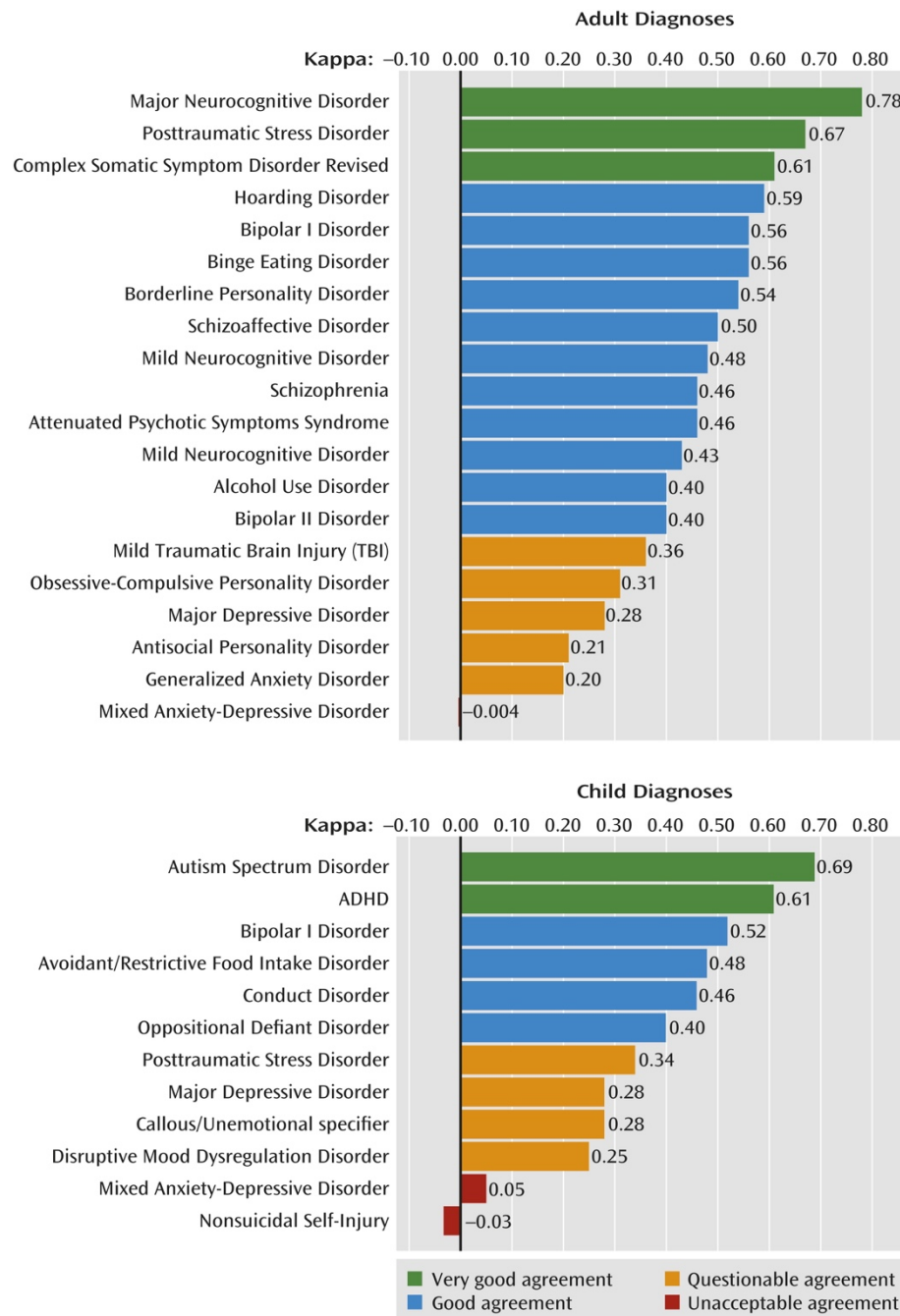


Figure 9.2. Inter-rater reliability of diagnostic scales from the DSM-5 measured by the Kappa statistic. The Kappa statistic can be interpreted as follows: ≤ 0 indicates no agreement and 0.01-0.20 as none to minimal agreement, 0.21-0.40 as fair agreement, 0.41-0.60 as moderate agreement, 0.61-0.80 as substantial agreement, and 0.91-1 as almost perfect agreement. Adapted from Freedman et al., 2013.

We, as researchers, should always be reminded that achieving large sample sizes for robust and reproducible brain-behaviour associations will only be useful if the phenotype of interest is reliable. Phenotypes that exhibit poor-to-moderate reliability will require increasing sample sizes or training data to successfully predict brain-behaviour relationships. Indeed, smaller sample sizes with high reliability tend to have better predictive performances compared to larger samples that have good-to-moderate reliability. A sample size of $N = 400$ with excellent phenotypic

reliability can outperform the prediction of age compared to a substantially larger sample size of $N = 4,450$ with moderate reliability derived from the UKBiobank (Gell et al., 2023). Therefore, acquiring very large sample sizes will be beneficial for brain-behaviour predictions if *both* the imaging and phenotypic data are highly reliable; otherwise, data that share suboptimal reliability may not offer any significant gains in predicting behavioural phenotypes. If we further consider the joint reliability of brain features (e.g., here, functional connectivity) and behavioural phenotypes expressed as $\sqrt{ICC_X ICC_Y}$, then poor phenotypic reliability combined with low imaging reliability would require significantly greater sample sizes ($N > 5,000$) to detect small-to-moderate effect sizes as opposed to having excellent joint reliability which would require a sample size of $N = 200$ to observe reproducible brain-behaviour relationships (Nikolaidis et al., 2022).

Although resting-state fMRI has provided an exciting opportunity to elucidate the functional brain organisation and its relevance to behavioural phenotypes across populations, brain states, and non-human species (see **Chapter 1**), its application has rendered potential trade-offs in the search for functional connectivity-based biomarkers.

9.4 Trade-offs of Resting-State fMRI

Low participant demand and wide availability mean that resting-state fMRI paradigms are particularly suitable for studying brain functions and dysfunctions across neurodevelopmental and psychiatric conditions. The majority of the *empirical* work presented in this thesis have focused on improving the reproducibility of the individual functional connectome using resting-state conditions with the exception of the last study, which applied a movie-watching paradigm to study individual and group-average functional brain responses at multiple spatial and temporal scales. In **Chapter 8**, I showed that an emotionally evocative movie can serve as an ecologically valid, rich, and engaging candidate of emotional processing and regulation during adolescence, even with a movie length of less than 5 min. Although preliminary, I showed that female adolescents experiencing subclinical depressive symptoms exhibited temporal autocorrelated patterns in the VAN and left vIPFC that could be modulated by depressive symptomatology and adaptive emotion regulation. In general, naturalistic connectomes tend to yield better brain-behaviour predictions than resting-state conditions as individual differences in functional connectivity may be influenced by trait-like features (Finn and Bandettini, 2021b). While researchers are bound to select an appropriate naturalistic stimulus, movie annotations such as luminance and human faces have been shown to be predictive of trait-like cognitive phenotypes (Finn and Bandettini, 2021b; see **Chapter 8**).

On the other hand, task-induced fMRI paradigms better capture individual properties of the functional connectome and their underlying associations with behavioural phenotypes (Zhao et al., 2023; Jiang et al., 2020; Greene et al., 2018a; Finn et al., 2017; Gratton et al., 2016; Finn et al., 2015; see **Chapter 5 (Sections 5.2-5.3)**). Certain tasks tend to yield stronger brain-behaviour relationships that are robust and generalisable compared to other tasks and/or resting-state conditions as they are usually designed to capture specific behavioural processes such as sustained attention, decision-making, reward processing, personality traits, intellectual ability, and executive functioning (Zhao et al., 2023; Finn, 2021a). As elaborated in **Chapter 5 (Sections 5.2-5.3)**, combining multiple task states has been shown to boost the CPM-based predictions of behavioural phenotypes, relative to deriving the task-based connectomes in isolation (Jiang et al., 2020; Gao et al., 2019; Greene et al., 2018a). There are two important underlying points to highlight: (1) there may not be a single brain state that encapsulates the functional organisation of the brain (Finn, 2021a; Finn et al., 2017); and (2) aggregating more task-based conditions does not necessarily enhance brain-behaviour predictions. Instead, more predictive tasks tend to contribute to more meaningful features (e.g., edges) in a predictive model (Jiang et al., 2020).

In this thesis, I showed how resting-state fMRI conditions can be used to successfully “fingerprint” an individual from their functional connectome, rescue “high-movement” participants, and predict brain-behaviour relationships in developing youths aged as early as 6 years old. Others have been able to differentiate an individual’s connectome during infancy as early as 1 month old, albeit observing poorer connectome distinctiveness compared to adolescence and adulthood, especially across between-session scans (King et al., 2023; Hu et al., 2022; Dufford et al., 2021). In **Chapter 4**, I further performed bootstrapping and scrubbing of resting-state fMRI timeseries to sample “low-motion” timepoints — procedures which were performed on the assumption that the timepoints do not need to be contiguous as opposed to task-based timeseries as the timepoints are stimulus-locked (Shinn et al., 2023; Arbabshirani et al., 2014). Subsequently, the ability to acquire fMRI data in the absence of an external stimulus has encouraged researchers to increase the amount of imaging data collected per participant from minutes to hours, as a means to allow precision or “deep” mapping of the individual functional connectome (Gratton et al., 2020; Gordon et al., 2017; Laumann et al., 2015). For example, the Midnight Scan Club [MSC] dataset collected more than 20 hours of fMRI data in 10 participants (OpenNeuro: [10.18112/openneuro.ds000224.v1.0.4](https://openneuro.org/datasets/10.18112/openneuro.ds000224.v1.0.4); Gratton et al., 2020). The TRT reliability of the functional connectome improves with longer acquisition times (Noble et al., 2019; Noble et al., 2017; see **Chapter 1 (Section 1.6.3)**), benefitting subcortical regions, such as cerebellar structures, in particular, since these brain regions are fundamentally susceptible to sources of physiological noise (Gratton et al., 2020; Shah et al., 2016).

Since the inception of resting-state fMRI and the emergence of functional connectivity in 1995, the field of neuroimaging has gained favourable insights into the functional brain organisation (see **Chapter 1 (Sections 1.1-1.3)**). However, we should not forget that the “resting” brain means it is unconstrained, which could confound the detection of reproducible functional connectivity patterns, especially in challenging populations such as developing youths, based on a number of factors including: (1) *resting conditions* such as eyes closed, eyes opened, and fixation (Zuo et al., 2015; Patriat et al., 2013; see **Chapter 1 (Section 1.6.3)**); (2) *wakefulness* such as sleepiness and drowsiness (Tagliazucchi and Laufs, 2014); (3) *spontaneous thoughts* such as “mind-wondering” (Kucyi et al., 2016; Christoff et al., 2009); (4) *mood states* (Harrison et al., 2008); and (5) *in-scanner motion* (Power et al., 2015; Satterthwaite et al. 2012; Kelly et al., 2012; see **Chapter 3 (Section 3.3)**). Another caveat of resting-state fMRI lies in disentangling the amount of explained variance that reflects trait-like versus state-like components in functional connectivity (Finn et al., 2017; Geerlings et al., 2015b). Features derived from resting-state functional connectivity have also demonstrated to weaken the prediction of cognitive measures than individual characteristics related to age, gender, and intracranial volume (Omidvarnia et al., 2023). While these individual characteristics increase the predictive performance, aggregating them with resting-state fMRI features marginally improves the phenotypic predictions (Omidvarnia et al., 2023). It is also possible that there is a ceiling when improving resting-state functional connectivity estimates; for example, I observed that the components that describe two optimal fingerprint-informed pipelines did not yield 100% identification accuracy (see **Chapter 2**). Of note, these components that form part of the optimal pipelines do not reflect a final set of factors, since there are parameters that I have not examined in **Chapter 2**. In addition, CPM-based predictions of three phenotypes did not necessarily improve (or disimprove) across several data pre-/post-processing factors (see **Chapter 6**). I return to this in **Section 9.5**. While the interactions between sample size, phenotypic measurements, and resting-state fMRI paradigms are deemed essential for reproducible functional connectivity estimates, methodological variability in fMRI pipelines still remains a dominant challenge.

9.5 Methodological Variability in fMRI Pipelines

Methodological variability in fMRI pipelines is often referred to as the “researcher degrees of freedom” (Wicherts et al. 2016; Simmons et al., 2011). In functional neuroimaging, these degrees of freedom are present during the data pre- and post-processing stages, thus contributing to a growing number of false positive (type I error) outcomes and inflated effect sizes (Graff et al., 2022; Dhamala et al., 2022; Li et al., 2021; Botvinik-Neher et al., 2020; Dadi et al., 2019; Parkes et al., 2018; Poldrack et al., 2017; Ciric et al., 2017; Shirer et al., 2015; Yan et al., 2013a; Carp, 2012). The major sources of variability in fMRI pipelines are typically observed due to either the

inclusion/exclusion of certain steps or their sequential order of a procedure. In this thesis, I have maintained identical pre-processing pipelines, even when I employed fMRIPrep for the last study, to alleviate potential sources of variation in the functional connectivity estimates, followed by the same Shen ($k = 268$) parcellation scheme to extract the fMRI timeseries. However, this is not the case in the wider context of neuroimaging; for example, at the pre-processing stage, GSR remains a debatable head motion mitigation strategy which means that some laboratories consistently apply it while others do not (see **Chapters 1 and 2**). Similarly, spatial smoothing varies according to the choice of the Gaussian kernel but also, depending on whether one is interested in the whole-brain, ROI-based, or network-based functional connectome (Mikl et al., 2008). Another form of variability lies in applying a head motion threshold to identify “high-motion” individuals; as I have elaborated in **Chapters 3 and 4**, consortia and typical neuroimaging studies tend to apply an arbitrary mean FD/rmsFD threshold which can be interpreted as either liberal (> 0.30 mm) or stringent (< 0.10 mm). At the post-processing stage, the choice of parcellating the functional brain varies largely on laboratory preferences, research questions, and desired resolution of the effects being investigated. Another common source of variability lies in thresholding statistical maps (Marek et al., 2022; Botvinik-Neher et al., 2020; Poldrack et al., 2017) and analysis software (Bowring et al., 2019). Together, these highlight the “toothbrush problem”² that persists in functional neuroimaging — when it comes to analysis pipelines for functional connectomics, many laboratories and researchers feel very strongly about only using their own. One implication of this is the necessary duplication of data analyses on the same raw data, in laboratories around the world. When we consider the size of some of the datasets available runs to the tens of thousands (e.g., ABCD Study, UKBiobank, HBN), this entails considerable unnecessary energy use. This then leads to a considerable amount of flexibility across functional connectomics-based pipelines. Hence, I hope that the work conducted in this thesis will help move the field from a surfeit of analysis options towards a consensus approach, which will boost reproducibility while reducing the resources spent using only one’s own toothbrush.

Subsequently, it is important to remember that controlling for covariates (e.g., socioeconomic status, sex, age, education, head motion) in deriving functional connectivity estimates imposes a quandary as researchers have the freedom to select the covariates they wish to account for although some may be conservative (i.e., controlling for too many confounds) whilst others could be flexible in their approach (controlling for least potential confounds) (Hyatt et al., 2020). Hence, I have covaried the same variables in **Chapters 4 and 6** when estimating, predicting, or

²Here, I am drawing on Poldrack & Yarkoni (2016), who cited Walter Mischel’s characterisation of the field of Psychology: “Psychologists treat other peoples’ theories like toothbrushes — no self-respecting person wants to use anyone else’s” (Mischel 2008, cited in Poldrack and Yarkoni, 2016).

generalising brain-behaviour relationships from the HBN and ABIDE. Although covarying for potential confounds could drastically change the statistical significance of outcomes and their underlying effect sizes, it does not always guarantee that their effects are completely removed. A prominent example of methodological variability in fMRI pipelines can be appreciated in a study which employed 70 independent teams to evaluate 9 predefined hypotheses using a single dataset; while there were some levels of agreement between groups, no two teams employed similar analysis workflows which contributed to the variability in functional connectivity estimates (Botvinik-Neher et al., 2020).

In terms of connectome-based predictions of behavioural phenotypes, predictive models such as CPM are generally trained via parameter tuning (see **Appendix E** for the list of parameters), which could increase the likelihood of finding a meaningful brain-behaviour association above chance if too many combinations are formulated during feature selection and model building. In **Chapter 6**, I employed connectome-based fingerprinting to examine the robustness of the CPM-based predictions and generalisations. I showed that FSIQ was generalisable to an independent sample when the individual functional connectomes were derived from fine-grained parcellation schemes and bagging. Recently, a study has shown that “enhancement attacks” — minor alterations in the functional connectome introduced to falsely boost the predictive performance of a phenotype — can dramatically inflate the relationships between observed and predicted IQ in ways that cannot be easily detected visually (Rosenblatt et al., 2023). Rosenblatt et al. (2023) showed that, by introducing individual-specific patterns that share a significant association with every participant’s IQ to create the enhanced individual connectomes, boosted the IQ predictions from 0.18 (using the original connectome) to 0.97 (using the enhanced connectomes), whilst the correlations between the original and enhanced connectomes maintained near-perfect correlations of $r \approx 0.99$ (Figure 9.3). Together with my findings in **Chapter 6**, this raises the alarm that the robustness of predictive frameworks such as CPM should be prioritised as manipulating the functional connectome to improve brain-behaviour associations can be harmful without proper justifications.

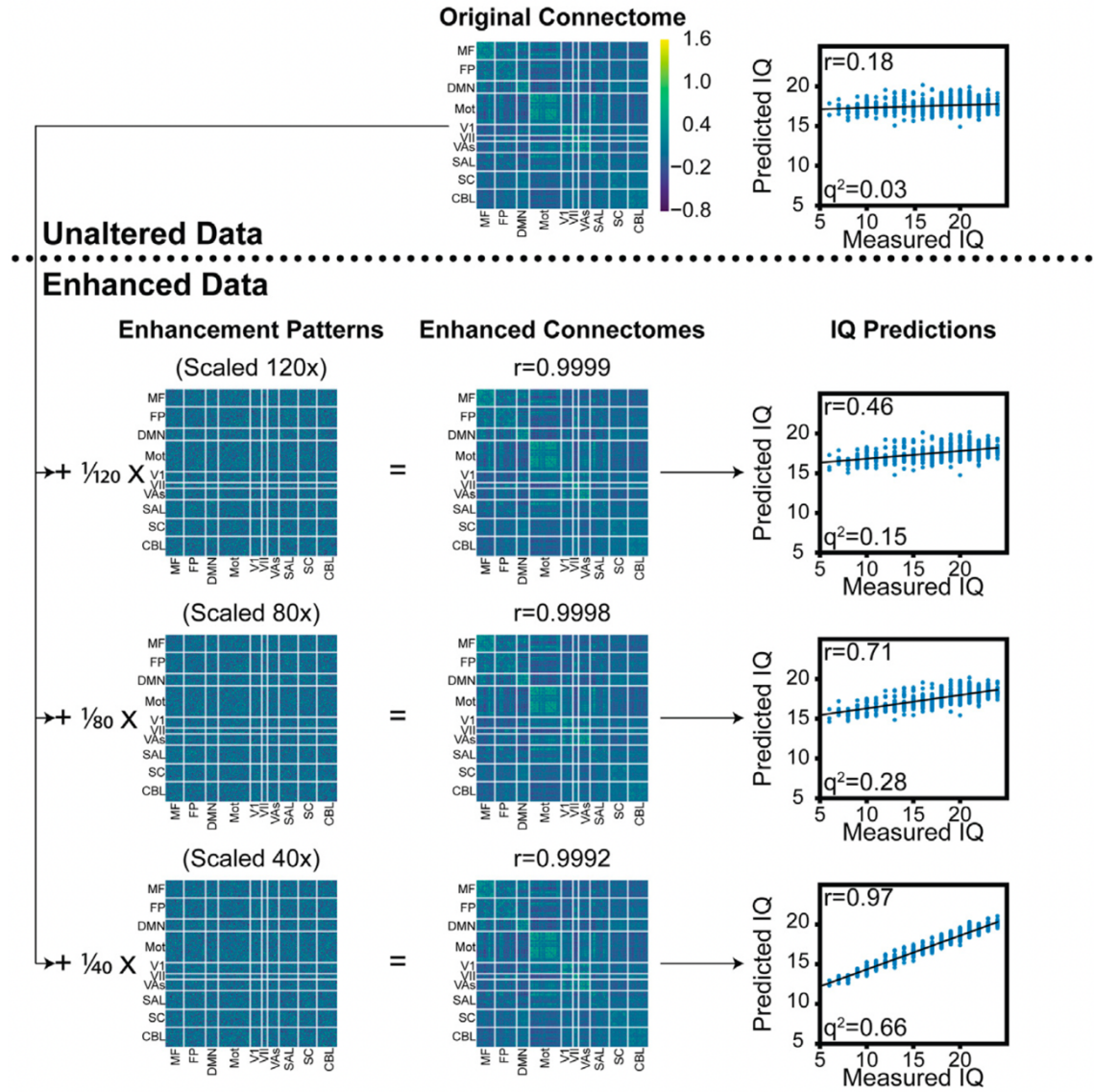


Figure 9.3. Subtle manipulations of the functional connectome by introducing “enhancement attacks” in the CPM to predict a chosen behavioural phenotype such as IQ obtained from the HCP. Model performance is quantified by the Pearson’s correlation coefficient (r) and cross-validated R^2 denoted by q^2 between observed and predicted IQ. MF: medial frontal; FP: frontoparietal; DMN: default mode; MOT: motor; VI: visual I; VII: visual II; Vas: visual associations; SAL: salience; SC: subcortical; CBL: cerebellum. Adapted from Rosenblatt et al., 2023.

The other question that arises is whether the fingerprint identification procedure always produces the “true” proportion of successfully identified individuals in a given cohort. There are scenarios whether fingerprint ID accuracy as well as TRT reliability can produce misleading conclusions (Figure 9.4; Milham et al., 2021). As such, a discriminability (D) metric has been proposed which calculates the probability of the distance between-sessions within each individual being lower than the distances to other individuals (Wang et al., 2020). An interesting property of D is that it can be equivalent to the ICC under the Gaussian assumption for univariate data although it can be broadly applied to assess the reliability of high-dimensional data given that it is a non-

parametric multivariate generalisation of the ICC. The advantage of D is that it provides stronger statistical power with increasing number of participants and repeated measurements compared to the fingerprint identification procedure (Wang et al., 2020). It also generates discriminability estimates that tend to be constant as the sample size grows whereas fingerprint ID accuracy decreases non-linearly with more participants (Wang et al., 2020), consistent with my findings in **Chapter 2**. Therefore, D could provide complementary information about the identifiability of an individual's functional connectome together with the fingerprint ID accuracy — as a means to index the reproducibility of functional connectivity estimates. This means the factors that improve both metrics could better indicate whether they can enhance the validity of the functional connectome, i.e., strengths of brain-behaviour relationships.

