ciftify: A framework for surface-based analysis of legacy MR acquisitions

Erin W. Dickie[1], Alan Anticevic[2,3,4,5] Dawn E. Smith[1], Timothy S. Coalson[6], Mathuvanthi Manogaran[1], Navona Calarco[1], Joseph D. Viviano[1], Matthew F. Glasser[6,7], David C. Van Essen[6], Aristotle N. Voineskos[1,8]

1. Kimel Family Translational Imaging-Genetics Laboratory, Campbell Family Mental Health Research Institute, Centre of Addiction and Mental Health, Toronto Canada
2. Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut, USA
3. Division of Neurocognition, Neurocomputation, & Neurogenetics (N3), Yale University School of Medicine, New Haven, Connecticut, USA
4. Interdepartmental Neuroscience Program, Yale University, New Haven, Connecticut, USA;
5. Department of Psychology, Yale University, New Haven, Connecticut, USA
6. Departments of Radiology and Neuroscience, Washington University School of Medicine, St Louis, USA
7. St. Luke's Hospital, Chesterfield MO USA
8. Department of Psychiatry, University of Toronto, Toronto, Canada

HIGHLIGHTS

- the ciftify package allows for grayordinate-based (CIFTI format) analysis of non-Human Connectome Project (i.e. legacy) MR acquisitions
- The workflow and dependencies are distributed as a Docker container, following the BIDS-app interface
- Additional ciftify utilities aid in downstream analysis of CIFTI images
- We intend for this work to offer bridging solution for legacy data that will allow many researchers to adopt CIFTI format analyses

ABSTRACT

The preprocessing pipelines of the Human Connectome Project (HCP) were made publicly available for the neuroimaging community to apply the HCP analytic approach to data from non-HCP sources. The HCP analytic approach is surface-based for the cerebral cortex, uses the CIFTI "grayordinate" file format, provides greater statistical sensitivity than traditional volume-based analysis approaches, and allows for a more neuroanatomically-faithful representation of data. However, the HCP pipelines require the acquisition of specific images (namely T2w and field map) that historically have often not been acquired. Massive amounts of this 'legacy' data could benefit from the adoption of HCP-style methods. However, there is currently no published framework, to our knowledge, for adapting HCP preprocessing to "legacy" data. Here we present the ciftify project, a parsimonious analytic framework for adapting key modules from the HCP pipeline into existing structural workflows using FreeSurfer's recon_all structural and existing functional preprocessing workflows. Within this framework, any functional dataset with an accompanying (i.e. T1w) anatomical data can be analyzed in CIFTI format. To simplify usage for new data, the workflow has been bundled with

fMRIPrep following the BIDS-app framework. Finally, we present the package and comment on future neuroinformatics advances that may accelerate the movement to a CIFTI-based grayordinate framework.

1.0 Introduction

The Human Connectome Project (HCP) has incorporated major technical advances at many steps of neuroimaging data acquisition and analysis (Glasser et al., 2016b; Van Essen et al., 2013). At the level of MR acquisition, the HCP used multi-band MR pulse sequences, which increased both the temporal and spatial resolution of MR data (Uğurbil et al., 2013; Van Essen et al., 2012). In addition, the HCP project took great care to utilize state-of-the-art approaches to correct for MR field bias and image distortions apparent across MR modalities (Glasser et al., 2016b). At the level of MR analysis, as part of the minimal preprocessing pipelines, the HCP introduced a novel Connectivity Informatics Technology Initiative file format (CIFTI; https://www.nitrc.org/projects/cifti/) for conducting analyses in the “grayordinate” framework. In CIFTI format, data from cerebral cortical gray matter is stored in relation to 2-dimensional surface meshes, whereas subcortical data is maintained in 3-dimensions, within the same file, by representing only subcortical gray matter voxels. The HCP also introduced a powerful visualization tool, Connectome Workbench, with an accompanying suite of command-line functions that allow for the manipulation of CIFTI, GIFTI and NIFTI format images (Marcus et al., 2013).

A 2D cortical surface-based approach to MR analysis of cortical signals provides several advantages over the more commonly used 3D volume-based approach as previously described (Argall et al., 2006; Fischl, 2012; Goebel et al., 2006; Zijdenbos et al., 2002). Most notable benefits include better adherence to the inherent geometry of cortical surfaces, increased statistical power (Anticevic et al., 2008; Argall et al., 2006; Coalson et al., 2018; Fischl et al., 1999; Jo et al., 2007; Tucholka et al., 2012), removal of the deleterious effects of volume-based smoothing (which markedly erodes spatial localization (Coalson et al., 2018), superior visualization (Van Essen, 2012), and finally a simplified, more compact framework for multimodal analysis (such as analyses combining surface-based anatomical features, e.g., cortical thickness, and fMRI). If should be noted that these benefits have been reported using data with lower spatial and temporal resolutions than the HCP. For these reasons, surface-based analyses have provided new insights into how the human cortex is functionally organized within humans at both the population level (Glasser et al., 2016a; Margulies et al., 2016; Mueller et al., 2013; Van Essen and Glasser, 2018; Yeo et al., 2011) and individual level (Braga and Buckner, 2017; Glasser et al., 2016a; Wang et al., 2015).

Although the benefits of surface-based over volume-based registration have been widely recognized for nearly a decade and despite the continued development and promulgation of surface-based tools by several groups (Argall et al., 2006; Fischl, 2012; Goebel et al., 2006; Van Essen, 2012; Zijdenbos et al., 2002), the adoption of surface-based analyses by the neuroimaging community has been slow. One reason for this inertia is that many widely used

tools for MR preprocessing (e.g., SPM, FSL) as well as widely adopted pipeline tools (e.g., NIAK, CPAC) do not offer “out of the box” workflows with surface-based registration steps. These tools are not designed to support the CIFTI format, or provide visualization frameworks for combined 2D/3D geometry. The HCP consortium released its pipelines as part of the Minimal Preprocessing Pipeline GitHub project (Glasser et al., 2013). One key requirement of the Minimal Preprocessing Pipeline is acquisition of a high-resolution T2-weighted image, used to generate high-quality surface reconstructions (Glasser et al., 2013) and myelin maps (Glasser et al., 2014; Glasser and Van Essen, 2011), that was not typically acquired in legacy human MR protocols. Therefore, many ‘legacy’ acquisitions often collected without high-resolution T2 images or fieldmaps cannot be processed using the HCP’s pipelines.

It is estimated that tens of thousands of participants are scanned annually as part of research studies, and this number has been growing for over 25 years (Smith, 2012). Thanks to data sharing consortia such as the International Neuroimaging Data-Sharing Initiative (INDI, (Milham et al., 2018)) and OpenfMRI (Poldrack et al., 2013), thousands of legacy datasets are publicly available (Eickhoff et al., 2016; Poldrack and Gorgolewski, 2014). The National Institute of Mental Health (NIMH) have also committed to data sharing via the NIMH Data Archive (NDA; ndar.nih.gov). These legacy datasets are drawn from healthy individuals plus a wide variety of clinical populations and developmental stages. While some clinical and developmental populations are now being scanned according to HCP acquisition standards, it may be years before these sample sizes will be large enough to answer some of today’s most pressing questions. The HCP requirements can be viewed either as a barrier or as an opportunity. If viewed as an opportunity, there is a concomitant need to develop tools for HCP-style analyses applicable to legacy MR data available today (and likely the near future) Therefore, maximally leveraging large legacy datasets is important for clinical research, especially for those attempting to characterize disease heterogeneity (Choudhury et al., 2014).

Here, we address the opportunity to leverage important innovations of the HCP pipelines to enable integration of two decades of existing legacy human neuroimaging data into the CIFTI grayordinate-based framework. To expand the utility of HCP-style methods, we present the ciftify package for grayordinate-based (CIFTI format) analysis of legacy acquisitions that have already been processed using FreeSurfer. (Datasets that use other cortical segmentation methods or have no surface-based processing at all are considered in the discussion.) Ciftify translates two key modules of the HCP Minimal Preprocessing Pipeline: the FreeSurfer-to-Connectome Workbench conversion, and the fMRI surface projection, into simple command line tools. Integrating these two steps will work from FreeSurfer-based anatomical outputs (generated from only T1w images) to convert current volume-based fMRI analysis pipelines into a grayordinate-based one. These tools allow researchers to move their analyses to the surface while ensuring the opportunity to analyze non-HCP quality data. Below, we describe the ciftify package and discuss its expected use case. We also introduce additional tools for running and interpreting group-level analyses in CIFTI format.

2.0 The ciftify preprocessing workflow and BIDS-app

A diagram of the ciftify preprocessing workflow is given in Figure 1. As a precursor to ciftify, surfaces are generated from T1w anatomical images using FreeSurfer's recon_all function (Fischl, 2012), and fMRI runs are preprocessed using other software as discussed below. The BIDS-app Docker container will use FMRIPrep (Esteban et al., 2019a, 2019b) for these preprocessing steps if they have not already been run. Anatomical data is converted from FreeSurfer to CIFTI formats, and MNI inter-subject anatomy-based registration and resampling is performed by the ciftify_recon_all function (using FSL's FNIRT). Next, fMRI acquisitions are projected to the surface, and subcortical data are resampled by the cifitfy_subject_fmri function. Quality assurance visualizations for these steps are generated with the cifti_vis_recon_all and cifti_vis_fmri utilities.

