The Human Connectome Project: Progress and Prospects

By David C. Van Essen, Ph.D., and Matthew F. Glasser, Ph.D.

Editor’s Note: As the first phase of one of the most ambitious projects in the history of neuroscience comes to a close, one early and influential leader and his younger colleague explain its evolution and underpinnings. Its goal “is to build a ‘network map’ that will shed light on the anatomical and functional connectivity within the healthy human brain, as well as to produce a body of data that will facilitate research into brain disorders such as dyslexia, autism, Alzheimer’s disease, and schizophrenia.”
Understanding the human brain in health and disease represents a grand scientific challenge for the 21st century and beyond. How does a collection of 90 billion neurons interconnected by 150 trillion synapses give rise to the extraordinary capabilities of human behavior and the amazing diversity of talents among the billions of people populating our earth? Recent years have seen exciting progress in addressing these fascinating questions, but achieving a deeper understanding of exactly how the human brain functions and what goes awry in various disorders remains a profoundly demanding endeavor. The more we learn, the more we appreciate how much is left to learn.

Progress in neuroscience has benefited greatly from increasingly powerful methods for acquiring, analyzing, and sharing data. For human brain studies, many noninvasive neuroimaging methods have emerged in recent decades. Among these, magnetic resonance imaging (MRI) has become a workhorse technology because of the diversity of information attainable using the same scanner to acquire images. Four main types of MRI are particularly germane here:

- **Structural MRI** provides simple but high resolution images of the brain, helpful in making geometrical models of brain structures necessary for modern brain imaging analysis and in analyzing subtle aspects of brain architecture, such as the thickness of the cerebral cortex or the amount of neuronal insulation, myelin, within the grey matter.

- **Task-activated functional MRI (tfMRI)** reveals which brain regions are relatively more activated or deactivated during performance of behavioral tasks, based on the blood oxygen level dependent (BOLD) MRI signal, which is modulated by neural activity via a process of neurovascular coupling.

- **Diffusion MRI (dMRI)** analyzes diffusion of water molecules within the white matter to infer the orientation of axonal fiber bundles; this information can then be used to infer long-distance white matter tracts (tractography) that connect distant gray matter locations, often called ‘structural connectivity.’

- **Resting-state fMRI (rfMRI)** correlates spontaneous fluctuations in the BOLD signal while subjects lie in the scanner and let their minds wander, to reveal how parts of the brain activate or deactivate together at rest — thereby enabling estimates of ‘functional connectivity.’
Brain imaging studies have typically focused on only one of the above approaches, and it has been challenging to integrate information across approaches and across different studies.

**Getting Off the Ground**

In 2009, leaders at the National Institutes of Health (NIH) recognized an opportunity to build on these advances by promoting the systematic characterization of human brain connectivity and its relationship to behavior. With Michael Huerta (associate director of the National Institute of Mental Health at the time) playing a key role, the NIH Neuroscience Blueprint announced a competition for the Human Connectome Project (HCP), with the overarching objectives of acquiring, analyzing, and freely sharing information about brain circuitry and connectivity gathered by noninvasively imaging healthy young adults. This funding opportunity spurred interest among neuroscientists around the world.

The scope of the project clearly warranted a multi-institutional consortium, leading to a complex courtship among potential dancing partners. A group of us at Washington University (WashU), having strong neurobiological and neuroinformatics expertise, joined with a University of Minnesota (UMinn) team of MR pulse-sequence experts led by Kamil Ugurbil; an Oxford University group of neuroimaging analysis experts led by Steve Smith; and investigators at several other institutions with expertise in magnetoencephalography (MEG). In 2010, our “WU-Minn-Ox” consortium received a $30 million HCP award from NIH. A separate HCP grant was awarded to investigators at MGH and UCLA to build a special scanner customized for diffusion imaging.

Six years later, the ‘young adult HCP’ is drawing to a close, as successor projects, targeting the full human lifespan as well as diseases of the brain, continue. We aim here to summarize some of the major accomplishments of the HCP, in methods development as well as emerging scientific discoveries, and suggest what the future may hold for human connectomics.