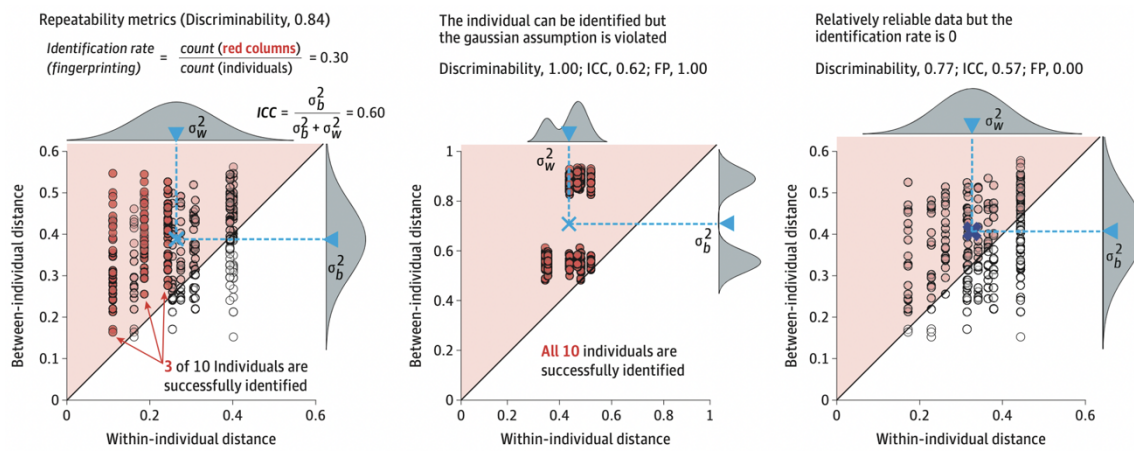


Figure 9.4. Discriminability, TRT Reliability (ICC), and Fingerprinting (FP) of the functional connectome. Left Panel: First scenario where FP reveals a low ID accuracy of 30%, ICC shows a “good” reliability of 60%, and Discriminability is high with a value of 84%. Middle Panel: Second scenario where both Discriminability and fingerprint ID accuracy are optimal (100%), but the ICC shows a “good” reliability of 62%. Right Panel: Third scenario where no individual has been successfully identified, i.e., fingerprint ID accuracy = 0%, but the TRT reliability (ICC = 57%) and Discriminability (77%) are fair/moderate. Adapted from Milham et al., 2021.

Although the disagreement in fMRI pipelines is evident in the literature, there has been some degree of consensus in applying Open-Source analysis workflows such as fMRIPrep (Esteban et al., 2019), C-PAC (Craddock et al., 2013c), and NiPreps (<https://www.nipreps.org>) aimed at harmonising and reproducing functional connectivity-based analyses (Ramduny et al., 2023). Achieving perfect agreement amongst researchers may not be feasible as some variability is bound to be present across populations. For example, applying the same parcellation scheme to derive the functional connectome during infancy and adulthood is not ideal as the size of an infant brain is smaller and less functionally developed. Similarly, developing youths tend to exhibit excessive in-scanner motion compared to adults (Satterthwaite et al., 2012); applying identical head motion thresholds for both populations will lead to a greater number of youths being excluded from the cohort compared to adults. Finally, methodological variability can also be

intertwined with the dataset being analysed due to the curse of dimensionality — analysing or extracting features in high-dimensional spaces (e.g., neuroimaging data) that do not occur in low-dimensional environments — a common curse for predictive models (Mwangi et al., 2014). While addressing the methodological variability in fMRI pipelines, it is also important to be aware that the rate of developmental changes could pose a challenge in capturing individual differences in functional connectivity that are reproducible across different populations.

9.6 Rate of Developmental Changes

Development contains a number of sensitive periods during which the brain exhibits increased plasticity, and patterns of segregation and integration emerge (early childhood → adulthood; see **Chapter 1 (Section 1.2)**). These changes shape the development of behaviours that are necessary for later stages of life (Cao et al., 2017). A key challenge for neuroimaging is that the developmental trajectories of the structural and functional brains generally differ due to the rate of their underlying maturational processes, and are influenced by several biological and environmental factors (Figure 9.5; Sydnor et al., 2021; Shaw et al., 2008). The work carried out in this thesis focuses on developing populations with fMRI data obtained at different scan intervals. For example, in **Chapter 2**, I used the resting-state scans from five different datasets (7-43 years) acquired across several days, months, and even years, to perform the connectome-based fingerprint identification procedure. This means individuals in some of the datasets underwent larger developmental changes compared to others. In **Chapter 4**, I used the UPSM group from CoRR (*between*-session scans) in addition to the HBN (*within*-session scans) to obtain the resting-state fMRI data, which means that the scans reflected short-term and long-term developing changes in the cohorts from 6 to 21 years. Although connectome-based fingerprinting and bagging were successfully applied in both datasets, the potential impact of developmental change was ignored, potentially impacting the results obtained; further research is required to determine how best to capture both short-term and long-term developmental changes in such studies.

Sydnor et al. (2023) elegantly revealed the presence of age-related changes in the intrinsic fMRI activity along the sensorimotor-association (S-A) axis during adolescence. This S-A axis features a cortical hierarchy that allows neurodevelopmental plasticity to unfold in youths aged from 8 to 17 years (Sydnor et al., 2023; Sydnor et al., 2021). Moreover, the cortical regions that follow the S-A axis showed a significant relationship between intrinsic fMRI activity indexed by amplitude fluctuations and environmental scores such as neighbourhood factors with *greater* amplitudes found in the association cortex and *lower* amplitudes pronounced in the sensory and motor cortices (Sydnor et al., 2023). An important aspect of studying developmental changes lies in the

presence of heterogeneity across the whole-brain. Individual variations in functional connectivity patterns are heterogeneous in both the heteromodal association cortex — higher variability across brain regions which subserve an integrative role in the cortex, and unimodal cortices — lower variability across brain regions which are rather functionally specific (Xu et al., 2019; Mueller et al., 2013). While these heterogeneous patterns can be observed in individuals as young as ~31-42 postmenstrual weeks, they are non-uniform across the whole brain with *minimal* decrease in the individual variability patterns of the default mode, frontoparietal, and limbic networks, *moderate* decrease in the visual, attention, and subcortical networks followed by *significant* decrease in the sensorimotor networks (Xu et al., 2019). In neurotypical adults, the frontoparietal and attention networks demonstrate the *highest* level of functional variability followed by the default mode network with *moderate* variability, and sensorimotor-visual systems exhibiting *least* level of variability (Mueller et al., 2013). It is also important to remember that the ability to detect these developmental changes is often limited by the spatial and temporal resolutions of the chosen imaging modality. For example, fMRI exhibits a good spatial (~2-3 mm) but poor temporal (~5-6 seconds) resolutions (Figure 9.5). Detecting individual differences in functional connectivity that are reproducible would thus improve our ability to understand their underlying short-term and long-term associations with individual behavioural phenotypes in developing populations — perhaps with the contribution of a multimodal MRI approach.

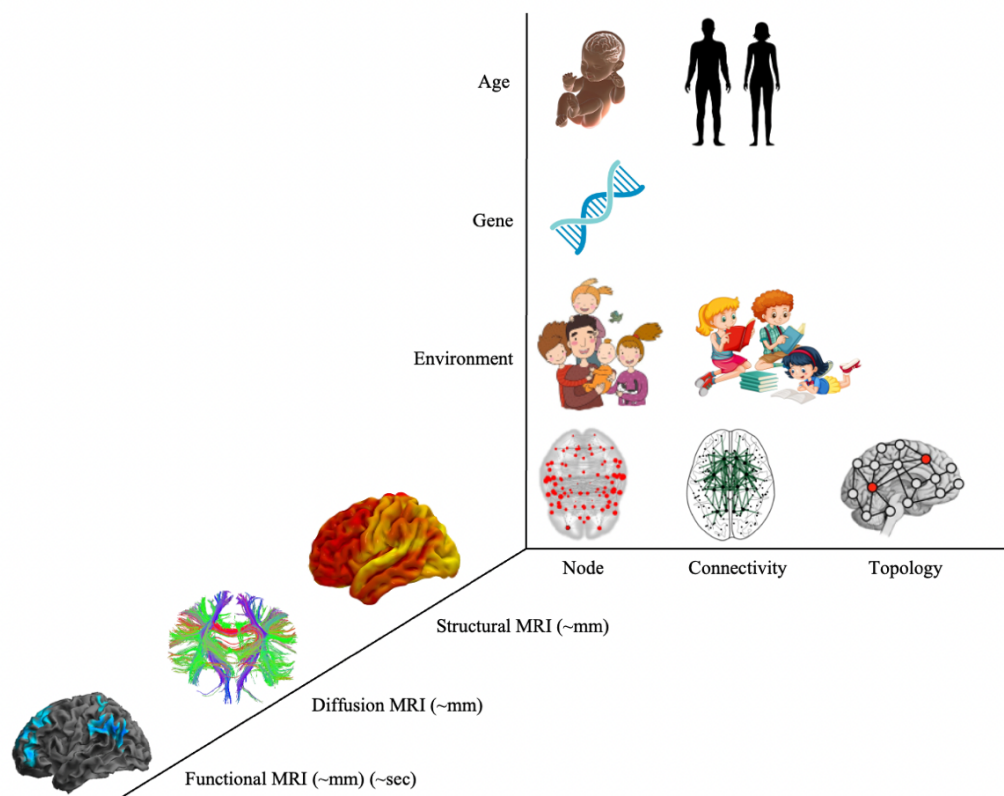


Figure 9.5. Factors capturing the rate of developmental changes such demographics, genetics, environment, and imaging modalities. Adapted from Cao et al., 2017.

9.7 A Quest for Multimodal MRI Applications

In searching for reproducible functional connectivity measures, I showed that there are individuals whose functional connectomes are always misidentified regardless of the data pre-/post-processing factors (see **Chapter 2**). In optimising the differentiation of an individual's connectome, I demonstrated that fingerprint ID accuracy was as high as 94% and the degree of confusion for misidentifying an individual was as low as 2%. The inability to achieve perfect connectome-based identification could potentially be explained by the anatomical variability that is present across individuals. At the same time, this motivates the study of structural connectomes as they share an ability to predict trait anxiety levels in youths and adults although these connectome-based predictions deserve methodological considerations to assess their reproducibility and generalisability in independent samples (Yoo et al., 2022; Feng et al., 2022; Zhang et al., 2019; see **Chapter 6**). Recently, suicide risk was predicted using multimodal connectomes (i.e., structural and functional connectivity) in a typical sample of late-life depression individuals (Gao et al., 2023). Individual patterns of structural and functional connectivity showed an ability to explain 14% and 30% of variance in suicide risk, respectively, in addition to classifying late-life depression participants into three distinct groups (i.e., non-suicidal, suicide-attempt, ideation/plan) with an accuracy of up to 90% that were generalisable in an independent dataset (Gao et al., 2023). Structural connectomes can successfully identify individuals during a perinatal period with an identifiability rate of 62% (Ciarrusta et al., 2022), which was greater than the self-self similarity of 10% derived from the individual functional connectome (Ciarrusta et al., 2022). Therefore, the individual structural connectome exhibits stable properties that are present in early development. High-resolution structural connectivity and morphological features can also enable stronger differentiation of an individual compared to the functional connectome, although the properties that uniquely identify individuals from a group may exhibit poor *inter*-individual differences in behavioural phenotypes (Mansour et al., 2021). The difference between individual identification and behavioural prediction derived from high-resolution multimodal connectivity features could thus better capture and provide meaningful insights into the *inter*-individual variations that are present in a range of mental health conditions (Mansour et al., 2021). The combination of structural and functional connectomes could inform us why some individuals remain difficult to ID using functional connectome fingerprints alone, and why improving functional connectome fingerprints do not necessarily strengthen or worsen the connectome-based phenotypic predictions, especially in developing youths.

The structure-function coupling has further revealed unique and stable individual characteristics that correspond to individual traits of cognitive and general behavioural domains which are

mostly relevant in the association cortex (Sun et al., 2022). These coupling patterns have an inherent ability to identify an individual from a cohort not only at rest but also across different brain states (Griffa et al., 2022). Although the structure-function coupling can be stronger for some brain networks compared to others (Suárez et al., 2020; Baum et al., 2019), it can serve as an indicator of *inter*-individual differences in macroscopic brain organisation (Griffa et al., 2022). The unique features, which underlie structure-function coupling, are both predictive of certain behavioural phenotypes and heritable though they can be influenced by age and sex (Gu et al., 2021). While the structure-function coupling has been found to be the highest in the visual and subcortical structures, and weakest in some transmodal areas (e.g., default mode, limbic), it could also reflect developmental changes that are localised within the frontoparietal and default mode regions (Baum et al., 2020; Honey et al., 2009). Given that the alignment of brain structure and function may not be perfect due to the presence of *intra*- and *inter*-variability, individual differences in structure-function coupling could provide complementary information in understanding the individual brain organisation across development, brain states, and clinical conditions (Suárez et al., 2020).

9.8 Limitations

The limitations of the four *empirical* projects have been described in their respective Chapters. Here, I describe the limitations of this thesis within the wider context of functional neuroimaging.

- ◆ The four *empirical* research projects have employed datasets with varying sample sizes, which makes it difficult to directly compare reproducibility across projects. Using a single large-*N* dataset, such as the ABCD Study, in future work could be a solution to this problem, since the ABCD consortium includes longitudinal multimodal imaging and deeply phenotyped data in developing youths. Of note, the minimally preprocessed ABCD Study samples would still substantially increase the computational resources required. Ideally, the consortium would share data preprocessed with a standardised pipeline such as fMRIPrep, addressing the toothbrush problem mentioned above, and reducing the duplication of effort and expenditure of computational resources across laboratories.
- ◆ The foundation of this thesis primarily relies on using datasets derived from resting-state fMRI conditions. Since task-based fMRI conditions have been shown to enhance individual differences in functional connectivity and to predict individual behavioural phenotypes across several domains, task-induced connectomes should be considered to provide a clearer picture of the reproducibility of functional connectivity-based measures in developing youths. A future line of work is required to examine the impact of bagging on the temporal

autocorrelated patterns of the fMRI signals in resting-state and task-based conditions as well as assessing the validity of the bagged functional connectomes in task-induced fMRI paradigms.

- ◆ The resting-state and naturalistic fMRI data used in this thesis have been acquired from relatively short scan lengths (~5 min) which proved to be sufficient in detecting individual patterns of functional connectivity that are unique, stable, and predictive of behavioural phenotypes. Nonetheless, future studies should determine if longer scan duration (e.g., > 10 min) could further boost the performances of the frameworks employed in this thesis. Similarly, it would be noteworthy to assess the efforts of several concatenating strategies that have been proposed in the literature to maximise the number of timepoints collected per individuals using the HCP and/or ABCD Study.
- ◆ While Chapter 8 employed a movie-watching paradigm to study emotional processing and regulation at multiple spatial and temporal scales, I did not quantify changes in emotional valence during the movie. Moreover, I used a relatively short animated movie clip (The Present) which is 4.18 min in length including the ending credits to study brain synchronisation, “static” and moment-to-moment functional connectivity patterns, and self-similarity structure of the BOLD fMRI signals. It remains important to assess how emotional valence scores, stimulus features (e.g., face expressions, scenes), and longer emotionally evocative movies influence the ability to capture spatial and temporal functional brain responses and their relationships with depressive symptomatology as well as adaptive and maladaptive emotion regulation strategies.
- ◆ For the last *empirical* research, I focused on a community-based sample of female adolescents who reported subclinical depressive symptoms on the Depression subscale of the Beck Youth Inventory (BYI-D). Although I adopted a dimensional approach and observed a wide range of BYI-D scores (ranging from 2 to 45), my findings may not be comparable to those obtained using categorical depression diagnoses or (e.g., melancholic depression, MDD) adult age groups. The use of the community sample exhibiting subclinical symptoms, combined with the small sample size, may have negatively impacted power and my ability to detect significant and robust brain-behaviour relationships.
- ◆ The four *empirical* projects employed both cross-sectional and longitudinal datasets, which can increase the difficulty of comparing the stability of the functional connectome and its potential phenotypic relationships. While the CoRR datasets that were employed in **Chapter**

2 included resting-state fMRI scans acquired over durations spanning several days to years, the HBN samples employed in **Chapters 4 and 6** obtained the fMRI scans within the same session. The impact of these temporal factors was not directly investigated. Similarly, a community-based sample of female adolescents exhibiting subclinical depressive symptoms was recruited to investigate naturalistic emotional processing and regulation cross-sectionally as presented in **Chapter 8**. An important caveat of cross-sectional studies is that the reliability of functional connectivity estimates could be influenced by subject-related variances such as stress, hormones, and sleep (Telzer et al., 2018).

- ◆ Most *empirical* projects in this thesis used Open Science data. This means I did not make certain methodological decisions (e.g., acquisition sequences, behavioural protocols) myself, and was restricted in terms of the analyses that I could apply following the hypotheses that I could propose (e.g., in terms of phenotypes I could examine). Another phenomenon refers to dataset decay, where researchers within the neuroimaging community re-analyse existing datasets with similar and/or different statistical testing procedures that may increase the occurrence of false positive results (Thompson et al., 2020). Given that methodological variability persists due to the lack of consensus in the pre-/post-processing strategies of fMRI data, re-analysing the same dataset could increase the already large number of statistical maps that exists due to different thresholds, multiple comparison corrections, ROI definitions, or neuroimaging software (e.g., FSL vs. SPM) amongst other decision points (Poldrack et al., 2017).

Beyond the future directions suggested by a consideration of these limitations, there are other potential research avenues that warrant consideration, including *meta-matching frameworks*, *multivariate techniques*, *multiverse pipelines*, and *sex differences*.

9.9 Future Directions

The *empirical* research conducted in this thesis focused on varying sample sizes from typical (e.g., $N = 24$ from the NYUado in CoRR) to consortia-level (e.g., $N = 540$ from the HBN). In this thesis, I have employed the components that describe two optimal connectome-based fingerprinting pipelines across five datasets ($N = 24-82$) to assess the robustness of connectome-based phenotypic predictions and generalisations in large-scale datasets. Given that a chosen phenotype P_1 derived from a typical sample-size study can be correlated, to certain extent, with another phenotype P_2 in a large-scale dataset, a predictive model trained on P_2 could be translated to predict P_1 . This meta-matching framework has the potential to extend my work by deriving robust and reproducible brain-behaviour associations with typical samples sizes (e.g., CoRR) by

applying meaningful features obtained from consortia datasets (e.g., HBN, ABIDE). Predictive models can exploit the functional connectivity features from these consortia ($N > 10,000$) to predict unseen phenotypes in datasets with typical sizes ($N \leq 200$) (He et al., 2022). Meta-matching models have recently been shown to boost brain-behaviour predictions with tens of participants compared to classical approaches (He et al., 2022), and have the ability to predict generalisable global cognitive functioning in psychiatric populations with gains in prediction accuracy ranging from 7% to 291% (Chopra et al., 2022). While the meta-matching framework shows there are common neurobiological features that are shared across psychiatric conditions for predicting cognition, such brain-behaviour predictions appear to be predominantly driven by the association networks related to the frontoparietal and default mode networks (Chopra et al., 2022). This framework will benefit the research conducted in this thesis by performing brain-behaviour associations using the relevant features in the functional connectome of the HBN that are predictive of a chosen behavioural phenotype to applying them to the independent datasets in CoRR. These brain-based features can further be investigated in terms of their generalisability within the context of CPM — as a means to improve the predictions of age, general psychopathology, and intellectual ability in an independent sample.

The *empirical* research that was conducted in this thesis mainly focused on developing, assessing, and validating univariate patterns of functional connectivity derived at multiple spatial and temporal scales. Given that the aim of this thesis lies in improving the reproducibility of functional connectivity-based measures in neurodevelopmental and psychiatric conditions, the application of multivariate techniques could provide complementary insights into the brain-behaviour associations that may not otherwise be detected by univariate models. Although multivariate methods can increase the dimensionality of imaging and phenotypic features in addition to reducing model interpretability, methods such as support vector regression (SVR; a machine learning model that can be used in classification problems or assigning classes when the data is not linearly separable) and CCA can yield stronger brain-behaviour associations compared to univariate models (Marek et al., 2022). Marek et al. (2022) showed that the maximum out-of-sample BWAS between functional connectivity and crystallised intelligence derived from SVR was more robust ($r = 0.39$) compared to the univariate correlation ($r = 0.16$) in the ABCD Study. Given that brain features are typically related to some extent, their multivariate patterns could boost their effect sizes instead of studying them individually, thus showing merit to improve the robustness of brain-behaviour associations (Rosenberg and Finn, 2022). Future work can include multivariate techniques such as CCA and SVR to further predict and generalise brain-behaviour relationships, especially in the context of general psychopathology — the CBCL scores which indexed externalising behaviours in **Chapter 6** were less sensitive to methodological factors and head motion strategies in developing youths.

As discussed above, a single research question or hypothesis can be answered using a range of functional connectivity analysis frameworks and pipelines — leading to increased methodological flexibility or “researcher degrees of freedom”. One solution is to establish multiverse analytic procedures (Dafflon et al., 2022; Botvinik-Nezer and Wager, 2022). They provide a platform for identifying the underlying relationships between different analytical factors that rest on data pre-/post-processing techniques such as regression, classification, motion correction strategies, and parcellation schemes. The variations in these methodological choices could provide meaningful insights into both the homogeneity and idiosyncrasy of different pipelines in addition to understanding how the multiverse space could be mapped to a low dimensional environment to control the amount of analyses being carried out. Importantly, we, as researchers, can focus on the strengths of different pipelines that are sparsely located in the multiverse space to form an ensemble analysis pipeline. Another advantage of establishing analytic pipelines in a multiverse space is the ability to evaluate the generalisability of a specific pipeline compared to other analysis methods by comparing their respective performances as well as assessing the generalisability of the pipelines in out-of-sample datasets to prevent overfitting a chosen dataset instead of attempting all the possible pipeline combinations. The data pre-/post-processing factors identified in **Chapter 2** and head motion strategies proposed in **Chapter 4** could be treated as a multiverse analysis pipeline for brain-behaviour predictions and generalisations as I have shown in **Chapter 6**. Nonetheless, future work is required to test whether such multiverse pipeline can derive robust and reproducible brain-behaviour associations in (sub) clinical populations such as those investigated in **Chapter 8** and extending other fMRI conditions (e.g., task-based, naturalistic).

During the *empirical* work conducted in this thesis, I estimated, predicted, and generalised brain-behaviour relationships by covarying sex differences using the HBN and ABIDE. To advance our search for functional connectivity-based biomarkers in neurodevelopmental and psychiatric conditions, treating sex as a biological variable of relevance is important; the connectome distinctiveness appears to be delayed in male compared to female youths (Kauffman et al., 2017; see **Chapter 1 (Section 1.7.2)**), and connectome-based predictions of memory functioning differ in male and female adults based on their underlying within- and between-network connections (Ju et al., 2023). In several behavioural paradigms, sex differences in learning, spatial navigation, decision-making, fear conditioning, or emotional processing do, in fact, reflect differential sex-dependent strategies (Shansky and Murphy, 2021; Shansky, 2018; Shansky and Woolley, 2016). Studying sex as a biological variable is also relevant across neuropsychiatric conditions as these disorders manifest differently in males and females by their age of onset, symptom severity, and treatment response. On the other hand, male and female individuals reveal overlapping patterns of brain-behaviour relationships across cognitive domains though the strengths of these

associations differ between the two sexes (Dhamala et al., 2021). Predictive models can also probe shared neurobiological features pertaining to internalising and externalising behaviours across sexes such that perturbations in functional connectivity patterns may represent unique psychiatric-related behaviours across males and females (Dhamala et al., 2023). Finally, there are sex differences in the functional network topology — association networks including the ventral attention, default mode, and frontoparietal networks probe the greatest *inter*-individual variations in functional topology in youths, wherein males exhibit greater network loadings than females in the default mode and frontoparietal networks whereas females express higher network loadings in the attention networks compared to males (Shanmugan et al., 2022). Another line of investigation requires the distinction between gender and sex differences on outcome measures (Regitz-Zagrosek and Gebhard, 2022). Given that gender and sex often have separable effects on clinical outcomes, the work carried out in this thesis could benefit from the examination of the distinct and potentially interacting effects of gender and sex on functional connectivity estimates.

