Richard Cameron Craddock
Research Statement

Introduction

Through integrating novel data analysis and neuroimaging techniques, I aim to transform the standard of care for mental health disorders by identifying brain-based biomarkers. Once found, objective biological tests of disease presence and severity will make it possible to track disease progression, parse hetereogeniety of affected populations and predict treatment outcomes. Although the tools and resources that I build can be applied broadly, I have primarily focused on connectomics, which represents the brain as a graph of its functional and structural connections [1]. This is an emerging area of neuroscience that relies on large-scale computation to infer brain connections from neuroimaging data and map them to inter-individual variations in phenotype (i.e. sex, IQ, cognitive abilities, indices of mental health) [2, 3]. My contributions in this field include **generating very large high quality datasets**, **developing high capacity tools and resources to efficiently analyze them**, and **creating innovative experimental paradigms to test generated hypothesis**. I have taken an interdisciplinary approach to addressing these issues that integrate aspects of computer engineering, data science, medical imaging, and neuroscience. Consistent with the tenets of open science, I openly share all of the ideas, software, and data developed through these research efforts.

Previous and ongoing research

Novel data analyses for clinical connectomics. For my dissertation research, I developed supervised and unsupervised learning frameworks for identifying connectome-based biomarkers of mental health disorders. I applied support vector classification and novel feature selection algorithms to learn a classifier of depression from functional MRI (FMRI) data and achieved 95% prediction accuracy [4]. To make this analysis tractable, the dimensionality of the problem was reduced by focusing on the connectivity of 15 brain areas that were chosen based on the existing literature. In order to perform similar analyses on the whole brain, without introducing biases from previous results or anatomical brain atlases, I later developed a spectral clustering approach for learning functionally homogenous brain areas directly from FMRI data [5]. The result was a multi-scale brain atlas offering various levels of dimensionality reduction while retaining more information about the connectome than anatomical atlases. During my postdoctoral fellowship, I worked with colleagues to develop a method based on time-lag correlation to map brain areas whose blood flow is restricted after stroke [6], which has been subsequently patented [7]. In my current position I have worked with collaborators to build a framework for connectome wide association studies [8] and to study the impact of connectome size on prediction analyses [9]. I have also authored reviews that educate researchers on the benefits and methodological concerns for applying machine learning in connectomes research [1, 2, 3, 10, 11, 12].

Examining temporal dynamics of brain networks. Although most connectome analyses assume that brain networks are static, it is well known that their structure can vary with time or cognitive state. During my post-doctoral fellowship, I utilized support vector regression to learn predictive models of brain networks that could be applied to data from a different time, cognitive state, or individual to measure the generalization of the network across conditions [13]. Using this technique, I was able to show that dynamics in brain connectivity follow the functional hierarchy of sensory integration proposed in the 1980's. This research was awarded a 1st place poster award in functional neuroimaging at the 2011 meeting of the International Society for Magnetic Resonance in Medicine (ISMRM), the de facto conference for MRI physics. More recently, I have continued this research by working with colleagues to investigate how boundaries between functionally defined brain areas change over time [14].

Examining network dynamics with real-time FMRI. Connectomic analyses of brain disorders typically focus on identifying differences in graph structure, but failure to appropriately regulate brain networks has been implicated in several disorders including ADHD, depression, social anxiety, traumatic brain injury and many others. During my post-doctoral fellowship, I developed a multivariate method for tracking the activity of networks from FMRI data in real-time [15]. The output of this system can be used to develop a brain-computer interface, provide neurofeedback to the participant, or enable a variety of other dynamic experiments. I am currently using this technique to identify the mechanisms and phenotypic correlates of network dysregulation across a variety of mental health disorders and symptom profiles. I have received a 2010 NARSAD Young Investigators Award from the Brain and Behavior Foundation and an ongoing 2012 Biobehavioral Research Award for Innovative New Scientists (BRAINS) R01 from the National Institutes of Mental Health to support this research.

