REVIEW

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]

*Correspondence:
ccraddock@nki.rfmh.org
[1]Computational Neuroimaging Lab, Center for Biomedical Imaging and Neuromodulation, Nathan Kline Institute for Psychiatric Research, 140 Old Orangeburg Rd, 10962 Orangeburg, New York, USA
Full list of author information is available at the end of the article

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

1 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[1] where thousands of brain imaging scans, each of which consist of hundreds of observations of thousands of variables, are being collected and openly shared through a combination of grass-root 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], Pediatric Imaging, Neurocognition and Genetics (PING) Study [7], Philadelphia Neurodevelopmental Cohort [8], Brain Genomics Superstruct (GSP) [9], 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,

---

[1]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.

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 1,000,000s 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 (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 is lying quietly and not performing an experimenter specified task; 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.

2 The Connectome Analysis Paradigm
In 2005 Sporn and Hagmann [19, 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 that indicates 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 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, 21, 17] 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.

2.1 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. But, 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 are commonly reduced by combining voxels into larger brain areas for analysis. This is accomplished using either boundaries derived from anatomical landmarks [23, 24], regions containing homogeneous cyto-architecture 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 that they impose on the clustering solution [27, 28, 26, 29, 30, 31, 22]. There is some evidence in the literature that hierarchical clustering based methods perform best [28, 22], 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 [27, 22].

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 [32, 17], 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 [33, 17].

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 a network that can be modeled. Cross-validation methods have been proposed to systematically search for the best model [37, 38, 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]. But 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, 48, 49, 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 has already shown promise for extracting dynamic network modes using temporal ICA, although really large number 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, although initial work has shown that parcellations of at least some brain regions from dynamic iFC are consistent with what is found with static [49].

2.2 Mapping intra and inter -individual variation

The ultimate goals of connectomics is to map the brain's functional architecture and to annotate it with the cognitive or behavioral functions that 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, 56, 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 that describes 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, 63, 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, [66] were able to identify functional connectivity subgraphs that had 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 these methods are under active development and have not been widely adapted by the connectomics community.

Another approach for graph similarity using all the vertices involves computing a set of *graph-invariants* such as node centrality, modality, 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.

2.2.1 Predictive modeling

Resting state fMRI and iFC analyses are commonly applied to studying 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 than mass univariate techniques, all while lessening 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. The brain graphs can be represented by any of the previously discussed features. The learnt 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 have been applied to analyzing brain graphs in this manner, but by far the most commonly employed has been support vector machines [72, 54]. Although these methods offer excellent prediction accuracy, they are often black boxes, for which the information that is used to make the predictions is not easily discernible. The extraction of neuroscientifically meaningful information from the learnt model can be achieved by employing sparse methods [73] and feature selection methods [62] to reduce the input variables to only those that are 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 that permit 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 doesn't consider the prevalence of the disorder in the population, it doesn't provide an accurate picture of how well a clinical diagnostic 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]. See [54] for 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 quantifying 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 and to instead focus on dimensional measures of symptoms [79]. In the context of predictive modeling this translates into change in focus toward regression models, which to date have been underutilized for analyzing 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 finding the features that are most important for defining groups. Another problem is regularizing the clustering solution to make sure that it is relevant to the phenotypes under evaluation. These issues can be resolved using semi-supervised techniques or "multi-way" methods that incorporate phenotypic information to 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].

2.3 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 preprocessing 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, among other parameters. 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 [85, 86, 87, 88, 89] and denoising strategies [90, 91] 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 [92]. 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 [88]. 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[93]. 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 [86, 87, 88, 89]. 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 [94, 95]. 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.

3 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 were 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*. Since, in most cases, processing each dataset can be performed independently, these pipelines can be executed in parallel on large scale high-performance computing architectures, such as multi-core workstations or multi-workstation clusters [96, 97, 98, 99, 100, 101]. The construction of these pipelines are made possible by the modularity of most neuroimaging packages (e.g., AFNI [102], ANTs [103], FSL [104], SPM [105]), in which each processing step is implemented by separate functionality, and by their reliance on the NIFTI standard [106], which allows tools from different packages to be inter-mixed. Some of the steps of the pipeline are independent as well, and many of the toolsets are multithreaded, 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 speedup by the number of pipelines that are run in parallel, but in practice this is not quite obtainable due to overhead incurred by the increased competition for resources (Amdahl's Law [107]). A major advantage of this strategy is that it does not require any modifications to the existing neuroimaging tools and easily scales 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 [108]. Instead, when shared or institutional computing resources are not available, 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) [109] and the NITRC Computational Environment (NTRIC-CE) [110] eliminate many of the challenges associated with installing and maintaining open source tools. Preprocessing a single dataset (structural MRI and functional MRI for a single participant)

using the C-PAC AMI costs around $1.50 on Amazon EC2 using memory optimized compute nodes with 8 processors and 30 gigabytes of RAM, this could cost as little as $0.20 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 the functional connectivity analyses rely on 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 GPUs architectures [111]. It is well established that GPU computing systems can achieve similar computation throughputs (floating point operations per second or flops) as computing clusters, using less expensive equipment and less power [111, 112]. Currently, tools that offer GPU implementations are BROCCOLI [113], freesurfer [114] and FSL [115]. BROCCOLI has achieved 195× speedup for nonlinear registration compared to the fastest multi-threaded implementation and 33× speedup for permutation testing [116]; the GPU implementation of freesurfer achieves a 6× speedup for cortical extraction [114]; a GPU implementation achieved 100× speedup for diffusion tractography [115]; and experiments with calculating functional connectivity using GPUs found a mean 250× speedup over a CPU implementation [117]. The speedups for permutation testing enable more accurate tests of statistical significance, as well as the objective comparison of statistical methods [118]. For example, the speedup afforded by GPUs made it possible to perform an in-depth evaluation of the specificity of statistical parameter mapping for task fMRI analyses in 10 days, a simulation that would have taken 100 years on standard processors [119]. 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., [120]) are simplifying the latter.

