dMRI and fMRI Estimated Connectomic Collection

Wilson Tang[1], Ross Lawrence[2*], Eric Bridgeford[1], Fang Cai[1], Jialin Kang[1], Chuankai Luo[1],Zhen Hu [1], Xuemin Zhu[1]

August 9, 2020

1. Johns Hopkins University *corresponding author(s): Firstname Lastname (email@address)

Abstract

Understanding the variability in brain connectivity is essential to understanding the human brain. Currently brains are commonly modeled as a connectome by modeling brain regions as nodes and connections as edges. Currently, there is a gap between the raw data, diffusion magnetic resonance imaging (dMRI) files , and actual connectomes, which can be complex and variable in development and usage. The m2g python package, from our Johns Hopkins Open Connectome Project, is a new end to end MRI connectome estimation pipeline that can be run and has been tested in large MRI datasets. Using this package we have generated a large set of standardized connectomes using publically available, single shell dMRI and fMRI data (32 datasets). This large amount of BIDs formatted data and the resulting connectomes can be found at our website `https://m2g-data-web.readthedocs.io/en/gh-pages/` and can be directly used by researchers.

Background & Summary

Brain connectivity research is essential to furthering our understanding of the human brain. The advent of magnetic resonance imaging has allowed scientists to study the brain. To study brain connectivity, brains are often modelled as 'connectomes' (a functional or structural map of connectivity), or brain graphs, by defining regions of the brain as nodes and the strength of connections between regions as edges. While great efforts have been made to increase the availability of MRI datasets for research, there is a significant barrier to convert from available or even private MRI datasets into connectomes that can be analyzed. Currently, because of the lack of standardization, most studies and labs build and run individualized pipelines[1–12], reproducibility and comparison between these datasets are difficult.

The M2G pipeline was built to address this issue, using a combination of open source tools and parcellations, the pipeline and hyper-parameters have been tuned to act perform all steps of this analysis required for connectomes. The pipeline can accept both raw unprocessed dMRI and T1w images (M2G-d) or raw fMRI and T1w images (M2G-f) and generate diffusion or functional connectomes respectively. Utilizing this powerful new pipeline we have prepared a large set of BIDs[13, 14] formatted, publically available dMRI and fMRI datasets and their connectome results for the public to use to investigate various relationships and questions about our brain. In addition, we have validated the outputs of this pipeline and the resulting dataset for individual and population variability using two metrics: discriminability [55] and connection consistency within and across studies. These demonstrate that the connectomes have reduced sources of variability and follow expected trends in connection consistency (More detail on these can be found in the Technical Validation section).

At a baseline, this is one of the largest, open databases of connectomes[15] to date. In addition, the standardization of this datasets, pipeline, and connectome results significantly improves the reproducibility and comparative power of any studies that uses our connectome results, enabling further mega-analyses of multi-model connectomics data[16, 17] and investigation of other important neuroscience questions.

We have characterized our dataset's underlying dMRI and fMRI 1 and provided some initial technical validation on the batch connectomes themselves. These analyses of multi-model connectonics data also provide insight some characterization of the connectomes properties as a whole for the users.

Methods

1 Data Preparation Overview

All data was acquired from publically available datasets via the following general steps:

Step 1: Request Access if Required

Step 2: Format/Check BIDS[13, 14] Formatting

Step 3: Run Sample of Dataset Locally and Visually Inspect Output

Step 4: Upload Dataset to S3 and Run Batch Analysis

Step 5: Process and Analyze Connectome Results

Multiple datasets during data preparation were not BIDS complaint and required manual file order adjustment to follow the BIDs standard.

2 M2G Pipeline Overview

A critical aspect of generating the connectomes and part of steps 3 and 4 is the M2G pipeline, a standardized connectome pipeline python package that works on both diffusion (dMRI) and functional (fMRI) magnetic resonance imaging data. (Please refer to the citation for more detail on the pipeline)[20].

2.1 dMRI Pipeline Overview

Step 1: Registration:Basic preprocessing of raw data (skull strip, eddy correction) and multiple registrations (alignment spatially of atlas and NIfTI image) using both FSL's "standard"[19, 21, 22, 36] linear and non-linear boundary based registration pipeline.

Step 2: Tractography: Uses DiPY's[18] csa or csd orientation distribution function (odf) model estimator to generate a model of the diffusion signal. The peak directions of this model in 3-d space are tracked with either DiPy's local or deterministic tracking algorithms to create streamlines representing brain connections.

Step 3: Graph Generation: Connectomes (graphs) are formed by deterministic tracking along fiber streamlines into sub-regions based on spatial[29] proximity or neuroanatomical[24–28, 30–33] similarity.

2.2 fMRI Pipeline Overview

Step 1: Preprocessing: AFNI[34] tools for skull stripping to extract brain and FSL[35, 36] tools for slice timing and motion correction.