Building large-scale datasets. Answering the next generation of neuroscientific questions using data science techniques will require very large and phenotypically rich datasets. My colleagues and I are taking a multifaceted open-science approach for building data sharing repositories for amassing these datasets and making them broadly accessible. In my current position, I am involved in the NKI Enhanced Rockland Sample (NKI-ERS) and Healthy Brain Network (HBN), which are both large-scale data collection efforts that combine comprehensive neuroimaging with a diverse assessment battery. The NKI-ERS is a consortium of four NIH grants (my R01 and 3 others) that are pooling resources to collect data from thousands of individuals, with a variety of mental health disorders, symptoms, and cognitive abilities, from throughout the lifespan [16]. In addition to partially funding the effort, I oversee the neuroinformatics infrastructure used to house and analyze the data and assist with image acquisition related issues. The HBN is a foundation-funded initiative that aims to collect data from 10,000 children and adolescents, across the spectrum of mental health and cognitive abilities, from the greater New York City area. I am the director of imaging for this endeavor, which involves specifying the neuroimaging protocols, overseeing image acquisition, as well as managing and analyzing the data.

Complimentary to these data collection efforts, my colleagues and I are actively aggregating data from smaller experiments across many different labs and openly sharing the results. Among these efforts is the International Neuroimaging Datasharing Initiative (INDI), to which I contribute data as well as assist in the data curation process [17]. The data from the NKI, HBN, and INDI initiatives are openly shared on the Internet with limited restriction beyond necessary privacy protections. Despite the open access of these data repositories, the size and complexity of the processing necessary to make the data suitable for further analysis presents an insurmountable hurdle for many would-be researchers. During my post-doctoral fellowship my colleagues and I created the Preprocessed Connectomes Project to calculate and share preprocessed versions of data available through INDI [18]. The results have been used in 50 publications, 3 PhD dissertations, 3 Master's theses, and a patent.

Open source software for high capacity data processing and analysis. Conventional neuroimaging tools were developed in an era typified by small scale data analysis and do not have the capacity to efficiently process very large datasets. My team at the Child Mind Institute is building the Configurable Pipeline for the Analysis of Connectomes (CPAC) open source software package to combine these tools into pipelines for automated processing and managing their efficient execution on high-performance computing resources [19]. Methods are chosen for inclusion into CPAC based on painstaking evaluations of the variety of techniques proposed in the literature (e.g., [20, 21, 22]). When an open source implementation of a method is unavailable, or is overly resource intensive, we develop our own with careful optimization. CPAC additionally includes features that enable plurality by allowing multiple configurations to be run simultaneously without needlessly duplicating computation, as well as the ability to restart pipelines after an error or reconfiguration without recomputing unaffected pipeline steps. Due to the success of CPAC, we received a contract from the National Institutes of Mental Health to integrate

it with the National Database for Autism Research data sharing repository and perform a proof-of-concept for performing neuroimaging data analysis using cloud computing [23].

Assessing the quality of neuroimaging data and determining which data should be included in an analysis are outstanding issues. In an effort to begin to address this problem, my team and I are developing the Quality Assessment Protocol (QAP), an open source software package for estimating many of the measures proposed in the literature [24]. Ultimately the goal is to use these measures to learn automated classifiers for indicating poor quality data and to this end we have also developed a repository of QAP measures, along with manually applied quality labels, calculated on data from openly shared data repositories.

Future research

In addition to continuing support of the INDI, CPAC, and PCP open science initiatives; my future research will focus on large-scale connectomics analysis and utilizing real-time FMRI to develop dynamic experiments for measuring network function.

Large scale analyses of connectomes to delineate the neurobiology of mental health disorders. Mental health disorders are syndromes diagnosed by the presence of a set of symptoms rather than an objective measure of biology. As a consequence, a substantial amount of variation exists between individuals who receive the same diagnosis, confounding the ability to map disorders to brain structure and function. Overcoming this hurdle will require new approaches to analyzing this data that focus on dimensions of symptoms rather than categorical labels (regression models) and redefine clinical classifications based on connectome profiles (clustering). In addition to working with data typified by high dimensionality and few observations ($p >> n$), new methods will need to integrate disparate data (i.e. different image modalities, physiological measures, genetics, and assessment data) to fully characterize the subtle neurobiological signatures of mental health disorders. I have begun to build the infrastructure for this research agenda by amassing the large and phenotypically rich datasets needed and developing computational tools for efficiently processing the data. This line of research will involve applying existing and developing new data analysis frameworks, the careful comparison of methods in terms of predictive power and reproducibility of results, and harnessing new computational paradigms (e.g. cloud computing, GPUs) to push the limits of the size and sophistication of analyses that can be performed.

