GIGA*n*SCIENCE

REVIEW   **Open Access**

Connectomics and new approaches for analyzing human brain functional connectivity

R Cameron Craddock[1,2*], Rosalia L Tungaraza[1] and Michael P Milham[1,2]

Abstract

Estimating the functional interactions between brain regions and mapping those connections to corresponding inter-individual differences in cognitive, behavioral and psychiatric domains are central pursuits for understanding the human connectome. The number and complexity of functional interactions within the connectome and the large amounts of data required to study them position functional connectivity research as a "big data" problem. Maximizing the degree to which knowledge about human brain function can be extracted from the connectome will require developing a new generation of neuroimaging analysis algorithms and tools. This review describes several outstanding problems in brain functional connectomics with the goal of engaging researchers from a broad spectrum of data sciences to help solve these problems. Additionally it provides information about open science resources consisting of raw and preprocessed data to help interested researchers get started.

Keywords: Human connectome, Functional MRI, Brain graphs, Open data, Open science

Introduction

With its new emphasis on collecting larger datasets, data sharing, deep phenotyping, and multimodal integration, neuroimaging has become a data-intensive science. This is particularly true for connectomics[a] where thousands of brain imaging scans, each consisting of hundreds of observations of thousands of variables, are being collected and openly shared through a combination of grass-roots initiatives (e.g. the 1000 Functional Connectomes Project (FCP) [1], the International Neuroimaging Data-sharing Initiative (INDI) [2]) and large-scale international projects (the Human Connectome Project (HCP) [3,4], the Brainnetome [5], the Human Brain Project in EU known as CONNECT [6], the Pediatric Imaging, Neurocognition and Genetics (PING) Study [7], the Philadelphia Neurodevelopmental Cohort [8], the Brain Genomics Superstruct Project (GSP) [9], the National Database for Autism Research (NDAR) [10], and the Nathan Kline Institute Rockland Sample [11]). Although this deluge of complex

data promises to enable the investigation of neuroscientific questions that were previously inaccessible, it is quickly overwhelming the capacity of existing tools and algorithms to extract meaningful information from the data. This combined with a new focus on discovery science is creating a plethora of opportunities for data scientists from a wide range of disciplines such as computer science, engineering, mathematics, statistics, etc., to make substantial contributions to neuroscience. The goal of this review is to describe the state-of-the-art in connectomics research and enumerate opportunities for data scientists to contribute to the field.

The human connectome is a comprehensive map of the brain's circuitry, which consists of brain areas, their structural connections and their functional interactions. The connectome can be measured with a variety of different imaging techniques, but magnetic resonance imaging (MRI) is the most common in large part due to its near-ubiquity, non-invasiveness, and high spatial resolution [12]. As measured by MRI brain areas are patches of cortex (approximately $1cm^2$ area) [13] containing millions of neurons (calculated from [14]); structural connections are long range fiber tracts that are inferred from the motion of water particles measured by diffusion weighted MRI (dMRI); and functional interactions are inferred from synchronized brain activity measured by functional MRI

*Correspondence: ccraddock@nki.rfmh.org
[1]Center for Biomedical Imaging and Neuromodulation, Nathan Kline Institute for Psychiatric Research, 140 Old Orangeburg Rd, 10962 Orangeburg, New York, USA
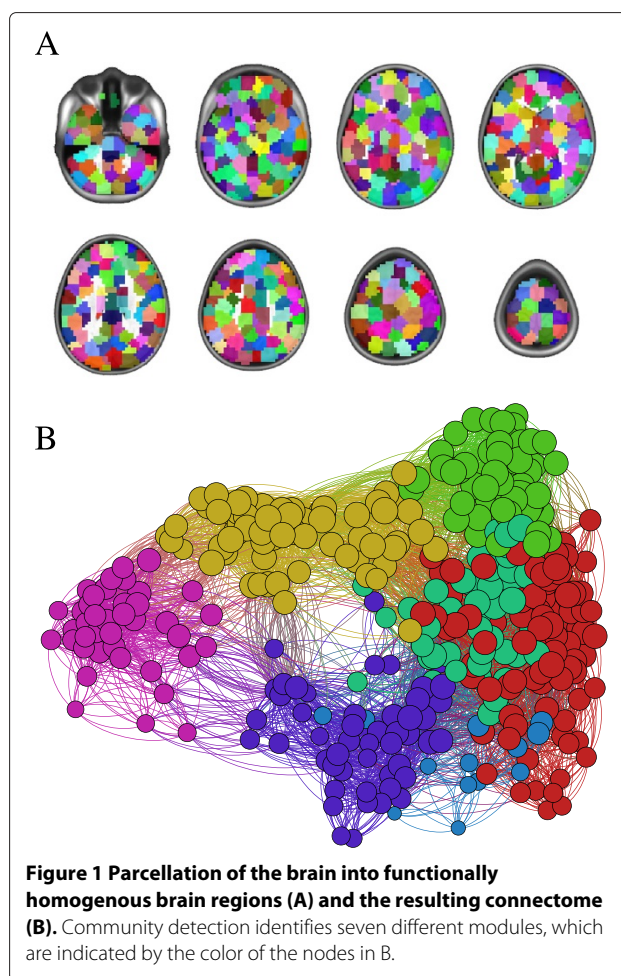[2]Center for the Developing Brain, Child Mind Institute, 445 Park Ave, 10022 New York, New York, USA

(fMRI) [15]. Addressing the current state-of-the-art for both functional and structural connectivity is well beyond the scope of a single review. Instead, this review will focus on functional connectivity, which is particularly fast-growing and offers many exciting opportunities for data scientists.

The advent of functional connectivity analyses has popularized the application of discovery science to brain function, which marks a shift in emphasis from hypothesis testing, to supervised and unsupervised methods for learning statistical relationships from the data [1]. Since functional connectivity is inferred from statistical dependencies between physiological measures of brain activity (i.e. correlations between the dependent variables), it can be estimated without an experimental manipulation. Thus, functional connectivity is most commonly estimated from "resting state" fMRI scans, during which the study participant lies quietly and does not perform any experimenter specified tasks; when estimated in this way, it is referred to as intrinsic functional connectivity (iFC) [16]. Once iFC is estimated, data mining techniques can be applied to identify iFC patterns that covary with phenotypes, such as indices of cognitive abilities, personality traits, or disease state, severity, and prognosis, to name a few [17]. In a time dominated by skepticism about the ecological validity of psychiatric diagnoses [18], iFC analyses have become particularly important for identifying subgroups within patient populations by similarity in brain architecture, rather than similarity in symptom profiles. This new emphasis in discovery necessitates a new breed of data analysis tools that are equipped to deal with the issues inherent to functional neuroimaging data.

Review
The connectome analysis paradigm
In 2005 Sporns [19] and Hagmann [20] independently and in parallel coined the term *the human connectome*, which embodies the notion that the set of all connections within the human brain can be represented and understood as graphs. In the context of iFC, graphs provide a mathematical representation of the functional interactions between brain areas: nodes in the graph represent brain areas and edges indicate their functional connectivity (as illustrated in Figure 1). While general graphs can have multiple edges between two nodes, brain graphs tend to be simple graphs with a single undirected edge between pairs of nodes (i.e. the direction of influence between nodes is unknown). Additionally edges in graphs of brain function tend to be weighted - annotated with a value indicating the similarity between nodes. Analyzing functional connectivity involves 1) preprocessing the data to remove confounding variation and to make it comparable across datasets, 2) specification of brain areas to be used as nodes, 3) identification of edges from the iFC between nodes, and

Figure 1 Parcellation of the brain into functionally homogenous brain regions (A) and the resulting connectome (B). Community detection identifies seven different modules, which are indicated by the color of the nodes in B.

4) analysis of the graph (i.e. the structure and edges) to identify relationships with inter- or intra- individual variability. All of these steps have been well covered in the literature by other reviews [12,17,21] and repeating that information provides little value. Instead we will focus on exciting areas in the functional connectomics literature that we believe provide the greatest opportunities for data scientists in this quickly advancing field.

