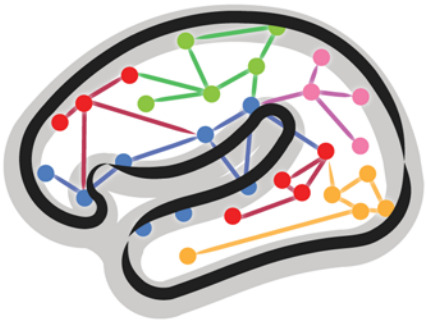


HUMAN Connectome PROJECT

WU-Minn HCP Quarter 3 (Q3) Data Release: Reference Manual

23 September 2013

Table of Contents

Introduction and overview 5

Should I be interested in the HCP Q3 data release? 5

What's in the HCP Q3 data release? 6

What's different from the Q2 release? 7

Multiple levels of data processing. 8

How do I access HCP data? 8

How can I analyze and visualize HCP data? 9

What's in store for future releases? 10

How were the HCP datasets acquired and processed? 10

What if I want to adapt HCP protocols to my own scanner and research projects? 11

Open access and restricted data - IMPORTANT! 11

What about technical support, bug reports, and feature requests? 12

HCP Data Release Updates Wiki and Data Patches 12

Gradient nonlinearities in unprocessed datasets 13

HCP Data Sizes 14

How to access HCP data 16

How to explore HCP data in ConnectomeDB 17

How to download HCP data 20

First time Aspera Download Setup 21

Handling downloaded HCP data 24

Connectome in a Box 25

MR scanner and other hardware 26

Summary of imaging protocols 28

Structural session 28

Resting-state fMRI (rfMRI) 28

Task-evoked fMRI (tfMRI) 29

Diffusion imaging (dMRI) 30

Full scanning protocols (PDF) 30

Summary of Behavioral and individual difference measures 31

Standard Operating Procedures (SOPs) 33
Standard two-day schedule for subject visits 33
Data in the Q3 release 34
Standard session structure 34
Physiological Data 35
Image Reconstruction and DICOM to NIFTI conversion 35
Initial QC and File Transfer to ConnectomeDB 37
Directory structure for unprocessed data 38
Preprocessing pipelines 42
Overview of HCP Minimal Processing Pipelines 42
HCP Structural Pipelines 43
HCP Functional Pipelines 45
HCP Diffusion Pipeline 47
Directory structure for preprocessed data 48
ICA-FIX denoised rfMRI data 51
Fix and Fix_extended datasets 52
Individual (within-subject) tfMRI analysis data 53
Group-average functional data (Unrelated 40 and Related 120) 54
Task fMRI group analysis 55
Resting-state group analysis and functional connectivity 56
File Names and Directory Structure for Group-average fMRI Data 57
Task fMRI files and protocol details 60
tfMRI scripts and data files 60
Details of tfMRI tasks 63
Behavioral and individual difference measures details 69
FreeSurfer measures 71
Descriptions and References for HCP Behavioral and Individual Difference Measures 71
Subject Information 71
MR Sessions 72
Alertness 72
Cognition 72

Emotion 76
Motor 77
Personality 78
Physical Function 78
Psychiatric, Substance Abuse, and Life Function 79
Sensory 80
References....82
Appendices 90
Appendix 1. HCP scan protocols 90
Appendix 2. Matlab code for voxel-wise correction of dMRI gradients 90
Appendix 4. HCP Protocol Standard Operating Procedures (SOPs) 90
Appendix 5: Data Acquisition Information for an Exemplar Subject 90
Appendix 6. Task fMRI E-Prime Key Variables....90
Appendix 7. Abbreviations for HCP Behavioral and Individual Difference Measures 90

Introduction and overview

This document provides information and guidance on how to use data released by the WU-Minn HCP consortium as of the Quarter 3 (Q3) release (September 2013). Given the richness of the HCP datasets and their utility for a wide range of research purposes, it is important that potential users understand what data are in the current release, how the datasets are organized, how they can be accessed, what has changed since the Q2 release (June 2013), and what's in store for future quarterly releases.

Should I be interested in the HCP Q3 data release? Over the next several years, the HCP will make available an enormous amount of data to enable exploration of human brain circuits, their relationship to individual behavior, and their heritability and genetic underpinnings. In the long term, the HCP data will thus be of interest to investigators in many fields, including neuroimaging, neuroanatomy, psychology, and network modeling - to name just a few.

The HCP aims to study 400 subjects per year, with a target of 1,200 subjects studied before the grant ends in 2015. Subjects from families with twins and non-twin siblings are being scanned on the same scanner using the same protocol for every subject. To date, HCP has provided a Q1 release (March 2013), a Q2 release (June 2013) and the current Q3 release (September 2013). The next data release is planned for winter 2014 (see below).

For Q3, no major changes were made between Q2 and Q3 in the preprocessing pipelines for each of the MRI modalities. However, an improved image reconstruction algorithm was used for diffusion (dMRI) and functional MRI (fMRI) data for Q3 subjects scanned on or after April 27, 2013 (a total of 26 subjects, many of whom are related). The new reconstruction method has slightly less spatial blurring of the unprocessed images, which is particularly important for dMRI analysis but of limited (negligible) impact on fMRI data. Therefore, we are also performing "retro" reconstruction and preprocessing for the dMRI data (but not the fMRI data) for all pre-April 27 Q3 HCP subjects; these additional dMRI datasets will be released soon, and retro reconstructed Q1-Q2 dMRI data will be part of the winter 2014 release.

The Q3 release includes imaging data from 87 extensively scanned participants (over 225 total for Q1-Q3). Besides the unprocessed and 'minimally preprocessed' data, the Q3 data also includes denoised rfMRI and task-fMRI analyzed datasets as described below.

The Q3 release includes 4D volume and surface grayordinate rfMRI data 'cleaned' of structured artifacts for all Q1-Q3 subjects. This is based on a new automatic method for denoising fMRI data developed by HCP investigators at FMRIB that combines independent component analysis (ICA) with a hierarchical fusion of classifiers approach called FIX (FMRIB's ICA-based Xnoiseifier) (Smith *et al*., 2013; Salimi-Khorshidi *et al.*, in preparation). See ICA-FIX denoised

rfMRI datasets below for more information on the ICA-FIX data and the steps involved in generating them.

Also available for all Q1-Q3 subjects are the results of individual task fMRI analysis based on cross-run (level 2) general linear modeling using FSL's FEAT. See Individual (within-subject) tfMRI analysis data below for more information.

Investigators interested in extensively processed functional data can now access data from two group-average analyses that include resting state functional connectivity MRI (fcMRI) and task functional MRI (tfMRI). The group-average analyses were performed on two groups of subjects that were released in Q1/Q2: 40 unrelated subjects (unrelated 40, or U40) and 120 subjects, in which many subjects are related to others in the group (the 'related' 120, or R120). See Group-average functional data below for more information.

As part of the group-average data we are, for the first time, releasing correlation matrices ("dense" functional connectome files) that are generated from the group-average analyses. Full correlation matrix and mean gray signal regressed versions of the group-average functional connectivity matrix are available for the U40 and R120 groups. Because of their large size (33 GB each), these dense connectome files are accessible as 1) download packages, separate from the rest of the group-average data, and 2) as remote file links in Connectome Workbench to retrieve the data as needed from ConnectomeDB without download (see Connectome Workbench tutorial for the Group-Average Data available for download in a zip archive with the data on the ConnectomeDB splash page).

If you are interested in accessing the HCP data, please continue reading the following important information.

What's in the HCP Q3 data release? The Q3 data include high-resolution MR scans from 87 healthy adults and four imaging modalities: structural images (T1w and T2w), resting-state fMRI (rfMRI), task fMRI (tfMRI), and high angular resolution diffusion imaging (dMRI). The majority of the behavioral and individual difference data acquired from each subject is also available. The current release expands the total (Q1-Q3) number of released healthy participant data for multiple imaging modalities and/or behavioral measures to 232. This includes 26 datasets with the complete HCP protocol (T1w and T2w MRI, rfMRI, tfMRI, dMRI, and behavioral measures) and over 200 more with all modalities except dMRI.

Please note that dMRI data is being released for 26 Q3 subjects only. dMRI data for these subjects was collected on or after April 27, 2013, when an improved image reconstruction algorithm for dMRI data was implemented. Unprocessed and minimally preprocessed dMRI datasets "retroactively" reconstructed with the new algorithm for all Q1-Q3 subjects will be available soon.

What's different from the Q2 release?

- ***Denoising of structured artifacts in rfMRI data***. Q1-Q3 resting-state fMRI data is available that has been cleaned of structured noise through a process that pairs independent component analysis (MELODIC) with a new FSL tool FIX, that automatically removes artifactual or "bad" components (Smith *et al*., 2013; Salimi-Khorshidi *et al.*, in preparation) (see ICA-FIX denoised rfMRI data below for details).

- ***Individual task fMRI analysis released on Q1-Q3 subjects***. Individual (within-subject) tfMRI data analyzed using FSL's FEAT is available for all Q1-Q3 subjects (see Individual (within-subject) tfMRI analysis data below for details).

- ***Group-average functional data on two groups.*** Group average resting-state functional connectivity and task fMRI data on 40 Unrelated (U40) and 120 subjects, some of whom are related, (R120) are available. These correlation matrix ("dense" connectome) files for full correlation and correlation with the mean gray signal regressed out, and also more compact PCA-based spatial eigenmap files that can be used to reconstruct the dense connectomes (see Group-average functional data (Unrelated 40 and Related 120) below for details).

- ***More behavioral and individual difference data***. More fields of open access data including mental state and sleep scoring are available. Restricted data including handedness, color vision, visual acuity and sensitive restricted data including drug screen data for all Q1-Q3 subjects are available for qualified investigators who are approved for restricted access. See Restricted data elements below for more details.

- ***Data updates wiki and Q1/Q2 data patch.*** A publically accessible wiki page documenting all known bugs and missing files in each HCP data release has been created and a patch script containing fixed data files for Q1/Q2 data is available on the ConnectomeDB splash page.

- ***New image reconstruction algorithm***. As noted above, since the Q2 release, we made improvements in the reconstruction algorithm that has been applied to dMRI and fMRI data for subjects scanned on or after April 27, 2013. The new reconstruction method has slightly less spatial blurring of the unprocessed images that will significantly improve the 26 Q3 subjects collected after April 27, 2013, some of whom are related. (see Image Reconstruction and DICOM to NIFTI conversion below for details).

  Note: For users of previously released (Q1 or Q2) HCP dMRI data using the older reconstruction algorithm, we strongly recommend not mixing previously released HCP diffusion data with the Q3 diffusion data for your analyses as there are differing noise (smoothness) properties between the Q1/Q2 and Q3 datasets.

Remember to be sure to keep track of version numbers when analyzing HCP data. Version numbers of the relevant HCP pipelines used to process the data in each HCP dataset are available in the release notes directory packaged in each dataset archive (see Preprocessing pipelines below). For example, the archive 100307_3T_tfMRI_EMOTION_preproc.zip contains the file release-notes-tfMRI_EMOTION_preprocess.txt:

100307_3T_tfMRI_EMOTION_preproc.zip

> Wed Jun 5 11:17:20 CDT 2013
> Structural Pipeline v2.0
> Functional Pipeline v2.0
> Execution 1

Multiple levels of data processing. As in previous releases, the Q3 imaging data includes "**unprocessed**" data (in NIFTI format) plus "**minimally preprocessed**" datasets. We encourage investigators to use the preprocessed datasets, in which spatial distortions have been minimized and data have been aligned across modalities and across subjects using appropriate volume-based and surface-based registration methods. The methods used for this preprocessing, (see Preprocessing Pipelines section below, were implemented by the HCP consortium through an intensive two-year refinement and optimization process. If your specific research interests make it important to use the unprocessed data, it is critical to be aware of certain characteristics that require special preprocessing (see Gradient Nonlinearities in Unprocessed Datasets).

The preprocessed structural datasets include cortical surfaces, myelin maps, and other data that are suitable for a variety of morphometric analyses. Denoised versions of individual resting-state fMRI scans (by ICA-FIX) are available in two datasets: the "FIXcomplete" dataset that includes 4D volume and grayordinate timeseries, the ICA components used in the denoising, and a report on the FIX run; the compact "FIX" dataset that includes only the grayordinate timeseries for each scan.

In addition, the HCP has more extensively processed group-average resting-state fMRI and task-fMRI data on groups of a) 40 unrelated Q1/Q2 subjects (U40) and b) 120 Q1/Q2 'related' subjects (R120, mixture of related and unrelated subjects). The extensively processed rfMRI data for the individual subjects used in this analysis are not being released at this time because key methods are still being evaluated and refined (e.g., temporal preprocessing and correcting for residual motion artifacts).

How do I access HCP data? Many HCP imaging files are large, and they must be maintained in a well-organized directory structure in order to prevent serious confusion or corruption when analyzing the data. For the Q3 release, the total amount of imaging data for all modalities is

about 3 terabytes (about 30 GB per subject), bringing the total amount of released imaging data from Q1-Q3 to ~6 terabytes.

HCP provides three ways to access part or all of these data:

(i) ***"Connectome in a Box"***. Investigators wanting access to data from many or all Q3 subjects (and/or Q1/Q2 subjects) and most or all imaging modalities are encouraged to use the Connectome in a Box option. A hard drive containing all of the available imaging data will be shipped within four weeks* of receiving payment for the hardware and delivery costs (see *Connectome in a Box* below). ***Note: For Q3, we are postponing fulfillment of Connectome in a Box until all "retro" reconstruction of dMRI data for Q3 is complete (October 2013).*** Institutions or imaging centers having many investigators who want access to the data should consider this option to facilitate sharing of HCP open access data (see HCP Data Sizes below); however, subject privacy provisions must be put in place in this case (see cautionary note about data sharing below).

(ii) ***Download Packages from ConnectomeDB***. The HCP provides an easy way to select standard packages based on modality of interest and a standard set of subjects. This includes selection of datasets by modality for one subject, 10, 40, and 80 unrelated subjects. Downloading is mediated by software that enables high-speed file transfer. Download times will depend on bandwidth and overall user demand, and there is a 500 GB limit on archive size that can be requested (see How to download HCP data).

(iii) The **ConnectomeDB** database user interface allows filtering of subjects based on a range of selection criteria (see How to access HCP data). The user can then download packages for just subjects selected by the filtering. Data filtering in ConnectomeDB is still under active development, but is increasing in its utility as the capabilities of the data mining interface are expanding and as the number of subjects increases with each quarterly release.

Why is access to some data restricted? Since the subjects include twins and their non-twin siblings, it will be important for many analyses to know the family status of the subjects under study. **Family status and other potentially sensitive information are part of the Restricted Data that is available only to qualified investigators after signing the Restricted Data Use Terms (see below). To protect the privacy of our HCP subjects, anyone wanting to use the Restricted Data MUST understand and agree to comply with the important constraints that are imposed.** To enable investigators to analyze data without being concerned about family structure issues, we have selectable groups of 10, 40 and 80 unrelated subjects for download (see How to download HCP data).

How can I analyze and visualize HCP data? Most of the preprocessed HCP data files are in standard formats supported by major neuroimaging platforms, including NIFTI (Neuroimaging Informatics Technology Initiative, see http://nifti.nimh.nih.gov) format for standard 3D or 4D

volume data and GIFTI (Geometry Informatics Technology Initiative) format for surface data. Importantly, the preprocessing pipelines for fMRI data (resting-state and task-fMRI) also include files that are in a 'grayordinate' representation that combines cortical surface vertices and subcortical gray matter voxels into a single file using the recently introduced CIFTI data format (see Glasser *et al.* 2013) These grayordinate CIFTI files offer major advantages in terms of efficiency in handling very large and complex datasets.

The Connectome Workbench platform can be used to visualize the outputs of HCP preprocessing and additional processing pipelines, including CIFTI, NIFTI and GIFTI files. Connectome Workbench is freely available (current version available at http://humanconnectome.org/connectome/get-connectome-workbench.html). A tutorial and associated dataset is available (requires ConnectomeDB login and signing of data use terms). Connectome Workbench also includes many command-line options (wb_command) that support a variety of analysis steps, including processing of CIFTI files.

What's in store for future releases? Future releases will not only add to the total number of subjects studied, but will include the more extensively analyzed datasets and will have progressively enhanced data mining and visualization capabilities. We also plan to make our data processing pipelines available in coming releases.

Commencing in the summer of 2013, a subset of 100 HCP subjects are undergoing MEG/EEG scans acquired at St. Louis University. Another subset of 200 HCP subjects will be scanned using a 7T scanner at the University of Minnesota, commencing in the winter of 2013. Data from these additional modalities will be made available in future releases.

The next HCP data release is scheduled for winter 2014 and will include data from at least 400 subjects scanned in the first year of the HCP production phase. This change from the quarterly release cycle followed up to the Q3 release will enable the generation and release of more extensively processed data, processing pipelines, and new data mining features on all released subjects, in addition to more unprocessed and minimally preprocessed datasets on newly scanned subjects. It also reflects a sense that quarterly releases will have reduced incremental scientific value as the total number of available subjects continues to increase.

How were the HCP datasets acquired and processed? Besides the documentation provided below, there are published descriptions of various aspects of the HCP (Van Essen *et al*., 2012; Marcus *et al*., 2011). Detailed descriptions of many HCP methods are available in eight papers (Van Essen *et al*., 2013, Ugurbil *et al*., 2013, Glasser *et al*., 2013, Smith *et al*., 2013, Barch *et al*., 2013, Sotiropolous *et al*., 2013, Marcus *et al*., 2013, and Larson-Prior *et al*., 2013) now available online in a special issue of NeuroImage in 2013. See http://www.humanconnectome.org/documentation/citations.html for updated lists of publications.

What if I want to adapt HCP protocols to my own scanner and research projects? Appendix I: Protocol Guidance and HCP Session Protocols provides advice and suggestions on this issue, which depends on the particular scanner system you are using as well as your research objectives.

Open access and restricted data - IMPORTANT!

The HCP provides imaging, behavioral, and demographic data from a large population of healthy adults. This poses special challenges for protecting the privacy of participants, especially because it is a family study including twins and their siblings. Unless these data are properly managed, there is a risk that some participants might be recognizable to family members and others (restricted data). In addition, some of the data elements (sensitive data) collected might harm or embarrass participants if they were to be inadvertently identified.

To protect the privacy of our participants, the HCP has implemented a two-tiered plan for data sharing, with different provisions for handling Open Access data and Restricted Data.

Open access data (all imaging data and most of the behavioral data) is available to those who register and agree to the Open Access Data Use Terms. This includes agreement to comply with institutional rules and regulations (see below), **which may require you to get your IRB's approval to use the data.**

Restricted data elements that could be potentially used to identify subjects include family structure (twin or non-twin status and number of siblings); birth order; age by year; handedness; ethnicity and race; body height, weight, and BMI; and a number of other categories (see Behavioral and individual difference measures details below). Restricted data also includes sensitive data (also called tier 2 restricted data) that could harm or embarrass participants if they were to be inadvertently identified, such as psychiatric information from the SSAGA telephone diagnostic interview, drug test results, and physiological data. We have built the distinction between the restricted and sensitive classes of Restricted data elements into the user interface of the ConnectomeDB dashboard used to explore and filter on the behavioral data (see below).

Each qualified investigator wanting to use Restricted Data must agree to the Restricted Data Use Terms. These terms explain how Restricted Data may and may not be used and shared, and they reiterate the need for compliance with institutional requirements. **They include major limitations on how Restricted Data can be incorporated into publications and public presentations**. Click here for additional information about required investigator qualifications and exemplar use case scenarios.

You must comply with your institutional rules and regulations regarding research on human subjects. Different institutions operate under different national and state laws.

Before you use any HCP data (either *Open Access or Restricted Data)* for research, you should ask your institution's Institutional Review Board (IRB), Ethics Committee, or office that oversees research on human subjects to determine whether you need institutional approval in order to use the data. On request, HCP can provide a certificate stating that you have accepted the Data Use Terms (Open Access or Restricted Access, as appropriate).

What about technical support, bug reports, and feature requests?

We anticipate a wide range of questions, suggestions, and discussion points as HCP data and software become freely available to the community. Users are strongly encouraged to join the HCP Users mailing list (hcp-users@humanconnectome.org) by signing up at http://www.humanconnectome.org/contact/ or by checking the appropriate box when registering to download HCP data.

Contributions to the hcp-users mailing list will be monitored and responded to by investigators and staff in the WU-Minn HCP consortium. Often this will entail prompt responses to answer questions or suggest solutions to technical problems. As with mailing lists for other brain mapping platforms (e.g., FSL, FreeSurfer), investigators outside the HCP consortium are encouraged to respond as well. Bug reports and feature requests will be entered into the issue tracking system used by HCP software developers by trained HCP staff.

If you are not currently a member of the hcp-users mailing list, you can submit a feature request, question, or suggestion directly at: http://humanconnectome.org/contact/feature-request.php. Comments submitted this way will be posted to the hcp-users mailing list so that all can benefit from the information provided.

HCP Data Release Updates Wiki and Data Patches

As mentioned above, a wiki page has been created to document Known Bugs/Missing files and Planned fixes in the HCP data released to date. This wiki will be updated frequently as we discover new bugs and files missing in the datasets. We recommend checking the wiki page before you do any major analyses with HCP data so that you are aware of any issues with files you may want to use. On the wiki, we also describe our plans to fix bugs/redistribute missing files and provide links to patches the HCP is releasing periodically to replace the files affected. Currently, a Q1/Q2 data patch Python script is being released to fix most issues found in the Q1/Q2 data. HCP patch scripts are written to automatically search through your downloaded or Connectome in a Box directory structure for files that need to be fixed and replaces them without modifying the structure. The Q1/Q2 data patch is available for download on the ConnectomeDB splash page.

Gradient nonlinearities in unprocessed datasets

If you intend to download the unprocessed NIFTI data, read on! Unusually large gradient nonlinearities are present in the raw images for ALL modalities, and it is important to correct for the spatial distortions they cause.