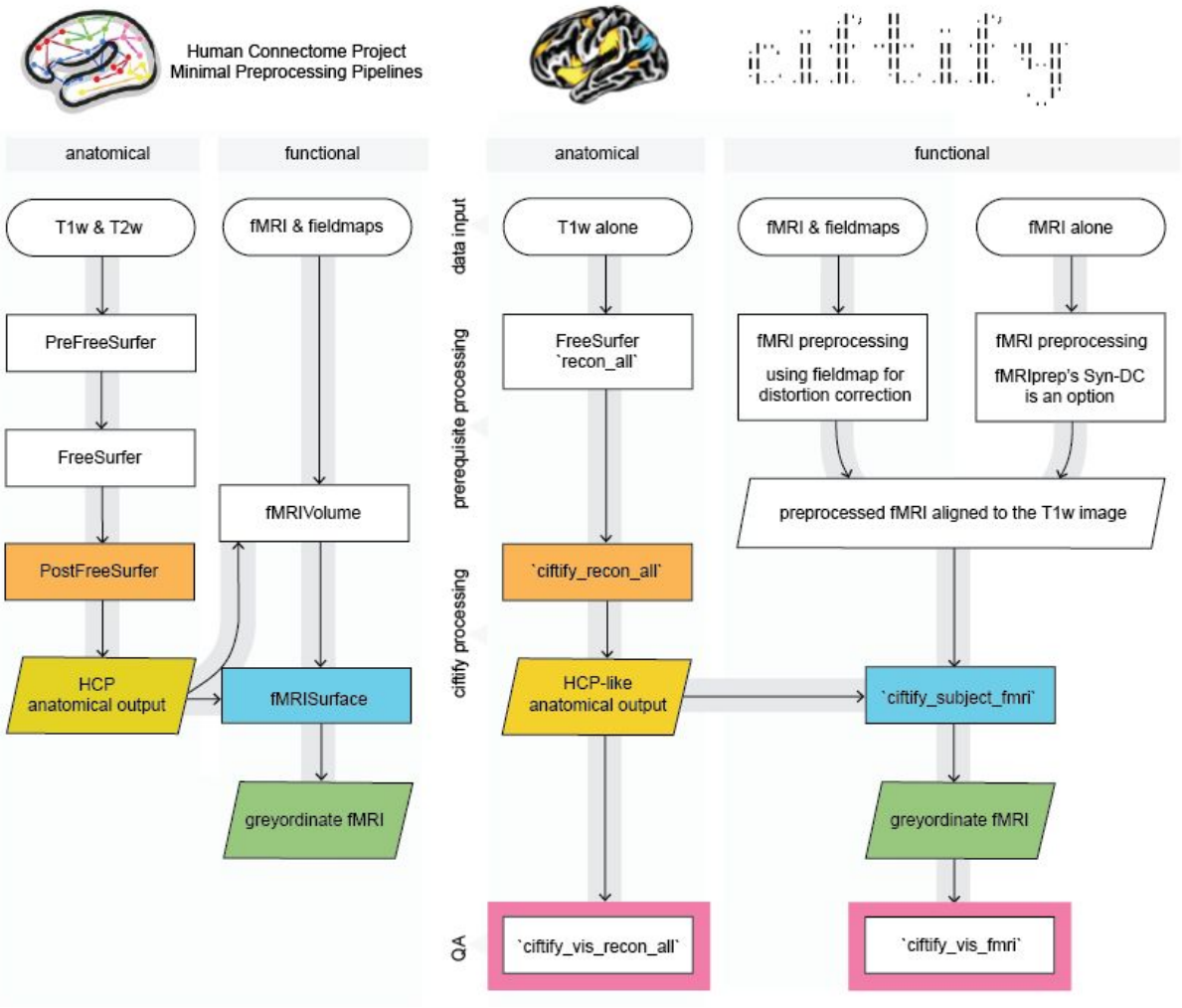

Figure 1. Key modules from HCP minimal processing pipelines (left) were adapted to create the ciftify preprocessing workflow (right). Within the fmriprep_ciftify BIDSapp the prerequisite processing steps listed here are performed by fmriprep. Additional usage options, including the

ability of the user to input pre-calculated transforms to standard space, are given at https://edickie.github.io/ciftify/.

2.1 Step 0: Prerequisite processing to ciftify

The ciftify approach is achieved starting from the standard FreeSurfer recon-all output derived from at least one raw T1w image. The requirements for the T1w image is that is can be used to produce freesurfer surface outputs passing quality assurance. Therefore, like FreeSurfer, we recommend a resolution no greater than 1mm isotropic. The ciftify approach to fMRI should, in theory, work on any Freesurfer output along with any fMRI volume that has been preprocessed with any fMRI preprocessing pipeline, assuming that 1) no volume-based smoothing has been applied and 2) the image has not been non-rigidly warped into standard space (Note: optional, expert user options allow for a standard space functional image to be provided with additional files detailing the standard space warp performed, see online documentation at https://edickie.github.io/ciftify/). To ensure the quality of the surface mapping step, it is recommended that preprocessing prior to ciftify_subject_fmri include steps to correct for EPI signal distortions. The fMRIPrep pipeline (Esteban et al., 2019a, 2019b) incorporates many recommended practices for these fMRI preprocessing steps, as well as running the FreeSurfer pipeline (version 6.0). Therefore, the fmriprep_ciftify BIDS-app will run FMRIPrep to accomplish these preprocessing steps. Those wishing to run ciftify without the FMRIPrep preprocessing base can run the individual ciftify steps manually. More instructions for doing so are given at https://edickie.github.io/ciftify/#/tutorials/example-usage. Recommendations for EPI distortion correction step are detailed further in *4.2 Functional preprocessing considerations*.

Importantly, the quality of surface-mapped MRI is strongly influenced by the registration between the functional volume and the anatomical data, which suffers in areas where the EPI image is distorted. HCP pipelines include two preprocessing steps to maximize the quality of registration across MR modalities: the acquisition and use of b0 field inhomogeneity maps (Glasser et al., 2013), and the use of Boundary-Based Registration (BBR) for cross-modal alignment (Greve and Fischl, 2009). The FMRIPrep pipeline is a “glass box” workflow that allows for a simplified user interface for the implementation of these recommended approaches for fieldmap-based EPI distortion correction and Boundary-Based Registration (BBR) for cross-modal alignment. In addition, FMRIPrep offers an option for “fieldmap-less” distortion correction, which estimates the areas of distortion using a combination of a participant’s T1w anatomical data and an average fieldmap template (Treiber et al., 2016; Wang et al., 2017).

2.2 Step 1: The participant anatomical workflow.

The first step in the ciftify workflow is accomplished by `ciftify_recon_all`, a command line utility adapted from the PostFreeSurferPipeline module of HCP’s Minimal Preprocessing Pipeline. Outputs from FreeSurfer are converted to GIFTI and CIFTI format. In turn, surface-based alignment of the cortical mesh is performed using the MSMSulc algorithm

(Robinson et al., 2018) followed by resampling to a 32k standard. In line with the HCP's Minimal Preprocessing Pipeline, `ciftify_recon_all` outputs a directory structure where output files are divided into subfolders representing analysis "spaces". More detail on these "spaces" is given in Glasser and colleagues (2013). In general, the MNINonLinear/fsLR32k combines the benefits of MNI volumetric registration (for subcortical structures) with surface-level registration based on sulcal anatomy for the cortex. This 91,282 standard-mesh "grayordinates" space, resampled to an approximate '32k' (average ~2mm spacing), is the HCP's standard space for human fMRI and multimodal analyses (note that HCP fMRI was acquired with 2mm voxel resolution). The surface meshes for fsLR32k space differ from the fsavarge meshes (used in FreeSurfer pipelines) not only in resolution (fsaverage5 has ~4mm spacing), but also (and more importantly) by providing mirror-symmetry between the left and right meshes (i.e., vertex 1024 on the left mesh corresponds in geographic location to vertex 1024 in the right mesh (Van Essen, 2012). This mirroring facilitates the analysis of symmetries and asymmetries in brain connectivity and diverse cortical features. Importantly, the CIFTI format also includes the FreeSurfer volumetric segmentation of subcortical and cerebellar structures.

For the purpose of retaining data provenance, `ciftify_recon_all` also produces a log file that records information about the input arguments, software environment, and information copied from the FreeSurfer logs about the environment in which recon_all was run. These logs also record all executed command line calls, so that the user can review the intermediate steps performed.

Due to the widespread use of FreeSurfer in the scientific community, it is anticipated that many investigators may already have recon_all outputs generated for their specific projects, with appropriate quality assurance completed. For such studies, the process to convert this data to CIFTI format can be achieved relatively quickly. However, if a high-resolution T2w image is available (i.e. a T2w image of voxel size matched to the T1w image), it is recommended that the HCPPipelines anatomical modules are run (PreFreeSurfer, FreeSurfer, PostFreeSurfer) instead of FreeSurfer and cifitify_recon_all (see *4.1 Anatomical preprocessing considerations*).

The anatomical workflow also produces quality assurance (QA) images, using the `cifti_vis_recon_all` utility, for reconstructed surfaces and subcortical masks (see Figure 2A). These images are analogous to those produced for other FreeSurfer quality assurance frameworks (such as FreeSurfer's own QATools). Some investigators may find these outputs to be a valuable adjunct to FreeSurfer recon_all outputs. An optional flag will output an accompanying ".scene" file, which allows the user to browse these QA views interactively in the connectome-workbench viewer. These QA views include slice representations with pial and white surface contour overlays, as well as the reconstructed surfaces (derived from the midthickness surface) with FreeSurfer's automatic parcellation (aparc) anatomical labels viewed from multiple standard angles. These images are generated using both the native and MNI space version of the outputs. Also, slices showing the masks generated by the FreeSurfer Automatic Segmentation (aseg) are presented in native T1w space so that these masks can be evaluated.