9.10 Conclusions

The overarching aim of this thesis is to improve the reproducibility of functional connectivity-based measures, particularly in challenging developmental populations. I first demonstrated that functional connectome fingerprinting offers a way to compare, benchmark, and adjudicate functional connectivity analysis pipelines — as a means to assess the reproducibility of the individual functional connectome. This approach offers advantages over the traditional estimation of TRT reliability using the ICC. Using the connectome-based fingerprint identification procedure as a testbed of reproducibility, I identified a set of pre-/post-processing factors that improved the identifiability of an individual's functional connectome. Second, I implemented bagging or bootstrap aggregation for estimating the functional connectomes using low-motion timepoints. Bagging generated reproducible functional connectomes and maximised sample inclusivity by salvaging participants who would normally be discarded due to excess head movement. This enabled the retention of a large sample size, desirable for analyses of typical brain-behaviour associations. Third, I examined whether the pre-/post-processing factors that maximise fingerprint identification and head motion mitigation strategies (i.e., bagging and scrubbing) that maximise participant inclusivity with their motion-limited timeseries could improve connectome-based predictions of age, general psychopathology, and intellectual ability in developing youths. Unexpectedly, most factors I examined neither boosted nor worsened the phenotypic predictions nor their generalisability to an independent sample, with the exception of fine-grained brain parcellations and bagging, which improved generalisability. Finally, I took a different approach to reproducibility by exploring whether brain activity during a naturalistic movie-watching paradigm could be modulated by depressive symptom severity and emotion

regulation strategies in female adolescents exhibiting subclinical depressive symptoms. By proposing a movie-watching analysis framework, I showed that depressive symptomatology and emotion regulation strategies may shape how female adolescents process emotional information over time based on affect-laden events that reflect real-life scenarios.

By *empirically* examining the reproducibility of the functional connectome as defined by Botvinik and Wager (2022) in consortia and community-based datasets, I conclude that there is no single factor that can optimise functional connectivity-based measures in isolation. Improving the reproducibility of functional connectivity estimates is just one route to boosting the reproducibility of brain-behaviour relationships; strategies can be applied to either side of the brain-behaviour equation. These strategies include adequate study design paradigms, robust and reproducible methodological pipelines, and reliable behavioural phenotypes. Although there is some consensus within the neuroimaging community on how to address the potential sources of variability in functional connectivity estimates, I agree that it is not realistic to achieve perfect agreement during data collection, pre-processing, and post-processing stages of a study. I strongly recommend the neuroimaging community employ preprocessed Open-Source datasets (if their research questions and hypotheses permit) and/or apply publicly available methodological pipelines to minimise some of the variations during decision-making. On the basis of this thesis, I further suggest that sample size (in the thousands) is not the only route to generate reproducible functional connectivity estimates, nor robust and reproducible brain-behaviour relationships. In addition, I argue that it is important to think about the validity of the functional connectome when developing methods to improve its reproducibility — an individual's functional connectome can be reproducible but exhibit low validity, thus having low utility to observe meaningful brain-behaviour relationships.

The contributions made by this thesis are: (1) connectome-based fingerprinting can serve as a testbed for reproducible functional connectivity-based analysis methods; (2) bagging can achieve a balance between data quality standards and sample size requirements by retaining “high-movement” participants for robust and reproducible brain-behaviour association studies; (3) while the strengths of connectome-based phenotypic predictions appear robust to methodological choices, bagging improves the generalisability of the CPM-based predictions; and (4) naturalistic movie-watching paradigms offer a rich context for the exploration of how phenotypic variables, such as depressive symptoms, affect information processing evoked by real-life scenarios. **A final recommendation arising from this work is that the neuroimaging community should remember to put the participants at the heart of their research questions — doing so means that we, as researchers, should apply the lens of reproducibility to our efforts to develop study design protocols, methodological pipelines, and reliable phenotypic measurements.**

My hope is that centring this perspective will enable us to move a step closer to discovering functional connectivity-based biomarkers that will make a real difference to the lives of our participants and their families.

References

- Ahmed, S., Bittencourt-Hewitt, A., & Sebastian, C. (2015). Neurocognitive bases of emotion regulation development in adolescence. *Developmental Cognitive Neuroscience*, 15, 11-25.
- Abbas, K., Liu, M., Venkatesh, M., Amico, E., Kaplan, A., Ventresca, M. et al. (2021). Geodesic distance on optimally regularized functional connectomes uncovers individual fingerprints. *Brain Connectivity*, 11 (5), 333-348.
- Abrams, D., Ryali, S., Chen, T., Chordia, P., Khouzam, A., Levitin, D., & Menon, V. (2013). Inter-subject synchronization of brain responses during natural music listening. *European Journal Of Neuroscience*, 37 (9), 1458-1469.
- Achenbach, T., & Edelbrock, C. (1978). The classification of Child psychopathology: A review and analysis of empirical efforts. *Psychological Bulletin*, 85 (6), 1275–1301.
- Achenbach, T., & Rescorla, L. (2001). *Manual for the ASEBA School-Age Forms & Profiles: An integrated system of multi-informant assessment*. Burlington, VT: University of Vermont.
- Airan, R., Vogelstein, J., Pillai, J., Caffo, B., Pekar, J., & Sair, H. (2016). Factors affecting characterization and localization of interindividual differences in functional connectivity using MRI. *Human Brain Mapping*, 37 (5), 1986-1997.
- Albert, N., Robertson, E., & Miall, R. (2009). The resting human brain and motor learning. *Current biology: CB*, 19 (12), 1023-1027.
- Aldao, A., Nolen-Hoeksema, S., & Schweitzer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review*, 30 (2), 217-237.
- Alexander, L., Escalera, J., Ai, L., Andreotti, C., & Febre, K. et al. (2017). An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Sci Data* 4, 170181.
- Aliko, S., Huang, J., Gheorghiu, F., Meliss, S., & Skipper, J. (2020). A naturalistic neuroimaging database for understanding the brain using ecological stimuli. *Scientific Data*, 7 (347).
- Allen, E., Erhardt, E., Damaraju, E., Gruner, W., Segall, J., & Silva, R. et al. (2011). A Baseline for the Multivariate Comparison of Resting-State Networks. *Frontiers In Systems Neuroscience*, 5.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).
- Aminoff, E., Kveraga, K., & Bar, M. (2013). The role of the parahippocampal cortex in cognition. *Trends In Cognitive Sciences*, 17 (8), 379-390.
- Anand, A., Li, Y., Wang, Y., Wu, J., Gao, S., & Bukhari, L. et al. (2005). Activity and Connectivity of Brain Mood Regulating Circuit in Depression: A Functional Magnetic Resonance Study. *Biological Psychiatry*, 57 (10), 1079-1088.

- Anderson, J., Ferguson, M., Lopez-Larson, M., & Yurgelun-Todd, D. (2011). Reproducibility of Single-Subject Functional Connectivity Measurements. *American Journal Of Neuroradiology*, 32 (3), 548-555.
- Anderson, S., Kelley, K., & Maxwell, S. (2017). Sample-Size Planning for More Accurate Statistical Power: A Method Adjusting Sample Effect Sizes for Publication Bias and Uncertainty. *Psychological Science*, 28 (11), 1547-1562.
- Andrews-Hanna, J., Snyder, A., Vincent, J., Lustig, C., Head, D., Raichle, M. E., & Buckner, R. (2007). Disruption of large-scale brain systems in Advanced Aging. *Neuron*, 56 (5), 924–935.
- Arslan, S., Ktena, S., Makropoulos, A., Robinson, E., Rueckert, D., & Parisot, S. (2018). Human brain mapping: A systematic comparison of parcellation methods for the human cerebral cortex. *Neuroimage*, 170, 5-30.
- Arbabshirani, M., Damaraju, E., Phlypo, R., Plis, S., Allen, E., Ma, S., et al. (2014). Impact of autocorrelation on functional connectivity. *NeuroImage*, 102, 294–308.
- Astle, D., Holmes, J., Kievit, R., & Gathercole, S. (2021). Annual research review: The transdiagnostic revolution in neurodevelopmental disorders. *Journal of Child Psychology and Psychiatry*, 63 (4), 397-417.
- Bailey, S., Aboud, K., Nguyen, T., & Cutting, L. (2018). Applying a network framework to the Neurobiology of Reading and Dyslexia. *Journal of Neurodevelopmental Disorders*, 10 (37).
- Bakhit, M., Fujii, M., Hiruta, R., Yamada, M., Iwami, K., Sato, T., & Saito, K. (2020). The superior frontal longitudinal tract: a connection between the dorsal premotor and the dorsolateral prefrontal cortices. *Scientific Reports*, 10 (1).
- Baldassano, C., Chen, J., Zadbood, A., Pillow, J., Hasson, U., & Norman, K. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron*, 95 (3).
- Baldassarre, A., Metcalf, N., Shulman, G., & Corbetta, M. (2019). Brain Networks' functional connectivity separates aphasic deficits in stroke. *Neurology*, 92 (2).
- Bandettini, P., Gonzalez-Castillo, J., Handwerker, D., Taylor, P., Chen, G., & Thomas, A. (2022). The challenge of bwas: Unknown unknowns in feature space and variance. *Med*, 3 (8), 526-531.
- Barron, D., Gao, S., Dadashkarimi, J., Greene, A., Spann, M., Noble, S. et al. (2020). Transdiagnostic, connectome-based prediction of memory constructs across psychiatric disorders. *Cerebral Cortex*, 31 (5), 2523-2533.
- Bartels, A., & Zeki, S. (2005). Brain Dynamics during natural viewing conditions—a new guide for mapping connectivity in vivo. *NeuroImage*, 24 (2), 339–349.
- Bartko, J. (1976). On various intraclass correlation reliability coefficients. *Psychological Bulletin*, 83 (5), 762-765.

- Baum, G., Cui, Z., Roalf, D., Ciric, R., Betzel, R., & Larsen, B. et al. (2019). Development of structure–function coupling in human brain networks during youth. *Proceedings Of The National Academy Of Sciences*, *117* (1), 771-778.
- Beauregard, M., Paquette, V., & Le'vesque, J. (2006). Dysfunction in the neural circuitry of emotional self-regulation in major depressive disorder. *Neuroreport*, *17* (8), 843-846.
- Beaty, R., Kenett, Y., Christensen, A., Rosenberg, M., Benedek, M., Chen, Q. et al. (2018). Robust prediction of individual creative ability from brain functional connectivity. *Proceedings of the National Academy of Sciences*, *115* (5), 1087-1092.
- Beck, J., Beck, A., Jolly, J., & Steer, R. (2005). *Beck youth inventories for children and adolescents*. Psychopath.
- Beckmann, C., DeLuca, M., Devlin, J., & Smith, S. (2005). Investigations into resting-state connectivity using independent component analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360* (1457), 1001-1013.
- Bellec, P., Benhajali, Y., Carbonell, F., Dansereau, C., Albouy, G., & Pelland, M. et al. (2015). Impact of the resolution of brain parcels on connectome-wide association studies in fMRI. *Neuroimage*, *123*, 212-228.
- Bellec, P., Rosa-Neto, P., Lyttelton, O., Benali, H., & Evans, A. (2010). Multi-level bootstrap analysis of stable clusters in resting-state fmri. *NeuroImage*, *51* (3), 1126–1139.
- Berkman, E., & Lieberman, M. (2009). Using Neuroscience to Broaden Emotion Regulation: Theoretical and Methodological Considerations. *Social And Personality Psychology Compass*, *3* (4), 475-493.
- Bernal, B., Ardila, A., & Rosselli, M. (2015). Broca's area network in language function: a pooling-data connectivity study. *Frontiers in Psychology*, *6*.
- Birn, R., Molloy, E., Patriat, R., Parker, T., Meier, T., & Kirk, G. et al. (2013). The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *Neuroimage*, *83*, 550-558.
- Biswal, B., Mennes, M., Zuo, X., Gohel, S., Kelly, C., & Smith, S. et al. (2010). Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences*, *107* (10), 4734-4739.
- Biswal, B., Taylor, P., & Ulmer, J. (2001). Use of jackknife resampling techniques to estimate the confidence intervals of fmri parameters. *Journal of Computer Assisted Tomography*, *25* (1), 113-120.
- Biswal, B., Zerrin Yetkin, F., Haughton, V., & Hyde, J. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic Resonance in Medicine*, *34* (4), 537-541.

- Bolt, T., Wang, S., Nomi, J., Setton, R., Gold, B., Frederick, B. et al. (2023). A unified physiological process links global patterns of functional MRI, respiratory activity, and autonomic signaling. *bioRxiv*. 2023.01.19.524818.
- Botvinik-Nezer, R., Holzmeister, F., Camerer, C., Dreber, A., Huber, J., & Johannesson, M. et al. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, 582, 84-88.
- Botvinik-Nezer, R., & Wager, T. (2022). Reproducibility in neuroimaging analysis: Challenges and Solutions. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*.
- Bowring, A., Maumet, C., & Nichols, T. (2019). Exploring the impact of analysis software on task fmri results. *Human Brain Mapping*, 40 (11), 3362-3384.
- Boyle, R., Connaughton, M., McGlinchey, E., Knight, S., De Looze, C., Carey, D. et al. (2023). Connectome-based predictive modelling of Cognitive Reserve using task-based functional connectivity. *European Journal of Neuroscience*, 57 (3), 490-510.
- Brauchli, C., Leipold, S., & Jäncke, L. (2020). Diminished large-scale functional brain networks in absolute pitch during the perception of naturalistic music and audiobooks. *Neuroimage*, 216, 116513
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24 (2), 123-140.
- Bressler, S., & Menon, V. (2010). Large-scale brain networks in cognition: emerging methods and principles. *Trends in Cognitive Sciences*, 14 (6), 277-290.
- Brosnan, M., Shalev, N., Ramduny, J., Sotiropoulos, S., & Chechlac, M. (2022). Right fronto-parietal networks mediate the neurocognitive benefits of enriched environments. *Brain Communications*, 4 (2).
- Bruchhage, M., Ngo, G., Schneider, N., D'Sa, V., & Deoni, S. (2020). Functional connectivity correlates of infant and early childhood cognitive development. *Brain Structure and Function*, 225 (2), 669-681.
- Brunswik, E., 1947. Perception and the Representative Design of Psychological Experiments. University of California Press, Berkeley, CA.
- Brunswik, E., 1949. Systematic and representative design of psychological experiments with results in physical and social perception. In: Neyman, J. (Ed.), *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, Berkeley, CA, 143-202.
- Bryce, N., Flournoy, J., Guassi Moreira, J., Rosen, M., Sambook, K., Mair, P., & McLaughlin, K. (2021). Brain Parcellation Selection: An overlooked decision point with meaningful effects on individual differences in resting-state functional connectivity. *NeuroImage*, 243, 118487.
- Buch, A., & Liston, C. (2021). Dissecting diagnostic heterogeneity in depression by integrating neuroimaging and genetics. *Neuropsychopharmacology*, 46 (1), 156-175.

- Buckner, R., Krienen, F., Castellanos, A., Diaz, J., & Yeo, B.T. (2011). The organization of the human cerebellum estimated by intrinsic functional connectivity. *Journal Of Neurophysiology*, 106 (5), 2322-2345.
- Bullmore, E., Barnes, A., Bassett, D., Fornito, A., Kitzbichler, M., Meunier, D., & Suckling, J. (2009). Generic aspects of complexity in brain imaging data and other biological systems. *NeuroImage*, 47 (3), 1125-1134.
- Button, K. (2016). Statistical Rigor and the Perils of Chance. *Eneuro*, 3 (4), ENEURO.0030-16.2016.
- Button, K., Ioannidis, J., Mokrysz, C., Nosek, B., Flint, J., Robinson, E., & Munafò, M. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14 (5), 365-376.
- Byrge, L., Dubois, J., Tyszka, J., Adolphs, R., & Kennedy, D. (2015). Idiosyncratic Brain Activation Patterns Are Associated with Poor Social Comprehension in Autism. *Journal Of Neuroscience*, 35 (14), 5837-5850.
- Byrge, L., & Kennedy, D. (2019). High-accuracy individual identification using a “thin slice” of the functional connectome. *Network Neuroscience*, 3 (2), 363-383.
- Byrge, L., & Kennedy, D. (2018). Identifying and characterizing systematic temporally-lagged bold artifacts. *NeuroImage*, 171, 376-392.
- Cai, B., Zhang, G., Zhang, A., Xiao, L., Hu, W., Stephen, J. et al. (2021). Functional connectome fingerprinting: Identifying individuals and predicting cognitive functions via autoencoder. *Human Brain Mapping*, 42 (9), 2691-2705.
- Calhoun, V., Miller, R., Pearlson, G., & Adalı, T. (2014). The Chronnectome: Time-Varying Connectivity Networks as the Next Frontier in fMRI Data Discovery. *Neuron*, 84 (2), 262-274.
- Califf, R. (2018). Biomarker definitions and their applications. *Experimental Biology and Medicine*, 243 (3), 213–221.
- Campbell, K., Grady, C., Ng, C., & Hasher, L. (2012). Age differences in the frontoparietal cognitive control network: Implications for distractibility. *Neuropsychologia*, 50 (9), 2212-2223.
- Campbell, S., & Macqueen, G. (2004). The role of the hippocampus in the pathophysiology of major depression. *Journal of psychiatry & neuroscience : JPN*, 29 (6), 417–426.
- Campbell, O., Vanderwal, T., & Weber, A. (2022b). Fractal-based analysis of fmri bold signal during naturalistic viewing conditions. *Frontiers in Physiology*, 12.
- Campbell, O., & Weber, A. (2022a). Monofractal analysis of Functional Magnetic Resonance Imaging: An introductory review. *Human Brain Mapping*, 43 (8), 2693-2706.
- Cantlon, J., & Li, R. (2013). Neural Activity during Natural Viewing of Sesame Street Statistically Predicts Test Scores in Early Childhood. *Plos Biology*, 11 (1), e1001462.
- Cao, M., Huang, H., & He, Y. (2017). Developmental connectomics from infancy through early childhood. *Trends in Neurosciences*, 40 (8), 494-506.

- Carp, J. (2012). On the plurality of (methodological) worlds: Estimating the analytic flexibility of fmri experiments. *Frontiers in Neuroscience*, 6 (149).
- Casey, B., Cannonier, T., Conley, M., Cohen, A., Barch, D., & Heitzeg, M. et al. (2018). The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience*, 32, 43-54.
- Castellanos, F., & Aoki, Y. (2016). Intrinsic functional connectivity in attention-deficit/hyperactivity disorder: A science in development. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 1 (3), 253–261.
- Castellanos, F., Di Martino, A., Craddock, R., Mehta, A., & Milham, M. (2013). Clinical applications of the functional connectome. *Neuroimage*, 80, 527-540.
- Cecchetti, L., & Handjaras, G. (2022). Reproducible brain-wide association studies do not necessarily require thousands of individuals. PsyArXiv.
- Chai, X., Castanon, A., Ongur, D., & Whitfield-Gabrieli, S. (2012). Anticorrelations in resting state networks without global signal regression. *Neuroimage*, 59 (2), 1420-1428.
- Chamberlain, T., & Rosenberg, M. (2022). Propofol selectively modulates functional connectivity signatures of sustained attention during rest and narrative listening. *Cerebral Cortex*, 32 (23), 5362-5375.
- Chambers, C., Feredoes, E., Muthukumaraswamy, S., & Etchells, P. (2014). Instead of “playing the game” it is time to change the rules: Registered reports at AIMS neuroscience and beyond. *AIMS Neuroscience*, 1 (1), 4–17.
- Chang, L., Jolly, E., Cheong, J., Rapuano, K., Greenstein, N., Chen, P., & Manning, J. (2021). Endogenous variation in ventromedial prefrontal cortex state dynamics during naturalistic viewing reflects affective experience. *Science Advances*, 7 (17).
- Chen, J., Leong, Y., Honey, C., Yong, C., Norman, K., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, 20 (1), 115–125.
- Chen, G., Shin, Y., Taylor, P., Glen, D., Reynolds, R., Israel, R., & Cox, R. (2016). Untangling the relatedness among correlations, part I: Nonparametric approaches to inter-subject correlation analysis at the group level. *Neuroimage*, 142, 248-259.
- Chen, G., Taylor, P., Haller, S., Kircanski, K., Stoddard, J., Pine, D. et al. (2017). Intraclass correlation: Improved modeling approaches and applications for neuroimaging. *Human Brain Mapping*, 39 (3), 1187-1206.
- Chen, Z., Xia, M., Zhao, Y., Kuang, W., Jia, Z., & Gong, Q. (2021). Characteristics of Intrinsic Brain Functional Connectivity Alterations in Major Depressive Disorder Patients With Suicide Behavior. *Journal Of Magnetic Resonance Imaging*, 54 (6), 1867-1875.
- Cho, J., Korchmaros, A., Vogelstein, J., Milham, M., & Xu, T. (2020). Impact of concatenating fMRI data on reliability for functional connectomics. *Neuroimage*, 226, 117549.

- Choe, A., Jones, C., Joel, S., Muschelli, J., Belegu, V., & Caffo, B. et al. (2015). Reproducibility and Temporal Structure in Weekly Resting-State fMRI over a Period of 3.5 Years. *PLOS ONE*, 10 (10), e0140134.
- Chong, J., Ng, K., Tandi, J., Wang, C., Poh, J., Lo, J. et al. (2019). Longitudinal changes in the cerebral cortex functional organization of healthy elderly. *The Journal of Neuroscience*, 39 (28), 5534–5550.
- Chopra, S., Dhamala, E., Lawhead, C., Ricard, J., Orchard, E., An, L. et al. Reliable and generalizable brain-based predictions of cognitive functioning across common psychiatric illness. medRxiv 2022.12.08.22283232.
- Christoff, K., Gordon, A., Smallwood, J., Smith, R., & Schooler, J. (2009). Experience sampling during fmri reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences*, 106 (21), 8719-8724.
- Churchill, N., Spring, R., Grady, C., Cimprich, B., Askren, M., Reuter-Lorenz, P. et al. (2016). The suppression of scale-free fmri brain dynamics across three different sources of effort: Aging, Task Novelty and task difficulty. *Scientific Reports*, 6 (1).
- Ciarrusta, J., Christiaens, D., Fitzgibbon, S., Dimitrova, R., Hutter, J., Hughes, E. et al. (2022). The developing brain structural and functional connectome fingerprint. *Developmental Cognitive Neuroscience*, 55, 101117.
- Cicchetti, D. V., & Sparrow S. A. (1981). Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *American journal of mental deficiency*, 86 (2), 127-137.
- Ciric, R., Wolf, D., Power, J., Roalf, D., Baum, G., & Ruparel, K. et al. (2017). Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *Neuroimage*, 154, 174-187.
- Ciuciu, P., Varoquaux, G., Abry, P., Sadaghiani, S., & Kleinschmidt, A. (2012). Scale-free and multifractal time dynamics of fmri signals during rest and task. *Frontiers in Physiology*, 3.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20 (1), 37–46.
- Cole, M., Reynolds, J., Power, J., Repovs, G., Anticevic, A., & Braver, T. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature Neuroscience*, 16 (9), 1348-1355.
- Colenbier, N., Van de Steen, F., Uddin, L., Poldrack, R., Calhoun, V., & Marinazzo, D. (2020). Disambiguating the role of blood flow and global signal with partial information decomposition. *Neuroimage*, 213, 116699.
- Constable, R., Scheinost, D., Finn, ES., Shen, X., Hampson, M., Winstanley, F. et al. (2013). Potential use and challenges of functional connectivity mapping in intractable epilepsy. *Frontiers in Neurology*, 4 (39).