Step 2: Registration: FSL[22, 37, 42] to perform nonlinear boundary based registration of fMRI image on desikan space (or whatever atlas is desired).

Step 3: Nuisance Correction: CPAC code that implements general linear model with regressors for quadratic detrending[44, 45], top five white-matter and cerebrospinal fluid principal components[38, 40], and Friston 24-parameter regressors[41]. Low-frequency drift is removed below 0.01 Hz[43], and the first 15 seconds of the fMRI sequence are discarded[39].

Step 4: Graph Generation: Custom python code that compute average time-series for all voxels per region of interest ROI(e.g brain regions in the atlas such as cerebellum), correlations between hemispheric pairs of ROIs, and finally rank transforms the magnitude of the correlation from smallest to largest[55].

A bit of arithmetic provides the definition of T in this setting, again immediately yielding the null distribution and p-value using the `TDist` family of functions from the `stats` package in R.

Code availability

The data processing code to generate the connectome data can be found at `https://github.com/neurodata/m2g`

Connectome data is provided as an .csv adjacency matrix and can be processed by many of the multiple graph analysis packages out there. We recommend using `https://github.com/neurodata/graspy` in combination with statistical packages to analyze the connectomes.

Data Records

We ran M2G using the desikan parcellation [24–28, 30–33] on two main sources of data: the Consortium of Reliability and Reproducibility (30 datasets)[57] and a variety of public datasets found through OpenFMRI (6), in addition to 4 other datasets (KKI2009, NKIENH, ABIDII, and CNPLS). Demographic description of each datasets can be found in 1

The datasets are organized with the Brain Imaging Data Structure (BIDS[13, 14]) - a standard specification for organizing multi-scan, multi-individual, and multi-modal studies[13, 14]. Our data is processed to be in BIDS[13, 14] format from the original shared source (CoRR, openFMRI, and etc...) and is hosted on in an s3 bucket, we provide a site to detail various s3 commands that can download the data: `https://m2g-data-web.readthedocs.io/en/gh-pages/index.html`.

The website is organized by source with the majority being found under CoRR, each dataset has a link back to the original dataset, a quick overview. To access the data we have provided per dataset four s3 commands to download different parts of the dataset as follow:

- Single Sample/Subject of Data
- Single Sample/Subject of Results
- All Data/Subjects of Data
- All Data/Subjects of Results

The main connectome data we provide is (an visualized example can be seen in Figure 1:

- Connectome Adjacency Matrix: Square Matrix representing a finite graph with each element indicated the connection strength between the ROIs at each index. Location: dataset/subject/ses/dwi/roi_connectome (.csv)

We also provide multiple intermediates generated during the pipeline, which we elect not to detail except for a subset of useful pipeline quality assurance images that visualizes various steps in the pipeline(dMRI pipeline only).

- FAST[46] Segmentation: Visual inspection of the segmentation of NiFTI images into grey matter, white matter, and cerebrospinal fluid across 9 brain slices (3 per coronal, sagital, and axial planes). Location: dataset/subject/session/qa/reg/ (qa_fast.png)

- Skull Strip QA: Visual overlay of success of FSL[22] in extracting brain mask vs. skull across 9 representatives brain slices (3 per coronal, sagital, and axial planes). Location: dataset/subject/session/qa/skullstrip/ (.png)

- Tractography QA: Visual heatmap of peak directions in odf models used to estimate diffusion signal across 9 brain slices (3 per coronal, sagital, and axial planes). Location: dataset/subject/session/qa/tensor/ (Tractography_Model_Peak_Directions.png)

- Final Connectome Heatmap: Heatmap of connectome adjacency matrix. Location: dataset/subject/ses/qa/graphs_plotting (.png)

Finally, we provide the BIDs[13, 14] formatted raw data in the corresponding folders(See**??** for a breakdown of dMRI vs. fMRI datasets):

- dMRI Anatomical Image: Basic MRI image using T1W weighting to show anatomical tissue. Location: dataset/subject/ses/anat/ (T1w.nii.gz)

- dMRI Diffusion Weighted Image(dwi): MRI image based on measuring brownian motion of water per tissue voxel. Location: dataset/subject/ses/dwi/ (.nii.gz)

- dMRI Diffusion Weighted Image(dwi) bval: Associated metadata with DWI representing gradient directions used during measurement. Location: - dataset/subject/ses/dwi/ (.bval)

- dMRI Diffusion Weighted Image(dwi) bvec: Associated metadata with DWI representing diffusion weighting used during measurement. Location: - dataset/subject/ses/dwi/ (.bvec)

- fMRI Functional Image: MRI image measuring brain activity by measuring changes in blood flow over time. Location: dataset/subject/ses/func/ (.nii.gz)

Technical Validation

We use discriminability as our main statistic of technical validation for all datasets that had repeat measurements (see Table 1).