Modeling functional interactions within the connectome
Defining the nodes to use for a connectivity graph is a well described problem that has become an increasingly active area of research [22]. From a neuroscientific perspective there is meaningful spatial variation in brain function that exists at resolutions much finer than what can be measured using modern non-invasive neuroimaging techniques. However, connectivity graphs generated at the spatial resolution of these techniques are too large to be wieldy and there is insufficient fine-grained information about brain function to interpret connectivity results at that level. For that reason, the number of nodes in the connectome is commonly reduced by way of combining

voxels into larger brain areas for analysis. This is accomplished using either boundaries derived from anatomical landmarks [23,24], regions containing homogeneous cytoarchitecture as determined by post-mortem studies [25], or from clusters determined by applying unsupervised learning methods to functional data [26,27]. The latter approach tends to be preferred since it is not clear that brain function respects anatomical subdivisions, and similar cells may support very different brain functions [27]. Quite a few clustering approaches have been applied to the problem of parcellating brain data into functionally homogenous brain areas, each varying in terms of the constraints they impose on the clustering solution [22,26-31]. The literature provides some evidence that hierarchical clustering based methods perform best [22,28], but no single clustering level has emerged as optimal. Instead, it appears as though there is a range of suitable clustering solutions from which to choose [22,27].

Once the nodes of a connectivity graph have been chosen, the functional connectivity between them is estimated from statistical dependencies between their time courses of brain activity. Although a variety of bivariate and multivariate methods have been proposed for this purpose [17,32], there is a lot of room for new techniques that provide better estimates of the dependencies, or provide more information about the nature of these dependencies. iFC is most commonly inferred using bivariate tests for statistical dependence, typically Pearson's correlation coefficient [16]. Since these methods only consider two brain areas at the time, they cannot differentiate between direct and indirect relationships. For example the connection $A \leftrightarrow C$ in the triangle $A \leftrightarrow B$, $B \leftrightarrow C$, $A \leftrightarrow C$ may be due to the variance that A and C both share with B (an indirect connection), rather than variance that is shared uniquely by the two independent of B (a direct connection). Indirect relationships can be excluded from the graph using partial correlation, or inverse covariance matrix estimation, but regularization estimators must be employed for large number of brain areas [17,33].

Tests of statistical dependencies between brain regions only provide information about whether or not two nodes are connected, but it should be possible to construct a more precise mathematical description of the relationship between brain areas [34]. Several different modeling techniques have been proposed to this end. Model confirmatory approaches such as structural equation modeling (SEM) [35] and dynamic causal modeling (DCM) [36] can offer fairly detailed descriptions of node relationships, but, they rely on the pre-specification of a model and are limited in the size of network that can be modeled. Cross-validation methods have been proposed to systematically search for the best model [37-39], but simulations have shown that those methods do not necessarily converge to the correct model [40]. Granger causality is another exploratory, data-driven modeling technique that has been particularly popular due to its promise of identifying causal relationships between nodes based on temporal lags between them [41]. However, the assumptions underlying Granger causality do not quite fit with fMRI data [32], where delays in the time-courses between regions may be more reflective of some physiological phenomena, such as a perfusion deficit [42], rather than causal relationships between brain areas. Alternatively, brain connectivity can be inferred from a multivariate regression that is solved using either dimensionality reduction [34] or regularization [43]. These more precise mathematical models of connectivity have shown great promise for testing hypotheses of brain organization [43], predicting response to rehabilitation after stroke data [44], and as biomarkers of disease [45].

Functional interactions within the connectome are commonly considered to be static over the course of an imaging experiment, but a growing body of research has demonstrated that connectivity between brain regions changes dynamically over time [46]. While most studies have measured connectivity within a short window of the fMRI time-course that is moved forward along time [47-50] other methods have been employed with similar results [51,52]. Several problems must be overcome in order to reliably measure changing functional connectivity patterns from the inherently slow and poorly sampled fMRI signal. First, the variance of correlation estimates increases with decreasing window size, meaning that unless proper statistical controls are utilized, the observed dynamics may arise solely from the increased variance [53]. This issue may be mitigated using the new higher speed imaging methods, which have already shown promise for extracting dynamic network modes using temporal independent component analysis (tICA), although large numbers of observations are still necessary [52]. Node definition is another issue, as it is unclear whether brain areas defined from static iFC are appropriate for dynamic iFC; however, initial work has shown that parcellations of at least some brain regions from dynamic iFC are consistent with what is found with static [49].

Mapping intra- and inter-individual variation

The ultimate goal of connectomics is to map the brain's functional architecture and to annotate it with the cognitive or behavioral functions they subtend. This latter pursuit is achieved by a group level analysis in which variations in the connectome are mapped to inter-individual differences in phenotype [21], clinical diagnosis [54], or intra-individual responses to experimental perturbations (such as the performance of different tasks) [55-57]. Several different analyses have been proposed for accomplishing these goals, and they all require some mechanism for comparing brain graphs [17].

Approaches to comparing brain graphs can be differentiated based on how they treat the statistical relationships between edges. One such approach, referred to as "bag of edges", is to treat each edge in the brain graph as a sample from some random variable. Thus, a set of N brain graphs each with M edges will have N observations for each of the M random variables. In this case, the adjacency (or similarity) matrix describing the brain graphs can be flattened into a vector representation and any of the well explored similarity or dissimilarity metrics can be applied to the data [12]. One of the benefits of this representation is the ability to treat each edge as independent of all other edges and to compare graphs using mass univariate analysis, in which a separate univariate statistical test (e.g. t-test, anova, or ancova) is performed at each edge. This will result in a very large number of comparisons and an appropriate correction for multiple comparisons, such as Network-Based Statistic [58], Spatial Pairwise Clustering [58], Statistical Parametric Networks [59], or group-wise false discovery rate [60], must be employed to control the number of false positives. Alternatively, the interdependencies between edges can be modeled at the node level using multivariate distance matrix regression (MDMR) [61], or across all edges using machine learning methods [62-64].

Despite the successful application of this technique, a drawback of representing a brain graph as a bag of edges is that it throws away all information about the structure of the graph. Alternative methods such as Frequent Subgraph Mining (FSM) rely on graph structure to discover features that better discriminate between different groups of graphs [65]. For instance, Bogdanov *et al.* [66] were able to identify functional connectivity subgraphs with a high predictive power for high versus low learners of motor tasks. A recent comprehensive review [67] outlines other approaches that take the graph structure into account e.g. the graph edit distance and a number of different graph kernels. All of these methods are under active development and have not yet been widely adapted by the connectomics community.

Another approach for estimating graph similarity using all the vertices involves computing a set of *graph-invariants* such as node centrality, modality, and global efficiency, among others, and using the values of these measures to represent the graph [68,69]. Depending on the invariant used, this approach may permit the direct comparison of graphs that are not aligned. Another advantage is that invariants substantially reduce the dimensionality of the graph comparison problem. On the other hand, representing the graph using its computed invariants throws away information about that graph's vertex labels [70]. Moreover, after computing these invariants it is often unclear how they can be interpreted biologically. It is important that the invariant used matches the

relationships represented by the graph. Since edges in functional brain graphs represent statistical dependencies between nodes and not anatomical connections, many of the path-based invariants do not make sense, as indirect relationships are not interpretable [68]. For example, the relationships $A \leftrightarrow B$ and $B \leftrightarrow C$ do not imply that there is a path between nodes A and C; if a statistical relationship between A and C were to exist they would be connected directly.

Predictive modeling Resting state fMRI and iFC analyses are commonly applied to the study of clinical disorders and, to this end, the ultimate goal is the identification of biomarkers of disease state, severity, and prognosis [54]. Prediction modeling has become a popular analysis method because it most directly addresses the question of biomarker efficacy [62,63,67]. Additionally, the prediction framework provides a principled means for validating multivariate models that more accurately deal with the statistical dependencies between edges compared to mass univariate techniques, all while reducing the need to correct for multiple comparisons.

The general predictive framework involves learning a relationship between a *training* set of brain graphs and a corresponding categorical or continuous variable. Brain graphs can be represented by any of the previously discussed features. The learned model is then applied to an independent *testing* set of brain graphs to decode or *predict* their corresponding value of the variable. These values are compared to their "true" values to estimate *prediction accuracy* - a measure of how well the model generalizes to new data. Several different strategies can be employed to split the data into training and testing datasets, although leave-one-out cross-validation has high variance and should be avoided [71].