2.3 Step 2: The participant functional workflow

Once surface anatomy has been defined for a particular subject, the `ciftify_subject_fmri` utility is employed. The ciftify_subject_fmri stage maps preprocessed fMRI volumes to a specific subject's fsLR23k "grayordinate" space. This utility, adapted from the fMRIsurface module of HCP's Minimal Preprocessing Pipeline, projects from volume to the surface in a weighted, ribbon-constrained manner, where the value assigned to each vertex is calculated as a weighted average of the voxels encompassed (or partially encompassed) by the cortical ribbon. Excessively noisy voxels, defined according to their local coefficient of variation, are excluded from the mapping. Importantly, this surface mapping approach differs from those used in other known pipelines - such as FreeSurfer, that resample only from voxels that intersect with the midthickness surface. Another notable difference from previous approaches is that data from subcortical structures are resampled within their anatomical masks to approximately 32k subcortical voxels (including cerebellum). The final output is a CIFTI mapped (dtseries.nii) version of the functional data, in MNINonLinear, fsLR32k space. Note that the fMRI input volume should be "minimally" preprocessed with no volume smoothing and no non-rigid registration to standard space (else that same registration would need to be applied to the surfaces too, which is not supported by ciftify). From this step, optional additional CIFTI smoothing can be added (2D surface smoothing for the cortical ribbon and parcel-constrained smoothing for subcortical volumes), though see the cautionary note in Coalson et al (2018) about smoothing of any kind.

Like the anatomical workflow, the functional workflow also produces visualizations for QA of the volume-to-cortex mapping step as well as subcortical resampling (see Figure 2B), using the cifti_vis_fmri utility. These images include slice overlays of the functional volume with the pial surface and views of the unsmoothed functional signal on the surface.

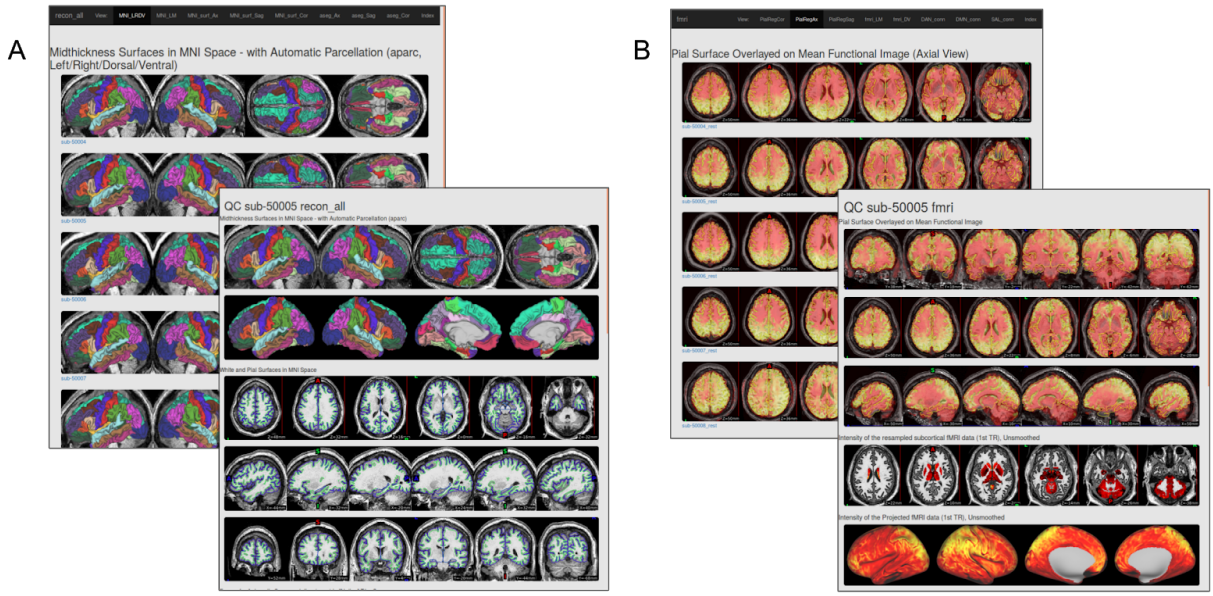

Figure 2. Example quality assurance (QA) pages created using A) cifti_vis_recon_all and B) cifti_vis_fmri. These example QA webpages can be viewed at https://edickie.github.io/ciftify/demo/qc_recon_all/index.html.

2.4 Generating a group quality assurance interface

These images are then combined into HTML pages so that collections of images can be viewed together in a browser window. Importantly, the HTML pages present the QA images in both subject-level pages (where all images related to one scanning session are presented together) as well as population-level index pages. These pages present the same view for all subjects in a study together. When the population is displayed beside each other, on the same page, QA failures can be identified more easily.

The cifti_vis utilities employ wb_command's '-show-scene' functionally to generate images for visual inspection. An optional flag will output an accompanying ".scene" file, which allows the user to browse these QA view options interactively in the connectome-workbench viewer. This framework does not require an external display nor MatLab proprietary software, which allows for easier generation on high-performance clusters, thus providing an advantage compared to other QA image-generating tools.

2.5 Programming Environment and Usage

The ciftify package contains a set of command-line utilities coded in Python. All code is publicly available under the MIT license at https://github.com/edickie/ciftify. The ciftify Python package (version 2.3.2) can be installed locally using the Python package management system (i.e., pip). Like the HCP pipeline, the ciftify package depends upon publicly available MR analysis software packages (FSL, FreeSurfer and Connectome Workbench) for various conversion and image manipulation steps. Also in concordance with the HCP Pipeline, the default surface registration algorithm is MSMSulc, dependent on the MSM registration tool (https://github.com/ecr05/MSM_HOCR/releases), and the Python packages docopt (docopt.org), nibabel (Brett et al., 2019), nilearn (Abraham et al., 2014), pybids (Yarkoni et al., 2019), NumPy (Walt et al., 2011), SciPy, PyYAML (https://pyyaml.org/), pandas (McKinney, 2010), seaborn (Waskom et al., 2017) and Matplotlib (Caswell et al., 2019) are dependencies. For further usage instructions and example code, see https://github.com/edickie/ciftify/wiki.

This complete programming environment is available as a BIDS app (Gorgolewski et al., 2017) available on Docker Hub (https://hub.docker.com/r/tigrlab/fmriprep_ciftify/). Therefore, the workflow should be relatively straightforward to set up and run for any dataset organized in accordance with the Brain Imaging Data Structure (BIDS; (Gorgolewski et al., 2016).

Example Docker usage:

```
docker run -ti --rm \
    -v /path/to/bids/raw/data:/bids_in:ro \
```

```
    -v /path/to/output:/derivatives \
    tigrlab/fmriprep_ciftify:latest /bids_in /derivatives participant
```

As with all BIDS apps the <bids_in> should be the top level bids directory containing all data participants data for the study. The <derivatives> output directory path will be populated with three subfolders of outputs (freesurfer/ fmriprep/ and ciftify/). If one has already completed part of the processing (i.e. FreeSurfer) these processing steps will be skipped by the BIDS app if the output folders are already present. The "bids_in" input argument provides a path to the dataset to be analyzed (read-only), which must conform to the BIDS standard (see the BIDS specification (http://bids.neuroimaging.io) and Supplementary Table 1 for more details of the input file structure). The above command would run the 'participant' level workflow for all available functional data in the dataset in a serial fashion. However, additional flags can be added select specific participants, sessions or tasks for processing. These flags are valuable for splitting up participant level processing into smaller tasks to be submitted to a computing cluster.

For example, one could select to only process the resting state task data from session 015 of the MyConnectome data (as in the above figure), using the following command:

```
docker run -ti --rm \
    -v /path/to/local/ds000031:/bids_in:ro \
    -v /filepath/to/output:/derivatives \
    tigrlab/fmrirep_ciftify:latest /bids_in /derivatives participant \
    --participant_label=01 \
    --task_label=rest \
    --session_label=015
```

Once all participants have been run, a final 'group' level mode will quickly write group level (index) quality assurance html pages.

```
docker run -ti --rm \
    -v /path/to/data/dir:/data:ro \
    -v /filepath/to/output/dir:/out \
    tigrlab/fmriprep_ciftify:latest /bids_in /derivatives group
```

The ciftify Docker container is intended to be run on personal computers and cloud services; however, Docker requires root or root-like permissions, and therefore may not be permitted in many multi-user environments, like High-Performance Clusters, for security reasons. For those using High-Performance Clusters, we recommend using the Singularity software. A Singularity image can either be built directly from the Docker specification (Singularity version >= 2.5), or the docker2singularity converter can be used to convert a Docker image into a Singularity image. For more detailed instructions for each use-case see (https://edickie.github.io/ciftify/#/01_installation).

For example, to build a Singularity image from the DockerHub specification:

```
singularity build /my_images/fmriprep_ciftify_image.simg
docker://tigrlab/fmriprep_ciftify:latest
```

To run the Singularity image:

```
singularity run --cleanenv \
    -B /path/to/bids/raw/data:/bids_in:ro \
    -B /path/to/output:/derivatives \
    fmriprep_ciftify_image.simg /bids_in /derivatives participant
```

The output directory structure is in similar to that of the HCP-Pipelines, is organized into folders according to the "space" in which the files are registered. Outputs are comprise of ~122 files approx 670 MB per participant. (See Supplementary Table 2 for more detailed list of all files generated).