All HCP imaging data were acquired on a Siemens Skyra 3T scanner with a customized SC72 gradient insert that greatly improves the quality of diffusion imaging scans (the 'Connectome Skyra'). Higher performing gradients require compromises in bore diameter and gradient nonlinearities. Further, in custom-fitting the higher performing gradient set into a standard clinical system, technical limitations prevent centering of the subjects' heads in the bore isocenter. Consequently, the gradient nonlinearities associated with all Connectome Skyra scans exceed those of a conventional clinical 3T scanner. In the HCP processed datasets for all scan modalities (structural, fMRI, and dMRI), these distortions have been corrected for by spatially warping the images using gradient field information specific to the Connectome Skyra. The gradient unwarping code is available at https://github.com/ksubramz/gradunwarp/blob/master/Readme.md (Jovicich *et al.*, 2006). The gradient field nonlinearity coefficients for the Connectome Skyra are considered by Siemens to be proprietary information. To request access to these coefficients, please contact your Siemens collaboration manager or email Dingxin Wang at dingxin.wang@siemens.com.

For diffusion MRI, the gradient nonlinearities also cause voxel-by-voxel changes in the strength and orientation of the diffusion encoding gradients. Consequently, the effective b-values and b-vectors in all the primary data that you can download exhibit small variations from voxel to voxel. When analyzing the primary (unprocessed) datasets or the minimally preprocessed dMRI data you may use the code provided in Appendix 2: Matlab code for voxel-wise correction of dMRI gradients in your fitting routine in order to account for gradient spatial nonlinearities during fiber orientation estimation.

HCP Data Sizes

The MRI images collected for each HCP subject are notably high in spatial resolution and (for fMRI) in temporal resolution as well. Coupled with the long scan duration for each modality, the resultant image data files are very large. This "big data" is generally a good thing, enabling researchers to conduct types of data analyses that have never before been possible. However, the sheer size of the imaging data generated by the HCP protocol (total Q1- Q3 release data ~7 TB) requires some special considerations:

If you choose to obtain your data via download, prepare for long download times. The HCP using an optimized download client (Aspera) to make downloading as fast as possible, even when multiple users are downloading at the same time. However, depending on the amount of data you are downloading and your internet bandwidth, you may experience lengthy download times.

Housing the data will require a large amount of disk space. Even data on a few subjects (total data/subject: ~10 GB unprocessed, ~18 GB preprocessed, which includes ICA-FIX rfMRI) could fill up your hard drive quickly! Consider just how much data you need to "try out" the data, or do the analysis you want to do, and download just what you need. If you want all the data, consider ordering "Connectome in a Box" (see below).

Consider organizing investigators at your institution to obtain a single copy of the HCP data for local distribution. In many cases, several groups at a research

HCP Data Sizes (per Subject)

Session	Format	.zip File Size
Structural	Unprocessed	70 MB
	Preprocessed	1 GB
Resting State fMRI (each of 2 runs)	Unprocessed	2 GB
	Preprocessed	2.5 GB
	FIX (compact, both runs)	1.3 GB
	FIX_extended	2.8 GB
Task fMRI (avg per Task)	Unprocessed	490 MB
	Preprocessed	600 MB
	Analyzed	60 MB
(all 7 Tasks)	Unprocessed	3.4 GB
	Preprocessed	4.2 GB
	Analyzed	420 MB
Diffusion	Unprocessed	2.6 GB
	Preprocessed	790 MB
Group-Average on U40 and R120	Additionally Processed	200 MB
Group-Average "dense" connectomes (each of 4)	Additionally Processed	33 GB
Total (per Subject)	Unprocessed	7.4 GB
	Preprocessed	17.8 GB
	Both	25.2 GB
	Both+Analyzed	25.6 GB
Total (10 Subjects) (no Diffusion)	Unprocessed	101 GB
	Preprocessed	174 GB
	Both+Analyzed	279 GB
Total (40 Subjects) (no Diffusion)	Unprocessed	403 GB
	Preprocessed	697 GB
	Both+Analyzed	1.1 TB
Total (80 Subjects) (no Diffusion)	Unprocessed	836 GB
	Preprocessed	1.4 TB
	Both+Analyzed	2.3 TB
Total (All available datasets, including dMRI from 226 Subjects)	Unprocessed	2.2 TB
	Preprocessed	3.8 TB
	Both+Analyzed	6.3 TB

institution will want access to HCP data. We encourage users to organize efforts to obtain a local archive of the data via joint ordering of Connectome in a Box (recommended) or a single download that can be copied and used by many within your institution. With a modest amount of planning, groups that chose to do this will enjoy faster distribution and save valuable research time.

Important cautionary note about distributing HCP data: If you are going to distribute HCP data in any way, remember that **all data users**, even those who access HCP data through a local archive and not through ConnectomeDB, must register and sign the Open Access Data Use Terms on the HCP site before HCP data is used for research purposes. This step is required to be in compliance with regulations on human subjects research (other approval procedures in addition to signing the HCP Data Use Terms may also be required by your institution – see above). Those who distribute HCP data or use HCP data without signing the Data Use Terms are at risk for violation of their institution's human research policies and possible legal repercussions.

With these realities in mind, the HCP is distributing image data in two ways:

1. Via download through ConnectomeDB. HCP has made convenient data packages in 1 subject, 10, 40, or 80 unrelated subject groups by modality type to allow the user to "try out" the data without incurring a large download or data storage burden. The subjects in these set groups are unrelated to facilitate its use to investigators who want to analyze data without being concerned about family structure issues. The ConnectomeDB interface also allows the user to create their own subject groups of interest and download only the modalities necessary for analysis of that group.

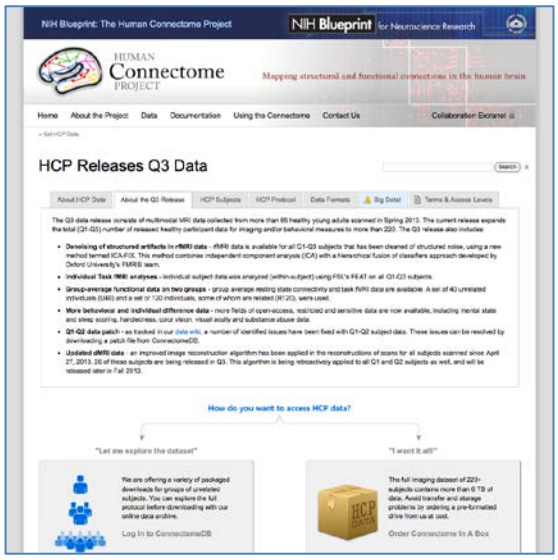
2. Q3 "Connectome in a Box". A complete Q3 Data Release preloaded onto 3TB hard drives, that can be ordered and shipped to you for a minimal cost (the cost of the hard drives + shipping).

In addition to the image data, the HCP has released additionally processed group-average resting-state fMRI and task-fMRI datasets acquired from 40 unrelated and 120 'related' subjects. This data is available both through download and as part of Q3 Connectome in a Box.

In addition to the download and Connectome in a Box option, because of their large size, we provide links to ~33 GB group-average functional connectivity data for visualization through the latest version of Connectome Workbench. An associated Connectome Workbench Tutorial for the Q3 Data Release is packaged with the Group-Average data available for download on the ConnectomeDB splash page.

As a reminder, the extensively processed data for the individual subjects used in this analysis are not being released for download at this time because key methods are still being evaluated and refined (e.g., temporal preprocessing and correcting for residual motion artifacts).

How to access HCP data



Go to the HCP data home page: http://humanconnectome.org/data/

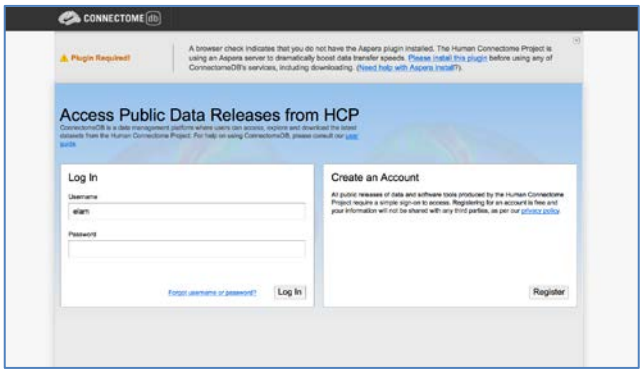
Here, you can get an overview of the Q3 data release, read the Open Access Data Use Terms, explore links to details of the HCP protocols, and access more useful resources for using HCP data.

Just below the top panel, you can decide if you would like to:

- Explore and/or download data in our database platform, ConnectomeDB

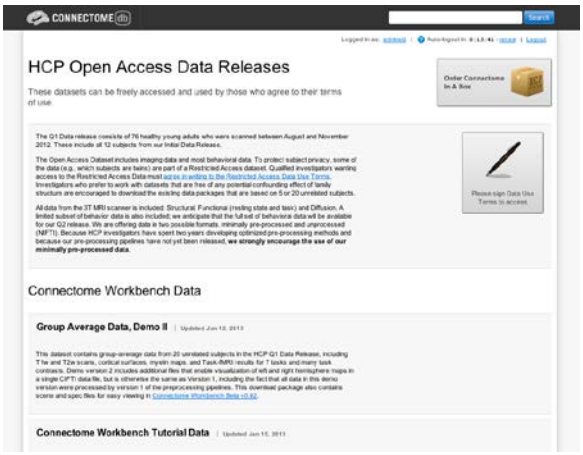
  or

- Order all the Q3 (and/or Q1/Q2) data via the "Connectome in a Box" option.



Clicking on either option takes you to a dialog to register for a ConnectomeDB account.

After you have created an account and are logged in, you must agree to the HCP Open Access Data Use Terms (required for access).



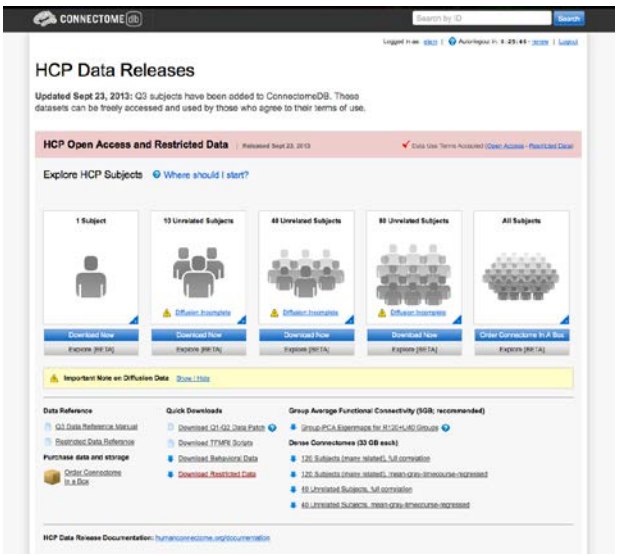
Once these steps are complete, you will launch the HCP Data Releases page (a.k.a. ConnectomeDB splash page) where you can begin to explore and download HCP data by selecting options for groups of subjects (see How to explore HCP data in ConnectomeDB and How to download HCP data below).

Previously released HCP data is also available in the 'More Datasets' section.

Note: Your login to ConnectomeDB times out every 30 minutes. Click the "renew" link to reset the time. (Auto-logout resets ConnectomeDB to the HCP Data Release page).

How to explore HCP data in ConnectomeDB

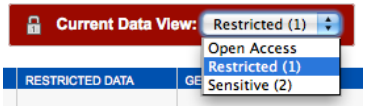
Upon login to ConnectomeDB, you are brought to the HCP Data Releases splash page, where you can:



- Download image data for groups of 1, 10, 40 or 80 unrelated subjects
    - Hover over the illustrations of the subjects to get a short description to help you decide which group is best for your purposes
- Explore the data for these groups in the ConnectomeDB user interface (currently in Beta testing, not all planned functionality is yet implemented)
- Download diffusion data for 26 Q3 subjects (click on Show in the yellow-highlighted Important Note on Diffusion Data)
- Download group-average functional data for the groups of 40 unrelated subjects and 120 'related' subjects
- Download group-average functional data for viewing in Connectome Workbench or other brain visualization software.
- Download the Data Patch for Q1-Q2 Data
- Download open access or restricted (if you have been approved for restricted access) behavioral data on all subjects
- Order all Q1-Q3 image data via Connectome in a Box
- Load a custom group of subjects that you created in a previous ConnectomeDB session
- Browse and download other HCP data that has been released previously

If you choose the "Explore [BETA]" option for 40 subjects, you will launch the HCP Dashboard filtered on 40 unrelated subjects with fully complete data.

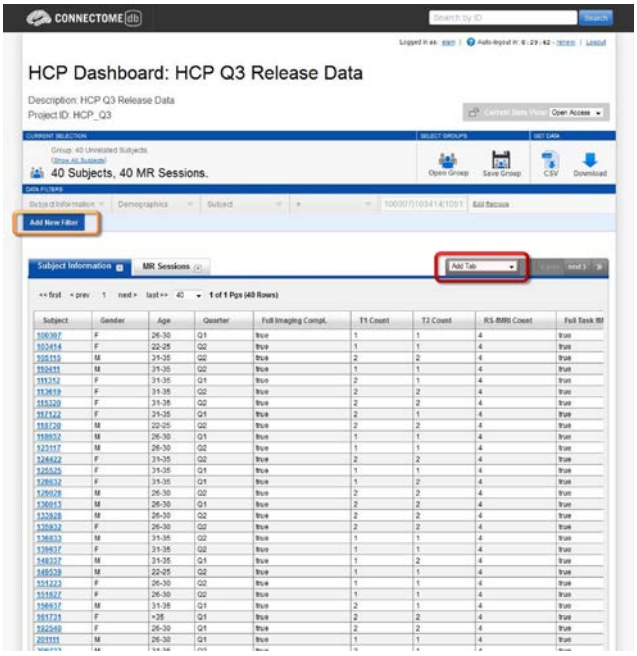
If you have Restricted data access, you will see a "Current Data View" box on the top right with a pulldown that allows you to choose the level of data you want to view: Open Access, Restricted (default, contains fields that are potentially



identifying), or Sensitive (contains fields potentially harmful or embarrassing to subjects). Restricted data view contains restricted data (column headers in red) and all open access data. Sensitive data view contains sensitive data (column headers in black) and restricted and open access data.

The HCP built this access level setting into the user interface to allow investigators to control what kinds of data appear for different use case scenarios. For example, one can show only Open Access data when giving a lecture to a class, but view Sensitive data when filtering subjects by positive drug test results for a research study on drug abuse.

Choose: Current Data View > Open Access.



To explore the behavioral data for these subjects, click on the “Add Tab” dropdown in the gray area on the right, midway down the page (circled in red in the figure, right).

Choose “Cognition” to see the scores for each of these 40 subjects for HCP measures of cognitive ability (see Behavioral and individual difference measures details). Descriptions of all the abbreviations for behavioral measured used in ConnectomeDB is available in Appendix VII – Abbreviations for HCP Behavioral and Individual Difference Measures.

Hover your mouse over the “Picture Sequence Unadjusted” column for more information on that measure and how it is scored.

You can filter the 40 unrelated subjects further. Here is an example:

- Under Data Filters, click on the “Add New Filter” button (circled in orange in the figure above)
- Click on the “Select Category” pulldown. and select “Cognition”.
- Select Assessment “Episodic Memory (Picture Sequence Memory)”.
- Select Attribute “NIH Toolbox Picture Sequence Memory Test Unadjusted Scale Score”,
- Select Operator “<”
- Type “95” in the final yellow field
- Click the “Apply” button

The filter will be applied and the filtered data will load into the subject list. For this filter, there are 8 Subjects out of the 40 unrelated subjects that have Picture Sequence Unadjusted Scale Scores that are less than 95.

To save a filtered group to reopen later, click on the “Save Group” button near at the top right and set a Group Name.

Your Groups are saved to your ConnectomeDB account and will be available when you login at a later time.

To download the behavioral and individual difference data for your selected group in a comma separated variable (.csv) file, click the “CSV” button.

To download the imaging data for your selected group, click the “Download” button.

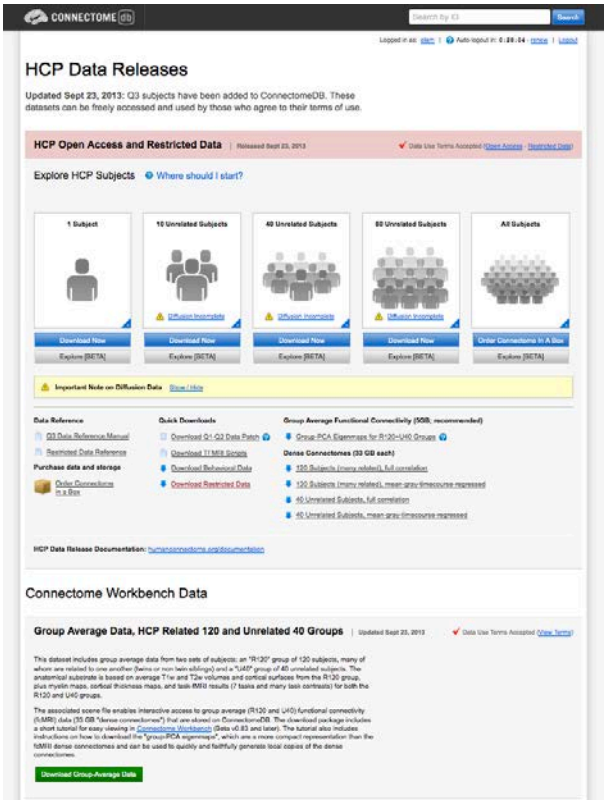
How to download HCP data

In order to facilitate access to the very large data files generated on every subject (see HCP Data Sizes above), we have prepackaged data into convenient group data packages. Datasets can be downloaded from ConnectomeDB in set packages or users may order all the Q3 data to be sent to them on hard drives (Connectome in a Box, see below).

Due to the large file sizes of the data, the HCP uses an Aspera server to dramatically boost data transfer speeds. Therefore, downloading data from the HCP requires you to have the Aspera plugin installed on your browser.

If you have not installed the plugin yet, a warning message will appear at the top of the browser window reminding you that you must Install the Aspera plugin before using any of ConnectomeDB's services, including downloading (see the second panel on page 12). Note: Installing the Aspera plugin to your browser requires you to restart your browser, therefore, we recommend installing the plugin *before* you explore HCP data in ConnectomeDB.

An Aspera install help guide is available on the HCP website. Note: One common problem with installing Aspera on the Mac platform for non-US users is that you must add US English to the Languages list (using the checkbox) to your System Preferences>Personal>Language & Text settings.

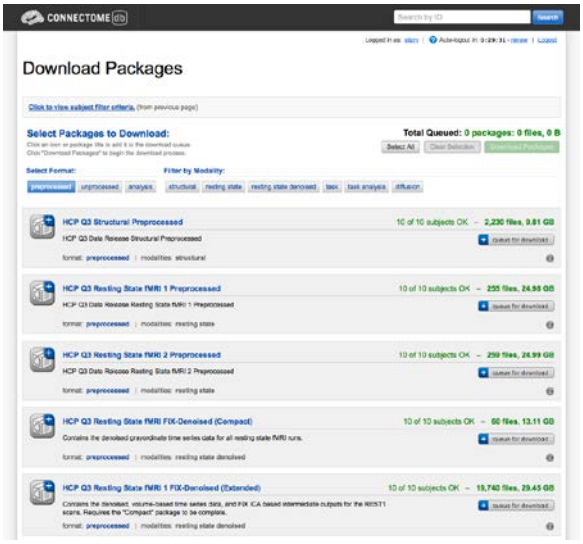


Upon login to ConnectomeDB, you are brought to the HCP Data Releases splash page, where you can:

- Download image data for groups of 1, 10, 40, or 80 unrelated subjects
    - Hover over the illustrations to get information to help you decide which group is best for your purposes
- Explore the data for these groups in the ConnectomeDB user interface (currently in Beta testing, not all planned functionality is yet implemented), see How to explore HCP data in ConnectomeDB
- Download diffusion data for 26 Q3 subjects (click on Show in the yellow Important Note on Diffusion Data section)
- Download group-average functional data

for the groups of 40 unrelated subjects and 120 'related' subjects

- Download group-average functional data for viewing in Connectome Workbench or other brain visualization software.
- Download the Data Patch for Q1-Q2 Data
- Download open access or restricted (if you have been approved for restricted access) behavioral data on all subjects
- Link to ordering all Q1-Q3 image data via Connectome in a Box
- Load a custom group of subjects that you created in a previous ConnectomeDB session
- Browse and download other HCP data that has been released previously



If you choose one of the "Download Now" options for 1, 10, 40 or 80 subjects, you will launch the Download Packages page (right).

First time Aspera Download Setup

If this is your first time downloading data from ConnectomeDB, **before you start selecting packages to download**, we highly recommend setting up the preferences in your Aspera Connect plugin to download to the appropriate place on your local or network hard drive.

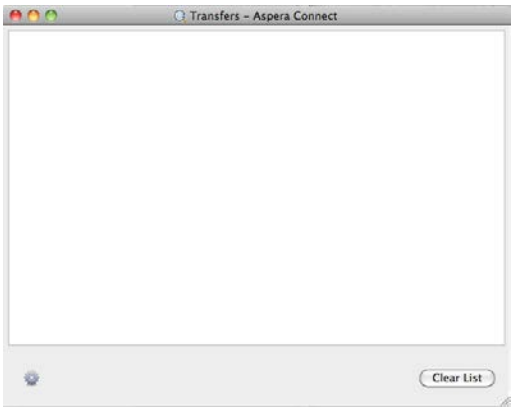
Note: the default Aspera Connect download location is your Desktop, which is likely not where you want your HCP data to go!
(Unfortunately HCP has no control over the Aspera defaults).

To set your Aspera Connect preferences:

- Search your computer for Aspera Connect using the Spotlight or Search functions in your operating system. Click on "Aspera Connect" to open the application.
- Open the Aspera Connect: Transfers window:
  - On Mac, in the Menu bar click on Window > Transfers, or use the shortcut **^T** (control + T)
  - On PC, in the Notification area at the very bottom right of your Desktop, click the ▲ button, then click the Aspera Connect icon:

- On Linux, extract the .gzip Aspera Connect plugin file you downloaded. Run the extracted script in a terminal window to launch Aspera Connect. The Aspera Connect icon should appear in your program notification area (often at the top or bottom right of your desktop), click on this icon and choose Transfers in the dropdown.