Real-time FMRI based dynamic experiments to characterize brain network function in health and disease across the lifespan. The real-time fMRI system that I developed provides me with a unique capability to perform neurofeedback experiments for characterizing and elucidating the underlying mechanisms of network dysregulation in mental health disorders. In the future, I plan to build on my current research into the phenotypic variation related to network regulation in adults, by looking at how these relationships change across the lifespan. To make the current experiment attractive to child and adolescent populations, I plan to develop video games that use the feedback signal to control some aspect of game play. But, since younger participants tend to move quite a bit while in the scanner, there is a concern that the head motion will degrade the feedback signal. Typical approaches for dealing with head motion involves modeling its impact on the data using a regression framework and subtracting the result [25]. Performing this regression in real-time involves an iterative method that refits the model every time a new data point is available, which results in variation in the quality of the model fit over time. To fix this problem I am working with previously proposed multi-echo FMRI acquisition techniques to differentiate neural signal from noise [26] and adapting this computationally intensive method so that it can be used in real-time. I am currently piloting this research with funding from a competitive supplement to my R01, and in the next few years I plan to secure additional funds for a full-scale evaluation of attentional networks in children.

References

[1] **R. Cameron Craddock**, Saad Jbabdi, Chao-Gan Yan, Joshua Vogelstein, F. Xavier Castellanos, Adriana Di Martino, Clare Kelly, Keith Heberlein, Stan Colcombe, and Michael P. Milham. Imaging human connectomes at the macroscale. *Nature Methods*, 10(6):524–539, Jun 2013. [DOI:10.1038/nmeth.2482] [PMID:23722212] [Open Access].

[2] Gaël Varoquaux and **R. Cameron Craddock**. Learning and comparing functional connectomes across subjects. *NeuroImage*, 80:405–415, Oct 2013. [DOI:10.1016/j.neuroimage.2013.04.007] [PMID:23583357] [Open Access].

[3] **R. Cameron Craddock**, Rosalia L Tungaraza, and Michael P Milham. Connectomics and new approaches for analyzing human brain functional connectivity. *GigaScience*, 4(1):13, March 2015. [DOI:10.1186/s13742-015-0045-x] [PMID:25810900] [Open Access].

[4] **R. Cameron Craddock**, Paul E. Holtzheimer, Xiaoping P. Hu, and Helen S. Mayberg. Disease state prediction from resting state functional connectivity. *Magnetic Resonance in Medicine*, 62(6):1619–1628, Dec 2009. [DOI:10.1002/mrm.22159] [PMID:19859933] [Open Access].

[5] **R. Cameron Craddock**, G. Andrew James, Paul E. Holtzheimer, Xiaoping P. Hu, and Helen S. Mayberg. A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human Brain Mapping*, 33(8):1914–1928, Aug 2012. [DOI:10.1002/hbm.21333] [PMID:21769991] [Open Access] [Open Source].

[6] Yating Lv, Daniel S. Margulies, **R. Cameron Craddock**, Xiangyu Long, Benjamin Winter, Daniel Gierhake, Matthias Endres, Kersten Villringer, Jochen Fiebach, and Arno Villringer. Identifying the perfusion deficit in acute stroke with resting-state functional magnetic resonance imaging. *Annals of Neurology*, 73(1):136–140, Jan 2013. [DOI:10.1002/ana.23763] [PMID:23378326].

[7] Daniel Margulies, Yating LV, **R. Cameron Craddock**, and Arno Villringer. Method of visualization of tissue perfusion by means of assessing bold signal fluctuations, an apparatus therefor and the use thereof, Jun 2013. US Patent 20,130,144,154.

[8] Z. Shehzad, C. Kelly, P. T. Reiss, **R. Cameron Craddock**, J. W. Emerson, K. McMahon, D. A. Copland, F. X. Castellanos, and M. P. Milham. A multivariate distance-based analytic framework for connectome-wide association studies. *NeuroImage*, 93 Pt 1:74–94, Jun 2014. [DOI:10.1016/j.neuroimage.2014.02.024] [PMID:24583255] [Open Access] [Open Source].

[9] P. Bellec, Y. Benhajali, F. Carbonell, C. Dansereau, G. Albouy, M. Pelland, **C. Craddock**, O. Collignon, J. Doyon, E. Stip, and P. Orban. Impact of the resolution of brain parcels on connectome-wide association studies in fMRI. *Neuroimage*, 123:212–228, Dec 2015. [DOI:10.1016/j.neuroimage.2015.07.071] [PMID:26241681] [Open Access].

