NeuroImage xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Review

The WU-Minn Human Connectome Project: An overview

David C. Van Essen [a,*], Stephen M. Smith [b], Deanna M. Barch [c], Timothy E.J. Behrens [b], Essa Yacoub [d], Kamil Ugurbil [d] for the WU-Minn HCP Consortium

[a] Department of Anatomy & Neurobiology, Washington University School of Medicine, 660 S. Euclid Avenue, St. Louis, MO 63110, USA
[b] University of Oxford, Functional MRI of Brain, Oxford OX3 9DU, UK
[c] Psychology Department, Washington University, St. Louis, MO 63105, USA
[d] Center for Magnetic Resonance Imaging, University of Minnesota, Minneapolis, MN 55455, USA

ARTICLE INFO

Article history:
Accepted 6 May 2013
Available online xxxx

ABSTRACT

The Human Connectome Project consortium led by Washington University, University of Minnesota, and Oxford University is undertaking a systematic effort to map macroscopic human brain circuits and their relationship to behavior in a large population of healthy adults. This overview article focuses on progress made during the first half of the 5-year project in refining the methods for data acquisition and analysis. Preliminary analyses based on a finalized set of acquisition and preprocessing protocols demonstrate the exceptionally high quality of the data from each modality. The first quarterly release of imaging and behavioral data via the ConnectomeDB database demonstrates the commitment to making HCP datasets freely accessible. Altogether, the progress to date provides grounds for optimism that the HCP datasets and associated methods and software will become increasingly valuable resources for characterizing human brain connectivity and function, their relationship to behavior, and their heritability and genetic underpinnings.

Contents

* Corresponding author. Fax: +1 314 747 3436.
E-mail address: vanessen@wustl.edu (D.C. Van Essen).

1053-8119/$ – see front matter © 2013 Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.neuroimage.2013.05.041

Introduction

A revolution in noninvasive neuroimaging methods over the past two decades has enabled the analysis and visualization of human brain structure, function, and connectivity in unprecedented detail. These advances make it feasible to systematically explore the human connectome, i.e., to generate maps of brain connectivity that are 'comprehensive' down to the spatial resolution of the imaging methods available.

In 2009, the NIH Neuroscience Blueprint Institutes and Centers announced a Request for Applications (RFA) targeted at characterizing the human connectome and its variability using cutting-edge neuroimaging methods. The RFA sought applications that addressed the dual objectives of accelerating advances in key technologies and applying these advances to a large population of healthy adults. In 2010, NIH awarded Human Connectome Project (HCP) grants to two consortia, one led by Washington University, the University of Minnesota, and Oxford University (the "WU-Minn" HCP consortium), and the other led by MGH and UCLA (the MGH-UCLA HCP consortium) (see http://www.neuroscienceblueprint.nih.gov/connectome/).

After summarizing the key objectives of the WU-Minn HCP consortium, this article provides an overview of results from our extensive efforts to refine and optimize the many methods used for data acquisition and analysis. MRI data acquisition protocols for scanning at 3 T were finalized[1] in August, 2012, and are now being used to acquire high-quality data from many subjects. In this article we highlight key methodological advances and summarize how these large and complex imaging and behavioral datasets are being acquired, processed, and shared. This sharing includes the release in March 2013 of data from 68 subjects scanned during the first quarter (Q1) of Phase II data collection. This dataset includes unprocessed and 'minimally preprocessed' data on all subjects, plus more extensively analyzed group-average data for several modalities.

Additional articles in this special issue go into greater detail in these specific areas and provide a wealth of information about our instrumentation and image acquisition methods (Ugurbil et al., 2013); preprocessing pipelines (Glasser et al., 2013b); diffusion imaging (Sotiropoulos et al., 2013b); resting-state fMRI (Smith et al., 2013); Q6 task-fMRI and behavior (Barch et al., 2013); MEG (Larson-Prior et al., Q7 2013); and informatics and quality control processes (Marcus et al., 2013). Other special issue articles describe progress by the MGH-UCLA HCP consortium.

HCP objectives

The WU-Minn HCP consortium aims to characterize human brain connectivity and function in a population of 1200 healthy adults and to enable detailed comparisons between brain circuits, behavior, and genetics at the level of individual subjects. Here, we summarize the overarching objectives and data acquisition plans of the HCP, which have not changed substantially since they were initially reported (Van Essen et al., 2012a).

Subjects

HCP subjects are drawn from a population of adult twins and their non-twin siblings, in the age range of 22–35 years. Studying sibships that include twins offers multiple advantages. Most obviously, it enables a systematic assessment of the heritability of neural circuits. Monozygotic (MZ) twins should have the greatest similarity because they are genetically nearly identical. Dizygotic (DZ) twins are no more related genetically than ordinary full siblings, but they share childhood environment, including in utero environment, to a greater degree. Combined analyses of MZ and DZ pairs will allow estimation of the extent to which genotype, shared environment, and non-shared influences each contribute to variation in traits. Including additional (non-twin) siblings provides a further increase in statistical power for analyzing heritability, distinguishing between genetic and environmental influences (Posthuma and Boomsma, 2000; Van Essen et al., 2012a) Q8 and relating genotype to phenotype.

Many aspects of brain circuitry and its relation to behavior are likely to involve small contributions from many genes, rather than dominant contributions from one or a few genes. Consequently, a large number of subjects will be needed in order to identify relationships between brain circuit phenotype and genotype. For practical reasons, our target number for the HCP is limited to 1200 subjects. This target reflects not only budget considerations but also logistical constraints associated with the number of scans feasible to carry out in a three-year period on a single dedicated 3 Tesla (3 T) scanner (see below). While 1200 subjects is small relative to many GWAS studies, the statistical power gained by studying twins and their siblings should nonetheless enable valuable exploratory genome-wide analyses of how specific genes, interacting genes, and genetic regulatory sequences may influence brain connectivity.

Imaging data

MR scanning includes four imaging modalities, acquired at resolutions that are notably high for a large-scale in vivo study: structural MRI, resting-state fMRI (rfMRI), task fMRI (tfMRI), and diffusion MRI (dMRI). All 1200 subjects will be scanned using all four of these modalities on a customized 3 T scanner at Washington University (WashU). Two hundred of the same subjects will also be scanned on a 7 T scanner at the University of Minnesota (UMinn), using the same four imaging modalities. A subset of 100 subjects will be studied using combined MEG/EEG (resting-state and task-evoked) carried out at St. Louis University (SLU).

[1] The WU-Minn consortium will also acquire MRI data at 7 T, using methods that are still under development (Ugurbil et al, 2013). The MEG protocol has recently been finalized and data acquisition is scheduled to begin in May, 2013 (Larson-Prior et al., 2013).

Please cite this article as: Van Essen, D.C., et al., The WU-Minn Human Connectome Project: An overview, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.05.041

Behavior

On the behavioral front, our objective is to capture a large amount of information about each subject across many behavioral domains, especially for measures that have the potential to covary in interesting ways (across subjects) with brain connectivity and function. A secondary objective is to use standardized behavioral tests as much as is feasible, to increase the prospects that findings based on the HCP data can in the future be related to other large-scale projects comparing brain and behavior.

Genetic data

Genetic analyses will be based on DNA extracted from blood samples acquired at the time of each subject's visit. Genotyping will be carried out in the final year of the project, for reasons of consistency (using a single platform), and also to obtain the greatest amount of data, given anticipated declines in price per sample.

Data sharing

The HCP is committed to making imaging and behavioral data freely available to the scientific community. Importantly, this includes not just the unprocessed ('raw') imaging data, but also data after the multiple levels of processing needed to analyze and interpret the data, e.g., to obtain maps of structural and functional connectivity at different spatial granularity. A second objective is to make the data available as soon as is feasible, via quarterly releases that allow time for data processing and quality control. A third objective is to enable flexible and powerful data mining via a user-friendly database and visualization platform. Family structure and other data will be handled by a restricted access data sharing process that imposes important constraints on what and how certain sensitive types of information can be shared and published (see below).

HCP progress

Here, we summarize progress since funding of the WU-Minn HCP consortium began (September, 2010), beginning with a brief summary of seven broad domains.

- *Subject recruitment, visits, and behavioral testing*. Many practical issues have been resolved to allow recruitment and visits to occur at a pace sufficient to study 1200 subjects over 3 years at a single imaging site, as discussed below.
- *3 T scanning protocol*. A two-year effort to develop and refine the scanning protocols for the 3 T Connectome Scanner has yielded important advances in each of the four MR-based imaging modalities (Ugurbil et al., 2013).
- *7 T scanning protocol*. An ongoing effort to improve data acquisition and preprocessing for the 7 T scanner will enable scanning of the 200 HCP subjects using higher spatial resolution than attainable on the 3 T Connectome Skyra. Scanning with the final 7 T protocols is scheduled to begin in the fall of 2013 (Ugurbil et al., 2013).
- *Minimal preprocessing pipelines*. Numerous innovations and refinements have been made in the many preprocessing steps needed to correct for spatial distortions, align data across modalities, and bring data into standard atlas spatial coordinate systems. These refinements are especially important for capitalizing on the high spatial resolution of the HCP datasets, but they are also likely to be of broad utility to other investigators and other large-scale projects in the neuroimaging community. These refinements have been consolidated into a set of well-defined preprocessing pipelines that consistently and reliably carry out distortion correction and spatial alignment for each of the four imaging modalities (Glasser et al., 2013b).
- *Analysis approaches*. Methods for later stages of image processing have advanced on many fronts and will continue to be refined over the remainder of the project. Some objectives, such as brain parcellation, inter-subject registration, and cross-modal comparisons are not only methodologically challenging, but will rely on extensive analysis of datasets generated by the HCP for their successful implementation. Some investigators external to the HCP consortium will elect to develop and apply their own analysis approaches to the unprocessed or minimally preprocessed HCP data. We anticipate that many others will prefer to take advantage of the optimized analyses being developed within the HCP and work with HCP data taken from "further along the analysis chain", so that they can begin working with the information level of most convenience to them — for example, starting with an HCP-derived "parcellated connectome" network matrix generated for individual subjects (Smith et al., 2013; Sotiropoulos et al., 2013b).
- *MEG*. Data acquisition protocols for MEG have been finalized, and scans are scheduled to commence in May, 2013. Many aspects of data analysis and cross-modal comparison will continue to be refined (Larson-Prior et al., 2013).
- *Informatics and data sharing*. The HCP has implemented two informatics platforms that will serve as workhorses for key aspects of data storage, access, analysis, and visualization. The ConnectomeDB database has been established for handling the large amounts of unprocessed and processed HCP data. The Connectome Workbench platform provides many novel visualization and analysis capabilities. Both platforms will continue to evolve and will jointly support an increasingly broad set of data mining capabilities over the next several years (Marcus et al., 2013). Q9

Subject recruitment, visits, and behavioral testing

Inclusion and exclusion criteria

Our primary participant pool comes from healthy individuals born Q11 in Missouri to families that include twins, based on data from the Missouri Department of Health and Senior Services Bureau of Vital Records. Additional recruiting efforts are used to insure that participants broadly reflect the ethnic and racial composition of the U.S. population as represented in the 2000 decennial census. We define 'healthy' broadly, aiming for a pool that is generally representative of the population at large, so that we can capture a wide range of variability in healthy individuals with respect to behavioral, ethnic, and socioeconomic diversity. We exclude sibships with individuals having severe neurodevelopmental disorders (e.g., autism), documented neuropsychiatric disorders (e.g., schizophrenia or depression) or neurologic disorders (e.g., Parkinson's disease). We also exclude individuals with illnesses such as diabetes or high blood pressure, as these might negatively impact neuroimaging data quality. Twins born prior to 34 weeks gestation and non-twins born prior to 37 weeks gestation are excluded, reflecting the higher incidence of prematurity in twins. We include individuals who are smokers, are overweight, or have a history of heavy drinking or recreational drug use without having experienced severe symptoms. This will facilitate future connectivity studies on psychiatric patients many of whom smoke, are overweight, or have subclinical substance use behaviors. Supplemental Table S1 lists all HCP inclusion and exclusion criteria.

Screening interviews

Initial telephone screening consists of a questionnaire to ascertain whether prospective participants meet the HCP inclusion criteria. If at least three family members (including one twin pair) meet the inclusion criteria and express willingness to participate, each is asked for verbal informed consent and given an extensive

Please cite this article as: Van Essen, D.C., et al., The WU-Minn Human Connectome Project: An overview, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.05.041

telephone interview, the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA, Bucholz et al., 1994). This instrument is used to confirm the absence of significant previously documented psychiatric illness and to obtain information about subthreshold psychiatric symptoms. To date, no participants who have passed the initial telephone screening have been subsequently excluded during the SSAGA. On average, approximately 6–7 families are screened in order to identify one family with a twin pair and at least one other sibling who meet all the inclusion criteria and are willing to participate. An average of 2.6 subjects per family completed visits in Q1. To prevent identification of families with unusual structures, the number of subjects in a family who can be studied has been set at a maximum of six, and no more than one pair of twins per family will be studied.

Two-day subject visits

Given the imperative of obtaining consistently high-quality data from a community population, it is important that the overall experience be as consistent as possible across participants and that it also be a positive one, without being unduly burdensome or stressful. Based on pilot studies, we established a schedule in which the standard procedure is for a participant to spend two days at WashU. In addition to the review and signature of the informed consent document at the beginning of Day 1, scans are also done in a consistent order (unless quality issues necessitate a rescan; see below). Before undergoing any actual scans, each participant has a practice session in a mock scanner to acclimate him or her to the scanner environment. The mock scanner session includes feedback on head motion following different types of instructed movements using a target strapped to the forehead, as well as training to minimize head motion while watching a film, which cuts off when head motion exceeds specific threshold. Day 1 includes a structural MRI session followed (after a break) by a session that includes first a resting-state and then a task-fMRI component. Day 2 includes a diffusion imaging scan followed by a second combined resting-state and task-fMRI session. The total duration of the standard four sessions is about 4 h, not counting set-up time. If any scan is judged unusable (see QC section below), we try to schedule an additional session during the initial visit or in a follow-up visit in order to reacquire the unusable scan.