Overall, our goal in connectomics is to map the hundreds of functionally distinct areas, or ‘parcels’ of the human brain and to understand how these areas are connected and how each contributes to our behavior. Additionally, we want to understand how the brain’s complex functional systems go awry in neurological and psychiatric diseases. Successfully addressing these challenging questions
requires acquiring and analyzing MRI data of the highest quality in both normal young adults individuals (the original HCP), developing and aging individuals, and in individuals with these diseases.

What Has the HCP Achieved?
The HCP has made substantial improvements in data acquisition, analysis, and sharing. In aggregate, these advances constitute a new ‘HCP-style’ neuroimaging paradigm whose seven core tenets offer a modern alternative to the traditional approach that has dominated the field for two decades. Here, we use these tenets as a framework for highlighting various important advances.

Tenet 1. Using multiple imaging modalities, collect as much data as possible from each subject and from as many subjects as possible.
To this end, the HCP collected 4 hours of imaging data per subject: about 1 hour each for the four modalities of structural MRI, tfMRI, rfMRI, and dMRI in four scan sessions over two days. This is an unprecedented amount of scan time for a large-scale project, but having lots of high quality multimodal data from each subject yields major benefits, as noted below.

Our original target was to study 1,200 healthy young adults. The final major HCP release, slated for Fall 2016, will meet this target with imaging data from 1,100 subjects, plus out-of-scanner behavior-only data from another 100. Moreover, the HCP subjects include twins and their non-twin siblings, enabling analyses of the heritability of brain and behavioral phenotypes to add invaluable ‘icing’ to this large ‘cake’ of imaging data!

Tenet 2. Maximize the resolution of imaging data and overall data quality.
The first two years of the HCP grant were devoted, in part, to improving the scanner hardware and developing better MRI pulse sequences, i.e., the specific sequences of radio frequency pulses and magnetic field modulations or gradients needed to generate an MR image. All subjects were scanned using a customized Siemens 3 Tesla (‘3T’) scanner at WashU that provided stronger gradients, leading to better signal to noise in diffusion imaging. (As an aside, Siemens indicated at the time the scanner would be ‘one-of-a-kind,’ but it turned out to be a precursor to the Siemens Prisma, a commercial scanner that also provides high gradient strength.)
Even more important than the physical scanner was the incorporation of ‘multi-band’ (a.k.a. ‘multi-slice’) pulse sequences for fMRI and dMRI. In fMRI, the HCP used this to increase spatial resolution as measured by the size of the ‘voxels’ (the equivalent of 2D pixels, but in 3D) to 2 mm voxel size, less than the average thickness of the cortical sheet and down from the conventional 3 mm or higher, as well as the temporal resolution (less than 1 second, down from more than 2 seconds), which aids in selective noise reduction (see Tenet 3). For dMRI, it enabled a spatial resolution of 1.25 mm, unprecedented for in vivo human studies where 2 mm had previously been considered high resolution.

**Tenet 3. Minimize distortions, blurring, and noise in each subject’s data.**

The unprocessed images coming from MR scanners are substantially distorted, in modality-specific ways. For the HCP, we used state-of-the-art existing methods to remove these distortions, and implemented new or refined methods to achieve further improvements.

Time-dependent fluctuations in the BOLD signal are a mix of (1) the neurobiological signals of interest; (2) unstructured temporal noise arising from random (‘Gaussian’) variation; and (3) structured noise from subject head movement, physiology (e.g., respiration), and transient artifacts related to MR physics. The traditional paradigm typically uses extensive smoothing in space and time to reduce unstructured temporal noise. However, this comes at a high cost, because smoothing also blurs signals across boundaries—between the gray matter (what we care about in fMRI) and the adjoining white matter and cerebrospinal fluid (CSF), across cortical folds (where voxels may be spatially close but are widely separated from each other along the cortical sheet), and across brain-area boundaries. The HCP-style paradigm keeps smoothing to a minimum, and constrains it within the cortical ribbon (for cortex) or within compartmental boundaries (for subcortical nuclei). Also, unstructured noise reduction can be better achieved by averaging within brain parcels (see Tenet 5).