4 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 made significant progress in assembling and openly sharing large datasets comprised of high-quality data from well-characterized populations. Before the data can be analyzed it must be preprocessed to remove nuisance variation and to make it comparable across individuals [92]. Additionally, the quality of the data must be assessed to determine if it is of 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 an man-power to accomplish. The Preprocessed Connectomes Project (PCP) [121], 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 that they will implement. The HCP takes advantage of the uniformity of their data collection to choose a single optimized preprocessing [122]. 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 [123, 124, 125, 126]. Several data-sharing resources for raw and preprocessed neuroimaging data are listed in table 1; 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) [127].

5 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 the common barriers of entry.

Author details
[1]Computational Neuroimaging Lab, 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.

References
1. Biswal, B.B., Mennes, M., Zuo, X.-N., Gohel, S., Kelly, C., Smith, S.M., Beckmann, C.F., Adelstein, J.S., Buckner, R.L., Colcombe, S., Dogonowski, A.-M., Ernst, M., Fair, D., Hampson, M., Hoptman, M.J., Hyde, J.S., Kiviniemi, V.J., Kötter, R., Li, S.-J., Lin, C.-P., Lowe, M.J., Mackay, C., Madden, D.J., Madsen, K.H., Margulies, D.S., Mayberg, H.S., McMahon, K., Monk, C.S., Mostofsky, S.H., Nagel, B.J., Pekar, J.J., Peltier, S.J., Petersen, S.E., Riedl, V., Rombouts, S.a.R.B., Rypma, B., Schlaggar, B.L., Schmidt, S., Seidler, R.D., Siegle, G.J., Sorg, C., Teng, G.-J., Veijola, J., Villringer, A., Walter, M., Wang, L., Weng, X.-C., Whitfield-Gabrieli, S., Williamson, P., Windischberger, C., Zang, Y.-F., Zhang, H.-Y., Castellanos, F.X., Milham, M.P.: Toward discovery science of human brain function. Proceedings of the National Academy of Sciences of the United States of America **107**(10), 4734–9 (2010). doi:10.1073/pnas.0911855107
2. Mennes, M., Biswal, B.B., Castellanos, F.X., Milham, M.P.: Making data sharing work: the FCP/INDI experience. NeuroImage **82**, 683–91 (2013). doi:10.1016/j.neuroimage.2012.10.064
3. Rosen, B., Wedeen, V.J., Horn, J.D.V., Fischl, B., Buckner, R.L., Wald, L., Hamalainen, M., Stufflebeam, S., Roffman, J., Shattuck, D.W., PM, T., Woods, R.P., Freimer, N., Bilder, R., , Toga, A.W.: The Human Connectome Project. In: Proceedings Organization for Human Brain Mapping 16th Annual Meeting, Barcelona (2010)
4. Van Essen, D.C., Ugurbil, K.: The future of the human connectome. NeuroImage **62**(2), 1299–310 (2012). doi:10.1016/j.neuroimage.2012.01.032
5. Jiang, T.: Brainnetome: a new -ome to understand the brain and its disorders. NeuroImage **80**, 263–72 (2013). doi:10.1016/j.neuroimage.2013.04.002
6. Assaf, Y., Alexander, D.C., Jones, D.K., Bizzi, A., Behrens, T.E.J., Clark, C.a., Cohen, Y., Dyrby, T.B., Huppi, P.S., Knoesche, T.R., Lebihan, D., Parker, G.J.M., Poupon, C., Anaby, D., Anwander, A., Bar, L., Barazany, D., Blumenfeld-Katzir, T., De-Santis, S., Duclap, D., Figini, M., Fischi, E., Guevara, P., Hubbard, P., Hofstetter, S., Jbabdi, S., Kunz, N., Lazeyras, F., Lebois, A., Liptrot, M.G., Lundell, H., Mangin, J.-F., Dominguez, D.M., Morozov, D., Schreiber, J., Seunarine, K., Nava, S., Riffert, T., Sasson, E., Schmitt, B., Shemesh, N., Sotiropoulos, S.N., Tavor, I., Zhang, H.G., Zhou, F.-L.: The CONNECT project: Combining macro- and micro-structure. NeuroImage **80**, 273–82 (2013). doi:10.1016/j.neuroimage.2013.05.055

7. Jernigan, T.L., McCabe, C., Chang, L., Akshoomoff, N., Newman, E., Dale, A.M., Ernst, T., Zijl, P.V., Kuperman, J., Murray, S., Bloss, C., Schork, N.J., Appelbaum, M., Gamst, A., Thompson, W., Bartsch, H., Keating, B., Amaral, D., Sowell, E., Kaufmann, W., Zijl, P.V., Mostofsky, S., Casey, B.J., Ruberry, E.J., Powers, A., Rosen, B., Kenet, T., Frazier, J., M.D., Kennedy, D., Gruen, J.: Pediatric Imaging, Neurocognition, and Genetics (PING) Study. http://pingstudy.ucsd.edu

8. Satterthwaite, T.D., Elliott, M.A., Ruparel, K., Loughead, J., Prabhakaran, K., Calkins, M.E., Hopson, R., Jackson, C., Keefe, J., Riley, M., Mentch, F.D., Sleiman, P., Verma, R., Davatzikos, C., Hakonarson, H., Gur, R.C., Gur, R.E.: Neuroimaging of the Philadelphia neurodevelopmental cohort. Neuroimage **86**, 544–553 (2014). [PubMed Central:PMC3947233] [DOI:10.1016/j.neuroimage.2013.07.064] [PubMed:23921101]