- Click the gear icon at the bottom left of the Aspera Connect — Transfers window to launch the Aspera preferences window.
- Click on the "Transfers" tab to setup where in your hard drive or network you would like your HCP data to download.
- Now, each time you download HCP data it will go to the directory you selected (unless you change the preference again).



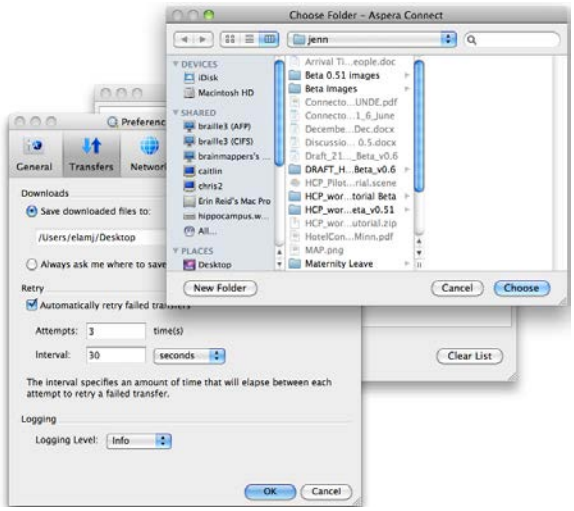
Downloading HCP data packages

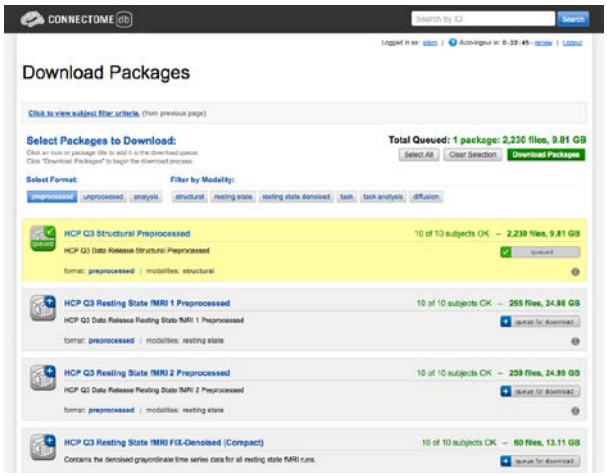
Now that you have Aspera setup to download to the correct location, go back to the "Downloading Packages" page in your browser.

Here, you can filter the image data available for download by format and modality by clicking the "Select Format" and "Filter by modality" buttons. To make selections for download, click the "add to queue" icons to the left of each dataset.

The total size of the data you have queued for download is at the upper and bottom right. Use this calculation to check against your available hard drive space to be sure you have enough space available for the files you intend to download.

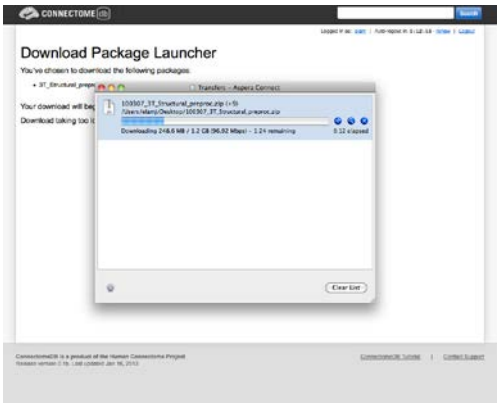
When you are done making your selections, click the Download Packages button at the upper right. A popup will remind you again what packages you selected and the total size of the files. Clicking "Download Now" will automatically open Aspera Connect and launch the download.





Note: If you have not setup Aspera Connect, as described above, by default the data will start downloading to your Desktop. If this not where you want the data to go, we recommend stopping the download by clicking the "X" button to the right of the download progress bar. Click the gear icon at the bottom left of the Aspera Connect: Transfers window to launch the Aspera preferences window (see above). Once you set the preferred download location, you will need to relaunch your download in ConnectomeDB, using your browser's back button to return to the "Downloading Packages" page.

If the download does not start automatically, use the link on the Download Package Launcher page in your browser to restart.



Once downloading has started, the blue bar in the Aspera Connect: Transfers window will show your progress (per subject file, not your overall progress).

Obviously, the more subjects and packages you selected for download, the longer your download will take (remember, this is big data!).

This is why we are offering the "Connectome in a Box" option for those who want data on many subjects (see below).

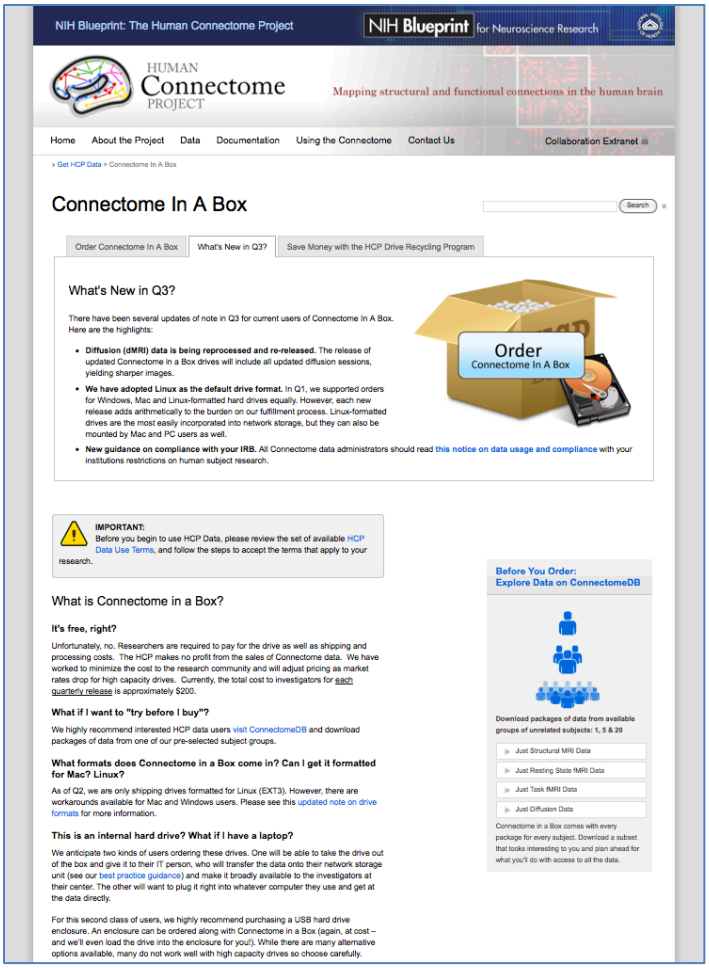
Handling downloaded HCP data

Now that you have downloaded the data you want, here's some tips for handling the data:

- Each download package consists of a .zip archive, and an md5 checksum. After you download the data you want, you can use the md5 file to verify the integrity of your downloaded file.
- To unzip the downloaded .zip archives, for Windows users we recommend using a utility such as 7-zip (available free) so that the directory structure of the unzipped files remains intact. The built-in unzip capabilities of Linux and MacOSX can be used, as they do not affect the directory structure of the unzipped files.
- Once the archives are downloaded and unzipped, many of the component files are gzipped (.gz files). You will need to unzip all.gz files in order to use them. For this, you need an application that is compatible with gzip. For Windows users, we again recommend 7-zip. Linux has support for gzip built in, and Mac users can use the Mac Gzip utility.
- Downloaded HCP data unzips to a set directory structure that is detailed below in Directory structure for unprocessed data, Directory structure for processed data, and in Appendix 3: File Names and Directory Structure for Unprocessed, Preprocessed, ICA-FIX denoised rfMRI, and Individual (within-subject) tfMRI analysis Data. The directory structure for downloaded HCP data is identical to what will be provided in HCP_Q3 Connectome in a Box.
- If you have downloaded a large amount of HCP data, we recommend sharing the data locally at your research institution. This will save download and organization time for your colleagues and relieve some of the burden on the ConnectomeDB download bandwidth available to other HCP users. Again, if you distribute HCP data in any way, **all users** (those who access HCP data) must register and sign the Open Access Data Use Terms in order to comply with regulations on human subjects research (see above).
- We are compiling a set of best practices for long-term data storage, which can be seen in our online documentation.

Connectome in a Box

Note: The Q3 Connectome in a Box order fulfillment is being postponed until October 2013, in order to allow for all dMRI images for Q3 subjects to be processed by the improved reconstruction algorithm and be included in the Q3 Connectome in a Box data.



An attractive option for getting HCP data is "Connectome in a Box". This option allows users to order hard drives containing the complete imaging data from chosen HCP data releases "at cost" (approximately $200/Quarter of data, plus shipping costs) and have them shipped to their address.

All versions of Connectome in a Box can be securely ordered from the HCP website.

Connectome in a Box contains both unprocessed and preprocessed NIFTI image data formats. The data is loaded onto a Linux-formatted SATA hard drive.

Once the user (or group of users) receive the drives, it can be added into a network storage unit and made available to all on the user's local network (provided **each user** registers and signs the Open Access Data Use Terms, see above). Alternatively, the drives can be simply plugged right into a computer and used as you would an external hard drive. For this option, we recommend purchasing a hard drive enclosure.

We anticipate as HCP data is updated, we will recommend updating versions of HCP Connectome in a Box via download of patches from ConnectomeDB. We recommend signing up for the HCP-Users mailing list at http://www.humanconnectome.org/contact/ to be apprised of these updates as they are released.

MR scanner and other hardware

Scanner hardware. All HCP subjects are scanned on a customized Siemens 3T "Connectome Skyra" housed at Washington University in St. Louis, using a standard 32-channel Siemens receive head coil and a "body" transmission coil designed by Siemens specifically for the smaller space available using the special gradients of the WU-Minn and MGH-UCLA Connectome scanners.

The scanner has a customized SC72 gradient insert and a customized body transmitter coil with 56 cm bore size (*diffusion*: Gmax = 100 mT/m, max slew rate = 91 mT/m/ms; *readout/imaging*: Gmax = 42 mT/m, max slew rate = 200 mT/m/ms). The HCP Skyra has the standard set of Siemen's shim coils (up to 2nd order).

Relative to a standard commercial Skyra, the customized hardware includes a gradient coil and gradient power amplifiers that together increase the maximum gradient strength from 40 mT/m to 100 mT/m on the WU-Minn 3T. This specifically benefits diffusion imaging, and on theoretical grounds (Ugurbil *et al*. 2013) it should provide significant gains over the standard 40 mT/m though not as much as the 300 mT/m customized gradients used by the MGH/UCLA HCP consortium.

Head motion and physiological monitoring. To address head motion, in most scan sessions we acquire dynamic head position information using an optical motion tracking camera system (Moire Phase Tracker, Kineticor). This system monitors head position precisely and in real-time using an infrared camera mounted in the scanner bore. Images of Moire interference fringes on a target affixed by clay to the bridge of the subject's nose are streamed in real time to a computer that displays the current position of the sensor and stores the positional information in a data file linked to the associated MRI scan. The stored file of head position and head movement can be used for post-hoc analyses. We also use it as a feedback trigger in dMRI scans to interrupt the movie being viewed whenever suprathreshold displacement and/or rapid head movement occur. Positional information can also be routed to the MRI scanner computer and can in principle be used prospectively to update the MRI slice prescription in real time (Zaitsev et al., 2006). However, prospective motion correction is not part of our 3 T HCP acquisition protocol because the technology became available only late in the HCP method development phase and was not sufficiently tested and developed before the data collection protocol was finalized (Van Essen *et al*., 2013).

We also acquire cardiac and respiratory signals associated with each scan, using a standard Siemens pulse oximeter placed on a digit and a respiratory belt placed on the abdomen. These signals are linked to scan onset using a trigger pulse generated by the pulse sequence. They are written to text files and assigned a unique file name that enables matching to the corresponding scan.

Visual projection and E-Prime computer. Visual stimuli are presented and participant responses are collected using a Dell Optiplex 790 computer, running an Intel Core i3-2100 with 8GB of RAM and 64-bit Windows 7 Enterprise SP1. The E-Prime version was E-Prime 2.0 Professional Production Release (2.0.10.242). Visual stimuli are projected with a NEC V260X projector onto a lucite screen at 1024x768 resolution, and viewed by the participant using a mirror mounted on the top of the head coil. Participant responses are registered on a customized fiber-optic button box.

Summary of imaging protocols

Structural, fMRI, and dMRI acquisitions were collected over 4 total imaging sessions, each approximately 1 hour in duration. Resting-state and task-fMRI data were collected in two sessions. Each session consisted of two resting-state acquisitions of approximately 15 minutes each, followed by task-fMRI acquisitions of varying durations (see below).

Vitamin E capsule on right side. A capsule of vitamin E was taped to the subject's right temple in every scan session, to enable definitive determination of the right side in the image data.

The following provides basic parameters for the main scan types in each session, and pertinent details about each session. A more complete set of imaging parameters can be found in the protocol exports from the scanner, available in Appendix 1. FOV positioning in all runs was handled in an automated manner using Siemens AutoAlign feature.

Structural session

Type	Series Description	Description	TR (ms)	TE (ms)	TI (ms)	Flip Angle	FOV (mm)	Voxel Size	BW (Hz/Px)	iPAT	Acquisition Time (min:sec)
T1w	T1w_MPR1	3D MPRAGE	2400	2.14	1000	8 deg	224x224	0.7 mm isotropic	210	2	7:40
T2w	T2w_SPC1	3D T2-SPACE	3200	565		variable	224x224	0.7 mm isotropic	744	2	8:24

Resting-state fMRI (rfMRI)

rfMRI data were acquired in four runs of approximately 15 minutes each, two runs in one session and two in another session, with eyes open with relaxed fixation on a projected bright cross-hair on a dark background (and presented in a darkened room). Within each session, oblique axial acquisitions alternated between phase encoding in a right-to-left (RL) direction in one run and phase encoding in a left-to-right (LR) direction in the other run.

Resting state images were collected with the following parameters:

Parameter	Value
Sequence	Gradient-echo EPI
TR	720 ms
TE	33.1 ms
flip angle	52 deg
FOV	208x180 mm (RO x PE)
Matrix	104x90 (RO x PE)

Parameter	Value
Slice thickness	2.0 mm; 72 slices; 2.0 mm isotropic voxels
Multiband factor	8
Echo spacing	0.58 ms
BW	2290 Hz/Px

Condition	Runs	Frames per run	Run Duration (min:sec)
REST (Resting-state)	4	1200	14:33

Task-evoked fMRI (tfMRI)

Following completion of rfMRI in each of the two fMRI scanning sessions, subjects were asked to complete tasks that were designed to activate a variety of cortical and subcortical networks. The following table provides a listing of the fMRI scans collected. For each scan type, one run was acquired with right-to-left phase encoding, and a second run with left-to-right phase encoding (in-plane FOV [field of view] rotation obtained by inverting both the RO (readout) and PE [phase encoding] gradient polarity).

T-fMRI data were acquired with the same EPI pulse sequence parameters as R-fMRI, except for the run duration information listed below. There are seven tasks (14 tfMRI runs) totaling one hour of total tfMRI scan time, with 3 tasks collected in one session and the remaining 4 tasks collected in another session.

Task	Runs	Frames per run	Run Duration (min:sec)
Working Memory	2	405	5:01
Gambling	2	253	3:12
Motor	2	284	3:34
Language	2	316	3:57
Social Cognition	2	274	3:27
Relational Processing	2	232	2:56
Emotion Processing	2	176	2:16

Scans necessary for field mapping correction were also acquired in each of the fMRI scan sessions.

EV (Explanatory Variable) files included in the dataset provide a set of conditions (and their associated timing) that can be used in the analysis of each task. An EPRIME_TAB.txt (tab-

delimited text spreadsheet) describes task parameters that can be used in various tfMRI analysis packages; see Task fMRI files and protocol details.

Diffusion imaging (dMRI)

Parameter	Value
Sequence	Spin-echo EPI
TR	5520 ms
TE	89.5 ms
flip angle	78 deg
refocusing flip angle	160 deg
FOV	210x180 (RO x PE)
matrix	168x144 (RO x PE)
slice thickness	1.25 mm, 111 slices, 1.25 mm isotropic voxels
Multiband factor	3
Echo spacing	0.78 ms
BW	1488 Hz/Px
Phase partial Fourier	6/8
b-values	1000, 2000, and 3000 s/mm^2

Other details: Diffusion gradients are monopolar. Oblique axial acquisitions alternate between right-to-left and left-to-right phase encoding directions in consecutive runs. Image reconstruction uses SENSE1 multi-channel (Sotiropolous *et al.*, 2013).

A full dMRI session includes 6 runs (each approximately 9 minutes and 50 seconds), representing 3 different gradient tables, with each table acquired once with right-to-left and left-to-right phase encoding polarities, respectively. Each gradient table includes approximately 90 diffusion weighting directions plus 6 b=0 acquisitions interspersed throughout each run. Diffusion weighting consisted of 3 shells of b=1000, 2000, and 3000 s/mm^2 interspersed with an approximately equal number of acquisitions on each shell within each run. The diffusion directions were obtained using a toolbox available from INRIA that returns uniformly distributed directions in multiple q-space shells. The directions are optimized so that every subset of the first M directions is also isotropic. References and the INRIA toolbox can be found at: http://www-sop.inria.fr/members/Emmanuel.Caruyer/q-space-sampling.php.

Full scanning protocols (PDF)

Complete scanning protocols for each imaging modality can be found in Appendix 1.

Summary of Behavioral and individual difference measures

HCP Subjects undergo many behavioral tests that are part of the NIH Toolbox battery and several Non-Toolbox behavioral measures. Behavioral measures are collected for the following domains (detailed descriptions of these measures can be found in Behavioral and individual difference measures details).

Alertness

- Cognitive Status
- Sleep

Cognition

- Episodic memory (Picture sequence and Verbal)
- Executive Function (Cognitive Flexibility and Inhibition)
- Fluid Intelligence
- Language (Reading decoding and Vocabulary comprehension)
- Processing Speed
- Self-regulation/Impulsivity
- Spatial Orientation
- Sustained Attention
- Working Memory

Emotion

- Emotion recognition
- Negative Affect
- Psychological Well-being
- Social Relationships
- Stress and Self-Efficacy

Motor

- Endurance
- Locomotion
- Dexterity
- Strength

Personality

- Five Factor Model (NEO-FFI)

Physical Function

- Blood Pressure*†
- Endocrine disorders*†
- Glucose Levels*†
- Hematocrit Levels*†
- Handedness*
- Height and Weight*†
- Menstrual Cycle and Hormonal Status*†
- Thyroid Function*†

Psychiatric, Substance Abuse, and Life Function

- Life function*
- Psychiatric Clinical Symptoms*†
- Parental Psychiatric or Neurological Illnesses*†
- Substance Use*
- Urine Drug Screen*

Sensory

- Audition
- Color Vision*
- Contrast Sensitivity
- Olfaction
- Pain
- Taste
- Visual Acuity*

*Available to users with access to Restricted Data only.

†To be included in a future data release.

Standard Operating Procedures (SOPs)

A major effort has been made to establish consistent procedures for all aspects of data acquisition and data processing. These are described in a set of Standard Operating Procedures (SOPs) that are included in Appendix 4. These SOPs provide a useful reference for investigators wanting to know more about exactly what was done. Any outdated versions of SOPs are retained and are available on request.

Standard two-day schedule for subject visits

Most subjects complete the full HCP protocol during a two-day visit. A typical schedule is shown below. Scan sessions are always done in the order shown below, unless problems with a scan necessitate a rescan in an extra imaging session.

Example WU-Minn HCP Subject Schedule

	Day 1	Day 2
7:30	Set-up and scanner QC	Set-up and scanner QC
8:30	Consent and Intake tests*	MR check and Diffusion session
	Mock scanner practice	
9:30	MR check and Structural session	
10:30	Non-Toolbox Behavioral Tests	NIH Toolbox Behavioral Tests
11:30	Lunch break	Lunch break
	Task practice and MR check	Task practice and MR check
12:30	R-fMRI session #1 (30 min)	R-fMRI session #2 (30 min)
1:30	T-fMRI session #1 (30 min)	T-fMRI session #2 (30 min)
2:30	Recognition task (ex-scanner)	Drug/alcohol tests, tobacco/alcohol 7d retrospective, satisfaction survey

*Drug/alcohol tests, mini mental state, sleep quality, menstrual info, HbA1c, TSH, hematocrit, blood for genotyping.

Data in the Q3 release

The third quarter of HCP data collection yielded 87 subjects with data from all or nearly all modalities and one with some behavioral tests. Combined with the Q1 and Q2 datasets, data from all or nearly all modalities is available for 226 subjects, including 26 datasets that include the complete HCP protocol.

The Q3 release contains dMRI scans for 26 subjects scanned on or after April 27, 2013 when the improved multiband image reconstruction algorithm was implemented (see below). Unprocessed and minimally preprocessed dMRI datasets based on the improved reconstructions for all Q1, Q2, and pre-April 27 Q3 subjects are currently being generated using the raw image data and will be released soon (by the scheduled winter 2014 data release).

In addition to the unprocessed and minimally preprocessed data of all modalities, the Q3 release contains additionally processed datasets that are further described in the linked sections below, including: ICA-FIX denoised rfMRI data, individual tfMRI data analyzed within-subject with FSL's FEAT, and group-average rfMRI and tfMRI data on 40 unrelated and 120 'related' subjects.

Standard session structure

Each subject was scanned in four regular sessions (~4 hours total), with the following shorthand labels:

Day 1:

- Session 1: Structural scans (two T1w and two T2w scans)
- Session 2: Two 15-min rfMRI scans (RL and LR phase encoding) and three tfMRI tasks (one RL and one LR scan for each task)

Day 2:

- Session 3: dMRI scan
- Session 4: Two 15-min rfMRI scans (RL and LR phase encoding) and the four remaining tfMRI tasks (one RL and one LR scan for each task)

Note: Counterbalancing the ordering of the different phase-encoding acquisitions for the resting-state fMRI scans (RL followed by LR in the first fMRI session; LR followed by RL in the second session) was adopted on October 1, 2012. Prior to that, both sets of rfMRI scans were acquired using the RL followed by LR order. The temporal order of actual scan acquisitions for each subject is reflected in the scan number in ConnectomeDB (and also as an explicit ScanOrder variable).