3.0 Running analyses in CIFTI format

After data has been preprocessed into CIFTI format, the researcher needs to manipulate CIFTI files, in order to extract measures or run group analyses. For a researcher experienced only in volume-based analysis, this may present a challenge, as code may need to be adapted or rewritten. With recent software developments, most (if not all) calculations previously run in the volume are now possible on the surface. The Connectome Workbench command-line utility (wb_command) offers an extensive suite of functions that are well documented and computationally efficient (Marcus et al., 2013; https://www.humanconnectome.org/software/connectome-workbench). In addition, the nibabel Python package (http://nipy.org/nibabel/) can read and write NIFTI, GIFTI and CIFTI formats, allowing for custom computations in the MatLab or Python (NumPy) environments. Connectome Workbench offers conversion utilities (-cifti-convert) that convert CIFTI data to a "fake" NIFTI (i.e., a file that can be read and written by NIFTI utilities, but that does not contain correct spatial information) so that analyses that have been written for the NIFTI file format can be run. The greatest challenge is presented by those analyses that need to take spatial information into account (e.g., calculations of the spatial extent of a statistical cluster). Fortunately, the command line version of FSL's statistical packages, such as FSL's linear modelling tools (FILM and FLAME), MELODIC and FSL's Permutation Analysis of Linear Models (Winkler et al., 2015, PALM; 2014), for groupwise statistical tests, handles surface data and cifti format.

3.1.0 Additional utilities and ciftify binder.org learning environment.

The ciftify package also contains several utilities to assist in CIFTI-based surface-level analyses (see Table 1). These utilities are not meant as a replacement for the functionality in Connectome Workbench. In many cases, they are simple utilities that chain together multiple calls to Connectome Workbench. In these cases, commands can be echoed to the user when

these utilities are run in with the “--verbose” or “--debug” flags. In some cases (such as the ciftify_meants), data is read into NumPy arrays using the NiBabel package. All code and examples are publicly available on our GitHub repository (https://github.com/edickie/ciftify), which we hope can serve as example workflows for researchers new to the CIFTI analyses.

Table 1. Additional utilities included in the ciftify package

Name	Description
`ciftify_meants`	Produces a comma separated values (*.csv) file mean voxel/vertex timeseries from a functional file <func> within a seed mask <seed>. <func> functional data input can be NIFTI or CIFTI (*.dscalar.nii or *.dtseries.nii). Seed mask input can be in NIFTI, CIFTI (*.dlabel.nii or *.dscalar.nii) or GIFTI.
`ciftify_seed_corr`	Produces a seed correlation map of the mean timeseries within the seed with every vertex/voxel in the functional file. Like `ciftify_meants`, can take a combination of NIFTI, CIFTI or GIFTI inputs. The output seed correlation map matches the file type (NIFTI or CIFTI) of the input functional file.
`ciftify_clean_img`	Performs a combination of signal “cleaning” steps (bandpass filtering, confound regression, dummy TR removal and smoothing) for use on CIFTI and NIFTI inputs. Signal cleaning settings can be specified in combination in a “.json” configuration file, or as command line argument.
`ciftify_statclust_report`	Creates a tabular statistical report from an input (*.dscalar.nii) map. Includes overlap of clusters with standard atlases
`ciftify_atlas_report`	Creates a tabular report of overlap between the input atlas or labels (*.dlabel.nii) with standard atlases
`ciftify_vol_result`	Simplified mapping of a NIFTI statistical map or atlas to fsLR32k CIFTI “grayordinate” space of a specified subject.

In contrast to the full ciftify preprocessing workflow, which requires more than 10 Gigabits of neuroimaging software prerequisites, these tools require only python as well as connectome-workbench. Therefore, they should be easier to install for new learners in a classroom or workshop setting. The programming environment is also available in an online interactive instance on binder.org (https://mybinder.org/v2/gh/edickie/ciftify/master). Future work intends to build interactive examples into this public, online, tutorial space.

3.1.1 ciftify_statclust_report

The ciftify_statclust_info utility generates a table (.csv output) listing statistical clusters defined from an input CIFTI (.dscalar.nii) statistical map. A CIFTI file of labelled clusters is also output. The table lists, for each cluster, the cluster ID and name (matching the label in the corresponding dlabel.nii output), as well as the surface area of that cluster (calculated on the midthickness surface). The table also reports the overlap of the cluster with three atlases: (i) the FreeSurfer anatomical atlas (Desikan et al., 2006); (ii) the Yeo seven resting state networks atlas (Yeo et al., 2011); and (iii) the Glasser Multimodal Parcellation atlas (Glasser et al., 2016a). For each peak, for each atlas, the table reports the label in which the peak vertex falls, as well as the proportion of the cluster (from the statistical input map) that overlaps with that index (see Figure 3).

label_name	label_idx	mean_value	area	DKT_overlap	Yeo7_overlap	MMP_overlap
LABEL_1	1	3.41	736	L_precuneus (61.5%); L_isthmuscingulate (38.5%);	7Networks_7 (100.0%);	L_31pv_ROI (32.4%); L_31pd_ROI (29.6%); L_v23ab_ROI (20.0%); L_d23ab_ROI (14.7%);

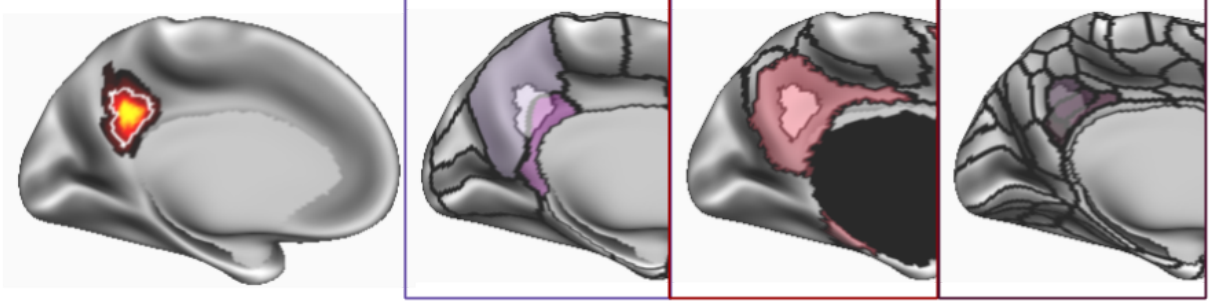

Figure 3. Example output table generated by ciftify_statclust_report. For this simulated statistical map (red-yellow shading), the output table summarizes a statistic map, according to input statistical thresholds, as one cluster (white outline). From the cluster, ID number, mean statistical value and the approximate surface area in mm^2. Then, for three surface atlases, we report atlas labels encompassing the proportion of the cluster's surface area that overlaps with this atlas label. In this example, 61.5% of the cluster overlaps the "L_precuneus" label of the FreeSurfer anatomical atlas (Desikan et al., 2006). Likewise, the entire cluster falls within Network 7 (i.e. the Default Mode Network) of the Yeo atlas (Yeo et al., 2011). Lastly, 32.4% of the cluster's surface area falls within the 'L_31pv_ROI' label from Glasser and colleagues Multimodal Parcellation atlas (Glasser et al., 2016a).

4.0 DISCUSSION

The brain imaging field has lacked a framework for leveraging the HCP preprocessing philosophy for non-HCP legacy data. The ciftify project presented here helps to close this gap and encourages wider adoption of surface-based analysis of MR data, specifically using the

CIFTI file format, for groups working with non-HCP legacy datasets. Allowing researchers to adapt the preprocessing pipelines they are most familiar with allows for 1) easier uptake of surface-based analysis insofar as we remove a steep learning curve and 2) a more flexible framework whereby we can test different preprocessing steps in order to optimize surface-level results across various acquisition settings. For this purpose, the workflow (fmriprep_ciftify), packaged in as a BIDS-app, adapts two critical sections from HCP's Minimal Preprocessing Pipelines and can be readily deployed on legacy datasets. This workflow also integrates the generation of quality assurance visualizations. In addition, the ciftify package contains a suite of tools to aid in the analysis and manipulation of data after it has been preprocessed into CIFTI format.

4.1 Anatomical preprocessing considerations

The goal of ciftify_recon_all was to start directly from default recon-all FreeSurfer outputs. Thus it offers a relatively simple starting point for groups to adopt and test: generating a CIFTI file of surface-level structural measures (such as cortical thickness) is just one command line call away from the workflow that many researchers are already familiar with. While this is an attractive prospect, we feel it important to stress that, when a high-resolution T2w image is available, the additional preprocessing steps incorporated into HCP's Minimal Preprocessing Pipeline (especially the adapted FreeSurfer approach that uses the T2w signal for better pial surface placement) will yield superior results. To be even more explicit, although the results generated from ciftify will retain many advantages over volume-based processing, they are not equivalent to the results generated from the HCP's minimal preprocessing pipelines. Note that by "high resolution" T2w image, we are referring to a T2w, or FLAIR acquisition matched in voxel size to the T1w image (i.e. no greater than 1mm isotropic). Hence many existing studies that acquired T2w or FLAIR scans with thick slices or gaps of 2-7mm would not be appropriate. Therefore, the ciftify_recon_all workflow is only recommended for cases when a high-resolution T2w image is not available. Note that the anatomical outputs from the HCP's Minimal Preprocessing Pipeline can be substituted for ciftify_recon_all outputs for those who would like to use ciftify's other tools (such as ciftify_subject_fmri).

