bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The Neuro Bureau ADHD-200 Preprocessed Repository

Pierre Bellec[a,b,c,*], Carlton Chu[a,d], François Chouinard-Decorte[a,b,e], Yassine Benhajali[a,b,f], Daniel S. Margulies[a,g], R. Cameron Craddock[a,h,i,*]

[a] *The Neuro Bureau*
[b] *Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal, Montréal, CA*
[c] *Département d'Informatique et de Recherche Opérationnelle, Université de Montréal, Montréal, CA*
[d] *Google DeepMind, London, UK*
[e] *Integrated Program in Neuroscience, McGill University, Montreal, CA*
[f] *Département d'Anthropologie, Université de Montréal, Montréal, CA*
[g] *Max Planck Research Group for Neuroanatomy & Connectivity, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*
[h] *Computational Neuroimaging Laboratory, Center for Biomedical Imaging and Neuromodulation, Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY, USA*
[i] *Center for the Developing Brain, Child Mind Institute, New York, NY, USA*

Abstract

In 2011, the "ADHD-200 Global Competition" was held with the aim of identifying biomarkers of attention-deficit/hyperactivity disorder from resting-state functional magnetic resonance imaging (rs-fMRI) and structural MRI (s-MRI) data collected on 973 individuals. Statisticians and computer scientists were potentially the most qualified for the machine learning aspect of the competition, but generally lacked the specialized skills to implement the necessary steps of data preparation for rs-fMRI. Realizing this barrier to entry, the Neuro Bureau prospectively collaborated with all competitors by preprocessing the data and sharing these results at the Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC) (`http://www.nitrc.org/frs/?group_id=383`). This "ADHD-200 Preprocessed" release included multiple analytical pipelines to cater to different philosophies of data analysis. The processed derivatives included denoised and registered 4D fMRI volumes, regional time series extracted from brain parcellations, maps of 10 intrinsic connectivity networks, fractional amplitude of low frequency fluctuation, and regional homogeneity, along with grey matter density maps. The data was used by several teams who competed in the ADHD-200 Global Competition, including the winning entry by a group of biostaticians. To the best of our knowledge, the ADHD-200 Preprocessed release was the first large public resource of preprocessed resting-state fMRI and structural MRI data, and remains to this day the only resource featuring a battery of alternative processing paths.

Keywords: preprocessed fMRI, data sharing, neuroimaging competition

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1. Introduction

In 2011, the "ADHD-200 Global Competition" was held with the aim of engaging researchers from a variety of analytical backgrounds to identify biomarkers of attention-deficit/hyperactivity disorder (ADHD) from resting-state functional magnetic resonance imaging (rs-fMRI) and structural MRI (s-MRI) data [1]. The competition made use of the "ADHD-200 Sample" data collection that was aggregated from eight independent sites and shared through the Intenational Neuroimaging Datasharing Initiative (INDI) [2]. The data includes rs-fMRI, structural MRI (s-MRI), and basic phenotypic information for 973 individuals: some typically-developing controls (TDC) and patients diagnosed with ADHD [1]. Competitors were given five and a half months to optimize a classification algorithm on training data (776 individuals) and submit their predicted clinical labels on test data for which diagnostic information was withheld. The competition data was distributed in a raw form and, before any analysis could begin, the images had to be preprocessed to make them comparable across individuals and reduce noise. These preprocessing steps present a significant hurdle for would-be competitors who do not have the specialist knowledge of neuroimaging methods, or access to high performance computing resources. Realizing this barrier to entry, the Neuro Bureau, a non-profit organization aimed at facilitating open science grassroots initiatives[1], prospectively collaborated with all competitors by preprocessing the data and sharing these results.

The "ADHD-200 Preprocessed" is a repository of preprocessed rs-fMRI and s-MRI data along with statistical derivatives from the ADHD-200 Sample. Rather than favoring a specific processing strategy, we followed a pluralistic approach by preprocessing the data using multiple pipelines (called "Athena", "Burner", and "NIAK") that differed in the toolsets used, the philosophy motivating choices of algorithms and parameters, and the statistical derivatives calculated. The Athena pipeline processed rs-fMRI and s-MRI images using a combination of AFNI [3] and FSL [4] neuroimaging toolkits. The Burner pipeline used SPM8 [5] to process s-MRI data for voxel-based morphometry. The NIAK pipeline processed rs-fMRI and s-MRI using the NeuroImaging Analysis Kit [6].

2. Organization and access to the repository

The ADHD-200 Preprocessed data was released in 2011 and can be downloaded from NITRC[2]. No data usage agreement is required to access or download the data, the only

*Corresponding authors.

Email addresses: `pierre.bellec@criugm.qc.ca` (Pierre Bellec), `carltonchu1@gmail.com` (Carlton Chu), `francois.chouinard@gmail.com` (François Chouinard-Decorte), `yanamarji@gmail.com` (Yassine Benhajali), `margulies@cbs.mpg.de` (Daniel S. Margulies), `ccraddock@nki.rfmh.org` (R. Cameron Craddock)

URL: `bellec.simexp-lab.org` (Pierre Bellec), `computational-neuroimaging-lab.org` (R. Cameron Craddock)

[1]See `http://www.neurobureau.org/mission-statement/` for the full mission statement. The Neurobureau is a non-profit organization registered in Germany

[2]`http://www.nitrc.org/frs/?group_id=383`

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

requirement is registering for a free NITRC account. This registration enables downloads to be tracked for usage statistics and users to be contacted in the event that errors are found in the dataset. The ADHD-200 Sample allows unrestricted data usage for non-commercial research purposes provided that the specific datasets included in an analysis be cited appropriately and that their funding sources be acknowledged[3]. There are no more restrictions placed on the preprocessed data or derivatives other than the request that the ADHD-200 Preprocessed Initiative is cited appropriately and that the specific pipeline is acknowledged in publications using the data. A forum is available on the Neuro Bureau's NITRC project page for users to ask questions or report problems[4]. Questions regarding data acquisition or phenotypic variables should be directed to INDI's support forum[5].

3. Contents of the repository

The ADHD-200 Preprocessed repository contains preprocessed outputs and derivatives for data from the ADHD-200 Sample, which includes 973 individuals (352 F) between the ages of 7 and 27 aggregated from 17 different studies conducted across 8 different sites (for a breakdown of age and sex by diagnosis, see Table 1). For each individual, phenotypic data includes sex, age, handedness, ADHD diagnosis (585 TDC, 362 ADHD, 26 with diagnosis unavailable), ADHD subtype (ADHD-combined, ADHD-inattentive, ADHD-hyperactive/impulsive), one of three different measures of ADHD severity, one of five measures of intelligence, co-morbid diagnoses, and whether or not they have used medication to treat their symptoms [1]. Imaging data for each individual includes one or more T1-weighted high-resolution s-MRI scan (s-MRI) and one or more rs-fMRI scan. The majority of data was acquired during a single imaging session, although a second session is available for 15 individuals from the Washington University at Saint Louis (WUSTL) site. There is a substantial amount of variation in data acquisition procedures across sites including the type of MRI system and scanning parameters, the length of the rs-fMRI scans, and the instructions given to participants prior to the scan (see Tables 2 and 3).

Nearly all of the imaging data from the ADHD-200 Sample was included in the preprocessing effort, though some individuals were excluded for poor quality or missing data[6]. The results of the preprocessing are made available as a collection of compressed tar files that are organized by pipeline, sites of data collection, training and test samples, as well as by derivatives. A group-level file containing the phenotypic data is available in comma-separated-values format (.csv).

Shared preprocessed data and extracted features include:

[3] `http://fcon_1000.projects.nitrc.org/indi/ADHD-200/`

[4] `http://www.nitrc.org/forum/forum.php?forum_id=2046`

[5] `http://www.nitrc.org/forum/forum.php?forum_id=1735`

[6] Further information regarding excluded data can be found at the respective pipeline wiki page: Athena: `http://www.nitrc.org/plugins/mwiki/index.php/neurobureau:AthenaPipeline#Excluded_Data`; Burner: `http://www.nitrc.org/plugins/mwiki/index.php/neurobureau:BurnerPipeline`; NIAK: `http://www.nitrc.org/plugins/mwiki/index.php/neurobureau:NIAKPipeline`.

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

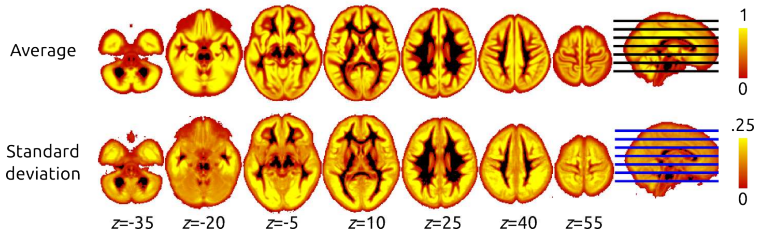

Figure 1: Average and standard deviation of the grey matter density maps generated by the Burner pipeline, for all subjects in the test subsample of ADHD-200 Preprocessed.

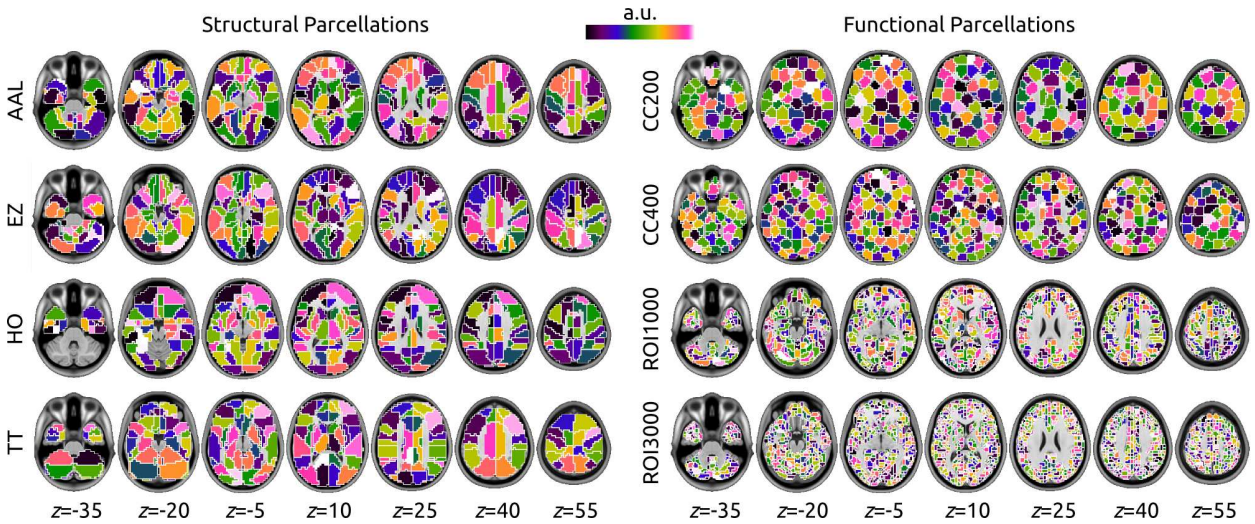

Figure 2: The brain parcellations used to generate regional time series in the NIAK (ROI1000 and ROI3000) and Athena (all other parcellations) pipelines. Each region was randomly assigned to one color in the colormap, and the in-plane outline of regions was painted white at 1 mm resolution.