In addition to these scan sessions, participants complete extensive behavioral assessment outside the scanner, during two sessions lasting a total of several hours (see Tables 2 and 3 in Barch et al., 2013). One set of measures, from the NIH Toolbox (http://www.nihtoolbox.org/) is typically done on visit Day 1, takes about 2 h and includes 19 subdomains within the broad domains of cognitive, emotional, motor, and sensory functions (see Barch et al., 2013, Table 2). The other session (~1.5 h duration) of 11 non-Toolbox measures is typically done on Day 2 and includes tests of vision (color vision, contrast sensitivity), attention, personality, episodic memory, emotion processing, spatial processing, fluid intelligence, and self-regulation (delay discounting). A variety of additional tests are used to characterize each participant's physical and mental state during the visit; see Supplemental Table S2 for a complete list. The order of these evaluations can vary somewhat within the visit, depending on scheduling considerations. At some time during the visit, participants are also asked for blood samples for genetic and other analyses, and for a saliva sample for genetic analysis if they decline to provide a blood sample. Blood samples for genetic analysis are shipped to the Rutgers University Cell and DNA Repository (http://www.rucdr.com) for extraction of DNA and creation of cell lines.

In general, the participants studied to date (through April, 2013) have tolerated the entire experience very well, including the extended time in a customized scanner with a reduced bore diameter (see below). The Q1 data release includes data from 76 subjects who visited through November, 2012. Complete or near-complete scans for all modalities were obtained from 68 of these subjects (see Supplemental Table S3). Reasons for partial or complete loss of imaging data include claustrophobia and physical size (body or head). Subjects who complete only the behavioral testing remain in the study if they do not meet other exclusion criteria, because the behavioral data alone may be of interest to some researchers.

A four-question satisfaction survey administered at the end of testing shows that participants report a very high level of satisfaction with their experiences. The majority of participants rate their experience as a 9 or 10 (out of 10) overall (Supplemental Table S4).

3 T connectome scanner — hardware, pulse sequences, and scanning protocols

3 T hardware

All HCP subjects are scanned on a customized Siemens 3 T "Connectome Skyra" at WashU, 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. 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 3 T. 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. For the specific method and diffusion weighting (b values) chosen in the WU-Minn consortium, 100 mT/m maximal gradient strength provides much of the gain that would be available at 300 mT/m (Ugurbil et al., 2013); the relative merit of each depends on the method and b-values employed. Thus, the two hardware systems provide complementary platforms for exploring the possible improvements that are available for tractography.

Placing the customized 100 mT/m gradient set into the Siemens 3 T Skyra system resulted in a clear inner bore diameter of 56 cm, smaller than the standard Siemens 3 T Skyra bore size (70 cm diameter) or a Siemens Trio 3 T Trio (60 cm diameter); in the absence of a custom designed patient table, this smaller bore necessitated the placement of the patient table higher in the bore, resulting in the subject's head not being centered along the gradient isocenter. As a consequence, all scans have gradient distortions larger than in a conventional scanner. These distortions have been corrected in HCP preprocessed data, but must be carried out separately by anyone starting with the unprocessed (raw) HCP scan data (see below).

Pulse sequences

The most significant pulse sequence development for the HCP was the implementation and optimization of slice-accelerated multiband (MB) acquisitions for fMRI and dMRI (Feinberg et al., 2010; Larkman et al., 2001; Moeller et al., 2008, 2010; Setsompop et al., 2012; Ugurbil et al., 2013). In general, multiband pulse sequences greatly increase the amount of data acquired per unit time, using a strategy of simultaneously exciting and acquiring multiple brain slices, which are then separated from one another during image reconstruction, based on the spatial sensitivity profiles of the multiple receive coils (32 channels for the HCP standard Siemens 3 T head coil). This efficiency increase can lead to substantially improved functional SNR (Feinberg et al., 2010; Smith et al., 2011), the ability to acquire more diffusion data points (Sotiropoulos et al., 2013b), and/or increases in spatial resolution for fMRI or dMRI (Ugurbil et al., 2013). The optimal multiband factor and other pulse sequence parameters depend on a complex set of trade-offs that entailed extensive piloting and analysis (Smith et al., 2013; Sotiropoulos et al., 2013b; Ugurbil et al., 2013). Piloting for the 3 T Connectome scanner was done at UMinn (CMRR) prior to shipping the scanner to WashU in May 2012. The multiband accelerated pulse sequences developed for the HCP project are available to interested

Please cite this article as: Van Essen, D.C., et al., The WU-Minn Human Connectome Project: An overview, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.05.041

sites (more than 60 as of February, 2013) using the Siemens "customer to peer" sequence distribution procedure. Implementation of multiband sequences for non-Siemens platforms (General Electric and Phillips) is ongoing as part of an additional HCP-funded effort.

Based on HCP piloting, we established an optimized fMRI protocol (both resting-state and task-evoked) on the Connectome Skyra that includes a multiband factor of 8, spatial resolution of 2 mm isotropic voxels, and a TR of 0.7 s (see Smith et al., 2013; Ugurbil et al., 2013). Each of the 2 hour-long sessions includes both resting-state and task fMRI. First, two 15-minute resting-state scans (eyes open and fixation on a cross-hair) are acquired with opposite phase encoding directions (L/R and R/L), for a total of 1 h of resting-state data over the two-day visit. Second, approximately 30 min of task-fMRI is acquired in each session, including 7 tasks split between the two sessions, for a total of 1 h of t-fMRI; each task is run twice, in opposing (L/R and R/L) phase-encoding directions (Barch et al., 2013). Parameters selected for diffusion imaging based on pilot data include a multiband factor of 3, nominal voxel size of 1.25 mm isotropic, and 270 diffusion weighted scans distributed equally over 3 shells defined with b-values of 1000, 2000 and 3000 s/mm^2 (Sotiropoulos et al., 2013b; Ugurbil et al, 2013). Scanning each subject for 55 min enables acquisition of 90 diffusion orientations per shell and a total of 18 b = 0 scans. Each scan is repeated along two phase encoding directions (L/R and R/L) to allow correction of susceptibility induced distortions. Combined with the spatial resolution of 1.25 mm isotropic, this yields exceptional data quality for in vivo whole brain diffusion imaging at 3 T (Sotiropoulos et al., 2013b; Ugurbil et al., 2013). Structural scans include a pair of T_1-weighted and a pair of T_2-weighted images, all acquired at 0.7 mm isotropic resolution (Glasser et al., 2013), plus ancillary scans, for a session duration of ~40 min. The higher resolution compared to standard 1 mm structural scans improves the fidelity of cortical surface reconstruction and provides higher quality myelin maps (Glasser et al., 2013a; see below). The high quality of the structural, fMRI and dMRI data is illustrated below and in other articles in this special issue.

Head motion and physiological monitoring

Head movements, even small in magnitude, can have deleterious effects on MR data quality for all modalities. Fortunately, our preliminary analyses indicate that head motion is relatively low in the majority of HCP subjects. To further address head motion, in most scan sessions we acquired 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.

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. These physiology datasets were not ready at the time of the initial Q1 data release but will be included for all available datasets at the time of the Q2 release for use by other investigators. 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 with the quarterly data releases.

Image reconstruction and conversion to unprocessed NIFTI data

The raw data from each scan is converted into standard (16 bit) DICOM images through a set of modality-specific reconstruction processes. The 16 bit DICOMs allow for an extended dynamic range of signal intensity values, which is advantageous with such multi-channel receiver arrays where signal intensity variations can be quite large. Major improvements to the standard reconstruction process have been made in order to improve the data quality (especially for dMRI, Sotiropoulos et al., 2013b) and to reduce the reconstruction time for the very large HCP datasets (Ugurbil et al., 2013).

DICOM files for each scan are converted to standard NIFTI format (using dcm2nii made available by Chris Rorden — http://www.mccauslandcenter.sc.edu/mricro/mricron/dcm2nii.html), and all scan types containing potentially identifiable facial features are defaced (Milchenko and Marcus, 2013), with visual QC inspection to confirm successful defacing. Conversion to NIFTI also removes date stamps and other potentially sensitive information. The resultant NIFTI files constitute the unprocessed datasets that are part of the quarterly data releases.

7 T hardware and pulse sequences

Scanning of 200 subjects at 7 T will be done at UMinn using a Siemens 7 T scanner. 7 T provides increases in both the image SNR (Vaughan et al., 2001) and functional contrast-to-noise (Yacoub et al., 2001), compared to lower fields. This in turn permits the acquisition of much higher resolution images. Additionally, higher fields increase the relative sensitivity to the microvasculature in BOLD-based functional images (Ogawa et al., 1992; Ugurbil et al., 2003b; Uludag et al., 2009), resulting in a smaller point spread function (Shmuel et al., 2007).

Refinement and optimization of 7 T pulse sequences for the HCP began in 2012 and will be finalized for the acquisition phase commencing in the fall of 2013. Initial pilot studies have focused on fMRI and have produced high quality images at higher spatial resolutions (~ 1 mm) than the 2 mm isotropic voxels used for fMRI data acquired in HCP subjects at 3 T. The functional contrast to noise at such high resolutions is not compromised, despite the ~8 times smaller voxel size, because of the aforementioned increases in image SNR and BOLD based contrast (Ugurbil et al., 2013). The acquisition of such high resolution images will result in lower temporal resolution than the 2 mm isotropic resolution 3 T data, because the many more slices needed to cover the entire brain results in a substantial increase in the TR. Further, the requirement of in-plane acceleration, due to the higher resolution images combined with the much shorter T_2* at 7 T, limits the achievable multiband factor, because it also relies on the coil's sensitivity profile to accelerate the acquisition. Despite this, early results (see Ugurbil et al, 2013) indicate that ~1 mm isotropic resolutions over the whole brain are feasible with a TR of around 2 s. Further optimization of image reconstruction for such high resolution images is ongoing, in order to address several technical issues (e.g., increased sensitivity to motion, increases in B_0 inhomogeneity, and larger fMRI data rates).

Data processing and preliminary analyses

Unprocessed images from MRI scanners invariably contain several types of spatial distortion, are not in a standard anatomical space, and are misaligned across modalities. They also contain various types of modality-specific noise, artifacts, and biases. Many stages of processing are needed before analyses of neurobiological interest can begin in earnest. In order to make best use of the high-resolution HCP

datasets, it is critical to compensate as much as possible for these distortions, biases, and artifacts, and also to acknowledge the potential impact of residual confounds.

Processing of the HCP MRI data is subdivided into two broad categories. During the first two years of the HCP, intensive efforts were put into optimizing a set of preprocessing steps that compensate for spatial distortions and perform other useful transformations and operations, but minimize the overt loss of data or modification of the time course of fMRI time series data. The optimization process entailed critical evaluation and comparisons of how various existing and new methods performed, then packaging the best methods into a set of preprocessing pipelines appropriate for consistent and systematic application to all HCP datasets. The resulting preprocessing pipelines provide substantial improvements used for each of the MRI modalities, including structural MRI, fMRI (both rfMRI and tfMRI), and dMRI. Some of the refinements have already been incorporated into the latest versions of FSL, FreeSurfer, and Connectome Workbench, three major software packages used by the HCP pipelines. The HCP minimal preprocessing pipelines are described in detail in four other articles in this special issue (Barch et al., 2013; Glasser et al., 2013b; Smith et al., 2013; Sotiropoulos et al., 2013b) and are summarized only briefly below.

A second category of processing includes various steps to remove noise and minimize artifacts and biases that are characteristic to each modality. For fMRI, one set of issues revolves around de-noising, and removal of motion confounds. Another involves brain parcellation and network analysis. For dMRI, key issues involve fiber orientation estimation followed by probabilistic tractography. These 'additional processing' methods are still under active development within the HCP.

In the following discussion of each separate modality, we summarize the progress achieved in preprocessing methods, the current status of additional analysis strategies, and examples of interesting preliminary results obtained for that modality. We start with analyses that can be carried out using structural MRI data alone, followed by rfMRI and dMRI (the modalities most informative about connectivity), and finally tfMRI and MEG (the modalities most closely related to brain function).

Structural MRI and cortical shape analyses

For each subject, the HCP acquires a pair of T_1-weighted (T_1w) scans and a pair of T_2-weighted (T_2w) scans, both at a spatial resolution of 0.7 mm isotropic voxels. Obtaining higher resolution than conventional 1 mm isotropic voxels is important because many HCP analyses rely on cortical surfaces that are as accurate as possible. Each structural scan is evaluated by a trained rater to assess overall quality (poor, fair, good, and excellent), based on visual inspection of tissue contrast, spatial blurring, ringing, and other possible artifacts. The only scans used for structural preprocessing pipelines and released to the community are those in which one or more good/excellent T_1w and T_2w scans were acquired in the same session (and accompanied by corresponding receive and transmit bias field maps that are used in preprocessing).

The HCP structural pipelines use FreeSurfer 5.1 software plus a series of customized steps that combine information from both T_1w and T_2w scans for more accurate white and pial surfaces. Fig. 1A shows a parasagittal slice through a T_1w scan from one HCP subject, along with surface contours for the 'pial' and 'white' surfaces generated by FreeSurfer. This illustrates the high quality of the structural images themselves and of the cortical segmentation, including regions where cortex is notably thin, such as the calcarine sulcus (red arrow) and precentral sulcus (black arrow). The fine detail in the cerebellum is also notable, as most lamellae and even many individual folia are discernible.