Discriminability indicates the probability that two observations within the same class are more similar [55]

$$D = Pr(||a_{ij} - a_{ij'}|| \leq ||a_{ij} - a_{i'j'}||) \tag{1}$$

In our dataset, discriminability provides a context of reliability for each resulting connectome. This is done by comparing each connectome, a_{ij}, to other connectomes from the same individual, $a_{ij'}$, and to all other connectomes belonging to other individuals, $a_{i'j'}$. Table 1 shows that discriminability is high, generally above 0.8 except for 1 dMRI and fMRI dataset. This consistency of discriminability shows that connectomes resulting from a scan for the same patient are more alike than scans compared to every other patient, providing a sense of robustness and reliability.

Our second metric of validation is a checking connection consistency within and across studies on connectomes to confirm that our connectomes follow expected real world behavior. Figure 2 is designed to test differences in magnitudes within and across studies by analyzing connection strengths across two groupings of connections: Ipsilat-eral (within hemisphere) vs. Contralateral (across hemisphere) and Homotopic(Mirrored ROI connections) vs. Heterotopic (ROI connections to non-mirroredROIs). Using a Wilcox-Mann-Whitney rank sum test we confirm that they follow expected behavior in the 4 following ways:

- When measuring within a modality (dMRI and fMRI): Ipsilateral connections are significantly different (100% p_value rejection rate) and stronger than Contralateral connections in fMRI connectomes than in dMRI connectomes.
- When measuring within a modality (dMRI and fMRI): Homotopic connections are significantly different (100% p_value rejection rate) and stronger than Heterotopic connections in dMRI connectomes than in fMRI connectomes.
- When measuring between modalities (dMRI vs. fMRI matched by study and patient): Ipsilateral connections still are significantly different (100% p_value rejection rate) and stronger than Contralateral connections in fMRI connectomes than in dMRI connectomes.
- When measuring between modalities (dMRI vs. fMRI matched by study and patient): Homotopic connections still are significantly different (100% p_value rejection rate) and stronger than Heterotopic connections in dMRI connectomes than in fMRI connectomes.

Thus these demonstrate that these connective strength differences between dMRI and fMRI connectomes hold across studies and patients.

To generate our characterization of some basic properties of our connectomes (shown in Figure 2), we used a Structural Independent Edge Model to analyze the generated connectome data.

2.3 Technical Validation: SIEM Model Statistics

Let $g = (V, E)$ be a graph, where V is the set of vertices that is shared for all i, and E_i is the set of binary undirected edges between pairs of vertices. Let A be a binary adjacency matrix where $a_{uv} = 1$ if and only if there is an edge

between u and v, that is, $(uv) \in E$. Assume g is a realization of a random graph $G \sim F$, which is sampled from a distribution F. We consider a random graph models that generalizes the stochastic block model, the structured independent edge models (SIEM).

The SIEM implies that $A \sim SIEM(P, \tau)$, where τ is a grouping of the $|E|$ edges in G into C non-overlapping communities, that is, $\cup_{i=1}^{C} \tau_i = E$, and $\tau_i \cap \tau_j = \emptyset$ for all $i \neq j$. $P = [p_1, p_2; p_1, p_2]$ represents the parameters for within and between edge group probabilities.

Our hypothesis test can be stated as follows:

$$H_0 : p_0 \leq p_1$$
$$H_A : p_0 > p_1.$$

Ignoring the subscript for notational convenience, when all edges in a group are sampled independently with probability p, then the number of edges in that group follows a binomial distribution. The maximum likelihood estimate of p is $\hat{p} = E[p] = \frac{k}{n}$, where k is the number of observed edges in the group, and n is the number of potential edges in the group.

We utilize large sample size theory to obtain a p-value. When n is sufficient large, $\hat{p}$ has a normal distribution, $\mathcal{N}(\mu_p, \sigma_p^2)$, where $\mu_p = p$ and $\sigma_p^2 = p(1-p)/n$. Plugging in $\hat{p}$ for p yields our estimate of the distribution of $\hat{p}$. Because $\hat{p}_1$ and $\hat{p}_2$ are different, testing whether they are statistically significantly different boils down to testing whether two Gaussians with different variances are different. We use the non-parametric Wilcoxon-Mann-Whitney Test for testing whether populations have a stochastic difference between edge distributions medians. This generates a statistic T, using a rank sum method with ranked (using relative connection magnitude) connectomes (represented as adjacency matrices).

$$T = R_i - \frac{n_i(n_i + 1)}{2},$$

where R_i is the sum of ranks in sample i, n_i is the sample size of sample i.

The null distribution, and therefore p-value, is available from scipy.stats.

$$H_0 : \delta^d \leq \delta^f$$
$$H_A : \delta^d > \delta^f.$$

For testing whether the difference between homotopic and heterotopic connectivity in the functional connectomes exceeds that of the diffusion connectomes, the across-modality test can be stated as follows:

$$H_0 : \delta^f \leq \delta^d$$
$$H_A : \delta^f > \delta^d.$$

Usage Notes

The main connectome data can be used as detailed in the Code Availability Section above, while the supporting "raw" data is primarily intended for running the pipeline.