- 3D grey matter density maps suitable for voxel-based morphometry – Athena and Burner (see Figure 1),
- 4D preprocessed resting-state fMRI data including limited intermediaries and quality assessment – Athena and NIAK,
- Average time series for brain regions from structurally defined parcellations – Athena (see Figures 2 and 3),
- Average time series for brain regions for regions defined by functional parcellation – Athena and NIAK (see Figures 2 and 3),
- Spatial maps for 10 intrinsic connectivity networks (ICNs), fractional amplitude of low frequency fluctuations (fALFF), and regional homogeneity (ReHo) – Athena (see Figure 4).

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

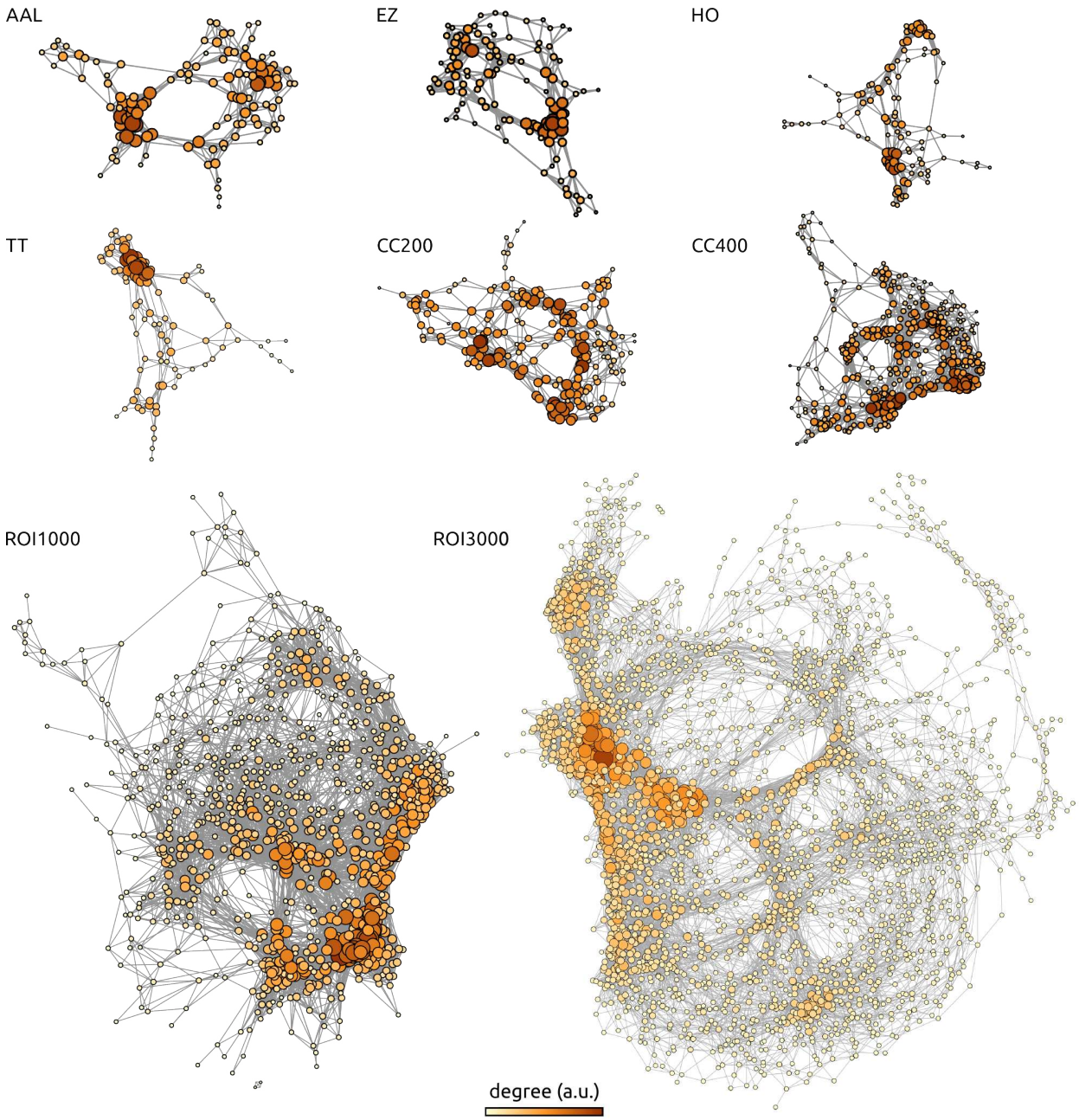

Figure 3: The average functional connectivity matrix (Pearson's correlation coefficient between regional time series) was generated across all individuals of the KKI site, for all parcellations of the release (see text for details). This matrix was further binarized by retaining connections with an average correlation larger than 0.3. The resulting binary adjacency matrices have been represented with an automated layout generated by Yfan Hu's multilevel algorithm, as implemented in the Gephi software [7]. The size and color of each node was set proportional to its degree, relative to the min and max inside the graph.

3.1. Athena Pipeline

The Athena pipeline[7] processed rs-fMRI and s-MRI images using a custom BASH script that combined AFNI [3] and FSL [4] neuroimaging toolkits and was run on the Athena

[7] `http://www.nitrc.org/plugins/mwiki/index.php/neurobureau:AthenaPipeline`

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

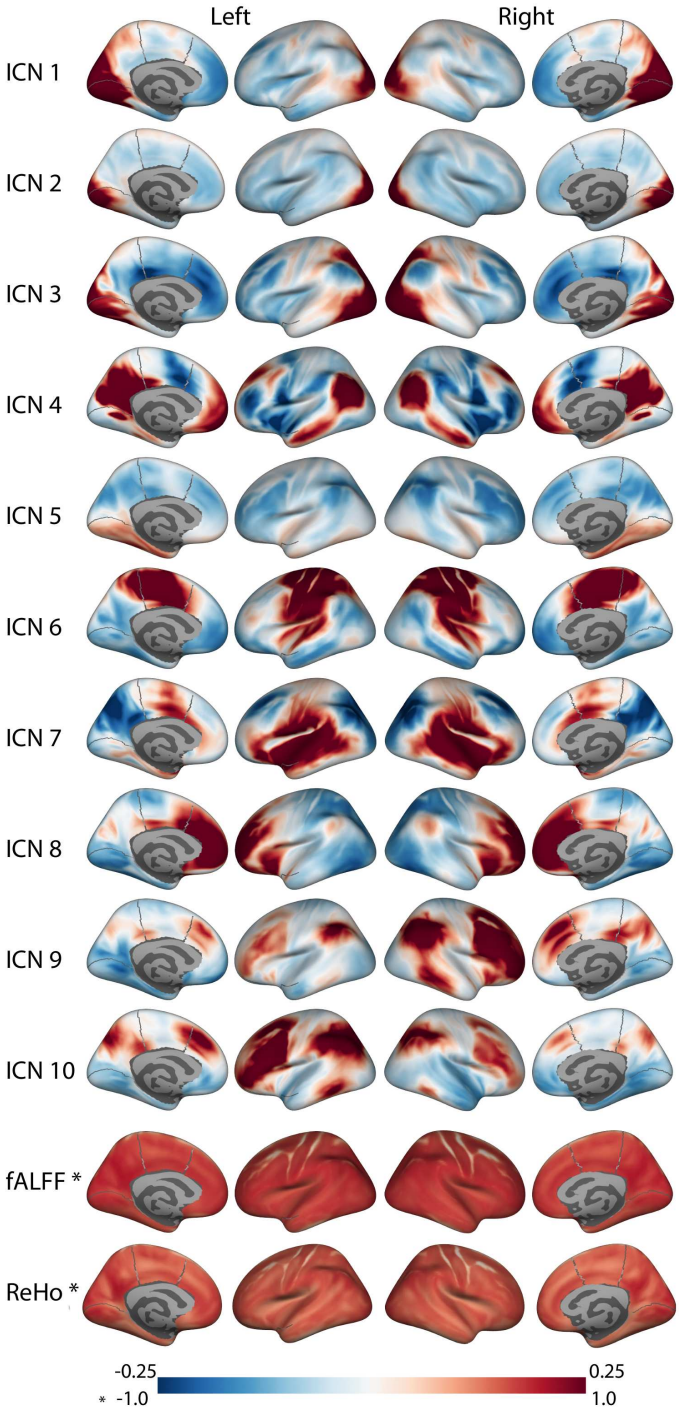

Figure 4: Derivatives from the Athena pipeline including ten intrinsic connectivity networks (ICNs), fractional amplitude of low-frequency fluctuations (fALFF), and regional homogeneity (ReHo). Asterisks (*) in the latter two derivatives denote the difference in colorbar max/min values, as indicated.

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

computer cluster at Virginia Tech's Advanced Research Computing center[8]. The processing scripts for each site are distributed in the repository, as well as on Github[9], along with output log files for each processed dataset.

3.1.1. Structural processing

Athena's s-MRI pipeline began with skull-stripping to remove non-brain tissue and background from the images [8] and segmenting the results into white matter (WM), cerebrospinal fluid (CSF), and grey matter (GM) probability maps [9]. A non-linear warp was calculated between the skull-off image and MNI space as represented by the NIHPD 4.5–18.5y age-specific assymmetric template [10] using a two step procedure that calculates a linear transform [11], that is subsequently refined by a non-linear registration procedure [12]. *Shared s-MRI outputs include:* skull-stripped whole-brain images and smoothed (by a 6 mm FWHM Gaussian) and unsmoothed GM density maps in MNI space at $1\times1\times1$ mm^3 resolution, along with the FSL fNIRT non-linear warp, as compressed NIfTI files (.nii.gz).

3.1.2. Functional processing

Preprocessing. Athena's rs-fMRI pipeline involved removing the first four volumes to allow for magnetization to reach equilibrium, site-specific slice timing correction to the middle slice, re-aligning each volume to the first volume to correct for motion [13], and calculating a linear transform between the mean functional volume and the corresponding s-MRI [11]. The rs-fMRI to s-MRI transform was then combined with the s-MRI to MNI non-linear warp to write the functional data into MNI152 space at $4 \times 4 \times 4$ mm^3 resolution. Mean WM and CSF signals extracted using the masks calculated during s-MRI processing were included along with 6 head motion parameters and a third-order polynomial in voxelwise nuisance regression models to remove variation due to physiological noise, head motion, and scanner drifts from the time series[14, 15]. The resulting denoised time series were band-pass filtered (0.009 Hz $< f < 0.08$ Hz) to limit the data to the frequencies implicated in resting state functional connectivity [16, 17] and then spatially smoothed with a 6 mm FWHM Gaussian filter. *Shared rs-fMRI outputs include:* denoised rs-fMRI volumes, with and without temporal bandpass filtering, in MNI space (compressed 4D NIfTIs, nii.gz), the mean rs-fMRI image and brain mask in template space (.nii.gz), and six parameter head motion traces (tab-separated values, AFNI .1D files).