Cortical myelin maps are another useful type of data that can be extracted from structural images by computing the ratio of the T_1w and T_2w image values at each voxel and mapping this ratio to the cortical surface (Glasser and Van Essen, 2011). Figs. 1B, C show myelin maps displayed on inflated hemispheres of the same subject. In general, the myelin maps for this and the other HCP subjects are higher in quality than those originally reported (Glasser et al., 2011), thanks to the higher spatial resolution (0.7 vs 1 mm isotropic voxels) coupled with several algorithmic improvements (Glasser et al., 2013a,b).

Registration to atlas space includes an initial volumetric registration to MNI152 space using FSL's linear FLIRT tool, followed by the nonlinear FNIRT algorithm, which does an excellent job of aligning subcortical structures. Cortical surface alignment benefits from a subsequent stage of surface-based registration to a population-average surface, using FreeSurfer to register each hemisphere to a separate left and right atlas surfaces based on matching of cortical folding patterns (Fischl et al., 1999). This is followed by registration to the Conte69 atlas, which brings the left and right hemispheres into precise geographic alignment using interhemispheric landmark-constrained registration (Van Essen et al., 2012). Accurate interhemispheric registration facilitates a variety of cross-hemisphere comparisons, such as the correspondence of myelin maps in the left and right hemispheres in individual subjects. For example, in Figs. 1B and C, eight vertices centered on hotspots of heavy myelin (MT+, FEF, and two others) are highlighted in the left hemisphere (black dots). The symmetry in the pattern of myelin content between the two hemispheres can be appreciated by comparing the location of corresponding vertices in the two hemispheres, which were selected to be centered on myelin hotspots in the left hemisphere (black circles) and are approximately centered on corresponding hotspots in the right hemisphere (blue dots).

A wide variety of morphometric and heritability analyses will be feasible to carry out using HCP structural datasets. Such analyses can capitalize on the high quality of HCP structural scans, surface reconstructions, and myelin maps; the associated behavioral data available for each subject; and the availability of family structure information (e.g., twin vs.or nontwin status). For example, Fig. 2 shows maps of cortical shape for two pairs of identical twins (A and B), displayed on the inflated atlas right hemisphere; these are FreeSurfer 'sulc' maps, in which bright regions represent gyral crowns and dark regions represent buried cortex (the darker the shading the deeper the sulcus). On visual inspection, there are many differences in these 'shape maps' (e.g., arrows and highlighted vertices). The differences in shape maps for identical twins (A1 vs A2; B1 vs B2) are comparable to those between unrelated individuals (either 'A' subject vs either 'B' subject). This is consistent with previous research suggesting that cortical folding patterns are only modestly heritable (Botteron et al., 2008), but extensive data on MZ and DZ twins and their siblings in the HCP datasets will enable detailed analysis of the heritability of cortical shape, myelin maps, and many other attributes, including the connectivity and functional data discussed below.

Resting-state fMRI

Preprocessing of fMRI data (both resting-state and task-fMRI) involves two pipelines, one carried out entirely on the volume data. The second involves mapping the data to cortical surfaces and subcortical gray-matter domains using the recently introduced CIFTI data format that offers several advantages (Glasser et al., 2013b; Marcus et al., 2013). CIFTI is predicated on the dual notion of (i) restricting data storage and analysis to just the gray matter domains of interest (hence bypassing the storage of white matter and non-brain data), and (ii) representing gray matter in a way that respects its natural geometry: surface vertices for cerebral cortex and voxels for subcortical gray matter. This is reflected by the term "grayordinate", which includes any surface vertex or subcortical voxel that represents gray matter.

Temporal filtering and de-noising

Neurobiologically relevant fluctuations, which ideally should be the only signals used to drive functional connectivity analyses, represent only a small fraction (~4%) of the total temporal variance in the

Please cite this article as: Van Essen, D.C., et al., The WU-Minn Human Connectome Project: An overview, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.05.041

minimally preprocessed datasets (Glasser et al., 2013b; Marcus et al., 2013). Hence, it is crucial to eliminate as much as possible the artifacts and noise, while preserving as much signal as possible. Our overall aim is to be thorough in removing aspects of the data that can be identified as artifact with reasonably strong specificity, while taking a more minimalist approach to removing ambiguous or mixed (signal + noise) data components. For example, the HCP does not apply temporal lowpass filtering, because the highest frequencies cannot be considered to only contain artifact. Similarly, very unaggressive highpass temporal filtering is applied, quite close in effect to linear detrending. In both cases, it is easy for researchers to subsequently apply their own, more aggressive, temporal filtering on the downloaded datasets, should they choose to do so.

One promising approach to removing structured artifacts from the minimally preprocessed data involves application of independent component analysis (ICA) denoising to each 15-minute rfMRI dataset. FSL's MELODIC tool (Beckmann and Smith, 2004) is used to decompose the data into multiple (typically ~230) components, each comprising a single spatial map and an associated timecourse. Some components represent artifacts such as head motion or cardiac pulsation, while others represent valid neuronally-related spontaneous fluctuations. A new tool called 'FIX' (FMRIB's ICA-based X-noiseifier; Salimi-Khorshidi et al., 2013, in preparation) is used to automatically classify components into "bad" versus "good". The bad components' timeseries are then regressed out of the data, along with various head-motion-related confound regressors. FIX has been hand-trained and tested on one hundred 15-minute HCP datasets, and has achieved better than 99% accuracy rate in correctly classifying components. The resulting resting-state network timeseries show exceptionally clean power spectra (Smith et al., 2013).

Despite the success of the above cleanup process for structured artifacts, spatially more global artifacts can remain in the data. This may include motion artifacts (Power et al., 2012) that are not fully removed by the above processing steps, and which may artifactually influence correlation-based estimates of functional connectivity. Ongoing analyses and discussions within as well as outside the HCP consortium may provide a better understanding of the residual global and motion confounds, as well as additional options for reducing them further.

Following preprocessing and artifact removal, an important next stage in HCP connectome analysis is the generation of "dense connectomes", either at single-subject or group level. A dense connectome is the full (voxels × voxels) or (grayordinates × grayordinates) correlation matrix obtained by correlating the timeseries of every brain voxel or every grayordinate with every other brain voxel or every grayordinate. These matrices are massive (190 GB and 32 GB respectively); the major data reduction by shifting from a voxel-based to grayordinate-based representation is immediately apparent.

Once dense functional connectomes have been generated for individuals or groups, they can be used in several neurobiologically interesting ways. Two powerful and complementary approaches involve seed-based correlation analysis and ICA-based analysis of network organization; both approaches are used extensively within the HCP consortium. Fig. 3 illustrates functional connectivity maps in an individual HCP subject for two seed locations, one in retrosplenial cortex (Fig. 3A, black arrow) and the other just a few mm more dorsal in posterior cingulate cortex (Fig. 3B, white arrow). Many of the regions that are strongly correlated (yellow, red) with the retrosplenial seed are poorly correlated or anti-correlated with the nearby seed in cingulate cortex (blue, purple; but note this is after regression of the mean gray-matter timecourse — see below). These striking differences in functional connectivity for nearby locations reflect several factors, including the high quality (and large amount) of data acquired from each subject; the use of preprocessing and analysis steps that respect the topology of the cortical sheet; and the advanced methods used to reduce noise and artifacts. These and many other comparisons that can be used during seed-based analyses take advantage of 'point-and-click' interrogation of remotely stored dense connectome datasets available in the Connectome Workbench visualization platform (see below).

Fig. 4 shows a functional connectivity map for a seed location in lateral parietal cortex, probing a dense connectome generated by concatenating rfMRI timeseries data from 20 HCP subjects. Several points merit comment. (i) The signal-to-noise improves substantially

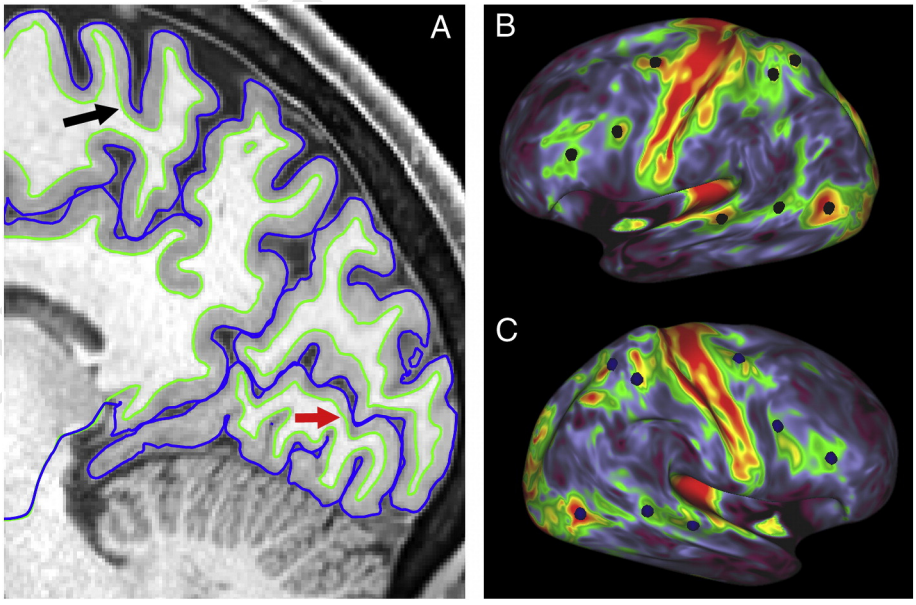


Fig. 1. A. Parasagittal slice through posterior cortex of T_1w image from subject A1 (study-specific code), with accurate pial and white surface contours, even where cortex is thin (arrows). The fidelity with which the FreeSurfer white and pial surfaces track the anatomical boundaries is much better than the initial surfaces generated by running FreeSurfer 5.1 on 1 mm isotopic T_1w data from the same subject (cf. Figs. 11, 12 in Glasser et al., 2013b). B, C. Myelin maps on inflated left and right hemispheres of subject A1. Highlighted vertices centered on myelin hotspots in the left hemisphere (B, black) have geographically corresponding vertices located within myelin hotspots in the right hemisphere (C, blue). The myelin maps illustrated here are improved over those available in the HCP Q1 data release by virtue of a step that reduces residual low spatial frequency biases by subtracting a highly smoothed population-average myelin map (see Glasser et al., 2013a Fig. 22 and associated text for details).

Please cite this article as: Van Essen, D.C., et al., The WU-Minn Human Connectome Project: An overview, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.05.041

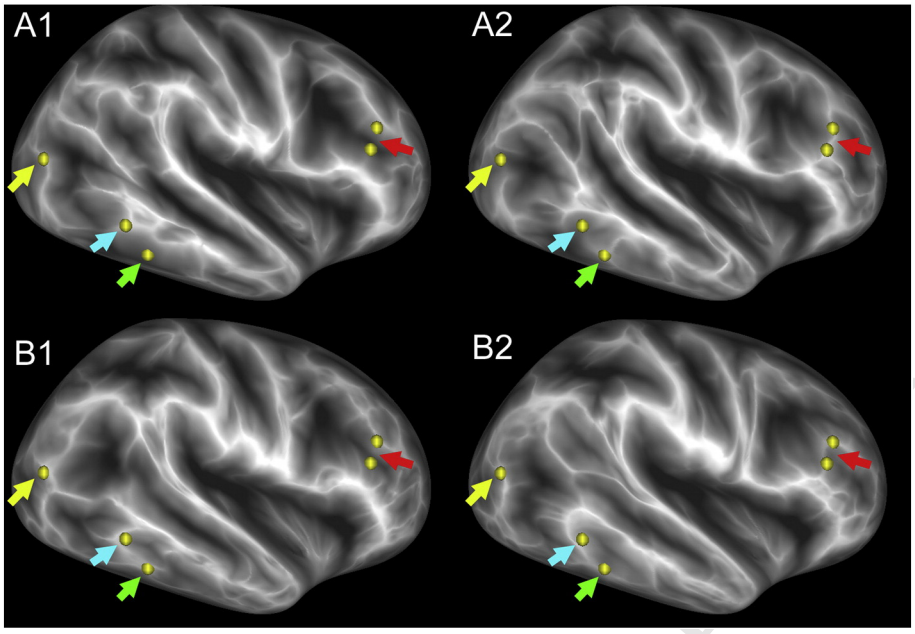


Fig. 2. Cortical shape features in identical twins. Highlighted vertices are locations on a gyral crown (white ridge) in twin A1 (yellow, blue arrows) or in twin A2 (red, green) but are deeper in a sulcus in the 'geographically corresponding' location in the other twin. Subjects are identified in a study-specific code (A1, A2, B1, B2) in conformance with the Restricted Access Data Use Terms (see below).

by virtue of the large group size. (ii) The functional connectivity hotspots associated with this location are spatially more blurred than equivalent maps derived from single subject datasets, owing to the fact that shape-based inter-subject registration can be inaccurate in aligning functionally defined areas, especially in regions of high folding variability. (iii) In contrast to Fig. 3, these correlations are estimated without regression of the mean gray timecourse. Hence, the anti-correlated regions (blue, purple) are smaller in extent, because the mean is not forced to be zero. The neurobiological interpretation of different types of representation (full correlation; correlation after mean gray-matter timecourse regression; and the partial correlation approach illustrated below) is not well understood, and none should be considered a perfect measure of direct anatomical connectivity. The analysis strategies that are neurobiologically most informative remain under active investigation (e.g., Smith, 2012; Smith et al., 2013).