- Cooper, R., & Ritchey, M. (2019). Cortico-hippocampal network connections support the multidimensional quality of episodic memory. *Elife*, 8, e45591.
- Corbetta, M., Patel, G., & Shulman, G. (2008). The Reorienting System of the Human Brain: From Environment to Theory of Mind. *Neuron*, 58 (3), 306-324.
- Corbetta, M., & Shulman, G. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3 (3), 201-215.
- Craddock, C., Benhajali, Y., Chu, C., Chouinard, F., Evans, A., Jakab, A., et al. (2013a). The Neuro Bureau Preprocessing Initiative: open sharing of preprocessed neuroimaging data and derivatives. *Front. Neuroinform. Conference Abstract: Neuroinformatics*.
- Craddock, R., Jbabdi, S., Yan, C., Vogelstein, J., Castellanos, F., & Di Martino, A. et al. (2013b). Imaging human connectomes at the macroscale. *Nature Methods*, 10 (6), 524-539.
- Craddock, C., Sikka, S., Cheung, B., Khanuja, R., Ghosh, S., & Yan, C. et al. (2013c). Towards Automated Analysis of Connectomes: The Configurable Pipeline for the Analysis of Connectomes (C-PAC). *Frontiers In Neuroinformatics*, 7.
- Cremers, H., Wager, T., & Yarkoni, T. (2017). The relation between statistical power and inference in fMRI. *PLOS ONE*, 12 (11), e0184923.
- Cuthbert, B., & Insel, T. (2013). Toward the future of psychiatric diagnosis: The seven pillars of rdoc. *BMC Medicine*, 11 (1).
- Dadi, K., Rahim, M., Abraham, A., Chyzyk, D., Milham, M., Thirion, B., & Varoquaux, G. (2019). Benchmarking functional connectome-based predictive models for resting-state fmri. *NeuroImage*, 192, 115-134.
- Dafflon, J., Da Costa, P., Váša, F., Monti, R., Bzdok, D., Hellyer, P. et al. (2022). A guided multiverse study of neuroimaging analyses. *Nature Communications*, 13 (1).
- Damaraju, E., Allen, E., Belger, A., Ford, J., McEwen, S., Mathalon, D. et al. (2014). Dynamic functional connectivity analysis reveals transient states of dysconnectivity in Schizophrenia. *NeuroImage: Clinical*, 5, 298–308.
- Damoiseaux, J., Rombouts, S., Barkhof, F., Scheltens, P., Stam, C., Smith, S., & Beckmann, C. (2006). Consistent resting-state networks across healthy subjects. *Proceedings of The National Academy of Sciences*, 103 (37), 13848-13853.
- Davatzikos, C. (2019). Machine learning in neuroimaging: Progress and challenges. *NeuroImage*, 197, 652-656.
- Demirtaş, M., Ponce-Alvarez, A., Gilson, M., Hagmann, P., Mantini, D., & Betti, V. et al. (2019). Distinct modes of functional connectivity induced by movie-watching. *Neuroimage*, 184, 335-348.
- Desikan, R., Ségonne, F., Fischl, B., Quinn, B., Dickerson, B., & Blacker, D. et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, 31 (3), 968-980.

- DeYoung, C., Sassenberg, T., Abend, R., Allen, T., Beaty, R., Bellgrove, M. et al. (2022). Reproducible between-person brain-behavior associations do not always require thousands of individuals. *PsyArXiv*.
- Dhamala, E., Yeo, B.T., & Holmes, A. (2022). One size does not fit all: Methodological considerations for Brain-based predictive modeling in psychiatry. *Biological Psychiatry*.
- Di, X., & Biswal, B. (2020). Intersubject consistent dynamic connectivity during natural vision revealed by functional MRI. *NeuroImage*, 216, 116698.
- Di, X., & Biswal, B. (2022a). Principal component analysis reveals multiple consistent responses to naturalistic stimuli in children and adults. *Human Brain Mapping*, 43 (11), 3332–3345.
- Di, X., Xu, T., Uddin, L., & Biswal, B. Individual differences in time-varying and stationary brain connectivity during movie watching from childhood to early adulthood: age, sex, and behavioral associations. *bioRxiv* 2023.01.30.526311.
- Di, X., Zhang, Z., Xu, T., & Biswal, B. (2022b). Dynamic and stationary brain connectivity during movie watching as revealed by functional MRI. *Brain Structure and Function*, 227 (7), 2299–2312.
- Di Martino, A., Fair, D., Kelly, C., Satterthwaite, T., Castellanos, F., & Thomason, M. et al. (2014a). Unraveling the Miswired Connectome: A Developmental Perspective. *Neuron*, 83 (6), 1335-1353.
- Di Martino, A., Kelly, C., Grzadzinski, R., Zuo, X., Mennes, M., & Mairena, M. et al. (2011). Aberrant Striatal Functional Connectivity in Children with Autism. *Biological Psychiatry*, 69 (9), 847-856.
- Di Martino, A., O'Connor, D., Chen, B., Alaerts, K., Anderson, J., Assaf, M., et al. (2017). Enhancing studies of the connectome in autism using the Autism Brain Imaging Data Exchange II. *Sci Data* 4, 170010.
- Di Martino, A., Yan, C., Li, Q., Denio, E., Castellanos, F., & Alaerts, K. et al. (2014b). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry*, 19 (6), 659-667.
- Dichter G. (2012). Functional magnetic resonance imaging of autism spectrum disorders. *Dialogues in clinical neuroscience*, 14 (3), 319-351.
- Dinga, R., Schmaal, L., Penninx, B.W.J.H., van Tol, M.J., & Veltman, D.J. et al. (2019). Evaluating the evidence for biotypes of depression: Methodological replication and extension of Drysdale et al. *NeuroImage: Clinical*, 22, 101796.
- Domaradzka, E., & Fajkowska, M. (2018). Cognitive Emotion Regulation Strategies in Anxiety and Depression Understood as Types of Personality. *Frontiers In Psychology*, 9.
- Dona, O., Hall, G., & Noseworthy, M. (2017a). Temporal fractal analysis of the rs-BOLD signal identifies brain abnormalities in autism spectrum disorder. *PLOS ONE*, 12 (12), e0190081.

Dona, O., Noseworthy, M., DeMatteo, C., & Connolly, J. (2017b). Fractal Analysis of Brain Blood Oxygenation Level Dependent (BOLD) Signals from Children with Mild Traumatic Brain Injury (mTBI). *PLOS ONE*, *12* (1), e0169647.

Dong, J., Jing, B., Ma, X., Liu, H., Mo, X., & Li, H. (2018a). Hurst exponent analysis of resting-state fmri signal complexity across the adult lifespan. *Frontiers in Neuroscience*, *12*.

Dong, D., Wang, Y., Chang, X., Luo, C., & Yao, D. (2018b). Dysfunction of large-scale brain networks in schizophrenia: A meta-analysis of resting-state functional connectivity. *Schizophrenia Bulletin*, *44* (1), 168–181.

Dosenbach, N., Fair, D., Cohen, A., Schlaggar, B., & Petersen, S. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences*, *12* (3), 99-105.

Dosenbach, N., Koller, J., Earl, E., Miranda-Dominguez, O., Klein, R., & Van, A. et al. (2017). Real-time motion analytics during brain MRI improve data quality and reduce costs. *Neuroimage*, *161*, 80-93.

Dosenbach, N., Nardos, B., Cohen, A., Fair, D., Power, J., Church, J. et al. (2010). Prediction of individual brain maturity using fmri. *Science*, *329* (5997), 1358-1361.

Driver, J., & Spence, C. (2000). Multisensory perception: Beyond modularity and convergence. *Current Biology*, *10* (20), R731-R735.

Drysdale, A., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., & Meng, Y. et al. (2017). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine*, *23* (1), 28-38.

Du, H., Liao, X., Lin, Q., Li, G., Chi, Y., Liu, X. et al. (2015). Test-retest reliability of graph metrics in high-resolution functional connectomics: A resting-state functional MRI study. *CNS Neuroscience & Therapeutics*, *21* (10), 802–816.

Dubois, J., Galdi, P., Han, Y., Paul, L., & Adolphs, R. (2018a). Resting-state functional brain connectivity best predicts the personality dimension of openness to experience. *Personality Neuroscience*, *1*.

Dubois, J., Galdi, P., Paul, L., & Adolphs, R. (2018b). A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373* (1756), 20170284.

Dudoit, S., & Fridlyand, J. (2003). Bagging to improve the accuracy of a clustering procedure. *Bioinformatics*, *19* (9), 1090-1099.

Dufford, A., Noble, S., Gao, S., & Scheinost, D. (2021). The instability of functional connectomes across the first year of life. *Developmental Cognitive Neuroscience*, *51*, 101007.

Durston, S., Tottenham, N., Thomas, K., Davidson, M., Eigsti, I., Yang, Y. et al. (2003). Differential patterns of striatal activation in young children with and without ADHD. *Biological Psychiatry*, *53* (10), 871–878.

- Easson, A., Fatima, Z., & McIntosh, A. (2019). Functional connectivity-based subtypes of individuals with and without autism spectrum disorder. *Network Neuroscience*, 3 (2), 344-362.
- Ebneabbasi, A., Mahdipour, M., Nejati, V., Li, M., Liebe, T., Colic, L. et al.. (2021). Emotion processing and regulation in major depressive disorder: A 7T resting-state fMRI study. *Human brain mapping*, 42 (3), 797–810.
- Ecker, C., Marquand, A., Mourao-Miranda, J., Johnston, P., Daly, E., & Brammer, M. et al. (2010). Describing the Brain in Autism in Five Dimensions—Magnetic Resonance Imaging-Assisted Diagnosis of Autism Spectrum Disorder Using a Multiparameter Classification Approach. *Journal of Neuroscience*, 30 (32), 10612-10623.
- Eickhoff, S., Milham, M., & Vanderwal, T. (2020). Towards clinical applications of movie fMRI. *Neuroimage*, 217, 116860.
- Eickhoff, S., Thirion, B., Varoquaux, G., & Bzdok, D. (2015). Connectivity-based parcellation: Critique and implications. *Human Brain Mapping*, 36 (12), 4771-4792.
- Eickhoff, S., Yeo, BT., & Genon, S. (2018). Imaging-based parcellations of the human brain. *Nature Reviews Neuroscience*, 19 (11), 672-686.
- Eke, A., Hermán, P., Bassingthwaighe, J., Raymond, G., Percival, D., Cannon, M. et al. (2000). Physiological time series: Distinguishing fractal noises from motions. *Pflügers Archiv - European Journal of Physiology*, 439 (4), 403-415.
- Elliott, M., Knodt, A., Cooke, M., Kim, M., Melzer, T., & Keenan, R. et al. (2019). General functional connectivity: Shared features of resting-state and task fMRI drive reliable and heritable individual differences in functional brain networks. *Neuroimage*, 189, 516-532.
- Emerson, R., Short, S., Lin, W., Gilmore, J., & Gao, W. (2015). Network-level connectivity dynamics of movie watching in 6-year-old children. *Frontiers in Human Neuroscience*, 9.
- Engen, H., & Anderson, M. (2018). Memory Control: A Fundamental Mechanism of Emotion Regulation. *Trends In Cognitive Sciences*, 22 (11), 982-995.
- Erk, S., Mikschl, A., Stier, S., Ciaramidaro, A., Gapp, V., Weber, B., & Walter, H. (2010). Acute and Sustained Effects of Cognitive Emotion Regulation in Major Depression. *Journal Of Neuroscience*, 30 (47), 15726-15734.
- Esteban, O., Markiewicz, C., Blair, R., Moodie, C., Isik, A., & Erramuzpe, A. et al. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, 16 (1), 111-116.
- Etkin, A., Büchel, C., & Gross, J. (2015). The neural bases of emotion regulation. *Nature Reviews Neuroscience*, 16 (11), 693-700.
- Expert, P., Lambiotte, R., Chialvo, D., Christensen, K., Jensen, H., Sharp, D., & Turkheimer, F. (2011). Self-similar correlation function in brain resting-state functional magnetic resonance imaging. *Journal Of The Royal Society Interface*, 8 (57), 472-479.

Fair, D., Bathula, D., Nikolas, M., & Nigg, J. (2012). Distinct neuropsychological subgroups in typically developing youth inform heterogeneity in children with ADHD. *Proceedings of The National Academy of Sciences*, 109 (17), 6769-6774.

Fair, D., Cohen, A., Power, J., Dosenbach, N., Church, J., & Miezin, F. et al. (2009). Functional Brain Networks Develop from a "Local to Distributed" Organization. *Plos Computational Biology*, 5 (5), e1000381.

Fair, D., Dosenbach, N., Church, J., Cohen, A., Brahmbhatt, S., Miezin, F. et al. (2007). Development of distinct control networks through segregation and integration. *Proceedings of the National Academy of Sciences*, 104 (33), 13507-13512.

Feng, G., Wang, Y., Huang, W., Chen, H., Dai, Z., Ma, G. et al. (2022). Methodological evaluation of individual cognitive prediction based on the brain white matter structural connectome. *Human Brain Mapping*, 43 (12), 3775-3791.

Feng, C., Wang, L., Li, T., & Xu, P. (2019). Connectome-based individualized prediction of loneliness. *Social Cognitive and Affective Neuroscience*, 14 (4), 353-365.

Feng, C., Yuan, J., Geng, H., Gu, R., Zhou, H., Wu, X., & Luo, Y. (2018). Individualized prediction of trait narcissism from whole-brain resting-state functional connectivity. *Human Brain Mapping*, 39 (9), 3701-3712.

Finn, ES. (2021a). Is it time to put rest to rest? *Trends in Cognitive Sciences*, 25 (12), 1021-1032.

Finn, ES., & Bandettini, P. (2021b). Movie-watching outperforms rest for functional connectivity-based prediction of behavior. *Neuroimage*, 235, 117963.

Finn, ES., Corlett, P., Chen, G., Bandettini, P., & Constable, R. (2018). Trait paranoia shapes inter-subject synchrony in brain activity during an ambiguous social narrative. *Nature Communications*, 9 (1).

Finn, ES., Glerean, E., Hasson, U., & Vanderwal, T. (2022). Naturalistic imaging: The use of ecologically valid conditions to study brain function. *Neuroimage*, 247, 118776.

Finn, ES., Glerean, E., Khojandi, A., Nielson, D., Molfese, P., Handwerker, D., & Bandettini, P. (2020). Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging. *NeuroImage*, 215, 116828.

Finn, ES., & Rosenberg, M. (2021). Beyond fingerprinting: Choosing predictive connectomes over reliable connectomes. *Neuroimage*, 239, 118254.

Finn, ES., Scheinost, D., Finn, D., Shen, X., Papademetris, X., & Constable, R. (2017). Can brain state be manipulated to emphasize individual differences in functional connectivity? *Neuroimage*, 160, 140-151.

Finn, ES., Shen, X., Holahan, J., Scheinost, D., Lacadie, C., & Papademetris, X. et al. (2014). Disruption of Functional Networks in Dyslexia: A Whole-Brain, Data-Driven Analysis of Connectivity. *Biological Psychiatry*, 76 (5), 397-404.

- Finn, ES., Shen, X., Scheinost, D., Rosenberg, M., Huang, J., & Chun, M. et al. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, 18 (11), 1664-1671.
- Finotelli, P., Dipasquale, O., Costantini, I., Pini, A., Baglio, F., Baselli, G., Dulio, P., & Cercignani, M. (2018). Exploring resting-state functional connectivity invariants across the lifespan in healthy people by means of a recently proposed graph theoretical model. *PloS one*, 13 (11).
- Fiuzat, E., Rhodes, S., & Murray, E. (2017). The Role of Orbitofrontal–Amygdala Interactions in Updating Action–Outcome Valuations in Macaques. *The Journal Of Neuroscience*, 37 (9), 2463-2470.
- Fogaça, M., & Duman, R. (2019). Cortical GABAergic Dysfunction in Stress and Depression: New Insights for Therapeutic Interventions. *Frontiers In Cellular Neuroscience*, 13.
- Forbes, M., Tackett, J., Markon, K., & Krueger, R. (2016). Beyond comorbidity: Toward a dimensional and hierarchical approach to understanding psychopathology across the life span. *Development and Psychopathology*, 28 (4pt1), 971-986.
- Fornito, A., & Bullmore, E. (2015). Reconciling abnormalities of brain network structure and function in schizophrenia. *Current Opinion in Neurobiology*, 30, 44-50.
- Fostering reproducible fMRI research. (2017). *Nature Neuroscience*, 20 (298).
- Fox, M., Corbetta, M., Snyder, A., Vincent, J., & Raichle, M. (2006). Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proceedings of The National Academy of Sciences*, 103 (26), 10046-10051.
- Fox, M., & Greicius, M. (2010). Clinical applications of resting state functional connectivity. *Frontiers in Systems Neuroscience*, 4.
- Fox, M., & Raichle, M. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience*, 8 (9), 700-711.
- Fox, M., Snyder, A., Vincent, J., Corbetta, M., Van Essen, D., & Raichle, M. (2005). From The Cover: The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of The National Academy of Sciences*, 102 (27), 9673-9678.
- Fox, M., Zhang, D., Snyder, A., & Raichle, M. (2009). The Global Signal and Observed Anticorrelated Resting State Brain Networks. *Journal of Neurophysiology*, 101 (6), 3270-3283.
- Freedman, R., Lewis, D., Michels, R., Pine, D., Schultz, S., Tamminga, C. et al. (2013). The initial field trials of DSM-5: New blooms and old thorns. *American Journal of Psychiatry*, 170 (1), 1–5.
- Frew, S., Samara, A., Shearer, H., Eilbott, J., & Vanderwal, T. (2022). Getting the nod: Pediatric head motion in a transdiagnostic sample during movie- and resting-state fMRI. *PLOS ONE*, 17 (4), e0265112.

- Gao, W., Chen, S., Biswal, B., Lei, X., & Yuan, J. (2018). Temporal dynamics of spontaneous default-mode network activity mediate the association between reappraisal and depression. *Social Cognitive And Affective Neuroscience*, 13 (12), 1235–1247.
- Gao, S., Greene, A., Constable, R., & Scheinost, D. (2019). Combining multiple connectomes improves predictive modeling of phenotypic measures. *NeuroImage*, 201, 116038.
- Gao, R., Peterson, E., & Voytek, B. (2017). Inferring synaptic excitation/inhibition balance from field potentials. *NeuroImage*, 158, 70–78.
- Gao, M., Wong, N., Lin, C., Huang, C., Liu, H., Toh, C. et al. (2023). Multimodal Brain Connectome-based prediction of suicide risk in people with late-life depression. *Nature Mental Health*, 1 (2), 100-113.
- Garnefski, N., Boon, S., & Kraaij, V. (2003). Relationships Between Cognitive Strategies of Adolescents and Depressive Symptomatology Across Different Types of Life Event. *Journal Of Youth And Adolescence*, 32 (6), 401-408.
- Garnefski, N., & Kraaij, V. (2007). The Cognitive Emotion Regulation Questionnaire. *European Journal Of Psychological Assessment*, 23 (3), 141-149.
- Garnefski, N., Kraaij, V., & Spinhoven, P. (2001). Negative life events, cognitive emotion regulation and emotional problems. *Personality And Individual Differences*, 30 (8), 1311-1327.
- Garnefski, N., Teerds, J., Kraaij, V., Legerstee, J., & van den Kommer, T. (2004). Cognitive emotion regulation strategies and depressive symptoms: differences between males and females. *Personality And Individual Differences*, 36 (2), 267-276.
- Garrity, A., Pearlson, G., McKiernan, K., Lloyd, D., Kiehl, K., & Calhoun, V. (2007). Aberrant “default mode” functional connectivity in schizophrenia. *American Journal of Psychiatry*, 164 (3), 450–457.
- Geerligs, L., Renken, R., Saliasi, E., Maurits, N., & Lorist, M. (2015a). A brain-wide study of age-related changes in functional connectivity. *Cerebral Cortex*, 25 (7), 1987–1999.
- Geerligs, L., Rubinov, M., Cam-CAN, & Henson, R. (2015b). State and trait components of functional connectivity: Individual differences vary with mental state. *The Journal of Neuroscience*, 35 (41), 13949-13961.
- Gell, M., Eickhoff, S., Omidvarnia, A., Küppers, V., Patil, K., Satterthwaite, T. et al. The burden of reliability: How measurement noise limits brain-behaviour predictions. *bioRxiv*. 2023.02.09.527898.
- Genon, S., Eickhoff, S., & Kharabian, S. (2022). Linking interindividual variability in brain structure to behaviour. *Nature Reviews Neuroscience*, 23 (5), 307-318.
- Glasser, M., Coalson, T., Robinson, E., Hacker, C., Harwell, J., & Yacoub, E. et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, 536 (7615), 171-178.

- Goldin, P., McRae, K., Ramel, W., & Gross, J. (2008). The Neural Bases of Emotion Regulation: Reappraisal and Suppression of Negative Emotion. *Biological Psychiatry*, 63 (6), 577-586.
- Goltermann, J., Winter, N., Gruber, M., Fisch, L., Richter, M., Grotegerd, D., et al. Cross-validation for the estimation of effect size generalizability in mass-Univariate Brain-wide association studies. bioRxiv. 2023.03.29.534696.
- Goodman, S., Fanelli, D., & Ioannidis, J. (2016). What does research reproducibility mean? *Science Translational Medicine*, 8 (341), 341ps12.
- Goodwin, L., & Leech, N. (2006). Understanding correlation: Factors that affect the size of R. *The Journal of Experimental Education*, 74 (3), 249–266.
- Gordon, E., Laumann, T., Gilmore, A., Newbold, D., Greene, D., Berg, J. et al. (2017). Precision functional mapping of individual human brains. *Neuron*, 95 (4), 791-807.
- Gorges, M., Roselli, F., Müller, H., Ludolph, A., Rasche, V., & Kassubek, J. (2017). Functional Connectivity Mapping in the Animal Model: Principles and Applications of Resting-State fMRI. *Frontiers in Neurology*, 8.
- Gorgolewski, K., Auer, T., Calhoun, V., Craddock, R., Das, S., & Duff, E. et al. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, 3 (1).
- Gorgolewski, K., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S., & Maumet, C. et al. (2015). NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Frontiers In Neuroinformatics*, 9.
- Gotts, S., Simmons, W., Milbury, L., Wallace, G., Cox, R., & Martin, A. (2012). Fractionation of social brain circuits in autism spectrum disorders. *Brain*, 135 (9), 2711–2725.
- Graff, K., Tansey, R., Ip, A., Rohr, C., Dimond, D., Dewey, D., & Bray, S. (2022). Benchmarking common preprocessing strategies in early childhood functional connectivity and intersubject correlation fMRI. *Developmental Cognitive Neuroscience*, 54, 101087.
- Gratton, C., Kraus, B., Greene, D., Gordon, E., Laumann, T., Nelson, S. et al. (2020). Defining individual-specific functional neuroanatomy for precision psychiatry. *Biological Psychiatry*, 88 (1), 28-39.
- Gratton, C., Laumann, T., Gordon, E., Adeyemo, B., & Petersen, S. (2016). Evidence for two independent factors that modify brain networks to meet task goals. *Cell Reports*, 17 (5), 1276-1288.
- Gratton, C., Nelson, S., & Gordon, E. (2022). Brain-behavior correlations: Two paths toward reliability. *Neuron*, 110 (9), 1446-1449.
- Grayson, D., & Fair, D. (2017). Development of large-scale functional networks from birth to adulthood: A guide to the neuroimaging literature. *Neuroimage*, 160, 15-31.