9. Buckner, R.L., Roffman, J.L., Smoller, J.W.: Brain Genomics Superstruct Project (GSP). doi:10.7910/DVN/25833

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

11. Nooner, K.B., Colcombe, S.J., Tobe, R.H., Mennes, M., Benedict, M.M., Moreno, A.L., Panek, L.J., Brown, S., Zavitz, S.T., Li, Q., Sikka, S., Gutman, D., Bangaru, S., Schlachter, R.T., Kamiel, S.M., Anwar, A.R., Hinz, C.M., Kaplan, M.S., Rachlin, A.B., Adelsberg, S., Cheung, B., Khanuja, R., Yan, C., Craddock, C.C., Calhoun, V., Courtney, W., King, M., Wood, D., Cox, C.L., Kelly, A.M., Di Martino, A., Petkova, E., Reiss, P.T., Duan, N., Thomsen, D., Biswal, B., Coffey, B., Hoptman, M.J., Javitt, D.C., Pomara, N., Sidtis, J.J., Koplewicz, H.S., Castellanos, F.X., Leventhal, B.L., Milham, M.P.: The nki-rockland sample: A model for accelerating the pace of discovery science in psychiatry. Frontiers in neuroscience **6**, 152 (2012). doi:10.3389/fnins.2012.00152

12. Craddock, R.C., Jbabdi, S., Yan, C.-G., Vogelstein, J.T., Castellanos, F.X., Di Martino, A., Kelly, C., Heberlein, K., Colcombe, S., Milham, M.P.: Imaging human connectomes at the macroscale. Nature methods **10**(6), 524–39 (2013). doi:10.1038/nmeth.2482

13. Varela, F., Lachaux, J.P., Rodriguez, E., Martinerie, J.: The brainweb: phase synchronization and large-scale integration. Nat. Rev. Neurosci. **2**(4), 229–239 (2001). [DOI:10.1038/35067550] [PubMed:11283746]

14. Aguirre, G.: Number of neurons in a voxel (2014). https://cfn.upenn.edu/aguirre/wiki/public:neurons_in_a_voxel

15. Behrens, T.E., Sporns, O.: Human connectomics. Curr. Opin. Neurobiol. **22**(1), 144–153 (2012). [PubMed Central:PMC3294015] [DOI:10.1016/j.conb.2011.08.005] [PubMed:21908183]

16. Biswal, B., Yetkin, F.Z., Haughton, V.M., Hyde, J.S.: Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn Reson Med **34**(4), 537–541 (1995). [PubMed:8524021]

17. Varoquaux, G., Craddock, R.C.: Learning and comparing functional connectomes across subjects. NeuroImage **80**, 405–415 (2013). doi:10.1016/j.neuroimage.2013.04.007

18. Kapur, S., Phillips, A.G., Insel, T.R.: Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? Mol. Psychiatry **17**(12), 1174–1179 (2012). [DOI:10.1038/mp.2012.105] [PubMed:22869033]

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

20. Hagmann, P.: From diffusion MRI to brain connectomics. PhD thesis (2005)

21. Kelly, C., Biswal, B.B., Craddock, R.C., Castellanos, F.X., Milham, M.P.: Characterizing variation in the functional connectome: promise and pitfalls. Trends Cogn. Sci. (Regul. Ed.) **16**(3), 181–188 (2012). [PubMed Central:PMC3882689] [DOI:10.1016/j.tics.2012.02.001] [PubMed:22341211]

22. Thirion, B., Varoquaux, G., Dohmatob, E., Poline, J.B.: Which fMRI clustering gives good brain parcellations? Front Neurosci **8**, 167 (2014). [PubMed Central:PMC4076743] [DOI:10.3389/fnins.2014.00167] [PubMed:25071425]

23. Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J.: An automated labeling system for subdividing the human cerebral cortex on {MRI} scans into gyral based regions of interest. NeuroImage **31**(3), 968–980 (2006). doi:10.1016/j.neuroimage.2006.01.021

24. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M.: Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage **15**(1), 273–289 (2002)

25. Eickhoff, S.B., Rottschy, C., Kujovic, M., Palomero-Gallagher, N., Zilles, K.: Organizational principles of human visual cortex revealed by receptor mapping. Cerebral Cortex **18**(11), 2637–2645 (2008). 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, J.L., Doyon, J., Benali, H.: Identification of large-scale networks in the brain using fMRI. Neuroimage **29**(4), 1231–1243 (2006). [DOI:10.1016/j.neuroimage.2005.08.044] [PubMed:16246590]

27. Craddock, R.C., James, G.A., Iii, P.E.H., Hu, X.P., Mayberg, H.S.: A whole brain fMRI atlas generated via spatially constrained spectral clustering. Human Brain Mapping **33**(8) (2012). doi:10.1002/hbm.21333.A

28. Blumensath, T., Jbabdi, S., Glasser, M.F., Van Essen, D.C., Ugurbil, K., Behrens, T.E., Smith, S.M.: Spatially constrained hierarchical parcellation of the brain with resting-state fMRI. Neuroimage **76**, 313–324 (2013). [PubMed Central:PMC3758955] [DOI:10.1016/j.neuroimage.2013.03.024] [PubMed:23523803]

29. Thirion, B., Flandin, G., Pinel, P., Roche, A., Ciuciu, P., Poline, J.B.: Dealing with the shortcomings of spatial normalization: multi-subject parcellation of fMRI datasets. Hum Brain Mapp **27**(8), 678–693 (2006). [DOI:10.1002/hbm.20210] [PubMed:16281292]

30. Zalesky, A., Fornito, A., Harding, I.H., Cocchi, L., Yucel, M., Pantelis, C., Bullmore, E.T.: Whole-brain anatomical networks: does the choice of nodes matter? Neuroimage **50**(3), 970–983 (2010). [DOI:10.1016/j.neuroimage.2009.12.027] [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, pp. 907–910 (2002). doi:10.1109/ISBI.2002.1029408

32. Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C.F., Nichols, T.E., Ramsey, J.D., Woolrich, M.W.: Network modelling methods for FMRI. Neuroimage **54**(2), 875–891 (2011). [DOI:10.1016/j.neuroimage.2010.08.063] [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 **59**(4), 3852–3861 (2012). [PubMed Central:PMC3288428] [DOI:10.1016/j.neuroimage.2011.11.054] [PubMed:22155039]

34. Friston, K.J.: Functional and effective connectivity in neuroimaging: A synthesis. Human Brain Mapping **2**(1-2), 56–78 (1994). doi:10.1002/hbm.460020107

35. Buchel, C., Friston, K.J.: Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. Cereb. Cortex **7**(8), 768–778 (1997). [PubMed:9408041]

36. Friston, K.J., Harrison, L., Penny, W.: Dynamic causal modelling. Neuroimage **19**(4), 1273–1302 (2003). [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 **25**(2), 462–470 (2005). [DOI:10.1016/j.neuroimage.2004.11.007] [PubMed:15784425]

38. Penny, W.D., Stephan, K.E., Daunizeau, J., Rosa, M.J., Friston, K.J., Schofield, T.M., Leff, A.P.: Comparing families of dynamic causal models. PLoS Comput. Biol. **6**(3), 1000709 (2010). [PubMed Central:PMC2837394] [DOI:10.1371/journal.pcbi.1000709] [PubMed:20300649]

39. James, G.A., Kelley, M.E., Craddock, R.C., Holtzheimer, P.E., Dunlop, B.W., Nemeroff, C.B., Mayberg, H.S., Hu, X.P.: Exploratory structural equation modeling of resting-state fMRI: applicability of group models to individual subjects. Neuroimage **45**(3), 778–787 (2009). [PubMed Central:PMC2653594] [DOI:10.1016/j.neuroimage.2008.12.049] [PubMed:19162206]

40. Lohmann, G., Erfurth, K., Müller, K., Turner, R.: Critical comments on dynamic causal modelling. NeuroImage **59**(3), 2322–2329 (2012). 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 **54**(2), 1043–1052 (2011). [PubMed Central:PMC2997120] [DOI:10.1016/j.neuroimage.2010.09.024] [PubMed:20850549]

42. Lv, Y., Margulies, D.S., Cameron Craddock, R., Long, X., Winter, B., Gierhake, D., Endres, M., Villringer, K., Fiebach, J., Villringer, A.: Identifying the perfusion deficit in acute stroke with resting-state functional magnetic resonance imaging. Ann. Neurol. **73**(1), 136–140 (2013). [DOI:10.1002/ana.23763] [PubMed:23378326]

43. Craddock, R.C., Milham, M.P., LaConte, S.M.: Predicting intrinsic brain activity. Neuroimage **82**, 127–136 (2013). [DOI:10.1016/j.neuroimage.2013.05.072] [PubMed:23707580]

44. James, G.A., Lu, Z.L., VanMeter, J.W., Sathian, K., Hu, X.P., Butler, A.J.: Changes in resting state effective connectivity in the motor network following rehabilitation of upper extremity poststroke paresis. Top Stroke Rehabil **16**(4), 270–281 (2009). [PubMed Central:PMC3595191] [DOI:10.1310/tsr1604-270] [PubMed:19740732]

45. Brodersen, K.H., Schofield, T.M., Leff, A.P., Ong, C.S., Lomakina, E.I., Buhmann, J.M., Stephan, K.E.: Generative embedding for model-based classification of fMRI data. PLoS Comput. Biol. **7**(6), 1002079 (2011). [PubMed Central:PMC3121683] [DOI:10.1371/journal.pcbi.1002079] [PubMed:21731479]

46. Hutchison, R.M., Womelsdorf, T., Allen, E.A., Bandettini, P.A., Calhoun, V.D., Corbetta, M., Della Penna, S., Duyn, J.H., Glover, G.H., Gonzalez-Castillo, J., Handwerker, D.A., Keilholz, S., Kiviniemi, V., Leopold, D.A., de Pasquale, F., Sporns, O., Walter, M., Chang, C.: Dynamic functional connectivity: promise, issues, and interpretations. Neuroimage **80**, 360–378 (2013)

47. Keilholz, S.D., Magnuson, M.E., Pan, W.J., Willis, M., Thompson, G.J.: Dynamic properties of functional connectivity in the rodent. Brain Connect **3**(1), 31–40 (2013). [PubMed Central:PMC3621313] [DOI:10.1089/brain.2012.0115] [PubMed:23106103]

48. Chang, C., Glover, G.H.: Time-frequency dynamics of resting-state brain connectivity measured with fMRI. Neuroimage **50**(1), 81–98 (2010). [PubMed Central:PMC2827259] [DOI:10.1016/j.neuroimage.2009.12.011] [PubMed:20006716]

49. Yang, Z., Craddock, R.C., Margulies, D.S., Yan, C.G., Milham, M.P.: Common intrinsic connectivity states among posteromedial cortex subdivisions: Insights from analysis of temporal dynamics. Neuroimage **93 Pt 1**, 124–137 (2014). [PubMed Central:PMC4010223] [DOI:10.1016/j.neuroimage.2014.02.014] [PubMed:24560717]

50. Allen, E.A., Damaraju, E., Plis, S.M., Erhardt, E.B., Eichele, T., Calhoun, V.D.: Tracking whole-brain connectivity dynamics in the resting state. Cerebral Cortex **24**(3), 663–676 (2014). 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, E.H., Barsalou, L., Keilholz, S.D.: Spatiotemporal dynamics of low frequency BOLD fluctuations in rats and humans. Neuroimage **54**(2), 1140–1150 (2011). [PubMed Central:PMC2997178] [DOI:10.1016/j.neuroimage.2010.08.030] [PubMed:20728554]