Useful libraries for doing other analysis on the "raw" data include many of the dependencies M2G is built upon such as Dipy, AFNI, FSL, nibabel, networkx, numpy scipy, scikit-learn, scikit-image, and nilearn. Visualization on these nifti "raw" data images can be done with programs such as FSLeyes.

Acknowledgements

The Acknowledgements should contain text acknowledging non-author contributors. Acknowledgements should be brief, and should not include thanks to anonymous referees and editors or effusive comments. Grant or contribution numbers may be acknowledged.

Author contributions

Each author's contribution to the work should be described briefly, on a separate line, in the Author Contributions section. Authors:

- Eric Bridgeford - Primary Statistical/Technical Validation Analysis Lead
- Fang Cai - Data Collection Organization, Technical Validation
- Jialin Kang - Data Collection Organization
- Ross Lawrence - Co-Author Corresponding Author, Manuscript Writer Data Collection Organization, Dataset Connectome Generation
- Chuankai Luo - Data Collection Organization, Storage, Website
- Wilson Tang - Co-Author?, Manuscript Writer, Data Collection Organization, Technical Validation
- Zhen Hu - Data Collection Organization, Technical Validation
- Xuemin Zhu - Data Collection Organization

Senior Author:

Joshua Vogelstein

Competing interests

A competing interests statement is required for all papers accepted by and published in *Scientific Data*. If there is no conflict of interest, a statement declaring this must still be included in the manuscript.

Figures and figures legends

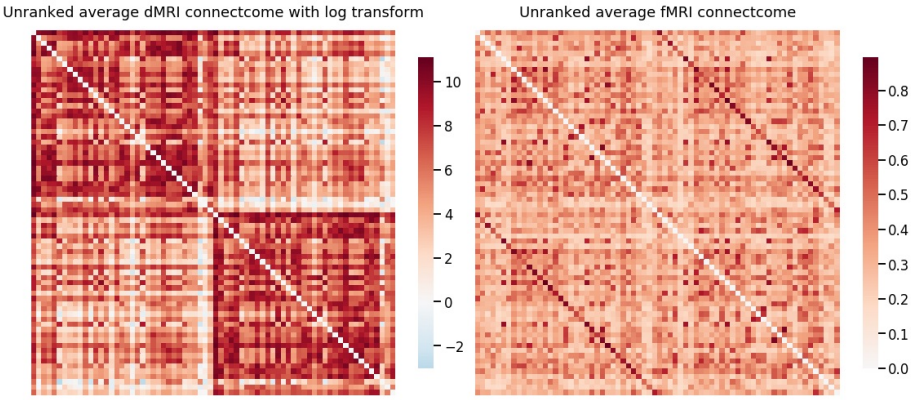
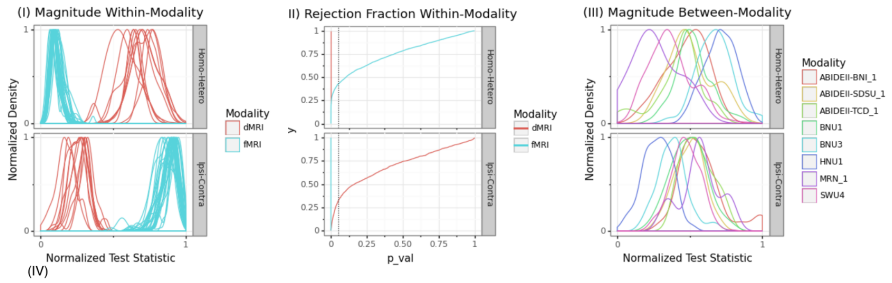


Figure 1: Mean dMRI and fMRI M2G Connectomes



(IV)

Test	dMRI	fMRI	Inter-modality
Ipsilateral stronger than Contralateral	100%	0%	100% of dMRI exceeded fMRI
Homotopic Stronger than Heterotopic	0%	100%	100% of fMRI exceeded dMRI

Figure 2: SIEM analysis of establishes multi-connectome properties are preserved both within and across studies. (I) Magnitude of difference kernel density estimate indicating if distributions of two connection groupings differ: Ipsilateral (within hemisphere) vs. Contralateral (across hemisphere) and Homotopic (Mirrored ROI connections) vs. Heterotopic (ROI connections to non-mirrored ROIs). dMRI Connectomes have stronger homo-hetero connective difference but weaker ipsi-contra connections than fMRI Connectomes and is inverted for fMRI to dMRI. (II) Rejection Fraction of the CDF of within modality comparisons on average indicate that essentially 100% of the time dMRI Connectomes have stronger homo-hetero connective difference but weaker ipsi-contra connections than fMRI Connectomes and is inverted for fMRI to dMRI. (III) For the eight studies with both dMRI and fMRI data, a paired comparison across modalities per individual for the same 2 connection groupings. Ipsilateral vs. Contralateral connective difference in dMRI exceed fMRI and Homotopic vs. Heterotopic connective differences in fMRI exceed dMRI. (IV) Table showing that 100% of the time all individual exhibited significant effects of each analysis described above matching Part (II), demonstrating consistency across both individuals and studies.