Another major objective is to use functional connectivity data for parcellating the brain into distinct parcels, or subdivisions. Classical parcellations of cortical areas and subcortical nuclei commonly assume that each parcel is topologically contiguous and is non-overlapping with neighboring parcels (aside from the experimental uncertainties in areal boundaries). Several approaches to brain parcellation based on functional connectivity have been explored, including methods based on spatial gradients (Cohen et al., 2008; Smith et al., 2013), snowball sampling (Wig et al., 2013); and region-growing (Blumensath et al., 2013). These efforts are still in early stages of development and must cope with two fundamental challenges: (i) the strength, or sharpness of transitions in functional connectivity vary widely and can be influenced by noise and biases in individual subjects; and (ii) the fidelity of inter-subject alignment using shape-based surface registration methods is imperfect in regions of high folding variability, resulting in misalignment and spatial blurring of functional connectivity gradients (cf. Robinson et al., 2013; Van Essen et al., 2012b).

ICA provides a powerful alternative approach to subdividing the brain into regions that functionally have a high degree of independence, but are not constrained to be topologically contiguous or non-overlapping. For example, Fig. 5A shows cortical surface maps of five example ICA components from a 22-component group-level ICA-based network analysis carried out on 20 HCP subjects (the same group as in Fig. 4). The ICA approach can support a much finer-grained spatial analysis involving hundreds of ICA components (see Fig. 10 below and Smith et al., 2013), but the coarser-grained analysis shown here is useful for illustrative purposes. ICA component 1 covers higher-level visual areas. ICA Component 7 includes the central visual field representation of V1 and V2, whereas component 3 mainly involves the peripheral visual field representation of these two areas. This fits with evidence for a major transition in functional connectivity that cuts across both V1 and V2 in their mid-eccentricity range (Yeo et al., 2011); it implies that network (parcel) boundaries defined by functional connectivity do not always respect classical areal boundaries (for other examples, see Yeo et al. 2011; Power et al., 2011; Van Essen and Glasser, in press). ICA components 12 and 15 include several parts of the default mode network, and support the hypothesis that this network includes functionally distinct subregions (Andrews-Hanna et al., 2010).

Fig. 5B illustrates how "parcellated connectomes" can be derived from the preceding ICA-based analysis. Each ICA component (parcel) has an associated timeseries (representing timeseries from voxels/grayordinates in that parcel), and the parcels × parcels network matrix can be generated, for example, just by correlating these $N_{parcels}$ timeseries with each other. The matrix entries below the diagonal represent the full correlation, whereas those above the diagonal represent the partial correlation matrix (each pairwise correlation is estimated after regressing out the other $N_{parcels}$-2 timeseries). The parcels are organized into groups that are most similar in their timeseries based on a hierarchical clustering analysis applied to the full correlation matrix. Both the full correlation matrix and the partial correlation matrix represent mathematically well-defined entities; however, as alluded to above, neither should be regarded as an explicit, validated indicator of direct anatomical connectivity, although significant values in the partial correlation matrix will hopefully have a high probability of reflecting genuine connections (Smith, 2012).

The preceding examples illustrate how parcellations can be generated and analyzed using group data, where the signal-to-noise is high. One strategy for the future will be to apply parcellations derived at the group-level (from multiple subjects' dense connectomes combined) to each individual subject. Then a parcellated connectome matrix could

Please cite this article as: Van Essen, D.C., et al., The WU-Minn Human Connectome Project: An overview, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.05.041

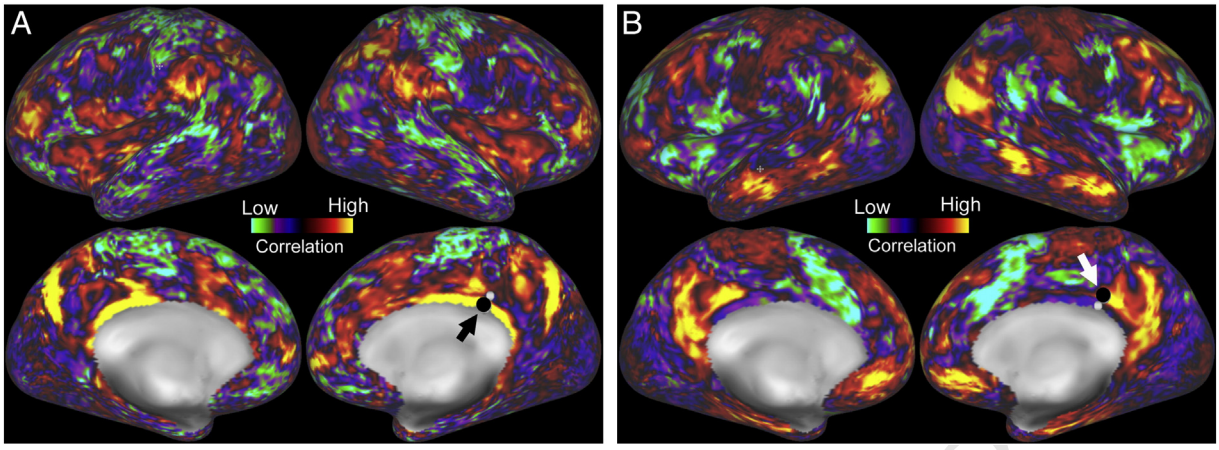


Fig. 3. A. A map of functional connectivity (after regression of the mean gray timecourse) in the left and right hemispheres of an individual HCP subject associated with a seed location in right retrosplenial cortex (black arrow, black circle). B. A functional connectivity map for a nearby seed location (white arrow, black circle) in cingulate cortex (part of the default mode network).

be generated based on the data from each subject. These subject-specific parcellated connectomes can then be averaged across subjects, or investigated to see how aspects of the matrices co-vary with behavioral or genetic factors. Such an approach offers the advantage of consistency based on a single parcellation based on a group average (a given parcel "means the same thing" in all subjects), but would not be optimal in compensating for intersubject differences in the size and location of each parcel.

Diffusion MRI analyses

The preprocessing and analysis of dMRI data involve a very different set of technical considerations than those just discussed for rfMRI. However, the overarching approach adopted by the HCP is similar: capitalize on the high quality of the acquired data by minimizing distortions, maximizing spatial registration, and addressing the residual confounds using the best methods available.

Extensive effort has been dedicated to improvements in preprocessing of the diffusion data, to improve fiber reconstruction (Sotiropoulos et al., 2013a). For example, combining data across multiple receive coils using a sensitivity-encoding method (SENSE-1) increases the dynamic range of the signal relative to the conventional root-sum-of-squares approach (Lenglet et al., 2012; Sotiropoulos et al., 2012, 2013b). We also developed a novel algorithm that greatly improves the correction of susceptibility and eddy-current induced distortions and the effects of subject motion (Andersson et al., 2012; Sotiropoulos et al., 2013b). The resultant preprocessed dMRI datasets are available to the community as part of the Q1 data release. Data from any individual shell ($b = 1000, 2000$, and 3000 s/mm^2) can be used with standard fiber reconstruction techniques, but methods that make use of all three shells will get the largest benefit. In the initial Q1 data release, the preprocessed dMRI data are in the coordinate system of the individual diffusion scans. However, for the Q2 data release (including a reprocessed Q1 dataset) and all future releases the data will be aligned to the native structural space in order to facilitate various cross-modal comparisons (see below).

HCP has developed novel fiber reconstruction algorithms that are optimized for multi-shell data (Jbabdi et al. 2012). These have not yet been applied to the full Q1 dMRI datasets, but they will be made available in future data releases. Probabilistic tractography has been applied

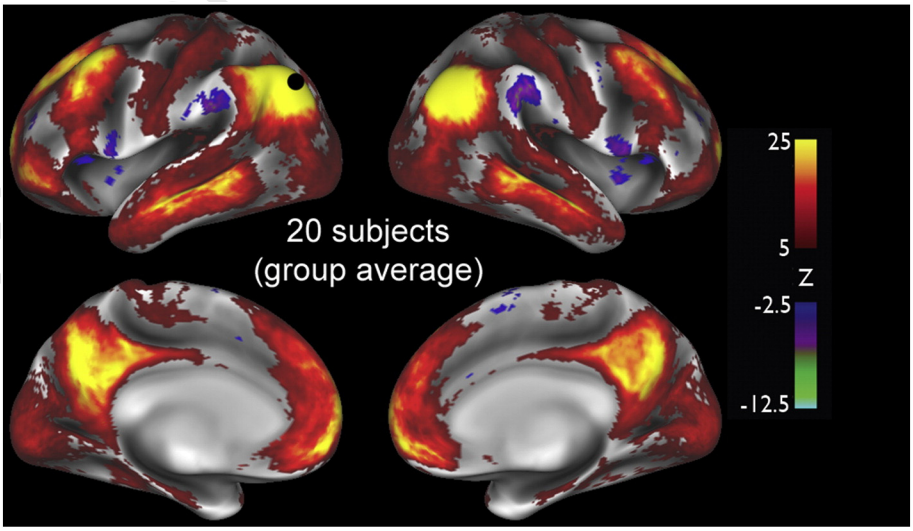


Fig. 4. A map of functional connectivity (full correlation converted to Z-statistics) in the left and right hemispheres associated with a seed location in the left parietal cortex (part of the default mode network), from a group average functional connectivity analysis (20 subjects from the HCP Q1 data release, but not the same as the standard '20 unrelated' subjects). Positive correlations are thresholded at $Z > 5$ and negative correlations are thresholded at $Z < -2.5$.
Adapted, with permission, from Smith et al. (2013).

Please cite this article as: Van Essen, D.C., et al., The WU-Minn Human Connectome Project: An overview, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.05.041

to some of these datasets using FSL's existing probabilistic tractography approaches to generate dense connectomes in grayordinate space (Behrens et al., 2007; Sotiropoulos et al., 2013b).

Fig. 6 shows representative fractional anisotropy and color-encoded principal diffusion direction images from the HCP dMRI data, compared with a more conventional 2 mm dataset (from a different subject). The improvement in anatomical detail is clearly visible.

The complex 3D trajectories resulting from probabilistic tractography analysis pose special challenges, in terms of the large size of the data files, the complex formats needed to encode probabilistically computed streamlines, and the need to visualize the 3D trajectories themselves, as well as where they intersect with cortical surfaces and subcortical nuclei. To this end, Connectome Workbench includes the capability for interactive 'point-and-click' visualization of probabilistic trajectories (Fig. 7A). This enables users to access the large trajectory files remotely by uploading only the trajectory data requested for the selected seed location. For example, Fig. 7A shows the connectivity trajectory for a seed location in the lateral prefrontal cortex. The figure shows a full 3D view of a probabilistic trajectory in a 'whole brain view' that includes brain slices and surface contours for a 3D reference frame (panel A left), and the trajectory's intersection with a single sagittal slice (panel A right). Panel B shows the average gray-to-gray connectivity from 9 subjects seeded at the same point on a pial (left) and inflated (right) hemisphere. Panel C shows average resting state functional connectivity from the same source location. These different views and datasets are easily integrated in a single Workbench screen that allows for yoked visualization of connectivity in each view.

Efforts will continue to further improve fiber orientation modeling as well as tractography algorithms that take advantage of the richness of the HCP data. While containing a wealth of information, dMRI connectomes will inevitably contain biases and errors resulting from limitations of the technique. Some of these are familiar (Jbabdi and Johansen-Berg, 2011), but generating and interpreting entire gray-to-gray connectomes bring new challenges. For example, a notable bias, present for clear geometric reasons, is that current tractography approaches are much more likely to trace to gyral crowns than to sulcal

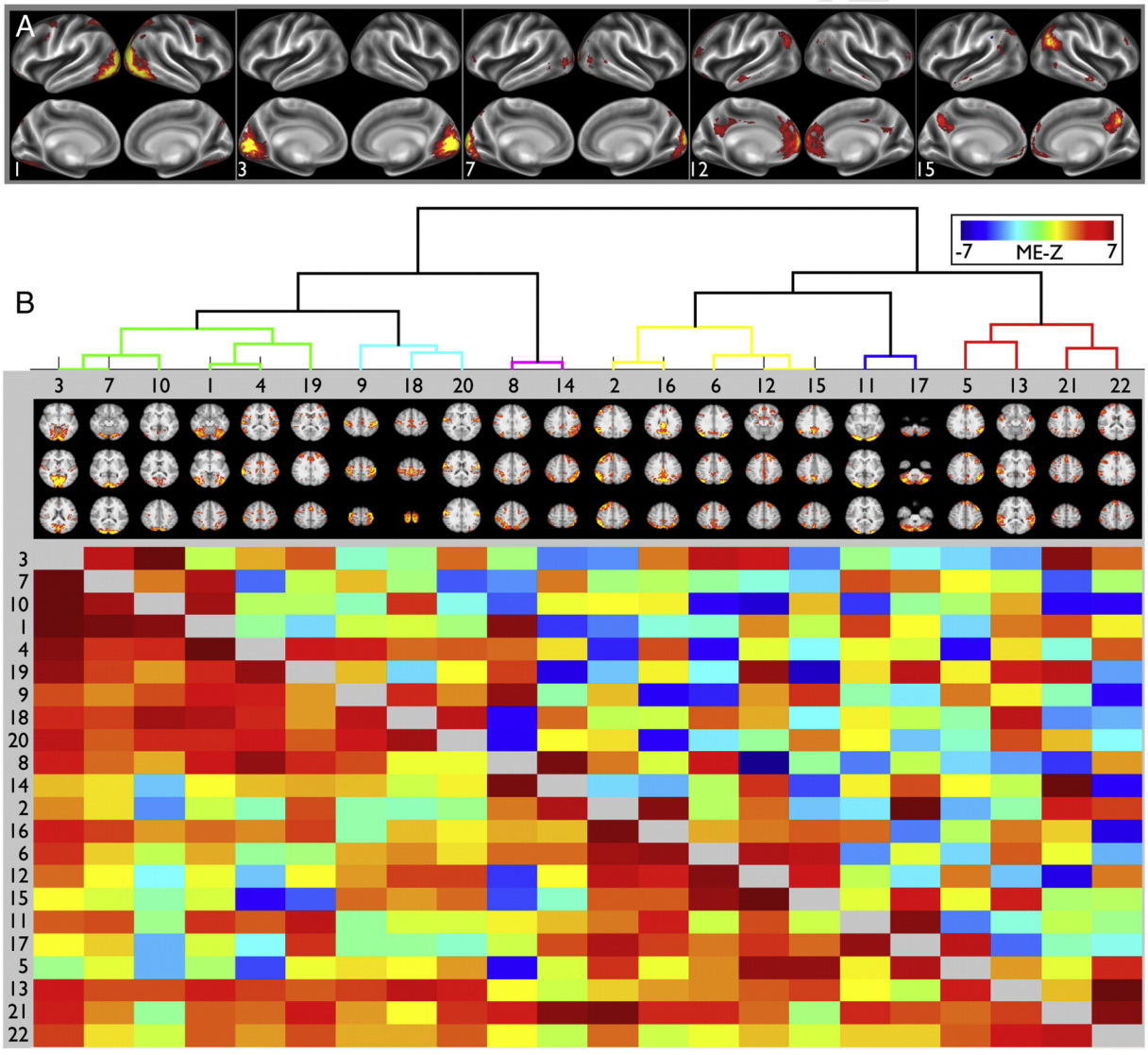


Fig. 5. A. Five example components from a 30-component ICA analysis (8 were discarded as being either artifact or being inconsistent across subjects) displayed on inflated cortical atlas surfaces. B. 22 × 22 correlation matrices (group-average parcellated connectomes) derived from the timeseries associated with the 22 group-ICA components. Full correlation is shown below the diagonal; partial correlation above the diagonal. Each row or column is the set of correlations (red, yellow) or anti-correlations (green, blue) between a single network matrix "node" and all other nodes; the nodes were reordered from the original ordering, according to a hierarchical clustering algorithm (depicted at the top). The network analysis and figure generation was carried out using the FSLNets package (fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSLNets).
Adapted from Smith et al. (2013).