- Greene, A., Gao, S., Scheinost, D., & Constable, R. (2018a). Task-induced brain state manipulation improves prediction of individual traits. *Nature Communications*, 9 (1).
- Greene, A., Horien, C., Barson, D., Scheinost, D., & Constable, R. (2023). Why is everyone talking about brain state? *Trends in Neurosciences*.
- Greene, D., Koller, J., Hampton, J., Wesevich, V., Van, A., Nguyen, A. et al. (2018b). Behavioral interventions for reducing head motion during MRI scans in children. *NeuroImage*, 171, 234-245.
- Greicius, M., Flores, B., Menon, V., Glover, G., Solvason, H., Kenna, H. et al. (2007). Resting-state functional connectivity in major depression: Abnormally increased contributions from subgenual cingulate cortex and thalamus. *Biological Psychiatry*, 62 (5), 429–437.
- Greicius, M., Srivastava, G., Reiss, A., & Menon, V. (2004). Default-mode network activity distinguishes Alzheimer's disease from healthy aging: Evidence from functional MRI. *Proceedings of the National Academy of Sciences*, 101 (13), 4637–4642.
- Greve, D., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, 48 (1), 63-72.
- Grieder, M., Wang, D., Dierks, T., Wahlund, L., & Jann, K. (2018). Default mode network complexity and cognitive decline in mild Alzheimer's disease. *Frontiers in Neuroscience*, 12.
- Griffa, A., Amico, E., Liégeois, R., Van De Ville, D., & Preti, M. G. (2022). Brain structure-function coupling provides signatures for task decoding and individual fingerprinting. *NeuroImage*, 250, 118970.
- Gross, J. (1998). The Emerging Field of Emotion Regulation: An Integrative Review. *Review Of General Psychology*, 2 (3), 271-299.
- Gruskin, D., Rosenberg, M., & Holmes, A. (2020). Relationships between depressive symptoms and brain responses during emotional movie viewing emerge in adolescence. *Neuroimage*, 216, 116217.
- Gu, Z., Jamison, K., Sabuncu, M., & Kuceyeski, A. (2021). Heritability and interindividual variability of regional structure-function coupling. *Nature Communications*, 12 (1).
- Guo, C., Kurth, F., Zhou, J., Mayer, E., Eickhoff, S., Kramer, J., & Seeley, W. (2012). One-year test-retest reliability of intrinsic connectivity network fMRI in older adults. *Neuroimage*, 61 (4), 1471-1483.
- Guo, C., Nguyen, V., Hyett, M., Parker, G., & Breakspear, M. (2015). Out-of-sync: disrupted neural activity in emotional circuitry during film viewing in melancholic depression. *Scientific Reports*, 5 (1).
- Harrison, B., Pujol, J., Ortiz, H., Fornito, A., Pantelis, C., & Yücel, M. (2008). Modulation of brain resting-state networks by sad mood induction. *PLoS ONE*, 3 (3), E1794.

- Hasson, U., Avidan, G., Gelbard, H., Vallines, I., Harel, M., Minshew, N., & Behrmann, M. (2009). Shared and idiosyncratic cortical activation patterns in autism revealed under continuous real-life viewing conditions. *Autism Research*, 2 (4), 220-231
- Hasson, U., Malach, R., & Heeger, D. (2010). Reliability of cortical activity during natural stimulation. *Trends In Cognitive Sciences*, 14 (1), 40-48.
- Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject Synchronization of Cortical Activity During Natural Vision. *Science*, 303 (5664), 1634-1640.
- Hausman, H., Hardcastle, C., Kraft, J., Evangelista, N., Boutzoukas, E., O'Shea, A. et al. (2022). The association between head motion during functional magnetic resonance imaging and executive functioning in older adults. *Neuroimage: Reports*, 2 (2), 100085.
- Haxby, J., Hoffman, E., & Gobbini, M. (2000). The distributed human neural system for face perception. *Trends In Cognitive Sciences*, 4 (6), 223-233.
- He, B. (2011). Scale-free properties of the functional magnetic resonance imaging signal during rest and task. *Journal of Neuroscience*, 31 (39), 13786-13795.
- He, T., An, L., Chen, P., Chen, J., Feng, J., Bzdok, D. et al. (2022). Meta-matching as a simple framework to translate phenotypic predictive models from big to small data. *Nature Neuroscience*, 25 (6), 795-804.
- He, Y., Byrge, L., & Kennedy, D. (2020). Nonreplication of functional connectivity differences in autism spectrum disorder across multiple sites and denoising strategies. *Human Brain Mapping*.
- He, C., Fan, D., Liu, X., Wang, Q., Zhang, H., & Zhang, H. et al. (2022). Insula network connectivity mediates the association between childhood maltreatment and depressive symptoms in major depressive disorder patients. *Translational Psychiatry*, 12 (1).
- He, C., Gong, L., Yin, Y., Yuan, Y., Zhang, H., & Lv, L. et al. (2019). Amygdala connectivity mediates the association between anxiety and depression in patients with major depressive disorder. *Brain Imaging And Behavior*, 13 (4), 1146-1159.
- Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence Based Nursing*, 18 (3), 66-67.
- Heilicher, M., Crombie, K., & Cisler, J. (2022). Test-retest reliability of fmri during an emotion processing task: Investigating the impact of analytical approaches on ICC values. *Frontiers in Neuroimaging*, 1, 859792.
- Honey, C., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J., Meuli, R., & Hagmann, P. (2009). Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences*, 106 (6), 2035-2040.
- Hong, S., Vogelstein, J., Gozzi, A., Bernhardt, B., Yeo, BT., Milham, M., & Di Martino, A. (2020). Towards Neurosubtypes in Autism. *Biological Psychiatry*.

- Horien, C., Greene, A., Shen, X., Fortes, D., Brennan-Wydra, E., Banarjee, C. et al. (2023). A generalizable connectome-based marker of in-scan sustained attention in Neurodiverse Youth. *Cerebral Cortex*, Bhac506.
- Horien, C., Shen, X., Scheinost, D., & Constable, R. (2019). The individual functional connectome is unique and stable over months to years. *Neuroimage*, 189, 676-687.
- Hsu, W., Rosenberg, M., Scheinost, D., Constable, R., & Chun, M. (2018). Resting-state functional connectivity predicts neuroticism and extraversion in novel individuals. *Social Cognitive and Affective Neuroscience*, 13 (2), 224-232.
- Hu, D., Wang, F., Zhang, H., Wu, Z., Zhou, Z., Li, G. et al. (2022). Existence of functional connectome fingerprint during infancy and its stability over months. *The Journal of Neuroscience*, 42 (3), 377-389.
- Huang, C., Rolls, E., Feng, J., & Lin, C. (2021). An extended human connectome project multimodal Parcellation Atlas of the human cortex and subcortical areas. *Brain Structure and Function*, 227 (3), 763-778.
- Husain, M., & Nachev, P. (2007). Space and the parietal cortex. *Trends In Cognitive Sciences*, 11 (1), 30-36.
- Hutchison, R. M., & Everling, S. (2012). Monkey in the middle: why non-human primates are needed to bridge the gap in resting-state investigations. *Frontiers in neuroanatomy*, 6, 29.
- Insel, T. (2014). The NIMH research domain criteria (rdoc) Project: Precision Medicine for Psychiatry. *American Journal of Psychiatry*, 171 (4), 395–397.
- Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D., Quinn, K. et al. (2010). Research domain criteria (rdoc): Toward a new classification framework for research on Mental Disorders. *American Journal of Psychiatry*, 167 (7), 748-751.
- Ioannidis, J. (2005). Why Most Published Research Findings Are False. *Plos Medicine*, 2 (8), e124.
- Jääskeläinen, I., Koskentalo, K., Balk, M., Autti, T., Kauramäki, J., Pomren, C., & Sams, M. (2008). Inter-Subject Synchronization of Prefrontal Cortex Hemodynamic Activity During Natural Viewing. *The Open Neuroimaging Journal*, 2 (1), 14-19.
- Jääskeläinen, I., Sams, M., Glerean, E., & Ahveninen, J. (2021). Movies and narratives as naturalistic stimuli in neuroimaging. *NeuroImage*, 224, 117445.
- Jalbrzikowski, M., Liu, F., Foran, W., Klei, L., Calabro, F., & Roeder, K. et al. (2020). Functional connectome fingerprinting accuracy in youths and adults is similar when examined on the same day and 1.5-years apart. *Human Brain Mapping*, 41 (15), 4187-4199.
- Jankowski, K., Batres, J., Scott, H., Smyda, G., Pfeifer, J., & Quevedo, K. (2018). Feeling left out: depressed adolescents may atypically recruit emotional salience and regulation networks during social exclusion. *Social Cognitive And Affective Neuroscience*, 13 (8), 863-876.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *Neuroimage*, 17 (2), 825-841.

Jiang, R., Calhoun, V., Zuo, N., Lin, D., Li, J., Fan, L. et al. (2018). Connectome-based individualized prediction of temperament trait scores. *NeuroImage*, 183, 366-374.

Jiang, R., Zuo, N., Ford, J., Qi, S., Zhi, D., Zhuo, C. et al. (2020). Task-induced brain connectivity promotes the detection of individual differences in brain-behavior relationships. *NeuroImage*, 207, 116370.

Jing, B., Long, Z., Liu, H., Yan, H., Dong, J., & Mo, X. et al. (2017). Identifying current and remitted major depressive disorder with the Hurst exponent: a comparative study on two automated anatomical labeling atlases. *Oncotarget*, 8 (52), 90452-90464.

Johnstone, T., van Reekum, C., Urry, H., Kalin, N., & Davidson, R. (2007). Failure to Regulate: Counterproductive Recruitment of Top-Down Prefrontal-Subcortical Circuitry in Major Depression. *Journal Of Neuroscience*, 27 (33), 8877-8884.

Jones, D., Machulda, M., Vemuri, P., McDade, E., Zeng, G., Senjem, M., ... Jack, C. (2011). Age-related changes in the default mode network are more advanced in Alzheimer disease. *Neurology*, 77 (16), 1524–1531.

Joormann, J., & Gotlib, I. (2010). Emotion regulation in depression: Relation to cognitive inhibition. *Cognition & Emotion*, 24 (2), 281-298.

Ju, S., Horien, C., Shen, X., Abuwarda, H., Trainer, A., Constable, R., & Fredericks, C. (2023). Connectome-based predictive modeling shows sex differences in brain-based predictors of memory performance. *Frontiers in Dementia*, 2, 1126016.

Kaiser, R., Andrews-Hanna, J., Wager, T., & Pizzagalli, D. (2015). Large-scale network dysfunction in major depressive disorder. *JAMA Psychiatry*, 72 (6), 603–611.

Karnath, H. (1997). Spatial orientation and the representation of space with parietal lobe lesions. *Philosophical Transactions Of The Royal Society Of London. Series B: Biological Sciences*, 352 (1360), 1411-1419.

Kaufmann, T., Alnaes, D., Doan, N., Brandt, C., Andreassen, O., & Westlye, L. (2017). Delayed stabilization and individualization in connectome development are related to psychiatric disorders. *Nature Neuroscience*, 20 (4), 513-515.

Kaya, S., & McCabe, C. (2019). What Role Does the Prefrontal Cortex Play in the Processing of Negative and Positive Stimuli in Adolescent Depression?. *Brain Sciences*, 9 (5), 104.

Keller, M., Mendoza-Quñones, R., Cabrera Muñoz, A., Iglesias-Fuster, J., Virués, A., & Zvyagintsev, M. et al. (2022). Transdiagnostic alterations in neural emotion regulation circuits – neural substrates of cognitive reappraisal in patients with depression and post-traumatic stress disorder. *BMC Psychiatry*, 22 (1).

- Kelly, C., Biswal, B., Craddock, R., Castellanos, F., & Milham, M. (2012). Characterizing variation in the functional connectome: promise and pitfalls. *Trends in Cognitive Sciences*, 16 (3), 181-188.
- Kennedy, D., & Courchesne, E. (2008). The intrinsic functional organization of the brain is altered in autism. *NeuroImage*, 39 (4), 1877-1885.
- Kessler, R., McGonagle, K., Swartz, M., Blazer, D., & Nelson, C. (1993). Sex and depression in the National Comorbidity Survey I: Lifetime prevalence, chronicity and recurrence. *Journal Of Affective Disorders*, 29 (2-3), 85-96.
- King, J., Prigge, M., King, C., Morgan, J., Weathersby, F., & Fox, J. et al. (2019). Generalizability and reproducibility of functional connectivity in autism. *Molecular Autism*, 10 (1).
- King, G., Truzzi, A., & Cusack, R. (2023). The confound of head position in within-session connectome fingerprinting in infants. *NeuroImage*, 265, 119808.
- Klein, D., Shankman, S., Lewinsohn, P., & Seeley, J. (2009). Subthreshold Depressive Disorder in Adolescents: Predictors of Escalation to Full-Syndrome Depressive Disorders. *Journal Of The American Academy Of Child & Adolescent Psychiatry*, 48 (7), 703-710.
- Klingbeil, J., Wawrzyniak, M., Stockert, A., & Saur, D. (2019). Resting-state functional connectivity: An emerging method for the study of language networks in post-stroke aphasia. *Brain and Cognition*, 131, 22–33.
- Kohno, S., Noriuchi, M., Iguchi, Y., Kikuchi, Y., & Hoshi, Y. (2015). Emotional discrimination during viewing unpleasant pictures: timing in human anterior ventrolateral prefrontal cortex and amygdala. *Frontiers In Human Neuroscience*, 9 (51).
- Kong, R., Yang, Q., Gordon, E., Xue, A., Yan, X., Orban, C. et al. (2021). Individual-specific areal-level parcellations improve functional connectivity prediction of behavior. *Cerebral Cortex*, 31 (10), 4477-4500.
- Kong, X., Zhen, Z., Li, X., Lu, H., Wang, R., Liu, L. et al. (2014). Individual differences in impulsivity predict head motion during magnetic resonance imaging. *PLoS ONE*, 9 (8).
- Koo, T., & Li, M. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal Of Chiropractic Medicine*, 15 (2), 155-163.
- Kriegeskorte, N., Simmons, W., Bellgowan, P., & Baker, C. (2009). Circular Analysis in Systems Neuroscience: The dangers of double dipping. *Nature Neuroscience*, 12 (5), 535-540.
- Kucyi, A., Esterman, M., Riley, C., & Valera, E. (2016). Spontaneous default network activity reflects behavioral variability independent of mind-wandering. *Proceedings of the National Academy of Sciences*, 113 (48), 13899-13904.
- Lake, E., Finn, ES., Noble, S., Vanderwal, T., Shen, X., Rosenberg, M. et al. (2019). The Functional Brain Organization of an individual allows prediction of measures of social abilities

transdiagnostically in autism and attention-deficit/hyperactivity disorder. *Biological Psychiatry*, 86 (4), 315-326.

Langs, G., Wang, D., Golland, P., Mueller, S., Pan, R., & Sabuncu, M. et al. (2015). Identifying Shared Brain Networks in Individuals by Decoupling Functional and Anatomical Variability. *Cerebral Cortex*, 26 (10), 4004-4014.

Larsen, J., Vermulst, A., Geenen, R., van Middendorp, H., English, T., & Gross, J. et al. (2012). Emotion Regulation in Adolescence. *The Journal Of Early Adolescence*, 33 (2), 184-200.

Laumann, T., Gordon, E., Adeyemo, B., Snyder, A., Joo, S., Chen, M. et al. (2015). Functional system and Areal Organization of a highly sampled individual human brain. *Neuron*, 87 (3), 657-670.

Lei, X., Zhao, Z., & Chen, H. (2013). Extraversion is encoded by scale-free dynamics of Default Mode Network. *NeuroImage*, 74, 52-57.

Li, X., Ai, L., Giavasis, S., Jin, H., Feczko, E., & Xu, T. et al. Moving Beyond Processing and Analysis-Related Variation in Neuroscience. bioRxiv 2021.12.01.470790.

Li, J., Bolt, T., Bzdok, D., Nomi, J., Yeo, BT., Spreng, R., & Uddin, L. (2019). Topography and behavioral relevance of the global signal in the human brain. *Scientific Reports*, 9 (1).

Li, Q., Ferraro, G., Pasquini, L., Peck, K., Makse, H., & Holodny, A. (2020). Core language brain network for fMRI language task used in clinical applications. *Network Neuroscience*, 4 (1), 134-154.

Li, F., He, N., Li, Y., Chen, L., Huang, X., & Lui, S. et al. (2014). Intrinsic Brain Abnormalities in Attention Deficit Hyperactivity Disorder: A Resting-State Functional MR Imaging Study. *Radiology*, 272 (2), 514-523.

Li, S., Hu, N., Zhang, W., Tao, B., Dai, J., Gong, Y., Tan, Y. et al. (2019). Dysconnectivity of multiple brain networks in schizophrenia: A meta-analysis of resting-state functional connectivity. *Frontiers in Psychiatry*, 10.

Li, K., Wisner, K., & Atluri, G. (2021). Feature selection framework for functional connectome fingerprinting. *Human Brain Mapping*, 42 (12), 3717–3732.

Light, S., Heller, A., Johnstone, T., Kolden, G., Peterson, M., Kalin, N., & Davidson, R. (2011). Reduced Right Ventrolateral Prefrontal Cortex Activity While Inhibiting Positive Affect Is Associated with Improvement in Hedonic Capacity After 8 Weeks of Antidepressant Treatment in Major Depressive Disorder. *Biological Psychiatry*, 70 (10), 962-968.

Liljequist, D., Elfving, B., & Skavberg Roaldsen, K. (2019). Intraclass correlation – A discussion and demonstration of basic features. *PLOS ONE*, 14 (7), e0219854.

Lin, Y., Baete, S., Wang, X., & Boada, F. (2020). Mapping brain–behavior networks using functional and structural connectome fingerprinting in the HCP Dataset. *Brain and Behavior*, 10 (6).

- Litwińczuk, M., Muhlert, N., Cloutman, L., Trujillo-Barreto, N., & Woollams, A. (2022). Combination of structural and functional connectivity explains unique variation in specific domains of cognitive function. *NeuroImage*, 262, 119531.
- Liu, S., Abdellaoui, A., Verweij, K., & van Wingen, G. (2023). Replicable Brain–phenotype associations require large-scale neuroimaging data. *Nature Human Behaviour*.
- Liu, J., Xu, P., Zhang, J., Jiang, N., Li, X., & Luo, Y. (2019). Ventral attention-network effective connectivity predicts individual differences in adolescent depression. *Journal Of Affective Disorders*, 252, 55-59.
- Long, Y., Cao, H., Yan, C., Chen, X., Li, L., & Castellanos, F. et al. (2020). Altered resting-state dynamic functional brain networks in major depressive disorder: Findings from the REST-meta-MDD consortium. *Neuroimage: Clinical*, 26, 102163.
- Long, Z., Jing, B., Guo, R., Li, B., Cui, F., Wang, T., & Chen, H. (2018). A Brainnetome Atlas Based Mild Cognitive Impairment Identification Using Hurst Exponent. *Frontiers In Aging Neuroscience*, 10.
- Lu, H., Zou, Q., Gu, H., Raichle, M., Stein, E., & Yang, Y. (2012). Rat brains also have a default mode network. *Proceedings of The National Academy of Sciences*, 109 (10), 3979-3984.
- Lynall, M., Bassett, D., Kerwin, R., McKenna, P., Kitzbichler, M., Muller, U., & Bullmore, E. (2010). Functional Connectivity and Brain Networks in Schizophrenia. *Journal of Neuroscience*, 30 (28), 9477–9487.
- Lyall, D., Quinn, T., Lyall, L., Ward, J., Anderson, J., Smith, D. et al. (2022). Quantifying bias in psychological and physical health in the UK Biobank Imaging sub-sample. *Brain Communications*, 4 (3).
- Lyons, K., Stevenson, R., Owen, A., & Stojanoski, B. (2020). Examining the relationship between measures of autistic traits and neural synchrony during movies in children with and without autism. *Neuroimage: Clinical*, 28, 102477.
- Madan, C. (2021). Scan Once, Analyse Many: Using Large Open-Access Neuroimaging Datasets to Understand the Brain. *Neuroinformatics*.
- Madhyastha, T., Peverill, M., Koh, N., McCabe, C., Flournoy, J., Mills, K. et al. (2018). Current methods and limitations for longitudinal fmri analysis across development. *Developmental Cognitive Neuroscience*, 33, 118-128.
- Maguire, E. (2012). Studying the freely-behaving brain with fmri. *NeuroImage*, 62 (2), 1170–1176.
- Mansour, S., Tian, Y., Yeo, B.T., Cropley, V., & Zalesky, A. (2021). High-resolution connectomic fingerprints: Mapping neural identity and behaviour. *NeuroImage*, 229, 117695.
- Mantwill, M., Gell, M., Krohn, S., & Finke, C. (2022). Brain connectivity fingerprinting and behavioural prediction rest on distinct functional systems of the human connectome. *Communications Biology*, 5 (1).

