# Toward discovery science of human brain function

The MIT Faculty has made this article openly available. *Please share* how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1073/pnas.0911855107">http://dx.doi.org/10.1073/pnas.0911855107</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>National Academy of Sciences (U.S.)</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Mon Feb 11 05:15:49 EST 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/61013">http://hdl.handle.net/1721.1/61013</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Toward discovery science of human brain function


Although it is being successfully implemented for exploration of the genome, discovery science has eluded the functional neuroimaging community. The core challenge remains the development of common paradigms for interrogating the myriad functional systems in the brain without the constraints of a priori hypotheses. Resting-state functional MRI (fMRI) constitutes a candidate approach capable of addressing this challenge. Imaging the brain during rest reveals large-scale, low-frequency (<0.1 Hz) fluctuations in the fMRI signal that are temporally correlated across functionally related areas. Referred to as functional connectivity, these correlations yield detailed maps of complex neural systems, collectively constituting an individual’s “functional connectome.” Reproducibility across datasets and individuals suggests the functional connectome has a common architecture, yet each individual’s functional connectome exhibits unique features, with stable, meaningful interindividual differences in connectivity patterns and strengths. Comprehensive mapping of the functional connectome, and its subsequent exploitation to discern genetic influences and brain–behavior relationships, will require multicenter collaborative datasets. Here we initiate this endeavor by gathering R-fMRI data from 1,414 volunteers collected independently at 35 international centers. We demonstrate a universal architecture of positive and negative functional connections, as well as consistent loci of inter-individual variability. Age and sex emerged as significant determinants. These results demonstrate that independent fMRI datasets can be aggregated and shared. High-throughput R-fMRI can provide quantitative phenotypes for molecular genetic studies and biomarkers of developmental and pathological processes in the brain. To initiate discovery science of brain function, the 1000 Functional Connectomes Project dataset is freely accessible at www.nitrc.org/projects/fcon_1000/.

Much like the challenge of decoding the human genome, the complexities of mapping human brain function pose a challenge to the functional neuroimaging community. As dem-
onstrated by the 1000 Genomes Project (1), the accumulation and sharing of large-scale datasets for data mining is necessary for the first phase of discovery science.

Although the neuroimaging community has traditionally focused on hypothesis-driven task-based approaches, resting-state functional MRI (R-fMRI) has recently emerged as a powerful tool for discovery science. Imaging the brain during rest reveals large-amplitude spontaneous low-frequency (<0.1 Hz) fluctuations in the fMRI signal that are temporally correlated across functionally related areas (2–5). A single R-fMRI scan (as brief as 5 min) can be used to interrogate a multitude of functional circuits simultaneously, without the requirement of selecting a priori hypotheses (6). Building on the term “connectome,” coined to describe the comprehensive map of structural connections in the human brain (7), we use “functional connectome” to describe the collective set of functional connections in the human brain.

Buttressed by moderate to high test–retest reliability (8–10) and replicability (11, 12), as well as widespread access, R-fMRI has overcome initial skepticism (13) regarding the validity of examining such an apparently unconstrained state (5, 8, 14). Recent R-fMRI studies have identified putative biomarkers of neuropsychiatric illness (12, 15–18), provided insight into the development of functional networks in the maturing and aging brain (19–22), demonstrated a shared intrinsic functional architecture (23) between humans and nonhuman primates (24, 25), and delineated the effects of sleep (26), anesthesia (27), and pharmacologic agents on R-fMRI measures (28, 29). Given the many sources of variability inherent in fMRI, the remaining challenge is to demonstrate the feasibility and utility of adopting a high-throughput model for R-fMRI, commensurate with the scale used by human genetics studies to have the power to detect both single gene and combinatorial genetic and environmental effects on complex phenotypes.

Accordingly, the 1000 Functional Connectomes Project was formed to aggregate existing R-fMRI data from collaborating centers throughout the world and to provide an initial demonstration of the ability to pool functional data across centers. As of December 11, 2009, the repository includes data from 1,414 healthy adult participants contributed by 35 laboratories (Table S1). The intent is to expand this open resource as additional data are made available.

Here we provide an initial demonstration of the feasibility of pooling R-fMRI datasets across centers. Specifically, we (i) establish the presence of a universal functional architecture in the brain, consistently detectable across centers; (ii) investigate the influence of center on R-fMRI measures; (iii) explore the potential impact of demographic variables (e.g., age, sex) on R-fMRI measures; and (iv) demonstrate the use of an intersubject variance–based method for identifying putative boundaries between functional networks.