Please cite this article as: Van Essen, D.C., et al., The WU-Minn Human Connectome Project: An overview, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.05.041

depths (Van Essen et al., 2013). Detailed comparisons in macaque monkeys with both histology (Sotiropoulos et al., 2013b; Van Essen et al. 2013) and invasive tracer studies (Jbabdi et al., 2013) will better inform our understanding of such biases, and the most attractive strategies for next generation tractography algorithms.

Task-fMRI (tfMRI) analyses

The HCP acquisition protocols include seven tfMRI paradigms, three of which (working memory, reward processing and motor processing) follow 30 min of rfMRI in one imaging session, and four of which (language, social cognition, relational processing and emotion processing) follow 30 min of fMRI in a second imaging session (Barch et al., 2013). The spatial preprocessing steps for tfMRI are identical to those used for rfMRI, both for the volume-based and surface-based aspects (Glasser et al., 2013b).

For the Q1 data release, we completed more extensive processing on all of the tfMRI data from 20 subjects who were unrelated to each other, using both volume and grayordinate-based (i.e., surface-based) tfMRI processing. The task modeling was carried out using FSL's FILM tool (FMRIB's Improved Linear Model, Woolrich et al., 2001), adapted for the grayordinate data such that FILM's spatial regularization of the temporal prewhitening is constrained to gray matter. Both approaches indicated excellent quality data from these paradigms, with clear group level activation as well as robust activation within individual subjects in many of the paradigms and contrasts. Here we provide two examples of this. Fig. 8 displays the results from the working memory task, a variant of the N-back task, with the specific contrast a high working memory load ("2-back") versus a low working memory load ("0-back"). The data for this task are acquired in ~10 min and show robust mixed-effects group level activation in dorsal frontal–parietal and cingulate systems typically associated with working memory and cognitive control, in both the volume and grayordinate analyses. Further, we see significant activation in these same regions in the majority of individual subjects, a result important for the individual difference and genetic analysis goals of the HCP.

As another example, Fig. 9 displays results from the language processing task developed by Binder et al. (2011), with the specific contrast being story processing versus math. These data are acquired in approximately 8 min, and show robust group level activation in anterior and inferior temporal regions, as well as ventral prefrontal regions typically associated with various components of language processing. As with the working memory task, we also see activation in these same regions in the majority of individual subjects. Taken together, these data illustrate our ability to acquire high quality tfMRI data from a range of paradigms. These data will provide rich information at both the group and individual subject level and offer complementary information for the parcellations and connectivity analyses from both the rfMRI and dfMRI acquisitions.

Cross-modal comparisons

The availability of information from multiple imaging modalities in individuals and group averages greatly increases the utility of the HCP datasets, and it will benefit from improved capabilities for cross-modal analysis and visualization. One such example has already been illustrated in which rfMRI-based functional connectivity is compared to dMRI-based structural connectivity (Fig. 7). Fig. 10 shows another example of cross-modal comparison that also illustrates the utility of being able to visualize fMRI data mapped to a cerebellar surface map. The top row shows the group-average task activation from the right-hand hand movement task, analyzed for the same group of 20 unrelated subjects shown in preceding figures. It includes activation in the expected location in the left motor cortex (left panel), and also at two distinct locations in dorsal and ventral cerebella matching published reports (Buckner et al., 2011). The bottom row shows a spatially corresponding ICA component from a 100-component group-level ICA-based network decomposition (with 82 'signal' components), carried out on 66 HCP subjects from the Q1 data release. The correspondence in spatial patterns between the rfMRI ICA component and the task-fMRI activation is striking.

More generally, there will be countless analyses that benefit from the ability to compare data across as well as within modalities, in individual subjects and in group averages. Besides having the data in a common spatial framework, it is also important for the data to be compactly represented (re-emphasizing the advantages of the CIFTI format over standard NIFTI volumes) and to take advantage of the flexible visualization options provided by Connectome Workbench.

Subcortical signals

Subcortical gray matter (excluding cerebellar cortex) constitutes about 8% of brain volume; remarkably, the many vital roles of subcortical nuclei in brain function are achieved with fewer than 1% of the total number of brain neurons (Azevedo et al., 2009). It is obviously very important that subcortical regions be well integrated into the HCP analyses of brain connectivity and function. Although not emphasized in the present article, the HCP data do include robust task activations and resting-state networks from the fMRI data (Barch et al., 2013; Smith et al., 2013). However, the SNR for subcortical regions is generally weaker than for cerebral and cerebellar cortical regions, in a large part because of their buried location relative to the 32-channel head coil (Ugurbil et al, 2013). In terms of visualization, recent advances in Connectome Workbench support montage views that display volume slices restricted to subcortical domains alongside surface views of cerebral and cerebellar cortex, thereby allowing each domain to be represented using a visualization format appropriate for its topology.

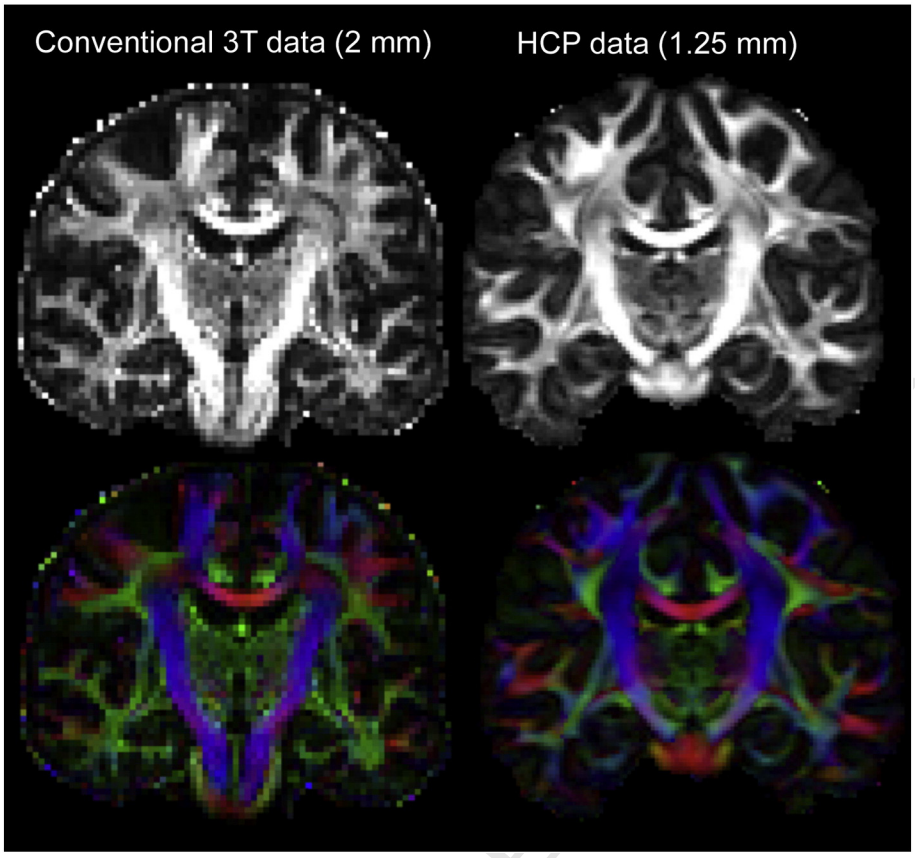
MEG acquisition and analysis

As noted previously, MEG will be acquired concurrently in 100 HCP subjects, starting in the spring of 2013. The obvious advantage of MEG over MRI is the much higher temporal resolution (milliseconds vs seconds), but it occurs at the expense of coarser spatial resolution (centimeters instead of millimeters). The session protocol includes resting-state scans (rMEG) plus three task-evoked scans (tMEG) involving a modified version of the working memory task being used in tfMRI, a modified version of the motor processing task being used in tfMRI, and a modification of a language task piloted during Phase I for tfMRI.

MEG data will be collected on a whole head MAGNES 3600 (4D Neuroimaging, San Diego, CA) system housed in a magnetically shielded room at Saint Louis University. Preprocessing to improve signal quality includes ICA-based removal of physiological artifacts (Escudero et al., 2007). An anatomy processing pipeline links MEG to structural MRI by co-registration of the MEG sensors to the anatomical coordinate system of the sMRI data, followed by generation of volume conduction models of the head to provide anatomical constraints for source localization analyses.

Accurate source reconstruction is a critical prerequisite for comparing electrophysiological results to those obtained from other imaging modalities. HCP will use three source reconstruction strategies, all supported by the FieldTrip Toolbox (Oostenveld et al., 2011). Resting state analyses will use a model-driven approach to computing the inverse solution. Specifically, weighted minimum-norm estimates (wMNE) will be used to generate computationally efficient and reliable projections of resting activity into source space (de Pasquale et al., 2010, 2012; Mantini et al., 2011). Task data will be analyzed using two beamformer reconstruction approaches, which are adaptive, data-driven methods for deriving the inverse solution from empirical evidence (sensor-space covariance or cross-spectral density). Linear constrained minimum variance beamformers (LCMV) reconstruct source space data in the time domain and are useful for inferring

Please cite this article as: Van Essen, D.C., et al., The WU-Minn Human Connectome Project: An overview, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.05.041



Q2 **Fig. 6.** The figure shows representative fractional anisotropy and color-encoded principal diffusion direction images from the HCP dMRI data, compared with a more conventional 2 mm dataset (from a different subject). The improvement in anatomical detail is clearly discernible. For example, many white matter tracts appear thicker (less partial voluming). The imaging protocol for the conventional data was as follows: Siemens 3 T Verio, 2 mm isotropic voxels, 64 slices, 60 directions, 2 averages with reversed phase encoding polarity, b = 1500 s/mm ^ 2, TE/TR = 86/10,000 ms, GRAPPA = 2, scan time = 20 min.

connectivity in oscillatory brain activity (Brookes et al., 2011; Schoffelen and Gross, 2009). Dynamic imaging of coherent source (DICS) reconstructs source-space data in the frequency domain (Gross et al., 2001; Van Veen et al., 1997).

Following source reconstruction, both seed-based and data-driven group-ICA methods will be used to analyze dynamic connectivity. Additional processing will be used to relate the electrophysiological connectivity matrices to the parcellations used for analyzing functional and structural connectomes. MEG source reconstructions may include up to ~8000 nodes (hence, electrophysiological connectivity estimates between 64 million node pairs). Dense connectivity matrices generated via fMRI or dMRI will have an order of magnitude more grayordinates, but a much smaller number (hundreds) of functionally or anatomically distinct parcels. For visualization, the electrophysiological data will be mapped onto this anatomically parcellated representation. The availability of resting and task MEG data in ConnectomeDB will enable the exploration of multiple features of the data using both existing and yet to be developed analysis techniques. In the future, more elaborate connectivity metrics are likely to become available.

Informatics and data sharing

The HCP has adopted a multifaceted approach to data sharing and Q39 data mining (Marcus et al., 2013). The Q1 data release (March 2013) includes three distinct levels of data analysis: the unprocessed image files (after image reconstruction and DICOM to NIFTI conversion); the minimally preprocessed data; and an additionally processed group average dataset. This amounts to ~2 terabytes in total for the 68 subjects. The final amount of HCP data may approach 1 petabyte once all acquisitions and analyses have been done on all 1200 subjects (including 7 T and MEG/EEG scans).

For the Q1 data release, the primary mode of access involves downloading pre-packaged archives organized by subject, scan modality, and level of processing. This includes several pre-packaged archives (a single subject; five unrelated subjects; and 20 unrelated subjects), which allows exploratory analyses without the complications of dealing with family structure (see below). Because data transfer can be notoriously slow when using standard ftp protocols, especially internationally, we adopted a UDP-based commercial high speed data transfer technology (Aspera fasp™), which has performed well in pilot testing and in the early stages of the Q1 data release. To date, the great majority of investigators have elected to download the minimally preprocessed datasets rather than the unprocessed NIFTI files, thereby capitalizing on the HCP preprocessing pipelines described above.