A variety of different machine learning algorithms has been applied to the analysis of brain graphs in this manner, but by far the most commonly employed has been support vector machines [54,72]. Although these methods offer excellent prediction accuracy, they are often black boxes, for which the information used to make the predictions is not easily discernible. The extraction of neuroscientifically meaningful information from the learned model can be achieved by employing sparse methods [73] and feature selection methods [62] to reduce the input variables to only those essential for prediction [17]. There is still considerable work to be performed in 1) improving the extraction of information from these models, 2) developing techniques permitting multiple labels to be considered jointly, and 3) developing kernels for measuring distances between graphs.

There are a few common analytical and experimental details that limit the utility of putative biomarkers

learned through predictive modeling analyses. Generalization ability is most commonly used to measure the quality of predictive models. However, since this measure does not consider the prevalence of the disorder in the population, it does not provide an accurate picture of how well a clinical diagnostic test based on the model would perform. This can be obtained from estimates of positive and negative predictive values [74,75] using disease prevalence information from resources such as Centers for Disease Control and Prevention Mortality and Morbidity Weekly Reports [76]. Castellanos *et al.* provide a reevaluation of generalizability metrics reported in the connectomics prediction literature up to 2013. Also, the majority of neuroimaging studies are designed to differentiate between an ultra-healthy cohort and a single severely-ill population, which further waters down estimates of specificity. Instead, it is also important to validate a biomarker's ability to differentiate between several different disease populations - an understudied area of connectomes research [18].

Most predictive modeling-based explorations of connectomes have utilized classification methods that are sensitive to noisy labels. This is particularly problematic given the growing uncertainty about the biological validity of classical categorizations of mental health disorders

[18]. This necessitates the use of methods that are robust to noisy labels [77,78]. Many such techniques require quantification of the uncertainty of each training example's label, which can be very difficult to estimate for clinical classifications. Another approach that is being embraced by the psychiatric community is to abandon classification approaches altogether, and to instead focus on dimensional measures of symptoms [79]. In the context of predictive modeling this translates into a change in focus toward regression models, which to date have been underutilized for the analysis of connectomes [54].

The aforementioned dissatisfaction with extant clinical categories opens up opportunities to redefine clinical populations based on their biology rather than symptomatology. This can be accomplished using unsupervised learning techniques to identify subpopulations of individuals based on indices of brain function and then identifying their associated phenotypes, as illustrated in Figure 2 [80]. Similar to predictive modeling, a major challenge of this approach is to find the features that are most important for defining groups. Another problem is regularizing the clustering solution to make sure it is relevant to the phenotypes under evaluation. These issues can be resolved using semi-supervised techniques or "multiway" methods that incorporate phenotypic information to
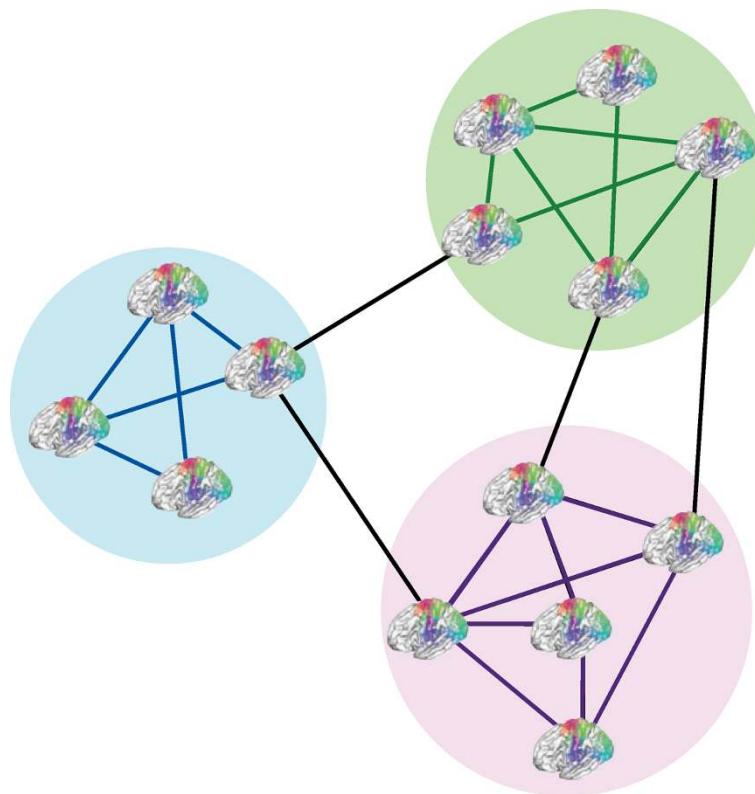
Figure 2 Identifying communities based on neurophenotypes. Brain glyphs provide succinct representations of whole brain functional connectivity [85].

guide clustering [81]. Along these lines, joint- or linked-ICA methods have been used to fuse different imaging modalities [82,83] as well as genetics and EEG data with imaging data [84].

Evaluating functional connectivity pipelines

Analyzing functional connectivity data requires the investigator to make a series of decisions that will impact the analysis results; examples include choosing the pre-processing strategy for removing noise, the parcellation method and scale for defining graph nodes, the measure for defining connectivity, and the features and methods for comparing connectivity across participants. Several different possibilities have been proposed for each of these steps and choosing the best analysis strategy is a critical problem for connectome researchers. The complexity of this problem is highlighted by observations that both uncorrected noise sources [86-90] and denoising strategies [91,92] can introduce artifactual findings. Ideally the choices for each of these parameters would be determined by maximizing the ability of the analysis to replicate some ground truth, but - as with most biomedical research - the ground truth is unknown. Simulations provide useful means for comparing the performance of different algorithms and parameter settings, but are limited by the same lack of knowledge that necessitates their use. Instead researchers are forced to rely on criteria such as prediction accuracy, reliability, reproducibility, and others for model selection [93]. Although most published evaluations of different connectivity analysis strategies focus on single optimization criterion in isolation, doing so may result in a sub-optimal choice. For example, head motion has high test-retest reliability, as do the artifacts that are induced by head motion [89]. As such, focusing solely on test-retest reliability may lead to the conclusion that motion correction should not be employed. Likewise, when learning a classifier for a hyperkinetic population, head motion-induced artifacts will improve prediction accuracy [94]. Instead, several - ideally orthogonal - metrics should be combined for model selection. For example, in the case of motion correction, metrics for model selection should include an estimate of residual head motion effects in the data [87-90]. Failure to include measures of prediction accuracy and reproducibility in the optimization might result in a strategy that is too aggressive and removes biological signal [95,96]. Going forward, the development of new frameworks and metrics for determining the best algorithms for connectivity analysis will continue to be a crucial area of research.

Computational considerations

Many of the advances in connectomics research have been spurred on by Moore's Law and the resulting rapid increase in the power and availability of computational resources. However, the amount of resources, time and memory required to process and analyze large connectomics datasets remains a significant barrier for many would-be connectomes researchers, hence providing another crucial area where computational researchers can contribute to connectomics research. The most common approach for automating high-throughput connectomics processing is to link existing neuroimaging tools together into software *pipelines*. In most cases, since processing each dataset can be performed independently, these pipelines can be executed in parallel on large-scale, high-performance computing (HPC) architectures, such as multi-core workstations or multi-workstation clusters [97-102]. The construction of these pipelines are made possible by the modularity of most neuroimaging packages (e.g., AFNI [103], ANTs [104], FSL [105], and SPM [106]), in which each processing step is implemented by separate functionality, and by their reliance on the NIfTI standard [107], which allows tools from different packages to be inter-mixed. Some steps of the pipeline are independent as well, and many of the toolsets are multi-threaded, providing further opportunities to speedup processing by taking advantage of multi-core systems. Using this strategy, the execution time for a large-scale analysis can theoretically be sped up by the number of pipelines that are run in parallel, but in practice this is not quite obtainable due to overheads incurred by the increased competition for resources (Amdahl's Law [108]). A major advantage of this strategy is that no modifications to the existing neuroimaging tools are required, plus it can be easily scaled to very large datasets, and it can take advantage of everything from relatively small multi-core systems to very large computing clusters. A disadvantage is that it requires access to large computational resources that are not always available, particularly at smaller research institutions, or in developing countries.