To reduce structured temporal noise, the fMRI fluctuations are decomposed into “components” using independent components analysis (ICA) and then a trained classifier algorithm (called ‘FIX’) automatically distinguishes between signal and noise components, allowing noise-related fMRI
fluctuations to be selectively removed. In contrast, the traditional paradigm uses less selective approaches like simply deleting noise-corrupted frames or removing the average MRI signal, which is controversial because this average contains both neural and noise effects.¹¹,¹² Further improvements are an area of active work.¹

**Tenet 4. Respect the natural geometry of brain structures.**

The traditional neuroimaging paradigm relies heavily on volume-based analyses based on a regular 3D array of voxels. This poses a serious problem for dealing with the highly convoluted, sheet-like cerebral cortex, which is better handled with geometric models that respect its topology. David Van Essen (DVE) first appreciated the compelling need for a ‘surface-based’ approach four decades ago when he began studying monkey visual cortex.¹³ This initially entailed making cortical ‘flat maps’ (collapsing the 3D cortex onto a 2D map) using a tedious manual pencil and tracing paper method.¹⁴ Two decades later, automated methodology finally emerged, based on computerized cortical surface reconstruction algorithms. For more accurate reconstructions, the HCP used high spatial resolution T1w and T2w scans at 0.7 mm (instead of the traditional T1w-only scans at 1 mm) processed by an adaptation of the FreeSurfer method.¹⁵ We also used the T1w/T2w ratio to generate cortical ‘myelin maps’ that provide valuable markers of cortical areas in many regions.⁷,¹⁶

Subcortical nuclei require a different approach because they are ‘blob-like’ rather than sheet-like, and occupy only a small fraction of total brain volume. To compactly represent subcortical gray matter as voxels and cerebral cortex as vertices (thus respecting the natural geometry of each domain), the HCP introduced ‘grayordinates’ as a concept and an associated ‘CIFTI’ file format. CIFTI files offer major advantages in data analysis and visualization.³ (Parenthetically, we note that cerebellar cortex is a folded sheet, but too thin to segment in individual subjects. Hence, it is, regrettably, treated as voxels rather than surface vertices in the current HCP grayordinate representation.)

**Tenet 5. Accurately align brain areas across subjects and across studies.**

Human brains are remarkably diverse, endowing us with our unique personalities and behaviors. Individual differences are especially pronounced for cerebral cortex as (1) the convolutions vary
dramatically across individuals (even in identical twins!), especially in higher cognitive regions, and (2) the boundaries of functionally defined areas vary in relation to cortical folds.

The traditional paradigm addresses individual variability using volume-based registration (alignment) of individuals to an atlas. However, even the widely used high-dimensional nonlinear registration algorithms are far from optimal because they fail to respect the topology of the cortical sheet. Van Essen has long contended that surface-based registration is fundamentally a more-sound approach.18

The widely used FreeSurfer surface registration method has provided a major advance over conventional volume-based methods in dealing with folding variability.19 However, FreeSurfer cannot achieve good functional alignment in regions where cortical areas are loosely correlated with specific folds. Two decades ago, Van Essen had mused about using functionally relevant features to improve alignment in such regions.20 This dream became a reality under HCP auspices, as Emma Robinson and Mark Jenkinson implemented the highly flexible Multimodal Surface Matching (MSM) algorithm, which we tuned to work well on HCP data, using myelin maps and resting-state networks to align the cerebral cortex across individuals.3,21

**Tenet 6. Use neuroanatomically accurate maps of brain areas.**

A fundamental neuroscientific challenge has been to subdivide the brain into neurobiologically accurate and functionally meaningful areas—i.e., to generate a ‘parts list.’ This ‘parcellation problem’ has been one of Van Essen’s abiding obsessions over his four decades as a cortical cartographer and quickly became Matthew Glasser’s as well after he began using brain imaging to study brain connectivity over a decade ago. For human cerebral cortex, the classical anatomists debated early in the 20th century whether there were about 50 or closer to 200 parcels.22-24 When Glasser joined Van Essen’s lab in 2008, they both were keen to use improved neuroimaging methods to attack the parcellation problem, and achieved their first success with the aforementioned myelin maps.16 The HCP offered a great opportunity to extend this to a truly multimodal approach using high quality data. Generating an improved cortical parcellation has been the overarching objective that drove our contributions to the methodological advances described in the preceding five tenets.
The figure below shows the fruits of these efforts in the form of the ‘HCP_MMP1.0’ multimodal cortical parcellation.\(^3\) It has 180 areas in each hemisphere, with a striking degree of left-right symmetry. This parcellation was initially generated by a semi-automated approach using data averaged across 210 HCP subjects, including cortical thickness, myelin maps, task-activation patterns, and functional connectivity, plus topographic organization revealed by rfMRI. Importantly, a fully automated areal classifier algorithm enabled highly accurate parcellation of the entire neocortex in individual subjects, even those with atypical topological arrangement of specific areas or who weren’t used to make the original map. This method reproduced the original parcellation in an independent group of another 210 subjects.