Tables

Study	Scanner	TR (s)	Age ± std	%Male	Indiv.	Ses.	dMRI #Scans	dMRI D	fMRI #Scans	fMRI D
ABIDEII-BNI_1[51]	P	3	39.5±15.1	100	29	1	29	N/A	29	N/A
ABIDEII-SDSU_1[50]	GE	2	13.3±3.0	92	25	1	25	N/A	25	N/A
ABIDEII-TCD_1[50]	P	2	15.6±3.1	100	21	1	21	N/A	21	N/A
IACAS_1[57]	GE	2	26.4±5.0	53	28	2	N/A	N/A	55	0.761039
BMB_1[57]	S	2.3	30.8±7.1	52	50	2-4	N/A	N/A	120	0.95137
BNU1[57]	S	2	23.0±2.3	53	57	2	144	0.926	106	0.94324
BNU2[57]	S	2.53	21.9±0.9	54	61	2	N/A	N/A	122	0.94324
BNU3[57]	S	2	22.5±2.1	50	46	3	46	N/A	144	0.8944
HNU1[57]	GE	2	24.4±2.3	50	30	10	298	0.966	300	0.94876
MRN1[57]	S	2	15.1±7.8	50	53	1-2	20	0.708333333	89	0.848214
IBATRT[57]	S	1.75	28.0±7.5	54	36	2-4	N/A	N/A	100	0.95392
IPCAS_1[57]	S	2	20.9±1.7	30	30	3-4	60	N/A	114	0.892127
IPCAS_2[57]	S	2.5	13.4±0.9	35	35	2	N/A	N/A	70	0.90714
IPCAS_3[57]	S	2	21.1±1.7	66	35	2-4	N/A	N/A	80	0.86022
IPCAS_4[57]	GE	2	23.2±1.6	50	20	2	N/A	N/A	40	0.914789
IPCAS_5[57]	S	2	18.3±0.5	100	22	3	N/A	N/A	66	0.847643
IPCAS_6[57]	S	2.5	23.0±2.0	50	2	45	N/A	N/A	90	0.9786
IPCAS_7[57]	S	2.5	11.6±3.1	57	72	2	N/A	N/A	144	0.9481866
IPCAS_8[56]	S	2	57.6±3.6	46	13	2	N/A	N/A	26	0.907051
JHNU[57]	S	2	23.3±3.7	70	30	2	N/A	N/A	60	0.866379
NKIENH[54]	S	N/A	42.5±19.6	40	192	1	192	N/A	N/A	N/A
NKI1[57]	S	N/A	34.4±12.8	0	20	2	36	0.924	N/A	N/A
NYU1[57]	S	2	29.4±8.4	40	25	3	N/A	N/A	75	0.931019
NYU2[57]	S	2	20.2±11.5	38	187	2-4	N/A	N/A	300	0.898203
SWU1[57]	S	2	21.6±1.7	30	20	6	N/A	N/A	119	0.929312
SWU2[57]	S	2	21.0±1.6	33	27	2	N/A	N/A	54	0.866096
SWU3[57]	S	2	20.4±1.6	63	24	2	N/A	N/A	48	0.958333
SWU4[57]	S	2	20.0±1.3	51	235	2	382	0.79	429	0.890612
UPSM[52]	S	1.5	15.1±2.8	48	100	2/3	N/A	N/A	230	0.806593
Utah1[48,49]	S	2	20.2±8.3	100	26	3	N/A	N/A	77	0.856687
UWM[57]	G	2.6	25.0±3.2	56	25	3	N/A	N/A	75	0.930833
XHCUMS[53]	S	3	52.0±6.3	50	24	5-14	120	0.9563	247	0.848269

Table 1: DataSet Demographics. Discriminability is above 0.8 for almost every single multi-measurement dataset for both fMRI and dMRI. D is discriminability. Scanners are G (GE), P(Phillips), or S (Siemens)

References

[1] Vince D Calhoun, Tulay Adali, Godfrey D Pearlson, and James J Pekar. A method for making group inferences from functional mri data using independent component analysis. *Human brain mapping*, 14(3):140–151, 2001.

[2] Yan Chao-Gan and Zang Yu-Feng. Dparsf: a matlab toolbox for "pipeline" data analysis of resting-state fmri. *Frontiers in systems neuroscience*, 4, 2010.

[3] R Cameron Craddock, G Andrew James, Paul E Holtzheimer III, Xiaoping P Hu, and Helen S Mayberg. A whole brain fmri atlas generated via spatially constrained spectral clustering. *Human brain mapping*, 33(8):1914–1928, 2012.