Since the preprocessing and analysis of large connectomics datasets are bursty in nature, they do not justify the large capital costs and maintenance burden of dedicated HPC infrastructures [109]. Instead, when shared or institutional computing resources are unavailable, cloud computing offers a "pay as you go" model that might be an economical alternative. Catalyzed by virtualization technology, systems such as the Amazon Elastic Compute Cloud and Google Compute Engine allow users to dynamically provision custom-configured HPC systems to perform an analysis. Pre-configured virtual machines such as the Configurable Pipeline for the Analysis of Connectomes Amazon Machine interface (C-PAC AMI) [110] and the NITRC Computational Environment (NITRIC-CE) [111] eliminate many of the challenges associated with installing and maintaining open source tools. Pre-processing a single dataset (structural MRI and functional MRI for a single participant) using the C-PAC AMI

costs around $2.50 on Amazon EC2 using computation-optimized compute nodes with 32 processors and 60 gigabytes of RAM, this could cost as little as $0.75 per dataset if more economical "spot" instances are utilized. The largest drawbacks to computing in the cloud are the time required for data transfers and the expense.

The previously described strategies for accelerating functional connectivity analyses rely on the data parallelism that exists between datasets, but there is quite a bit of parallelism that exists at the voxel level that can be exploited using graphics processing unit (GPU) architectures [112]. It is well established that GPU computing systems can achieve similar computation throughputs (floating point operations per second; FLOPS) as computing clusters, using less expensive equipment and less power [112,113]. Currently, tools that offer GPU implementations are BROCCOLI [114], freesurfer [115] and FSL [116]. Compared to the fastest multi-threaded implementation, BROCCOLI has achieved $195\times$ speedup for nonlinear registration and is $33\times$ faster for permutation testing [117]; the GPU implementation of freesurfer achieves a $6\times$ increase in speed for cortical extraction [115]; a GPU implementation achieved $100\times$ speedup for diffusion tractography [116]; and experiments with calculating functional connectivity using GPUs found a mean increase of $250\times$ more speed over a CPU implementation [118]. The speedups for permutation testing enable more accurate tests of statistical significance, as well as the objective comparison of statistical methods [119]. For example, the increase in speed afforded by GPUs made it possible to perform an in-depth evaluation of the specificity of statistical parameter mapping for task fMRI analyses in ten days; a simulation that would have taken 100 years on standard processors [120]. The major drawbacks of using GPUs for connectomes analysis are that few tools have been ported to these architectures and the additional level of programming sophistication required to develop software for GPUs, although programming libraries such as OpenCL (e.g. as described in Munshi *et al.* [121]) are simplifying the latter.

Open science resources for big data research

Significant barriers exists for "big data" scientists who wish to engage in connectomics research. The aforementioned imaging repositories have allowed significant progress to be made in assembling and openly sharing large datasets comprised of high-quality data from well-characterized populations. Before a dataset can be analyzed it must be preprocessed to remove nuisance variation and to make it comparable across individuals [93]. Additionally, the quality of the data must be assessed to determine if they are suitable for analysis. Both of these are daunting chores, and although several open source toolsets are available for performing these tasks, they

require a significant amount of domain-specific knowledge and labor to accomplish. The Preprocessed Connectomes Project (PCP) [122], the Human Connectome Project (HCP) [3,4], and others, are directly addressing this challenge by sharing data in its preprocessed form. The biggest challenge faced by these preprocessing initiatives is determining the preprocessing pipeline to be implemented. The HCP takes advantage of the uniformity its data collection to choose a single optimized pipeline [123]. Favoring plurality, the PCP approaches this problem by preprocessing the data using a variety of different processing tools and strategies. After an analysis is complete, the results can be compared to previous results from other analyses to assess their validity and to assist in their interpretation. Several hand-curated and automatically generated databases of neuroimaging results exist to aide in this effort [124-127]. Several data-sharing resources for raw and preprocessed neuroimaging data are listed in Section "List of resources for openly shared raw and processed neuroimaging data"; a nearly comprehensive index of open source software packages for working with neuroimaging data can be found at the Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC) [128].

List of resources for openly shared raw and processed neuroimaging data

- http://fcon_1000.projects.nitrc.org 1000 Functional Connectomes (FCP)*: Raw resting state functional MRI and structural MRI for more than 1200 healthy individuals from 33 different contributors [1].
- https://thedata.harvard.edu/dvn/dv/GSP Brain Genomics Superstruct Project (GSP)*: Raw resting state functional MRI, and structural MRI data, along with automated quality assessment and pre-computed brain morphometrics, and cognitive, personality, and behavior data for 1570 healthy, college-age individuals (18-35 years old) acquired using one of four MRI scanners. 1139 of the participants have second resting-state fMRI scans acquired from the same scanning session, and 69 have re-test scans [9].
- http://fcon_1000.projects.nitrc.org International Neuroimaging Datasharing Initiative (INDI): A follow-up to the 1000 Functional Connectomes Project, which shares raw resting state functional MRI, task-based functional MRI, structural MRI, and diffusion MRI data for 20 different projects; nine of which are being shared prospectively, as they are collected, and before publication. INDI contains data from a variety of different clinical populations and other experimental designs [2]. Notable examples are the http://fcon_1000.projects.nitrc.org/indi/adhd200 ADHD-200 [129], which contains 490 individuals

with ADHD and 598 typically developing controls, the http://fcon_1000.projects.nitrc.org/indi/ abideAutism Brain Imaging Data Exchange (ABIDE; 539 Autism and 573 healthy controls) [130], the http://fcon_1000.projects.nitrc.org/indi/CoRR/html/ Consortium for Reliability and Reproducibility (CoRR) [131], which contains test-retest datasets on over 1600 individuals, and the http://fcon_1000. projects.nitrc.org/indi/enhanced/ Enhanced Nathan Kline Institute-Rockland Sample [11], which is a community ascertained longitudinal sample with deep phenotyping.

- http://www.humanconnectomeproject.org/ Human Connectome Project (HCP): Raw and preprocessed resting state functional MRI, task functional MRI, structural MRI, diffusion MRI, deep phenotyping, and genetics data collected from a variety of individuals, including 1200 healthy adults (twins and non-twin siblings) by two consortia: one between Washington University St. Louis and University of Minnesota [4] and another between Massachusetts General Hospital and the University of Southern California [3]. The connectome projects are also developing and sharing imaging analysis pipelines and toolsets.

- http://ndar.nih.gov/ National Database for Autism Research (NDAR)*: An NIH-funded data repository of raw and preprocessed neuroimaging, phenotypic, and genomic data from a variety of different autism experiments [10].

- https://openfmri.org/ OpenFMRI: Raw and preprocessed data along with behavioral data for a variety of different task-based functional MRI experiments [132].

- http://pingstudy.ucsd.edu/ Pediatric Imaging, Neurocognition and Genetics (PING) Study: A multisite project that has collected "neurodevelopmental histories, information about mental and emotional functions, multimodal brain imaging data and genotypes for well over 1000 children and adolescents between the ages 3 and 20" [7]. Preprocessed structural and diffusion MRI data are also shared.

- http://www.med.upenn.edu/bbl/projects/pnc/ PhiladelphiaNeurodevelopmentalCohort.shtml Philadelphia Neurodevelopmental Cohort: Raw structural MRI, diffusion MRI, task functional MRI, resting state fMRI, cerebral blood flow, neuropsychiatric assessment, genotyping, and computerized neurocognitive testing data for 1445 individuals, 8-21 years old, including healthy controls and individuals with a variety of diagnoses [8].

- http://preprocessed-connectomes-project.github.io/ Preprocessed Connectomes Project (PCP)*: Preprocessed data, common statistical derivatives, and automated quality assessment measures for resting state fMRI, structural MRI, and diffusion MRI scans for data shared through INDI [122].

*These repositories contain data that is also available in INDI.