- Marcotte, K., Perlberg, V., Marrelec, G., Benali, H., & Ansaldo, A. (2013). Default-mode network functional connectivity in aphasia: Therapy-induced neuroplasticity. *Brain and Language*, 124 (1), 45–55.
- Marek, S., Tervo-Clemmens, B., Calabro, F., Montez, D., Kay, B., & Hatoum, A. et al. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature*, 603 (7902), 654-660.
- Marek, S., Tervo-Clemmens, B., Nielsen, A., Wheelock, M., Miller, R., & Laumann, T. et al. (2019). Identifying reproducible individual differences in childhood functional brain networks: An ABCD study. *Developmental Cognitive Neuroscience*, 40, 100706.
- Markiewicz, C., Gorgolewski, K., Feingold, F., Blair, R., Halchenko, Y., Miller, E. et al. (2021). The OpenNeuro resource for sharing of Neuroscience Data. *ELife*, 10.
- Margulies, D., Ghosh, S., Goulas, A., Falkiewicz, M., Huntenburg, J., & Langs, G. et al. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of The National Academy of Sciences*, 113 (44), 12574-12579.
- Margulies, D., Vincent, J., Kelly, C., Lohmann, G., Uddin, L., & Biswal, B. et al. (2009). Precuneus shares intrinsic functional architecture in humans and monkeys. *Proceedings of The National Academy of Sciences*, 106 (47), 20069-20074.
- Marquand, A., Wolfers, T., Mennes, M., Buitelaar, J., & Beckmann, C. (2016). Beyond lumping and splitting: A review of Computational approaches for stratifying psychiatric disorders. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 1 (5), 433-447.
- Maxim, V., Şendur, L., Fadili, J., Suckling, J., Gould, R., Howard, R., & Bullmore, E. (2005). Fractional gaussian noise, functional MRI and alzheimer's disease. *NeuroImage*, 25 (1), 141-158.
- McFadzean, R., Brosnahan, D., Hadley, D., & Mutlukan, E. (1994). Representation of the visual field in the occipital striate cortex. *British Journal Of Ophthalmology*, 78 (3), 185-190.
- McRae, K., Hughes, B., Chopra, S., Gabrieli, J., Gross, J., & Ochsner, K. (2010). The Neural Bases of Distraction and Reappraisal. *Journal Of Cognitive Neuroscience*, 22 (2), 248-262.
- Meer, J., Breakspear, M., Chang, L., Sonkusare, S., & Cocchi, L. (2020). Movie viewing elicits rich and reliable brain state dynamics. *Nature Communications*, 11 (1).
- Mellinger, C., & Hanson, T. (2021). Methodological Considerations for survey research: Validity, reliability, and Quantitative Analysis. *Linguistica Antverpiensia, New Series – Themes in Translation Studies*, 19, 172–190.
- Mennes, M., Biswal, B., Castellanos, F., & Milham, M. (2013). Making data sharing work: The FCP/INDI experience. *Neuroimage*, 82, 683-691.
- Menon, V. (2013). Developmental pathways to functional brain networks: Emerging principles. *Trends in Cognitive Sciences*, 17 (12), 627–640.

- Meskaldji, D., Preti, M., Bolton, T., Montandon, M., Rodriguez, C., Morgenthaler, A. et al. (2016). Predicting individual scores from resting state fmri using partial least squares regression. *IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 1311-1314.
- Mevel, K., Chételat, G., Eustache, F., & Desgranges, B. (2011). The default mode network in healthy aging and Alzheimer's disease. *International Journal of Alzheimer's Disease*, 2011, 1–9.
- Mihalik, A., Ferreira, F., Rosa, M., Moutoussis, M., Ziegler, G., & Monteiro, J. et al. (2019). Brain-behaviour modes of covariation in healthy and clinically depressed young people. *Scientific Reports*, 9 (1).
- Mikl, M., Mareček, R., Hlušík, P., Pavlicová, M., Drastich, A., Chlebus, P. et al. (2008). Effects of spatial smoothing on fmri group inferences. *Magnetic Resonance Imaging*, 26 (4), 490-503.
- Milham, M., Craddock, R., Son, J., Fleischmann, M., Clucas, J., & Xu, H. et al. (2018). Assessment of the impact of shared brain imaging data on the scientific literature. *Nature Communications*, 9 (1).
- Milham, M., Damien, F., Maarten, M., & Stewart, M. (2012). The ADHD-200 Consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. *Frontiers in Systems Neuroscience*, 6.
- Milham, M., Vogelstein, J., & Xu, T. (2021). Removing the reliability bottleneck in functional magnetic resonance imaging research to achieve clinical utility. *JAMA Psychiatry*, 78 (6), 587.
- Miller, K., Alfaro-Almagro, F., Bangerter, N., Thomas, D., Yacoub, E., & Xu, J. et al. (2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature Neuroscience*, 19 (11), 1523-1536.
- Monk, C., Peltier, S., Wiggins, J., Weng, S., Carrasco, M., Risi, S., & Lord, C. (2009). Abnormalities of intrinsic functional connectivity in autism spectrum disorders. *Neuroimage*, 47 (2), 764-772.
- Mueller, S., Wang, D., Fox, M., Pan, R., Lu, J., & Li, K. et al. (2015). Reliability correction for functional connectivity: Theory and implementation. *Human Brain Mapping*, 36 (11), 4664-4680.
- Mueller, S., Wang, D., Fox, M., Yeo, BT., Sepulcre, J., & Sabuncu, M. et al. (2013). Individual Variability in Functional Connectivity Architecture of the Human Brain. *Neuron*, 77 (3), 586-595.
- Mulders, P., van Eijndhoven, P., Schene, A., Beckmann, C., & Tendolkar, I. (2015). Resting-state functional connectivity in major depressive disorder: A Review. *Neuroscience Biobehavioral Reviews*, 56, 330–344.
- Mumford, J. (2012). A power calculation guide for fmri studies. *Social Cognitive and Affective Neuroscience*, 7 (6), 738–742.

- Munafò, M., Nosek, B., Bishop, D., Button, K., Chambers, C., Percie du Sert, N. et al. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1 (1).
- Murphy, K., Birn, R., Handwerker, D., Jones, T., & Bandettini, P. (2009). The impact of global signal regression on resting state correlations: Are anti-correlated networks introduced? *Neuroimage*, 44 (3), 893-905.
- Murphy, C., Wang, H., Konu, D., Lowndes, R., Margulies, D., Jefferies, E., & Smallwood, J. (2019). Modes of operation: A topographic neural gradient supporting stimulus dependent and independent cognition. *Neuroimage*, 186, 487-496.
- Mwangi, B., Tian, T., & Soares, J. (2014). A review of feature reduction techniques in neuroimaging. *Neuroinformatics*, 12 (2), 229-244.
- Nastase, S., Gazzola, V., Hasson, U., & Keysers, C. (2019). Measuring shared responses across subjects using intersubject correlation. *Social Cognitive And Affective Neuroscience*, 669–687.
- Nastase, S., Goldstein, A., & Hasson, U. (2020). Keep it real: Rethinking the primacy of experimental control in Cognitive Neuroscience. *NeuroImage*, 222, 117254.
- Nastase, S., Liu, Y., Hillman, H., Zadbood, A., Hasenfratz, L., Keshavarzian, N. et al. (2021). The “narratives” fmri dataset for evaluating models of naturalistic language comprehension. *Scientific Data*, 8 (250).
- Neufang, S., & Akhrif, A. (2020). Regional Hurst Exponent Reflects Impulsivity-Related Alterations in Fronto-Hippocampal Pathways Within the Waiting Impulsivity Network. *Frontiers In Physiology*, 11, 827.
- Nguyen, M., Vanderwal, T., & Hasson, U. (2019). Shared understanding of narratives is correlated with shared neural responses. *NeuroImage*, 184, 161–170.
- Nielsen, J., Zielinski, B., Fletcher, P., Alexander, A., Lange, N., & Bigler, E. et al. (2013). Multisite functional connectivity MRI classification of autism: ABIDE results. *Frontiers in Human Neuroscience*, 7.
- Nikolaidis, A., Chen, A., He, X., Shinohara, R., Vogelstein, J., Milham, M., & Shou, H. Suboptimal phenotypic reliability impedes reproducible human neuroscience. *bioRxiv* 2022.07.22.501193.
- Nikolaidis, A., Solon Heinsfeld, A., Xu, T., Bellec, P., Vogelstein, J., & Milham, M. (2020). Bagging improves reproducibility of functional parcellation of the human brain. *Neuroimage*, 214, 116678.
- Noble, S., Scheinost, D., & Constable, R. (2019). A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *Neuroimage*, 203, 116157.
- Noble, S., Spann, M., Tokoglu, F., Shen, X., Constable, R., & Scheinost, D. (2017). Influences on the Test-Retest reliability of Functional Connectivity MRI and its Relationship with Behavioral Utility. *Cerebral Cortex*, 27 (11), 5415-5429.

Nomi, J., Bzdok, D., Li, J., Bolt, T., Kornfeld, S., Goodman, Z. et al. Global signal topography in the human brain differs systematically across the lifespan. *bioRxiv* 2022.07.27.501804.

Nomi, J., & Uddin, L. (2015). Developmental changes in large-scale network connectivity in autism. *NeuroImage: Clinical*, 7, 732–741.

Nosek, B., Hardwicke, T., Moshontz, H., Allard, A., Corker, K., Dreber, A. et al. (2021). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, 73, 719–748.

Nummenmaa, L., Glerean, E., Viinikainen, M., Jääskeläinen, I., Hari, R., & Sams, M. (2012). Emotions promote social interaction by synchronizing brain activity across individuals. *Proceedings Of The National Academy Of Sciences*, 109 (24), 9599-9604.

Nummenmaa, L., Lahnakoski, J., & Glerean, E. (2018). Sharing the social world via intersubject neural synchronisation. *Current Opinion In Psychology*, 24, 7-14.

O'Connor, D., Lake, E., Scheinost, D., & Constable, R. (2021). Resample aggregating improves the generalizability of Connectome Predictive Modeling. *NeuroImage*, 236, 118044.

O'Connor, D., Potler, N., Kovacs, M., Xu, T., Ai, L., & Pellman, J. et al. (2017). The Healthy Brain Network Serial Scanning Initiative: a resource for evaluating inter-individual differences and their reliabilities across scan conditions and sessions. *Gigascience*, 6 (2).

Ohbayashi, M., Picard, N., & Strick, P. (2016). Inactivation of the Dorsal Premotor Area Disrupts Internally Generated, But Not Visually Guided, Sequential Movements. *The Journal Of Neuroscience*, 36 (6), 1971-1976.

Omidvarnia, A., Liégeois, R., Amico, E., Preti, M., Zalesky, A., & Van De Ville, D. (2022). On the spatial distribution of temporal complexity in resting state and task functional MRI. *Entropy*, 24 (8), 1148.

Omidvarnia, A., Sasse, L., Larabi, D., Raimondo, F., Hoffstaedter, F., Kasper, J. et al. Is resting state fmri better than individual characteristics at predicting cognition? *bioRxiv*. 2023.02.18.529076.

Opitz, D., & Maclin, R. (1999). Popular Ensemble Methods: An Empirical Study. *Journal Of Artificial Intelligence Research*, 11, 169-198.

Pan, W., Billings, J., Grooms, J., Shakil, S., & Keilholz, S. (2015). Considerations for resting state functional MRI and functional connectivity studies in rodents. *Frontiers in Neuroscience*, 9.

Pandya, M., Altinay, M., Malone, D., & Anand, A. (2012). Where in the Brain Is Depression?. *Current Psychiatry Reports*, 14 (6), 634-642.

Park, C., Lazar, N., Ahn, J., & Sornborger, A. (2010). A multiscale analysis of the temporal characteristics of resting-state fMRI data. *Journal Of Neuroscience Methods*, 193 (2), 334-342.

Park, D., & Reuter-Lorenz, P. (2009). The Adaptive Brain: Aging and neurocognitive scaffolding. *Annual Review of Psychology*, 60 (1), 173–196.

- Parkes, L., Fulcher, B., Yücel, M., & Fornito, A. (2018). An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. *Neuroimage*, *171*, 415-436.
- Patriat, R., Molloy, E., Meier, T., Kirk, G., Nair, V., & Meyerand, M. et al. (2013). The effect of resting condition on resting-state fMRI reliability and consistency: A comparison between resting with eyes open, closed, and fixated. *Neuroimage*, *78*, 463-473.
- Petersen, S., & Sporns, O. (2015). Brain Networks and Cognitive Architectures. *Neuron*, *88* (1), 207-219.
- Phạm, D., McDonald, D., Ding, L., Nebel, M., & Mejia, A. (2023). Less is more: Balancing noise reduction and data retention in fmri with data-driven scrubbing. *NeuroImage*, *270*, 119972.
- Pizzagalli, D., & Roberts, A. (2021). Prefrontal cortex and depression. *Neuropsychopharmacology*, *47* (1), 225-246.
- Poldrack, R., Baker, C., Durnez, J., Gorgolewski, K., Matthews, P., & Munafo, M. et al. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, *18* (2), 115-126.
- Poldrack, R., & Gorgolewski, K. (2017). OpenfMRI: Open sharing of task fMRI data. *Neuroimage*, *144*, 259-261.
- Poldrack, R., & Gorgolewski, K. (2014). Making big data open: data sharing in neuroimaging. *Nature Neuroscience*, *17* (11), 1510-1517.
- Poldrack, R., Laumann, T., Koyejo, O., Gregory, B., Hover, A., Chen, M. et al. (2015). Long-term neural and physiological phenotyping of a single human. *Nature Communications*, *6* (1).
- Poldrack, R., & Yarkoni, T. (2016). From Brain Maps to Cognitive Ontologies: Informatics and the Search for Mental Structure. *Annual Review Of Psychology*, *67* (1), 587-612.
- Poline, J., Breeze, J., Ghosh, S., Gorgolewski, K., Halchenko, Y., Hanke, M., et al. (2012). Data sharing in neuroimaging research. *Frontiers in neuroinformatics*, *6*, 9.
- Pomponio, R., Erus, G., Habes, M., Doshi, J., & Srinivasan, D. (2020). Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. *NeuroImage*, *208*, 116450.
- Pourtois, G., Degelder, B., Bol, A., & Crommelinck, M. (2005). Perception of Facial Expressions and Voices and of their Combination in the Human Brain. *Cortex*, *41* (1), 49-59.
- Power, J., Barnes, K., Snyder, A., Schlaggar, B., & Petersen, S. (2013). Steps toward optimizing motion artifact removal in functional connectivity MRI; a reply to Carp. *Neuroimage*, *76*, 439-441.
- Power, J., Barnes, K., Snyder, A., Schlaggar, B., & Petersen, S. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage*, *59* (3), 2142-2154.

- Power, J., Mitra, A., Laumann, T., Snyder, A., Schlaggar, B., & Petersen, S. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage*, 84, 320-341.
- Power, J., Plitt, M., Kundu, P., Bandettini, P., & Martin, A. (2017). Temporal interpolation alters motion in fMRI scans: Magnitudes and consequences for artifact detection. *PLoS ONE*, 12 (9), e0182939.
- Power, J., Schlaggar, B., & Petersen, S. (2015). Recent progress and outstanding issues in motion correction in resting state fMRI. *Neuroimage*, 105, 536-551.
- Preti, M., Bolton, T., & Van De Ville, D. (2017). The dynamic functional connectome: State-of-the-art and perspectives. *Neuroimage*, 160, 41-54.
- Pruim, R., Mennes, M., Van Rooij, D., Llera, A., Buitelaar, J., & Beckmann, C. (2015). ICA-Aroma: A robust ICA-based strategy for removing motion artifacts from fmri data. *NeuroImage*, 112, 267-277.
- Qi, S., Morris, R., Turner, J., Fu, Z., Jiang, R., Deramus, T. et al. (2020). Common and unique multimodal covarying patterns in autism spectrum disorder subtypes. *Molecular Autism*, 11 (1).
- Raichle, M., MacLeod, A., Snyder, A., Powers, W., Gusnard, D., & Shulman, G. (2001). A default mode of brain function. *Proceedings of The National Academy of Sciences*, 98 (2), 676-682.
- Rajmohan, V., & Mohandas, E. (2007). The limbic system. *Indian journal of psychiatry*, 49 (2), 132-139.
- Ramduny, J., Bastiani, M., Huedepohl, R., Sotiropoulos, S., & Chechlac, M. (2022). The association between inadequate sleep and accelerated brain ageing. *Neurobiology of Aging*, 114, 1–14.
- Ramduny, J., Garcia, M., & Kelly, C. (2023). Establishing a reproducible and sustainable analysis workflow. OSF Preprints.
- Rao, A., Monteiro, J., & Mourao-Miranda, J. (2017). Predictive modelling using neuroimaging data in the presence of confounds. *NeuroImage*, 150, 23-49.
- Raz, G., Touroutoglou, A., Wilson-Mendenhall, C., Gilam, G., Lin, T., & Gonen, T. et al. (2016). Functional connectivity dynamics during film viewing reveal common networks for different emotional experiences. *Cognitive, Affective, & Behavioral Neuroscience*, 16 (4), 709-723.
- Regitz-Zagrosek, V., & Gebhard, C. (2022). Gender medicine: Effects of sex and gender on cardiovascular disease manifestation and outcomes. *Nature Reviews Cardiology*, 20 (4), 236-247.
- Revisiting doubt in neuroimaging research. (2022). *Nature Neuroscience*, 25 (7), 833-834.

- Ricard, J., Parker, T., Dhamala, E., Kwas, J., Allsop, A., & Holmes, A. (2023). Confronting racially exclusionary practices in the acquisition and analyses of neuroimaging data. *Nature Neuroscience*, 26 (1), 4–11.
- Richiardi, J., Eryilmaz, H., Schwartz, S., Vuilleumier, P., & Van De Ville, D. (2011). Decoding brain states from fmri connectivity graphs. *NeuroImage*, 56 (2), 616-626.
- Rohr, C., Vinette, S., Parsons, K., Cho, I., Dimond, D., & Benischek, A. et al. (2016). Functional Connectivity of the Dorsal Attention Network Predicts Selective Attention in 4-7 year-old Girls. *Cerebral Cortex*.
- Rolls, E., Huang, C., Lin, C., Feng, J., & Joliot, M. (2020). Automated Anatomical Labelling Atlas 3. *NeuroImage*, 206, 116189.
- Rosen, M., Sheridan, M., Sambrook, K., Dennison, M., Jenness, J., & Askren, M. et al. (2017). Salience network response to changes in emotional expressions of others is heightened during early adolescence: relevance for social functioning. *Developmental Science*, 21 (3), e12571.
- Rosenberg, M., & Finn, ES. (2022). How to establish robust brain–behavior relationships without thousands of individuals. *Nature Neuroscience*, 25 (7), 835-837.
- Rosenberg, M., Finn, ES., Scheinost, D., Papademetris, X., Shen, X., Constable, R., & Chun, M. (2015). A neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience*, 19 (1), 165-171.
- Rosenberg, M., Hsu, W., Scheinost, D., Constable, R., & Chun, M. (2018). Connectome-based models predict separable components of attention in novel individuals. *Journal of Cognitive Neuroscience*, 30 (2), 160-173.
- Rosenberg, M., Scheinost, D., Greene, A., Avery, E., Kwon, Y., Finn, ES. et al. (2020). Functional connectivity predicts changes in attention observed across minutes, days, and months. *Proceedings of the National Academy of Sciences*, 117 (7), 3797-3807.
- Rosenberg, M., Zhang, S., Hsu, W., Scheinost, D., Finn, ES., Shen, X. et al. (2016). Methylphenidate modulates functional network connectivity to enhance attention. *Journal of Neuroscience*, 36 (37), 9547-9557.
- Rosenblatt, M., Rodriguez, R., Westwater, M., Dai, W., Horien, C., Greene, A. et al. (2023). Connectome-based machine learning models are vulnerable to subtle data manipulations. *Patterns*, 4, 100756.
- Rubenstein, J., & Merzenich, M. (2003). Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes, Brain And Behavior*, 2 (5), 255-267.
- Saad, Z., Gotts, S., Murphy, K., Chen, G., Jo, H., Martin, A., & Cox, R. (2012). Trouble at Rest: How Correlation Patterns and Group Differences Become Distorted After Global Signal Regression. *Brain Connectivity*, 2 (1), 25-32.
- Saarimäki, H. (2021). Naturalistic stimuli in affective neuroimaging: A Review. *Frontiers in Human Neuroscience*, 15.

- Saggar, M., & Uddin, L. (2019). Pushing the boundaries of psychiatric neuroimaging to ground diagnosis in biology. *Eneuro*, 6 (6), ENEURO.0384-19.2019.
- Salehi, M., Greene, A., Karbasi, A., Shen, X., Scheinost, D., & Constable, R. (2020). There is no single functional atlas even for a single individual: Functional parcel definitions change with task. *Neuroimage*, 208, 116366.
- Salmi, J., Roine, U., Glerean, E., Lahnakoski, J., Nieminen-von Wendt, T., & Tani, P. et al. (2013). The brains of high functioning autistic individuals do not synchronize with those of others. *Neuroimage: Clinical*, 3, 489-497.
- Saluja, G., Iachan, R., Scheidt, P., Overpeck, M., Sun, W., & Giedd, J. (2004). Prevalence of and Risk Factors for Depressive Symptoms Among Young Adolescents. *Archives Of Pediatrics & Adolescent Medicine*, 158 (8), 760.
- Sanders, A., Harms, M., Kandala, S., Marek, S., Somerville, L., Bookheimer, S. et al. (2023). Age-related differences in resting-state functional connectivity from childhood to adolescence. *Cerebral Cortex*, 33 (11), 6928–6942.
- Sarwar, T., Tian, Y., Yeo, BT., Ramamohanarao, K., & Zalesky, A. (2020). Structure-function coupling in the human connectome: A machine learning approach. *Neuroimage*, 226, 117609.
- Satterthwaite, T., Ciric, R., Roalf, D., Davatzikos, C., Bassett, D., & Wolf, D. (2017). Motion artifact in studies of functional connectivity: Characteristics and mitigation strategies. *Human Brain Mapping*, 40 (7), 2033-2051.
- Satterthwaite, T., Elliott, M., Gerraty, R., Ruparel, K., Loughead, J., & Calkins, M. et al. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage*, 64, 240-256.
- Satterthwaite, T., Wolf, D., Loughead, J., Ruparel, K., Elliott, M., & Hakonarson, H. et al. (2012). Impact of in-scanner head motion on multiple measures of functional connectivity: Relevance for studies of neurodevelopment in youth. *Neuroimage*, 60 (1), 623-632.
- Satterthwaite, T., Wolf, D., Roalf, D., Ruparel, K., Erus, G., & Vandekar, S. et al. (2015). Linked Sex Differences in Cognition and Functional Connectivity in Youth. *Cerebral Cortex*, 25 (9), 2383-2394.
- Schaefer, A., Kong, R., Gordon, E., Laumann, T., Zuo, X., & Holmes, A. et al. (2018). Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex*, 28 (9), 3095-3114.
- Schäfer, J., Naumann, E., Holmes, E., Tuschen-Caffier, B., & Samson, A. (2016). Emotion Regulation Strategies in Depressive and Anxiety Symptoms in Youth: A Meta-Analytic Review. *Journal Of Youth And Adolescence*, 46 (2), 261-276.