The ConnectomeDB database enables selection of subjects based on a large number of behavioral phenotype data types that are stored in the database and available for each subject. Currently, these search capabilities are mainly useful for selecting subgroups of subjects from the Q1 data release for download. This is at present of limited utility, given the relatively small number of subjects available for the first quarterly release. However, more extensive data mining capabilities will be added, and the number of subjects will of course increase with successive quarterly releases.

Datasets will be released on a quarterly basis in order to avoid data management problems that would arise if the data came out in smaller 'dribs and drabs'. Moreover, the extensive data processing and QC efforts that are essential for the data to be maximally useful to the community currently require several months between the end of a

Please cite this article as: Van Essen, D.C., et al., The WU-Minn Human Connectome Project: An overview, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.05.041

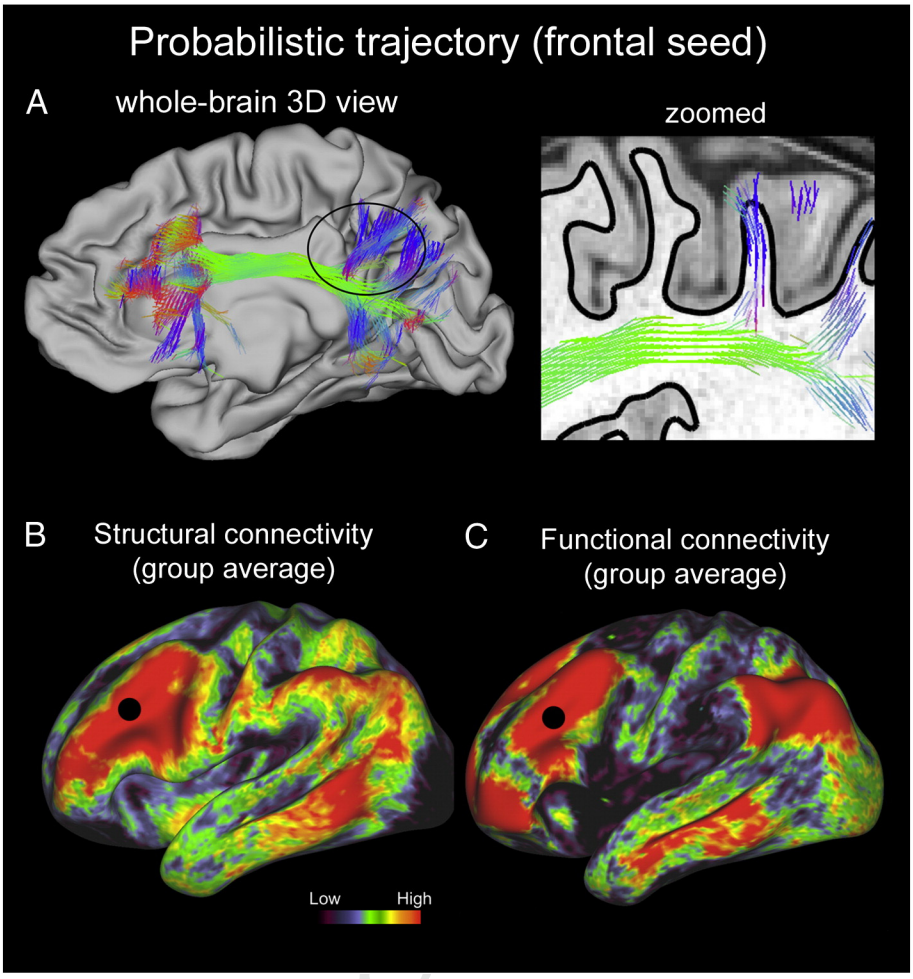


Fig. 7. Structural connectivity in an individual and in group averages and in comparison to functional connectivity. A. Connectivity trajectory visualization for a single HCP subject (100307). Probabilistic trajectories seeded from a single grayordinate in left frontal cortex and intersecting the white/gray matter boundary surface in at least one more location are shown on the left panel; the right hemisphere's midthickness surface provides a spatial reference. The inset (right) displays a part of the trajectories for a single sagittal slice, overlayed on a T_1w image (white/gray matter boundary shown with the black solid line). B. Structural connectivity values in a group average (9 HCP subjects) for the same seed location (black dot), viewed on the inflated cortical surface. The values are displayed using a logarithmic scale. C. Functional connectivity values for the same seed location, displayed on the inflated surface. The values correspond to the average functional connectivity of a group of 20 HCP subjects.

quarter's data acquisition and when the data are ready for release. Thus, each release will cover data acquired up until approximately three months prior to the release.

In general, our intent is for each quarter's data release to be incremental, by adding to datasets released in preceding quarters. However, between the Q1 and Q2 release, a number of significant refinements were made in the pipelines for each of the MRI modalities. Hence, the Q2 release will also include a complete regeneration of the minimally preprocessed data from Q1 along with the newly processed Q2 datasets. The differences between the original and reprocessed versions of the minimally preprocessed datasets are expected to be small (except for the aforementioned change in the coordinate space for dMRI data), but investigators who have already begun analyses using the initial Q1 datasets will need to be mindful of these changes before combining data for subjects acquired in different quarters.

Connectome Workbench is a platform that has been customized for analyzing and visualizing each of the MRI-based imaging modalities acquired for the HCP. It includes command-line utilities that support (along with FSL and FreeSurfer) many of the preprocessing pipelines and subsequent analysis functionality. Some of the capabilities of the Workbench visualization platform have been demonstrated in the figures contained in this paper and in the other HCP articles in this special issue. Workbench is especially well suited for handling grayordinate representations (surface vertices and gray-matter voxels) in the CIFTI format (see Glasser et al., 2013b; Marcus et al., 2013). Q40

Open access and restricted access datasets

To aid in the protection of participants' privacy, the HCP has adopted a two-tiered data access strategy (http://www.humanconnectome.org/data/data-use-terms/). Every investigator must agree to FieldTrip Toolbox. An additional set of Restricted Data Use Terms applies to an important subset of the non-imaging data and is essential for preventing any inappropriate disclosure of subject identity.

The released HCP data are not considered de-identified, insofar as certain combinations of HCP Restricted Data (available through a separate process) might allow identification of individuals as discussed below. It is accordingly important that all investigators who agree to Open Access Data Use Terms consult with their local IRB or Ethics Committee to determine whether the research needs to be approved or declared exempt. If needed and upon request, the HCP will provide a certificate stating that an investigator has accepted the HCP Open Access Data Use Terms.

Because HCP participants come from families with twins and non-twin siblings, there is a risk that combinations of information about an individual (e.g., age by year; body weight and height; handedness)

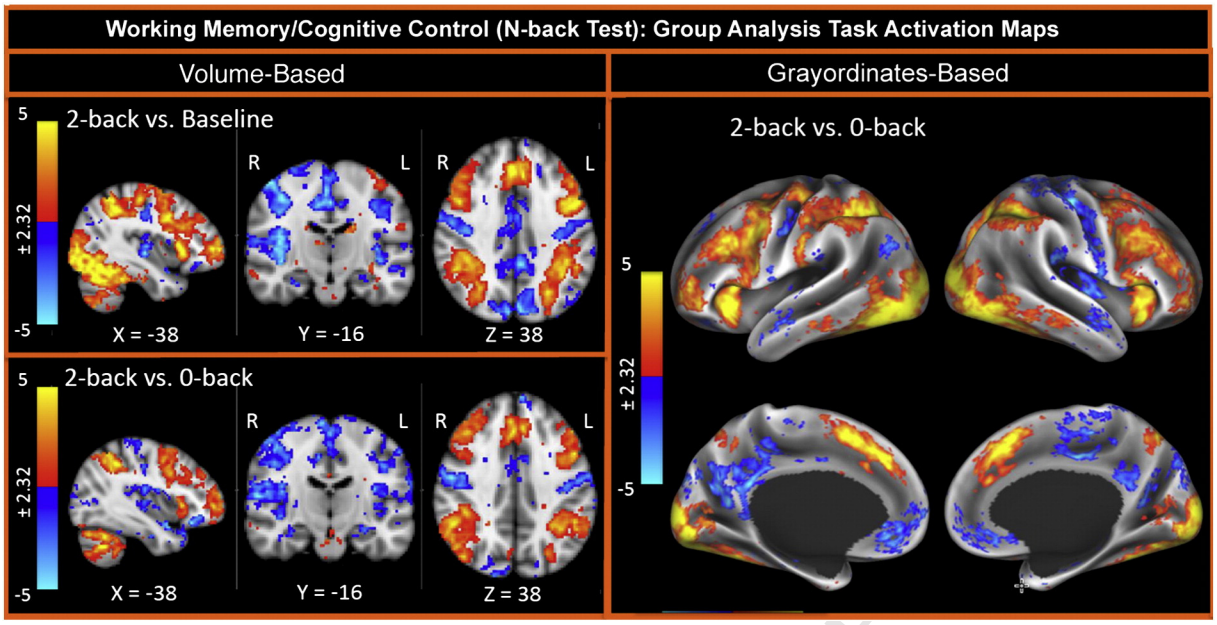


Fig. 8. Group-average task-fMRI from the working memory task. Adapted, with permission, from Barch et al. (2013).

Q3

might lead to inadvertent identification, particularly by other family members, if these combinations were publicly released. On the other hand, this information will be needed for many types of scientific inquiry aimed at characterizing the heritability of brain circuits and relating brain circuits to behavioral and demographic phenotypes. In order to minimize the risk of inappropriate disclosure of subject identity and yet maximize the usefulness of the data for research, all researchers who wish to make use of the HCP Restricted Access data elements (including *all* members of a given laboratory, not just the principal investigator) must agree in writing to a number of conditions, including the following:

- I agree to keep the data secure (password protected), to use the data responsibly, and to abide by the following terms
- I will not redistribute or share Restricted Data with others, including individuals in my laboratory, unless they have independently applied and been granted access to the Restricted Access data by the HCP.

I will abide by the following:

- No reporting of HCP Subject ID numbers when publishing or publicly reporting analyses that use Restricted data. I will not include any HCP-assigned subject IDs in any publication or public presentation that makes use of Restricted Data from individual subjects. I will instead assign my own study-specific subject IDs to each individual, e.g., subjects A, B, C, etc.
- Family structure is the ONLY Restricted Data element that can be reported for individual subjects in a publication or public presentation. When reporting family structure of subjects, individuals must be assigned study-specific subject IDs.
- If I publish data analyzed using Additional Restricted Data elements (including handedness, exact age, ethnicity, race, and body weight), each reported analysis must be based on at least 3 subjects, and the presentation of the data must not reveal the study-specific subject ID associated with any particular data point or value.

To mitigate any loss of transparency across studies, HCP will host a password-protected web page where investigators will be asked to load a key that maps their study-specific IDs to HCP ID subject IDs This resource will be accessible only to investigators granted access to Restricted Data and will facilitate comparison of results across different studies.

It is very important that everyone using Restricted Data understands and agrees to the full set of terms. Consistent compliance will be aided by general awareness among reviewers and editors as well as the scientific community in general. Examples of use case scenarios that may help investigators to understand how these terms apply to realistic scenarios are available at: http://www.humanconnectome.org/data/restricted-access/.

Genetic data based on genotyping (full-genome sequencing if feasible given cost–benefit tradeoffs) will be carried out in 2015. Data will be stored in dbGaP, and possibly also housed in ConnectomeDB. Great care will be taken to ensure that the genotyping data is handled with robust privacy protection while allowing data mining to benefit from information about population admixture derived from the genotyping data. This will include risk management for special cases (e.g., if the biological parents of an individual differs from that reported by participants), while ensuring that data analyses use genetically accurate relationships among siblings.

Some lessons learned

The HCP is one of many large-scale imaging projects currently underway around the world (see Craddock et al., in press), but it is distinctive if not unique in several important respects. One is the mandate to undertake major methodological improvements as a prelude to scanning a large number of subjects. Another is the unprecedented amount, quality, resolution and diversity of imaging modalities and other data types being systematically acquired. A third is the breadth of the data sharing and data mining efforts, commensurate with the richness and complexity of the data and the many levels of processing made available.

Given that the 5-year HCP grant is at its halfway point and is still in the early stages of systematic data collection and sharing, it would obviously be premature to declare the overall project a complete success. Nonetheless, the achievements to date are considerable, and the project remains on track relative to its original ambitious schedule. This reflects dedicated efforts and hard work by a large team that currently includes more than 100 investigators and technical staff from ten institutions in the consortium (Supplemental Table S5). Collectively, they provide

Please cite this article as: Van Essen, D.C., et al., The WU-Minn Human Connectome Project: An overview, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.05.041

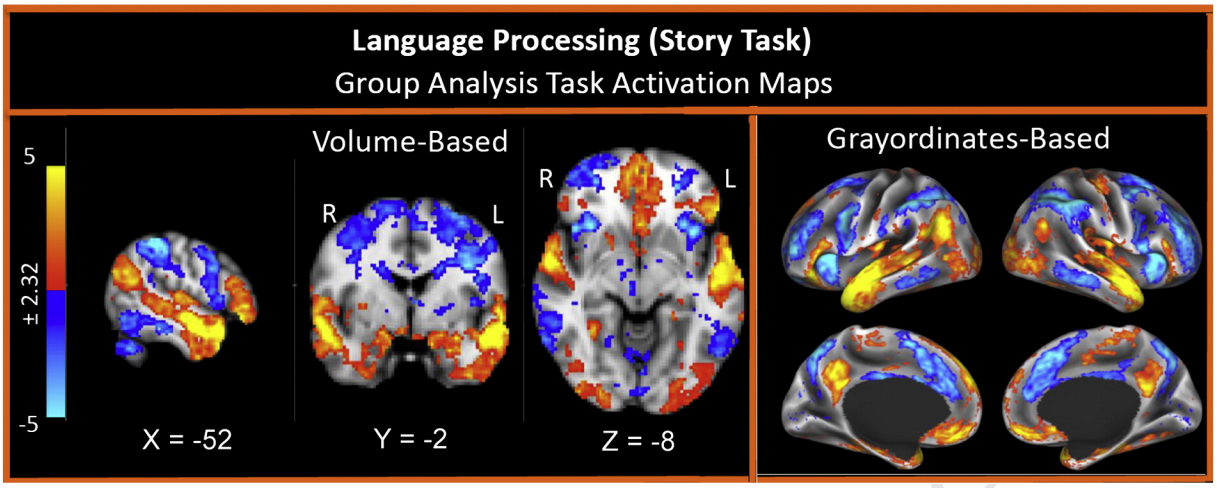


Fig. 9. Group-average task-fMRI from the language vs math task. Adapted, with permission, from Barch et al. (2013).

great breadth of expertise and intellectual perspectives needed to address the many facets of the project.