Conclusion

Functional connectomics is a "big data" science. As highlighted in this review, the challenge of learning statistical relationships between very high dimensional feature spaces and noisy or underspecified labels is rapidly emerging as rate-limiting steps for this burgeoning field and its promises to transform clinical knowledge. Accelerating the pace of discovery in functional connectivity research will require attracting data science researchers to develop new tools and techniques to address these challenges. It is our hope that recent augmentation of open science data-sharing initiatives with preprocessing efforts will catalyze the involvement of these researchers by reducing common barriers of entry.

Endnote

[a]Consistent with the literature, we use the term connectome to refer to the sum total of all connections in the human brain, and connectomics to refer to the scientific field dedicated to studying these connections.

References

1. Biswal BB, Mennes M, Zuo X-N, Gohel S, Kelly C, Smith SM, et al. Toward discovery science of human brain function. Proc Nat Acad Sci USA. 2010;107(10):4734–9. doi:10.1073/pnas.0911855107.
2. Mennes M, Biswal BB, Castellanos FX, Milham MP. Making data sharing work: the FCP/INDI experience. NeuroImage. 2013;82:683–91. doi:10.1016/j.neuroimage.2012.10.064.
3. Rosen B, Wedeen VJ, Horn JDV, Fischl B, Buckner RL, Wald L, et al. The Human Connectome Project. In: Proceedings Organization for Human Brain Mapping 16th Annual Meeting. Barcelona; 2010.
4. Van Essen DC, Ugurbil K. The future of the human connectome. NeuroImage. 2012;62(2):1299–310. doi:10.1016/j.neuroimage.2012.01.032.
5. Jiang T. Brainnetome: a new -ome to understand the brain and its disorders. NeuroImage. 2013;80:263–72. doi:10.1016/j.neuroimage.2013.04.002.
6. Assaf Y, Alexander DC, Jones DK, Bizzi A, Behrens TEJ, Clark Ca, et al. The CONNECT project: Combining macro- and micro-structure. NeuroImage. 2013;80:273–82. doi:10.1016/j.neuroimage.2013.05.055.
7. Jernigan TL, McCabe C, Chang L, Akshoomoff N, Newman E, Dale AM, et al, Pediatric Imaging Neurocognition and Genetics (PING) Study. Accessed 12 13 2014. http://pingstudy.ucsd.edu.
8. Satterthwaite TD, Elliott MA, Ruparel K, Loughead J, Prabhakaran K, Calkins ME, et al. Neuroimaging of the Philadelphia neurodevelopmental

cohort. Neuroimage. 2014;86:544–53. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3947233PMC3947233] [DOI:http://dx.doi.org/10.1016/j.neuroimage.2013.07.064] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/2392110123921101].

9. Buckner RL, Roffman JL, Smoller JW. Brain Genomics Superstruct Project (GSP). doi:10.7910/DVN/25833.

10. NIMH. National Database for Autism Research (NDAR). Accessed 12 13 2014. http://ndar.nih.gov.

11. Nooner KB, Colcombe SJ, Tobe RH, Mennes M, Benedict MM, Moreno AL, et al. The nki-rockland sample: A model for accelerating the pace of discovery science in psychiatry. Front Neurosc. 2012;6:152. doi:10.3389/fnins.2012.00152.

12. Craddock RC, Jbabdi S, Yan C-G, Vogelstein JT, Castellanos FX, Di Martino A, et al. Imaging human connectomes at the macroscale. Nat Methods. 2013;10(6):524–39. doi:10.1038/nmeth.2482.

13. Varela F, Lachaux JP, Rodriguez E, Martinerie J. The brainweb: phase synchronization and large-scale integration. Nat Rev Neurosci. 2001;2(4):229–39. [DOI:http://dx.doi.org/10.1038/3506755010.1038/35067550] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/1128374611283746].

14. Aguirre G. Number of neurons in a voxel. Accessed 12 13 2014. 2014. https://cfn.upenn.edu/aguirre/wiki/public:neurons_in_a_voxel.

15. Behrens TE, Sporns O. Human connectomics. Curr Opin Neurobiol. 2012;22(1):144–53. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3294015PMC3294015] [DOI:http://dx.doi.org/10.1016/j.conb.2011.08.00510.1016/j.conb.2011.08.005] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/2190818321908183].

16. Biswal B, Yetkin FZ, Haughton VM, Hyde JS. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn Reson Med. 1995;34(4):537–41. [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/85240218524021].

17. Varoquaux G, Craddock RC. Learning and comparing functional connectomes across subjects. NeuroImage. 2013;80:405–15. doi:10.1016/j.neuroimage.2013.04.007.

18. Kapur S, Phillips AG, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? Mol Psychiat. 2012;17(12):1174–9. [DOI:http://dx.doi.org/10.1038/mp.2012.10510.1038/mp.2012.105] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/2286903322869033].

19. Sporns O, Tononi G, Kötter R. The human connectome: A structural description of the human brain. PLoS Comput Biol. 2005;1(4):42. doi:10.1371/journal.pcbi.0010042.

20. Hagmann P. From diffusion MRI to brain connectomics. PhD thesis. Lausanne: STI; 2005. doi:10.5075/epfl-thesis-3230. http://vpaa.epfl.ch/page14976.html.

21. Kelly C, Biswal BB, Craddock RC, Castellanos FX, Milham MP. Characterizing variation in the functional connectome: promise and pitfalls. Trends Cogn Sci (Regul Ed). 2012;16(3):181–8. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3882689PMC3882689] [DOI:http://dx.doi.org/10.1016/j.tics.2012.02.00110.1016/j.tics.2012.02.001] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/2234121122341211].

22. Thirion B, Varoquaux G, Dohmatob E, Poline JB. Which fMRI clustering gives good brain parcellations? Front Neurosci. 2014;8:167. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4076743PMC4076743] [DOI:http://dx.doi.org/10.3389/fnins.2014.0016710.3389/fnins.2014.00167] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/2507142525071425].

23. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on {MRI} scans into gyral based regions of interest. NeuroImage. 2006;31(3):968–80. doi:10.1016/j.neuroimage.2006.01.021.

24. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage. 2002;15(1):273–89.

25. Eickhoff SB, Rottschy C, Kujovic M, Palomero-Gallagher N, Zilles K. Organizational principles of human visual cortex revealed by receptor mapping. Cereb Cortex. 2008;18(11):2637–45. doi:10.1093/cercor/bhn024. http://cercor.oxfordjournals.org/content/18/11/2637.full.pdf+html.

26. Bellec P, Perlbarg V, Jbabdi S, Pelegrini-Issac M, Anton JL, Doyon J, et al. Identification of large-scale networks in the brain using fMRI. Neuroimage. 2006;29(4):1231–43. [DOI:http://dx.doi.org/10.1016/j.neuroimage.2005.08.044] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/16246590].

27. Craddock RC, James GA, Iii PEH, Hu XP, Mayberg HS. A whole brain fMRI atlas generated via spatially constrained spectral clustering. Hum Brain Mapp. 2012;33(8). doi:10.1002/hbm.21333.A.

28. Blumensath T, Jbabdi S, Glasser MF, Van Essen DC, Ugurbil K, Behrens TE, et al. Spatially constrained hierarchical parcellation of the brain with resting-state fMRI. Neuroimage. 2013;76:313–24. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3758955] [DOI:http://dx.doi.org/10.1016/j.neuroimage.2013.03.024] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/23523803].

29. Thirion B, Flandin G, Pinel P, Roche A, Ciuciu P, Poline JB. Dealing with the shortcomings of spatial normalization: multi-subject parcellation of fMRI datasets. Hum Brain Mapp. 2006;27(8):678–93. [DOI:http://dx.doi.org/10.1002/hbm.20210] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/16281292].

30. Zalesky A, Fornito A, Harding IH, Cocchi L, Yucel M, Pantelis C, et al. Whole-brain anatomical networks: does the choice of nodes matter? Neuroimage. 2010;50(3):970–83. [DOI:http://dx.doi.org/10.1016/j.neuroimage.2009.12.027] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/20035887].

31. Flandin G, Kherif F, Pennec X, Riviere D, Ayache N, Poline J-B. Parcellation of brain images with anatomical and functional constraints for fmri data analysis. In: Biomedical Imaging, 2002. Proceedings. 2002 IEEE International Symposium On; 2002. p. 907–910. doi:10.1109/ISBI.2002.1029408.