In some cases, scans were acquired in an extra session (coded session 5 or higher, as appropriate). This was done if the data quality was inadequate in the regular scan or if technical problems prevented a full set of acquisitions in the regular session.

Physiological Data

We also acquire cardiac and respiratory signals associated with each functional scan, using a standard Siemens pulse oximeter placed on a digit and a respiratory belt placed on the abdomen. These signals are linked to scan onset using a trigger pulse generated by the pulse sequence. They are written to text files and assigned unique file names that enable matching to the corresponding scan. These raw files were processed to create a file format expected by FSL's PNM. They are combined into a 3-column physio text file starting from the onset of the first trigger pulse through to the end of the last acquired frame. They are all sampled equally at 400 Hz, meaning there are roughly 288 samples per frame for functional images. In the aggregate file the first column is used for the trigger pulse information, the second for pulse oximeter (i.e., cardiac) information, and the third column for respiratory information.

The physiology datasets are included for most of the functional Q1-Q3 datasets. However, in some collected functional sessions for some subjects, physiological data is not yet available. A listing of the subjects and sessions for which a functional session was collected, but physiological data is missing, as of the Q3 release is documented on the public HCP Data Release Updates wiki site. In some of these cases, there was a problem with the physiological data collection that is recoverable (e.g., a gap in the timing between the trigger pulse and pulse/respiratory signal); in other cases, the physiological data was not collected or is otherwise unrecoverable. We are working to resolve these cases and for those subjects whose data is recoverable, the physiological data will be provided in a to be released patch.

We anticipate releasing cardiac datasets for diffusion imaging on all subjects with a future release. Ongoing HCP analyses will compare resting-state and task-fMRI data with vs. without regression of physiological signals. If warranted by these analyses, additional data files reflecting such corrective steps may be included in future quarterly data releases.

Image Reconstruction and DICOM to NIFTI conversion

Image reconstruction of multiband images was implemented in ICE (the image reconstruction environment of the Siemens scanners) using a customized set of image reconstruction algorithms (see Ugurbil *et al*. 2013). While a fMRI or dMRI scan is ongoing, data are streamed to an offline ("remote") reconstruction computer for unaliasing of the multiband data, which streams the processed data back to the scanner for final processing and viewing. In some cases of technical problems, reconstruction was done at a later time using a "retro" reconstruction.

In the spring of 2013, several months after the commencement of HCP Phase II scanning in August 2012, an improved reconstruction algorithm was developed that reduced spatial blurring in fMRI and dMRI acquisitions. This improvement was deemed worthy of upgrading the reconstruction algorithm, which occurred in late April 2013. In brief, the original reconstruction algorithm performed the separation of the multi-band multi-slice images after transforming the acquired fully sampled data to frequency space along the read-out direction, and now the multi-band multi-slice separation is performed in k-space.

The new reconstruction method is being applied retrospectively to existing dMRI scans that were saved in a compact but unreconstructed data format. These unreconstructed volumes are available for the vast majority of dMRI sessions before the transition; if unreconstructed volumes are not available for a particular dataset, the originally reconstructed images will be released with the reconstruction version noted.

For fMRI scans prior to April 2013, unreconstructed scans were not preserved owing to technical considerations related to the time required to complete reconstruction at the scanner. However, concurrent with the change in reconstruction algorithm in April 2013, the reconstruction switched to using graphics-processing-unit (GPU) accelerated reconstruction, which reduces reconstruction times considerably, thus permitting saving of unreconstructed fMRI data from April 2013 onward.

Details regarding the reconstruction method used will be clearly communicated along with each HCP data release.

As noted earlier in this document, improvements to the image reconstruction algorithms for diffusion images were made in and applied in the streaming reconstructions for Q3 subjects scanned on or after April 27, 2013 (26 subjects, many of which are related). dMRI data for these subjects are being released as unprocessed and preprocessed datasets as of the Q3 release date. For all earlier Q3 subjects, and the Q1 and Q2 subjects, “retro” reconstruction and preprocessing for the dMRI data is ongoing and these datasets will be released soon.

The reconstructed DICOM files are uploaded to the IntraDB database. Image defacing is carried out (Milchenko and Marcus, 2013) on all DICOM files containing potentially recognizable facial features. This includes structural (T1w, T2w), Bias receive and Bias transmit scans.

Conversion of DICOM files to NIFTI format was carried out using the dcm2nii utility. This utility is a component of the MRIcron suite of tools developed by Chris Rorden: (http://www.nitrc.org/frs/?group_id=152).

Initial QC and File Transfer to ConnectomeDB

Structural scans. An experienced rater has evaluated each structural scan and designated it as excellent, good, fair, poor, or unusable based on criteria related to tissue contrast, blurriness, or banding artifacts. To be included in the data release and in the structural pipeline processing, at least one T1w and one T2w rated as good or excellent must have been acquired in the same session (without the subject having exited the scanner). If these criteria are not met using the scans from the initial structural session, an extra scan session is acquired during the subject's visit or (if necessary) during a subsequent visit.

Usability of other scans. Scans from other modalities (fMRI and Diffusion) are included in the data release only if they meet specified criteria that make them useable in conjunction with other scans of the same modality acquired from the subject.

All scans are permanently stored in an internal database (IntraDB). Scans that are considered usable for preprocessing are transferred to the public-facing ConnectomeDB based on information encoded in a set of scan-specific fields in IntraDB. This process also entails a reorganization of data from session-specific directories (reflecting the fact that data are acquired in multiple scan sessions) to a composite subject-specific directory in ConnectomeDB, with subdirectories appropriate for the scan modality. Files of a given modality are grouped with ancillary files (e.g., bias fields and field maps) from the same session as needed in order to facilitate preprocessing using standardized scripts and pipelines. Many file names are modified and standardized as part of this transfer and reorganization process, again to facilitate preprocessing.

A subject-specific spreadsheet is included with the structural download packages, containing information that may be useful when analyzing the data. Appendix 5: Data Acquisition Information for an Exemplar Subject shows a representative example.

Many aspects of QC are described in Marcus *et al.*, 2013.

Directory structure for unprocessed data

Unprocessed datasets downloaded from ConnectomeDB are organized into subject-specific and modality-specific archives. When unpacked, they are moved into a <subject_id>/unprocessed/3T/ subdirectory (by creating it if not present, or placing into existing directories if present). If you have ordered Connectome in a Box, the directory structure is the same as the unpacked downloaded data. A full list of file names in each directory and subdirectory is provided in Appendix 3: File Names and Directory Structure for Unprocessed, Preprocessed, ICA-FIX denoised rfMRI, and Individual (within-subject) tfMRI analysis Data, Section A.

Unprocessed data for exemplar subject 100307 unpacks to the following directory structure:

100307/unprocessed/3T/
- 100307_3T.csv
- Diffusion
- rfMRI_REST1_LR
- rfMRI_REST1_RL
- rfMRI_REST2_LR
- rfMRI_REST2_RL
- T1w_MPR1
- T2w_SPC1
- tfMRI_EMOTION_LR
- tfMRI_EMOTION_RL
- tfMRI_GAMBLING_LR
- tfMRI_GAMBLING_RL
- tfMRI_LANGUAGE_LR
- tfMRI_LANGUAGE_RL
- tfMRI_MOTOR_LR
- tfMRI_MOTOR_RL
- tfMRI_RELATIONAL_LR
- tfMRI_RELATIONAL_RL
- tfMRI_SOCIAL_LR
- tfMRI_SOCIAL_RL
- tfMRI_WM_LR
- tfMRI_WM_RL

The 3T/ subdirectory signifies that these data were acquired on the 3T Connectome Skyra at Wash U. For the subjects that are later scanned at 7T (200 of the 1200), the 7T data will unpack to a 7T/ subdirectory. The .csv file contains the subject-specific spreadsheet mentioned in the preceding section, which provides useful metadata about the scans acquired for this particular subject.

Within each subdirectory, files are named as indicated by the exemplars below for several scan types.

T1w_MPR1/
- 100307_3T_AFI.nii.gz
- 100307_3T_BIAS_32CH.nii.gz
- 100307_3T_BIAS_BC.nii.gz

100307_3T_FieldMap_Magnitude.nii.gz
100307_3T_FieldMap_Phase.nii.gz
100307_3T_T1w_MPR1.nii.gz

T2w_SPC1/
100307_3T_AFI.nii.gz
100307_3T_BIAS_32CH.nii.gz
100307_3T_BIAS_BC.nii.gz
100307_3T_FieldMap_Magnitude.nii.gz
100307_3T_FieldMap_Phase.nii.gz
100307_3T_T2w_SPC1.nii.gz

The FieldMap, BIAS, and AFI scans are ancillary files copied into multiple structural subdirectories, in order to facilitate using standardized scripts and pipelines.

If multiple structural scans for a given subject passed the QC criteria and are included in ConnectomeDB, they will be in separate subdirectories T1w_MPR2/ and/or T2w_SPC2, with uniquely named structural files (e.g. 100307_3T_T1w_MPR2.nii.gz and/or 100307_3T_T2w_SPC2.nii.gz). Each subdirectory will include copies of relevant ancillary files needed for preprocessing.

The four 15-min rfMRI scans for each subject are downloaded into separate subdirectories, because the initial preprocessing is carried out separately for each scan.

rfMRI_REST1_LR
100307_3T_BIAS_32CH.nii.gz
100307_3T_BIAS_BC.nii.gz
100307_3T_rfMRI_REST1_LR_SBRef.nii.gz
100307_3T_rfMRI_REST1_LR.nii.gz
100307_3T_rfMRI_REST1_LR_Physio_log.txt
100307_3T_SpinEchoFieldMap_LR.nii.gz
100307_3T_SpinEchoFieldMap_RL.nii.gz

rfMRI_REST1_RL
100307_3T_BIAS_32CH.nii.gz
100307_3T_BIAS_BC.nii.gz
100307_3T_rfMRI_REST1_RL_SBRef.nii.gz
100307_3T_rfMRI_REST1_RL.nii.gz
100307_3T_rfMRI_REST1_RL_Physio_log.txt
100307_3T_SpinEchoFieldMap_LR.nii.gz
100307_3T_SpinEchoFieldMap_RL.nii.gz

Here, the ancillary files are SpinEchoFieldMap and BIAS scans acquired at the beginning of the rfMRI session and copied into both the rfMRI_REST1_LR/ and rfMRI_REST1_RL/ directories to simplify preprocessing scripts.

For tfMRI scans, the 14 scans (7 tasks x two phase encoding directions) are downloaded into separate subdirectories, because the initial preprocessing is carried out separately for each scan. For example, the working memory (WM) LR phase encoding imaging data:

tfMRI_WM_LR
100307_3T_BIAS_32CH.nii.gz

100307_3T_BIAS_BC.nii.gz
100307_3T_SpinEchoFieldMap_LR.nii.gz
100307_3T_SpinEchoFieldMap_RL.nii.gz
100307_3T_tfMRI_WM_LR.nii.gz
100307_3T_tfMRI_WM_LR_SBRef.nii.gz

and the relevant E-Prime data are in appropriate subdirectories:

tfMRI_WM_LR/LINKED_DATA/EPRIME
100307_3T_REC_run2_TAB.txt
100307_3T_WM_run2_TAB.txt

tfMRI_WM_LR/LINKED_DATA/EPRIME/EVs
0bk_body.txt
0bk_cor.txt
0bk_err.txt
0bk_faces.txt
0bk_nlr.txt
etc….

and the relevant Physiological data are in the PHYSIO subdirectory:

tfMRI_WM_LR/LINKED_DATA/PHYSIO/100307_3T_tfMRI_WM_LR_Physio_log.txt

Note that inclusion of an ancillary file in any given directory does not necessarily mean that it is used as part of the current HCP preprocessing pipelines. For example, the BIAS and AFI scans are not used currently (for any modality) for the preprocessed data.

The 6 dMRI scans (3 b-values x two phase encoding directions) are in a single Diffusion/ subdirectory:

Diffusion/
100307_3T_BIAS_32CH.nii.gz
100307_3T_BIAS_BC.nii.gz
100307_3T_DWI_dir95_LR_SBRef.nii.gz
100307_3T_DWI_dir95_LR.bval
100307_3T_DWI_dir95_LR.bvec
100307_3T_DWI_dir95_LR.nii.gz
100307_3T_DWI_dir95_RL_SBRef.nii.gz
100307_3T_DWI_dir95_RL.bval
100307_3T_DWI_dir95_RL.bvec
100307_3T_DWI_dir95_RL.nii.gz
100307_3T_DWI_dir96_LR_SBRef.nii.gz
100307_3T_DWI_dir96_LR.bval
100307_3T_DWI_dir96_LR.bvec
100307_3T_DWI_dir96_LR.nii.gz
100307_3T_DWI_dir96_RL_SBRef.nii.gz
100307_3T_DWI_dir96_RL.bval
100307_3T_DWI_dir96_RL.bvec
100307_3T_DWI_dir96_RL.nii.gz

100307_3T_DWI_dir97_LR_SBRef.nii.gz
100307_3T_DWI_dir97_LR.bval
100307_3T_DWI_dir97_LR.bvec
100307_3T_DWI_dir97_LR.nii.gz
100307_3T_DWI_dir97_RL_SBRef.nii.gz
100307_3T_DWI_dir97_RL.bval
100307_3T_DWI_dir97_RL.bvec
100307_3T_DWI_dir97_RL.nii.gz

Preprocessing pipelines

The HCP preprocessing pipelines (detailed in Glasser *et al.* 2013) are specially designed to capitalize on the high quality data offered by the HCP. The minimal preprocessing pipelines for structural, functional, and diffusion MRI accomplish many low level tasks, including spatial artifact/distortion removal, surface generation, cross-modal registration, and alignment to standard space. The final standard space makes use of a recently introduced CIFTI file format and the associated grayordinates spatial coordinate system. This allows for combined cortical surface and subcortical volume analyses while reducing the storage and processing requirements for high spatial and temporal resolution data.

Version 2 preprocessing pipelines. As in Q2, all preprocessing of the Q3 data was done using Version 2 of the preprocessing pipelines as detailed in Glasser *et al.* 2013. Therefore all currently available released data have been processed using the same pipelines to maximize compatibility for analyses going forward.

As a reminder, Q1 data was released in March 2013 using Version 1 of the preprocessing pipelines. In the Q2 release, we re-released all Q1 subjects data reprocessed using Version 2 preprocessing pipelines. Therefore, we recommend for users to update any data acquired from HCP before the Q2 release (June 2013).

Each dataset includes Release Notes (in a Subject_ID/release-notes directory) that state the version number(s) of the relevant HCP pipelines used to process the data. For example, the archive 100307_3T_tfMRI_EMOTION_preproc.zip contains the file release-notes-tfMRI_EMOTION_preprocess.txt:

100307_3T_tfMRI_EMOTION_preproc.zip

> Wed Jun 5 11:17:20 CDT 2013
> Structural Pipeline v2.0
> Functional Pipeline v2.0
> Execution 1

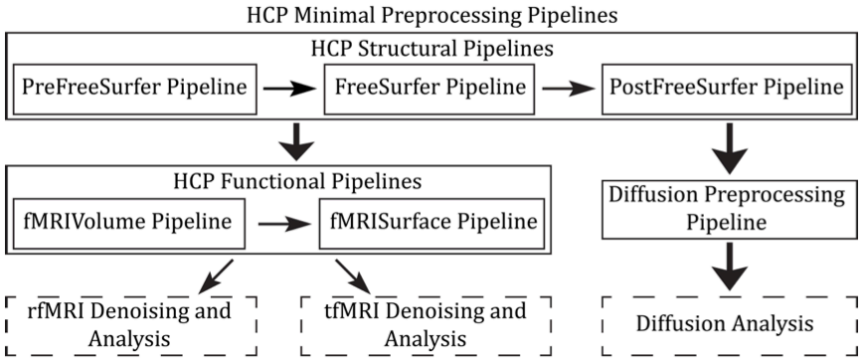
Improvements in processing pipelines that occur over time may warrant retrospective application of updated pipelines to previous data releases, as we have done with Version 2. The HCP will document any such changes in processing pipelines between versions.

Overview of HCP Minimal Processing Pipelines

Structural MR, fMRI, and dMRI datasets were processed by a set of 6 pipelines (see below) aimed at providing high quality volume and surface data. These pipelines use freely available software from the FSL (Jenkinson M *et al*., 2012; http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/) and

FreeSurfer (Dale AM *et al*., 1999; http://surfer.nmr.mgh.harvard.edu/) and Connectome Workbench (Marcus *et al*. 2013) image analysis suites, and are discussed in detail in (Glasser *et al.* 2013).

The figure to the right shows the overall workflow for preprocessing and data analysis in the HCP. The six minimal preprocessing pipelines include three structural pipelines (PreFreeSurfer, FreeSurfer, and PostFreeSurfer), two functional pipelines (fMRIVolume and fMRISurface), and a Diffusion Preprocessing pipeline.



Overview of HCP Minimal Processing Pipelines Glasser *et al.* 2013

A summary of the major processing steps for each pipeline is provided below and full descriptions are in Glasser *et al.* 2013. The following section, Directory structure for processed data summarizes the directory structure created by the preprocessing pipelines. Appendix 3B contains a complete list of all resulting preprocessing file names and subdirectories.

HCP Structural Pipelines

PreFreeSurfer pipeline

The main goals of the first structural pipeline, PreFreeSurfer, are to produce an undistorted "native" structural volume space for each subject, align the T1w and T2w images, perform a B1 (bias field) correction, and register the subject's native structural volume space to MNI space. Thus, there are two volume spaces in HCP data: 1) The subject's undistorted native volume space (rigidly "aligned" to the axes of MNI space), which is where volumes and areas of structures should be measured and where tractography should be performed, as this space is the best approximation of the subject's physical brain. 2) Standard MNI space is useful for comparisons across subjects and studies, particularly of subcortical data, which is more accurately aligned by nonlinear volume registration than is cortical data.

Steps:

1. Gradient distortion correction.
2. Coregistration and averaging of T1w and T2w runs.

3. ACPC (i.e. 6 dof) registration for T1w and T2w (FLIRT plus a customized script). The acpc registered distortion corrected T1w image is in the "native" volume space.
4. Initial brain extraction for T1w and T2w (FNIRT-atlas-based-brain-mask).
5. Field map distortion correction and registration of T2w to T1w using a customized FLIRT BBR algorithm.
6. Bias field correction using sqrt(T1w X T2w) (Rilling *et al*., 2011).
7. Atlas registration (FLIRT linear + FNIRT nonlinear to MNI152).

FreeSurfer pipeline

The FreeSurfer pipeline is based on FreeSurfer version 5.2, with a number of enhancements. The main goals of this pipeline are to segment the volume into predefined structures (including the subcortical parcels used in CIFTI), reconstruct white and pial cortical surfaces, and perform FreeSurfer's standard folding-based surface registration to their surface atlas (fsaverage).

Steps:

1. Downsampling of 0.7mm T1w to 1mm using splines.
2. Initial FreeSurfer steps (autorecon1 except -skullstrip).
3. Initialize FreeSurfer with skull registration using PreFreeSurfer brain mask for a more robust registration.
4. FreeSurfer skullstripping.
5. Further early FreeSurfer processing steps (-autorecon2 -nosmooth2 -noinflate2).
6. Adjustment of white matter surface placement using hires T1w (full 0.7mm data) with a customized mri_normalize, mris_make_surfaces, and fine tuning of T2w to T1w registration using bbregister.
7. Middle FreeSurfer steps (-smooth2 -inflate2 -sphere -surfreg -jacobian_white -avgcurv –cortparc).
8. Creation of pial surface using hires T1w (full 0.7mm data) and adjustment of pial surface placement using hires T2w (full 0.7mm data) with a customized mris_make_surfaces to remove vessels and dura. Then grey matter intensity normalization of T1w, regeneration of pial surface and adjustment with T2w.
9. Final FreeSurfer steps (-surfvolume -parcstats -cortparc2 -parcstats2 -cortribbon -segstats -aparc2aseg -wmparc -balabels -label-exvivo-ec).

PostFreeSurfer pipeline

The final structural pipeline, PostFreeSurfer, produces all of the NIFTI volume and GIFTI

surface files necessary for viewing the data in Connectome Workbench, along with applying the surface registration (to the Conte69 surface template (Van Essen et al., 2012b)), downsampling registered surfaces for connectivity analysis, creating the final brain mask, and creating myelin maps. There are three surface spaces in HCP data: the native surface mesh for each individual (~136k vertices, most accurate for volume to surface mapping), the high resolution Conte69 registered standard mesh (~164k vertices, appropriate for cross- subject analysis of high resolution data like myelin maps) and the low resolution Conte69 registered standard mesh (~32k vertices, appropriate for cross-subject analysis of low resolution data like fMRI or diffusion). These spaces are also shown in Figure 8. The 91,282 standard grayordinates (CIFTI) space is made up of a standard subcortical segmentation in 2mm MNI space (from the Conte69 subjects) and the 32k Conte69 mesh of both hemispheres (Figure 1). Following completion of the structural pipelines, the functional or diffusion pipelines may run.

Steps:

1. Creation of Connectome Workbench data files and spec files, creation of FreeSurfer segmentation brain mask, surface registration and downsampling, volume transformation of surfaces nonlinearly into MNI space. Creation of 2mm grayordinates space.
2. Creation of FreeSurfer ribbon files with full 0.7mm resolution data (masks of grey matter and white matter).
3. Myelin mapping and combination of all transforms for one step resampling of T1w and T2w images from original to MNI space (Glasser and Van Essen, 2011). Normalization of myelin maps (Glasser et al 2013).

HCP Functional Pipelines

fMRIVolume pipeline

The first functional pipeline, fMRIVolume, removes spatial distortions, realigns volumes to compensate for subject motion, registers the fMRI data to the structural, reduces the bias field, normalizes the 4D image to a global mean, and masks the data with the final brain mask. Standard volume-based analyses of the fMRI data could proceed from the output of this pipeline. Care is taken in the fMRIVolume pipeline to minimize the smoothing from interpolation, and no overt volume smoothing is done.