- Scheinost, D., Noble, S., Horien, C., Greene, A., Lake, E., Salehi, M. et al. (2019). Ten simple rules for predictive modeling of individual differences in neuroimaging. *NeuroImage*, 193, 35-45.
- Schmaal, L., Pozzi, E., Ho, T., van Velzen, L., Veer, I., Opel, N. et al. (2020). Enigma MDD: Seven Years of global neuroimaging studies of major depression through Worldwide Data Sharing. *Translational Psychiatry*, 10 (172).
- Schönbrodt, F., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, 47 (5), 609-612.
- Schurz, M., Wimmer, H., Richlan, F., Ludersdorfer, P., Klackl, J., & Kronbichler, M. (2015). Resting-state and task-based functional brain connectivity in developmental dyslexia. *Cerebral Cortex*, 25 (10), 3502–3514.
- Seghier, M. (2012). The Angular Gyrus. *The Neuroscientist*, 19 (1), 43-61.
- Segrin, C., & Rynes, K. (2009). The mediating role of positive relations with others in associations between depressive symptoms, social skills, and perceived stress. *Journal Of Research In Personality*, 43 (6), 962-971.
- Selten, M., van Bokhoven, H., & Nadif Kasri, N. (2018). Inhibitory control of the excitatory/inhibitory balance in psychiatric disorders. *F1000research*, 7, 23.
- Shah, L., Cramer, J., Ferguson, M., Birn, R., & Anderson, J. (2016). and reproducibility of individual differences in functional connectivity acquired during task and resting state. *Brain And Behavior*, 6 (5).
- Shanmugan, S., Seidlitz, J., Cui, Z., Adebimpe, A., Bassett, D., Bertolero, M. et al. (2022). Sex differences in the functional topography of association networks in Youth. *Proceedings of the National Academy of Sciences*, 119 (33).
- Shansky, R. (2018). Sex differences in behavioral strategies: Avoiding Interpretational Pitfalls. *Current Opinion in Neurobiology*, 49, 95–98.
- Shansky, R., & Murphy, A. (2021). Considering sex as a biological variable will require a global shift in science culture. *Nature Neuroscience*, 24 (4), 457–464.
- Shansky, R., & Woolley, C. (2016). Considering sex as a biological variable will be valuable for Neuroscience Research. *The Journal of Neuroscience*, 36 (47), 11817–11822.
- Shaw, P., Kabani, N., Lerch, J., Eckstrand, K., Lenroot, R., Gogtay, N. et al. (2008). Neurodevelopmental trajectories of the human cerebral cortex. *The Journal of Neuroscience*, 28 (14), 3586-3594.
- Shehzad, Z., Kelly, C., Reiss, P., Gee, D., Gotimer, K., & Uddin, L. et al. (2009). The Resting Brain: Unconstrained yet Reliable. *Cerebral Cortex*, 19 (10), 2209-2229.
- Shen, X., Finn, ES., Scheinost, D., Rosenberg, M., Chun, M., Papademetris, X., & Constable, T. (2017). Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nature Protocols*, 12 (3), 506-518.

- Shen, X., Tokoglu, F., Papademetris, X., & Constable, R. (2013). Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *Neuroimage*, 82, 403-415.
- Shinn, M., Hu, A., Turner, L., Noble, S., Preller, K., Ji, J. et al. (2023). Functional Brain Networks reflect spatial and temporal autocorrelation. *Nature Neuroscience*, 26 (5), 867–878.
- Shirer, W., Jiang, H., Price, C., Ng, B., & Greicius, M. (2015). Optimization of rs-fMRI Preprocessing for Enhanced Signal Noise Separation, Test-Retest Reliability, and Group Discrimination. *Neuroimage*, 117, 67-79.
- Short, S., Elison, J., Goldman, B., Styner, M., Gu, H., Connelly, M. et al. (2013). Associations between white matter microstructure and infants' working memory. *NeuroImage*, 64, 156–166.
- Shrout, P., & Fleiss, J. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86 (2), 420-428.
- Siegel, J., Mitra, A., Laumann, T., Seitzman, B., Raichle, M., Corbetta, M., & Snyder, A. (2016). Data quality influences observed links between functional connectivity and behavior. *Cerebral Cortex*, 27 (9), 4492-4502.
- Siegle, G., Thompson, W., Carter, C., Steinhauer, S., & Thase, M. (2007). Increased Amygdala and Decreased Dorsolateral Prefrontal BOLD Responses in Unipolar Depression: Related and Independent Features. *Biological Psychiatry*, 61 (2), 198-209.
- Silk, J., Steinberg, L., & Morris, A. (2003). Adolescents' Emotion Regulation in Daily Life: Links to Depressive Symptoms and Problem Behavior. *Child Development*, 74 (6), 1869-1880.
- Simmons, J., Nelson, L., & Simonsohn, U. (2011). False-Positive Psychology. Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22 (11), 1359-1366.
- Simony, E., Honey, C., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., & Hasson, U. (2016). Dynamic reconfiguration of the default mode network during narrative comprehension. *Nature Communications*, 7 (1).
- Sisk, L., & Gee, D. (2022). Stress and adolescence: vulnerability and opportunity during a sensitive window of development. *Current Opinion In Psychology*, 44, 286-292.
- Smith, S., Fox, P., Miller, K., Glahn, D., Fox, P., & Mackay, C. et al. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of The National Academy of Sciences*, 106 (31), 13040-13045.
- Smith, S., Jenkinson, M., Woolrich, M., Beckmann, C., Behrens, T., & Johansen-Berg, H. et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, 23, S208-S219.
- Smith, S., Vidaurre, D., Beckmann, C., Glasser, M., Jenkinson, M., & Miller, K. et al. (2013). Functional connectomics from resting-state fMRI. *Trends in Cognitive Sciences*, 17 (12), 666-682.

- Smith, J., Wilkey, E., Clarke, B., Shanley, L., Men, V., Fair, D., & Sabb, F. (2022). Can this data be saved? techniques for high motion in resting state scans of first grade children. *Developmental Cognitive Neuroscience*, 58, 101178.
- Sohal, V., & Rubenstein, J. (2019). Excitation-inhibition balance as a framework for investigating mechanisms in neuropsychiatric disorders. *Molecular Psychiatry*, 24 (9), 1248-1257.
- Solopchuk, O., Alamia, A., & Zénon, A. (2016). The Role of the Dorsal Premotor Cortex in Skilled Action Sequences. *The Journal Of Neuroscience*, 36 (25), 6599-6601.
- Somandepalli, K., Kelly, C., Reiss, P., Zuo, X., Craddock, R., & Yan, C. et al. (2015). Short-term test-retest reliability of resting state fMRI metrics in children with and without attention-deficit/hyperactivity disorder. *Developmental Cognitive Neuroscience*, 15, 83-93.
- Song, J., Birn, R., Boly, M., Meier, T., Nair, V., Meyerand, M., & Prabhakaran, V. (2014). Age-related reorganizational changes in modularity and functional connectivity of Human Brain Networks. *Brain Connectivity*, 4 (9), 662–676.
- Song, J., Desphande, A., Meier, T., Tudorascu, D., Vergun, S., & Nair, V. et al. (2012). Age-Related Differences in Test-Retest Reliability in Resting-State Brain Functional Connectivity. *PloS ONE*, 7 (12), e49847.
- Song, H., Park, B., Park, H., & Shim, W. (2021). Cognitive and Neural State Dynamics of narrative comprehension. *The Journal of Neuroscience*, 41 (43), 8972–8990.
- Sonkusare, S., Breakspear, M., & Guo, C. (2019). Naturalistic Stimuli in Neuroscience: Critically Acclaimed. *Trends In Cognitive Sciences*, 23 (8), 699-714.
- Spiers, H., & Maguire, E. (2007). Decoding human brain activity during real-world experiences. *Trends in Cognitive Sciences*, 11 (8), 356–365.
- Spisak, T., Bingel, U., & Wager, T. (2023). Multivariate BWAS can be replicable with moderate sample sizes. *Nature*, 615 (7951).
- Sporns O. (2013). Structure and function of complex brain networks. *Dialogues in clinical neuroscience*, 15 (3), 247–262.
- Steiger, J. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87 (2), 245–251.
- Strimbu, K., & Tavel, J. (2010). What do biomarkers mark? *Current Opinion in HIV and AIDS*, 5 (6), 463–466.
- Suárez, L., Markello, R., Betzel, R., & Misic, B. (2020). Linking structure and function in Macroscale Brain Networks. *Trends in Cognitive Sciences*, 24 (4), 302-315.
- Sun, L., Liang, X., Duan, D., Liu, J., Chen, Y., Wang, X. et al. (2022). Structural insight into the individual variability architecture of the functional brain connectome. *NeuroImage*, 259, 119387.

Supekar, K., Musen, M., & Menon, V. (2009). Development of Large-Scale Functional Brain Networks in Children. *Neuroimage*, 47, S109.

Supekar, K., Uddin, L., Khouzam, A., Phillips, J., Gaillard, W., & Kenworthy, L. et al. (2013). Brain Hyperconnectivity in Children with Autism and its Links to Social Deficits. *Cell Reports*, 5 (3), 738-747.

Sydnor, V., Larsen, B., Bassett, D., Alexander-Bloch, A., Fair, D., Liston, C. et al. (2021). Neurodevelopment of the association cortices: Patterns, mechanisms, and implications for psychopathology. *Neuron*, 109 (18), 2820-2846.

Sydnor, V., Larsen, B., Seidlitz, J., Adebimpe, A., Alexander-Bloch, A., Bassett, D. et al. (2023). Intrinsic activity development unfolds along a sensorimotor–association cortical axis in Youth. *Nature Neuroscience*, 26 (4), 638-649.

Sylvester, C., Barch, D., Corbetta, M., Power, J., Schlaggar, B., & Luby, J. (2013). Resting State Functional Connectivity of the Ventral Attention Network in Children With a History of Depression or Anxiety. *Journal Of The American Academy Of Child & Adolescent Psychiatry*, 52 (12), 1326-1336.

Szucs, D., & Ioannidis, J. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLOS Biology*, 15 (3), e2000797.

Szucs, D., & Ioannidis, J. (2020). Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990–2012) and of latest practices (2017–2018) in high-impact journals. *NeuroImage*, 221, 117164.

Tagliazucchi, E., & Laufs, H. (2014). Decoding Wakefulness Levels from Typical fMRI Resting-State Data Reveals Reliable Drifts between Wakefulness and Sleep. *Neuron*, 82 (3), 695-708.

Tamnes, C., Roalf, D., Goddings, A., & Lebel, C. (2018). Diffusion MRI of white matter microstructure development in childhood and adolescence: Methods, challenges and progress. *Developmental Cognitive Neuroscience*, 33, 161–175.

Tejavibulya, L., Rolison, M., Gao, S., Liang, Q., Peterson, H., Dadashkarimi, J. et al. (2022). Predicting the future of neuroimaging predictive models in Mental Health. *Molecular Psychiatry*, 27 (8), 3129-3137.

Telzer, E., McCormick, E., Peters, S., Cosme, D., Pfeifer, J., & van Duijvenvoorde, A. (2018). Methodological considerations for developmental longitudinal fmri research. *Developmental Cognitive Neuroscience*, 33, 149–160.

Tervo-Clemmens, B., Marek, S., Chauvin, R., Van, A., Kay, B., Laumann, T. et al. (2023). Reply to: Multivariate BWAS can be replicable with moderate sample sizes. *Nature*, 615 (7951).

Thirion, B., Pinel, P., Mériaux, S., Roche, A., Dehaene, S., & Poline, J. (2007). Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *Neuroimage*, 35 (1), 105-120.

Thomason, M., Brown, J., Dassanayake, M., Shastri, R., Marusak, H., Hernandez-Andrade, E. et al. (2014). Intrinsic functional brain architecture derived from graph theoretical analysis in the human fetus. *PLoS ONE*, 9 (5), e94423.

Thomason, M., Burrows, B., Gabrieli, J., & Glover, G. (2005). Breath holding reveals differences in fmri bold signal in children and adults. *NeuroImage*, 25 (3), 824–837.

Thomason, M., Dennis, E., Joshi, A., Joshi, S., Dinov, I., & Chang, C. et al. (2011). Resting-state fMRI can reliably map neural networks in children. *Neuroimage*, 55 (1), 165-175.

Thomason, M., Grove, L., Lozon, T., Vila, A., Ye, Y., Nye, M., et al. (2015). Age-related increases in long-range connectivity in fetal functional neural connectivity networks in utero. *Developmental Cognitive Neuroscience*, 11, 96–104.

Thomason, M., Scheinost, D., Manning, J., Grove, L., Hect, J., Marshall, N., et al. (2017). Weak functional connectivity in the human fetal brain prior to preterm birth. *Scientific Reports*, 7 (1), 39286.

Thompson, R. (1994). Emotion Regulation: A Theme in Search of Definition. *Monographs Of The Society For Research In Child Development*, 59 (2/3), 25.

Thompson, P., Stein, J., Medland, S., Hibar, D., Vasquez, A., Renteria, M. et al. (2014). The ENIGMA Consortium: Large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging and Behavior*, 8 (2), 153–182.

Thompson, W., Wright, J., Bissett, P., & Poldrack, R. (2020). Dataset decay and the problem of sequential analyses on open datasets. *eLife*, 9, e53498.

Thornburgh, C., Narayana, S., Rezaie, R., Bydlinski, B., Tylavsky, F., & Papanicolaou, A. et al. (2017). Concordance of the Resting State Networks in Typically Developing, 6-to 7-Year-Old Children and Healthy Adults. *Frontiers in Human Neuroscience*, 11.

Tian, L., Jiang, T., Liang, M., Li, X., He, Y., & Wang, K. et al. (2007). Stabilities of negative correlations between blood oxygen level-dependent signals associated with sensory and motor cortices. *Human Brain Mapping*, 28 (7), 681-690.

Tiego, J., Martin, E., DeYoung, C., Hagan, K., Cooper, S., Pasion, R. et al. (2022). Precision behavioral phenotyping as a strategy for uncovering the biological correlates of psychopathology. OSF Preprints.

Tomasi, D., Shokri-Kojori, E., & Volkow, N. (2016). Temporal Evolution of Brain Functional Connectivity Metrics: Could 7 Min of Rest be Enough?, *Cerebral Cortex*, 27, 4153-4165.

Tomasi, D., & Volkow, N. (2012). Resting functional connectivity of language networks: characterization and reproducibility. *Molecular Psychiatry*, 17 (8), 841-854.

Tomasi, D., & Volkow, N. (2011). Aging and functional brain networks. *Molecular Psychiatry*, 17 (5), 549–558.

Torre, J., & Lieberman, M. (2018). Putting Feelings Into Words: Affect Labeling as Implicit Emotion Regulation. *Emotion Review*, 10 (2), 116-124.

- Tovée, M. (1995). Face Recognition: What are faces for?. *Current Biology*, 5 (5), 480-482.
- Tozzi, L., Fleming, S., Taylor, Z., Raterink, C., & Williams, L. (2020). Test-retest reliability of the human functional connectome over consecutive days: Identifying highly reliable portions and assessing the impact of methodological choices. *Network Neuroscience*, 4 (3), 925-945.
- Trakoshis, S., Martínez-Cañada, P., Rocchi, F., Canella, C., You, W., & Chakrabarti, B. et al. (2020). Intrinsic excitation-inhibition imbalance affects medial prefrontal cortex differently in autistic men versus women. *Elife*, 9.
- Trost, W., Frühholz, S., Cochrane, T., Cojan, Y., & Vuilleumier, P. (2015). Temporal dynamics of musical emotions examined through intersubject synchrony of brain activity. *Social Cognitive And Affective Neuroscience*, 10 (12), 1705-1721.
- Turken, A. U., & Dronkers, N. F. (2011). The neural architecture of the language comprehension network: converging evidence from lesion and connectivity analyses. *Frontiers in systems neuroscience*, 5, 1.
- Uddin, L., Supekar, K., & Menon, V. (2013). Reconceptualizing functional brain connectivity in autism from a developmental perspective. *Frontiers in Human Neuroscience*, 7 (458).
- Uddin, L., Supekar, K., Ryali, S., & Menon, V. (2011). Dynamic Reconfiguration of Structural and Functional Connectivity Across Core Neurocognitive Brain Networks with Development. *Journal Of Neuroscience*, 31 (50), 18578-18589.
- Urchs, S., Armoza, J., Benhajali, Y., St-Aubin, J., Orban, P., & Bellec, P. (2017). MIST: A multi-resolution parcellation of functional brain networks. *MNI Open Research*, 1, 3.
- Van Dijk, K., Hedden, T., Venkataraman, A., Evans, K., Lazar, S., & Buckner, R. (2010). Intrinsic Functional Connectivity As a Tool For Human Connectomics: Theory, Properties, and Optimization. *Journal Of Neurophysiology*, 103 (1), 297-321.
- Van Dijk, K., Sabuncu, M., & Buckner, R. (2012). The influence of head motion on intrinsic functional connectivity MRI. *NeuroImage*, 59 (1), 431-438.
- Varoquaux, G., Raamana, P., Engemann, D., Hoyos-Idrobo, A., Schwartz, Y., & Thirion, B. (2017). Assessing and tuning brain decoders: Cross-validation, caveats, and Guidelines. *NeuroImage*, 145, 166-179.
- Vincent, J., Patel, G., Fox, M., Snyder, A., Baker, J., & Van Essen, D. et al. (2007). Intrinsic functional architecture in the anaesthetized monkey brain. *Nature*, 447 (7140), 83-86.
- Vincent, J., Snyder, A., Fox, M., Shannon, B., Andrews, J., Raichle, M., & Buckner, R. (2006). Coherent Spontaneous Activity Identifies a Hippocampal-Parietal Memory Network. *Journal of Neurophysiology*, 96 (6), 3517-3531.
- Vahdat, S., Darainy, M., Milner, T., & Ostry, D. (2011). Functionally Specific Changes in Resting-State Sensorimotor Networks after Motor Learning. *Journal of Neuroscience*, 31 (47), 16907-16915.

- Van Dijk, K., Hedden, T., Venkataraman, A., Evans, K., Lazar, S., & Buckner, R. (2010). Intrinsic Functional Connectivity As a Tool For Human Connectomics: Theory, Properties, and Optimization. *Journal of Neurophysiology*, 103 (1), 297-321.
- Van Dijk, K., Sabuncu, M., & Buckner, R. (2012). The influence of head motion on intrinsic functional connectivity MRI. *Neuroimage*, 59 (1), 431-438.
- Van Essen, D., Smith, S., Barch, D., Behrens, T., Yacoub E., Ugurbil, K., for the WU-Minn HCP Consortium. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, 80, 62-79.
- Vanderwal, T., Eilbott, J., & Castellanos, F. (2019). Movies in the magnet: Naturalistic paradigms in developmental functional neuroimaging. *Developmental Cognitive Neuroscience*, 36, 100600.
- Vanderwal, T., Eilbott, J., Finn, ES., Craddock, R., Turnbull, A., & Castellanos, F. (2017). Individual differences in functional connectivity during naturalistic viewing conditions. *Neuroimage*, 157, 521-530.
- Vanderwal, T., Eilbott, J., Kelly, C., Frew, S., Woodward, T., Milham, M., & Castellanos, F. (2021). Stability and similarity of the pediatric connectome as developmental measures. *NeuroImage*, 226, 117537.
- Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L., & Castellanos, F. (2015). Inscapes: A movie paradigm to improve compliance in functional magnetic resonance imaging. *Neuroimage*, 122, 222-232.
- Varikuti, D., Hoffstaedter, F., Genon, S., Schwender, H., Reid, A., & Eickhoff, S. (2016). Resting-state test–retest reliability of a priori defined canonical networks over different preprocessing steps. *Brain Structure and Function*, 222 (3), 1447–1468.
- Varoquaux, G. (2018). Cross-validation failure: Small sample sizes lead to large error bars. *NeuroImage*, 180, 68-77.
- Varoquaux, G., Raamana, P., Engemann, D., Hoyos-Idrobo, A., Schwartz, Y., & Thirion, B. (2017). Assessing and tuning brain decoders: Cross-validation, caveats, and Guidelines. *NeuroImage*, 145, 166-179.
- Veer, I., Beckmann, C., van Tol, M., Ferrarini, L., Milles, J., & Veltman, D. et al. (2010). Whole brain resting-state analysis reveals decreased connectivity in major depressive disorder. *Frontiers in Systems Neuroscience*, 4.
- Venkatesh, M., Jaja, J., & Pessoa, L. (2020). Comparing functional connectivity matrices: A geometry-aware approach applied to participant identification. *NeuroImage*, 207, 116398.
- Vijayakumar, N., Whittle, S., Yücel, M., Dennison, M., Simmons, J., & Allen, N. (2014). Thinning of the lateral prefrontal cortex during adolescence predicts emotion regulation in females. *Social Cognitive And Affective Neuroscience*, 9 (11), 1845-1854.

- Visted, E., Vøllestad, J., Nielsen, M., & Schanche, E. (2018). Emotion Regulation in Current and Remitted Depression: A Systematic Review and Meta-Analysis. *Frontiers In Psychology*, 9.
- von dem Hagen, E., Stoyanova, R., Baron-Cohen, S., & Calder, A. (2012). Reduced functional connectivity within and between 'social' resting state networks in autism spectrum conditions. *Social Cognitive And Affective Neuroscience*, 8 (6), 694-701.
- Vossel, S., Geng, J. J., & Fink, G. R. (2014). Dorsal and ventral attention systems: distinct neural circuits but collaborative roles. *The Neuroscientist*, 20 (2), 150-159.
- Waller, L., Walter, H., Kruschwitz, J., Reuter, L., Müller, S., Erk, S., & Veer, I. (2017). Evaluating the replicability, specificity, and generalizability of connectome fingerprints. *Neuroimage*, 158, 371-377.
- Wang, Z., Goerlich, K., Ai, H., Aleman, A., Luo, Y., & Xu, P. (2021). Connectome-based predictive modeling of individual anxiety. *Cerebral Cortex*, 31 (6), 3006-3020.
- Wang, J., Han, J., Nguyen, V., Guo, L., & Guo, C. (2017a). Improving the Test-Retest reliability of Resting State fMRI by Removing the Impact of Sleep. *Frontiers in Neuroscience*, 11.
- Wang, J., Ren, Y., Hu, X., Nguyen, V., Guo, L., Han, J., & Guo, C. (2017b). Test-retest reliability of functional connectivity networks during naturalistic fMRI paradigms. *Human Brain Mapping*, 38 (4), 2226-2241.
- Wang, Q., Xu, Y., Zhao, T., Xu, Z., He, Y., & Liao, X. (2021). Individual uniqueness in the neonatal functional connectome. *Cerebral Cortex*, 31 (8), 3701–3712.
- Wang, Z., Bridgeford, E., Wang, S, Vogelstein, J., & Caffo, B. Statistical Analysis of Data Repeatability Measures. arXiv. 2005.11911.
- Wanger, T., Janes, A., & Frederick, B. (2022). Spatial variation of changes in test–retest reliability of functional connectivity after global signal regression: The effect of considering hemodynamic delay. *Human Brain Mapping*, 44 (2), 668–678.
- Wechsler D. (2003). *Wechsler Intelligence Scale for Children*, 4th Edn San Antonia, TX: PsychCorp.
- Wechsler D. (2008). *Wechsler Adult Intelligence Scale*, 4th Edn San Antonia, TX: Pearson.
- Wei, S., Chang, M., Zhang, R., Jiang, X., Wang, F., & Tang, Y. (2018). Amygdala functional connectivity in female patients with major depressive disorder with and without suicidal ideation. *Annals Of General Psychiatry*, 17 (1).
- Wei, L., Jing, B., & Li, H. (2020). Bootstrapping promotes the RSFC-Behavior Associations: An application of individual cognitive traits prediction. *Human Brain Mapping*, 41 (9), 2302-2316.
- Wei, M., Qin, J., Yan, R., Bi, K., Liu, C., Yao, Z., & Lu, Q. (2015). Association of resting-state network dysfunction with their dynamics of inter-network interactions in depression. *Journal Of Affective Disorders*, 174, 527-534.