32. Smith SM, Miller KL, Salimi-Khorshidi G, Webster M, Beckmann CF, Nichols TE, et al. Network modelling methods for FMRI. Neuroimage. 2011;54(2):875–91. [DOI:http://dx.doi.org/10.1016/j.neuroimage.2010.08.063] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/20817103].

33. Ryali S, Chen T, Supekar K, Menon V. Estimation of functional connectivity in fMRI data using stability selection-based sparse partial correlation with elastic net penalty. Neuroimage. 2012;59(4):3852–61. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3288428] [DOI:http://dx.doi.org/10.1016/j.neuroimage.2011.11.054] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/22155039].

34. Friston KJ. Functional and effective connectivity in neuroimaging: A synthesis. Hum Brain Mapp. 1994;2(1-2):56–78. doi:10.1002/hbm.460020107.

35. Buchel C, Friston KJ. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. Cereb Cortex. 1997;7(8):768–78. [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/9408041].

36. Friston KJ, Harrison L, Penny W. Dynamic causal modelling. Neuroimage. 2003;19(4):1273–302. [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/12948688].

37. Zhuang J, LaConte S, Peltier S, Zhang K, Hu X. Connectivity exploration with structural equation modeling: an fMRI study of bimanual motor coordination. Neuroimage. 2005;25(2):462–70. [DOI:http://dx.doi.org/10.1016/j.neuroimage.2004.11.007] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/15784425].

38. Penny WD, Stephan KE, Daunizeau J, Rosa MJ, Friston KJ, Schofield TM, et al. Comparing families of dynamic causal models. PLoS Comput Biol. 2010;6(3):1000709. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2837394] [DOI:http://dx.doi.org/10.1371/journal.pcbi.1000709] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/20300649].

39. James GA, Kelley ME, Craddock RC, Holtzheimer PE, Dunlop BW, Nemeroff CB, et al. Exploratory structural equation modeling of resting-state fMRI: applicability of group models to individual subjects. Neuroimage. 2009;45(3):778–87. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2653594] [DOI:http://dx.doi.org/10.1016/j.neuroimage.2008.12.049] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/19162206].

40. Lohmann G, Erfurth K, Müller K, Turner R. Critical comments on dynamic causal modelling. NeuroImage. 2012;59(3):2322–9. doi:10.1016/j.neuroimage.2011.09.025.

41. Deshpande G, Santhanam P, Hu X. Instantaneous and causal connectivity in resting state brain networks derived from functional MRI data. Neuroimage. 2011;54(2):1043–52. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2997120] [DOI:http://dx.doi.org/10.

1016/j.neuroimage.2010.09.024] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/20850549].

42. Lv Y, Margulies DS, Cameron Craddock R, Long X, Winter B, Gierhake D, et al. Identifying the perfusion deficit in acute stroke with resting-state functional magnetic resonance imaging. Ann Neurol. 2013;73(1):136–40. [DOI:http://dx.doi.org/10.1002/ana.23763] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/23378326.].

43. Craddock RC, Milham MP, LaConte SM. Predicting intrinsic brain activity. Neuroimage. 2013;82:127–36. [DOI:http://dx.doi.org/10.1016/j.neuroimage.2013.05.072] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/23707580].

44. James GA, Lu ZL, VanMeter JW, Sathian K, Hu XP, Butler AJ. Changes in resting state effective connectivity in the motor network following rehabilitation of upper extremity poststroke paresis. Top Stroke Rehabil. 2009;16(4):270–81. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3595191] [DOI:http://dx.doi.org/10.1310/tsr1604-270] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/19740732].

45. Brodersen KH, Schofield TM, Leff AP, Ong CS, Lomakina EI, Buhmann JM, et al. Generative embedding for model-based classification of fMRI data. PLoS Comput Biol. 2011;7(6):1002079. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3121683] [DOI:http://dx.doi.org/10.1371/journal.pcbi.1002079] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/21731479].

46. Hutchison RM, Womelsdorf T, Allen EA, Bandettini PA, Calhoun VD, Corbetta M, et al. Dynamic functional connectivity: promise, issues, and interpretations. Neuroimage. 2013;80:360–78.

47. Keilholz SD, Magnuson ME, Pan WJ, Willis M, Thompson GJ. Dynamic properties of functional connectivity in the rodent. Brain Connect. 2013;3(1):31–40. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3621313] [DOI:http://dx.doi.org/10.1089/brain.2012.0115] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/23106103].

48. Chang C, Glover GH. Time-frequency dynamics of resting-state brain connectivity measured with fMRI. 2010;50(1):81–98. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2827259] [DOI:http://dx.doi.org/10.1016/j.neuroimage.2009.12.011] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/20006716].

49. Yang Z, Craddock RC, Margulies DS, Yan CG, Milham MP. Common intrinsic connectivity states among posteromedial cortex subdivisions: Insights from analysis of temporal dynamics. Neuroimage. 2014;93 Pt 1: 124–137. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4010223] [DOI:http://dx.doi.org/10.1016/j.neuroimage.2014.02.014] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/24560717].

50. Allen EA, Damaraju E, Plis SM, Erhardt EB, Eichele T, Calhoun VD. Tracking whole-brain connectivity dynamics in the resting state. 2014;24(3):663–76. doi:10.1093/cercor/bhs352. http://cercor.oxfordjournals.org/content/24/3/663.full.pdf+html.

51. Majeed W, Magnuson M, Hasenkamp W, Schwarb H, Schumacher EH, Barsalou L, et al. Spatiotemporal dynamics of low frequency BOLD fluctuations in rats and humans. Neuroimage. 2011;54(2):1140–50. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2997178] [DOI:http://dx.doi.org/10.1016/j.neuroimage.2010.08.030] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/20728554].

52. Smith SM, Miller KL, Moeller S, Xu J, Auerbach EJ, Woolrich MW, et al. Temporally-independent functional modes of spontaneous brain activity. Proc Nat Acad Sci. 2012;109(8):3131–6. doi:10.1073/pnas.1121329109. http://www.pnas.org/content/109/8/3131.full.pdf+html.

53. Handwerker DA, Roopchansingh V, Gonzalez-Castillo J, Bandettini PA. Periodic changes in fMRI connectivity. Neuroimage. 2012;63(3):1712–9. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4180175] [DOI:http://dx.doi.org/10.1016/j.neuroimage.2012.06.078] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/22796990].

54. Castellanos FX, Di Martino A, Craddock RC, Mehta AD, Milham MP. Clinical applications of the functional connectome. Neuroimage. 2013;80:527–40. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3809093] [DOI:http://dx.doi.org/10.1016/j.neuroimage.2013.04.083] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/23631991].

55. Shirer WR, Ryali S, Rykhlevskaia E, Menon V, Greicius MD. Decoding subject-driven cognitive states with whole-brain connectivity patterns. Cereb Cortex. 2012;22(1):158–65. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3236795] [DOI:http://dx.doi.org/10.1093/

cercor/bhr099] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/21616982].

56. Krienen FM, Yeo BT, Buckner RL. Reconfigurable task-dependent functional coupling modes cluster around a core functional architecture. Philos Trans R Soc Lond B Biol Sci. 2014;369(1653). [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4150301] [DOI:http://dx.doi.org/10.1098/rstb.2013.0526] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/25180304].

57. Cole MW, Bassett DS, Power JD, Braver TS, Petersen SE. Intrinsic and task-evoked network architectures of the human brain. Neuron. 2014;83(1):238–51. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4082806] [DOI:http://dx.doi.org/10.1016/j.neuron.2014.05.014] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/24991964].

58. Zalesky A, Cocchi L, Fornito A, Murray MM, Bullmore E. Connectivity differences in brain networks. NeuroImage. 2012;60(2):1055–62. doi:10.1016/j.neuroimage.2012.01.068.

59. Ginestet CE, Simmons A. Statistical parametric network analysis of functional connectivity dynamics during a working memory task. Neuroimage. 2011;55(2):688–704. [DOI:http://dx.doi.org/10.1016/j.neuroimage.2010.11.030] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/21095229].

60. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Ann Stat. 2001;29(4):1165–88. doi:10.1214/aos/1013699998.