[4] Zaixu Cui, Suyu Zhong, Pengfei Xu, Gaolang Gong, and Yong He. Panda: a pipeline toolbox for analyzing brain diffusion images. *Frontiers in human neuroscience*, 7:42, 2013.

[5] Alessandro Daducci, Stephan Gerhard, Alessandra Griffa, Alia Lemkaddem, Leila Cammoun, Xavier Gigandet, Reto Meuli, Patric Hagmann, and

Jean-Philippe Thiran. The connectome mapper: an open-source processing pipeline to map connectomes with mri. *PloS one*, 7(12), 2012.

[6] William R Gray, John A Bogovic, Joshua T Vogelstein, Bennett A Landman, Jerry L Prince, and R Jacob Vogelstein. Magnetic resonance connectome automated pipeline: an overview. *IEEE pulse*, 3(2):42–48, 2012.

[7] William Gray Roncal, Zachary H Koterba, Disa Mhembere, Dean M Kleissas, Joshua T Vogelstein, Randal Burns, Anita R Bowles, Dimitrios K Donavos, Sephira Ryman, Rex E Jung, et al. Migraine: Mri graph reliability analysis and inference for connectomics. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 313–316. IEEE, 2013.

[8] Xiao-Wei Song, Zhang-Ye Dong, Xiang-Yu Long, Su-Fang Li, Xi-Nian Zuo, Chao-Zhe Zhu, Yong He, Chao-Gan Yan, and Yu-Feng Zang. Rest: a toolkit for resting-state functional magnetic resonance imaging data processing. *PloS one*, 6(9), 2011.

[9] Jinhui Wang, Xindi Wang, Mingrui Xia, Xuhong Liao, Alan Evans, and Yong He. Gretna: a graph theoretical network analysis toolbox for imaging connectomics. *Frontiers in human neuroscience*, 9:386, 2015.

[10] Ting Xu, Zhi Yang, Lili Jiang, Xiu-Xia Xing, and Xi-Nian Zuo. A connectome computation system for discovery science of brain. *Science Bulletin*, 60(1):86–95, 2015.

[11] Chao-Gan Yan, Xin-Di Wang, Xi-Nian Zuo, and Yu-Feng Zang. Dpabi: data processing & analysis for (resting-state) brain imaging. *Neuroinformatics*, 14(3):339–351, 2016.

[12] Xi-Nian Zuo, Ting Xu, Lili Jiang, Zhi Yang, Xiao-Yan Cao, Yong He, Yu-Feng Zang, F Xavier Castellanos, and Michael P Milham. Toward reliable characterization of functional homogeneity in the human brain: preprocessing, scan duration, imaging resolution and computational space. *Neuroimage*, 65:374–386, 2013.

[13] Jesse A Brown and John D Van Horn. Connected brains and minds—the umcd repository for brain connectivity matrices. *Neuroimage*, 124:1238–1241, 2016.

[14] Krzysztof J Gorgolewski, Fidel Alfaro-Almagro, Tibor Auer, Pierre Bellec, Mihai Capotă, M Mallar Chakravarty, Nathan W Churchill, Alexander Li Cohen, R Cameron Craddock, Gabriel A Devenyi, et al. Bids apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. *PLoS computational biology*, 13(3):e1005209, 2017.

[15] Krzysztof J Gorgolewski, Tibor Auer, Vince D Calhoun, R Cameron Craddock, Samir Das, Eugene P Duff, Guillaume Flandin, Satrajit S Ghosh, Tristan Glatard, Yaroslav O Halchenko, et al. The brain imaging data structure,

a format for organizing and describing outputs of neuroimaging experiments. *Scientific data*, 3(1):1–9, 2016.

[16] Gaël Varoquaux and R Cameron Craddock. Learning and comparing functional connectomes across subjects. *NeuroImage*, 80:405–415, 2013.

[17] Diego Vidaurre, Romesh Abeysuriya, Robert Becker, Andrew J Quinn, Fidel Alfaro-Almagro, Stephen M Smith, and Mark W Woolrich. Discovering dynamic brain networks from big data in rest and task. *Neuroimage*, 180:646–656, 2018.

[18] Eleftherios Garyfallidis, Matthew Brett, Bagrat Amirbekian, Ariel Rokem, Stefan Van Der Walt, Maxime Descoteaux, and Ian Nimmo-Smith. Dipy, a library for the analysis of diffusion mri data. *Frontiers in neuroinformatics*, 8:8, 2014.

[19] Mark Jenkinson, Christian F Beckmann, Timothy EJ Behrens, Mark W Woolrich, and Stephen M Smith. Fsl. *Neuroimage*, 62(2):782–790, 2012.

[20] Gregory Kiar, Eric Bridgeford, Will Gray Roncal, Vikram Chandrashekhar, Disa Mhembere, Sephira Ryman, Xi-Nian Zuo, Daniel S Marguiles, R Cameron Craddock, Carey E Priebe, et al. A high-throughput pipeline identifies robust connectomes but troublesome variability. *bioRxiv*, page 188706, 2018.