There remains an ongoing debate regarding the best algorithms, parameters and software for accurate between-subject alignment. We argue that the best registration algorithm should achieve the best alignment for a given level of distortion, or the least distortion for a given level of alignment. This concept applies to both volume-based and surface-based alignment steps within the ciftify pipeline. For the volume-based transformation to standard (i.e. MNI) space, we prefer FNIRT. Compared to the ANTs implementation in fmriprep, FNIRT is tuned conservatively to generate less distortion and as a result, does not overfit to folding patterns (Coalson et al. 2018 see Supplementary Figures S11 and S12). This recommendation for volume-based alignment mirrors recent observations for surface-based alignment, where the FreeSurfer default alignment produces more spherical distortion and performs less well at aligning cortical areas than does MSM alignment that has been tuned for less distortion (the amount of distortion allowed was determined by optimizing for maximal task-based functional alignment (Robinson

et al. 2018; Robinson et al. 2014). However, as both volume and surface alignment tools continue to be developed, mechanisms need to be available for available tools to be validated in comparison to one another, within the context of pipelines like ciftify. Therefore, we have added expert-user options to change realignment settings or to override the realignment with a user-provided (FSL format) transform file.

4.2 Functional preprocessing considerations

Importantly, the quality of surface-mapped MRI is a product of the quality of the volume-to-surface mapping step. This, in turn, is a direct consequence of the registration between the functional volume and the anatomical data, which suffers in areas where the EPI image is distorted. The HCP adopted state of the art methods to maximize the quality of registration across MR modalities. These methods included upgrading sequences to acquire data at a voxel size smaller than average cortical thickness, the acquisition of data for correcting EPI distortion (Glasser et al., 2013), and the use of Boundary-Based Registration (BBR) for cross-modal alignment (Greve and Fischl, 2009). While it is clear that legacy acquisitions will generally not match the quality of HCP, it is important that investigators make preprocessing choices that will maximize the data quality of their dataset given its limitations.
A critical step for good fMRI-to-anatomical mapping is correcting for fMRI distortions. Correcting for susceptibility distortions in fMRI data can be done with several approaches. The preferred technique is to use data from additional field map or phase reversed spin echo scans. Phase encoding polarity techniques combine images with opposing phase encoding to calculate distortion maps (using tools such as FSL's TOPUP (Andersson et al., 2003). This is the technique currently used by the HCP. Second, field maps can be estimated using sequences that measure phase evolution in time between two close gradient echo acquisitions (Hutton et al., 2002). It is becoming common practice to incorporate one of these techniques into current acquisition protocols. However, proper incorporation of these sequences into a fMRI preprocessing pipeline can be challenging for a novice brain scientist, as they often require increased knowledge of the specific scan acquisition settings. The fMRIprep pipeline offers a helpful interface in this regard, where the appropriate EPI distortion software will be called as long as the input files are organized in accordance with the Brain Imaging Data Structure (BIDS). For older acquisitions, where no field map or reverse-phase encoded maps are available, FMRIPrep has implemented a displacement field estimation via nonlinear registration. Another experimental approach is the incorporation of mean field map images (if they are available) from the employed scanner. Future work will explore these and other distortion correction options in legacy fMRI data to ensure that the deformations they generate lead to net improved registrations over no correction.

Another important consideration is whether or not to smooth data on the surface or within the subcortical grey matter structures. Spatial smoothing increases signal to noise ratio, but at a cost to anatomical specificity. It is extremely important to avoid volume-based smoothing before the volume-to-surface mapping step because such smoothing (even with a small 4mm FWHM

kernel) is very deleterious to cortical anatomical specificity (Coalson et al., 2018). Once fMRI data is in cifti format, 2D surface-based smoothing can be applied to the cortical data and 3D parcel-constrained smoothing can be applied to the subcortical data. If smoothing is used, we recommend moderation, as even 4mm FWHM in the surface is mildly deleterious, particularly to cortical areas that are narrow (Coalson et al., 2018), but may provide an acceptable tradeoff to increase signal to noise ratio in small sample sizes (in large datasets like the HCP smoothing is often unnecessary). A preferable alternative for improving the signal to noise ratio are approaches like parcellation (Glasser et al., 2016a) or using Wishart-based filtering to preferentially smooth noise over spatiotemporally structured signal (Glasser et al., 2016b). Parcellation is especially preferable to smoothing for researchers who are interested in effects at the level of cortical areas.

4.3 The value of higher resolution data

While we advocate use of the HCP analytic approach to all MR datasets, we recognize the synergistic benefits from combining the HCP's analytic approach with the HCP's higher standards for MR data acquisition. In particular, higher spatial resolution for MR data, specifically a voxel size less than the mean cortical thickness (~2.6mm) allows for better separation of signal across sulci/gyri of opposite grey matter ribbons (Glasser et al., 2016b). That said, even fMRI resolution as low as 4mm benefits from the HCP-Style approach to cross-subject alignment and not smoothing in the volume (Coalson et al., 2018). In addition, increasing the temporal resolution of MR data allows for more effective cleaning for MR artifacts from motion or physiology (Glasser et al., 2017, 2016b; Salimi-Khorshidi et al., 2014).

4.4 A note about visualizing a volume-based result on the cortical surface

It is becoming more common for researchers to provide surface visualizations when reporting the results of a volume-based group analysis. This practice is beneficial, in comparison to slice representations, for the communication of whole brain results, and can drive more thoughtful interpretations that reflect large-scale networks. For this reason, ciftify_vol_result, a utility for doing simple NIFTI to CIFTI mapping, is included in the ciftify package. If a group average 'subject' (i.e. HCP_S1200_GroupAvg) is specified, 'average fiducial mapping', or mapping to average surfaces, is performed (Van Essen, 2005). This type of mapping, while fast and easy, suffers from the drawback that group average surfaces fail to closely overlap with the actual location of the cortical ribbon. Instead, they "drift" towards the inside of curvature (often towards white matter) especially in areas of variable folding patterns, such as the association cortex (Coalson et al., 2018). What is more troubling is that if the correspondence between MNI coordinates and cortical areas is not uniform throughout the cortex, results will be most severely impacted in regions of high individual variability and especially for gyral crowns and sulcal fundi. An alternative approach is to repeat the volume-to-surface mapping steps using the surfaces from a collection of subjects, and then to summarise (i.e., average) the results. This so-called "multi-fiducial mapping" (Van Essen, 2005) samples more uniformly from association cortex,

with the caveat that the effects of blurring from cross-subject misregistration will be applied twice (Coalson et al., 2018).

Unfortunately, no direct method for surface projection of a group average volume result is accurate. Once data has been averaged and smoothed in the volume, precise cortical localization cannot be recovered for cortical areas of high folding variability (i.e. association cortex) (Coalson et al., 2018) - although the magnitude of this problem may be minimal in some areas (like parts of the insula) where there is lower variability of folds and of areas relative to folds across subjects (Coalson et al., 2018). Therefore, the only way to accurately map the cortical areas for a statistical contrast of interest is following a surface-based preprocessing of individual subjects and surface-level group analysis. This fact is one of the major reasons we developed ciftify. This has implications for the appropriate combination of ciftify utilities when evaluating the result of a volume-based analysis: while it is technically possible to input a dense scalar file generated from ciftify_vol_result into ciftify_statclust_report, the resulting table should be interpreted with great caution. In particular, due to individual differences in cortical folding patterns and functional organization, the MMP1.0 parcellation cannot be accurately translated into MNI volume space (Coalson et al., 2018). Therefore, ciftify_peaktable is not recommended for the results from a volume-based analysis.

4.5 Future directions

While important advances in software development for the CIFTI file format have been achieved, a multitude of analysis workflows are more easily applied in volume than on the surface. Examples of comprehensive analysis packages include the NiLearn project for machine learning in Python, which contains several integrated utilities of NIFTI file manipulation, plotting, and time-series extraction (Abraham et al., 2014). Future work will be needed to integrate CIFTI file reading and writing into these programming environments. In addition, neuroinformatics platforms are well known for the systematic (machine-readable) reporting and meta-analysis of statistical maps in MNI (volumetric) space. Specific projects include neurosynth (neurosynth.org; Yarkoni et al., 2011), NeuroVault (https://neurovault.org/; Gorgolewski et al., 2015) and BrainMap (brainmap.org). Moreover, the greater neuroimaging community has aided these efforts by defining standards for reporting of these data as part of the neuroimaging data model (Maumet et al., 2016). The seeds for an alternative approach are provided by the BALSA database (balsa.wustl.edu, (Van Essen et al., 2017), a repository for sharing CIFTI-based results (via scene files in Connectome Workbench). Future work will be needed to develop methods for meta-analytic synthesis of these results.

Currently, the ciftify workflow employs surface anatomy (sulcal depth) based alignment to align cortical data across subjects. Recent advances show that improvements in cross-subject alignment can be gained using additional features such as resting-state connectivity (Conroy et al., 2013; Robinson et al., 2014; Tong et al., 2017) and/or surface myelin maps (Robinson et al., 2014). The HCP pipelines have implemented these advances in registration into their workflow, using the Multimodal Surface Matching registration tool (Robinson et al., 2018, MSM; 2014), to

incorporate fMRI and anatomical features in the “MSMAll” registration (Glasser et al., 2016b). A ciftify workflow using only T1w data cannot match the precision of the HCP Pipeline surfaces. The HCP surface refinement procedure requires the high-resolution T2w image as well as additional tools for registration. Individualized parcellations based on functionally-related features (such as resting-state connectivity) could potentially provide an additional mechanism for improving cross-participant correspondence when the signal is averaged within these parcels (Glasser et al., 2016a; Wang et al., 2015). Future work is needed to validate which cortical regions might be effectively parcellated using the features available in legacy datasets.