The HCP_MMP1.0 parcellation provides an invaluable neuroanatomical framework that can serve as a foundation for future brain imaging analyses. It is associated with a treasure trove of

---

180-area per hemisphere HCP_MMP1.0 multimodal human cortical parcellation. Reprinted, with permission, from Ref. 3. Data at http://balsa.wustl.edu/WN56.
multimodal information about each of the 360 areas in the group average and in individual subjects, which is ripe for exploration and data mining using the freely available HCP datasets (see Tenet 7).

**Tenet 7. Freely share data and analysis tools.**

The HCP was mandated by NIH to make its data freely available to the neuroscience community. This was music to Van Essen’s ears, as it resonated with his decades-long interest in data sharing, including early efforts with the SumsDB neuroimaging database, which exposed some of the challenges inherent in the enterprise. The HCP provided a great opportunity to take neuroimaging data sharing to a new level. We had the good fortune of teaming up with Dan Marcus (Van Essen’s former neurophysiology graduate student), who spearheaded the HCP neuroinformatics effort.

The main HCP database, ConnectomeDB, has served as a highly effective user-friendly platform for sharing the unprocessed and the much more complex (and useful) processed datasets. To date, more than 6,000 investigators have signed HCP Data Use Terms and downloaded more than 10 PB (10,000 TB) of HCP data.

There is also a growing need to share extensively analyzed neuroimaging data, particularly datasets associated with published figures—something ConnectomeDB was not designed to handle. Our lab recently launched the ‘BALSA’ database (http://balsa.wustl.edu), which builds on what worked well in SumsDB. BALSA is designed around ‘scenes’ that store all of the information needed to recapitulate exactly what is seen in a complex image viewed on a visualization platform. For example, the data used to generate Fig. 1 was uploaded into BALSA exactly as shown in the figure (including all of the annotations); the URL in the figure legend links directly to the corresponding scene in BALSA, and an archive containing 217 MB of associated data can be immediately downloaded. Currently, Connectome Workbench is the only neuroimaging platform that supports scene files and is compatible with BALSA, but this will hopefully change as other software developers recognize the utility of the scene-based approach to data sharing.

The number of publications acknowledging the use of HCP data currently exceeds 140 and is rapidly growing. A few brief examples illustrate the scope and diversity of emerging discoveries: (1) A comparison between resting-state functional connectivity and numerous behavioral and
demographic measures revealed a “positive–negative” phenotypic axis that covaries with the strength of functional connectivity in brain regions implicated in higher cognitive functions.\textsuperscript{30} (2) An analysis of resting-state functional networks revealed that regions of association cortex tend to overlap with at least two networks whereas somatosensory and visual regions are more isolated.\textsuperscript{31} (3) A model that links resting-state fMRI measurements to task activations in one set of subjects can be used to predict task activation maps for other subjects, suggesting that brain connectivity and function are coupled at the level of individual subjects.\textsuperscript{32} (4) Functional connectivity revealed in HCP data correlates with gene expression patterns in post-mortem human cortex.\textsuperscript{33}

\textbf{The Present and Future of Human Connectomics}

A number of large-scale projects are now building upon the success of the HCP by using its paradigm to study the brain in health and disease. One set, Lifespan Human Connectome projects, focuses on the healthy brain across the full human lifespan. They include the Lifespan Development project, which will study some 1,300 individuals in ages ranging from 5 to 21, and the Lifespan Aging project study, which will study 1,200 individuals older than 36.

Each project will include longitudinal scans on a subset of the population, complementing the cross-sectional approach and increasing the sensitivity for characterizing age-related changes. Earlier stages of development are being covered by the Lifespan Baby Connectome project, which will enroll 300 children (from birth to age 5) and by the Developing Human Connectome Project in the United Kingdom, which is studying prenatal and neonatal brain development.