52. Smith, S.M., Miller, K.L., Moeller, S., Xu, J., Auerbach, E.J., Woolrich, M.W., Beckmann, C.F., Jenkinson, M., Andersson, J., Glasser, M.F., Van Essen, D.C., Feinberg, D.A., Yacoub, E.S., Ugurbil, K.: Temporally-independent functional modes of spontaneous brain activity. Proceedings of the National Academy of Sciences **109**(8), 3131–3136 (2012). doi:10.1073/pnas.1121329109. http://www.pnas.org/content/109/8/3131.full.pdf+html

53. Handwerker, D.A., Roopchansingh, V., Gonzalez-Castillo, J., Bandettini, P.A.: Periodic changes in fMRI connectivity. Neuroimage **63**(3), 1712–1719 (2012). [PubMed Central:PMC4180175] [DOI:10.1016/j.neuroimage.2012.06.078] [PubMed:22796990]

54. Castellanos, F.X., Di Martino, A., Craddock, R.C., Mehta, A.D., Milham, M.P.: Clinical applications of the functional connectome. Neuroimage **80**, 527–540 (2013). [PubMed Central:PMC3809093]

[DOI:10.1016/j.neuroimage.2013.04.083] [PubMed:23631991]

55. Shirer, W.R., Ryali, S., Rykhlevskaia, E., Menon, V., Greicius, M.D.: Decoding subject-driven cognitive states with whole-brain connectivity patterns. Cereb. Cortex **22**(1), 158–165 (2012). [PubMed Central:PMC3236795] [DOI:10.1093/cercor/bhr099] [PubMed:21616982]

56. Krienen, F.M., Yeo, B.T., Buckner, R.L.: Reconfigurable task-dependent functional coupling modes cluster around a core functional architecture. Philos. Trans. R. Soc. Lond., B, Biol. Sci. **369**(1653) (2014). [PubMed Central:PMC4150301] [DOI:10.1098/rstb.2013.0526] [PubMed:25180304]

57. Cole, M.W., Bassett, D.S., Power, J.D., Braver, T.S., Petersen, S.E.: Intrinsic and task-evoked network architectures of the human brain. Neuron **83**(1), 238–251 (2014). [PubMed Central:PMC4082806] [DOI:10.1016/j.neuron.2014.05.014] [PubMed:24991964]

58. Zalesky, A., Cocchi, L., Fornito, A., Murray, M.M., Bullmore, E.: Connectivity differences in brain networks. NeuroImage **60**(2), 1055–1062 (2012). doi:10.1016/j.neuroimage.2012.01.068

59. Ginestet, C.E., Simmons, A.: Statistical parametric network analysis of functional connectivity dynamics during a working memory task. Neuroimage **55**(2), 688–704 (2011). [DOI:10.1016/j.neuroimage.2010.11.030] [PubMed:21095229]

60. Benjamini, Y., Yekutieli, D.: The control of the false discovery rate in multiple testing under dependency. The Annals of Statistics **29**(4), 1165–1188 (2001). doi:10.1214/aos/1013699998

61. Shehzad, Z., Kelly, C., Reiss, P.T., Cameron Craddock, R., Emerson, J.W., McMahon, K., Copland, D.A., Castellanos, F.X., Milham, M.P.: A multivariate distance-based analytic framework for connectome-wide association studies. Neuroimage **93 Pt 1**, 74–94 (2014). [PubMed Central:PMC4138049] [DOI:10.1016/j.neuroimage.2014.02.024] [PubMed:24583255]

62. Craddock, R.C., Holtzheimer, P.E., Hu, X.P., Mayberg, H.S.: Disease state prediction from resting state functional connectivity. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine **62**(6), 1619–28 (2009). doi:10.1002/mrm.22159

63. Dosenbach, N.U.F., Nardos, B., Cohen, A.L., Fair, D.a., Power, J.D., Church, J.a., Nelson, S.M., Wig, G.S., Vogel, A.C., Lessov-Schlaggar, C.N., Barnes, K.A., Dubis, J.W., Feczko, E., Coalson, R.S., Pruett, J.R., Barch, D.M., Petersen, S.E., Schlaggar, B.L.: Prediction of individual brain maturity using fMRI. Science (New York, N.Y.) **329**(5997), 1358–61 (2010). 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 **56**(2), 616–26 (2011). doi:10.1016/j.neuroimage.2010.05.081

65. Thoma, M., Cheng, H., Gretton, A., Han, J., Kriegel, H.-P., Smola, A., Song, L., Yu, P.S., Yan, X., Borgwardt, K.M.: Discriminative frequent subgraph mining with optimality guarantees. Stat. Anal. Data Min. **3**(5), 302–318 (2010). 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., Ville, D.V.D.: Machine Learning with Brain Graphs (April), 58–70 (2013)

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

69. Bullmore, E.T., Bassett, D.S.: Brain graphs: graphical models of the human brain connectome. Annual review of clinical psychology **7**, 113–40 (2011). doi:10.1146/annurev-clinpsy-040510-143934

70. Vogelstein, J.T., Roncal, W.G., Vogelstein, R.J., Priebe, C.E.: Graph classification using signal-subgraphs: Applications in statistical connectomics. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(7), 1539–1551 (2013). 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, V.N., Vapnik, V.: Statistical Learning Theory vol. 2. Wiley New York, ??? (1998)

73. Ryali, S., Supekar, K., Abrams, D.A., Menon, V.: Sparse logistic regression for whole-brain classification of fMRI data. Neuroimage **51**(2), 752–764 (2010). [PubMed Central:PMC2856747] [DOI:10.1016/j.neuroimage.2010.02.040] [PubMed:20188193]