There is rapid growth in the field of surface-based MR analysis. A growing number of analytic toolkits and methods use surface-based approaches, including novel 2D registration algorithms (Robinson et al., 2018; Tong et al., 2017), diffusion embedding(Margulies et al., 2016), and individualized parcellation(Glasser et al., 2016a; Kong et al., 2018; Schaefer et al., 2018; Wang et al., 2015)). However, due to differences in the underlying surface meshes (for example fsaverage5 vs fsLR32k vs CIVET), combining and comparing these methods remains technically challenging. The ciftify package allows for immediate application of tools and atlases developed using the HCP data (published in fsLR32k space) to a multitude of additional datasets. We believe the field will benefit from a simplified framework that promotes interoperability of different surface-based tools and facilitates evaluation of their performance on various type of MR acquistions (where diverse sources of noise may be present). We hope that the ciftify toolkit will provide a more standardized (and available) approach for the steps required to integrate fsLR32k atlases and tools into extended harmonized workflows.

4.6 Conclusions

The ciftify package offers a bridging solution for legacy data that will allow many researchers to adopt CIFTI format analyses. We intend for the analytic framework provided within the ciftify package to be a starting point for future bridging work. In parallel, the research community is increasingly embracing the HCP’s standards for MR data acquisition. Several initiatives, e.g. HCP Lifespan, the Connectomes Related to Human Disease, ABCD (Bjork et al., 2017) and the UK Biobank (Miller et al., 2016)) are now taking advantage of higher spatial and temporal resolution of MR data made possible using multi-band (multi-slice) acquisition (Ugurbil et al., 2013). Going forward, we hope that studies of such methodological quality will become the norm, rather than the exception. Whether using cifity for legacy data or CIFTI for HCP-quality data, it is our hope that an analytic shift from traditional volume-based approaches to surface-based (i.e. grayordinate) approaches will accelerate within the neuroimaging community.

ACKNOWLEDGEMENTS

ANV acknowledges support related to the present work the Canadian Institute for Health Research, Ontario Mental Health Foundation, and Centre for Addiction and Mental Health Foundation, the Ontario Ministry of Research and Innovation, the Canadian Foundation for

Innovation. MFG and DCVE acknowledge support from NIH (F30 MH097312 to MFG, RO1 MH-60974 to DCVE, and 3R24 MH108315 to DCVE).

REFERENCES

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., Varoquaux, G., 2014. Machine learning for neuroimaging with scikit-learn. Front. Neuroinform. 8. https://doi.org/10.3389/fninf.2014.00014

Andersson, J.L.R., Skare, S., Ashburner, J., 2003. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. Neuroimage 20, 870–888.

Anticevic, A., Dierker, D.L., Gillespie, S.K., Repovs, G., Csernansky, J.G., Van Essen, D.C., Barch, D.M., 2008. Comparing surface-based and volume-based analyses of functional neuroimaging data in patients with schizophrenia. Neuroimage 41, 835–848.

Argall, B.D., Saad, Z.S., Beauchamp, M.S., 2006. Simplified intersubject averaging on the cortical surface using SUMA. Hum. Brain Mapp. 27, 14–27.

Bjork, J.M., Straub, L.K., Provost, R.G., Neale, M.C., 2017. The ABCD Study of Neurodevelopment: Identifying Neurocircuit Targets for Prevention and Treatment of Adolescent Substance Abuse. Curr Treat Options Psych 4, 196–209.

Braga, R.M., Buckner, R.L., 2017. Parallel Interdigitated Distributed Networks within the Individual Estimated by Intrinsic Functional Connectivity. Neuron 95, 457–471.e5.

Brett, M., Hanke, M., Markiewicz, C., Côté, M.-A., McCarthy, P., Cheng, C., Halchenko, Y., Ghosh, S., Wassermann, D., Gerhard, S., Larson, E., Lee, G.R., Kastman, E., Cindee, M., Rokem, A., Morency, F.C., moloney, Cottaar, M., Millman, J., Markello, R., jaeilepp, Gramfort, A., Vincent, R.D., van den Bosch, J.J.F., Subramaniam, K., Raamana, P.R., Goncalves, M., Nichols, N., embaker, Basile, 2019. nipy/nibabel: 2.3.3. https://doi.org/10.5281/zenodo.2541736

Caswell, T.A., Droettboom, M., Hunter, J., Firing, E., Lee, A., Klymak, J., Stansby, D., de Andrade, E.S., Nielsen, J.H., Varoquaux, N., Root, B., Elson, P., May, R., Dale, D., Hoffmann, T., Lee, J.-J., Seppänen, J.K., McDougall, D., Straw, A., Hobson, P., Gohlke, C., Yu, T.S., Ma, E., Vincent, A.F., Silvester, S., Moad, C., Katins, J., Kniazev, N., Ariza, F., Ernest, E., 2019. matplotlib/matplotlib v3.0.3. https://doi.org/10.5281/zenodo.2577644

Choudhury, S., Fishman, J.R., McGowan, M.L., Juengst, E.T., 2014. Big data, open science and the brain: lessons learned from genomics. Front. Hum. Neurosci. 8, 239.

Coalson, T.S., Van Essen, D.C., Glasser, M.F., 2018. The impact of traditional neuroimaging methods on the spatial localization of cortical areas. Proc. Natl. Acad. Sci. U. S. A. 115, E6356–E6365.

Conroy, B.R., Singer, B.D., Guntupalli, J.S., Ramadge, P.J., Haxby, J.V., 2013. Inter-subject alignment of human cortical anatomy using functional connectivity. Neuroimage 81, 400–411.

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

Eickhoff, S., Nichols, T.E., Van Horn, J.D., Turner, J.A., 2016. Sharing the wealth: Neuroimaging data repositories. Neuroimage 124, 1065–1068.

Esteban, O., Blair, R., Markiewicz, C.J., Berleant, S.L., Moodie, C., Ma, F., Isik, A.I., Erramuzpe, A., Kent, J.D., Goncalves, M., DuPre, E., Sitek, K.R., Gomez, D.E.P., Lurie, D.J., Ye, Z., Salo, T., Valabregue, R., Amlien, I.K., Liem, F., Jacoby, N., Stojic, H., Poldrack, R.A., Gorgolewski, K.J., 2019a. poldracklab/fmriprep: 1.3.0.post2. https://doi.org/10.5281/zenodo.2565292

Esteban, O., Markiewicz, C.J., Blair, R.W., Moodie, C.A., Isik, A.I., Erramuzpe, A., Kent, J.D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S.S., Wright, J., Durnez, J., Poldrack, R.A., Gorgolewski, K.J., 2019b. fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat. Methods 16, 111–116.

Fischl, B., 2012. FreeSurfer. Neuroimage 62, 774–781.

Fischl, B., Sereno, M.I., Dale, A.M., 1999. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9, 195–207.

Glasser, M.F., Coalson, T.S., Bijsterbosch, J.D., Harrison, S.J., Harms, M.P., Anticevic, A., Van Essen, D.C., Smith, S.M., 2017. Using Temporal ICA to Selectively Remove Global Noise While Preserving Global Signal in Functional MRI Data. bioRxiv. https://doi.org/10.1101/193862

Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., Smith, S.M., Van Essen, D.C., 2016a. A multi-modal parcellation of human cerebral cortex. Nature 536, 171–178.

Glasser, M.F., Goyal, M.S., Preuss, T.M., Raichle, M.E., Van Essen, D.C., 2014. Trends and properties of human cerebral cortex: correlations with cortical myelin content. Neuroimage 93 Pt 2, 165–175.

Glasser, M.F., Smith, S.M., Marcus, D.S., Andersson, J.L.R., Auerbach, E.J., Behrens, T.E.J., Coalson, T.S., Harms, M.P., Jenkinson, M., Moeller, S., Robinson, E.C., Sotiropoulos, S.N., Xu, J., Yacoub, E., Ugurbil, K., Van Essen, D.C., 2016b. The Human Connectome Project's neuroimaging approach. Nat. Neurosci. 19, 1175–1187.

Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., Van Essen, D.C., Jenkinson, M., WU-Minn HCP Consortium, 2013. The minimal preprocessing pipelines for the Human Connectome Project. Neuroimage 80, 105–124.

Glasser, M.F., Van Essen, D.C., 2011. Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. J. Neurosci. 31, 11597–11616.

Goebel, R., Esposito, F., Formisano, E., 2006. Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. Hum. Brain Mapp. 27, 392–401.

Gorgolewski, K.J., Alfaro-Almagro, F., Auer, T., Bellec, P., Capotă, M., Chakravarty, M.M., Churchill, N.W., Cohen, A.L., Craddock, R.C., Devenyi, G.A., Eklund, A., Esteban, O., Flandin, G., Ghosh, S.S., Guntupalli, J.S., Jenkinson, M., Keshavan, A., Kiar, G., Liem, F., Raamana, P.R., Raffelt, D., Steele, C.J., Quirion, P.-O., Smith, R.E., Strother, S.C., Varoquaux, G., Wang, Y., Yarkoni, T., Poldrack, R.A., 2017. BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. PLoS Comput. Biol. 13, e1005209.

Gorgolewski, K.J., Auer, T., Calhoun, V.D., Craddock, R.C., Das, S., Duff, E.P., Flandin, G., Ghosh, S.S., Glatard, T., Halchenko, Y.O., Handwerker, D.A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B.N., Nichols, T.E., Pellman, J., Poline, J.-B., Rokem, A., Schaefer, G., Sochat, V., Triplett, W., Turner, J.A., Varoquaux, G., Poldrack, R.A., 2016. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci Data 3, 160044.

Gorgolewski, K.J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S.S., Maumet, C., Sochat, V.V., Nichols, T.E., Poldrack, R.A., Poline, J.-B., Yarkoni, T., Margulies, D.S., 2015. NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. Front. Neuroinform. 9, 8.