[21] John Mazziotta, Arthur Toga, Alan Evans, Peter Fox, Jack Lancaster, Karl Zilles, Roger Woods, Tomas Paus, Gregory Simpson, Bruce Pike, et al. A four-dimensional probabilistic atlas of the human brain. *Journal of the American Medical Informatics Association*, 8(5):401–430, 2001.

[22] Stephen M Smith, Mark Jenkinson, Mark W Woolrich, Christian F Beckmann, Timothy EJ Behrens, Heidi Johansen-Berg, Peter R Bannister, Marilena De Luca, Ivana Drobnjak, David E Flitney, et al. Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage*, 23:S208–S219, 2004.

[23] Mark W Woolrich, Saad Jbabdi, Brian Patenaude, Michael Chappell, Salima Makni, Timothy Behrens, Christian Beckmann, Mark Jenkinson, and Stephen M Smith. Bayesian analysis of neuroimaging data in fsl. *Neuroimage*, 45(1):S173–S186, 2009.

[24] Rahul S Desikan, Florent Ségonne, Bruce Fischl, Brian T Quinn, Bradford C Dickerson, Deborah Blacker, Randy L Buckner, Anders M Dale, R Paul Maguire, Bradley T Hyman, et al. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, 31(3):968–980, 2006.

[25] Matthew F Glasser, Timothy S Coalson, Emma C Robinson, Carl D Hacker, John Harwell, Essa Yacoub, Kamil Ugurbil, Jesper Andersson,

Christian F Beckmann, Mark Jenkinson, et al. A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171–178, 2016.

[26] Daniel Kessler, Michael Angstadt, Robert C Welsh, and Chandra Sripada. Modality-spanning deficits in attention-deficit/hyperactivity disorder in functional networks, gray matter, and white matter. *Journal of Neuroscience*, 34(50):16555–16566, 2014.

[27] JLSJ Lancaster. The talairach daemon, a database server for talairach atlas labels. *Neuroimage*, 5(4):238–242, 1997.

[28] Nikos Makris, Jill M Goldstein, David Kennedy, Steven M Hodge, Verne S Caviness, Stephen V Faraone, Ming T Tsuang, and Larry J Seidman. Decreased volume of left and total anterior insular lobule in schizophrenia. *Schizophrenia research*, 83(2-3):155–171, 2006.

[29] Disa Mhembere, William Gray Roncal, Daniel Sussman, Carey E Priebe, Rex Jung, Sephira Ryman, R Jacob Vogelstein, Joshua T Vogelstein, and Randal Burns. Computing scalable multivariate glocal invariants of large (brain-) graphs. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 297–300. IEEE, 2013.

[30] Susumu Mori, Setsu Wakana, Peter CM Van Zijl, and LM Nagae-Poetscher. *MRI atlas of human white matter*. Elsevier, 2005.

[31] Chandra S Sripada, Daniel Kessler, and Mike Angstadt. Lag in maturation of the brain's intrinsic functional architecture in attention-deficit/hyperactivity disorder. *Proceedings of the National Academy of Sciences*, 111(39):14259–14264, 2014.

[32] Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Olivier Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289, 2002.

[33] Liang Wang, Ryan EB Mruczek, Michael J Arcaro, and Sabine Kastner. Probabilistic maps of visual topography in human cortex. *Cerebral cortex*, 25(10):3911–3931, 2015.

[34] Robert W Cox. Afni: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research*, 29(3):162–173, 1996.

[35] Mark Jenkinson, Peter Bannister, Michael Brady, and Stephen Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2):825–841, 2002.

[36] Mark W Woolrich, Saad Jbabdi, Brian Patenaude, Michael Chappell, Salima Makni, Timothy Behrens, Christian Beckmann, Mark Jenkinson, and Stephen M Smith. Bayesian analysis of neuroimaging data in fsl. *Neuroimage*, 45(1):S173–S186, 2009.

[37] Jesper LR Andersson, Mark Jenkinson, Stephen Smith, and Jesper Andersson. Non-linear registration, aka spatial normalisation. 2007. *URL https://www. fmrib. ox. ac. uk/datasets/techrep/tr07ja2/tr07ja2. pdf*, 2013.

[38] Yashar Behzadi, Khaled Restom, Joy Liau, and Thomas T Liu. A component based noise correction method (compcor) for bold and perfusion based fmri. *Neuroimage*, 37(1):90–101, 2007.

[39] Molly G Bright, Christopher R Tench, and Kevin Murphy. Potential pitfalls when denoising resting state fmri data using nuisance regression. *Neuroimage*, 154:159–168, 2017.

[40] Rastko Ciric, Daniel H Wolf, Jonathan D Power, David R Roalf, Graham L Baum, Kosha Ruparel, Russell T Shinohara, Mark A Elliott, Simon B Eickhoff, Christos Davatzikos, et al. Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *Neuroimage*, 154:174–187, 2017.