Time series for structurally defined brain areas. Regional time series were extracted for the automated anatomical labeling (AAL) [18], Eickhoff-Zilles (EZ) [19], Harvard-Oxford (HO) [20–23], and Talairach and Tournoux (TT) [24] parcellations. The EZ parcellation was derived from the max-propagation parcellation distributed with the SPM Anatomy Toolbox[10] and was transformed into template space using the Colin 27 template (also distributed with the toolbox) as an intermediary. The HO parcellation was constructed from 25% thresholded

[8] `http://www.arc.vt.edu/`

[9] `https://github.com/preprocessed-connectomes-project/adhd200_athena_scripts`

[10] `http://www.fz-juelich.de/inm/inm-1/EN/Forschung/_docs/SPMAnatomyToolbox/SPMAnatomyToolbox_node.html`

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

cortical and subcortical max-propagation parcellations distributed with FSL. The parcellations were bisected into left and right hemispheres at the midline ($x = 0$), ROIs representing left/right WM, left/right GM, left/right CSF and brainstem were removed from the subcortical parcellation and then the subcortical and cortical ROIs were combined into a single parcellation. The AAL parcellation distributed with the SPM8 version of the AAL Toolbox[11] and the TT parcellation distributed with AFNI were coregistered and warped into template space. Each of the structural parcellations were resampled into the functional space using nearest-neighbor interpolation. The average time series within each parcel were extracted from both the filtered and unfiltered data, and are distributed in tab-separated values format (AFNI .1D). Each of the conformed ROI parcellations are available as compressed 3D NIfTI files (.nii.gz).

Time series for functionally defined parcellations. The CC200 and CC400 functional brain parcellations were constructed using a two-stage spatially-constrained spectral clustering procedure [25] applied to unfiltered preprocessed rs-fMRI data from a subset ($N = 650$) of the participants in the training dataset. Participants were chosen for inclusion based on registration quality and after excluding participants with more than 3 mm translation or 3 degrees rotations in their motion parameters. To reduce computation time, the clustering was restricted to grey matter using a group GM mask that was constructed by averaging individual GM masks derived from FreeSurfer automated segmentation [26]. Although 200 and 400 ROIs were specified in the functional parcellation procedure, the normalized cut algorithm resulted in 190 and 351 clusters respectfully. Time series were extracted for each parcellation from both the filtered and unfiltered data by averaging the voxel time series contained within each labeled region and are distributed in tab-separated values format (AFNI .1D). CC200 and CC400 brain parcellations are available as compressed 3D NIfTI files (.nii.gz).

ICN time series and spatial maps. Time series and spatial maps were derived for 10 group ICNs generated by [27], which were found to be consistent across resting-state datasets and a variety of neuroimaging tasks. Based on these template ICNs, we applied a modified dual-regression approach [28] to the unfiltered preprocessed data. A spatial multiple regression was first used to extract time series corresponding to each network. In a second step, each time course was independently correlated with whole-brain time series to generate subject-specific functional connectivity maps for each network. Alternatively, all time series were entered simultaneously into a multiple (temporal) regression, and the regression coefficients associated with each time series constituted the functional connectivity maps. The resulting ICN time series are distributed as tab-separated values (AFNI .1D) files and the spatial maps for both temporal regression approaches are distributed as compressed 4D NIfTI files (.nii.gz).

fALFF and ReHo. Whole brain fALFF maps were generated by dividing the variance of each voxel's bandpass-filtered time series by the variance of its unfiltered time series [29].

[11] `http://www.cyceron.fr/index.php/en/plateforme-en/freeware`

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

ReHo was estimated from the unfiltered data at each voxel by the Kendall's Coefficient of Concordance [30] between the voxel and its 26 face-, edge-, and corner- touching neighbors. The resulting fALFF and ReHo whole brain maps are distributed as compressed 3D NIfTI files (.nii.gz).

3.1.3. Quality Control

Images were visually inspected and attempts were made to fix gross misregistration errors by hand adjusting the offending images to center them on the anterior commisure and rotate them into rough correspondence with the MNI template. With the exception of a few datasets that were missing data, and one dataset that was corrupted, all of the ADHD-200 Sample was processed and released by Athena, regardless of data quality. This was done to accommodate differing opinions as to what qualifies as usable data, and to provide poor quality data that may be used by others to develop methods that are robust to noise. Files containing the six motion parameters for each rs-fMRI data and anatomical and mean EPI images in template space have been included in the release to enable users to determine high-motion data or poor registrations.

Additional quality metrics derived from the preprocessed s-MRI, GM masks, mean rs-fMRI images, fALFF maps, and FC maps were also included to help with the QC process. For each data type a mean and standard deviation image was calculated from all of the scans (in stereotaxic space) from all of the subjects. These images were used to perform a voxelwise z-score transformation on each data type for each subject $z_i = \frac{(v_i - m_i)}{\sigma_i}$. For each image (or map), the number of these z-scores whose absolute value exceeded 3 where summed to generate a quality score. The higher the resulting sum, the larger the number of voxels a image has with $|z| \geq 3$ and the more likely the images are outliers. Using this metric it is possible to rank images for a more directed search for poor quality scans. These metrics are distributed in participant-specific text files along with the BASH scripts that implement the procedure.

A rigorous visual inspection of registration has also been performed using the same procedures and raters (PB and YB) as the NIAK pipeline. The procedure entailed generating an online report[12] of registration quality for all subjects. The report includes a visualization of the individual structural images in standard space overlaid onto the corresponding brain template (MNI pediatric), as well as a visualization of the individual mean EPI image overlaid onto the corresponding structural image, both in standard space. These images were visually reviewed, and images with severe misregistraion or quality problems were marked as "fail" (7% failure rate). The outcome of the visual quality assessment along with all available phenotypic variables have been consolidated into a tabular-separated (.tsv) spreadsheet that can be downloaded from NITRC[13].

[12] `http://preprocessed-connectomes-project.org/adhd200_visual_qc_athena/`

[13] `https://www.nitrc.org/frs/download.php/9024/adhd200_preprocessed_phenotypics.tsv`

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

3.2. Burner Pipeline

The Burner pipeline[14] used SPM8 [5] to process s-MRI data for voxel-based morphometry [31] style analyses.

3.2.1. Structural processing

Processing began by segmenting s-MRI images into GM and WM probability maps using SPM8's unified segmentation procedure, which iteratively registers the data to a template and performs tissue classification until both are optimized [32]. Next, SPM8's DARTEL toolbox [33] was used to register the s-MRI of all participants into a common space using an iterative method. Initially, all WM and GM maps were rigidly aligned, and the initial GM and WM templates were created by averaging all aligned maps. Then, all WM and GM maps were non-linearly registered to the templates. New templates were created after each such iteration of registration. The procedure was repeated six times (i.e. template creation and registration) to generate sharper templates and warping all participant WM and GM maps to the template space. The final (6th iteration) non-linear deformations were applied to each participant's GM probability maps to transform them into the space of the population average at $1.5 \times 1.5 \times 1.5$ mm^3 resolution and modulated to conserve the global tissue volumes after normalization. The resulting grey matter density maps are distributed as 3D NIfTI files (.nii).

3.2.2. Quality Control

Stringent quality control was not applied to the data in order to accommodate different opinions on what constitutes poor quality data. Images for four participants were excluded after visual inspection by Dr. Chu because they were determined to be of insufficient quality for further processing.

3.3. NIAK Pipeline

The NIAK[15] is a collection of workflows, implemented in the Pipeline System for Octave and Matlab (PSOM) [34], that perform s-MRI and rs-fMRI processing using a combination of generic medical image processing modules, the MINC tools[16], and custom Matlab/Octave scripts. The ADHD-200 Sample was processed using NIAK version 0.6.4.1, running on a server of the Canadian Brain Imaging Research Platform (CBRAIN) [35]. The NIAK is distributed as an open-source software under MIT license and the code is available on NITRC[17] and Github[18]. The processing scripts for ADHD200 are available on github[19]. The log files for execution were included with the derivatives and can be accessed through the PSOM interface[20].

[14] `http://www.nitrc.org/plugins/mwiki/index.php/neurobureau:BurnerPipeline`
[15] `http://www.nitrc.org/plugins/mwiki/index.php/neurobureau:NIAKPipeline`
[16] `http://en.wikibooks.org/wiki/MINC`
[17] `http://www.nitrc.org/frs/?group_id=411`
[18] `https://github.com/SIMEXP/niak`
[19] `https://github.com/preprocessed-connectomes-project/adhd200_niak_scripts`
[20] `http://psom.simexp-lab.org/how_to_use_psom.html`

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

3.3.1. Structural processing

The NIAK implements a variant of the CIVET pipeline [36]. Each individual s-MRI scan was first corrected for intensity non-uniformities [37] and the brain was extracted using a region growing algorithm [38]. Individual scans were then linearly registered (9 parameters) with the T1 MNI symmetric template [10], restricted to the brain with the previous mask. Note that, by selecting a symmetric template, it is possible to study functional connectivity between homotopic regions by simply flipping the x axis in stereotaxic space, e.g. [39]. The s-MRI scans were again corrected for intensity non-uniformities in stereotaxic space, this time restricted to the template brain mask. An individual brain mask was extracted a second time on this improved image [38] and combined with template priors. An iterative non-linear registration was estimated between the linearly registered s-MRI and the template space, restricted to the brain mask [40]. A final brain mask of the T1 image in native space was extracted from the template brain mask by inverting the linear and non-linear transformation. This final mask was used for registration between rs-fMRI and sMRI data (see below). *Shared s-MRI outputs include:* non-uniformity corrected T1 volumes in native and stereotaxic space (after linear or non-linear transformations) at 1 mm isotropic resolution and brain masks in all spaces, in compressed NIFTI format (.nii.gz), as well as the linear and non-linear transformations from native to template space, as .xfm MINC files.

3.3.2. Functional processing

Preprocessing. The NIAK rs-fMRI pipeline involved removing the first three volumes to allow for magnetization to reach equilibrium, site-specific slice timing correction to the middle slice, and estimating the parameters of a rigid-body motion between each time frame and the median volume of a run, followed by spatial resampling across frames. The fMRI time series were then corrected from slow time drifts (high-pass filter with a 0.01 Hz cut-off, using a discrete cosines transform) and physiological noise using an automated labeling of noise components in an individual independent component analysis, ICA [41]. Finally, the median volume of one selected fMRI run for each subject was coregistered (restricted to the brain) with the corresponding s-MRI scan using Minctracc [40]. The rs-fMRI to s-MRI transform and s-MRI to template (non-linear) transform were combined to resample the rs-fMRI volumes into MNI space at a 3 mm isotropic resolution and the results were spatially smoothed with a 6 mm FWHM Gaussian filter. *Shared rs-fMRI outputs include:* denoised rs-fMRI volumes in MNI space (compressed 4D NIfTIs, nii.gz), the mean / standard deviation rs-fMRI volumes and brain mask in native and template space (.nii.gz), six parameter head motion traces (HDF5 .mat files) as well as individual ICA reports (.pdf).

Time series for functionally defined regions. A region-growing algorithm [42] based on the iterative merging of mutual-nearest-neighbours was implemented to generate functional brain parcellations. The spatial dimension was selected arbitrarily by specifying the size where the growing process should stop, measured in mm^3. Two parameters (1000 mm^3 and 330 mm^3) were selected, resulting in the ROI1000 and ROI3000 parcellations, which include roughly 1000 and 3000 ROIs covering the grey matter, respectively. The region growing was applied on the time series concatenated across all participant's rs-fMRI data (after correction to

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

zero mean and unit variance) from the KKI site (training data only). The homogeneity of regions was thus maximized on average for all subjects, and the regions were identical for all subjects. To limit the amount of memory required by the region-growing procedure, it was applied seperatedly in each of the 116 areas of the AAL template [18]. The average time series for each ROI were extracted for both parcellations and are distributed in individual HDF5 (.mat) files. The ROI1000 and ROI3000 parcellations are also available as compressed 3D NIfTI files (.nii.gz).