- Wei, M., Qin, J., Yan, R., Li, H., Yao, Z., & Lu, Q. (2013). Identifying major depressive disorder using Hurst exponent of resting-state brain networks. *Psychiatry Research: Neuroimaging*, 214 (3), 306-312.
- Weiner, K., & Zilles, K. (2016). The anatomical and functional specialization of the fusiform gyrus. *Neuropsychologia*, 83, 48-62.
- Weissenbacher, A., Kasess, C., Gerstl, F., Lanzenberger, R., Moser, E., & Windischberger, C. (2009). Correlations and anticorrelations in resting-state functional connectivity MRI: A quantitative comparison of preprocessing strategies. *Neuroimage*, 47 (4), 1408-1416.
- Weng, S., Wiggins, J., Peltier, S., Carrasco, M., Risi, S., Lord, C., & Monk, C. (2010). Alterations of resting state functional connectivity in the default network in adolescents with autism spectrum disorders. *Brain Research*, 1313, 202-214.
- Whelan, R., & Garavan, H. (2014). When optimism hurts: Inflated predictions in psychiatric neuroimaging. *Biological Psychiatry*, 75 (9), 746-748.
- Whitfield-Gabrieli, S., Thermenos, H., Milanovic, S., Tsuang, M., Faraone, S., McCarley, R. et al. (2009). Hyperactivity and hyperconnectivity of the default network in schizophrenia and in first-degree relatives of persons with schizophrenia. *Proceedings of the National Academy of Sciences*, 106 (4), 1279–1284.
- Wicherts, J., Veldkamp, C., Augusteijn, H., Bakker, M., Van Aert, R., & Van Assen, M. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid P-hacking. *Frontiers in Psychology*, 7 (1832).
- Wilkinson, M., Dumontier, M., Aalbersberg, I. Appleton, G., Axton, M., Baak, A. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018.
- Willems, R., Nastase, S., & Milivojevic, B. (2020). Narratives for neuroscience. *Trends in Neurosciences*, 43 (5), 271–273.
- Willinger, D., Karipidis, I., Neuer, S., Emery, S., Rauch, C., & Häberling, I. et al. (2022). Maladaptive Avoidance Learning in the Orbitofrontal Cortex in Adolescents With Major Depression. *Biological Psychiatry: Cognitive Neuroscience And Neuroimaging*, 7 (3), 293-301.
- Wilms, I., Petersen, A., & Vangkilde, S. (2013). Intensive video gaming improves encoding speed to visual short-term memory in young male adults. *Acta Psychologica*, 142 (1), 108-118.
- Woo, C., Chang, L., Lindquist, M., & Wager, T. (2017). Building better biomarkers: Brain models in translational neuroimaging. *Nature Neuroscience*, 20 (3), 365-377.
- Wylie, G., Genova, H., DeLuca, J., Chiaravalloti, N., & Sumowski, J. (2014). Functional magnetic resonance imaging movers and shakers: Does subject-movement cause sampling bias? *Human Brain Mapping*, 35 (1), 1–13.

- Xia, C., Ma, Z., Ciric, R., Gu, S., Betzel, R., & Kaczkurkin, A. et al. (2018). Linked dimensions of psychopathology and connectivity in functional brain networks. *Nature Communications*, 9 (1).
- Xia, M., Si, T., Sun, X., Ma, Q., Liu, B., Wang, L. et al. (2019). Reproducibility of functional brain alterations in major depressive disorder: Evidence from a multisite resting-state functional MRI study with 1,434 individuals. *NeuroImage*, 189, 700–714.
- Xu, Y., Cao, M., Liao, X., Xia, M., Wang, X., Jeon, T. et al. (2019). Development and emergence of individual variability in the functional connectivity architecture of the preterm human brain. *Cerebral Cortex*, 29 (10), 4208-4222.
- Yamashita, A., Yahata, N., Itahashi, T., Lisi, G., Yamada, T., & Ichikawa, N. et al. (2019). Harmonization of resting-state functional MRI data across multiple imaging sites via the separation of site differences into sampling bias and measurement bias. *PLoS Biology*, 17 (4), e3000042.
- Yan, C., Chen, X., Li, L., & Zang, Y. (2019). Reduced default mode network functional connectivity in patients with recurrent major depressive disorder. *Proceedings of the National Academy of Sciences*, 116 (18), 9078–9083.
- Yan, C., Cheung, B., Kelly, C., Colcombe, S., Craddock, R., & Di Martino, A. et al. (2013a). A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. *Neuroimage*, 76, 183-201.
- Yan, C., Craddock, R. C., Zuo, X., Zang, Y., & Milham, M. (2013b). Standardizing the intrinsic brain: towards robust measurement of inter-individual variation in 1000 functional connectomes. *NeuroImage*, 80, 246-262.
- Yang, E., Milisav, F., Kopal, J., Holmes, A., Mitsis, G., Misic, B. et al. (2023). The default network dominates neural responses to evolving movie stories. *Nature Communications*, 14 (1).
- Yarkoni, T., Poldrack, R., Nichols, T., Van Essen, D., & Wager, T. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, 8 (8), 665-670.
- Yao, H., Guet-McCreight, A., Mazza, F., Moradi Chameh, H., Prevot, T., & Griffiths, J. et al. (2022). Reduced inhibition in depression impairs stimulus processing in human cortical microcircuits. *Cell Reports*, 38 (2), 110232.
- Yeung, A., More, S., Wu, J., & Eickhoff, S. (2022). Reporting details of neuroimaging studies on individual traits prediction: A literature survey. *NeuroImage*, 256, 119275.
- Yeo, BT., Krienen, F., Sepulcre, J., Sabuncu, M., Lashkari, D., & Hollinshead, M. et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106 (3), 1125-1165.
- Yip, S., Scheinost, D., Potenza, M., & Carroll, K. (2019). Connectome-based prediction of Cocaine Abstinence. *American Journal of Psychiatry*, 176 (2), 156-164.

- Yoo, C., Park, S., & Kim, M. (2022). Structural connectome-based prediction of trait anxiety. *Brain Imaging and Behavior*, 16 (6), 2467-2476.
- Yoo, K., Rosenberg, M., Hsu, W., Zhang, S., Li, C., Scheinost, D. et al. (2018). Connectome-based predictive modeling of attention: Comparing different functional connectivity features and prediction methods across datasets. *NeuroImage*, 167, 11-22.
- Young, K., Sandman, C., & Craske, M. (2019). Positive and Negative Emotion Regulation in Adolescence: Links to Anxiety and Depression. *Brain Sciences*, 9 (4), 76.
- Zanto, T., & Gazzaley, A. (2013). Fronto-parietal network: flexible hub of cognitive control. *Trends in Cognitive Sciences*, 17 (12), 602-603.
- Zeng, L., Wang, D., Fox, M., Sabuncu, M., Hu, D., Ge, M. et al. (2014). Neurobiological basis of head motion in Brain Imaging. *Proceedings of the National Academy of Sciences*, 111 (16), 6058-6062.
- Zhang, Z., Allen, G., Zhu, H., & Dunson, D. (2019). Tensor network factorizations: Relationships between brain structural connectomes and traits. *NeuroImage*, 197, 330-343.
- Zhang, S., Hu, S., Chao, H., & Li, C. (2015). Resting-State Functional Connectivity of the Locus Coeruleus in Humans: In Comparison with the Ventral Tegmental Area/Substantia Nigra Pars Compacta and the Effects of Age. *Cerebral Cortex*, 26 (8), 3413-3427.
- Zhang, J., Kucyi, A., Raya, J., Nielsen, A., Nomi, J., Damoiseaux, J. et al. (2021). What have we really learned from functional connectivity in clinical populations? *NeuroImage*, 242, 118466.
- Zhang, B., Lin, P., Shi, H., Öngür, D., Auerbach, R., & Wang, X. et al. (2016). Mapping anhedonia-specific dysfunction in a transdiagnostic approach: an ALE meta-analysis. *Brain Imaging And Behavior*, 10 (3), 920-939.
- Zhang, H., Zhao, Y., Cao, W., Cui, D., Jiao, Q., Lu, W. et al. (2020). Aberrant functional connectivity in resting state networks of ADHD patients revealed by independent component analysis. *BMC Neuroscience*, 21 (39).
- Zhou, Y., Liang, M., Tian, L., Wang, K., Hao, Y., Liu, H., et al. (2007). Functional disintegration in paranoid schizophrenia using resting-state fmri. *Schizophrenia Research*, 97 (1–3), 194–205.
- Zuo, X., Anderson, J., Bellec, P. et al. (2014a). An open science resource for establishing reliability and reproducibility in functional connectomics. *Sci Data* 1, 140049.
- Zuo, X., Biswal, B., & Poldrack, R. (2019). Editorial: and Reproducibility in Functional Connectomics. *Frontiers in Neuroscience*, 13.
- Zuo, X., He, Y., Betzel, R., Colcombe, S., Sporns, O., & Milham, M. (2017). Human Connectomics across the Life Span. *Trends in Cognitive Sciences*, 21 (1), 32-45.

Zuo, X., Kelly, C., Adelstein, J., Klein, D., Castellanos, F., & Milham, M. (2010a). Reliable intrinsic connectivity networks: Test-retest evaluation using ICA and dual regression approach. *Neuroimage*, 49 (3), 2163-2177.

Zuo, X., Kelly, C., Di Martino, A., Mennes, M., Margulies, D., & Bangaru, S. et al. (2010b). Growing Together and Growing Apart: Regional and Sex Differences in the Lifespan Developmental Trajectories of Functional Homotopy. *Journal of Neuroscience*, 30 (45), 15034-15043.

Zou, Q., Miao, X., Liu, D., Wang, D., Zhuo, Y., & Gao, J. (2015). Reliability comparison of spontaneous brain activities between BOLD and CBF contrasts in eyes-open and eyes-closed resting states. *Neuroimage*, 121, 91-105.

Zuo, X., & Xing, X. (2014b). Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: A systems neuroscience perspective. *Neuroscience & Biobehavioral Reviews*, 45, 100-118.

Appendices

Appendix A



Coláiste na Tríonóide, Baile Átha Cliath
Trinity College Dublin

Ollscoil Átha Cliath | The University of Dublin

F.A.O. Clare Kelly
SPREC092017-01

School of Psychology Research Ethics Committee

12th September 2017

Dear Clare,

The School of Psychology Research Ethics Committee has reviewed your application entitled "Leveraging Open Data for Neurodevelopmental Disorders" and I am pleased to inform you that it was approved.

Please note that you will be required to submit a completed **Project Annual Report Form** on each anniversary of this approval, until such time as an **End of Project Report Form** is submitted upon completion of the research. Copies of both forms are available for download from the Ethics section of the School website.

Adverse events associated with the conduct of this research must be reported immediately to the Chair of the Ethics Committee.

Yours sincerely,

A handwritten signature in black ink, appearing to read 'Richard Carson'.

Richard Carson
Chair,
School of Psychology Research Ethics Committee

Scoil na Síceolaíochta

Dámh na nEolaíochtaí Sóisialta agus Daonna,
Áras an Phiarsaigh,
Coláiste na Tríonóide,
Baile Átha Cliath,
Ollscoil Átha Cliath,
Baile Átha Cliath 2, Éire.

School of Psychology

Faculty of Arts, Humanities and Social Sciences,
Trinity College Dublin,
The University of Dublin,
Dublin 2, Ireland.

+353 1 896 1886
psychology@tcd.ie
www.tcd.ie/psychology

Appendix B

F.A.O. Clare Kelly
Approval ID: SPREC0102015-01

School of Psychology Research Ethics Committee

28th October 2015

Dear Clare,

The School of Psychology Research Ethics Committee has reviewed your application entitled "Depressive Symptoms and Coping in Adolescence" and I am pleased to inform you that it was approved.

Please note that you will be required to submit a completed **Project Annual Report Form** on each anniversary of this approval, until such time as an **End of Project Report Form is submitted** upon completion of the research. Copies of both forms are available for download from the Ethics section of the School website.

Please note that you must be familiar with and adhere to the attached 'Safety Protocol for Adults'.

Adverse events associated with the conduct of this research must be reported immediately to the Chair of the Ethics Committee.

Yours sincerely,



Richard Carson
Chair,
School of Psychology Research Ethics Committee

Appendix C

Recommendations	
Factors	Influence on fingerprint accuracy
GSR	Increases by regressing out the GS in the fMRI timeseries
Parcellation Scheme	Increases with higher resolutions (Shen 268 → Schaefer 1000)
Parcel Resolution	Increases as parcel resolution increases (7 → 1095)
Parcel Coverage	Increases with cortical regions Decreases with subcortical and cerebellar structures
Network Organisation	Increases with medial frontal and frontoparietal networks Increases with medial frontal + frontoparietal networks Decreases with visual, subcortical and limbic networks
Discriminative Edges	Increases with higher number of differentiating edges (top 10%) Decreases with lower number of differentiating edges (top 1%)
Scan Duration	Increases with longer scan lengths (0.5 min → full scan duration)
Amount of fMRI Data	Decreases with increasing sample sizes (20 → 264)

Table A.1. Recommendations for describing an optimal pipeline to reliably measure the functional connectome at the individual level. The recommendations follow the underlying factors that influence fingerprint accuracy at the data pre- and post-processing stages as well as during the data acquisition phase.

Appendix D

Bagging applies a popular bootstrapping sampling strategy by generating L random samples of size B *with* replacement from an initial dataset of a given size N (Figure A.1). The bootstrap samples are generated from a training set to produce multiple train sets or “bootstraps” across a number of iterations (e.g., ≥ 100) (Varoquaux et al., 2017; Opitz and Maclin, 1999; Breiman, 1996) — which means that a value or instance may be repeated more than once in a bootstrap sample. The bootstrap samples should typically satisfy the statistical properties of *representativity* and *independence*. Here, representativity infers that the size N of a given dataset should be powered enough to explain the majority of the complexity of the distribution so that the distribution of the bootstrap samples can approximate to the real distribution of the dataset. Subsequently, independence means that the size N of a given dataset should be large enough related to the size B of the bootstrap samples to prevent the samples from being strongly correlated with the initial dataset. With the two statistical properties in mind, the bootstrap samples can be used to compute the confidence intervals (CIs) as a function of some observations. The bootstrap samples can then be trained independently and in parallel using the weak learners. Based on the goal of the model, e.g., regression, a weak learner is fitted for each bootstrap sample (L times) and then aggregated by averaging their predictions to obtain an ensemble model — a process called soft voting. In the case of classification, the class that is returned by each weak learner from the L bootstrap samples is treated as a vote and the class that receives the highest or majority of votes is output by the ensemble model — a process called hard voting.

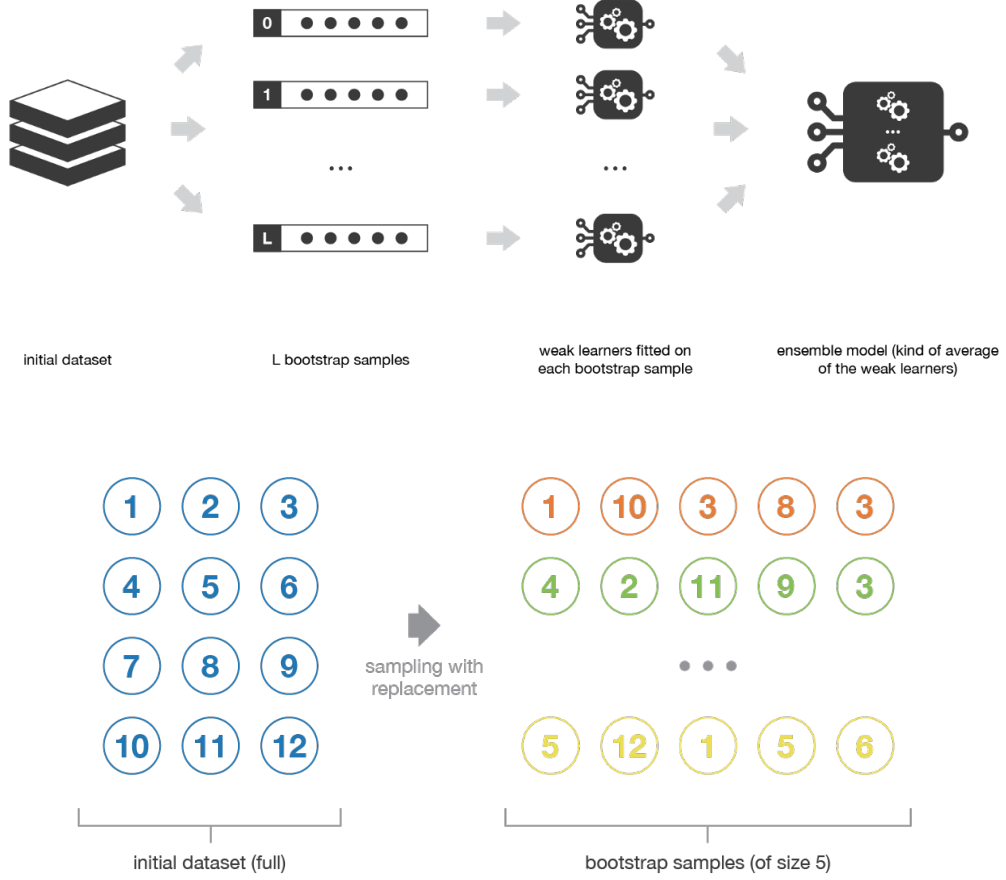


Figure A.1. Bagging framework which applies a bootstrap sampling strategy to create an ensemble model.

In practice, we have L bootstrap samples of size B ,

$$\{z_1^1, z_2^1, z_3^1, \dots, z_B^1\}, \{z_1^2, z_2^2, z_3^2, \dots, z_B^2\}, \dots, \{z_1^L, z_2^L, z_3^L, \dots, z_B^L\} \quad (\text{A.1})$$

where z_b^L denotes the b -th observation derived from the L -th bootstrap sample. We then fit a weak learner denoted by $w_l(\bullet)$ to each of the L bootstrap samples,

$$w_1(\bullet), w_2(\bullet), \dots, w_L(\bullet) \quad (\text{A.2})$$

and aggregate them by simply averaging the L weak learners to produce an ensemble model for

$$\text{regression: } s_L(\bullet) = \frac{1}{L} \sum_{l=1}^L w_l(\bullet) \quad (\text{A.3})$$

$$\text{classification: } s_L(\bullet) = \underset{k}{\operatorname{argmax}} [\operatorname{card}(l \mid w_l(\bullet) = k)] \quad (\text{A.4})$$

Appendix E

CPM expects an $M \times M$ functional connectivity matrix for each participant in a given dataset, where M typically represents the brain areas obtained from a parcellation scheme and the values correspond to the edge or connection between any two brain regions followed by a $N \times 1$ vector, where N refers to the behavioural scores of the participants. As described in **Chapter 5 (Section 5.1)**, the CPM protocol comprises the following key steps: (1) feature selection; (2) feature summarisation; (3) model building; and (4) assessment of the prediction significance. The CPM framework takes the functional connectivity matrix derived from the whole-brain connectome and behavioural score of every participant and splits the dataset into training and testing sets. In the training set, each edge in the connectivity matrices is related to the behavioural scores using a variant of linear regression, where the most meaningful edges that share significant positive, negative, or combined brain-behaviour associations (e.g., $P < 0.05$) are selected to build the predictive model. It is also possible to select network-specific connectomes instead of the whole-brain to compute the functional connectivity matrices on the basis of *a priori* hypotheses. In other words, the feature selection and summarisation steps can focus on only one functional network so that their relevant edges can be retained during model fitting. For each participant, the meaningful edges are then summarised into a single value by adding their respective correlation strengths. Hence, the predictive model assumes linearity ($y = mx + b$) between the summarised brain connectivity data and behavioural scores. Finally, the summary values are computed for every participant in the testing set to obtain the predicted behavioural scores for the unseen participants. The performance of the predictive model can be derived by comparing the predicted behavioural scores with their true values. In order to assess the significance of the model prediction, permutation testing is applied to generate a null distribution of the test statistic by randomly shuffling the behavioural scores of the participants but preserving the structure of the connectivity matrices (e.g., Participant A's functional connectivity matrix is paired with Participant B's behavioural score) over a number of iterations (e.g., 100-10,000). The statistical significance is then obtained by calculating the percentage of the permutations that are greater or equal to the true prediction estimate.

CPM has an in-built cross validation (CV) procedure such as leave-one-out [LOOCV] and k -fold CV to allow the generalisability of model predictions in unseen participants. In an effort to predict an individual behavioural phenotype derived from patterns of functional connectivity, CV divides a dataset D into training and testing sets that are non-overlapping to allow independence between feature selection and model prediction (Scheinost et al., 2019; Varoquaux et al., 2017; Whelan and Garavan, 2014). While the training set is used to build the model using a form of regression, the testing set assesses the generalisability of the model (Yeung et al., 2022). For example, k -fold

CV ensures that D is randomly divided into k samples of equal size, where the $k-1$ samples correspond to the training set and the remaining k samples represent the testing set. This assignment is performed k times to allow each of the k subsets to serve as the testing set once. The model performance can then be assessed by averaging the predictions across the k combinations of training and testing sets. In LOOCV, D is split into training and testing sets, where only one sample is used as the testing data and the remaining data are used to train the model. This process is performed n times, where n is the size of D , and model performance is evaluated by averaging across n possible combinations. LOOCV is generally used when D is small (i.e., lower n) as it is less computationally intensive and increases the size of the training data albeit the potential risk of overfitting (Scheinost et al., 2019). The application of k -fold CV, however, depends on minimising the bias-variance trade-off given that k is likely to have an inverse effect on sample size (Scheinost et al., 2019; Varoquaux et al., 2017). While it is essential to note that there is no one-size-fits-all k value, a higher k decreases bias by minimising the difference between the predicted and observed behavioural measure in addition to increasing the variability of the model prediction under different subsets of the training data. Hence, a balance is needed to have enough training data to reduce potential bias during model fitting and also, sufficient testing data to reduce the amount of variance during model prediction. In theory, 5 or 10-fold CV and 10-20% LOOCV (i.e., leave out 10-20% of the dataset to serve as the testing set) have been recommended on the basis of a given sample size (Scheinost et al., 2019; Varoquaux et al., 2017).

It is important to remember that CPM focuses on predicting linear relationships between brain connectivity features and behavioural scores, rather than non-linear (e.g., quadratic, cubic) associations. These linear predictive models could further generate a range of predicted behavioural measures that is lower compared to their true observed scores such that the smallest scores could be overestimated, and the highest scores could be underestimated (Shen et al., 2017). CPM requires the following parameter tuning:

1. Type of CV to be performed (e.g., LOOCV, k -fold)
2. Number of iterations for CV (e.g., 100-1000)
3. Method to control for confounds (e.g., adjust with partial correlation during feature selection, adjust with multiple regression during model fitting, adjust at both feature selection and model fitting steps)
4. Method to relate functional connectivity patterns to the given behavioural scores (e.g., Pearson's correlation, Spearman's Rank correlation, partial correlation, robust regression)
5. Feature selection type (e.g., P -value, sparsity)

6. Feature selection threshold (e.g., P -value = 0-1 or sparsity = 0-100%)
7. Consistency of a significantly correlated edge with the target behavioural variable to fall in all of the k -fold CV (e.g., 0-100%)
8. Number of iterations to perform random permutations (e.g., 1000-10,000)