Given the size of the consortium and the multi-faceted nature of the endeavor, a number of operating principles and practices have proved (and will continue to be) especially useful. Here, we comment briefly on a few lessons learned and insights gained about the process of coordinating efforts by a consortium that is both geographically dispersed and highly diverse in its expertise, in hopes that some of these might be useful in other contexts.

Teams and working groups

At the beginning of the project we established seven operational teams (OTs) to organize the work of the consortium: Hardware, pulse sequences, and preprocessing; dMRI; rfMRI and tfMRI; MEG/EEG; recruitment, behavior, and genotyping; cross-modal integration and network modeling; and informatics (http://www.humanconnectome.org/about/teams.html). To promote high levels of coordination and collaboration across the different sites, and to give equal weight to potentially different scientific approaches to the work of the consortium, each OT is co-led by senior investigators from different institutions. Many consortium members participate in multiple teams, further aiding in cross-fertilization of ideas and in coordination with work across the teams. In many situations, ad hoc working groups have been established to address focused issues that typically included a subset of one or two operational teams. For example, the informatics efforts reached such a level of complexity after the first year that the team no longer met as a whole, and instead evolved into more than a half dozen working groups focused on specific and tractable components of the project (e.g., preprocessing pipelines; data storage needs; visualization software; and computational infrastructure). These working groups form as needed and disband when their work is completed. A Steering and Operations Committee that includes the

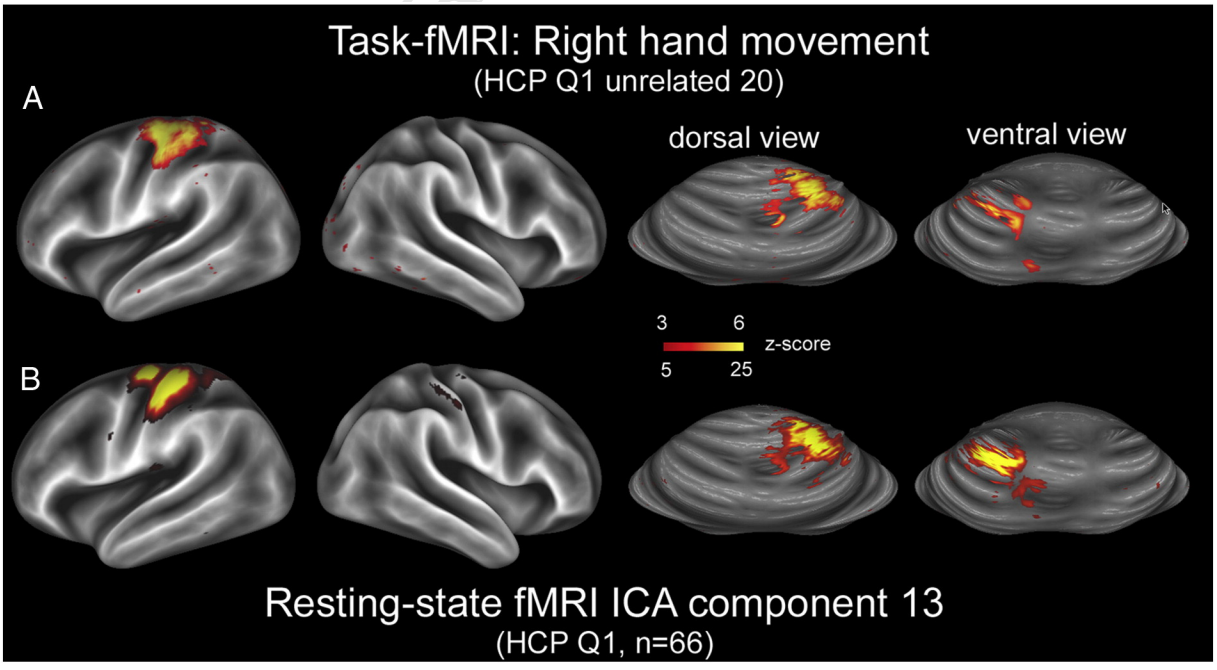


Fig. 10. A. Task-fMRI activation from the right-hand movement task carried out on the Q1 unrelated 20 subjects, mapped onto the group-average cerebral surfaces (first two panels) and onto the inflated cerebellar atlas surface that has been mapped to the MNI atlas stereotaxic space (Van Essen, 2009). B. Resting-state fMRI component 13 from a 100-dimensional ICA decomposition (with 82 components judged to be signal), applied to the 66 subjects in the HCP Q1 data release having four rfMRI runs.

Please cite this article as: Van Essen, D.C., et al., The WU-Minn Human Connectome Project: An overview, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.05.041

chairs and co-chairs of each Operations Team, as well as additional senior faculty advisors, provides overall coordination of the HCP effort as well as guidance on general questions, e.g., how the consortium will handle publications and share data.

In addition to frequent teleconferences and literally hundreds of thousands of emails among team members, the planning, data analysis and consensus-building necessary to develop a unified HCP approach to data collection has benefitted greatly from semiannual face-to-face meetings of all (in autumn) or many (in spring) HCP colleagues from around the world. These 'All-Hands' and 'Many-Hands' meetings have proven particularly valuable for addressing complex issues in an open forum that allows the domain experts time to drill down into the technical details while also allowing the broader consortium membership to gain valuable familiarity with key technical challenges and how they could be addressed. They also helped engender respect for the unique contributions that each team and each individual has brought to the table, including technical, conceptual and organizational skills and abilities. In addition, they provide us with an opportunity to interact with and receive feedback from our NIH Program and Science Officers and our External Advisory Panel members (Supplemental Table S5), who are also invited and who participate regularly.

These general organizational approaches were complemented by the promotion of a mindset of striving for improvements at every step and in every way possible. The established investigators joined the consortium with vast amounts of invaluable experience, but also with the baggage of sometimes relying on standard methods based on 'conventional wisdom' about how best to acquire, process, analyze, and interpret data. By encouraging all consortium members to challenge standard assumptions, then let the pilot data and results from proposed analyses drive the decision; many improvements have been realized. Some of the advances, such as the decision to use multiband imaging, have had a large impact on their own. Many other refinements represent incremental improvements individually, but the concatenation of many small increments has led to large gains in the aggregate. This applies to the extensive efforts to refine pulse sequences, image reconstruction algorithms, and also to the preprocessing and analysis pipelines. A number of these refinements have already been incorporated into other analysis platforms, including FSL, FreeSurfer, and Connectome Workbench, so that the benefits extend well outside the HCP proper.

HCP prospects

At the time this article was submitted, the WU-Minn HCP is at the midway point of the 5-year grant. It is also in a transitional period, with an increasing focus on standardized data acquisition and data sharing, but with important methods refinement efforts are still continuing. The Q1 data release constitutes only ~6% of the target number of 1200 subjects. Moreover, the more advanced stages of data analysis which are essential for characterizing structural and functional connectivity are still being refined and optimized. The companion articles in this special issue report many encouraging preliminary results as well as methodological advances, but not surprisingly they do not yet report major neuroscientific discoveries. We expect this to change dramatically over the next several years, as the HCP generates and shares an immense amount of neuroimaging, behavioral, and genotyping data, and also provides more extensively processed data — e.g., 'dense connectomes' and 'parcellated connectomes' from individual subjects as well as group averages. This should lead to a variety of important discoveries about brain connectivity, its relation to behavior and to other aspects of brain function, and its genetic underpinnings. We couple our optimism about the utility of the HCP datasets with the need to manage expectations and to acknowledge the technical limitations associated with each of the imaging modalities used by the HCP. For example, fMRI scans can be impacted by signals "bleeding across" opposing banks of sulci. Tractography has a bias for showing stronger connections with gyral blades compared to sulcal banks and fundi. Hence, for both modalities, the effective spatial resolution does not always achieve that implied by the size of the acquired voxels. Efforts to characterize brain circuits in individuals and in group averages must be mindful of these limitations as well as the strengths of the HCP datasets.

It is instructive to consider the aggregate amount of imaging information obtained via each modality in individual HCP subjects and what that may imply about the overall ability to characterize brain connectivity and its variability. The hour's worth of rfMRI scanning accumulated per subject yields ~5000 frames (TRs) of data for each of the ~90,000 grayordinates that represent the anatomical substrate on which a dense functional connectome is generated. If, hypothetically, each time point could encode just 2 bits of information that was statistically independent of other time points and other grayordinates, then the theoretical upper bound would be about 1 gigabit (10^9 bits) of information per subject. However, given the strong correlations in time (owing to the slow hemodynamic response function) and in space (neighboring grayordinates tend to be highly correlated), the actual amount of information is presumably much smaller, perhaps by around two orders of magnitude. If so, the amount of information about brain circuits provided by rfMRI would be in the range of 10^7 bits per HCP subject. An alternative assessment that yields a similar estimate comes from considering the covariance matrix of the fMRI timeseries, which presumably should be more reproducible across different scan sessions than the timeseries itself. At 2 mm resolution the covariance matrix contains ~8×10^9 ($90{,}000^2$) elements, or ~4×10^{10} information bits if there are 2 bits per element. If spatial correlations typically extend over ~50–100 grayordinates (e.g., patches ~15–20 mm in diameter), this would also suggest about 10^7 information bits per subject. For the 7 T HCP scans, the smaller voxel size attainable (~1 mm^3) will increase the number of spatial elements about 8-fold, but the anticipated temporal resolution will be lower by 2- or 3-fold, suggesting that the total amount of information may be about 2-fold greater. It will be interesting to refine such estimates in the future (and to make analogous estimates for other modalities such as dMRI), but even this rough ballpark assessment is of some interest. It suggests that MRI-based connectivity analyses have the potential to discriminate connectivity 'brainprints' among large numbers of individuals, albeit not unique for every individual on the planet.

A brief comparison with human genomics is also informative (cf. Van Essen and Ugurbil, 2012). The spectacular successes of the human genome project have enabled extraordinarily accurate sequencing (99.99% or better) of the ~3 billion bp of the human genome. However, the level of nucleotide diversity across individuals is remarkably low (only about 1 part in 1000; Jorde and Wooding, 2004; Tishkoff and Kidd, 2004). Hence, high sensitivity to sequence variants is critical for being able to characterize individual genomic differences and to relate these differences to phenotypes of interest. In contrast, the accuracy with which human brain connectivity can be quantitatively assessed is much lower than for genome sequencing, but the degree of individual variability is far greater. At a macroscopic level, we know that individual cortical areas vary in surface area by two-fold or more across individuals (cf. Van Essen et al., 2012b), and evidence from the macaque monkey suggests that the strength of pathways between any pair of cortical areas can vary by one or two orders of magnitude (Markov et al., 2011). But how pronounced are the individual differences in human brain connectivity that contribute to distinct behavioral phenotypes or that derive from distinct genotypes? These are empirical questions that will be addressed with increasing sensitivity as additional HCP datasets are acquired and analyzed over the next several years.

In this overall context, we are optimistic that major insights will emerge from mining of HCP data. In broad strokes, this will include (i) more accurate charting of brain parcellations, brain networks,

Please cite this article as: Van Essen, D.C., et al., The WU-Minn Human Connectome Project: An overview, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.05.041

and their dynamics; (ii) improved quantitative characterizations of network variability across individuals; and (iii) correlations between behavioral phenotypes and brain networks that provide a deeper understanding of the neural basis of individual variability. These insights will in turn provide an invaluable substrate for characterizing circuit abnormalities in a variety of brain disorders that afflict humankind.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2013.05.041.

Q42 Uncited references

Botteron, 2008
Folstein et al., 1975
Fox et al., 2005
Sotiropoulos et al., in press
Ugurbil et al., 2003a

Acknowledgments

We thank the current and past members of the WU-Minn HCP consortium (Supplemental Table S5) for their dedicated efforts on this project. We especially thank Matthew F. Glasser and Stam Sotiropoulos for their contributions to many of the analyses illustrated herein and Dr. Sandra Curtiss for overall project management as Q44 well as comments on the manuscript. The project was supported by an NIH grant 1U54MH091657, funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University; the Biotechnology Research Center (BTRC) P41 EB015894 from NIBIB, and the NINDS Institutional Center Core Grant P30 NS076408

Q45 References

Andersson, J.L.R., Xu, J., Yacoub, E., Auerbach, E., Moeller, S., Ugurbil, K., 2012. A comprehensive Gaussian process framework for correcting distortions and movements in diffusion images. INSERM Annual Meeting, Melbourne, Australia, May.

Andrews-Hanna, J.R., Reidler, J.S., Sepulcre, J., Poulin, R., Buckner, R.L., 2010. Functional–anatomic fractionation of the brain's default network. Neuron 65, 550–562.