3.3.3. Quality control

Outputs of the NIAK pipeline were subjected to a careful visual inspection and the QC reports, along with head motion statistics, are available on the NIAK description page [21]. Estimates of the maximum motion (translation and rotation) between consecutive functional volumes for each rs-fMRI dataset were inspected to categorize the datasets as containing minimal (<1mm or degree), moderate (2 to 3 mm or degrees) or severe motion (>3 mm or degrees). The individual brain registration of the NIAK pipeline were visually inspected using online QC reports[22], similar to those generated for the Athena pipeline. When sub-standard registration outcomes were identified, a parameter controlling the non-uniformity correction of the s-MRI was adjusted and the analysis was repeated until the coregistration results were satisfactory. Satisfactory results could not be achived with some datasets and have been indicated as "Fail" for QC (5.2% failure rate) in the .tsv spreadsheet including QC assessments, available on NITRC[23].

4. Usage recommendations

In line with the original purpose of the ADHD-200 Global competition, most of the publications that have used the ADHD-200 Preprocessed initiative data have proposed new machine learning techniques for predicting ADHD diagnosis or subtype. The most frequently used derivatives have been Athena preprocessed fMRI volumes [e.g. 43, 44] and Athena regional time series [e.g. 45, 46]. Works have used Burner pipeline data both in isolation [e.g. 47–49] or in combination with the Athena functional derivatives [e.g. 50–52]. A fewer number of publications have also used Athena ICN, ReHo and fALFF maps [e.g. 53] and the NIAK high resolution (either ROI1000 or ROI3000) regional time series in their main analysis [e.g. 54, 55]. Researchers have also found value in this initiative beyond the processed data, such as the CC200 or CC400 functional brain parcellations [e.g. 56, 57] and the Athena processing scripts [.e.g 58]. The three PhD dissertations [59–61] and three master's theses [62–64] that have used this resource have all focused on disease state prediction and data dimensionality reduction techniques.

To the best of our knowledge, there are no published comparisons between the results generated by the Athena and NIAK pipeline. The two pipelines conceptually implement

[21] `http://www.nitrc.org/plugins/mwiki/index.php/neurobureau:NIAKPipeline#Quality_control_of_the_preprocessing_-_Training_dataset`

[22] `http://preprocessed-connectomes-project.org/adhd200_visual_qc_niak/`

[23] `https://www.nitrc.org/frs/download.php/9024/adhd200_preprocessed_phenotypics.tsv`

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 1: **ADHD-200 participants by site.** BHBU: Bradley Hospital/ Brown University, KKI: Kennedy Krieger Institute, NI: NeuroIMAGE sample, NYU: New York University Child Study Center, OHSU: Oregon Health Sciences University, PKU: Peking University, Pitt: University of Pittsburgh, WUSTL: Washington University at Saint Louis, avg.: average. *Diagnostic labels are currently not available for BHBU, they have been listed as TDC in the table, but not included in the totals.

		TDC		ADHD	
Site	Sex	N	Age Range (avg.)	N	Age Range (avg.)
BHBU	F	17*	8 - 18 (13.8)	0	-
	M	9*	12 - 18 (16.1)	0	-
KKI	F	28	8 - 12 (10.3)	10	8 - 13 (9.9)
	M	41	8 - 13 (10.4)	15	8 - 13 (10.1)
NI	F	25	12 - 26 (18.8)	5	12 - 20 (15.2)
	M	12	13 - 25 (17.9)	31	11 - 21 (17.1)
NYU	F	55	7 - 18 (12.2)	34	7 - 17 (10.1)
	M	56	7 - 18 (12.0)	117	7 - 18 (11.2)
OHSU	F	40	7 - 12 (9.0)	13	7 - 11 (8.9)
	M	30	7 - 12 (9.5)	30	7 - 12 (8.9)
PKU	F	59	8 - 15 (10.9)	10	9 - 16 (10.9)
	M	84	8 - 15 (11.8)	92	8 - 17 (12.2)
Pitt	F	44	10 - 20 (15.7)	1	15
	M	50	10 - 19 (14.5)	3	14 - 17 (15.7)
WUSTL	F	28	7 - 22 (11.3)	0	-
	M	33	7 - 22 (11.5)	0	-
Totals	F	279*	7 - 26 (12.3)	73	7 - 20 (10.4)
	M	306*	7 - 25 (12.1)	288	7 - 21 (11.9)

very similar steps, with key operations including the non-linear volumetric registration of individual structural MRIs in two different variants of the MNI space, and the registration of individual EPI and T1 images. The most striking differences between the pipelines are in the software that they used and the opinions that drove the parameter choices. The Athena implemented nuisance variable regression [14, 15] (motion parameters, WM and GM signals), while NIAK implemented an automated labeling of structured noise in a spatial ICA [41]. The Athena strategy is fairly standard in the rs-fMRI community, while ICA-based noise attenuation techniques are less common. Athena provided time series for a variety of low-dimension anatomical and functional brain parcellations, whereas NIAK favored much higher-resolution (several thousands) brain parcels. There are also differences in the type of data that each pipeline offers, Athena provides a wide variety of statistical derivatives calculated from both the functional and structural data, whereas Burner provided structural derivatives only, and NIAK primarily focused on function. Beyond personal ideology, the availability of particular features, or allegiances to certain software tools there is no clear reason to prefer one pipeline over the other. It has recently become clear that even subtle

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

variations of processing environments can impact the end-results of processing pipelines [65], even when the same processing is replicated. The variety of data available through this initiative enables researchers to compare the robustness of their tools or analysis results to different processing choices and derivatives.

The ADHD-200 Preprocessed sample also provides ample opportunity for analyses besides the identification of rs-fMRI based biomarkers of ADHD [66]. With 585 TDC participants between the ages of 7 and 26, and the inclusion of intelligence measures, the ADHD-200 Preprocessed is a valuable resource for mapping developmental trajectories [67–69] and other sources of inter-individual variation [70]. Perhaps most exciting are new methods that cluster individuals based on connectivity profiles [71, 72], which are providing new hope for using neuroimaging data to parse the heterogeneity within mental health disorders [73]. One of the outstanding needs for neuroimaging, and connectomics in particular, is the development and validation of new analytical tools and processing strategies [66, 74, 75]. In the service of this aim, the ADHD-200 Preprocessed repository has the necessary components to become a benchmark dataset for evaluating new tools as they are proposed.

The two biggest challenges for using the ADHD-200 Preprocessed data are head motion [76–81] and inter-site variation in the acquisition equipment, parameters, and experimental procedures [82, 83]. A variety of different approaches have been proposed for addressing head motion in hyperkinetic populations [76, 84], and in the ADHD-200 Sample in particular [79], that should be considered when analyzing the data. At the very least, some statistic that characterizes individual motion (such as root mean square deviation [85]) should be included as a nuisance regressor in the group-level model [78, 80]. Differences in the manner in which data was collected at each site can introduce additive and multiplicative effects (batch effects) to the data, which may obscure the underlying biological signal [82, 83]. Including a regressor for acquisition protocol (see Tables 2 and 3 for a summary of the different protocols), the average pairwise correlation between all regions in the brain (GCOR) [86], or the whole-brain average of the feature under inquiry [83], have all been shown to be effective for dealing with inter-site variation.

5. Discussion and conclusions

The ADHD-200 Preprocessed initiative was successful in terms of its primary objectives: the derivatives shared in the repository were effectively used by many researchers during and after the ADHD-200 Global Competition, with over 10,500 downloads by more than 600 users, as well as 52 resulting publications [34, 43–58, 82, 87–120], 3 PhD theses [59–61], 3 master's dissertations [62–64], and 1 patent [121] derived from the release in just over three years. Further publications are either in press or under review, and an updated list of publications will be maintained in a public Mendeley group[24]. Although there was clearly a peak in usage around the ADHD-200 Global competition, there has been a sustained amount of downloads and publications since then, see Figure 5, which we take as a demonstration of a long-term interest from the community in this resource.

[24] `https://www.mendeley.com/groups/4198361/adhd-200-preprocessed/`

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

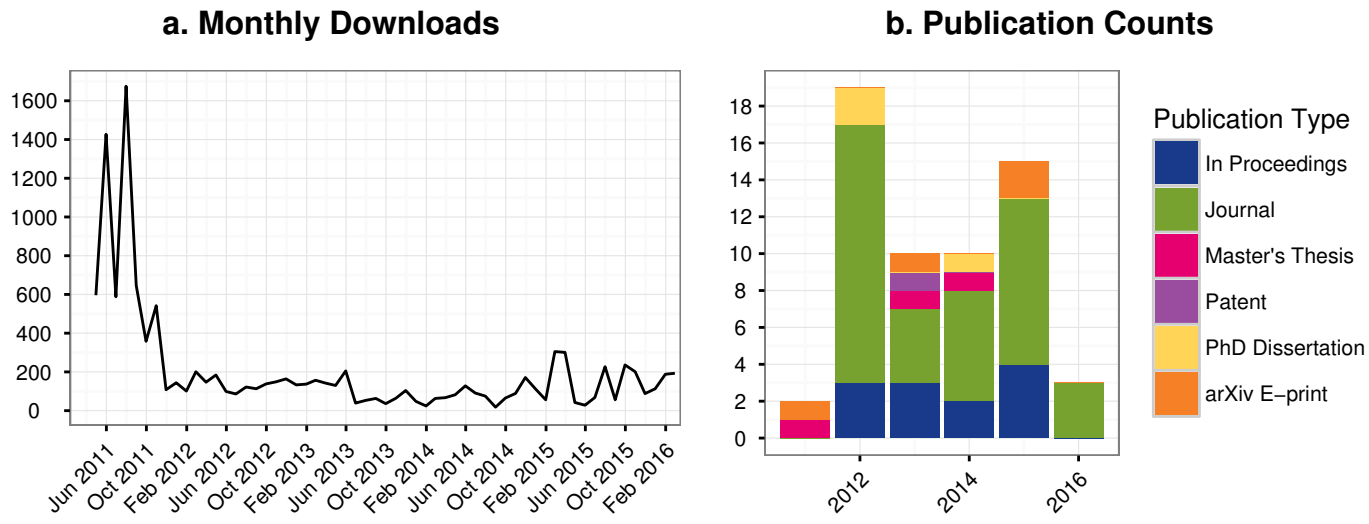

Figure 5: Statistics on download and citations of the ADHD200 preprocessed initiative.

The ADHD-200 preprocessed has helped to expand the boundaries of the traditional neuroimaging community, with several publications in core engineering and statistics journals that do not routinely feature neuroimaging applications, such as Statistica Sinica [55] or the Journal of the American Statistical Association [104], amongst others [e.g. 43, 47–50, 54, 89–91, 99–102, 113, 114]. In particular, the winning team of the ADHD-200 Global Competition was based at the Johns Hopkins Biostatistics Department and used ADHD-200 Preprocessed to develop their diagnostic algorithm [51]. We also found interesting that a handful of methodological publications used the ADHD-200 Preprocessed sample as one application in a series of benchmarks that could be as varied as positron emission tomography in Alzheimer's disease [55], the Yeast gene regulatory network [102] or gene expression in brain tissues of patients with HIV-1 associated neurocognitive disorders [91]. This suggests that fully processed, easy-to-access imaging samples could help validate general-purpose methods on a wider scope of applications.