74. Grimes, D.A., Schulz, K.F.: Uses and abuses of screening tests. Lancet **359**(9309), 881–884 (2002). [DOI:10.1016/S0140-6736(02)07948-5] [PubMed:11897304]

75. Altman, D.G., Bland, J.M.: Diagnostic tests 2: Predictive values. BMJ **309**(6947), 102 (1994). [PubMed Central:PMC2540558] [PubMed:8038641]

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

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

78. Scott, C., Blanchard, G., Handy, G., Pozzi, S., Flaska, M.: Classification with Asymmetric Label Noise: Consistency and Maximal Denoising. ArXiv e-prints (2013). 1303.1208

79. Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D.S., Quinn, K., Sanislow, C., Wang, P.: Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am J Psychiatry **167**(7), 748–751 (2010). [DOI:10.1176/appi.ajp.2010.09091379] [PubMed:20595427]

80. Gates, K.M., Molenaar, P.C., Iyer, S.P., Nigg, J.T., Fair, D.A.: Organizing heterogeneous samples using community detection of GIMME-derived resting state functional networks. PLoS ONE **9**(3), 91322 (2014). [PubMed Central:PMC3958357] [DOI:10.1371/journal.pone.0091322] [PubMed:24642753]

81. Mørup, M.: Applications of tensor (multiway array) factorizations and decompositions in data mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **1**(1), 24–40 (2011). doi:10.1002/widm.1

82. Franco, A.R., Ling, J., Caprihan, A., Calhoun, V.D., Jung, R.E., Heileman, G.L., Mayer, A.R.: Multimodal and multi-tissue measures of connectivity revealed by joint independent component analysis. Selected Topics in Signal Processing, IEEE Journal of **2**(6), 986–997 (2008)

83. Groves, A.R., Beckmann, C.F., Smith, S.M., Woolrich, M.W.: Linked independent component analysis for multimodal data fusion. Neuroimage **54**(3), 2198–2217 (2011). [DOI:10.1016/j.neuroimage.2010.09.073]

[PubMed:20932919]

84. Calhoun, V.D., Liu, J., Adalı, T.: A review of group ica for fmri data and ica for joint inference of imaging, genetic, and erp data. Neuroimage **45**(1), 163–172 (2009)

85. Birn, R.M.: The role of physiological noise in resting-state functional connectivity. NeuroImage **62**(2), 864–870 (2012). doi:10.1016/j.neuroimage.2012.01.016

86. Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E.: Spurious but systematic correlations in functional connectivity {MRI} networks arise from subject motion. NeuroImage **59**(3), 2142–2154 (2012). doi:10.1016/j.neuroimage.2011.10.018

87. Dijk, K.R.A.V., Sabuncu, M.R., Buckner, R.L.: The influence of head motion on intrinsic functional connectivity {MRI}. NeuroImage **59**(1), 431–438 (2012). doi:10.1016/j.neuroimage.2011.07.044. Neuroergonomics: The human brain in action and at work

88. Yan, C.-G., Cheung, B., Kelly, C., Colcombe, S., Craddock, R.C., Di Martino, A., Li, Q., Zuo, X.-N., Castellanos, F.X., Milham, M.P.: A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. Neuroimage **76**, 183–201 (2013)

89. Satterthwaite, T.D., Wolf, D.H., Loughead, J., Ruparel, K., Elliott, M.A., Hakonarson, H., Gur, R.C., Gur, R.E.: Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. Neuroimage (2012)

90. Murphy, K., Birn, R.M., Handwerker, D.A., Jones, T.B., Bandettini, P.A.: The impact of global signal regression on resting state correlations: Are anti-correlated networks introduced? NeuroImage **44**(3), 893–905 (2009). doi:10.1016/j.neuroimage.2008.09.036

91. Saad, Z.S., Gotts, S.J., Murphy, K., Chen, G., Jo, H.J., Martin, A., Cox, R.W.: Trouble at rest: how correlation patterns and group differences become distorted after global signal regression. Brain Connect **2**(1), 25–32 (2012). [PubMed Central:PMC3484684] [DOI:10.1089/brain.2012.0080] [PubMed:22432927]

92. Strother, S.C.: Evaluating fmri preprocessing pipelines. Engineering in Medicine and Biology Magazine, IEEE **25**(2), 27–41 (2006)

93. Satterthwaite, T.D., Elliott, M.A., Gerraty, R.T., Ruparel, K., Loughead, J., Calkins, M.E., Eickhoff, S.B., Hakonarson, H., Gur, R.C., Gur, R.E., et al.: An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. NeuroImage (2012)

94. LaConte, S., Anderson, J., Muley, S., Ashe, J., Frutiger, S., Rehm, K., Hansen, L.K., Yacoub, E., Hu, X., Rottenberg, D., et al.: The evaluation of preprocessing choices in single-subject bold fmri using npairs performance metrics. NeuroImage **18**(1), 10–27 (2003)

95. Strother, S.C., Anderson, J., Hansen, L.K., Kjems, U., Kustra, R., Sidtis, J., Frutiger, S., Muley, S., LaConte, S., Rottenberg, D.: The quantitative evaluation of functional neuroimaging experiments: the npairs data analysis framework. NeuroImage **15**(4), 747–771 (2002)

96. Dinov, I., Lozev, K., Petrosyan, P., Liu, Z., Eggert, P., Pierce, J., Zamanyan, A., Chakrapani, S., Van Horn, J., Parker, D.S., Magsipoc, R., Leung, K., Gutman, B., Woods, R., Toga, A.: Neuroimaging study designs, computational analyses and data provenance using the LONI pipeline. PLoS ONE **5**(9) (2010)