![Fig. 1.](http://brainvis.wustl.edu)

**Fig. 1.** Independent center-, age-, and sex-related variations detected in R-fMRI measures of functional connectivity and amplitude fluctuation. The first row depicts group-level maps for representative seed-based (column 1) and ICA-based (column 3) functional connectivity analyses (SI Results), as well as fALFF (column 2). Group-level maps were derived from one-way ANOVA across 1,093 participants from 24 centers (factor: center; covariates: age and sex). All group-level maps were corrected for multiple comparisons at the cluster level using Gaussian random-field theory ($Z > 2.3; P < 0.05$, corrected). For each measure, the second row shows robust between-center concordances (Kendall’s $W$), with the voxelwise coefficients of variation above the diagonal and the voxelwise means below the diagonal. Kendall’s $W$ concordance between any two centers was calculated across all voxels in the brain mask for the mean (or coefficient of variation) connectivity map across all participants included in each center. Rows 3, 4, and 5 depict voxels exhibiting significant effects of center, age, and sex, respectively, as detected by one-way ANOVA. “Male” refers to significantly greater connectivity (or amplitude, i.e., fALFF) in males; similarly, “female” refers to significantly greater connectivity (or amplitude) in females. “Older” refers to significantly increasing connectivity (or amplitude) with increasing age, whereas “younger” refers to significantly increasing connectivity (or amplitude) with decreasing age. “Pos” refers to positive functional connectivity, and “neg” refers to negative functional connectivity. The PCC seed region is indicated by a white dot. (Top Left) Surface map legend: LL, left lateral; RL, right lateral; LM, left medial; RM, right medial. All surface maps are rendered on the PALS-B12 atlas in CARET (http://brainvis.wustl.edu).
Results

We applied three distinct analytic methods commonly used in the R-fMRI literature: seed-based functional connectivity, independent component analysis (ICA), and frequency-domain analyses. Across the three approaches, we found evidence of (i) a universal intrinsic functional architecture in the human brain, (ii) center-related variation in R-fMRI measures, and (iii) consistent effects of age and sex on R-fMRI measures, detectable across centers despite the presence of center-related variability (Fig. 1). Specifically, seed-based correlational analyses revealed highly consistent patterns of functional connectivity across centers for both the “default mode” (30) and “task-positive” networks (31), supporting a universal functional architecture (Fig. S1). Similarly, a data-driven, temporal concatenation ICA approach, combined with dual regression (32–34), revealed consistent patterns of functional connectivity across centers for 20 spatially independent functional networks (Fig. 1 and Figs. S2 and S3). In addition, for each of the functional connectivity measures, within-center coefficient of variation maps showed a high degree of concordance across centers (Fig. S4). This suggests that common loci of variation exist: centers demonstrated a high degree of agreement on which connections are characterized by relative variance or invariance. Despite the high degree of concordance between centers, there were appreciable center-related variations in the strength of functional connectivity throughout the brain (8). The effect of center was especially prominent in regions exhibiting greater interregional connection strength, because these have the least within-center variability (See SI Results and Fig. S5 for further discussion of center-related variability.) However, even when taking this center-related variability into account, robustly reliable effects of age and sex remained appreciable (Fig. 2 and Figs. S1 and S2). (See SI Results and Fig. S6 for an examination of the impact of sample size on effects of age and sex.)

The detection of sex differences was particularly noteworthy, because these differences are rarely appreciated in the R-fMRI literature (35). Sexual dimorphism in human genomic expression (36) is known to affect numerous physiological variables that can influence the fMRI signal (37, 38). For example, males and females differ in terms of hemoglobin concentrations and hematocrit (39). However, global variables such as these do not explain the regionally specific sex-related phenomenon noted in the present work. Hormonal effects (e.g., estrogen), operating both during brain development (40) and acutely (41), are known to have regional specificity (42), making them potential contributors to the differences observed. Given the discovery nature of the present work and the lack of prior coordination among centers, the specific sex differences that we observed should be interpreted with caution until replicated in an independent sample.

Along with examining patterns of functional connectivity, we measured the amplitude of low-frequency fluctuations at each voxel using two common periodogram-based measures: amplitude of low frequency fluctuation (ALFF; total power <0.1 Hz) (2, 17, 43) and fractional ALFF (fALFF; total power <0.1 Hz/total power in the measured spectrum) (44). Concordant with previous work, the dominance of low-frequency fluctuations was consistently noted within gray matter regions, but not white matter (44). As with our analyses of functional connectivity, despite clear evidence of center-related effects, we were again able to demonstrate age- and sex-related differences in the magnitude of low-frequency fluctuations in various regions, particularly medial wall structures (Fig. 2 and Fig. S7).