The impact of the ADHD-200 Preprocessed repository demonstrated the need for reducing computational barriers to participation in discovery neuroscience, including but not limited to machine learning competitions based on neuroimaging data. While the ADHD-200 Preprocessed initiative will have a long-term impact on that need, we believe that a much larger-scale effort will be necessary to unlock the full potential of openly shared neuroimaging data in the service of accelerating neuroscience research. In line with the grassroots, open science ethos of the Neurobureau, new contributors interested in sharing derivatives of ADHD-200 or other open imaging data repository can contact us via our web-based forum[25]. We are actively seeking new contributions to the Preprocessed Connectomes Project, notably

[25] `http://www.nitrc.org/forum/?group_id=383`

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

during Brainhack events[122]. An important current project is the ABIDE preprocessed initiative[26] [69]. This new resource, still under development, will include 16 different processing strategies for rs-fMRI processing, implemented across 4 different software packages, and 2 different pipelines for structural MRI processing. The release will feature an harmonized organization of processed derivatives across packages and extensive quality control. We are planning to continue to expand this line on work on other data sources in the near future. Our hope is that ADHD-200 Preprocessed and future related efforts will critically help fMRI researchers to identify optimal analytical paths for a given task.

Acknowldegements. The authors would like to thank the ADHD-200 Consortium for assembling and sharing the ADHD-200 Sample and for hosting the ADHD-200 Global competition. RCC would like to thank the developers of AFNI and FSL, and the Advanced Research Computing at Virginia Tech for computational support. The scripts used to calculate ReHo were generously provided by Dr. Xinian Zuo. CC would like to thank the developers of SPM. PB would like to thank the developers of MINC tools, the CIVET pipeline, the CBRAIN computational infrastructure, and the computational resources provided by Compute Canada[27] and CLUMEQ[28], which is funded in part by NSERC (MRS), FQRNT, and McGill University.

References

[1] M. P. Milham, D. Fair, M. Mennes, S. H. Mostofsky, The adhd-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience, Frontiers in Systems Neuroscience 6 (62), ISSN 1662-5137, doi:\bibinfo{doi}{10.3389/fnsys.2012.00062}, URL `http://www.frontiersin.org/systems_neuroscience/10.3389/fnsys.2012.00062/full`.

[2] M. Mennes, B. B. Biswal, F. X. Castellanos, M. P. Milham, Making data sharing work: the FCP/INDI experience, Neuroimage 82 (2013) 683–691.

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

[4] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney, et al., Advances in functional and structural MR image analysis and implementation as FSL, Neuroimage 23 (2004) S208–S219.

[5] J. Ashburner, G. Barnes, C. Chen, J. Daunizeau, G. Flandin, K. Friston, D. Gitelman, S. Kiebel, J. Kilner, V. Litvak, et al., SPM8 manual, Functional Imaging Laboratory, Institute of Neurology .

[6] P. Bellec, F. M. Carbonell, V. Perlbarg, C. Lepage, O. Lyttelton, V. Fonov, A. Janke, J. Tohka, A. C. Evans, A neuroimaging analysis kit for Matlab and Octave, in: Proceedings of the 17th International Conference on Functional Mapping of the Human Brain, In Press+, 2011.

[7] M. Bastian, S. Heymann, M. Jacomy, et al., Gephi: an open source software for exploring and manipulating networks., ICWSM 8 (2009) 361–362.

[8] S. M. Smith, Fast robust automated brain extraction, Hum Brain Mapp 17 (3) (2002) 143–155.

[9] Y. Zhang, M. Brady, S. Smith, Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm, IEEE Trans Med Imaging 20 (1) (2001) 45–57.

[10] V. Fonov, A. C. Evans, K. Botteron, C. R. Almli, R. C. McKinstry, D. L. Collins, B. D. C. Group, et al., Unbiased average age-appropriate atlases for pediatric studies, NeuroImage 54 (1) (2011) 313–327.

[26] `http://preprocessed-connectomes-project.org/abide/`
[27] `https://computecanada.org/`
[28] `http://www.clumeq.mcgill.ca/`

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[11] M. Jenkinson, P. Bannister, M. Brady, S. Smith, Improved optimization for the robust and accurate linear registration and motion correction of brain images, Neuroimage 17 (2) (2002) 825–841.

[12] J. L. Andersson, M. Jenkinson, S. Smith, Non-linear registration, aka Spatial normalisation FMRIB technical report TR07JA2, FMRIB Analysis Group of the University of Oxford .

[13] R. W. Cox, A. Jesmanowicz, Real-time 3D image registration for functional MRI, Magn Reson Med 42 (6) (1999) 1014–1018.

[14] T. E. Lund, K. H. Madsen, K. Sidaros, W. L. Luo, T. E. Nichols, Non-white noise in fMRI: does modelling have an impact?, Neuroimage 29 (1) (2006) 54–66.

[15] M. D. Fox, A. Z. Snyder, J. L. Vincent, M. Corbetta, D. C. Van Essen, M. E. Raichle, The human brain is intrinsically organized into dynamic, anticorrelated functional networks, Proc. Natl. Acad. Sci. U.S.A. 102 (27) (2005) 9673–9678.

[16] B. Biswal, F. Z. Yetkin, V. M. Haughton, J. S. Hyde, Functional connectivity in the motor cortex of resting human brain using echo-planar MRI, Magn Reson Med 34 (4) (1995) 537–541.

[17] D. Cordes, V. M. Haughton, K. Arfanakis, J. D. Carew, P. A. Turski, C. H. Moritz, M. A. Quigley, M. E. Meyerand, Frequencies contributing to functional connectivity in the cerebral cortex in "resting-state" data, AJNR Am J Neuroradiol 22 (7) (2001) 1326–1333.

[18] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, M. Joliot, Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain, Neuroimage 15 (1) (2002) 273–289.

[19] S. B. Eickhoff, K. E. Stephan, H. Mohlberg, C. Grefkes, G. R. Fink, K. Amunts, K. Zilles, A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data, Neuroimage 25 (4) (2005) 1325–1335.

[20] J. M. Goldstein, L. J. Seidman, N. Makris, T. Ahern, L. M. O'Brien, V. S. Caviness, D. N. Kennedy, S. V. Faraone, M. T. Tsuang, Hypothalamic abnormalities in schizophrenia: sex effects and genetic vulnerability, Biol. Psychiatry 61 (8) (2007) 935–945.

[21] R. S. Desikan, F. Segonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, R. J. Killiany, An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest, Neuroimage 31 (3) (2006) 968–980.

[22] N. Makris, J. M. Goldstein, D. Kennedy, S. M. Hodge, V. S. Caviness, S. V. Faraone, M. T. Tsuang, L. J. Seidman, Decreased volume of left and total anterior insular lobule in schizophrenia, Schizophr. Res. 83 (2-3) (2006) 155–171.

[23] J. A. Frazier, S. Chiu, J. L. Breeze, N. Makris, N. Lange, D. N. Kennedy, M. R. Herbert, E. K. Bent, V. K. Koneru, M. E. Dieterich, S. M. Hodge, S. L. Rauch, P. E. Grant, B. M. Cohen, L. J. Seidman, V. S. Caviness, J. Biederman, Structural brain magnetic resonance imaging of limbic and thalamic volumes in pediatric bipolar disorder, Am J Psychiatry 162 (7) (2005) 1256–1265.

[24] J. L. Lancaster, M. G. Woldorff, L. M. Parsons, M. Liotti, C. S. Freitas, L. Rainey, P. V. Kochunov, D. Nickerson, S. A. Mikiten, P. T. Fox, Automated Talairach atlas labels for functional brain mapping, Human brain mapping 10 (3) (2000) 120–131.

[25] R. C. Craddock, G. A. James, P. E. Holtzheimer, X. P. Hu, H. S. Mayberg, A whole brain fMRI atlas generated via spatially constrained spectral clustering, Human brain mapping 33 (8) (2012) 1914–1928.

[26] A. M. Dale, B. Fischl, M. I. Sereno, Cortical surface-based analysis. I. Segmentation and surface reconstruction, Neuroimage 9 (2) (1999) 179–194.

[27] S. M. Smith, P. T. Fox, K. L. Miller, D. C. Glahn, P. M. Fox, C. E. Mackay, N. Filippini, K. E. Watkins, R. Toro, A. R. Laird, et al., Correspondence of the brain's functional architecture during activation and rest, Proceedings of the National Academy of Sciences 106 (31) (2009) 13040–13045.

[28] C. Beckmann, C. Mackay, N. Filippini, S. Smith, Group comparison of resting-state {FMRI} data using multi-subject {ICA} and dual regression, NeuroImage 47, Supplement 1 (2009) S148 –, ISSN 1053-8119, doi:\bibinfo{doi}{http://dx.doi.org/10.1016/S1053-8119(09)71511-3}, URL `http://www.sciencedirect.com/science/article/pii/S1053811909715113`, organization for Human

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Brain Mapping 2009 Annual Meeting.

[29] Q. H. Zou, C. Z. Zhu, Y. Yang, X. N. Zuo, X. Y. Long, Q. J. Cao, Y. F. Wang, Y. F. Zang, An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: fractional ALFF, J. Neurosci. Methods 172 (1) (2008) 137–141.

[30] B. B. S. M. G. Kendall, The Problem of m Rankings, The Annals of Mathematical Statistics 10 (3) (1939) 275–287, ISSN 00034851, URL `http://www.jstor.org/stable/2235668`.

[31] J. Ashburner, K. J. Friston, Voxel-based morphometry–the methods, Neuroimage 11 (6 Pt 1) (2000) 805–821.

[32] J. Ashburner, K. J. Friston, Unified segmentation, Neuroimage 26 (3) (2005) 839–851.

[33] J. Ashburner, A fast diffeomorphic image registration algorithm, Neuroimage 38 (1) (2007) 95–113.

[34] P. Bellec, S. Lavoie-Courchesne, P. Dickinson, J. P. Lerch, A. P. Zijdenbos, A. C. Evans, The pipeline system for Octave and Matlab (PSOM): a lightweight scripting framework and execution engine for scientific workflows., Frontiers in neuroinformatics 6, ISSN 1662-5196, doi:\bibinfo{doi}{10.3389/fninf.2012.00007}, URL `http://dx.doi.org/10.3389/fninf.2012.00007`.

[35] T. Sherif, P. Rioux, M.-E. Rousseau, N. Kassis, N. Beck, R. Adalat, S. Das, T. Glatard, A. C. Evans, CBRAIN: a web-based, distributed computing platform for collaborative neuroimaging research, Frontiers in neuroinformatics 8.

[36] Y. Ad-Dabbagh, O. Lyttelton, J. Muehlboeck, C. Lepage, D. Einarson, K. Mok, O. Ivanov, R. Vincent, J. Lerch, E. Fombonne, et al., The CIVET image-processing environment: a fully automated comprehensive pipeline for anatomical neuroimaging research, in: Proceedings of the 12th annual meeting of the organization for human brain mapping, M. Corbetta, S45, 2006.

[37] J. G. Sled, A. P. Zijdenbos, A. C. Evans, A nonparametric method for automatic correction of intensity nonuniformity in MRI data, Medical Imaging, IEEE Transactions on 17 (1) (1998) 87–97.

[38] J. G. Park, C. Lee, Skull stripping based on region growing for magnetic resonance brain images, NeuroImage 47 (4) (2009) 1394–1407.

[39] X.-N. Zuo, C. Kelly, A. Di Martino, M. Mennes, D. S. Margulies, S. Bangaru, R. Grzadzinski, A. C. Evans, Y.-F. Zang, F. X. Castellanos, M. P. Milham, Growing together and growing apart: regional and sex differences in the lifespan developmental trajectories of functional homotopy, J. Neurosci. 30 (45) (2010) 15034–15043.