61. Shehzad Z, Kelly C, Reiss PT, Cameron Craddock R, Emerson JW, McMahon K, et al. A multivariate distance-based analytic framework for connectome-wide association studies. Neuroimage. 2014;93 Pt 1:74–94. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4138049] [DOI:http://dx.doi.org/10.1016/j.neuroimage.2014.02.024] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/24583255].

62. Craddock RC, Holtzheimer PE, Hu XP, Mayberg HS. Disease state prediction from resting state functional connectivity. Magn Reson Med. 2009;62(6):1619–28. doi:10.1002/mrm.22159.

63. Dosenbach NUF, Nardos B, Cohen AL, Fair DA, Power JD, Church JA, et al. Prediction of individual brain maturity using fMRI. Sci (New York, N.Y.) 2010;329(5997):1358–61. doi:10.1126/science.1194144.

64. Richiardi J, Eryilmaz H, Schwartz S, Vuilleumier P, Van De Ville D. Decoding brain states from fMRI connectivity graphs. NeuroImage. 2011;56(2):616–26. doi:10.1016/j.neuroimage.2010.05.081.

65. Thoma M, Cheng H, Gretton A, Han J, Kriegel H-P, Smola A, et al. Discriminative frequent subgraph mining with optimality guarantees. Stat Anal Data Min. 2010;3(5):302–18. doi:10.1002/sam.v3:5.

66. Bogdanov P, Dereli N, Bassett D. Learning about Learning: Human Brain Sub-Network Biomarkers in fMRI Data. arXiv preprint arXiv: … 2014. arXiv:1407.5590v1.

67. Richiardi J, Achard S, Bunke H, Van De Ville D. Machine learning with brain graphs: Predictive modeling approaches for functional imaging in systems neuroscience. Signal Process Mag IEEE. 2013;30(3):58–70. doi:10.1109/MSP.2012.2233865.

68. Rubinov M, Sporns O. Complex network measures of brain connectivity: uses and interpretations. NeuroImage. 2010;52(3):1059–69. doi:10.1016/j.neuroimage.2009.10.003.

69. Bullmore ET, Bassett DS. Brain graphs: graphical models of the human brain connectome. Ann Rev Clin Psychol. 2011;7:113–40. doi:10.1146/annurev-clinpsy-040510-143934.

70. Vogelstein JT, Roncal WG, Vogelstein RJ, Priebe CE. Graph classification using signal-subgraphs: Applications in statistical connectomics. IEEE Trans Pattern Anal Mach Intell. 2013;35(7):1539–51. doi:10.1109/TPAMI.2012.235.

71. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: with Applications in R. Springer Texts in Statistics: Springer; 2014. http://books.google.com/books?id=at1bmAEACAAJ.

72. Vapnik VN, Vapnik V. Statistical Learning Theory vol. 2. New York: Wiley; 1998.

73. Ryali S, Supekar K, Abrams DA, Menon V. Sparse logistic regression for whole-brain classification of fMRI data. Neuroimage. 2010;51(2):752–64. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2856747] [DOI:http://dx.doi.org/10.1016/j.neuroimage.2010.02.040] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/20188193].

74. Grimes DA, Schulz KF. Uses and abuses of screening tests. Lancet. 2002;359(9309):881–4. [DOI:http://dx.doi.org/10.1016/S0140-

6736(02)07948-5] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/11897304].

75. Altman DG, Bland JM. Diagnostic tests 2: Predictive values. BMJ. 1994;309(6947):102. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2540558] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/8038641].

76. Centers for Disease Control and Prevention. Morbidity and Mortality Weekly Report (MMWR). Accessed 12 13 2014. http://www.cdc.gov/mmwr/.

77. Lugosi G. Learning with an unreliable teacher. Pattern Recognit. 1992;25(1):79–87. doi:10.1016/0031-3203(92)90008-7.

78. Scott C, Blanchard G, Handy G, Pozzi S, Flaska M. ArXiv e-prints. 2013. 1303.1208.

79. Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am J Psychiat. 2010;167(7):748–51. [DOI:http://dx.doi.org/10.1176/appi.ajp.2010.09091379] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/20595427].

80. Gates KM, Molenaar PC, Iyer SP, Nigg JT, Fair DA. Organizing heterogeneous samples using community detection of GIMME-derived resting state functional networks. PLoS ONE. 2014;9(3):91322. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3958357] [DOI:http://dx.doi.org/10.1371/journal.pone.0091322] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/24642753].

81. Mørup M. Applications of tensor (multiway array) factorizations and decompositions in data mining. Wiley Interdiscip Rev: Data Min Knowl Discov. 2011;1(1):24–40. doi:10.1002/widm.1.

82. Franco AR, Ling J, Caprihan A, Calhoun VD, Jung RE, Heileman GL, et al. Multimodal and multi-tissue measures of connectivity revealed by joint independent component analysis. Selected Topics Signal Process IEEE J. 2008;2(6):986–97.

83. Groves AR, Beckmann CF, Smith SM, Woolrich MW. Linked independent component analysis for multimodal data fusion. Neuroimage. 2011;54(3):2198–217. [DOI:http://dx.doi.org/10.1016/j.neuroimage.2010.09.073] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/20932919].

84. Calhoun VD, Liu J, Adalı T. A review of group ica for fmri data and ica for joint inference of imaging, genetic, and erp data. Neuroimage. 2009;45(1):163–72.

85. Bottger J, Schurade R, Jakobsen E, Schaefer A, Margulies DS. Connexel visualization: a software implementation of glyphs and edge-bundling for dense connectivity data using brainGL. Front Neurosci. 2014;8:15. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3941704] [DOI:http://dx.doi.org/10.3389/fnins.2014.00015] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/24624052].

86. Birn RM. The role of physiological noise in resting-state functional connectivity. NeuroImage. 2012;62(2):864–70. doi:10.1016/j.neuroimage.2012.01.016.

87. Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE. Spurious but systematic correlations in functional connectivity {MRI} networks arise from subject motion. NeuroImage. 2012;59(3):2142–54. doi:10.1016/j.neuroimage.2011.10.018.

88. Dijk KRAV, Sabuncu MR, Buckner RL. The influence of head motion on intrinsic functional connectivity {MRI}. NeuroImage. 2012;59(1):431–8. doi:10.1016/j.neuroimage.2011.07.044. Neuroergonomics: The human brain in action and at work.

89. Yan C-G, Cheung B, Kelly C, Colcombe S, Craddock RC, Di Martino A, et al. A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. Neuroimage. 2013;76:183–201.

90. Satterthwaite TD, Wolf DH, Loughead J, Ruparel K, Elliott MA, Hakonarson H. Impact of in-scanner head motion on multiple measures of functional connectivity: Relevance for studies of neurodevelopment in youth. NeuroImage. 2012;60(1):623–632. doi:10.1016/j.neuroimage.2011.12.063.

91. Murphy K, Birn RM, Handwerker DA, Jones TB, Bandettini PA. The impact of global signal regression on resting state correlations: Are anti-correlated networks introduced? NeuroImage. 2009;44(3):893–905. doi:10.1016/j.neuroimage.2008.09.036.

92. Saad ZS, Gotts SJ, Murphy K, Chen G, Jo HJ, Martin A, et al. Trouble at rest: how correlation patterns and group differences become distorted after global signal regression. Brain Connect. 2012;2(1):25–32. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3484684]

[DOI:http://dx.doi.org/10.1089/brain.2012.0080] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/22432927].

93. Strother SC. Evaluating fmri preprocessing pipelines. Eng Med Biol Mag IEEE. 2006;25(2):27–41.

94. Satterthwaite TD, Elliott MA, Gerraty RT, Ruparel K, Loughead J, Calkins ME, Eickhoff SB, Hakonarson H, Gur RC, Gur RE, Wolf DH. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. NeuroImage. 2013;64(0):240–256. doi:10.1016/j.neuroimage.2012.08.052.

95. LaConte S, Anderson J, Muley S, Ashe J, Frutiger S, Rehm K, et al. The evaluation of preprocessing choices in single-subject bold fmri using npairs performance metrics. NeuroImage. 2003;18(1):10–27.