Beyond data pooling for statistical analyses, we demonstrate the potential to use high-throughput datasets to develop normative maps of functional systems in the brain, which is a prerequisite for clinical applications. Specifically, we exploit a key property of functional connectivity maps, the presence of well-differentiated borders between functionally distinct regions (45). The voxelwise measures of coefficients of variation for each type of functional connectivity map delineate putative functional boundaries based on the presence of marked variability in func-

Fig. 2. Illustrative areas exhibiting age- and sex-related variation in R-fMRI properties. Significant group-level variance in functional connectivity maps was explained by age and sex (cluster-based Gaussian random-field corrected; Z > 2.3; P < 0.05). For each of three methods (seed-based, fALFF, and ICA), variance in connectivity strength explained by age (Left) and sex (Right) is illustrated both anatomically and graphically. Age-related differences are represented as scatterplots. Sex-related differences are represented as histograms depicting the distributions of resting-state functional connectivity (RSFC) values for males and females separately. Vertical lines indicate peak values. Corresponding topographical brain areas are indicated with dots. "Male" refers to significantly greater connectivity (or amplitude, i.e., fALFF) in males; similarly, "female" refers to significantly greater connectivity (or amplitude) in females. "Older" refers to significantly increasing connectivity (or amplitude) with increasing age, whereas "younger" refers to significantly increasing connectivity (or amplitude) with decreasing age.
Discussion

The present work represents a watershed event in functional imaging: demonstration of the feasibility of sharing and pooling functional data across multiple centers, alongside the establishment of an open-access data repository. We have demonstrated (i) the presence of a universal functional architecture, with remarkable stability in the functional connectome and its loci of variation across participants and centers; (ii) evidence of systematic sex differences in R-fMRI measures, as well as age-related gradients even in middle adulthood; and (iii) a method for highlighting the complex array of putative functional boundaries between networks from which normative maps can be developed. Future work should focus on using the functional connectome to catalog phenotypic diversity in brain–behavior relationships.

Functional connectivity is both related to and distinct from anatomic connectivity. Specifically, a recent study reported that a structural core appears to play “a central role in integrating information across functionally segregated brain regions” (23). As such, our finding of a universal functional architecture was not unexpected. But structure and function are not completely coupled, as illustrated by the robust homotopic (i.e., contralateral) functional connectivity for such regions as the primary visual cortex or the amygdala, both of which lack direct callosal projections (24, 46). Such findings imply that functional connectivity is subserved by polysynaptic as well as monosynaptic anatomic circuits. In addition, functional connectivity exhibits dynamic properties that are absent in structural connectivity. For instance, functional connectivity is modulated by cognitive (47) and emotional state (48), sleep (26), whereas structural connectivity is grossly unaffected by such factors. In short, the presence of a demonstrable structural connection does not necessitate that of a functional connection, nor does the demonstration of a functional connection imply the presence of a direct structural connection.

Task-based fMRI and R-fMRI approaches have complementary roles in the study of human brain function. Task-based approaches require sufficient a priori knowledge to articulate specific hypotheses, and they are invaluable in refining such hypotheses. But when the knowledge base is insufficient, task-based approaches may be compared to candidate gene studies, which have had limited success when applied to complex genetic disorders. In contrast, genome-wide association studies are increasingly providing initial findings for complex traits (49) and diseases that are subsequently validated through replication, extension, and deep sequencing (50). Our demonstration that R-fMRI data can be aggregated and pooled, and that variability among individuals can be explained in terms of specific subject variables (e.g., sex, age), suggests that this approach can provide quantitative phenotypes to be integrated into molecular studies.

Our results must be considered in light of several limitations of the present study. First, we used a convenience sample comprising previously collected data from an array of centers, without prior coordination of acquisition parameters or scanning conditions. Although the robustness of our results attests to the consistency of intrinsic brain activity, it still represents a potential underestimate of the true across-center consistency. Our demographic data warrant caution, because centers were heterogeneous with respect to male:female ratio, mean age, and age range. Our findings should motivate more systematic exploration of these variables, because future high-throughput imaging studies will need to take such factors into account.