Steps:

1. Gradient distortion correction.
2. FLIRT-based motion correction using the “SBRef” volume as the target (Smith SM *et al.*, submitted).

3. TOPUP-based field map preprocessing using Spin echo field map (for each day of each BOLD run)
4. Distortion correction and EPI to T1w registration of the "SBRef" volume using a customized FLIRT BBR algorithm for distortion correction and bbregister to fine tune the registration. A Jacobian file representing the intensity modulations of the distortion correction is stored but not applied to the data.
5. One step spline resampling from the original EPI frames to atlas space including all transforms (motion, EPI distortion, EPI to T1w from FLIRT BBR, fine tuning of EPI to T1w with bbregister, nonlinear T1w to MNI).
6. Intensity normalization to mean of 10000 (like in FEAT) and bias field removal. Brain mask based on FreeSurfer segmentation.

fMRISurface pipeline

The main goal of the second functional pipeline, fMRISurface, is to bring the timeseries from the volume into the CIFTI grayordinates standard space. This is accomplished by mapping the voxels within the cortical grey matter ribbon onto the native cortical surface, transforming them according to the surface registration onto the 32k Conte69 mesh, and mapping the set of subcortical grey matter voxels from each subcortical parcel in each individual to a standard set of voxels in each atlas parcel. The result is a standard set of grayordinates in every subject (i.e. the same number in each subject, and with spatial correspondence) with 2mm average surface vertex and subcortical volume voxel spacing. These data are smoothed with surface (novel algorithm, see below) and parcel constrained smoothing of 2mm FWHM to regularize the mapping process. The output of these pipelines is a CIFTI dense timeseries that can be used for subsequent resting-state or task fMRI analyses.

Steps:

1. Cortical ribbon-based volume to surface mapping (from MNI space 2mm volume to MNI space native mesh surface).
    1. Voxels between white and pial surface are included in mapping. Voxels that are only partially between white and pial surface are weighted according to their partial volume that is between the two surfaces.
    2. Voxels with a temporal coefficient of variance greater than 0.5 standard deviations of their local neighborhood (sigma=5mm) are excluded from volume to surface mapping. The practical effect of this is to remove any small vessels and brain rim voxels remaining in the ribbon.
    3. Transformation of timeseries from native mesh to fs_LR registered 32k mesh (2mm average vertex spacing) in a single step.
2. Surface-based smoothing (2mm FWHM).

3. Subcortical parcel-constrained smoothing and atlas resampling (2mm FWHM).
4. Creation of dense timeseries (on a standard set of brainordinates).

HCP Diffusion Pipeline

The diffusion preprocessing pipeline does the following: normalizes the b0 image intensity across runs; removes EPI distortions, eddy-current-induced distortions, and subject motion; corrects for gradient-nonlinearities; registers the diffusion data with the structural; brings it into 1.25mm structural space; and masks the data with the final brain mask.

Diffusion processing pipeline

1. Basic preprocessing: Intensity normalization across runs, preparation for later modules.
2. 'TOPUP' algorithm for EPI distortion correction.
3. 'EDDY' algorithm for eddy current and motion correction.
4. Gradient nonlinearity correction, calculation of gradient bvalue/bvector deviation.
5. Registration of mean b0 to native volume T1w with FLIRT BBR+bbregister and transformation of diffusion data, gradient deviation, and gradient directions to 1.25mm structural space. Brain mask based on FreeSurfer segmentation.

Note, in the initial Q1 data release (March 2013), the preprocessed dMRI data were in the coordinate system of the individual diffusion scans. For the Q2 release (including the reprocessed Q1 dataset), this Q3 release, and all future releases, the data are aligned to the native structural space in order to facilitate various cross-modal comparisons.

Directory structure for preprocessed data

Preprocessed datasets downloaded from ConnectomeDB unpack into subdirectories under the i <subject_id>/ parent directory (by creating them if not present, or placing into existing directories if present). If you have ordered Connectome in a Box, the directory structure is the same as the unpacked downloaded data.

Preprocessing generates thousands of files, many of which are of little or no use. Therefore, in the preprocessed dataset package for each subject HCP releases a subset of files that are likely to be of general use to investigators. A list of file names in each directory and subdirectory is provided in Appendix 3B.

Structural data

- T1w/ contains T1w and T2w volume data
- T1w/Native/ contains FreeSurfer surfaces in their native mesh and original dimensions after rigid-body rotation to AC-PC alignment.
- MNINonLinear/ contains cortical surfaces and other data volumetrically registered to MNI152 space (using nonlinear FNIRT) followed by surface registration to Conte69 '164k_fs_LR' mesh (Van Essen *et al*. 2011) (via FreeSurfer fsaverage as an intermediate).
- MNINonLinear/Native/ replicates some of the files in T1w/Native/ but contains additional files used during surface-based registration.
- MNINonLinear/xfms/ contains files encoding the transformation between acpc and MNINonLinear volumetric space
- MNINonLinear/fsaverage_LR32k contains files spatially downsampled to a 32k mesh (average vertex spacing of ~2 mm), which is useful for analyses of rfMRI and dMRI connectivity data.

fMRI data

- MNINonLinear/Results/ contains volumetric and CIFTI grayordinates data for rfMRI scans (15 min each), motion parameters, and physiological data in four subdirectories,
    - rfMRI_REST1_RL
    - rfMRI_REST1_LR
    - rfMRI_REST2_RL
    - rfMRI_REST2_LR

plus volumetric and CIFTI grayordinates data and physiological data for 7 pairs of tfMRI scans (each task run once with right-to-left and once with left-to-right phase encoding.

- tfMRI_EMOTION_RL
- tfMRI_EMOTION_LR
- tfMRI_GAMBLING_RL
- tfMRI_GAMBLING_LR
- tfMRI_LANGUAGE_RL
- tfMRI_LANGUAGE_LR
- tfMRI_MOTOR_RL
- tfMRI_MOTOR_LR
- tfMRI_RELATIONAL_RL
- tfMRI_RELATIONAL_LR
- tfMRI_SOCIAL_RL
- tfMRI_SOCIAL_LR
- tfMRI_WM_RL
- tfMRI_WM_LR

Each of the directories above contain an .fsf file and a "EVs" directory containing explanatory variables. These files can be used to run first-level analyses in the FSL 5 program FEAT. The directories above also contain a TAB.txt file containing the timing of events, so that researchers might create their own explanatory variables of interest.

The MNINonLinear/Results/ directory contain seven other directories, one for each task:

- tfMRI_EMOTION
- tfMRI_GAMBLING
- tfMRI_LANGUAGE
- tfMRI_MOTOR
- tfMRI_RELATIONAL
- tfMRI_SOCIAL
- tfMRI_WM

These directories contain an .fsf file that can be used to run analysis across the two runs of each task. They also contain a script (prepare_level2_feat_analysis.sh) that must be run first, to set up appropriate registration matrices in the lower-level outputs.

Note: HCP's results of individual (within-subject) tfMRI analysis are available in download packages separate from the preprocessed datasets. When unpackaged, these data will integrate into the fMRI directory structure (MNINonLinear/Results/…). See Individual (within-subject) tfMRI analysis results data and Appendix 3D for more details.

- **Motion parameters.** Estimates of motion parameters are saved into two different files:

Movement_Regressors.txt and Movement_Regressors_dt.txt. The first file (Movement_Regressors.txt) contains 12 variables. The first six variables are the motion parameters estimates from a rigid-body transformation to the SBRef image acquired at the start of each fMRI scan.

- trans_x (mm)
- trans_y (mm)
- trans_z (mm)
- rot_x (deg)
- rot_y (deg)
- rot_z (deg)

The second six variables are temporal derivatives of those motion parameters

- trans_dx
- trans_dy
- trans_dz
- rot_dx
- rot_dy
- rot_dz

The second file (Movement_Regressors_dt.txt) contains 12 variables derived by removing the mean and linear trend from each variable in Movement_Regressors.txt

Diffusion Data

- Diffusion data includes diffusion weighting, direction, time series, brain mask, and gradient nonlinearity data, with the following files and directory structure:
  - T1w/ contains T1w_acpc_dc_restore_1.25.nii.gz (structural volume sampled at the same resolution as the diffusion data)
  - T1w/Diffusion contains the following subdirectories and files:
    - bvals (contains the diffusion weighting (b-value) for each volume)
    - bvecs (contains the diffusion direction (b-vector) for each volume)
    - data.nii.gz (preprocessed diffusion time series file)
    - nodif_brain_mask.nii.gz (brain mask in diffusion space)
    - grad_dev.nii.gz (contains the effects of gradient nonlinearities on the bvals and bvecs for each voxel)

ICA-FIX denoised rfMRI data

Here, we describe the approach we took to generate denoised resting-state fMRI data that is now available as additionally processed datasets for HCP subjects. For further details on this method see Smith *et al*., 2013; Salimi-Khorshidi *et al.*, in preparation, and Griffanti *et al.*, in preparation).

For rfMRI data, the HCP investigated the use of independent component analysis (ICA)-based artifact removal, to remove non-neural spatiotemporal components from each (highpass filtered) 15-minute run of rfMRI data. ICA is a powerful approach for decomposition of fMRI data as a summation of “good” and “bad” components, where each component comprises a weighted set of voxels (the component's spatial map), along with a single timeseries that is common to those voxels identified (Beckmann and Smith, 2004). Once ICA has identified a number of artifactual components, the data can be “cleaned” by subtracting these components from the data.

In working with the high quality/high temporal quantity HCP data, HCP investigators have developed a new approach that combines ICA with a more complex automated component classifier referred to as FIX (FMRIB's ICA-based X-noisifier) that has been specifically trained on HCP data. The FIX approach and initial results of classification accuracy are described in detail in (Salimi-Khorshidi *et al.,* in preparation), and the effects of the ICA + FIX cleanup (and optimal methods to remove the bad components from the data) are evaluated in detail in (Griffanti *et al.*, in preparation).

For HCP data, ICA is run using FSL’s MELODIC with automatic dimensionality estimation (MELODIC estimates how many components the given quality and quantity of data will support being separated from each other); this dimensionality is limited to a maximum of 250. These components are fed into FIX, which classifies components into “good” vs. “bad”. Bad components are then removed from the data. All of this is run using the volumetric data, rather than the grayordinate version of the data, because many artifacts are inherently 3D and do not respect tissue boundaries. The same set of artifactual processes is then removed from the (already created) grayordinates version of the data, by first applying the same highpass temporal filtering, and then regressing the bad components' timeseries out. For both volume and surface cleanup, the cleanup is done in a “non-aggressive” manner — both the good and bad component timeseries are regressed into the data, and then the resulting bad spatial maps are multiplied by the associated timeseries and subtracted from the original dataset. Thus, in this non-aggressive approach, only the unique variance associated with the bad components is removed from the data. Applying “aggressive” cleanup means removing all variance associated with the bad components, and not just the unique part, relative to non-artifact component timeseries. We have taken the more conservative non-aggressive approach to avoid removing variance of interest from the data, with the understanding that our cleanup will be less effective for more global types of noise whose variance is shared across good and bad components (this

decision will be revisited in future cleanup investigations). As part of this cleanup, we also used 24 confound timeseries derived from the motion estimation (the 6 rigid-body parameter timeseries, their backwards-looking temporal derivatives, plus all 12 resulting regressors squared — Satterthwaite *et al.*, 2013). The motion parameters have the temporal highpass filtering applied to them and are then regressed out of the data aggressively, as they are not expected to contain variance of interest.

Data from 25 HCP subjects (100 rfMRI runs) were hand-labeled, to train FIX. The average number of components per 15-minute run estimated by ICA was 229; of these, on average 24 components were hand-classified as "good" and the remainder as "bad". Leave-one-subject-out testing of the classifier (i.e., leaving one subjects' 4 runs out, training FIX on the other 19 subjects, and testing on the 4 runs left out) resulted in a mean accuracy of 99.3% in identifying "good" components correctly, and 99% accuracy in identifying "bad" components correctly (with corresponding median values being 100% and 99.3%). The balance between these two accuracies can be adjusted through the setting of a single controlling parameter; one has the option to choose to give greater importance to the accuracy of identification of good or of bad components. It is likely that an important factor in the very high accuracy of FIX classification is the high HCP data quality (and quantity — number of voxels and timepoints in each run), meaning that ICA is able to do a good job of separating multiple signal and noise components. See Smith *et al*., 2013 for examples of good and bad components and results of the cleanup. The cleanup reduces the resting-state network timeseries amplitude by ~ 30%, but, despite that reduction, does not reduce the effective group-level results (as judged by group-level statistics applied to both RSN spatial maps and network matrices).

Fix and Fix_extended datasets

Denoised rfMRI data is being released as two different datasets for the 4 rfMRI runs for each subject: the "fix" dataset contains ICA-FIX denoised rfMRI grayordinate surface timeseries only (compact version); the "fix_extended" dataset contains ICA-FIX denoised rfMRI volumetric timeseries, and ICA component data (for evaluation of FIX). Each of these datasets will unpackage into the <SubjectID>/MNINonLinear/Results/ subdirectories for the 4 rfMRI scans. A list of file names in each directory and subdirectory for both datasets is provided in Appendix 3C.

As an example, for the REST1_LR run:

rfMRI_REST1_LR_Atlas_hp2000_clean.dtseries.nii is the grayordinates timeseries file.

rfMRI_REST1_LR_hp2000_clean.nii.gz is the volumetric timeseries file.

We recommend using the fix compact package for general analysis of resting state data across many or all HCP subjects. If you would like to evaluate the FIX cleanup, we recommend downloading fix_extended for a limited number of subjects, in addition to fix (compact).

Individual (within-subject) tfMRI analysis data

The grayordinate surface-based results of individual (within-subject) tfMRI analysis are available in download packages for each of the seven tasks, separate from the preprocessed datasets. When unpackaged, these data will integrate into the fMRI directory structure (MNINonLinear/Results/…). See Appendix 3D for a list of file names in the directory and subdirectory structure.

The within-subject fixed-effects grayordinate-based analyses generate output files that index mean effects for an individual subject averaged across the two scan runs for a specific task. The output files are placed in the .feat subdirectory of the per task (not per run) directories for each task, e.g. MNINonLinear/Results/tfMRI_EMOTION/tfMRI_EMOTION_hp200_s4_level2.feat

The grayordinate-based output of a within-subject task fMRI was performed as follows. First, the minimal preprocessing pipeline handled motion correction, distortion correction, registration to standard space, and generation of a grayordinate time series for each individual task run. Then the grayordinate data for individual task runs was processed in a "Level 1" analysis that included high-pass filtering at 200s, and smoothing of 4 mm FWHM. This smoothing was conducted on the surface for the surface-based grayordinate data, and was constrained by segmentation (parcel) borders for the subcortical grayordinate data. Model estimation and temporal autocorrelation was handled using a version of FSL's 'film_gls' equipped to handle surface-based data in its internal spatial smoothing. The two runs for each task and subject were then combined in a "Level 2" fixed-effects analysis. The results of this "Level 2" analysis are what is contained in the task fMRI analysis dataset for each task and subject. These outputs are in standard HCP grayordinate space and can be viewed in Connectome Workbench. The surface-based grayordinates datasets are output in CIFTI format (e.g. file name: <SubjectID>_tfMRI_EMOTION_level2_hp200_s4.dscalar.nii).

The volume-based results of individual (within-subject) tfMRI analysis will be available later in the fall 2013. Volume-smoothed data is released on an interim basis even though it has drawbacks in statistical power and spatial resolution compared to surface-smoothed data. HCP anticipates that future releases of processed tfMRI data (beyond minimal pre-processing) will include only CIFTI-format, surface-smoothed data, once the bottlenecks to routine use of CIFTI are ironed out.

When available, the volume-based datasets will be available in NIFTI format (e.g. file name: <SubjectID>_tfMRI_EMOTION_level2_hp200_s4.nii). These outputs will be registered in standard volumetric space and will be viewable in Connectome Workbench.

Group-average functional data (Unrelated 40 and Related 120)

A number of processing pipelines are currently being implemented and refined by the HCP that make use of the minimally processed data to carry out further analyses at the individual and group levels. For example, the earlier HCP Q1 and Q2 Data Releases included an additionally processed demonstration dataset based on group-average data acquired from 20 unrelated subjects. With data from more subjects and improved analysis methods now in hand, we have generated a dataset containing two new functional group-average analyses based on subjects scanned in Q1/Q2: 40 Unrelated (U40) and 120 subjects, including many subjects that are related (R120), (one can view the subjects included in these analyses using the "Open group" function on the ConnectomeDB dashboard). This new group-average dataset is available for download through the links in the "Group Averages" section of the ConnectomeDB splash page.

This dataset includes:

- R120 group-average structural scans (T1w, T2w), cortical surfaces, and myelin maps (Related 120 data is being provided as a structural atlas due to its larger size)
- U40 and R120 group-average task fMRI analysis results for 7 tasks and many task contrasts
- CIFTI format files compatible for viewing and comparison in Workbench Beta v0.83
- Group-average functional connectivity matrix ("dense" functional connectome, the grayordinate × grayordinate full correlation matrix), computed with and without mean gray timecourse regression (MGTR), for the U40 and R120 groups (available as separate download packages due to their size).

  Because of their large size (33 GB each) these four dense functional connectome files are released separately from the rest of the group average data. Alternatively to besides downloading these large files, two other options are provided for the group-average dense connectomes:

  We recommend that you download the group-PCA eigenmaps package (5 GB for all four files combined: HCP_rfMRI_MIGP-GroupPCA_Eigenmaps_120related_40unrelated.zip). To create the dense connectomes locally, simply run wb_command -cifti–correlation <PCA Series> <dense connectome> -fisher-z. If you have less than ~35GB of RAM, use the -mem-limit option to limit the RAM usage. wb_command is the command line program distributed with Connectome Workbench.

  If you prefer to view these dense connectome files in Connectome Workbench, you do not need to download them. The data are accessible in Workbench by remote

access to the ConnectomeDB database. A Connectome Workbench tutorial for the Q1-Q2 group-average Related 120 and Unrelated 40 data is included as a PDF file in the **HCP_Q1-2_GroupAvg_Related120_Unrelated40_v1** archive available for download on the ConnectomeDB splash page.

The dataset is customized for viewing in the latest version of Connectome Workbench (http://www.humanconnectome.org/connectome/get-connectome-workbench.html), but most of the files are in standard NIFTI and GIFTI format and can also be viewed using other brain-mapping software platforms. A number of files, including the task fMRI files and the rfMRI functional connectivity files, are in a relatively new CIFTI file format (http://www.nitrc.org/projects/cifti/) that can incorporate surface vertices (both hemispheres) plus subcortical gray-matter voxels into a single file format. The CIFTI format is also used to represent surface maps (e.g., myelin maps) for both the left and right hemispheres in a single *.dscalar.nii file type and cortical parcellations for both hemispheres are contained in a single *.dlabel.nii file. The scenes file for Workbench distributed as part of this dataset makes use of these CIFTI files, but the downloaded datasets also include conventional hemisphere-specific GIFTI files (*.shape.gii; *.label.gii).

Task fMRI group analysis

For both the U40 and R120 group-averages, each of the seven tasks was processed using FLAME 1 (FMRIB's Local Analysis of Mixed Effects stage 1) mixed effects applied across subjects. We are currently releasing the grayordinates-based results that incorporate surface-constrained smoothing that respects the topology of the cortical sheet and the boundaries of subcortical anatomical parcels (see below and Barch *et al*., 2013 for details).

We will also be making the FLAME 1 data processed using the volume-smoothed data for the U40 and R120 group-averages available later in fall 2013. As noted above in Individual (within-subject) tfMRI analysis data, volume-smoothed data will be released on an interim basis even though it has drawbacks in statistical power and spatial resolution compared to surface-smoothed data. HCP anticipates that future releases of processed tfMRI data (beyond minimal pre-processing) will include only CIFTI-format, surface-smoothed data, once the bottlenecks to routine use of CIFTI are ironed out.

The grayordinates-based analysis began with outputs of the HCP “fMRISurface” pipeline (see above) in which the data from the cortical gray matter ribbon were projected onto the surface and then onto registered surface meshes with a standard number of vertices. Subcortical data were also projected to a set of subcortical grey matter parcel voxels, and when combined with the surface data formed the standard grayordinates space (Glasser *et al.* 2013). Smoothing of the left and right hemisphere time series and autocorrelation estimates (from FILM) were done on the surface using a geodesic Gaussian algorithm. Subcortical gray matter time series were smoothed within defined gray matter parcels. Smoothing of the grayordinate data was done in

two steps, but for a total of 4mm FWHM. Activity estimates were computed for the preprocessed functional time series from each run using a general linear model (GLM) implemented in FSL's FILM (Woolrich et al., 2001). Predictors were convolved with a double gamma "canonical" hemodynamic response function (Glover, 1999) to generate the main model regressors. To compensate for slice-timing differences and variability in the HRF delay across regions, temporal derivative terms derived from each predictor were added to each GLM and were treated as confounds of no interest. Subsequently, both the 4D time series and the GLM design were temporally filtered with a Gaussian-weighted linear highpass filter with a (soft) cutoff of 200 seconds. Finally, the time series was prewhitened within FILM to correct for autocorrelations in the fMRI data. Surface-based autocorrelation estimate smoothing was incorporated into FSL's FILM at a sigma of 5mm.

For the volume-based analysis to be released soon, spatial smoothing was applied using an unconstrained 3D Gaussian kernel of FWHM=4mm. The volume-based run-level analysis was carried out identically to the grayordinated-based analysis described above aside from spatial smoothing steps.

The GLM Model Design for each task is described in Barch *et al.*, 2013. Fixed-effects analyses were conducted using FSL's FEAT to estimate the average effects across runs within-subjects. Mixed-effects analyses treating subjects as random effects were conducted using FSL's FLAME to estimate the average effects of interest for the group. Volume-based group-level analyses were carried out using voxelwise comparisons in MNI space. The grayordinates-based subject-level and group-level analyses were done identically to the volume-based analysis except that cross-run and cross-subject statistical comparisons occurred in the better-aligned standard grayordinates space. NIFTI-1 matrices were processed separately for left and right surface and subcortical volume data, and surface outputs were converted to GIFTI at the conclusion of analysis. Subject-level and group-level z-statistic maps were combined from left and right hemisphere cortical and subcortical gray matter into the recently introduced CIFTI data format (http://www.nitrc.org/projects/cifti/; see (Glasser *et al.* 2013) for visualization using the Connectome Workbench platform (Marcus *et al.*, 2013).