96. Strother SC, Anderson J, Hansen LK, Kjems U, Kustra R, Sidtis J, et al. The quantitative evaluation of functional neuroimaging experiments: the npairs data analysis framework. NeuroImage. 2002; 15(4):747–71.

97. Dinov I, Lozev K, Petrosyan P, Liu Z, Eggert P, Pierce J. Neuroimaging study designs, computational analyses and data provenance using the loni pipeline. PLoS ONE. 2010;5(9):13070. doi:10.1371/journal.pone.0013070.

98. Yan C-G, Zang Y-F. DPARSF: A MATLAB Toolbox for "Pipeline" Data Analysis of Resting-State fMRI. Front Syst Neurosci. 2010;4:13.

99. Bellec P, Lavoie-Courchesne S, Dickinson P, Lerch JP, Zijdenbos AP, Evans AC. The pipeline system for Octave and Matlab (PSOM): a lightweight scripting framework and execution engine for scientific workflows. Front Neuroinform. 2012;6:7.

100. Lavoie-Courchesne S, Rioux P, Chouinard-Decorte F, Sherif T, Rousseau M-E, Das S, et al. Integration of a neuroimaging processing pipeline into a pan-canadian computing grid. J Phys Conf Ser. 2012;341(1). doi:10.1088/1742-6596/341/1/012032.

101. Gorgolewski K, Burns CD, Madison C, Clark D, Halchenko YO, Waskom ML, et al. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. Front Neuroinform. 2011;5:13.

102. Craddock C, Sikka S, Cheung B, Khanuja R, Ghosh SS, Yan C, et al. Towards automated analysis of connectomes: The configurable pipeline for the analysis of connectomes (c-pac). Front Neuroinform. 2013;(42). doi:10.3389/conf.fninf.2013.09.00042.

103. Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res Int J. 1996;29(3):162–73.

104. Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med Image Anal. 2008;12(1):26–41.

105. Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage. 2004;23 Suppl 1:208–19. doi:10.1016/j.neuroimage.2004.07.051.

106. Friston KJ, Holmes AP, Worsley KJ, Poline J-P, Frith CD, Frackowiak RSJ. Statistical parametric maps in functional imaging: A general linear approach. Hum Brain Mapp. 1994;2(4):189–210. doi:10.1002/hbm.460020402.

107. Cox RW, Ashburner J, Breman H, Fissell K, Haselgrove C, Holmes CJ, et al. A (sort of) new image data format standard: NifTI-1. In: Proceedings Organization of Human Brain Mapping 10th Annual Meeting, Budapest, Hungary. Budapest, Hungary; 2004.

108. Amdahl GM. Validity of the single processor approach to achieving large scale computing capabilities. In: Proceedings of the April 18-20, 1967, Spring Joint Computer Conference. AFIPS '67 (Spring). New York, NY, USA: ACM; 1967. p. 483–485. doi:10.1145/1465482.1465560. http://doi.acm.org/10.1145/1465482.1465560.

109. O'Driscoll A, Daugelaite J, Sleator RD. 'big data', hadoop and cloud computing in genomics. J Biomed Inform. 2013;46(5):774–81. doi:10.1016/j.jbi.2013.07.001.

110. CPAC. Configurable Pipeline for the Analysis of Connectomes Amazon Machine Instance. Accessed 12 13 2014. 2014. https://github.com/FCP-INDI/ndar-dev/blob/master/aws_walkthrough.md.

111. NITRC. NITRC Computational Environment. Accessed 01 14 2015. 2014. https://aws.amazon.com/marketplace/pp/B00AW0MBLO/ref=mkt_ste_l2_hls_f1?nc2=h_l3_hl.

112. Eklund A, Andersson M, Knutsson H. fMRI analysis on the GPU-possibilities and challenges. Comput Methods Prog Biomed. 2012;105(2):145–61. doi:10.1016/j.cmpb.2011.07.007.

113. Hernandez D. Now You Can Build Google's $1M Artificial Brain on the Cheap. Wired. 2013;6(3):413–421.

114. Eklund A, Dufort P, Villani M, LaConte S. Broccoli: Software for fast fmri analysis on many-core cpus and gpus. Front Neuroinform. 2014;8:24.

115. Delgado J, Moure JC, Vives-Gilabert Y, Delfino M, Espinosa A, Gomez-Anson B. Improving the execution performance of FreeSurfer : a new scheduled pipeline scheme for optimizing the use of CPU and GPU resources. Neuroinformatics. 2014;12(3):413–21.

116. Hernandez M, Guerrero GD, Cecilia JM, Garcia JM, Inuggi A, Jbabdi S, et al. Accelerating fibre orientation estimation from diffusion weighted magnetic resonance imaging using GPUs. PLoS ONE. 2013;8(4):61892.

117. Eklund A, Dufort P, Forsberg D, Laconte SM. Medical image processing on the GPU - Past, present and future. Med Image Anal. 2013;17(8): 1073–94. doi:10.1016/j.media.2013.05.008.

118. Eklund A, Friman O, Andersson M, Knutsson H. A gpu accelerated interactive interface for exploratory functional connectivity analysis of fmri data. In: Image Processing (ICIP), 2011 18th IEEE International Conference On; 2011. p. 1589–1592. doi:10.1109/ICIP.2011.6115753.

119. Eklund A, Andersson M, Knutsson H. Fast random permutation tests enable objective evaluation of methods for single-subject FMRI analysis. Int J Biomed Imaging. 2011;2011:627947. doi:10.1155/2011/627947.

120. Eklund A, Andersson M, Josephson C, Johannesson M, Knutsson H. Does parametric fMRI analysis with SPM yield valid results? An empirical study of 1484 rest datasets. NeuroImage. 2012;61(3):565–78. doi:10.1016/j.neuroimage.2012.03.093.

121. Munshi A, Gaster B, Mattson TG, Fung J, Ginsburg D. OpenCL Programming Guide, 1st edn. Boston, MA: Addison-Wesley Professional; 2011.

122. Craddock RC, Bellec P. Preprocessed Connectomes Project (PCP). Accessed 12 13 2014. 2014. http://preprocessed-connectomes-project. github.io.

123. Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, et al. The minimal preprocessing pipelines for the Human Connectome Project. Neuroimage. 2013;80:105–24. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3720813] [DOI:http://dx.doi.org/10.1016/j.neuroimage.2013.04.127] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/23668970].

124. Fox PT, Lancaster JL. Opinion: Mapping context and content: the BrainMap model. Nat Rev Neurosci. 2002;3(4):319–21. [DOI:http://dx.doi. org/10.1038/nrn789] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/ 11967563].

125. Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD. Large-scale automated synthesis of human functional neuroimaging data. Nat Methods. 2011;8(8):665–70. [PubMed Central:http://www.ncbi.nlm.nih. gov/pmc/articles/PMC3146590] [DOI:http://dx.doi.org/10.1038/nmeth. 1635] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/21706013].

126. Gorgolewski C, Yarkoni T, Schwarz Y, Maumet C, Margulies D. Neurovault. Accessed 12 13 2014. 2014. http://www.neurovault.org.

127. Toro R. Brainspell. Accessed 12 13 2014. 2014. http://brainspell.org.

128. of Health Blueprint for Neuroscience Research, N.I. Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC). Accessed 12 13 2014. 2006. http://www.nitrc.org.

129. Milham MP, Fair D, Mennes M, Mostofsky SH. The adhd-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. Front Syst Neurosci. 2012;6(62). doi:10.3389/fnsys.2012.00062.

130. Di Martino A, Yan CG, Li Q, Denio E, Castellanos FX, Alaerts K, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol Psychiat. 2014;19(6): 659–67. [PubMed Central:http://www.ncbi.nlm.nih.gov/pmc/articles/ PMC4162310] [DOI:http://dx.doi.org/10.1038/mp.2013.78] [PubMed:http://www.ncbi.nlm.nih.gov/pubmed/23774715].

131. Zuo X-N, Anderson JS, Bellec P, Birn RM, Biswal BB, Blautzik J, et al. An open science resource for establishing reliability and reproducibility in functional connectomics. Sci Data. 2014;1.

132. Poldrack R. OpenfMRI. Accessed 12 13 2014. 2014. https://openfmri.org/.