Azevedo, F.A., Carvalho, L.R., Grinberg, L.T., Farfel, J.M., Ferretti, R.E., Leite, R.E., Jacob Filho, W., Lent, R., Herculano-Houzel, S., 2009. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. J. Comp. Neurol. 513, 532–541.

Beckmann, C.F., Smith, S.M., 2004. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans. Med. Imaging 23, 137–152.

Behrens, T.E., Berg, H.J., Jbabdi, S., Rushworth, M.F., Woolrich, M.W., 2007. Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? NeuroImage 34, 144–155.

Binder, J.R., Gross, W.L., Allendorfer, J.B., Bonilha, L., Chapin, J., Edwards, J.C., Grabowski, T.J., Langfitt, J.T., Loring, D.W., Lowe, M.J., Koenig, K., Morgan, P.S., Ojemann, J.G., Rorden, C., Szaflarski, J.P., Tivarus, M.E., Weaver, K.E., 2011. Mapping anterior temporal lobe language areas with fMRI: a multicenter normative study. NeuroImage 54, 1465–1475.

Blumensath T, Jbabdi S, Glasser MF, Van Essen DC, Ugurbil K, Behrens TE, Smith SM (2013) Spatially constrained hierarchical parcellation of the brain with resting-state fMRI. http://dx.doi.org/10.1016/jneuroimage201303024.

Botteron, K.N., 2008. Regional specificity of traumatic stress-related cortical reduction: further evidence from a twin study of post-traumatic stress disorder. Biol. Psychiatry 63, 539–541.

Botteron, K.N., Dierker, D., Todd, R., Alexopolous, J., Seung, D., Han, K., Nishino, T., Reid, E., Todorov, A., Van Essen, D.C., 2008. Human vs. computer algorithm choices in identifying identical twin pairs based on cortical shape characteristics — who's better? Org Human Brain Mapping Abstract #533.

Brookes, M.J., Hale, J.R., Zumer, J.M., Stevenson, C.M., Francis, S.T., Barnes, G.R., Owen, J.P., Morris, P.G., Nagarajan, S.S., 2011. Measuring functional connectivity using MEG: methodology and comparison with fcMRI. NeuroImage 56, 1082–1104.

Buckner, R.L., Krienen, F.M., Castellanos, A., Diaz, J.C., Yeo, B.T., 2011. The organization of the human cerebellum estimated by intrinsic functional connectivity. J. Neurophysiol. 106, 2322–2345.

Cohen, A.L., Fair, D.A., Dosenbach, N.U.F., Miezin, F.M., Dierker, D., Van Essen, D.C., Schlaggar, B.L., Petersen, S.E., 2008. Defining functional areas in individual human brains using resting functional connectivity MRI. NeuroImage 41, 45–57.

Q46 Craddock, R.C., Jbabdi, S., Yan, C.G., Vogelstein, J., Castellanos, F.X., Di Martino, A., Kelly, C., Heberlein, K., Colcombe, S., Milham, M.P., 2013. Imaging human connectomes at the macroscale. Nat. Methods (in press).

de Pasquale, F., Della Penna, S., Snyder, A.Z., Lewis, C., Mantini, D., Marzetti, L., Belardinelli, P., Ciancetta, L., Pizzella, V., Romani, G.L., Corbetta, M., 2010. Temporal dynamics of spontaneous MEG activity in brain networks. Proc. Natl. Acad. Sci. U.S.A. 107, 6040–6045.

de Pasquale, F., Della Penna, S., Snyder, A.Z., Marzetti, L., Pizzella, V., Romani, G.L., Corbetta, M., 2012. A cortical core for dynamic integration of functional networks in the resting human brain. Neuron 74, 753–764.

Escudero, J., Hornero, R., Abasolo, D., Fernandez, A., Lopez-Coronado, M., 2007. Artifact removal in magnetoencephalogram background activity with independent component analysis. Biomedical Engineering, IEEE Transactions on 54, 1965–1973.

Feinberg, D.A., Moeller, S., Smith, S.M., Auerbach, E., Ramanna, S., Glasser, M.F., Miller, K.L., Ugurbil, K., Yacoub, E., 2010. Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. PLoS One 5, e15710.

Fischl, B., Sereno, M., Tootell, R., Dale, A., 1999. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp. 8, 272–284.

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. J. Psychiatr. Res. 12, 189–198.

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

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

Glasser, M.F., Goyal, M.S., Press, T.M., Raichle, M.E., Van Essen, D.C., 2013. Trends and properties of human cerebral cortex: correlations with cortical myelin content. NeuroImage. http://dx.doi.org/10.1016/n.neuroimage.2013.03.060 (Special issue on In Vivo Brodmann Mapping, Epub ahead of print).

Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T., Fischl, B., Andersson, J., Xu, J., Jbabdi, S., Webster, M., Polimeni, J., Van Essen, D.C., Jenkinson, M., 2013. The minimal preprocessing pipelines for the Human Connectome Projects. NeuroImage (Special issue on Mapping the Connectome). Q47

Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, A., Salmelin, R., 2001. Dynamic imaging of coherent sources: studying neural interactions in the human brain. Proc. Natl. Acad. Sci. 98, 694–699.

Jbabdi, S., Johansen-Berg, H., 2011. Tractography: where do we go from here? Brain Connect. 1, 169–183.

Jorde, L.B., Wooding, S.P., 2004. Genetic variation, classification and 'race'. Nat. Genet. 36, S28–S33.

Larkman, D.J., Hajnal, J.V., Herlihy, A.H., Coutts, G.A., Young, I.R., Ehnholm, G., 2001. Use of multicoil arrays for separation of signal from multiple slices simultaneously excited. J. Magn. Reson. Imaging 13, 313–317.

Larson-Prior, L.J., Oostenveld, R., Della Penna, S., Michalareas, G., Prior, F., Babajani-Feremi, A., Marzetti, L., de Pasquale, F., Di Pompeo, F., Stout, J., Woolrich, M., Luo, Q., Bucholz, R., Fries, P., Pizzella, V., Romani, G.L., Corbetta, M., Snyder, A.Z., 2013. Adding dynamics to the Human Connectome Project with MEG and EEG. NeuroImage (Special issue on Mapping the Connectome). Q48

Lenglet, C., Abosch, A., Yacoub, E., De Martino, F., Sapiro, G., Harel, N., 2012. Comprehensive in vivo mapping of the human basal ganglia and thalamic connectome in individuals using 7 T MRI. PLoS One 7, e29153.

Mantini, D., Della Penna, S., Marzetti, L., de Pasquale, F., Pizzella, V., Corbetta, M., Romani, G.L., 2011. A signal-processing pipeline for magnetoencephalography resting-state networks. Brain Connectivity 1, 49–59.

Markov, N.T., Misery, P., Falchier, A., Lamy, C., Vezoli, J., Quilodran, R., Gariel, M.A., Giroud, P., Ercsey-Ravasz, M., Pilaz, L.J., Huissoud, C., Barone, P., Dehay, C., Toroczkai, Z., Van Essen, D.C., Kennedy, H., Knoblauch, K., 2011. Weight consistency specifies regularities of macaque cortical networks. Cereb. Cortex 21, 1254–1272.

Milchenko, M., Marcus, D., 2013. Obscuring surface anatomy in volumetric imaging data. Neuroinformatics 11, 65–75.

Moeller, S., Auerbach, E., Van de Moortele, P.-F., Adriany, G., Ugurbil, K., 2008. fMRI with 16 fold reduction using multibanded multislice sampling. Proc. Int. Soc. Magn. Reson. Med. 16.

Moeller, S., Yacoub, E., Olman, C.A., Auerbach, E., Strupp, J., Harel, N., Ugurbil, K., 2010. Multiband multislice GE-EPI at 7 T, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. Magn. Reson. Med. 63, 1144–1153.

Ogawa, S., Tank, D.W., Menon, R., Ellermann, J.M., Kim, S.G., Merkle, H., Ugurbil, K., 1992. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. Proc. Natl. Acad. Sci. U.S.A. 89, 5951–5955.

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. (156869), 152010 (EPub).

Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L., Petersen, S.E., 2011. Functional network organization of the human brain. Neuron 72, 665–678.

Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2012. Steps toward optimizing motion artifact removal in functional connectivity MRI; a reply to Carp. NeuroImage. Q49

Robinson, E., Jbabdi, S., Andersson, J., Smith, S., Glasser, M., Van Essen, D., Burgess, G., Harms, M., Barch, D., Jenkinson, M., 2013. Multimodal surface matching: fast and generalisable cortical registration using discrete optimisation. Proc Information Processing in Medical Imaging. Springer.

Schoffelen, J.M., Gross, J., 2009. Source connectivity analysis with MEG and EEG. Hum. Brain Mapp. 30, 1857–1865.

Setsompop, K., Gagoski, B.A., Polimeni, J.R., Witzel, T., Wedeen, V.J., Wald, L.L., 2012. Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. Magn. Reson. Med. 67, 1210–1224.

Shmuel, A., Yacoub, E., Chaimow, D., Logothetis, N.K., Ugurbil, K., 2007. Spatio-temporal point-spread function of fMRI signal in human gray matter at 7 T. NeuroImage 35, 539–552.

Smith, S., 2012. The future of fMRI connectivity. NeuroImage 62, 1257–1266.

Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C.F., Nichols, T.E., Ramsey, J.D., Woolrich, M.W., 2011. Network modelling methods for FMRI. NeuroImage 54, 875–891.

Q50 Smith, S.M., Andersson, J., Auerbach, E.J., Beckmann, C.F., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D.A., Griffanti, L., Harms, M.P., Kelly, M., Laumann, T., Miller, K.L., Moeller, S., Petersen, S., Power, J., Salimi-Khorshidi, G., Snyder, A.Z., Van Essen, D.C., Glasser, M.F., 2013. Resting-state fMRI in the Human Connectome Project. NeuroImage (Special issue on Mapping the Connectome).

Sotiropoulos, S.N., Behrens, T.E., Jbabdi, S., 2012. Ball and rackets: inferring fiber fanning from diffusion-weighted MRI. NeuroImage 60, 1412–1425.

Sotiropoulos, S.N., Chen, C., Dikranian, K., Jbabdi, S., Behrens, T.E., Van Essen, D.C., Glasser, M.F., 2013. Comparison of diffusion MRI predictions and histology in the macaque brain. ISMRM Abstract.

Q51 Sotiropoulos, S.N., Moeller, S., Jbabdi, S., Xu, J., Andersson, J.L., Auerbach, E., Yacoub, E., Feinberg, D., Setsompop, K., Wald, L.L., Behrens, T.E., Ugurbil, K., Lenglet, C., 2013. Effects of image reconstruction on fibre orientation mapping from multi-channel difusion MRI: reducing the noise floor using SENSE. Magn. Reson. Med. (in press).

Q52 Sotiropoulos, S.N., Jbabdi, S., Xu, J., Andersson, J.L., Moeller, S., Auerbach, E.J., Glasser, M.F., Hernandez, M., Sapiro, G., Jenkinson, M., Feinberg, D.A., Yacoub, E., Lenglet, C., Van Essen, D.C., Ugurbil, K., Behrens, T.E.J., 2013. Advances in diffusion MRI acquisition and processing in the Human Connectome Project. NeuroImage (Special issue on Mapping the Connectome).

Tishkoff, S.A., Kidd, K.K., 2004. Implications of biogeography of human populations for 'race' and medicine. Nat. Genet. 36, S21–S27.

Ugurbil, K., Toth, L., Kim, D.S., 2003. How accurate is magnetic resonance imaging of brain function? Trends Neurosci. 26, 108–114.

Ugurbil, K., Adriany, G., Andersen, P., Chen, W., Garwood, M., Gruetter, R., Henry, P.G., Kim, S.G., Lieu, H., Tkac, I., Vaughan, T., Van De Moortele, P.F., Yacoub, E., Zhu, X.H., 2003. Ultrahigh field magnetic resonance imaging and spectroscopy. Magn. Reson. Imaging 21, 1263–1281.

Ugurbil, K., et al., 2013. Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. NeuroImage (Special issue on Mapping the Connectome). Q53

Van Essen, D.C., 2009. Lost in localization — but found with foci?! NeuroImage 48, 14–17.

Van Essen, D.C., et al., 2012. The Human Connectome Project: a data acquisition perspective. NeuroImage 62, 2222–2231.

Van Essen, D.C., Glasser, M.F., Dierker, D., Harwell, J., Coalson, T., 2012. Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. Cereb. Cortex 22, 2241–2262. http://dx.doi.org/10.1093/cercor/bhr2291.

Van Veen, B.D., van Drongelen, W., Yuchtman, M., Suzuki, A., 1997. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Trans. Biomed. Eng. 44, 867–880.

Vaughan, J.T., Garwood, M., Collins, C.M., Liu, W., DelaBarre, L., Adriany, G., Andersen, P., Merkle, H., Goebel, R., Smith, M.B., Ugurbil, K., 2001. 7 T vs. 4 T: RF power, homogeneity, and signal-to-noise comparison in head images. Magn. Reson. Med. 46, 24–30.

Woolrich, M.W., Ripley, B.D., Brady, M., Smith, S.M., 2001. Temporal autocorrelation in univariate linear modeling of FMRI data. NeuroImage 14, 1370–1386.

Yacoub, E., Shmuel, A., Pfeuffer, J., Van De Moortele, P.F., Adriany, G., Andersen, P., Vaughan, J.T., Merkle, H., Ugurbil, K., Hu, X., 2001. Imaging brain function in humans at 7 T. Magn. Reson. Med. 45, 588–594.

Zaitsev, M., Dold, C., 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.

Please cite this article as: Van Essen, D.C., et al., The WU-Minn Human Connectome Project: An overview, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.05.041