[40] D. L. Collins, P. Neelin, T. M. Peters, A. C. Evans, Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space., Journal of computer assisted tomography 18 (2) (1994) 192–205.

[41] V. Perlbarg, P. Bellec, J.-L. Anton, M. Pélégrini-Issac, J. Doyon, H. Benali, CORSICA: correction of structured noise in fMRI by automatic identification of ICA components, Magnetic resonance imaging 25 (1) (2007) 35–46.

[42] P. Bellec, V. Perlbarg, S. Jbabdi, M. Pélégrini-Issac, J.-L. Anton, J. Doyon, H. Benali, Identification of large-scale networks in the brain using fMRI, Neuroimage 29 (4) (2006) 1231–1243.

[43] S.-F. Liang, T.-H. Hsieh, P.-T. Chen, M.-L. Wu, C.-C. Kung, C.-Y. Lin, F.-Z. Shaw, Differentiation between resting-state fMRI data from ADHD and normal subjects: Based on functional connectivity and machine learning, in: 2012 International conference on Fuzzy Theory and Its Applications (iFUZZY2012), IEEE, 294–298, URL `http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=6409719`, 2012.

[44] S. Carmona, E. Hoekzema, F. X. Castellanos, D. García-García, A. Lage-Castellanos, K. R. A. Van Dijk, F. J. Navas-Sánchez, K. Martínez, M. Desco, J. Sepulcre, Sensation-to-cognition cortical streams in attention-deficit/hyperactivity disorder., Human brain mapping 36 (7) (2015) 2544–57, ISSN 1097-0193, URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4484811{\&}tool=pmcentrez{\&}rendertype=abstract`.

[45] A. Fujita, D. Y. Takahashi, A. G. Patriota, J. R. Sato, A statistical test to identify differences in clustering structures, ArXiv e-prints (2013) 1311.6732.

[46] S. Dey, A. R. Rao, M. Shah, Attributed graph distance measure for automatic detection of attention deficit hyperactive disordered subjects, Frontiers in Neural Circuits 8, ISSN 1662-5110, URL `http:`

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

//journal.frontiersin.org/Journal/10.3389/fncir.2014.00064/abstract.

[47] B. Rangarajan, S. Suresh, B. S. Mahanand, Identification of Potential Biomarkers in the Hippocampus Region for the Diagnosis of ADHD using PBL-McRBFN Approach, 13th International Conference on Control, Automation, Robotics and Vision, (ICARCV 2014) 2.

[48] B. Rangarajan, K. Subramaian, S. Suresh, Importance of phenotypic information in ADHD diagnosis, 2015 International Conference on Cognitive Computing and Information Processing(CCIP) (MARCH) (2015) 1–6, doi:\bibinfo{doi}{10.1109/CCIP.2015.7100722}, URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7100722.

[49] B. S. Mahanand, R. Savitha, S. Suresh, Computer Aided Diagnosis of ADHD Using Brain Magnetic Resonance Images, in: S. Cranefield, A. Nayak (Eds.), AI 2013: Advances in Artificial Intelligence, vol. 8272 of *Lecture Notes in Computer Science*, Springer International Publishing, Cham, ISBN 978-3-319-03679-3, 386–395, doi:\bibinfo{doi}{10.1007/978-3-319-03680-9}, URL http://link.springer.com/10.1007/978-3-319-03680-9, 2013.

[50] X. Ji, W. Cheng, J. Zhang, T. Ge, L. Sun, Y. Wang, J. Feng, Increased Coupling in the Saliency Network is the main cause/effect of Attention Deficit Hyperactivity Disorder, ArXiv e-prints URL http://arxiv.org/abs/1112.3496.

[51] A. Eloyan, J. Muschelli, M. B. Nebel, H. Liu, F. Han, T. Zhao, A. D. Barber, S. Joel, J. J. Pekar, S. H. Mostofsky, B. Caffo, Automated diagnoses of attention deficit hyperactive disorder using magnetic resonance imaging., Frontiers in systems neuroscience 6 (2012) 61, ISSN 1662-5137, URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3431009{\&}tool=pmcentrez{\&}rendertype=abstract.

[52] E. Olivetti, S. Greiner, P. Avesani, Statistical independence for the evaluation of classifier-based diagnosis, Brain Informatics ISSN 2198-4018, doi:\bibinfo{doi}{10.1007/s40708-014-0007-6}, URL http://link.springer.com/10.1007/s40708-014-0007-6.

[53] J. R. Sato, M. Q. Hoexter, A. Fujita, L. A. Rohde, Evaluation of pattern recognition and feature extraction methods in ADHD prediction., Frontiers in systems neuroscience 6 (2012) 68, ISSN 1662-5137, doi:\bibinfo{doi}{10.3389/fnsys.2012.00068}, URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3449288{\&}tool=pmcentrez{\&}rendertype=abstract.

[54] M. Ahn, H. Shen, W. Lin, H. Zhu, A Sparse Reduced Rank Framework for Group Analysis of Functional Neuroimaging Data, Statistica Sinica (JANUARY), ISSN 10170405, doi:\bibinfo{doi}{10.5705/ss.2013.232w}, URL http://www3.stat.sinica.edu.tw/statistica/J25N1/J25N117/J25N117.html.

[55] S. Yang, Z. Lu, X. Shen, P. Wonka, J. Ye, Fused Multiple Graphical Lasso, SIAM Journal on Optimization 25 (2) (2015) 916–943, ISSN 1052-6234, doi:\bibinfo{doi}{10.1137/130936397}, URL http://epubs.siam.org/doi/abs/10.1137/130936397.

[56] S. Dey, A. R. Rao, M. Shah, Exploiting the brain's network structure in identifying ADHD subjects., Frontiers in systems neuroscience 6 (2012) 75, ISSN 1662-5137, URL http://www.frontiersin.org/Journal/10.3389/fnsys.2012.00075/abstract.

[57] D. Dai, J. Wang, J. Hua, H. He, Classification of ADHD children through multimodal magnetic resonance imaging., Frontiers in systems neuroscience 6 (2012) 63, ISSN 1662-5137, URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3432508{\&}tool=pmcentrez{\&}rendertype=abstract.

[58] A. Fujita, D. Y. Takahashi, J. B. Balardin, J. R. Sato, Correlation between graphs with an application to brain networks analysis, Tech. Rep., URL http://arxiv.org/abs/1512.06830, 2015.

[59] J. B. Colby, Development of human brain connectivity in health and disease, Phd dissertation, University of California at Los Angeles, URL http://escholarship.org/uc/item/2p3471tj{\#}page-2, 2012.

[60] S. Dey, Automatic Detection of Brain Functional Disorder Using Imaging Data, Phd dissertation, University of Central Florida, 2014.

[61] B. Zhang, Dimension Reduction and Classification for High Dimensional Complex Data, Ph.D. thesis, North Carolina State University, URL http://www.lib.ncsu.edu/resolver/1840.16/9022, 2012.

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[62] N. A. van Galen Last, Cortical Parcellation and Classification using PageRank Clustering and the Small-Worldness of ADHD, Master's thesis, Delft University of Technolgy, URL `uuid:f352d7fb-5316-448b-ab58-8aeb3a45e8e5`, 2011.

[63] P. Wang, Machine Learning Approaches for Disease State Classification from Neuroimaging Data, Masters thesis, Auburn University, URL `http://etd.auburn.edu/etd/handle/10415/3623`, 2013.

[64] M. C. Vidal, Análise da estrutura de clusterização das redes de conectividade funcional do cérebro para investigar as bases das desordens do espectro autista, Ph.D. thesis, Universidade de São Paulo, São Paulo, 2014.

[65] T. Glatard, L. B. Lewis, R. Ferreira da Silva, R. Adalat, N. Beck, C. Lepage, P. Rioux, M.-E. Rousseau, T. Sherif, E. Deelman, N. Khalili-Mahani, A. C. Evans, Reproducibility of neuroimaging analyses across operating systems, Front. Neuroinform. 9 (2015) 12.

[66] R. C. Craddock, R. L. Tungaraza, M. P. Milham, Connectomics and new approaches for analyzing human brain functional connectivity, Gigascience 4 (2015) 13.

[67] N. U. Dosenbach, B. Nardos, A. L. Cohen, D. A. Fair, J. D. Power, J. A. Church, S. M. Nelson, G. S. Wig, A. C. Vogel, C. N. Lessov-Schlaggar, K. A. Barnes, J. W. Dubis, E. Feczko, R. S. Coalson, J. R. Pruett, D. M. Barch, S. E. Petersen, B. L. Schlaggar, Prediction of individual brain maturity using fMRI, Science 329 (5997) (2010) 1358–1361.

[68] X. N. Zuo, C. Kelly, A. Di Martino, M. Mennes, D. S. Margulies, S. Bangaru, R. Grzadzinski, A. C. Evans, Y. F. Zang, F. X. Castellanos, M. P. Milham, Growing together and growing apart: regional and sex differences in the lifespan developmental trajectories of functional homotopy, J. Neurosci. 30 (45) (2010) 15034–15043.

[69] A. Di Martino, D. A. Fair, C. Kelly, T. D. Satterthwaite, F. X. Castellanos, M. E. Thomason, R. C. Craddock, B. Luna, B. L. Leventhal, X. N. Zuo, M. P. Milham, Unraveling the miswired connectome: a developmental perspective, Neuron 83 (6) (2014) 1335–1353.

[70] C. Kelly, B. B. Biswal, R. C. Craddock, F. X. Castellanos, M. P. Milham, Characterizing variation in the functional connectome: promise and pitfalls, Trends Cogn. Sci. (Regul. Ed.) 16 (3) (2012) 181–188.

[71] O. Miranda-Dominguez, B. D. Mills, S. D. Carpenter, K. A. Grant, C. D. Kroenke, J. T. Nigg, D. A. Fair, Connectotyping: Model Based Fingerprinting of the Functional Connectome, PLoS ONE 9 (11) (2014) e111048, doi:\bibinfo{doi}{10.1371/journal.pone.0111048}, URL `http://dx.doi.org/10.1371%2Fjournal.pone.0111048`.

[72] E. S. Finn, X. Shen, D. Scheinost, M. D. Rosenberg, J. Huang, M. M. Chun, X. Papademetris, R. T. Constable, Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity, Nat. Neurosci. 18 (11) (2015) 1664–1671.

[73] F. X. Castellanos, A. Di Martino, R. C. Craddock, A. D. Mehta, M. P. Milham, Clinical applications of the functional connectome, Neuroimage 80 (2013) 527–540.

[74] R. C. Craddock, M. P. Milham, S. M. LaConte, Predicting intrinsic brain activity, Neuroimage 82 (2013) 127–136.

[75] G. Varoquaux, R. C. Craddock, Learning and comparing functional connectomes across subjects, Neuroimage 80 (2013) 405–415.

[76] J. D. Power, K. A. Barnes, A. Z. Snyder, B. L. Schlaggar, S. E. Petersen, Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion, Neuroimage 59 (3) (2012) 2142–2154.

[77] K. R. Van Dijk, M. R. Sabuncu, R. L. Buckner, The influence of head motion on intrinsic functional connectivity MRI, Neuroimage 59 (1) (2012) 431–438.

[78] T. D. Satterthwaite, D. H. Wolf, J. Loughead, K. Ruparel, M. A. Elliott, H. Hakonarson, R. C. Gur, R. E. Gur, Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth, Neuroimage 60 (1) (2012) 623–632.