[41] Karl J Friston, Steven Williams, Robert Howard, Richard SJ Frackowiak, and Robert Turner. Movement-related effects in fmri time-series. *Magnetic resonance in medicine*, 35(3):346–355, 1996.

[42] Douglas N Greve and Bruce Fischl. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, 48(1):63–72, 2009.

[43] Martin A Lindquist, Ji Meng Loh, Lauren Y Atlas, and Tor D Wager. Modeling the hemodynamic response function in fmri: efficiency, bias and mis-modeling. *Neuroimage*, 45(1):S187–S198, 2009.

[44] Anne M Smith, Bobbi K Lewis, Urs E Ruttimann, Q Ye Frank, Teresa M Sinnwell, Yihong Yang, Jeff H Duyn, and Joseph A Frank. Investigation of low frequency drift in fmri signal. *Neuroimage*, 9(5):526–533, 1999.

[45] Jody Tanabe, David Miller, Jason Tregellas, Robert Freedman, and Francois G Meyer. Comparison of detrending methods for optimal fmri preprocessing. *NeuroImage*, 15(4):902–907, 2002.

[46] Yongyue Zhang, Michael Brady, and Stephen Smith. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging*, 20(1):45–57, 2001.

[47] Xi-Nian Zuo, Jeffrey S Anderson, Pierre Bellec, Rasmus M Birn, Bharat B Biswal, Janusch Blautzik, John CS Breitner, Randy L Buckner, Vince D

Calhoun, F Xavier Castellanos, et al. An open science resource for establishing reliability and reproducibility in functional connectomics. *Scientific data*, 1:140049, 2014.

[48] Jeffrey S Anderson, T Jason Druzgal, Alyson Froehlich, Molly B DuBray, Nicholas Lange, Andrew L Alexander, Tracy Abildskov, Jared A Nielsen, Annahir N Cariello, Jason R Cooperrider, et al. Decreased interhemispheric functional connectivity in autism. *Cerebral cortex*, 21(5):1134–1146, 2011.

[49] Jeffrey S Anderson, Jared A Nielsen, Alyson L Froehlich, Molly B DuBray, T Jason Druzgal, Annahir N Cariello, Jason R Cooperrider, Brandon A Zielinski, Caitlin Ravichandran, P Thomas Fletcher, et al. Functional connectivity magnetic resonance imaging classification of autism. *Brain*, 134(12):3742–3754, 2011.

[50] Adriana Di Martino, David O'connor, Bosi Chen, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Joshua H Balsters, Leslie Baxter, Anita Beggiato, Sylvie Bernaerts, et al. Enhancing studies of the connectome in autism using the autism brain imaging data exchange ii. *Scientific data*, 4(1):1–15, 2017.

[51] Adriana Di Martino, Chao-Gan Yan, Qingyang Li, Erin Denio, Francisco X Castellanos, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Susan Y Bookheimer, Mirella Dapretto, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6):659–667, 2014.

[52] Kai Hwang, Michael N Hallquist, and Beatriz Luna. The development of hub architecture in the human functional brain network. *Cerebral Cortex*, 23(10):2380–2393, 2013.

[53] Jie Lu, Hesheng Liu, Miao Zhang, Danhong Wang, Yanxiang Cao, Qingfeng Ma, Dongdong Rong, Xiaoyi Wang, Randy L Buckner, and Kuncheng Li. Focal pontine lesions provide evidence that intrinsic functional connectivity reflects polysynaptic anatomical pathways. *Journal of Neuroscience*, 31(42):15065–15071, 2011.

[54] Kate Brody Nooner, Stanley Colcombe, Russell Tobe, Maarten Mennes, Melissa Benedict, Alexis Moreno, Laura Panek, Shaquanna Brown, Stephen Zavitz, Qingyang Li, et al. The nki-rockland sample: a model for accelerating the pace of discovery science in psychiatry. *Frontiers in neuroscience*, 6:152, 2012.

[55] Shangsi Wang, Zhi Yang, Michael Milham, Cameron Craddock, Xi-Nian Zuo, Carey E Priebe, and Joshua T Vogelstein. Optimal experimental design for generating reference connectome datasets. In *Organization for Human Brain Mapping, 21st Annual Meeting*, 2015.

[56] Gao-Xia Wei, Hao-Ming Dong, Zhi Yang, Jing Luo, and Xi-Nian Zuo. Tai chi chuan optimizes the functional organization of the intrinsic human brain architecture in older adults. *Frontiers in Aging Neuroscience*, 6:74, 2014.

[57] Xi-Nian Zuo, Jeffrey S Anderson, Pierre Bellec, Rasmus M Birn, Bharat B Biswal, Janusch Blautzik, John CS Breitner, Randy L Buckner, Vince D Calhoun, F Xavier Castellanos, et al. An open science resource for establishing reliability and reproducibility in functional connectomics. *Scientific data*, 1:140049, 2014.