Important note about the Related 120 (R120) dataset. As noted above, many of the 120 subjects in the R120 dataset are related. However, family structure *was not* taken into account in the tfMRI group average analyses of R120. Therefore, the associated standard errors and p-values are optimistic. Since the p-values are not valid, any false discovery rate (FDR) inferences based on them will not be valid either. We plan for standard errors and p-values to be revised in future related datasets once family structure is incorporated into the analysis.

Resting-state group analysis and functional connectivity

Resting-state fMRI data from the U40 and R120 groups of subjects were analyzed using steps outlined below that have yielded two versions of a group-average functional connectivity matrix

("dense" functional connectome). One version of the dense connectome uses the full correlation matrix. For the other, the mean gray signal was removed prior to grayordinatewise cross-correlation before generating the dense connectome.

For group average analysis, each of the four resting state runs for each subject in the group was cleaned (denoised) using 24-parameter motion regression followed by ICA+FIX denoising (Smith *et al.*, 2013; Salimi-Khorshidi *et al.*, in preparation). The resulting cleaned runs were combined across the 40 unrelated and 120 related subjects respectively using variance normalization of the timeseries (using the same approach as MELODIC (Beckmann and Smith, 2004)), followed by a group-PCA approach (MIGP - MELODIC's Incremental Group-PCA) that approximates full temporal concatenation of all subjects' data followed by PCA, outputting the strongest 4500 spatial eigenvectors (PCA components, weighted by the eigenvalues).

Important Note: When using seed-based methods to explore group-average functional connectivity data, some seeds have a "mound" of high correlation surrounded by a "moat" of low correlation. This "mound and moat" pattern reflects the interpolation methods used during image reconstruction and preprocessing; for additional information, see http://www.humanconnectome.org/documentation/mound-and-moat-effect.

File Names and Directory Structure for Group-average fMRI Data

The group-average functional dataset (HCP_Q1-2_GroupAvg_120related_40unrelated.zip) unpacks to HCP_Q1-2_GroupAvg_120related_40unrelated/ . It includes the following files:

- Group-average R120 surface/volume data and R120 and U40 myelin maps generated from the T1w and T2w images:
    - Q1-2_R120_AverageT1w_restore.nii.gz
    - Q1-2_R120_AverageT2w_restore.nii.gz
    - Q1-2_R120.thickness.32k_fs_LR.dscalar.gii
    - Q1-2_R120.corrThickness.32k_fs_LR.dscalar.gii
    - Q1-2_R120.curvature.32k_fs_LR.dscalar.gii
    - Q1-2_R120.L.inflated.32k_fs_LR.surf.gii
    - Q1-2_R120.L.midthickness_std.32k_fs_LR.shape.gii
    - Q1-2_R120.L.midthickness.32k_fs_LR.surf.gii
    - Q1-2_R120.L.very_inflated.32k_fs_LR.surf.gii
    - Q1-2_R120.MyelinMap_BC.32k_fs_LR.dscalar.nii
    - Q1-2_R120.R.inflated.32k_fs_LR.surf.gii
    - Q1-2_R120.R.midthickness_std.32k_fs_LR.shape.gii
    - Q1-2_R120.R.midthickness.32k_fs_LR.surf.gii
    - Q1-2_R120.R.very_inflated.32k_fs_LR.surf.gii
    - Q1-2_R120.sulc.32k_fs_LR.dscalar.nii
    - Q1-2_U40.thickness.32k_fs_LR.dscalar.gii

  - Q1-2_U40.corrThickness.32k_fs_LR.dscalar.gii
  - Q1-2_U40.MyelinMap_BC.32k_fs_LR.dscalar.nii

- Group-average Task fMRI grayordinates-based FEAT-level 3 analyzed files (CIFTI format):
  - HCP_Q1-Q2_R120_tfMRI_ALLTASKS_level3_hp200_s4.dscalar.nii
  - HCP_Q1-Q2_U40_tfMRI_ALLTASKS_level3_hp200_s4.dscalar.nii

    These are CIFTI files in which spatial smoothing (4mm FWHM) was constrained to 'grayordinates' (cortical surface vertices and subcortical voxels). By avoiding spatial blurring across anatomical compartments (gray matter vs white matter and CSF) and across sulcal banks, this provides the highest spatial fidelity attainable using these data.

- Reference files generated from other studies:
  - parcellations_VGD11b.L.32k_fs_LR.dlabel.nii
  - parcellations_VGD11b.L.32k_fs_LR.border
  - parcellations_VGD11b.R.32k_fs_LR.border
  - Parcels_GV11.L.32k_fs_LR.foci
  - Parcels_GV11.R.32k_fs_LR.foci
  - RSN-networks.32k_fs_LR.dlabel.nii

The parcellations_VGD11b.L.32k_fs_LR.dlabel.nii and .border files are derived from a composite cortical parcellation containing 52 distinct areas accurately mapped to the fs_LR atlas surface and based on architectonic or retinotopic fMRI maps (Van Essen *et al.* 2011). Abbreviations associated with labels of cortical areas (FRB08, OFP03, etc.) refer to the publication that defined a particular cortical area in the composite map (cf. Table 3 of Van Essen *et al*. 2011).

The Parcels_GV11.*.32k_fs_LR.foci files are provided to display of foci located at the center of mass of the probabilistic cytoarchitectonic areas (originally defined by Amunts and Zilles) used by Glasser and Van Essen 2011 for comparison to areas defined by myelin mapping.

The RSN-networks.L.32k_fs_LR.dlabel.nii file displays resting state network cortical parcellation maps from Yeo *et al.* 2011 (7 and 17 network maps) and the Resting State network consensus communities (with and without gaps in the communities filled) from Power *et al.* 2011.

- Files for easily opening Q1-2 Group Average data in Connectome Workbench:
  - HCP_Q1-2_120related_40unrelated_DATA.32k_fs_LR.wb.spec
  - HCP_Q1-2_120related_40unrelated.scene

- Release Notes file:

- ReleaseNotes_HCP_Q1-2_GroupAvg_120related_40unrelated.txt

To be released later in fall 2014, Group-average Task fMRI volume-based FEAT-level 3 analyzed files (CIFTI and NIFTI formats):

- Q1-Q2_120related_tfMRI_ALLTASKS_level3_hp200_s4_vol-smooth.dscalar.nii
- Q1-Q2_40unrelated_tfMRI_ALLTASKS_level3_hp200_s4_vol-smooth.dscalar.nii

  These are also in CIFTI format, but spatial smoothing (4mm FWHM) was done in standard volume space that includes blurring across sulcal banks and between gray-matter and non-gray-matter voxels.
- Q1-Q2_120related _tfMRI_ALLTASKS_level3_hp200_s4_vol-smooth.nii.gz
- Q1-Q2_40unrelated _tfMRI_ALLTASKS_level3_hp200_s4_vol-smooth.nii.gz

  These are the same as the preceding file but was converted from CIFTI to NIFTI format to enable viewing with platforms that are not yet CIFTI-compliant.

Task fMRI files and protocol details

Task-Evoked Functional Brain Activity

Our primary goals in including task-evoked functional MRI (tfMRI) in the HCP are to: 1) help identify as many "nodes" as possible that can guide, validate and interpret the results of the connectivity analyses that will be conducted on resting state fMRI (rfMRI), resting state MEG (rMEG) and diffusion data; 2) to allow a comparison of network connectivity in a task context to connectivity results generated using rfMRI; and 3) to relate signatures of activation magnitude or location in key network nodes to individual differences in performance, psychometric measures, or other phenotypic traits. To accomplish these goals, we developed a battery of tasks that can identify node locations in as wide a range of neural systems as is feasible within realistic time constraints (see Barch *et al.* 2013 for more detail).

We assess seven major domains that sample the diversity of neural systems that will be of interest to a wide range of individuals in the field, including: 1) visual, motion, somatosensory, and motor systems; 2) category specific representations; 3) working memory/cognitive control systems; 4) language processing (semantic and phonological processing); 5) social cognition (Theory of Mind); 6) relational processing; and 7) emotion processing. These tasks are described in more detail below and in Barch *et al*. 2013. Stimuli were projected onto a computer screen behind the subject's head within the imaging chamber. The screen was viewed by a mirror positioned approximately 8 cm above the subject's face.

tfMRI scripts and data files

Script files are run in E-Prime 2.0 Professional to present task fMRI stimuli and collect behavioral responses in the scanner. If you would like to run HCP tasks in your own research project, or examine the stimuli used in HCP tasks, the task stimulus script archives can be obtained through the Download Task FMRI Scripts link on the ConnectomeDB splash page.

Tab-delimited versions of the E-Prime data files (TAB.txt) are included in the released data. TAB.txt files are named according to the task condition that they describe and are contained within the directories for each of the two runs within each task (each phase encoding direction). A brief description of the key variables in those files can be found in Appendix 6: Task fMRI E-Prime Key Variables. The original edat files will not be available, because they may contain identifying information.

In addition, we include EV .txt files derived from the TAB.txt files (defined above) in the released data. EV files are explanatory variables (predictors) in FSL format (3-columns: onset, duration, and amplitude). There is a separate EV directory for each of the two runs within each task. Examples of the EV files for each task are detailed below.

The preprocessed data also includes .fsf files for each task. The fsf file is the setup or configuration file for running GLM-based fMRI analyses in 'FEAT' (FMRIB's Expert Analysis Tool: http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FEAT). The Lev1 .fsf files contain setup information necessary to run GLM analyses on the timeseries data for an individual scan run. Lev2 .fsf files contain setup information to run GLM analyses combining multiple scan runs for an individual participant. Lev3 .fsf files (not included in the release) can be created to setup GLM analyses across multiple participants.

HCP's results of individual (within-subject, level 2) tfMRI analysis for each task are available in download packages separate from the preprocessed datasets. These tfMRI datasets include a design.fsf file showing the setup information that was used in the FEAT GLM analysis for each task. See Individual (within-subject) tfMRI analysis results data and Appendix 3D for more details.

Here are examples of the EV files for each task and phase encoding direction in the appropriate tfMRI_[task_phaseencodingdirection]/LINKED_DATA/EPRIME directory (e.g. tfMRI_WM_LR/LINKED_DATA/EPRIME):

Working Memory

EVs/0bk_body.txt	BLOCKED	onset of 0Back body block condition
EVs/0bk_faces.txt	BLOCKED	onset of 0Back faces block condition
EVs/0bk_places.txt	BLOCKED	onset of 0Back places block condition
EVs/0bk_tools.txt	BLOCKED	onset of 0Back tools block condition
EVs/2bk_body.txt	BLOCKED	onset of 2Back body block condition
EVs/2bk_faces.txt	BLOCKED	onset of 2Back faces block condition
EVs/2bk_places.txt	BLOCKED	onset of 2Back places block condition
EVs/2bk_tools.txt	BLOCKED	onset of 2Back tools block condition
EVs /0bk_cor.txt	EVENT	onset of correct trials in 0Back blocks
EVs /0bk_err.txt	EVENT	onset of error trials in 0Back blocks
EVs /0bk_nlr.txt	EVENT	onset of trials in 0Back blocks with no response
EVs /2bk_cor.txt	EVENT	onset of correct trials in 2Back blocks
EVs /2bk_err.txt	EVENT	onset of error trials in 2Back blocks
EVs /2bk_nlr.txt	EVENT	onset of trials in 0Back blocks with no response
EVs /all_bk_cor.txt	EVENT	onset of correct trials in 0- and 2Back blocks
EVs/all_bk_err.txt	EVENT	onset of error trials in both 0- and 2Back blocks

Gambling

EVs/win.txt	BLOCKED	Onset of mostly reward blocks

EVs /loss.txt	BLOCKED	Onset of mostly loss blocks
EVs /win_event.txt	EVENT	Onset of reward trials
EVs /loss_event.txt	EVENT	Onset of loss trials
EVs /neutral.txt	EVENT	Onset of neutral trials

Motor

EVs /cue.txt	BLOCKED	Onset of task cues
EVs /lf.txt	BLOCKED	Onset of left foot blocks
EVs /rf.txt	BLOCKED	Onset of right foot blocks
EVs /lh.txt	BLOCKED	Onset of left hand blocks
EVs /rh.txt	BLOCKED	Onset of right hand blocks
EVs /t.txt	BLOCKED	Onset of tongue blocks

Language

EVs /story.txt	BLOCKED	Onset of story blocks
EVs /math.txt	BLOCKED	Onset of math blocks

Social Cognition

EVs /mental.txt	BLOCKED	Onset of mental interaction blocks
EVs /rnd.txt	BLOCKED	Onset of random interaction blocks
EVs /mental_resp.txt	EVENT	Onset of trials rated as mental interaction
EVs /other_resp.txt	EVENT	Onset of trials not rated as mental interaction

Relational Processing

EVs /relation.txt	BLOCKED	Onset of relational blocks
EVs /match.txt	BLOCKED	Onset of match blocks

Emotion Processing

EVs /fear.txt	BLOCKED	Onset of emotional face blocks
EVs /neut.txt	BLOCKED	Onset of shape blocks

Details of tfMRI tasks

Working Memory

The category specific representation task and the working memory task are combined into a single task paradigm. Participants were presented with blocks of trials that consisted of pictures of places, tools, faces and body parts (non-mutilated parts of bodies with no "nudity"). Within each run, the 4 different stimulus types were presented in separate blocks. Also, within each run, ½ of the blocks use a 2-back working memory task and ½ use a 0-back working memory task (as a working memory comparison). A 2.5 second cue indicates the task type (and target for 0-back) at the start of the block. Each of the two runs contains 8 task blocks (10 trials of 2.5 seconds each, for 25 seconds) and 4 fixation blocks (15 seconds). On each trial, the stimulus is presented for 2 seconds, followed by a 500 ms inter-task interval (ITI).

Conditions (Blocked)	
0-back faces	2-back faces
0-back places	2-back places
0-back tools	2-back tools
0-back body parts	2-back body parts

Conditions (Event-Related)	
0-back correct trials	2-back correct trials
0-back error trials	2-back error trials
0-back no response trials	2-back no response trials

Additional Contrasts. These event types can be combined to create two categories of contrasts.

Working Memory Contrasts
0-back contrast (activity combined across conditions 1-4)
2-back contrast (activity combined across conditions 5-8)
2-back versus 0-back contrast (2-back contrast minus 0-back contrast)

Category Contrasts
Faces contrast (0-back faces plus 2-back faces)
Places contrast (0-back places plus 2-back places)
Tools contrast (0-back tools plus 2-back tools)
Body contrast (0-back body plus 2-back body)

Potential Additional Event Related Contrasts: Researchers can also use the TAB.txt E-Prime data files to generate the following potential event-related contrasts:

1. Targets
    a. For 2-back tasks, targets are 2-back repeats
    b. For 0-back tasks, targets match the cue stimulus
2. Non-targets
    a. For 2-back tasks, non-targets are novel items
    b. For 0-back tasks, non-targets do not match the cue stimulus
3. Lures
    a. For 2-back tasks, lures are 1-back or 3-back repeats
    b. For 0-back tasks, lures are repeated stimuli that do not match the cue stimulus

Recognition Memory

After participants exit the scanner from the session that includes the working memory tasks, they are given a Remember, Know, New item recognition test for the faces and places that were presented during the working memory task. Responses to this recognition memory test can be used to create events to analyze the working memory trials as a function of whether the item was subsequently recognized (remember or know) or not (new). This is referred to as a subsequent memory analysis. The REC_run1_TAB.txt and REC_run2_TAB.txt files needed for this analysis are in the */tfMRI_WM_LR/ and */tfMRI_WM_RL/ data directories.

The recognition memory data are included for most of the Q1-Q3 datasets. However, in some subjects, recognition memory data was not available in the Q2 release. A listing of the subjects and sessions for which recognition memory data is missing for one or both of the working memory runs in the Q2 release is documented on the public HCP Data Release Updates wiki site. In some of these cases, there was a problem with the data collection that is recoverable; in other cases, the recognition memory data was not collected or is otherwise unrecoverable. We are working to resolve these cases and for those subjects whose data is recoverable, the recognition memory data will be provided in a to be released patch.

References for Working Memory: Localizer (Drobyshevsky *et al.* 2006); Reliable across subjects (Drobyshevsky *et al.* 2006) and time (Caceres *et al*. 2009).

References for Category-Specific Representations: Faces, Places, Tools and Body Parts: Localizer (Downing *et al.* 2001; Peelen and Downing 2005; Taylor *et al.* 2007; Fox *et al.* 2009); Reliable across subjects (Downing *et al.* 2001; Fox *et al.* 2009) and time (Peelen and Downing 2005; Kung *et al.* 2007).

Gambling

This task was adapted from the one developed by Delgado and Fiez (Delgado *et al*. 2000).

Participants play a card guessing game where they are asked to guess the number on a mystery card (represented by a "?") in order to win or lose money. Participants are told that potential card numbers range from 1-9 and to indicate if they think the mystery card number is more or less than 5 by pressing one of two buttons on the response box. Feedback is the number on the card (generated by the program as a function of whether the trial was a reward, loss or neutral trial) and either: 1) a green up arrow with "$1" for reward trials, 2) a red down arrow next to -$0.50 for loss trials; or 3) the number 5 and a gray double headed arrow for neutral trials. The "?" is presented for up to 1500 ms (if the participant responds before 1500 ms, a fixation cross is displayed for the remaining time), following by feedback for 1000 ms. There is a 1000 ms ITI with a "+" presented on the screen. The task is presented in blocks of 8 trials that are either mostly reward (6 reward trials pseudo randomly interleaved with either 1 neutral and 1 loss trial, 2 neutral trials, or 2 loss trials) or mostly loss (6 loss trials pseudo-randomly interleaved with either 1 neutral and 1 reward trial, 2 neutral trials, or 2 reward trials). In each of the two runs, there are 2 mostly reward and 2 mostly loss blocks, interleaved with 4 fixation blocks (15 seconds each).

Conditions (Blocked)
Mostly reward blocks
Mostly loss blocks

Conditions (Event-Related)
Reward trials
Loss trials
Neutral trials

References for Gambling Task: Reliable across subjects and robust activation in fMRI (Delgado *et al*. 2000; May *et al*. 2004; Tricomi *et al*. 2004; Forbes *et al*. 2009)

Motor

This task was adapted from the one developed by Buckner and colleagues (Buckner *et al*. 2011; Yeo *et al*. 2011)**.** Participants are presented with visual cues that ask them to either tap their left or right fingers, or squeeze their left or right toes, or move their tongue to map motor areas. Each block of a movement type lasted 12 seconds (10 movements), and is preceded by a 3 second cue. In each of the two runs, there are 13 blocks, with 2 of tongue movements, 4 of hand movements (2 right and 2 left), and 4 of foot movements (2 right and 2 left). In addition, there are 3 15-second fixation blocks per run. This task contains the following events, each of which is computed against the fixation baseline.

Conditions (Blocked)

Left finger blocks
Right finger blocks
Left toe blocks
Right toe blocks
Tongue movement

References for Motor Task: Localizer (Morioka *et al*. 1995; Bizzi *et al*. 2008; Buckner *et al*. 2011; Yeo *et al*. 2011).

Language Processing

This task was developed by Binder and colleagues (Binder *et al.* 2011) and uses the E-prime scripts provided by these investigators. The task consists of two runs that each interleave 4 blocks of a story task and 4 blocks of a math task. The lengths of the blocks vary (average of approximately 30 seconds), but the task was designed so that the math task blocks match the length of the story task blocks, with some additional math trials at the end of the task to complete the 3.8 minute run as needed. The story blocks present participants with brief auditory stories (5-9 sentences) adapted from Aesop's fables, followed by a 2-alternative forced-choice question that asks participants about the topic of the story. The example provided in the original Binder paper (p. 1466) is "*For example, after a story about an eagle that saves a man who had done him a favor, participants were asked, "Was that about revenge or reciprocity?"* The math task also presents trials aurally and requires subjects to complete addition and subtraction problems. The trials present subjects with a series of arithmetic operations (e.g., "fourteen plus twelve"), followed by "equals" and then two choices (e.g., "twenty-nine or twenty-six"). Participants push a button to select either the first or the second answer. The math task is adaptive to try to maintain a similar level of difficulty across participants. For more details on the task, please see (Binder *et al.* 2011).

Conditions (Blocked)
Story
Math

References for Language Task: Reliable across subjects and robust activation (Binder *et al.* 2011).

Social Cognition (Theory of Mind)

Participants were presented with short video clips (20 seconds) of objects (squares, circles, triangles) that either interacted in some way, or moved randomly on the screen. These videos were developed by either Castelli and colleagues (Castelli *et al.* 2000) or Martin and colleagues (Wheatley *et al.* 2007). After each video clip, participants judge whether the objects had a

mental interaction (an interaction that appears as if the shapes are taking into account each other's feelings and thoughts), Not Sure, or No interaction (i.e., there is no obvious interaction between the shapes and the movement appears random). Each of the two task runs has 5 video blocks (2 Mental and 3 Random in one run, 3 Mental and 2 Random in the other run) and 5 fixation blocks (15 seconds each).

Conditions (Blocked)
Random interaction
Mental interaction

References for the Social Cognition Task: Reliable across subjects and robust activation (Castelli *et al*. 2000; Castelli *et al.* 2002; Wheatley *et al.* 2007; White *et al.* 2011).

Relational Processing

This task was adapted from the one developed by Christoff and colleagues (Smith *et al.* 2007). The stimuli are 6 different shapes filled with 1 of 6 different textures. In the relational processing condition, participants are presented with 2 pairs of objects, with one pair at the top of the screen and the other pair at the bottom of the screen. They are told that they should first decide what dimension differs across the top pair of objects (differed in shape or differed in texture) and then they should decide whether the bottom pair of objects also differ along that same dimension (e.g., if the top pair differs in shape, does the bottom pair also differ in shape). In the control matching condition, participants are shown two objects at the top of the screen and one object at the bottom of the screen, and a word in the middle of the screen (either "shape" or "texture"). They are told to decide whether the bottom object matches either of the top two objects on that dimension (e.g., if the word is "shape", is the bottom object the same shape as either of the top two objects. For both conditions, the subject responds yes or no using one button or another. For the relational condition, the stimuli are presented for 3500 ms, with a 500 ms ITI, and there are four trials per block. In the matching condition, stimuli are presented for 2800 ms, with a 400 ms ITI, and there are 5 trials per block. Each type of block (relational or matching) lasts a total of 18 seconds. In each of the two runs of this task, there are 3 relational blocks, 3 matching blocks and 3 16-second fixation blocks.