Complementing these Lifespan Connectome initiatives is a diverse set of projects focusing on brain disorders, under the general umbrella of the Connectomes Related to Human Disease effort. More than 10 have been funded to date, and more are anticipated. They are using the HCP-style paradigm to address many diseases and disorders, including schizophrenia and Alzheimer’s disease.

Data sharing for the above projects will be under the umbrella of the Connectome Coordination Facility (CCF; http://humanconnectome.org/ccf). As a major extension of ConnectomeDB, the CCF will provide a common gateway for investigators wanting access to these datasets. Other large-
scale projects, such as the recently launched ABCD (Adolescent Brain Cognitive Development) project, will also emulate the HCP-style paradigm, but use a different data sharing venue.

Importantly, all of these projects entail significant adaptations in data acquisition to deal with practical limitations. Because the original HCP’s ‘luxury’ of obtaining four hours of imaging data over a two-day visit is in general not feasible for these other projects, they aim to acquire about 90 minutes of imaging data per subject.

Studying younger and older populations and populations with brain disorders also entails using shorter individual scan durations than the 15 minutes tolerated by the healthy young adults of HCP. We can be optimistic that these ‘HCP-fast’ acquisition protocols will be sufficient to increase the sensitivity for characterizing what may be subtle changes across the lifespan and differences between disease and control groups. However, it will remain critical to be vigilant in distinguishing neurobiologically interesting phenomena from artifacts related to such potential confounds as group differences in head motion or physiological noise inside the scanner. Efforts to better compensate for such confounds and further improve MRI data quality will be important to maintain and accelerate progress.34

Finally, and on a personal level, the opportunity to contribute to the HCP and to several successor projects represents a true highlight of Van Essen’s decades as a neuroscientist and a cortical cartographer; it provided Glasser with a unique opportunity to learn from an international consortium of experts while making significant contributions to the effort. In our opinion, the HCP represents team science at its best. Its success reflects an exceptionally talented and dedicated group of well over 100 investigators and technical staff who energetically debated ideas and options, then enthusiastically pulled together to carry out the hard work of gathering, analyzing, and sharing a rich dataset that should serve the neuroscience community well for many years to come.
Bios

David C. Van Essen, Ph.D., is the Alumni Endowed Professor in the Department of Neuroscience at Washington University in St. Louis. He has served as editor-in-chief of the *Journal of Neuroscience*, founding chair of the Organization for Human Brain Mapping, and president of the Society for Neuroscience. He is a fellow of the AAAS and received the Peter Raven Lifetime Achievement Award from the St. Louis Academy of Science and the Krieg Cortical Discoverer Award from the Cajal Club. Van Essen is internationally known for his research on the structure, function, connectivity, and development of cerebral cortex in humans and nonhuman primates. He led the Human Connectome Project (HCP) and co-leads two Lifespan HCP projects that will acquire and share data on brain circuitry in childhood and in aging.

Matthew F. Glasser, Ph.D., a medical student at Washington University in St. Louis, completed his Ph.D. training with David Van Essen. Glasser has over a decade of experience in brain imaging research with a focus on brain anatomy and brain imaging methods development, and has authored or coauthored 41 peer-reviewed articles. He is best known for his work on reconstructing the arcuate fasciculus, the main connection between the brain’s language areas; for developing novel or improved methods for mapping cortical areas, such as mapping the amount of neuronal insulation, called myelin, of the cortical grey matter based on clinical T1-weighted / T2-weighted MRI images; and for producing a new multi-modal map of the human cerebral cortex. Glasser is pursuing clinical training in neuroradiology to be a physician-scientist neuroradiologist.

Acknowledgments. We thank Sandra Curtiss for comments on the manuscript. Supported by the Human Connectome Project, WU-Minn Consortium (1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; NIH F30 MH097312, MH-60974, and the McDonnell Center for Systems Neuroscience at Washington University.

No financial conflicts to disclose.
References


34. M.D. Tisdall, M. Reuter, A. Qureshi, R.L. Buckner, B. Fischl, and A.J. van der Kouwe, "Prospective motion correction with volumetric navigators (vNavs) reduces the bias and..."