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

98. Bellec, P., Lavoie-Courchesne, S., Dickinson, P., Lerch, J.P., Zijdenbos, A.P., Evans, A.C.: The pipeline system for Octave and Matlab (PSOM): a lightweight scripting framework and execution engine for scientific workflows. Front Neuroinform **6**, 7 (2012)

99. Lavoie-Courchesne, S., Rioux, P., Chouinard-Decorte, F., Sherif, T., Rousseau, M.-E., Das, S., Adalat, R., Doyon, J., **Craddock**, **R. C.** ., Margulies, D., C. Chu, O.L., Evans, A.C., Bellec, P.: Integration of a neuroimaging processing pipeline into a pan-canadian computing grid. Journal of Physics: Conference Series **341(1)** (2012). doi:10.1088/1742-6596/341/1/012032

100. Gorgolewski, K., Burns, C.D., Madison, C., Clark, D., Halchenko, Y.O., Waskom, M.L., Ghosh, S.S.: Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. Front Neuroinform **5**, 13 (2011)

101. Craddock, C., Sikka, S., Cheung, B., Khanuja, R., Ghosh, S.S., Yan, C., Li, Q., Lurie, D., Vogelstein, J., Burns, R., Colcombe, S., Mennes, M., Kelly, C., Di Martino, A., Castellanos, F.X., Milham, M.: Towards automated analysis of connectomes: The configurable pipeline for the analysis of connectomes (c-pac). Frontiers in Neuroinformatics (42) (2013). doi:10.3389/conf.fninf.2013.09.00042

102. Cox, R.W.: AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Computers and biomedical research, an international journal **29**(3), 162–73 (1996)

103. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med Image Anal **12**(1), 26–41 (2008)

104. Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M.: Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage **23 Suppl 1**, 208–19 (2004). doi:10.1016/j.neuroimage.2004.07.051

105. Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.-P., Frith, C.D., Frackowiak, R.S.J.: Statistical parametric maps in functional imaging: A general linear approach. Human Brain Mapping **2**(4), 189–210 (1994). doi:10.1002/hbm.460020402

106. Cox, R.W., Ashburner, J., Breman, H., Fissell, K., Haselgrove, C., Holmes, C.J., Lancaster, J.L., Rex, D.E., Smith, S.M., Woodward, J.B., Strother, S.C.: 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)

107. Amdahl, G.M.: 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), pp. 483–485.

ACM, New York, NY, USA (1967). doi:10.1145/1465482.1465560.
http://doi.acm.org/10.1145/1465482.1465560

108. O'Driscoll, A., Daugelaite, J., Sleator, R.D.: 'big data', hadoop and cloud computing in genomics. Journal of Biomedical Informatics **46**(5), 774–781 (2013). doi:10.1016/j.jbi.2013.07.001

109. CPAC: Configurable Pipeline for the Analysis of Connectomes Amazon Machine Instance (2014). http://thecloudmarket.com/owner/015424812512?owner=015424812512&search_term=&order=created_at%20desc

110. NITRC: NITRC Computational Environment (2014). https://aws.amazon.com/marketplace/pp/B00AW0MBLO/ref=mkt_ste_l2_hls_f1?nc2=h_l3_hl

111. Eklund, A., Andersson, M., Knutsson, H.: fMRI analysis on the GPU-possibilities and challenges. Computer methods and programs in biomedicine **105**(2), 145–61 (2012). doi:10.1016/j.cmpb.2011.07.007

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

113. Eklund, A., Dufort, P., Villani, M., LaConte, S.: Broccoli: Software for fast fmri analysis on many-core cpus and gpus. Frontiers in Neuroinformatics **8**, 24 (2014)

114. Delgado, J., Moure, J.C., 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 **12**(3), 413–421 (2014)

115. Hernandez, M., Guerrero, G.D., Cecilia, J.M., Garcia, J.M., Inuggi, A., Jbabdi, S., Behrens, T.E., Sotiropoulos, S.N.: Accelerating fibre orientation estimation from diffusion weighted magnetic resonance imaging using GPUs. PLoS ONE **8**(4), 61892 (2013)

116. Eklund, A., Dufort, P., Forsberg, D., Laconte, S.M.: Medical image processing on the GPU - Past, present and future. Medical image analysis **17**(8), 1073–1094 (2013). doi:10.1016/j.media.2013.05.008

117. 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, pp. 1589–1592 (2011). doi:10.1109/ICIP.2011.6115753

118. Eklund, A., Andersson, M., Knutsson, H.: Fast random permutation tests enable objective evaluation of methods for single-subject FMRI analysis. International journal of biomedical imaging **2011**, 627947 (2011). doi:10.1155/2011/627947

119. 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 **61**(3), 565–78 (2012). doi:10.1016/j.neuroimage.2012.03.093

120. Munshi, A., Gaster, B., Mattson, T.G., Fung, J., Ginsburg, D.: OpenCL Programming Guide, 1st edn. Addison-Wesley Professional, ??? (2011)

121. Craddock, R.C., Bellec, P.: Preprocessed Connectomes Project (PCP) (2014). http://preprocessed-connectomes-project.github.io

122. Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., Van Essen, D.C., Jenkinson, M.: The minimal preprocessing pipelines for the Human Connectome Project. Neuroimage **80**, 105–124 (2013). [PubMed Central:PMC3720813] [DOI:10.1016/j.neuroimage.2013.04.127] [PubMed:23668970]

123. Fox, P.T., Lancaster, J.L.: Opinion: Mapping context and content: the BrainMap model. Nat. Rev. Neurosci. **3**(4), 319–321 (2002). [DOI:10.1038/nrn789] [PubMed:11967563]

124. Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D.: Large-scale automated synthesis of human functional neuroimaging data. Nat. Methods **8**(8), 665–670 (2011). [PubMed Central:PMC3146590] [DOI:10.1038/nmeth.1635] [PubMed:21706013]