Conditions (Blocked)
Relational processing
Matching

References for the Relational Processing Task: Localizer (Smith *et al.* 2007).

Emotion Processing

This task was adapted from the one developed by Hariri and colleagues (Hariri *et al.* 2002). Participants are presented with blocks of trials that either ask them to decide which of two faces presented on the bottom of the screen match the face at the top of the screen, or which of two shapes presented at the bottom of the screen match the shape at the top of the screen. The faces have either an angry or fearful expression. Trials are presented in blocks of 6 trials of the same task (face or shape), with the stimulus presented for 2000 ms and a 1000 ms ITI. Each block is preceded by a 3000 ms task cue ("shape" or "face"), so that each block is 21 seconds including the cue. Each of the two runs includes 3 face blocks and 3 shape blocks, with 8 seconds of fixation at the end of each run.

Conditions (Blocked)
Face
Shape

Note: A bug was written into the E-prime script for the EMOTION task, such that the task stopped short of the last three trials of the last task block in each run. This bug was not discovered until data had been collected on several participants. Consequently, the BOLD images and E-Prime data for the EMOTION task are shorter than our original design described above.

References for the Emotion Processing Task: Localizer (Hariri *et al.* 2002); Moderate reliability across time (Manuck *et al*. 2007).

Behavioral and individual difference measures details

There are important individual differences in patterns of connectivity and there is evidence that this variability is associated with alterations in important cognitive and behavioral variables that affect real world function. Thus, the HCP is collecting behavioral measures of a range of motor, sensory, cognitive and emotional processes that will delineate a core set of functions relevant to understanding the relationship between brain connectivity and human behavior.

The HCP collects many behavioral measures developed for the NIH Toolbox Assessment of Neurological and Behavioral function and several additional measures to assess domains not covered by the NIH Toolbox. Our goal in choosing assessments was to identify and utilize a reliable and well-validated battery of measures that assess a wide range of human functions and behaviors in a reasonable amount of time (3-4 hours total, to satisfy subject burden considerations). For more information on the rationale behind the development of the behavioral and individual differences batteries used in HCP see Barch *et al*. 2013.

Many of the behavioral and individual difference measures are included in the data that is released under the Open Access Data Use Terms. Due to their sensitive and possibly identifying nature, several of the measures are being released only to those who have been approved for Restricted Data. Restricted Data categories are marked with a * below.

Some behavioral measure data (marked with a † below) are not currently ready for release, but will be made available for all subjects in a future release.

Details on the NIH Toolbox behavioral measures included in the HCP behavioral battery can be found at http://www.nihtoolbox.org and are individually linked in the table below. Details and references for each behavioral measure are described below by domain.

HCP battery of behavioral and individual difference measures

Domain	Subdomain (Measure Name)
Alertness	Cognitive Status (Mini Mental Status Exam)
	Sleep (Pittsburgh Sleep Questionnaire)
Cognition	Episodic Memory (Picture Sequence Memory)
	Executive Function/Cognitive Flexibility (Dimensional Change Card Sort)
	Executive Function/Inhibition (Flanker Inhibitory Control and Attention Task)
	Fluid Intelligence (Penn Progressive Matrices)
	Language/Reading Decoding (Oral Reading Recognition)
	Language/Vocabulary Comprehension (Picture Vocabulary)
	Processing Speed (Pattern Comparison Processing Speed)

Domain	Subdomain (Measure Name)
	Self-regulation/Impulsivity (Delay Discounting)
	Spatial Orientation (Variable Short Penn Line Orientation Test)
	Sustained Attention (Short Penn Continuous Performance Test)
	Verbal Episodic Memory (Penn Word Memory Test)
	Working Memory (List Sorting)
Emotion	Emotion Recognition (Penn Emotion Recognition Test)
	Negative Affect (Sadness, Fear, Anger) (self-report)
	Psychological Well-being (Positive Affect, Life Satisfaction, Meaning and Purpose) (self-report)
	Social Relationships (Social Support, Companionship, Social Distress, Positive Social Development) (self-report)
	Stress and Self-Efficacy (Perceived Stress, Self-Efficacy) (self-report)
Motor	Endurance (2 minute walk test)
	Locomotion (4-meter walk test)
	Dexterity (9-hole Pegboard)
	Strength (Grip Strength Dynamometry)
Personality	Five Factor Model (NEO-FFI)
Physical Function	Blood Pressure[*†]
	Endocrine disorders and age of onset (self-report)[*†]
	Glucose levels (Hemoglobin A1c)[*†]
	Hematocrit Levels[*†]
	Height and Weight[*†]
	Menstrual Cycle and Hormonal Status[*†]
	Thyroid Function (Thyroid Stimulating Hormone Levels)[*†]
Psychiatric, Substance Abuse, and Life Function	Life Function (Achenbach Adult Self-Report, Syndrome Scales and DSM-Oriented Scale)[*]
	Psychiatric Clinical Symptoms (Semi-Structured Assessment for the Genetics of Alcoholism-SSAGA)[*†]
	Parental Psychiatric or Neurologic illnesses[*†]
	Substance Use (breathalyzer, self-report of current alcohol and tobacco use)[*]
	Urine Drug Screen[*]
Sensory	Audition (Words in Noise)
	Color Vision (Farnsworth Test)[*]

Domain	Subdomain (Measure Name)
	Contrast Sensitivity (Mars Contrast Sensitivity)
	Olfaction (Odor Identification Test)
	Pain (Pain Intensity and Interference Surveys) (self-report)
	Taste (Regional Taste Intensity Test)
	Visual Acuity (Electronic Visual Acuity System)*

*Access to this restricted data is granted to approved users only.

†To be included in a future data release.

Appendix 7: Abbreviations for HCP Behavioral and Individual Difference Measures provides definitions and descriptions of the short names and typical values for each of the measures.

Results of open access behavioral measures for all tested Q1-Q3 subjects are available as a comma separated variable (.csv) file on the ConnectomeDB splash page (accessible after log in at https://db.humanconnectome.org/). The behavioral measures are also available per subject and can be filtered upon in the ConnectomeDB HCP Dashboard (accessible by clicking the Explore [BETA] buttons from the ConnectomeDB splash page).

FreeSurfer measures

In addition to the behavioral measures, individual difference measures for physical brain features (morphometry statistics) that are automatically generated by the FreeSurfer processing pipeline are included as a searchable domain in ConnectomeDB. References to the FreeSurfer definitions for these statistics are included in the ConnectomeDB data dictionary.

Descriptions and References for HCP Behavioral and Individual Difference Measures

Subject Information

Demographics

Each HCP participant is given a participant identification number for tracking and asked a number of demographic questions including gender, age, educational level, twin status, race, and ethnicity. Gender and five year-age range are available as Open Access data. Educational level will be open access and available in a future data release. Due to the possible identification of participants when several data types are known and used in combination, the other demographic data are available as restricted data.

Study Completion

For each subject, we record the data completeness for each study imaging session and measurement to enable the ability to search for complete datasets for analysis.

Handedness

Handedness is assessed using the Edinburgh Handedness questionnaire (Oldfield 1971). Due to the possible identification of participants when several data types are known and used in combination, handedness is available as restricted data.

MR Sessions

Session Information

Detailed information on each of the magnetic resonance scans collected on each participant can be viewed and searched and ConnectomeDB.

Alertness

Cognitive Status and Sleep

Participants are administered the Mini Mental Status Exam (Folstein *et al.*, 1975) as a broad measure of cognitive status (participants are excluded if they score below a 27) (Crum *et al.*, 1993) and the Pittsburgh Sleep Questionnaire (Buysse *et al.*, 1989) as a measure of sleep quality. Summary scores are reported.

Cognition

Episodic Memory

Non-verbal episodic memory is assessed using the NIH Toolbox Picture Sequence Memory Test.

Executive Function

Executive function is measured by two tests in the NIH Toolbox battery. Cognitive flexibility is assessed using the NIH Toolbox Dimensional Change Card Sort Test (DCCS) and attention and inhibitory control are accessed using the NIH Toolbox Flanker Inhibitory Control and Attention Test.

Fluid Intelligence

Although the Toolbox contains measures of crystallized IQ (e.g., vocabulary acquisition), an aspect of IQ strongly influenced by educational opportunities, and measures of executive function (which are both theoretically and empirically related to fluid intelligence), it does not contain a specific measure of fluid intelligence. This construct is strongly linked to specific functional outcomes and to variations in neuronal structure and function in humans (Duncan *et al.* 2000; Duncan 2003; Duncan 2005). The most commonly used measure of fluid intelligence is Raven's Progressive Matrices (Prabhakaran *et al.* 1997; Christoff *et al.* 2001; Gray *et al.* 2003; Conway *et al.* 2005; Gray *et al.* 2005; Wendelken *et al.* 2008). We use Form A of an abbreviated version of the Raven's developed by Gur and colleagues (Bilker *et al.* 2012). Participants are presented with patterns made up of 2x2, 3x3 or 1x5 arrangements of squares, with one of the squares missing. The participant must pick one of five response choices that best fits the missing square on the pattern. The task has 24 items and 3 bonus items, arranged in order of increasing difficulty. However, the task discontinues if the participant makes 5 incorrect responses in a row.

Language

Language skills are accessed using the NIH Toolbox Oral Reading Recognition Test and the NIH Toolbox Picture Vocabulary Test.

Processing Speed

Speed of participants mental processing is measured using the NIH Toolbox Pattern Comparison Processing Speed Test.

Self-Regulation/Impulsivity (Delay Discounting)

Delay discounting describes the undervaluing of rewards that are delayed in time. It is illustrated by the fact that humans (and other animals) will often choose a smaller immediate reward over an objectively larger, but delayed reward. We use a version of the discounting task that identifies 'indifference points' at which a person is equally likely to choose a smaller reward (e.g., $100) sooner versus a larger reward later (e.g., $200 in 3 years). Based on the work of Green and Myerson (Estle *et al*. 2006; Green *et al*. 2007), we use an adjusting-amount approach, in which delays are fixed and reward amounts are adjusted on a trial-by-trial basis based on participants' choices, to rapidly hone in on indifference points. This approach has been repeatedly validated to provide reliable estimates of delay discounting (Estle *et al.* 2006). As a summary measure, we use an area-under-the-curve discounting measure (AUC) that provides a valid and reliable index of how steeply an individual discounts delayed rewards (Myerson *et al.* 2001).

Detailed Description of Delay Discounting Measure

In this task, participants are presented with two choices on each trial – a smaller amount "today" or a larger amount at a later point in time. Participants make choices at each of 6 delays (1 month, 6 months, 1 year, 3 years, 5 years and 10 years) and for two delayed amounts ($200 and $40,000). For each combination of delay and amount of delayed reward (e.g., $200 in 1 month or $40,000 in 6 months), participants make 5 choices, and the value that would have been used for the immediate amount in a 6th choice is taken as the indifference point for that condition. The participants make all five choices for a particular combination of delay and amount before moving on to the next combination of delay and amount. The order is as follows:

Delayed amount of $200 dollars

- Today versus 6 months
- Today versus 3 years
- Today versus 1 month
- Today versus 5 years
- Today versus 10 years
- Today versus 1 year

Delayed amount of $40,000 dollars

- Today versus 6 months
- Today versus 3 years
- Today versus 1 month
- Today versus 5 years
- Today versus 10 years
- Today versus 1 year

The first choice at each delay is between the delayed amount ($200 or $40,000) and an immediate amount equal to ½ the delayed amount (e.g., $100 today or $200 in 1 month, $20,000 today or $40,000 in one month). The size of the adjustment after the first choice is always ½ the amount of the immediate value on the first choice (e.g., a change of $50 if the first immediate amount is $100). If the subject chooses the immediate amount, then the immediate amount is reduced on the next choice (e.g., $50 today versus $200 in 1 month). If the subject chooses the delayed amount, then the immediate amount is increased (e.g., $150 today versus $200 in 1 month). The amount of change on each subsequent choice is ½ the amount of the prior change (e.g., $25 on the 3rd trial), regardless of whether the subject chooses the immediate or the delayed amount. This procedure rapidly hones in on the amount of immediate gain that is close to the subjective value of the delayed gain.

This design means that for all the choices with $200 dollars as the delayed amount, the first choice will always be between $100 today, and $200 in the specified time period. The second

choice will always increment or decrement the immediate value by $50. The third choice will always increment or decrement the immediate value by $25. The fourth choice will always increment or decrement the immediate value by $12.50. The fifth choice will always increment or decrement the immediate value by $6.25. The "sixth" choice value, which is never presented to the subject, but is entered in the database, is always an increment or decrement of $3.125 from the immediate value on the 5th choice. Similarly, for all the choices with $40,000 dollars as the delayed amount, the first choice will always be between $20,000 today, and $40,000 in XX time period. The second choice will always increment or decrement the immediate value by $10,000. The third choice will always increment or decrement the immediate value by $5,000. The fourth choice will always increment or decrement the immediate value by $2,500. The fifth choice will always increment or decrement the immediate value by $1,250. The "sixth" choice value, which is never presented to the subject, but is entered in the database, will always be an increment or decrement of $625 from the immediate value on the 5th choice.

Thus, for the $200 amount, we will have 6 Subject Values:

- $SV_{1mo.2}$
- $SV_{6mo.2}$
- $SV_{1yr.2}$
- $SV_{3yr.2}$
- $SV_{5yr.2}$
- $SV_{10yr.2}$

Thus, for the $40,000 amount, we will have 6 Subject Values:

- $SV_{1mo.40}$
- $SV_{6mo.40}$
- $SV_{1yr.40}$
- $SV_{3yr.40}$
- $SV_{5yr.40}$
- $SV_{10yr.40}$

We compute an Area under the curve measure for each of the two amounts as described below.

Area under the Curve for $200 = $((1+SV_{1mo.2})/(120*200)) + ((SV_{1mo.2}+SV_{6mo.2})/(48*200)) + ((SV_{6mo.2}+SV_{1yr.2})/(40*200)) + ((SV_{1yr.2}+SV_{3yr.2})/(10*200)) + ((SV_{3yr.2}+SV_{5yr.2})/(10*200)) + ((SV_{5yr.2}+SV_{10yr.2})/(4*200))$

Area under the Curve for $40,000 = $((1+SV_{1mo.4})/(120*40,000)) + ((SV_{1mo.4}+SV_{6mo.4})/(48*40,000)) + ((SV_{6mo.4}+SV_{1yr.4})/(40*40,000)) + ((SV_{1yr.4}+SV_{3yr.4})/(10*40,000)) + ((SV_{3yr.4}+SV_{5yr.4})/(10*40,000)) + ((SV_{5yr.4}+SV_{10yr.4})/(4*40,000))$

Spatial Orientation

The NIH Toolbox does not contain any measures of visual-spatial processing. Thus, we are measuring spatial orientation processing using the Variable Short Penn Line Orientation Test (Gur *et al.* 2001a; Gur *et al.* 2010). Participants are shown two lines with different orientations. They have to rotate one of the lines (a moveable blue one) so that is parallel to the other line (a fixed red line). The rotation of the blue line is accomplished by clicking buttons on the keyboard that rotate the lines either clockwise or counterclockwise. Across trials, the lines vary in their relative location on the screen, though the distance between the centers of the two lines is always the same. The length of the red line is always the same, but the length of the blue line can be either short or long. There are a total of 24 trials.

Sustained Attention

We measure continuous sustained attention using the Short Penn Continuous Performance Test (CPT, Number/Letter Version) (Gur *et al.* 2001a; Gur *et al.* 2001b; Gur *et al.* 2010). Participants see vertical and horizontal red lines flash on the computer screen. In one block, they must press the spacebar when the lines form a number and in the other block they press the spacebar when the lines form a letter. The lines are displayed for 300 ms followed by a 700 ms ITI. Each block contains 90 stimuli and lasts for 1.5 minutes.

Verbal Episodic Memory

The NIH Toolbox contains a measure of non-verbal episodic memory. Thus, we are assessing verbal episodic memory using Form A of the Penn Word Memory Test (Gur *et al.* 2001a; Gur *et al.* 2010). Participants are shown 20 words and asked to remember them for a subsequent memory test. They are then shown 40 words (the 20 previously presented words and 20 new words matched on memory related characteristics). They decide whether they have seen the word previously by choosing among “definitely yes,” “probably yes,” “probably no,” and “definitely no.”

Working Memory

We measure working memory using the NIH Toolbox List Sorting Working Memory Test.

Emotion

Emotion Recognition

The NIH Toolbox contains only self-report measures of emotional function. Thus, in order to obtain a behavioral measure of emotion processing, we are using the Penn Emotion

Recognition Test (Gur *et al.* 2001a; Gur *et al.* 2010). Participants are presented with 40 faces, one at a time. They are asked to choose what emotion the face is showing from five choices: Happy, Sad, Angry, Scared and No Feeling. Half of the faces are males and half are females. There are 8 faces each that have a happy, sad, angry, scared or no feeling expression.

Negative Affect

Self-reported unpleasant feelings or emotions including anger, fear and sadness are accessed by the NIH Toolbox Negative Affect surveys.

Psychological Well-being

Measures of general life satisfaction, meaning and purpose, and positive affect are self-reported through the NIH Toolbox Psychological Well-Being surveys.

Social Relationships

The structure, extent, and quality of each participant's social relationships including self-reported social support, companionship, social distress, and positive social development are accessed by NIH Toolbox Social Relationships survey measures.

Stress and Self-Efficacy

Self-reported perception of each participant's psychological stress and ability to cope with difficult situations are accessed by the NIH Toolbox Stress and Self-Efficacy surveys.

Motor

Endurance

Sub-maximal cardiovascular endurance is measured using the NIH Toolbox 2-Minute Walk Endurance Test.

Locomotion

Gait speed as an indicator of the health of many physiological systems is measured by the NIH Toolbox 4- Meter Walk Gait Speed Test.

Dexterity

Manual dexterity is measured using the NIH Toolbox 9-Hole Pegboard Dexterity Test.

Strength

Full force grip strength is measured by the NIH Toolbox Grip Strength Test.

Personality

Five Factor Model

There is consensus that a five factor model captures the major facets of human personality across cultures (Heine and Buchtel 2009): a) neuroticism; b) extroversion/introversion; c) agreeableness; d) openness; and e) conscientiousness (Goldberg 1993; McCrae and Costa 2008). We are administering the 60 item version of the Costa and McRae Neuroticism/Extroversion/Openness Five Factor Inventory (NEO-FFI), which has shown excellent reliability and validity (McCrae and Costa 2004). This measure was available as part of the Penn Computerized Cognitive Battery (Gur *et al*. 2001a; Gur *et al.* 2010).

Physical Function

All Physical Function measures are being released as Restricted Access data in a future data release.

Blood Pressure

Each participant's blood pressure is taken and recorded at their visit.

Endocrine disorders

Participants are asked to self-report if they have ever been diagnosed with an endocrine disorder and, if so, what age they were at the onset of the disorder.

Glucose, Hematocrit Levels

We take a blood sample of every participant for various tests. We measure Hemoglobin A1c as a measure of glucose levels over time and hematocrit levels to assess the volume percentage of red blood cells in the blood.

Height and Weight

Self-reported height and weight are recorded for each participant.

Menstrual Cycle and Hormonal Status

Female participants are asked questions about their hormonal status and menstrual cycle during the intake interview at their first in person session.

Thyroid Function

Thyroid Stimulating Hormone level in the blood is measured as an endocrine measure.

Psychiatric, Substance Abuse, and Life Function

All Psychiatric, Substance Abuse, and Life Function data is being released to Restricted Access approved users only. Psychiatric clinical symptoms and parental psychiatric illness data will be released in a future data release.

Life Function

The NIH toolbox contains self-report measures of a number of important domains of experience, including positive and negative affect, stress, anxiety, depression and social support. To obtain additional self-report information on an even broader variety of domains, we also administer the Achenbach Adult Self-Report (ASR) for Ages 18-59 (Achenbach 2009). Specifically, we administer the 123 items from Section VIII. These can be used to generate the ASR Syndrome Scales and the ASR Diagnostic and Statistical Manual (DSM)-Oriented Scales.

Psychiatric Clinical Symptoms

As part of the screening and assessment process, all participants are given a comprehensive assessment of psychiatric and substance use history over the phone, using the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) (Bucholz *et al*., 1994). The SSAGA is a well-validated diagnostic instrument used in numerous previous large-scale studies (Bucholz *et al*., 1994; Hesselbrock *et al*., 1999). It assesses a range of diagnostic categories (substance, mood, anxiety, eating disorders and adult ADHD), as well as antisocial personality disorder, using both DSM-IV criteria and either RDC or ICD-criteria, and provides information about both current and lifetime experiences. This instrument also contains the Fagerstrom Test for Nicotine Dependence (Heatherton *et al.*, 1991; Kozlowski *et al.*, 1994).

Parental Psychiatric or Neurological Illnesses

Participants are given a brief assessment of parental history of psychiatric and neurological disorders (yes/no for schizophrenia or psychosis, depression, bipolar, anxiety that needed treatment, drug or alcohol problems, Alzheimer's Disease or dementia, Parkinson's disease, or Tourette's syndrome).

Substance Use and Drug Screen

Participants are also given a breathalyzer and a urine drug screen (cocaine, THC, opiates, amphetamine, methamphetamine, oxycontin) on each day of testing. These drug screens were not used as an exclusion criterion, but rather for characterization. In addition, on the last day of testing, participants fill out a seven day retrospective report of alcohol and tobacco use.

Sensory

Audition

Hearing, or audition, is assessed using the NIH Toolbox Words-in-Noise Test.