Despite the promise of R-fMRI, some theoretical and pragmatic issues need to be addressed. Examples include the determination of...
the origins and biological significance of spontaneous low-frequency fluctuations of neuronal and hemodynamic activity, the impact of intrinsic activity on evoked responses (and vice versa), and the ideal means of acquiring, processing, and analyzing R-fMRI data. Nevertheless, the potential of discovery science is vast, from the development of objective measures of brain functional integrity to help guide clinical diagnoses and decision-making, to tracking treatment response and assessing the efficacy of treatment interventions. Finally, whereas the present work examines functional connectivity alone, future studies may combine R-fMRI with other modalities (e.g., EEG, magnetoencephalography, diffusion-tensor imaging, volumetrics) and genetics to achieve a more complete understanding of the human brain.

All data and analytic tools used in the present work will be made available at www.nitrc.org/projects/econ_1000/. We anticipate that the open availability of the 1000 Functional Connectomes dataset will recruit the broad participation and collaboration among the scientific community necessary for successful implementation of discovery-based science of human brain function. In addition, we hope that it will further advance the ethos of data sharing and collaboration initiated by such efforts as fMRIDC (www.fmriddc.org), FBIRN (www.brincommunity.org), OASIS (www.oasis-brains.org), BrainScape (www.brainscape.org), and BrainMap (www.brainmap.org).

Methods

Resting-state fMRI scans were aggregated from 35 community-based datasets (n = 1,414). The present analysis was restricted to 24 centers (n = 1,093; 21 published, 3 unpublished; mean age <60 years; only participants over age 18; one scan per participant; duration: 2.2–20 min; n = 970 at T, n = 123 at 1.5 T; voxel size, 1.5–8 mm within plane; slice thickness, 3–8 mm). Each contributor’s respective ethics committee approved submission of deidentified data. The institutional review boards of NYU Langone Medical Center and New Jersey Medical School approved the receipt and dissemination of the data.

For functional connectivity, we used seed-based correlation analysis, based on six previously identified seed regions (31), and model-free ICA, using temporal concatenation to generate group-level components and dual regression to generate individual participant maps. For amplitude measures at each voxel, we used a FEAT-based ALFF (7, 43) and its normalized variant, SALFF (44).

Standard image preprocessing was performed (i.e., motion correction, spatial filtering with FWHM = 6 mm, 12-dof affine transformation to MN152 stereotactic space). For seed-based correlation approaches and dual regression following ICA analysis, nuisance signals (e.g., global signal, WM, CSF, motion parameters) were regressed out. Temporal filtering was tailored for each analytic approach (29, 31, 32, 44).

ICA components for dual regression analyses were determined by (i) low-dimensional (20 components) temporal concatenation ICA carried out 25 times (each with 18 participants randomly selected from each of 17 centers with minimum of 165 time points) and (ii) low-dimensional (20 components) meta-ICA, a second concatenation-based ICA using the component sets produced by the 25 runs (see SI Results for a description of an alternative method). For each participant, dual regression (32–44) was performed using the 20 components identified by the meta-ICA (Fig. S3), yielding a connectivity map for each component.

Aggregate statistical analyses of center, sex, and age effects were based on a generalized linear model implementation of one-way ANOVA (factor: center; covariates: age and sex). To identify functional boundaries, we calculated voxelwise coefficients of variation across all 1,093 participants, and ranked each voxel based on the absolute value of its coefficient of variation.

ACKNOWLEDGMENTS. We thank David Kennedy and www.nitrc.org for supporting the 1000 Functional Connectomes Project data release, Aivi Snyder for providing helpful insights and advice concerning project goals, and Cameron Craddock for helpful advice on this study. Financial support for the 1000 Functional Connectomes project was provided by grants from the National Institute of Mental Health (R01MH083246 and R01MH081218 to F.X.C. and M.P.M.), National Institute on Drug Abuse (R03DA024775, to C.K.; R01DA016979, to F.X.C.), Autism Speaks, National Institute of Neurological Disorders and Stroke (RO1NS049176, to B.B.), and the Howard Hughes Medical Institute (to J.S.A. and R.L.B.), as well as gifts to the NYU Child Study Center from the Stavros Niarchos Foundation, Leon Levy Foundation, Joseph P. Healy, Linda and Richard Schaps, and Jill and Bob Smith and an endowment provided by Phylis Green and Randolph Cöwen. NITRC is funded by the National Institutes of Health’s Blueprint for Neurosciences Research (neuroscienceblueprint.nih.gov) (Contract N02-EB-6-4281, to TCG, Inc.).