125. Gorgolewski, C., Yarkoni, T., Schwarz, Y., Maumet, C., Margulies, D.: Neurovault (2014). http://www.neurovault.org

126. Toro, R.: Brainspell (2014). http://www.brainspell.org

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

128. Bottger, J., Schurade, R., Jakobsen, E., Schaefer, A., Margulies, D.S.: Connexel visualization: a software implementation of glyphs and edge-bundling for dense connectivity data using brainGL. Front Neurosci **8**, 15 (2014). [PubMed Central:PMC3941704] [DOI:10.3389/fnins.2014.00015] [PubMed:24624052]

129. Milham, M.P., Fair, D., Mennes, M., Mostofsky, S.H.: The adhd-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. Frontiers in Systems Neuroscience **6**(62) (2012). doi:10.3389/fnsys.2012.00062

130. Di Martino, A., Yan, C.G., Li, Q., Denio, E., Castellanos, F.X., Alaerts, K., Anderson, J.S., Assaf, M., Bookheimer, S.Y., Dapretto, M., Deen, B., Delmonte, S., Dinstein, I., Ertl-Wagner, B., Fair, D.A., Gallagher, L., Kennedy, D.P., Keown, C.L., Keysers, C., Lainhart, J.E., Lord, C., Luna, B., Menon, V., Minshew, N.J., Monk, C.S., Mueller, S., Muller, R.A., Nebel, M.B., Nigg, J.T., O'Hearn, K., Pelphrey, K.A., Peltier, S.J., Rudie, J.D., Sunaert, S., Thioux, M., Tyszka, J.M., Uddin, L.Q., Verhoeven, J.S., Wenderoth, N., Wiggins, J.L., Mostofsky, S.H., Milham, M.P.: The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol. Psychiatry **19**(6), 659–667 (2014). [PubMed Central:PMC4162310] [DOI:10.1038/mp.2013.78] [PubMed:23774715]

131. Zuo, X.-N., Anderson, J.S., Bellec, P., Birn, R.M., Biswal, B.B., Blautzik, J., Breitner, J.C.S., Buckner, R.L., Calhoun, V.D., Castellanos, F.X., Chen, A., Chen, B., Chen, J., Chen, X., Colcombe, S.J., Courtney, W., Craddock, R., Martino, A.D., Dong, H.-M., Fu, X., Gong, Q., Gorgolewski, K.J., Han, Y., He, Y., He, Y., Ho, E., Holmes, A., Hou, X.-H., Huckins, J., Jiang, T., Jiang, Y., Kelley, W., Kelly, C., King, M., LaConte, S.M., Lainhart, J.E., Lei, X., Li, H.-J., Li, K., Li, K., Lin, Q., Liu, D., Liu, J., Liu, X., Liu, Y., Lu, G., Lu, J., Luna, B., Luo, J., Lurie, D., Mao, Y., Margulies, D.S., Mayer, A.R., Meindl, T., Meyerand, M.E., Nan, W., Nielsen, J.A., O'Connor, D., Paulsen, D., Prabhakaran, V., Qi, Z., Qiu, J., Shao, C., Shehzad, Z., Tang, W., Villringer, A., Wang, H., Wang, K., Wei, D., Wei, G.-X., Weng, X.-C., Wu, X., Xu, T., Yang, N., Yang, Z., Zang, Y.-F., Zhang, L., Zhang, Q., Zhang, Z., Zhang, Z., Zhao, K., Zhen, Z., Zhou, Y., Zhu, X.-T., Milham, M.P.: An

open science resource for establishing reliability and reproducibility in functional connectomics. Scientific Data (in press)

132. Poldrack, R.: OpenfMRI (2014). https://openfmri.org/

Figures

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.

Figure 2 Identifying communities based on neurophenotypes. Brain glyphs provide succinct representations of whole brain functional connectivity [128].

Tables

Table 1 List of resources for openly shared raw and processed neuroimaging data. *These repositories contain data that is also available in INDI.

1000 Functional Connectomes (FCP)*: Raw resting state functional MRI and structural MRI for more than 1200 healthy individuals from 33 different contributors [1].

Brain Genomics Superstruct Project (GSP)*: Raw resting state functional MRI, and structural MRI data, along with automated quality assessment and pre-computed brain morphometrics, along with cognitive, personality, and behavior data for 1570 college age healthy individuals (18 - 35 years old) acquired using 1 of 4 MRI scanners. 1139 of the participants have a second resting-state fMRI scans acquired from the same scan and 69 have re-test scans [9].

International Neuroimaging Datasharing Initiative (INDI): A followup to the 1,000 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. Contains data from a variety of different clinical populations and other experimental designs [2]. Notable examples are the ADHD-200 [129], which contains 490 individuals with ADHD and 598 typically developing controls, the Autism Brain Imaging Data Exchange (ABIDE; 539 Autism and 573 healthy controls) [130], the Consortium for Reliability and Reproducibility (CoRR) [131], which contains test-retest datasets on over 1600 individuals, and the Enhanced Nathan Kline Institute - Rockland Sample [11], which is a community ascertained longitudinal sample with deep phenotyping.

Human Connectome Project (HCP): Raw and preprocessed resting state functional MRI, task functional MRI, structural MRI, diffusion MRI, deep phenotyping, and genetics collected on a variety of individuals including 1,200 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 University of Southern California [3]. The connectome projects are also developing and sharing imaging analysis pipelines and toolsets.

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].

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

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.

Philadelphia Neurodevelopmental Cohort: Raw structural MRI, diffusion MRI, task functional MRI, resting state fMRI, cerebral blood flow, neuropsychiatric assessment, genotyping, and computerized neurocognitive testing for 1445 individuals, 8 - 21 years old, including healthy controls and individuals with a variety of diagnoses [8].

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 [121].