Color Vision and Contrast Sensitivity

The NIH Toolbox does not measure color vision or contrast sensitivity. We are assessing color vision using the Farnsworth Test, a valid and reliable measure that provides more quantitative information than the commonly used Ishihara Test (Cole *et al*. 2007). In this task, participants order 15 colored blobs as a function of what they think are the closest matching colors. Based on the results, participants are classified as having Normal color vision, Protan (reduced sensitivity to red light), Deutan (reduced sensitivity to green light) or Tritian (reduced sensitivity to blue light) color vision problems. Because color vision deficits are relatively rare and can be identifying, results from this assessment are Restricted Access.

We are assessing contrast sensitivity using the Mars Contrast Sensitivity Test (Arditi *et al*. 2005), a brief, valid and reliable measure that improves upon the traditional Pelli-Robson measure (Dougherty *et al.* 2005; Haymes *et al.* 2006; Thayaparan *et al.* 2007).

Olfaction

Smell, or olfaction, is assessed using the NIH Toolbox Odor Identification Test.

Pain

Participants' self-reported experience of pain is measured by the NIH Toolbox Pain Interference and Pain Intensity Surveys.

Taste

Taste perception is measured by the NIH Toolbox Regional Taste Intensity Test.

Visual Acuity

To assess visual acuity, we are using the Electronic Visual Acuity (EVA) system provided by the Jaeb Center for Health Research and running the Electronic Early Treatment of Diabetic Retinopathy (E-ETDR) protocol (Beck *et al*., 2003; Moke *et al.*, 2001). Subjects are scored by the EVA Numerator and Denominator, which are approximations of the Snellen ratio of visual acuity. The EVA Numerator is the virtual distance in feet (always 20) from the EVA system display that the subject's visual acuity was tested. Our test is conducted with the EVA Model 10-WIN system at 3 meters (~10 feet from the subject), but is calibrated to reflect values equivalent to visual acuity testing at 20 feet. Letters are displayed and the subject (with corrected vision, if applicable) is asked to read the letters. The subject's ability to read progressively smaller letters is compared with the ability of a person with normal (20/20 Snellen ratio) vision to read letters at 20 feet. The EVA Denominator is the distance in feet that a person with normal (20/20 Snellen ratio) vision can read letters as well as the subject (with corrected vision, if applicable) can read letters at a virtual distance of 20 feet (value of the EVA numerator). An EVA Denominator value less than 20 means the subject has better than normal vision, a value greater than 20 means the subject has poorer than normal vision. Because visual acuity can be identifying, results from this assessment are Restricted Access.

References

Achenbach TM (2009). The Achenbach System of Empirically Based Assessement (ASEBA): Development, Findings, Theory, and Applications. Burlington, VT: University of Vermont Research Center for Children, Youth and Families.

Andersson JLR, Skare S, Ashburner J (2003). How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage* 20(2): 870-888.

Arditi A (2005). Improving the design of the letter contrast sensitivity test. *Invest Ophthalmol Vis Sci* 46(6): 2225-2229.

Barch DM, Burgess GC, Harms MP, Petersen SE, Schlaggar BL, Corbetta M, Glasser MF, Curtiss S, Dixit S, Feldt C, Nolan D, Bryant E, Hartley T, Footer O, Bjork JM, Poldrack R, Smith SM, Johansen-Berg H, Snyder AZ, Van Essen DC (2013). Function in the human connectome: Task-fMRI and individual differences in behavior. *NeuroImage* 80: 169-189.

Beck RW, Moke PS, Turpin AH, Ferris FL, SanGiovanni JP, Johnson CA, Birch EE, Chandler DL, Cox TA, Blair RC, Kraker RT (2003). A computerized method of visual acuity testing: adaptation of the early treatment of diabetic retinopathy study testing protocol. *Am J Ophthalmol* 135: 194-205.

Beckmann CF, Smith SM (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans Med Imaging.* 23(2): 137-152.

Bilker WB, Hansen JA, Brensinger CM, Richard J, Gur RE, Gur RC (2012). Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *Assessment* 19(3): 354-369.

Binder JR, Gross WL, Allendorfer JB, Bonilha L, Chapin J, Edwards JC, Grabowski TJ, Langfitt JT, Loring DW, Lowe MJ, Koenig K, Morgan PS, Ojemann JG, Rorden C, Szaflarski JP, Tivarus ME, Weaver KE (2011). Mapping anterior temporal lobe language areas with fMRI: a multicenter normative study. *NeuroImage* 54(2): 1465-1475.

Bizzi A, Blasi V, Falini A, Ferroli P, Cadioli M, Danesi U, Aquino D, Marras C, Caldiroli D, Broggi G (2008). Presurgical functional MR imaging of language and motor functions: validation with intraoperative electrocortical mapping. *Radiology* 248(2): 579-589.

Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger Jr. JI, Reich T, Schmidt I, Schuckit MA (1994). A new, semi-structured psychiatric interview for use in

genetic linkage studies: a report on the reliability of the SSAGA. *J Stud Alcohol* 55: 149-158.

Buckner RL, Krienen FM, Castellanos A, Diaz JC, Yeo BT (2011). The organization of the human cerebellum estimated by intrinsic functional connectivity. *J Neurophysiol* 106(5): 2322-2345.

Buysse DJ, Reynolds III CF, Monk TH, Berman SR, Kupfer DJ (1989). The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res* 28: 193-213.

Caceres A, Hall DL, Zelaya FO, Williams SC, Mehta MA (2009). Measuring fMRI reliability with the intra-class correlation coefficient. *Neuroimage* 45(3): 758-768.

Castelli F, Frith C, Happé F, Frith U (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain* 125(Pt 8): 1839-1849.

Castelli F, Happé F, Frith U, Frith C (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage* 12(3): 314-325.

Christoff K, Prabhakaran V, Dorfman J, Zhao Z, Kroger JK, Holyoak KJ, Gabrieli JD (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *NeuroImage* 14(5): 1136-1149.

Cole BL (2007). Assessment of inherited colour vision defects in clinical practice. *Clin Exp Optom* 90(3): 157-175.

Conway AR, Kane MJ, Bunting MF, Hambrick DZ, Wilhelm O, Engle RW (2005). Working memory span tasks: A methodological review and user's guide. *Psychon Bull Rev* 12(5): 769-786.

Crum RM, Anthony JC, Bassett SS, Folstein MF (1993). Population-based norms for the Mini-Mental State Examination by age and educational level. *JAMA* 269: 2386-2391.

Dale AM, Fischl B, Sereno MI (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage* 9(2): 179-194.

Delgado MR, Nystrom LE, Fissell C, Noll DC, Fiez JA (2000). Tracking the hemodynamic responses to reward and punishment in the striatum. *J Neurophysiol* 84: 3072-3077.

Dougherty BE, Flom RE, Bullimore MA (2005). An evaluation of the Mars Letter Contrast Sensitivity Test. *Optom Vis Sci* 82(11): 970-975.

Downing PE, Jiang Y, Shuman M, Kanwisher N (2001). A cortical area selective for visual processing of the human body. *Science* 293(5539): 2470-2473.

Drobyshevsky A, Baumann SB, Schneider W (2006). A rapid fMRI task battery for mapping of visual, motor, cognitive, and emotional function. *NeuroImage* 31(2): 732-744.

Duncan J (2003). Intelligence tests predict brain response to demanding task events. *Nat Neurosci* 6(3): 207-208.

Duncan J (2005). Frontal lobe function and general intelligence: why it matters. *Cortex* 41(2): 215-217.

Duncan J, Seitz RJ, Kolodny J, Bor D, Herzog H, Ahmed A, Newell FN, Emslie H (2000). A neural basis for general intelligence. *Science* 289(5478): 457-460.

Estle SJ, Green L, Myerson J, Holt DD (2006). Differential effects of amount on temporal and probability discounting of gains and losses. *Mem Cognit* 34(4): 914-928.

Folstein MF, Folstein SE, McHugh PR (1975). Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 12,189-198.

Forbes EE, Hariri AR, Martin SL, Silk JS, Moyles DL, Fisher PM, Brown SM, Ryan ND, Birmaher B, Axelson DA, Dahl RE (2009). Altered striatal activation predicting real-world positive affect in adolescent major depressive disorder. *Am J Psychiatry* 166(1): 64-73.

Fox CJ, Iaria G, Barton JJ (2009). Defining the face processing network: optimization of the functional localizer in fMRI. *Hum Brain Mapp* 30(5): 1637-1651.

Glasser MF, Van Essen DC (2011). Mapping human cortical areas in vivo based on myelin content as revealed by T1-and T2-weighted MRI. *J Neurosci* 31:11597-11616.

Glasser MF, Sotiropoulos SN, Wilson JA, Coalson T, Fischl B, Andersson JL, Xu J, Jbabdi S, Webster M, Polimeni JR, Van Essen DC, Jenkinson M (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* 80:105-124.

Goldberg LR (1993). The structure of phenotypic personality traits. *Am Psychol* 48: 26-34.

Gray JR, Chabris CF, Braver TS (2003). Neural mechanisms of general fluid intelligence. *Nat Neurosci* 6(3): 316-322.

Gray JR, Burgess GC, Schaefer A, Yarkoni T, Larsen RJ, Braver TS (2005). Affective personality differences in neural processing efficiency confirmed using fMRI. *Cogn Affect Behav Neurosci* 5(2): 182-190.

Green L, Myerson J, Shah AK, Estle SJ, Holt DD (2007). Do adjusting-amount and adjusting-delay procedures produce equivalent estimates of subjective value in pigeons? *J Exp Anal Behav* 87(3): 337-347.

Griffanti L, Salimi-Khorshidi G, Beckmann CF, Bijsterbisch J, Auerbach EJ, Douaud G, Zsoldos E, Ebmeier KP, Filippini N, Mackay C, Moeller S, Xu J, Yacoub E, Baselli G, Ugurbil K, Miller KL, Smith SM (2013). Automated artefact removal and accelerated fMRI acquisition for improved Resting State Network imaging (in preparation).

Gur RC, Ragland JD, Moberg PJ, Turner TH, Bilker WB, Kohler C, Siegel SJ, Gur RE (2001). Computerized neurocognitive scanning: I. Methodology and validation in healthy people. *Neuropsychopharmacology* 25(5): 766-776.

Gur RC, Ragland JD, Moberg PJ, Bilker WB, Kohler C, Siegel SJ, Gur RE (2001). Computerized neurocognitive scanning: II. The profile of schizophrenia. *Neuropsychopharmacology* 25(5): 777-788.

Gur RC, Richard J, Hughett P, Calkins ME, Macy L, Bilker WB, Brensinger C, Gur RE (2010). A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: standardization and initial construct validation. *J Neurosci Methods* 187(2): 254-262.

Hariri AR, Brown SM, Williamson DE, Flory JD, de Wit H, Manuck SB (2006). Preference for immediate over delayed rewards is associated with magnitude of ventral striatal activity. *J Neurosci* 26: 13213-13217.

Haymes SA, Roberts KF, Cruess AF, Nicolela MT, LeBlanc RP, Ramsey MS, Chauhan BC, Artes PH (2006). The letter contrast sensitivity test: clinical evaluation of a new design. *Invest Ophthalmol Vis Sci* 47(6): 2739-2745.

Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO (1991). The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. *Br J Addict* 86: 1119-1127.

Heine SJ, Buchtel EE (2009). Personality: the universal and the culturally specific. *Annu Rev Psychol* 60: 369-394.

Hesselbrock M, Easton C, Bucholz KK, Schuckit M, Hesselbrock V (1999). A validity study of the SSAGA— a comparison with the SCAN. *Addiction* 94: 1361-1370.

Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM (2012). FSL. *NeuroImage* 62: 782-90.

Jovicich J, Czanner S, Greve D, Haley E, Kouwe A, Gollub R, Kennedy D, Schmitt F, Brown G, MacFall J, Fischl B, Dale A (2006). Reliability in Multi-Site Structural MRI Studies: Effects of Gradient Non-linearity Correction on Phantom and Human Data. *NeuroImage* 30(2): 436-43.

Kozlowski LT, Porter CQ, Orleans CT, Pope MA, Heatherton T (1994). Predicting smoking cessation with self-reported measures of nicotine dependence: FTQ, FTND, and HSI. *Drug Alcohol Depend* 34: 211-216.

Kung CC, Peissig JJ, Tarr MJ (2007). Is region-of-interest overlap comparison a reliable measure of category specificity? *J Cogn Neurosci* 19(12): 2019-2034.

Larson-Prior LJ, Oostenveld R, Della Penna S Michalareas G, Prior F, Babajani-Feremi, Schoffelen J-M, Marzetti L, de Pasquale F, Di Pompeo F, Stout J, Moolrich M, Luo Q, Bucholz R, Fries P, Pizzella V, Romani GL, Corbetta M, Snyder AZ (2013). Adding dynamics to the Human Connectome Project with MEG. *NeuroImage* 80: 190-201.

Manuck SB, Brown SM, Forbes EE, Hariri AR (2007). Temporal stability of individual differences in amygdala reactivity. *Am J Psychiatry* 164(10): 1613-1614.

Marcus DS, Harms MP, Snyder AZ, Jenkinson M, Wilson JA, Glasser MF, Barch DM, Archie KA, Burgess GC, 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 JS, Curtiss SW, Van Essen DC (2013). Human Connectome Project informatics: Quality control, database services, and data visualization. *NeuroImage* 80: 144-168.

May JC, Delgado MR, Dahl RE, Stenger VA, Ryan ND, Fiez JA, Carter C (2004). Event-related functional magnetic resonance imaging of reward-related brain circuitry in children and adolescents. *Biol psychiatry* 55(4): 359-366.

McCrae RR, Costa PT (2004). A contemplated revision of the NEO Five Factor Inventory. *Pers Individ Diff* 36(3): 587-596.

McCrae RR, Costa PT (2008). The five factor theory of personality. *Handbook of Personality: Theory and Research*. O. P. John, R. W. Robins and L. A. Pervin. New York, Guilford: 159-181.

Milchenko M, Marcus D (2013). Obscuring surface anatomy in volumetric imaging data. *Neuroinformatics* 11: 65-75.

Moke PS, Turpin AH, Beck RW, Holmes JM, Repka MX, Birch EE, Hertle RW, Kraker RT, Miller JM, Johnson CA (2001). Computerized method of visual acuity testing: adaptation of the amblyopia treatment study visual acuity testing protocol. *Am J Ophthalmol* 132: 903-909.

Morioka T, Yamamoto T, Mizushima A, Tombimatsu S, Shigeto H, Hasuo K, Nishio S, Fujii K, Fukui M (1995). Comparison of magnetoencephalography, functional MRI, and motor evoked potentials in the localization of the sensory-motor cortex. *Neurol Res* 17(5): 361-367.

Myerson J, Green L, Warusawitharana M (2001). Area under the curve as a measure of discounting. *J Exp Anal Behav* 76(2): 235-243.

Oldfield RC (1971). The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia, 9(1): 97-113.

Peelen MV, Downing PE (2005). Within-subject reproducibility of category-specific visual activation with functional MRI. *Hum Brain Mapp* 25(4): 402-408.

Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, Vogel AC, Laumann TO, Miezin FM, Schlaggar BL, Petersen SE (2011). Functional network organization of the human brain. *Neuron* 72(4): 665-78.

Prabhakaran V, Smith JA, Desmond JE, Glover GH, Gabrieli JD (1997). Neural substrates of fluid reasoning: an fMRI study of neocortical activation during performance of the Raven's Progressive Matrices Test. *Cogn Psychol* 33(1): 43-63.

Rilling JK, Glasser MF, Jbabdi S, Andersson J, Preuss TM (2012). Continuity, divergence, and the evolution of brain language pathways. *Front Evol Neurosci* 3: 11.

Salimi-Khorshidi G, Douaud G, Beckmann CF, Bijsterbisch J, Duff E, Glasser MF, Griffanti L, Smith SM (2013). Automatic Denoising of Functional MRI Data: Combining Independent Component Analysis and Hierarchical Fusion of Classifiers (in preparation).

Satterthwaite TD, Elliott MA, Gerraty RT, Ruparel K, Loughead J, Calkins ME, Eickhoff SB, Hakonarson H, Gur RC, Gur RE, Wolf DH (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage*, 64(0): 240-256.

Smith R, Keramatian K, Christoff K (2007). Localizing the rostrolateral prefrontal cortex at the individual level. *Neuroimage* 36(4): 1387-1396.

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23(S1): 208-19.

Smith SM, Andersson J, Auerbach EJ, Beckmann CF, Bijsterbosch J, Douaud G, Duff E, Feinberg DA, Griffanti L, Harms MP, Kelly M, Laumann T, Miller KL, Moeller S, Petersen

SE, Power J, Salimi-Khorshidi G, Snyder AZ, Vu A, Woolrich MW, Xu J, Yacoub E, Ugurbil K, Van Essen DC, Glasser MF (2013). Resting-state fMRI in the Human Connectome Project. *NeuroImage* 80: 144-168.

Sotiropoulos SN, Moeller S, Jbabdi S, Xu J, Andersson JL, Auerbach EJ, Yacoub E, Feinberg D, Setsompop K, Wald LL, Behrens TEJ, Ugurbil K, Lenglet C (2013). Effects of Image reconstruction on fibre orientation mapping from multichannel diffusion MRI: Reducing the noise floor using SENSE. *Mag Res Med* EPub ahead of print 7 Feb 2013.

Sotiropoulos SN, Jbabdi S, Xu J, Andersson JL, Moeller S, Auerbach EJ, Glasser MF, Hernandez M, Sapiro G, Jenkinson M, Feinberg D, Yacoub E, Lenglet C, Van Essen DC, Ugurbil K, Behrens TEJ (2013). Advances in diffusion MRI acquisition and processing in the Human Connectome Project. *NeuroImage* 80: 125-143.

Taylor JC, Wiggett AJ, Downing PE (2007). Functional MRI analysis of body and body part representations in the extrastriate and fusiform body areas. *J Neurophysiol* 98(3): 1626-1633.

Thayaparan K, Crossland MD, Rubin GS (2007). Clinical assessment of two new contrast sensitivity charts. *Br J Ophthalmol* 91(6): 749-752.

Tricomi EM, Delgado MR, Fiez JA (2004). Modulation of caudate activity by action contingency. *Neuron* 41(2): 281-292.

Ugurbil K, Xu J, Auerbach EJ, Moeller S, Vu A, Duarte-Carvajalino JM, Lenglet C, WU X, Schmitter S, Van de Moortele PF, Strupp J, Sapiro G, De Martino F, Wang D, Harel N, Garwood M, Chen L, Feinberg DA, Smith SM, Miller KL, Sotiropoulos SN, Jbabdi S, Andersson JL, Behrens TEJ, Glasser MF, Van Essen DC, Yacoub E (2013). Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *NeuroImage* 80: 80-104.

Van Essen DC, Glasser MF, Dierker DL, Harwell J, Coalson T (2012). Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. *Cereb Cortex* 22(10): 2241-62.

Van Essen DC, Smith SM, Barch DM, Behrens TEJ, Yacoub E, Ugurbil K (2013). The WU-Minn Human Connectome Project: An Overview. *NeuroImage* 80: 62-79.

Wheatley T, Milleville SC, Martin A (2007). Understanding animate agents: distinct roles for the social network and mirror system. *Psychol Sci* 18(6): 469-474.

White SJ, Coniston D, Rogers R, Frith U (2011). Developing the Frith-Happe animations: a quick and objective test of Theory of Mind for adults with autism. *Autism Res* 4(2): 149-154.

Wendelken C, Nakhabenko D, Donohue SE, Carter CS, Bunge SA (2008). Brain is to thought as stomach is to ??: investigating the role of rostrolateral prefrontal cortex in relational reasoning. *J Cogn Neurosci* 20(4): 682-693.

Woolrich MW, Jbabdi S, Patenaude B, Chappell M, Makni S, Behrens T, Beckmann C, Jenkinson M, Smith SM (2009). Bayesian analysis of neuroimaging data in FSL. *NeuroImage* 45(1 Suppl): S173-S186.

Yeo BT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zöllei L, Polimeni JR, Fischl B, Liu H, Buckner RL (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol* 106(3): 1125-1165.

Zaitsev M, Dold, Sakas G, Hennig J, Speck O (2006). Magnetic resonance imaging of freely moving objects: prospective real-time motion correction using an external optical motion tracking system. *NeuroImage* 31:1038–1050.

** Further details on the HCP data acquisition and analysis methods as of March 2013 are available in eight manuscripts published in a special Connectome issue of *NeuroImage* (Van Essen *et al*., 2013, Ugurbil *et al*., 2013, Glasser *et al.* 2013, Smith *et al*., 2013, Barch *et al*., 2013, Sotiropolous *et al*., 2013, Marcus *et al*., 2013, and Larson-Prior *et al*., 2013).

Appendices

Appendix 1. HCP scan protocols

Download Appendix 1 here:
http://humanconnectome.org/documentation/Q3/HCP_Q3_Release_Appendix_I.pdf

Appendix 2. Matlab code for voxel-wise correction of dMRI gradients

Download Appendix 2 here:
http://humanconnectome.org/documentation/Q3/HCP_Q3_Release_Appendix_II.pdf

Appendix 3. File Names and Directory Structure for Unprocessed, Preprocessed, ICA-FIX denoised rfMRI, and Individual (within-subject) tfMRI FEAT-analyzed Data

Download Appendix 3 here:
http://humanconnectome.org/documentation/Q3/HCP_Q3_Release_Appendix_III.pdf

Appendix 4. HCP Protocol Standard Operating Procedures (SOPs)

Download Appendix 4 here:
http://humanconnectome.org/documentation/Q3/HCP_Q3_Release_Appendix_IV.pdf

Appendix 5: Data Acquisition Information for an Exemplar Subject

Download Appendix 5 here:
http://humanconnectome.org/documentation/Q3/HCP_Q3_Release_Appendix_V.pdf

Appendix 6. Task fMRI E-Prime Key Variables

Download Appendix 6 here:
http://humanconnectome.org/documentation/Q3/HCP_Q3_Release_Appendix_VI.pdf

Appendix 7. Abbreviations for HCP Behavioral and Individual Difference Measures

Download Appendix 7 here:
http://humanconnectome.org/documentation/Q3/HCP_Q3_Release_Appendix_VII.pdf