[79] D. A. Fair, J. T. Nigg, S. Iyer, D. Bathula, K. L. Mills, N. U. Dosenbach, B. L. Schlaggar, M. Mennes, D. Gutman, S. Bangaru, J. K. Buitelaar, D. P. Dickstein, A. Di Martino, D. N. Kennedy, C. Kelly, B. Luna, J. B. Schweitzer, K. Velanova, Y. F. Wang, S. Mostofsky, F. X. Castellanos, M. P. Milham, Distinct neural signatures detected for ADHD subtypes after controlling for micro-movements

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

in resting state functional connectivity MRI data, Front Syst Neurosci 6 (2012) 80.

[80] C. G. Yan, B. Cheung, C. Kelly, S. Colcombe, R. C. Craddock, A. Di Martino, Q. Li, X. N. Zuo, F. X. Castellanos, M. P. Milham, A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics, Neuroimage 76 (2013) 183–201.

[81] C. G. Yan, R. C. Craddock, Y. He, M. P. Milham, Addressing head motion dependencies for small-world topologies in functional connectomics, Front Hum Neurosci 7 (2013) 910.

[82] E. Olivetti, S. Greiner, P. Avesani, ADHD diagnosis from multiple data sources with batch effects., Frontiers in systems neuroscience 6 (2012) 70, ISSN 1662-5137, URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3465911{\&}tool=pmcentrez{\&}rendertype=abstract.

[83] C. G. Yan, R. C. Craddock, X. N. Zuo, Y. F. Zang, M. P. Milham, Standardizing the intrinsic brain: towards robust measurement of inter-individual variation in 1000 functional connectomes, Neuroimage 80 (2013) 246–262.

[84] T. D. Satterthwaite, M. A. Elliott, R. T. Gerraty, K. Ruparel, J. Loughead, M. E. Calkins, S. B. Eickhoff, H. Hakonarson, R. C. Gur, R. E. Gur, D. H. Wolf, An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data, Neuroimage 64 (2013) 240–256.

[85] M. Jenkinson, Measuring Transformation Error by RMS Deviation, Internal Technical Report TR99MJ1, Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, Department of Clinical Neurology, Oxford University, Oxford, UK, 1999.

[86] Z. S. Saad, R. C. Reynolds, H. J. Jo, S. J. Gotts, G. Chen, A. Martin, R. W. Cox, Correcting brain-wide correlation differences in resting-state FMRI, Brain Connect 3 (4) (2013) 339–352.

[87] A. Tabas, E. Balaguer-Ballester, L. Igual, Spatial discriminant ICA for RS-fMRI characterisation, in: 2014 International Workshop on Pattern Recognition in Neuroimaging, IEEE, 1–4, URL http://ieeexplore.ieee.org.proxy.wexler.hunter.cuny.edu/articleDetails.jsp?arnumber=6858546, 2014.

[88] M. Lifshitz, D. S. Margulies, A. Raz, Lengthy and Expensive? Why the Future of Diagnostic Neuroimaging May Be Faster, Cheaper, and More Collaborative Than We Think, AJOB Neuroscience 3 (4) (2012) 48–50, ISSN 2150-7740, URL http://www.tandfonline.com/doi/abs/10.1080/21507740.2012.721466?journalCode=uabn20.

[89] L. Li, X. Zhang, Parsimonious Tensor Response Regression, ArXiv e-prints (2015) 1501.07815doi:\bibinfo{doi}{arXiv:1501.07815}.

[90] X. Li, H. Zhou, L. Li, Tucker Tensor Regression and Neuroimaging Analysis, ArXiv e-prints URL http://arxiv.org/abs/1304.5637.

[91] W. Liu, X. Luo, High-dimensional Sparse Precision Matrix Estimation via Sparse Column Inverse Operator, ArXiv e-prints (2012) 1203.3896.

[92] A. dos Santos Siqueira, C. E. Biazoli Junior, W. E. Comfort, L. A. Rohde, J. R. Sato, Abnormal Functional Resting-State Networks in ADHD: Graph Theory and Pattern Recognition Analysis of fMRI Data, BioMed Research International 2014 (2014) 1–10, ISSN 2314-6133, doi:\bibinfo{doi}{10.1155/2014/380531}, URL http://www.hindawi.com/journals/bmri/2014/380531/.

[93] X. Han, Y. Zhong, L. He, P. S. Yu, L. Zhang, The Unsupervised Hierarchical Convolutional Sparse Auto-Encoder for Neuroimaging Data Classification, in: Y. Guo, K. Friston, F. Aldo, S. Hill, H. Peng (Eds.), Brain Informatics and Health, Springer International Publishing, 156–166, doi:\bibinfo{doi}{10.1007/978-3-319-23344-4{_}16}, URL http://link.springer.com/10.1007/978-3-319-23344-4{_}16, 2015.

[94] X. Wang, Y. Jiao, T. Tang, H. Wang, Z. Lu, Altered regional homogeneity patterns in adults with attention-deficit hyperactivity disorder., European journal of radiology 82 (9) (2013) 1552–7, ISSN 1872-7727, URL http://www.sciencedirect.com/science/article/pii/S0720048X13002040.

[95] J. W. Bohland, S. Saperstein, F. Pereira, J. Rapin, L. Grady, Network, anatomical, and non-imaging measures for the prediction of ADHD diagnosis in individual subjects., Frontiers in systems neuroscience 6 (2012) 78, ISSN 1662-5137, URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3527894{\&}tool=pmcentrez{\&}rendertype=abstract.

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[96] C.-W. Chang, C.-C. Ho, J.-H. Chen, ADHD classification by a texture analysis of anatomical brain MRI data., Frontiers in systems neuroscience 6 (2012) 66, ISSN 1662-5137, URL http://www.frontiersin.org/Journal/10.3389/fnsys.2012.00066/abstract.

[97] W. Cheng, X. Ji, J. Zhang, J. Feng, Individual classification of ADHD patients by integrating multi-scale neuroimaging markers and advanced pattern recognition techniques., Frontiers in systems neuroscience 6 (2012) 58, ISSN 1662-5137, URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3412279{\&}tool=pmcentrez{\&}rendertype=abstract.

[98] J. B. Colby, J. D. Rudie, J. A. Brown, P. K. Douglas, M. S. Cohen, Z. Shehzad, Insights into multimodal imaging classification of ADHD., Frontiers in systems neuroscience 6 (2012) 59, ISSN 1662-5137, URL http://www.frontiersin.org/Journal/10.3389/fnsys.2012.00059/abstract.

[99] K. Subramanian, A. K. Das, S. Sundaram, S. Ramasamy, A meta-cognitive interval type-2 fuzzy inference system and its projection based learning algorithm, Evolving Systems ISSN 1868-6478, doi:\bibinfo{doi}{10.1007/s12530-013-9102-9}, URL http://link.springer.com/10.1007/s12530-013-9102-9.

[100] M. Hou, B. Chaib-draa, HIERARCHICAL TUCKER TENSOR REGRESSION : APPLICATION TO BRAIN IMAGING DATA ANALYSIS, in: IEEE International Conference on Image Processing (ICIP '15), Quebec, Canada, 2015.

[101] G. Deshpande, P. Wang, D. Rangaprakash, B. Wilamowski, Fully Connected Cascade Artificial Neural Network Architecture for Attention Deficit Hyperactivity Disorder Classification From Functional Magnetic Resonance Imaging Data, IEEE Transactions on Cybernetics 45 (12) (2015) 2668–2679, ISSN 2168-2267, doi:\bibinfo{doi}{10.1109/TCYB.2014.2379621}, URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7001645.

[102] Y. She, Y. He, D. Wu, Learning Topology and Dynamics of Large Recurrent Neural Networks, IEEE Transactions on Signal Processing 62 (22) (2014) 5881–5891, ISSN 1053-587X, doi:\bibinfo{doi}{10.1109/TSP.2014.2358956}, URL http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=6914572.

[103] S. Lavoie-Courchesne, P. Rioux, F. Chouinard-Decorte, T. Sherif, M. E. Rousseau, S. Das, R. Adalat, J. Doyon, C. Craddock, D. Margulies, C. Chu, O. Lyttelton, A. C. Evans, P. Bellec, Integration of a neuroimaging processing pipeline into a pan-canadian computing grid, Journal of Physics: Conference Series 341 (1) (2012) 012032+, URL http://dx.doi.org/10.1088/1742-6596/341/1/012032.

[104] Z. Chen, C. Leng, Dynamic Covariance Models, Journal of the American Statistical Association (2015) 1–55ISSN 0162-1459, doi:\bibinfo{doi}{10.1080/01621459.2015.1077712}, URL http://www.tandfonline.com/doi/abs/10.1080/01621459.2015.1077712http://www.tandfonline.com/doi/full/10.1080/01621459.2015.1077712.

[105] M. Nuñez-Garcia, S. Simpraga, M. A. Jurado, M. Garolera, R. Pueyo, L. Igual, FADR: Functional-Anatomical Discriminative Regions for Rest fMRI Characterization, in: L. Zhou, L. Wang, Q. Wang, Y. Shi (Eds.), Machine Learning in Medical Imaging, Springer International Publishing, 61–68, doi:\bibinfo{doi}{10.1007/978-3-319-24888-2{_}8}, URL http://link.springer.com/10.1007/978-3-319-24888-2{_}8, 2015.

[106] B. Solmaz, S. Dey, A. R. Rao, M. Shah, ADHD classification using bag of words approach on network features, in: D. R. Haynor, S. Ourselin (Eds.), Medical Imaging 2012: Image Processing. Edited by Haynor, vol. 8314, 83144T, URL http://adsabs.harvard.edu/abs/2012SPIE.8314E.164S, 2012.

[107] A. Anderson, P. K. Douglas, W. T. Kerr, V. S. Haynes, A. L. Yuille, J. Xie, Y. N. Wu, J. a. Brown, M. S. Cohen, Non-negative matrix factorization of multimodal MRI, fMRI and phenotypic data reveals differential changes in default mode subnetworks in ADHD, NeuroImage ISSN 10538119, doi:\bibinfo{doi}{10.1016/j.neuroimage.2013.12.015}, URL http://dx.doi.org/10.1016/j.neuroimage.2013.12.015.

[108] S. K. Bakhtiari, G.-A. Hossein-Zadeh, Subspace-based Identification Algorithm for characterizing causal networks in resting brain., NeuroImage 60 (2) (2012) 1236–49, ISSN 1095-9572, URL http://www.sciencedirect.com/science/article/pii/S105381191200016X.

[109] J. R. Sato, D. Y. Takahashi, M. Q. Hoexter, K. B. Massirer, A. Fujita, Measuring network's entropy

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

in ADHD: a new approach to investigate neuropsychiatric disorders., NeuroImage 77 (2013) 44–51, ISSN 1095-9572, URL http://www.ncbi.nlm.nih.gov/pubmed/23571416.

[110] S. Kyeong, S. Park, K.-A. Cheon, J.-J. Kim, D.-H. Song, E. Kim, A New Approach to Investigate the Association between Brain Functional Connectivity and Disease Characteristics of Attention-Deficit/Hyperactivity Disorder: Topological Neuroimaging Data Analysis., PloS one 10 (9) (2015) e0137296, ISSN 1932-6203, URL http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0137296.

[111] J. R. Sato, M. Q. Hoexter, X. F. Castellanos, L. A. Rohde, Abnormal brain connectivity patterns in adults with ADHD: a coherence study., PloS one 7 (9) (2012) e45671, ISSN 1932-6203, doi:\bibinfo{doi}{10.1371/journal.pone.0045671}, URL http://dx.plos.org/10.1371/journal.pone.0045671.