Greve, D.N., Fischl, B., 2009. Accurate and robust brain image alignment using boundary-based registration. Neuroimage 48, 63–72.
Hutton, C., Bork, A., Josephs, O., Deichmann, R., Ashburner, J., Turner, R., 2002. Image distortion correction in fMRI: A quantitative evaluation. Neuroimage 16, 217–240.
Jo, H.J., Lee, J.-M., Kim, J.-H., Shin, Y.-W., Kim, I.-Y., Kwon, J.S., Kim, S.I., 2007. Spatial accuracy of fMRI activation influenced by volume- and surface-based spatial smoothing techniques. Neuroimage 34, 550–564.
Kong, R., Li, J., Orban, C., Sabuncu, M.R., Liu, H., Schaefer, A., Sun, N., Zuo, X.-N., Holmes, A.J., Eickhoff, S.B., Yeo, B.T.T., 2018. Spatial Topography of Individual-Specific Cortical Networks Predicts Human Cognition, Personality, and Emotion. Cereb. Cortex. https://doi.org/10.1093/cercor/bhy123
Marcus, D.S., Harms, M.P., Snyder, A.Z., Jenkinson, M., Wilson, J.A., Glasser, M.F., Barch, D.M., Archie, K.A., Burgess, G.C., Ramaratnam, M., Hodge, M., Horton, W., Herrick, R., Olsen, T., McKay, M., House, M., Hileman, M., Reid, E., Harwell, J., Coalson, T., Schindler, J., Elam, J.S., Curtiss, S.W., Van Essen, D.C., WU-Minn HCP Consortium, 2013. Human Connectome Project informatics: quality control, database services, and data visualization. Neuroimage 80, 202–219.
Margulies, D.S., Ghosh, S.S., Goulas, A., Falkiewicz, M., Huntenburg, J.M., Langs, G., Bezgin, G., Eickhoff, S.B., Castellanos, F.X., Petrides, M., Jefferies, E., Smallwood, J., 2016. Situating the default-mode network along a principal gradient of macroscale cortical organization. Proc. Natl. Acad. Sci. U. S. A. 113, 12574–12579.
Maumet, C., Auer, T., Bowring, A., Chen, G., Das, S., Flandin, G., Ghosh, S., Glatard, T., Gorgolewski, K.J., Helmer, K.G., Jenkinson, M., Keator, D.B., Nichols, B.N., Poline, J.-B., Reynolds, R., Sochat, V., Turner, J., Nichols, T.E., 2016. Sharing brain mapping statistical results with the neuroimaging data model. Sci Data 3, 160102.
McKinney, W., 2010. Data Structures for Statistical Computing in Python, in: van der Walt, S., Millman, J. (Eds.), Proceedings of the 9th Python in Science Conference. pp. 51–56.
Milham, M.P., Craddock, R.C., Son, J.J., Fleischmann, M., Clucas, J., Xu, H., Koo, B., Krishnakumar, A., Biswal, B.B., Castellanos, F.X., Colcombe, S., Di Martino, A., Zuo, X.-N., Klein, A., 2018. Assessment of the impact of shared brain imaging data on the scientific literature. Nat. Commun. 9, 2818.
Miller, K.L., Alfaro-Almagro, F., Bangerter, N.K., Thomas, D.L., Yacoub, E., Xu, J., Bartsch, A.J., Jbabdi, S., Sotiropoulos, S.N., Andersson, J.L.R., Griffanti, L., Douaud, G., Okell, T.W., Weale, P., Dragonu, I., Garratt, S., Hudson, S., Collins, R., Jenkinson, M., Matthews, P.M., Smith, S.M., 2016. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Nat. Neurosci. 19, 1523–1536.
Mueller, S., Wang, D., Fox, M.D., Yeo, B.T.T., Sepulcre, J., Sabuncu, M.R., Shafee, R., Lu, J., Liu, H., 2013. Individual variability in functional connectivity architecture of the human brain. Neuron 77, 586–595.
Poldrack, R.A., Barch, D.M., Mitchell, J.P., Wager, T.D., Wagner, A.D., Devlin, J.T., Cumba, C., Koyejo, O., Milham, M.P., 2013. Toward open sharing of task-based fMRI data: the OpenfMRI project. Front. Neuroinform. 7, 12.
Poldrack, R.A., Gorgolewski, K.J., 2014. Making big data open: data sharing in neuroimaging. Nat. Neurosci. 17, 1510–1517.
Robinson, E.C., Garcia, K., Glasser, M.F., Chen, Z., Coalson, T.S., Makropoulos, A., Bozek, J., Wright, R., Schuh, A., Webster, M., Hutter, J., Price, A., Cordero Grande, L., Hughes, E., Tusor, N., Bayly, P.V., Van Essen, D.C., Smith, S.M., Edwards, A.D., Hajnal, J., Jenkinson, M., Glocker, B., Rueckert, D., 2018. Multimodal surface matching with higher-order

smoothness constraints. Neuroimage 167, 453–465.
Robinson, E.C., Jbabdi, S., Glasser, M.F., Andersson, J., Burgess, G.C., Harms, M.P., Smith, S.M., Van Essen, D.C., Jenkinson, M., 2014. MSM: a new flexible framework for Multimodal Surface Matching. Neuroimage 100, 414–426.
Salimi-Khorshidi, G., Douaud, G., Beckmann, C.F., Glasser, M.F., Griffanti, L., Smith, S.M., 2014. Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. Neuroimage 90, 449–468.
Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.-N., Holmes, A.J., Eickhoff, S.B., Yeo, B.T.T., 2018. Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. Cereb. Cortex 28, 3095–3114.
Smith, K., 2012. Brain imaging: fMRI 2.0. Nature 484, 24–26.
Tong, T., Aganj, I., Ge, T., Polimeni, J.R., Fischl, B., 2017. Functional density and edge maps: Characterizing functional architecture in individuals and improving cross-subject registration. Neuroimage 158, 346–355.
Treiber, J.M., White, N.S., Steed, T.C., Bartsch, H., Holland, D., Farid, N., McDonald, C.R., Carter, B.S., Dale, A.M., Chen, C.C., 2016. Characterization and Correction of Geometric Distortions in 814 Diffusion Weighted Images. PLoS One 11, e0152472.
Tucholka, A., Fritsch, V., Poline, J.-B., Thirion, B., 2012. An empirical comparison of surface-based and volume-based group studies in neuroimaging. Neuroimage 63, 1443–1453.
Uğurbil, K., Xu, J., Auerbach, E.J., Moeller, S., Vu, A.T., Duarte-Carvajalino, J.M., Lenglet, C., Wu, X., Schmitter, S., Van de Moortele, P.F., Strupp, J., Sapiro, G., De Martino, F., Wang, D., Harel, N., Garwood, M., Chen, L., Feinberg, D.A., Smith, S.M., Miller, K.L., Sotiropoulos, S.N., Jbabdi, S., Andersson, J.L.R., Behrens, T.E.J., Glasser, M.F., Van Essen, D.C., Yacoub, E., WU-Minn HCP Consortium, 2013. Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. Neuroimage 80, 80–104.
Van Essen, D.C., 2012. Cortical cartography and Caret software. Neuroimage 62, 757–764.
Van Essen, D.C., 2005. A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. Neuroimage 28, 635–662.
Van Essen, D.C., Glasser, M.F., 2018. Parcellating Cerebral Cortex: How Invasive Animal Studies Inform Noninvasive Mapmaking in Humans. Neuron 99, 640–663.
Van Essen, D.C., Smith, J., Glasser, M.F., Elam, J., Donahue, C.J., Dierker, D.L., Reid, E.K., Coalson, T., Harwell, J., 2017. The Brain Analysis Library of Spatial maps and Atlases (BALSA) database. Neuroimage 144, 270–274.
Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, W.U.-M.H., 2013. The WU-Minn Human Connectome Project: an overview. Neuroimage 80, 62–79.
Van Essen, D.C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T.E.J., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S.W., Della Penna, S., Feinberg, D., Glasser, M.F., Harel, N., Heath, A.C., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S.E., Prior, F., Schlaggar, B.L., Smith, S.M., Snyder, A.Z., Xu, J., Yacoub, E., WU-Minn HCP Consortium, 2012. The Human Connectome Project: a data acquisition perspective. Neuroimage 62, 2222–2231.
Walt, S. van der, Colbert, S.C., Varoquaux, G., 2011. The NumPy Array: A Structure for Efficient Numerical Computation. Comput. Sci. Eng. 13, 22–30.
Wang, D., Buckner, R.L., Fox, M.D., Holt, D.J., Holmes, A.J., Stoecklein, S., Langs, G., Pan, R., Qian, T., Li, K., Baker, J.T., Stufflebeam, S.M., Wang, K., Wang, X., Hong, B., Liu, H., 2015. Parcellating cortical functional networks in individuals. Nat. Neurosci. 18, 1853–1860.

Wang, S., Peterson, D.J., Gatenby, J.C., Li, W., Grabowski, T.J., Madhyastha, T.M., 2017. Evaluation of Field Map and Nonlinear Registration Methods for Correction of Susceptibility Artifacts in Diffusion MRI. Front. Neuroinform. 11, 17.

Waskom, M., Botvinnik, O., O'Kane, D., Hobson, P., Lukauskas, S., Gemperline, D.C., Augspurger, T., Halchenko, Y., Cole, J.B., Warmenhoven, J., de Ruiter, J., Pye, C., Hoyer, S., Vanderplas, J., Villalba, S., Kunter, G., Quintero, E., Bachant, P., Martin, M., Meyer, K., Miles, A., Ram, Y., Yarkoni, T., Williams, M.L., Evans, C., Fitzgerald, C., Brian, Fonnesbeck, C., Lee, A., Qalieh, A., 2017. mwaskom/seaborn: v0.8.1 (September 2017). https://doi.org/10.5281/zenodo.883859

Winkler, A.M., Ridgway, G.R., Webster, M.A., Smith, S.M., Nichols, T.E., 2014. Permutation inference for the general linear model. Neuroimage 92, 381–397.

Winkler, A.M., Webster, M.A., Vidaurre, D., Nichols, T.E., Smith, S.M., 2015. Multi-level block permutation. Neuroimage 123, 253–268.

Yarkoni, T., Markiewicz, C.J., de la Vega, A., Gorgolewski, K.J., Salo, T., Halchenko, Y.O., McNamara, Q., DeStasio, K., Poline, J.-B., Petrov, D., Hayot-Sasson, V., Nielson, D.M., Carlin, J., Kiar, G., Whitaker, K., DuPre, E., Wagner, A., Ivanov, A., Tirrell, L.S., Jas, M., Hanke, M., Poldrack, R., Esteban, O., Appelhoff, S., Holdgraf, C., Staden, I., Rokem, A., Thirion, B., Kleinschmidt, D.F., Lee, J.A., Visconti di Oleggio Castello, M., Notter, M.P., Roca, P., Blair, R., 2019. bids-standard/pybids: 0.7.1. https://doi.org/10.5281/zenodo.2555449

Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D., 2011. Large-scale automated synthesis of human functional neuroimaging data. Nat. Methods 8, 665–670.

Yeo, B.T.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., Fischl, B., Liu, H., Buckner, R.L., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. 106, 1125–1165.

Zijdenbos, A.P., Forghani, R., Evans, A.C., 2002. Automatic "pipeline" analysis of 3-D MRI data for clinical trials: application to multiple sclerosis. IEEE Trans. Med. Imaging 21, 1280–1291.

Supplementary Table 1: the expected input directory structure (two options)	
The expected input structure (BIDS raw)	
Description	Filename
T1w image	sub-{subject}/anat/sub-{subject}_T1w.nii.gz sub-{subject}/anat/sub-{subject}_T1w.json
Functional image	sub-{subject}/func/sub-{subject}_task-{task}_bold.nii.gz sub-{subject}/func/sub-{subject}_task-{task}_bold.json
Alternative Derivative input structure (if working from BIDS common derivatives) *Note: is these file are not present in the /derivatives/ folder, then they will be generated by fMRIPrep*	
Freesurfer Outputs	derivatives/freesurfer/sub-{subject}/
Preprocessed T1w Image (i.e. the image the functional data is registered to)	derivatives/{func-pipeline}/sub-{subject}/anat/sub-{subject}_desc-preproc_T1w.nii.gz
Preprocessed functional image	derivatives/{func-pipeline}/sub-{subject}/func/sub-{subject}_task-{task}_space-T1w_desc-preproc_bold.nii.gz

Supplementary Table 2: Details of the ciftify output directory structure *Similar to the output structure of the HCP-Pipelines*			
Output type	3D space	2D space	Filename
T1w space volumes (NIfTI files)	T1w	N/A	derivatives/ciftify/sub-{subject}/T1w/ ├── T1w.nii.gz ├── wmparc.nii.gz ├── brainmask_fs.nii.gz ├── aparc+aseg.nii.gz └── aparc.a2009s+aseg.nii.gz
native (i.e. FreeSurfer mesh) surfaces	T1w	fsnative	derivatives/ciftify/sub-{subject}/T1w/Native/ └──sub-{subject}.{hemisphere}.{surface}.native.surf.gii

32k surfaces (GIFTI),	T1w	fsLR32k	derivatives/ciftify/sub-{subject}/T1w/fsaverage_LR32k/ └──sub-{subject}.{hemisphere}.{surface}.32k_fs_LR.surf.gii
labels & anatomical data (CIFTI)	T1w	fsLR32k	derivatives/ciftify/sub-{subject}/T1w/fsaverage_LR32k/ ├- sub-{subject}.{metric}.164k_fs_LR.dscalar.nii └─sub-{subject}.{parc}.164k_fs_LR.dlabel.nii
164k surfaces (GIFTI)	T1w	fsLR164k	derivatives/ciftify/sub-{subject}/T1w/ └──sub-{subject}.{hemisphere}.{surface}.164k_fs_LR.surf.gii
164k labels & anatomical data (CIFTI)	T1w	fsLR164k	derivatives/ciftify/sub-{subject}/T1w/ ├──sub-{subject}.{metric}.164k_fs_LR.dscalar.nii └──sub-{subject}.{parc}.164k_fs_LR.dlabel.nii
MNINonLinear space volumes (NIfTI)	MNI	N/A	derivatives/ciftify/sub-{subject}/MNINonLinear/ ├── T1w.nii.gz ├── wmparc.nii.gz ├── brainmask_fs.nii.gz ├── aparc+aseg.nii.gz └── aparc.a2009s+aseg.nii.gz
FSL transforms from T1w space	T1w to MNI (transforms)	N/A	derivatives/ciftify/sub-{subject}/MNINonLinear/xfms ├── NonlinearReg_fromlinear.log ├── Standard2T1w_warp_noaffine.nii.gz ├── T1w2StandardLinear.mat └── T1w2Standard_warp_noaffine.nii.gz *Note: if user supplied transform options are given, the input transforms will be copied to this location*
ROIs for subcortical resampling	MNI	N/A	derivatives/ciftify/sub-{subject}/ MNINonLinear/ ROIs/ ├── Atlas_ROIs.2.nii.gz -> *a copy (or shorcut) to the template ROIs* └── ROIs.2.nii.gz -> *derived from the participant freesurfer outputs, resampled to 2mm isotropic voxels*
Fsnative mesh surfaces, labels and anatomical results	MNI	fsnative	derivatives/ciftify/sub-{subject}/MNINonLinear/Native/ └──sub-{subject}.{hemisphere}.{surface}.native.surf.gii
native (i.e. FreeSurfer	MNI	fsnative	derivatives/ciftify/sub-{subject}/MNINonLinear/Native ├── sub-{subject}.{hemisphere}.roi.native.shape.gii -> *medial wall roi calculated from the cortical thickness data*

mesh) surfaces spheres for resampling from freesurfer to HCP's "fsaverageLR"			├── sub-{subject}.{hemisphere}.sphere.MSMSulc.native.surf.gii -> MSMSulc registration sphere (fsnative to fsLR via MSMSulc) ├── sub-{subject}.{hemisphere}.sphere.native.surf.gii -> *fsnative sphere from FreeSurfer* ├── sub-{subject}.{hemisphere}.sphere.reg.native.surf.gii -> *registration sphere (fsnative to fsaverage)* ├── sub-{subject}.{hemisphere}.sphere.reg.reg_LR.native.surf.gii -> *registration sphere (fsnative to fsLR via FreeSurfer registration)*
fsLR32k surfaces (GIFTI)	MNI	fsLR32k	derivatives/ciftify/sub-{subject}/MNINonLinear/fsaverage_LR32k/ └──sub-{subject}.{hemisphere}.{surface}.32k_fs_LR.surf.gii
fsLR32k labels & anatomical data (CIFTI)	MNI	fsLR32k	derivatives/ciftify/sub-{subject}/MNINonLinear/fsaverage_LR32k/ ├──sub-{subject}.{metric}.32k_fs_LR.dscalar.nii └──sub-{subject}.{parc}.32k_fs_LR.dlabel.nii
164k surfaces (GIFTI	MNI	fsLR164k	derivatives/ciftify/sub-{subject}/MNINonLinear/ └──sub-{subject}.{hemisphere}.{surface}.164k_fs_LR.surf.gii
164k surfaces (GIFTI), labels & anatomical data (CIFTI)	MNI	fsLR164k	derivatives/ciftify/sub-{subject}/MNINonLinear/ ├──sub-{subject}.{metric}.164k_fs_LR.dscalar.nii └──sub-{subject}.{parc}.164k_fs_LR.dlabel.nii
ciftify functional workflow outputs (are nested within the ciftify directory			
Preprocessed image and FSL transforms (NIFTI, .mat)	MNI		derivatives/ciftify/sub-{subject}/MNINonLinear/Results/task-{task}/ ├── task-rest_bold.nii.gz └── native/ ├── mat_EPI_to_T1.mat -> *FSL linear transform to T1w* └── mat_EPI_to_TAL.mat -> *FSL linear transform to MNI*
Preprossed functional data (CIFTI)	MNI	fsLR32k	derivatives/ciftify/sub-{subject}/MNINonLinear/Results/task-{task}/ ├── ciftify_subject_fmri.log ├── task-{task}_Atlas_s0.dtseries.nii └── task-{task}_bold_Atlas_s{fwhm}.dtseries.nii -> *if optional smoothing was specified*
T1w space is registered to the average of available T1w images MNI space is a non-linearly transformed to the MNI template {hemisphere} can be L = left or R = right {surface} can be white, pial, midthickness, inflated, vinflated, sphere or flat			

{metric} is the type of anatomical data converted from FreeSurfer (thickness, sulc, curvature) or a calculation of distortion after surface registration (ArealDistortion_FS, ArealDistortion_MSMSulc, EdgeDistortion_MSMSulc)
{parc} is the parcellation name converted from FreeSurfer (aparc, aparc.a2009s, BA_exvivo)
{task} is the fMRI task