[112] D. Y. Takahashi, J. R. Sato, C. E. Ferreira, A. Fujita, Discriminating different classes of biological networks by analyzing the graphs spectra distribution., PloS one 7 (12) (2012) e49949, ISSN 1932-6203, URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3526608{\&}tool=pmcentrez{\&}rendertype=abstract.

[113] L. He, X. Kong, P. S. Yu, A. B. Ragin, Z. Hao, X. Yang, DuSK: A Dual Structure-preserving Kernel for Supervised Tensor Learning with Applications to Neuroimages, in: Proc of the Thirteenth SIAM International Conference on Data Mining (SDM 2013), Philadelphia, PA, 127–135, doi:\bibinfo{doi}{http://epubs.siam.org/doi/abs/10.1137/1.9781611973440.15}, URL http://epubs.siam.org/doi/abs/10.1137/1.9781611973440.15, 2013.

[114] X. Kong, P. S. Yu, X. Wang, A. B. Ragin, Discriminative Feature Selection for Uncertain Graph Classification, in: Proc of the Thirteenth SIAM International Conference on Data Mining (SDM 2013), Philadelphia, PA, URL http://arxiv.org/abs/1301.6626, 2013.

[115] Y. Yao, W. L. Lu, B. Xu, C. B. Li, C. P. Lin, D. Waxman, J. F. Feng, The increase of the functional entropy of the human brain with age., Scientific reports 3 (2013) 2853, ISSN 2045-2322, URL http://www.nature.com/srep/2013/131009/srep02853/full/srep02853.html.

[116] A. Fujita, D. Y. Takahashi, A. G. Patriota, J. R. Sato, A non-parametric statistical test to compare clusters with applications in functional magnetic resonance imaging data., Statistics in medicine ISSN 1097-0258, URL http://www.ncbi.nlm.nih.gov/pubmed/25185759.

[117] P. T. Reiss, L. Huo, Y. Zhao, C. Kelly, R. T. Ogden, Wavelet-domain Regression and Predictive Inference in Psychiatric Neuroimaging, The SelectedWorks of Philip T. Reiss URL http://works.bepress.com/phil{_}reiss/29.

[118] M. Nachamai, Sub-Type Discernment of Attention Deficit Hyperactive Disorder in Children using a Cluster Partitioning Algorithm, Indian Journal of Science and Technology 9 (8).

[119] J. A. Brown, J. D. V. Horn, Connected brains and mindsThe {UMCD} repository for brain connectivity matrices, NeuroImage 124, Part B (2016) 1238 – 1241, ISSN 1053-8119, doi:\bibinfo{doi}{http://dx.doi.org/10.1016/j.neuroimage.2015.08.043}, URL http://www.sciencedirect.com/science/article/pii/S1053811915007624, sharing the wealth: Brain Imaging Repositories in 2015.

[120] D. Yu, L. Kong, I. Mizera, Partial functional linear quantile regression for neuroimaging data analysis, Neurocomputing (2016) –ISSN 0925-2312, doi:\bibinfo{doi}{http://dx.doi.org/10.1016/j.neucom.2015.08.116}, URL http://www.sciencedirect.com/science/article/pii/S0925231216001181.

[121] S. Dey, R. Rao, M. Shah, B. Solmaz, Method and system for modeling and processing fmri image data using a bag-of-words approach, URL http://www.google.com/patents/US20130211229, 2013.

[122] R. Cameron Craddock, D. S Margulies, P. Bellec, B. Nolan Nichols, S. Alcauter, F. A Barrios, Y. Burnod, C. J Cannistraci, J. Cohen-Adad, B. De Leener, S. Dery, J. Downar, K. Dunlop, A. R Franco, C. Seligman Froehlich, A. J Gerber, S. S Ghosh, T. J Grabowski, S. Hill, A. Sólon Heinsfeld, R. Matthew Hutchison, P. Kundu, A. R Laird, S.-L. Liew, D. J Lurie, D. G McLaren, F. Meneguzzi, M. Mennes, S. Mesmoudi, D. O'Connor, E. H Pasaye, S. Peltier, J.-B. Poline, G. Prasad, R. Fraga Pereira, P.-O. Quirion, A. Rokem, Z. S Saad, Y. Shi, S. C Strother, R. Toro, L. Q Uddin, J. D Van Horn, J. W Van Meter, R. C Welsh, T. Xu, Brainhack: a collaborative workshop for the open neuroscience community, Gigascience 5 (2016) 16.

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 2: **Structural MRI acquisition parameters by site.** Seq: imaging sequence, FA: flip angle, TE: echo time, TR: repetition time, TI: inversion recovery delay, PA: parallel acquisition, Res: voxel resolution, BHBU: Bradley Hospital/ Brown University, KKI: Kennedy Krieger Institute, NI: NeuroIMAGE sample, NYU: New York University Child Study Center, OHSU: Oregon Health Sciences University, PKU: Peking University, Pitt: University of Pittsburgh, WUSTL: Washington University at Saint Louis, Trio: Siemens TIM Trio 3T, Allegra: Siemens Allegra, Avanto: Siemens Avanto, MPRAGE: magnetization prepared rapid gradient echo, S: sensitivity encoding (SENSE), G: generalized auto-calibrating partially parallel acquisition (GRAPPA)

Site	Scanner	Seq	FA	TE	TR	TI	PA	Res.
BHBU	Trio 3T	3D MPRAGE	9°	2.98 ms	2250 ms	900 ms	None	$1.00 \times 1.00 \times 1.00$ mm^3
KKI	Phillips 3T	3D MPRAGE	8°	3.7 ms	3500 ms	1000 ms	S $\times 2$	$1.00 \times 1.00 \times 1.00$ mm^3
NI	Avanto 1.5T	3D MPRAGE	7°	2.95 ms	2730 ms	1000 ms	G $\times 2$	$1.00 \times 1.00 \times 1.00$ mm^3
NYU	Allegra 3T	3D MPRAGE	7°	3.25 ms	2530 ms	1100 ms	None	$1.30 \times 1.00 \times 1.30$ mm^3
OHSU	Trio 3T	3D MPRAGE	10°	3.58 ms	2300 ms	900 ms	None	$1.00 \times 1.00 \times 1.10$ mm^3
PKU 1	Trio 3T	3D MPRAGE	7°	3.39 ms	2530 ms	1100 ms	None	$1.30 \times 1.00 \times 1.30$ mm^3
PKU 2	Trio 3T	3D MPRAGE	7°	3.45 ms	2530 ms	1100 ms	None	$1.00 \times 1.00 \times 1.00$ mm^3
PKU 3 (1)	Trio 3T	3D MPRAGE	12°	3.67 ms	2000 ms	1100 ms	None	$0.94 \times 0.94 \times 1.00$ mm^3
PKU 3 (2)	Trio 3T	3D MPRAGE	10°	2.60 ms	1950 ms	900 ms	None	$1.00 \times 1.00 \times 1.30$ mm^3
PKU 3 (3)	Trio 3T	3D MPRAGE	7°	3.37 ms	2530 ms	1100 ms	None	$1.00 \times 1.00 \times 1.33$ mm^3
PKU 3 (4)	Trio 3T	3D MPRAGE	12°	3.92 ms	1770 ms	1100 ms	None	$0.50 \times 0.50 \times 1.00$ mm^3
PKU 3 (5)	Trio 3T	3D MPRAGE	8°	2.89 ms	845 ms	600 ms	None	$1.02 \times 1.02 \times 1.30$ mm^3
Pitt	Trio 3T	3D MPRAGE	8°	3.43 ms	2100 ms	1050 ms	None	$1.00 \times 1.00 \times 1.00$ mm^3
WUSTL	Trio 3T	3D MPRAGE	8°	3.08 ms	2400 ms	1000 ms	G $\times 2$	$1.00 \times 1.00 \times 1.00$ mm^3

Table 3: **Resting state fMRI acquisition parameters by site.** Seq: imaging sequence, FA: flip angle, TE: echo time, TR: repetition time, PA: parallel acquisition, N_{slc}: number of slices, Th.: slice thickness, Slc. Acq.: slice acquisition order, N_{TR}: number of measurements (TRs), BHBU: Bradley Hospital/ Brown University, KKI: Kennedy Krieger Institute, NI: NeuroIMAGE sample, NYU: New York University Child Study Center, OHSU: Oregon Health Sciences University, PKU: Peking University, Pitt: University of Pittsburgh, Pitt 2: U. Pitt. parameters used for acquiring the testing data, WUSTL: Washington University at Saint Louis, EPI: echo planar imaging, PACE: Prospective Acquisition CorrEction (EPI with prospective motion correction), S: sensitivity encoding (SENSE), G: generalized autocalibrating partially parallel acquisition (GRAPPA), int+: slices were acquired interleaved ascending, seq+: slices were acquired sequentially ascending, var.: the number of measurements varies across datasets, fixate: participants were asked to keep their eyes open and fixate on an image, closed: participants were asked to keep their eyes closed, open: participants were asked to keep their eyes open.

Site	Seq	FA	TE	TR	PA	N_{slc}	Th.	Slc. Acq.	Resolution	N_{TR}	Instructions
BHBU	PACE	90°	25 ms	2000 ms	None	35	3 mm	int+	3.0×3.0 mm^2	256	fixate
KKI	EPI	75°	30 ms	2500 ms	S ×3	47	3 mm	seq+	3.0×3.0 mm^2	128	fixate
NI	EPI	80°	40 ms	1960 ms	G ×2	37	3 mm	int+	3.5×3.5 mm^2	266	eyes closed
NYU	EPI	90°	15 ms	2000 ms	None	33	4 mm	int+	3.0×3.0 mm^2	180	eyes closed
OHSU	EPI	90°	30 ms	2500 ms	None	36	3.8 mm	int+	3.8×3.8 mm^2	82	fixate
PKU 1	EPI	90°	30 ms	2000 ms	None	33	4.2 mm	int+	3.1×3.1 mm^2	240	closed or fixate
PKU 2	EPI	90°	30 ms	2000 ms	None	33	3.6 mm	int+	3.1×3.1 mm^2	240	closed or fixate
PKU 3	EPI	90°	30 ms	2000 ms	None	30	4.5 mm	int+	3.44×3.44 mm^2	240	closed or fixate
Pitt	EPI	70°	29 ms	1500 ms	G ×2	29	4.0 mm	int+	3.1×3.1 mm^2	200	open or closed
Pitt 2	EPI	90°	30 ms	3000 ms	None	46	3.5 mm	int+	3.8×3.8 mm^2	128	open or closed
WUSTL	EPI	90°	27 ms	2500 ms	None	32	4.0 mm	int+	4.0×4.0 mm^2	var.	fixate

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/037044; this version posted June 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 4: **Summary of the characteristics of all brain parcellations used in the ADHD-200 Preprocessed release.** Sizes for the parcels are reported in mm^3.

Name	Type	# parcels	mean size	std size	min size	max size
AAL	structural	116	16726	11896	768	55552
EZ	structural	116	15880	11059	1344	52608
HO	structural	111	14540	15342	64	99200
TT	structural	97	17106	16164	64	70400
CC200	functional	190	11351	2001	2880	17856
CC400	functional	351	6144	1207	64	10048
ROI1000	functional	954	1404	366	27	2781
ROI3000	functional	2843	471	109	27	1026