[10] F. Xavier Castellanos, Adriana Di Martino, **R. Cameron Craddock**, Ashesh D. Mehta, and Michael P. Milham. Clinical applications of the functional connectome. *NeuroImage*, 80:527–540, Oct 2013. [DOI:10.1016/j.neuroimage.2013.04.083] [PMID:23631991] [Open Access].

[11] C. Kelly, B. B. Biswal, **R. C. Craddock**, F. X. Castellanos, and M. P. Milham. Characterizing variation in the functional connectome: promise and pitfalls. *Trends in Cognitive Science (Regular Edition)*, 16(3):181–188, Mar 2012. [DOI:10.1016/j.tics.2012.02.001] [PMID:22341211] [Open Access].

[12] A. Di Martino, D. A. Fair, C. Kelly, T. D. Satterthwaite, F. X. Castellanos, M. E. Thomason, **R. C. Craddock**, B. Luna, B. L. Leventhal, X. N. Zuo, and M. P. Milham. Unraveling the Miswired Connectome: A Developmental Perspective. *Neuron*, 83(6):1335–1353, Sep 2014. [DOI:10.1016/j.neuron.2014.08.050] [PMID:25233316] [Open Access].

[13] **R. Cameron Craddock**, Michael P. Milham, and Stephen M. LaConte. Predicting intrinsic brain activity. *NeuroImage*, 82:127–136, Nov 2013. [DOI:10.1016/j.neuroimage.2013.05.072] [PMID:23707580].

[14] Z. Yang, **R. C. Craddock**, D. S. Margulies, C. G. Yan, and M. P. Milham. Common intrinsic connectivity states among posteromedial cortex subdivisions: Insights from analysis of temporal dynamics. *NeuroImage*, 93 Pt 1:124–137, Jun 2014. [DOI:10.1016/j.neuroimage.2014.02.014] [PMID:24560717] [Open Access].

[15] **R. Cameron Craddock**, Jonathan M. Lisinski, Pearl Chiu, Helen S. Mayberg, and Stephen M. LaConte. Real-time tracking and biofeedback of the default mode network. In *Proceedings Organization of Human Brain Mapping 18th Annual Meeting, Beijing*, pages 648, MT, Beijing, 2012.

[16] Kate Brody Nooner, Stanley J. Colcombe, Russell H. Tobe, Maarten Mennes, Melissa M. Benedict, Alexis L. Moreno, Laura J. Panek, Shaquanna Brown, Stephen T. Zavitz, Qingyang Li, Sharad Sikka, David Gutman, Saroja Bangaru, Rochelle Tziona Schlachter, Stephanie M. Kamiel, Ayesha R. Anwar, Caitlin M. Hinz, Michelle S. Kaplan, Anna B. Rachlin, Samantha Adelsberg, Brian Cheung, Ranjit Khanuja, Chaogan Yan, **R. Cameron Craddock**, Vincent Calhoun, William Courtney, Margaret King, Dylan Wood, Christine L. Cox, A. M. Clare Kelly, Dawn Thomsen, Bharat Biswal, Barbara Coffey, Matthew J. Hoptman, Daniel C. Javitt, Nunzio Pomara, John J. Sidtis, Harold S. Koplewicz, Francisco Xavier Castellanos, Bennet L. Leventhal, and Michael P. Milham. 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] [PMID:23087608] [Open Access].

[17] M. Mennes, B. B. Biswal, F. X. Castellanos, and M. P. Milham. Making data sharing work: the FCP/INDI experience. *Neuroimage*, 82:683–691, Nov 2013.

[18] P. Bellec, C. Chu, F. Chouinard-Decorte, Y. Benhajali, D. S. Margulies, and **R. C. Craddock**. The Neuro Bureau ADHD-200 Preprocessed repository. *Neuroimage*, Jul 2016. [DOI:10.1016/j.neuroimage.2016.06.034] [PMID:27423255][Data Sharing] [Open Access] [Preprint].

[19] **R. Cameron Craddock**, Sharad Sikka, Brian Cheung, Ranjit Khanuja, Satra Ghosh, Chao-Gan YAN, Qingyang Li, Daniel Lurie, Joshua Vogelstein, Randal Burns, Stan Colcombe, Maarten Mennes, Clare Kelly Adriana Di Martino, F. Xavier Castellanos, and Michael Milham. Towards automated analysis of connectomes: The Configurable Pipeline for the Analysis of Connectomes. In *Neuroinformatics 2013, Stockholm, Sweden*, Stockholm, Sweden, 2013.

[20] C. G. Yan, B. Cheung, C. Kelly, S. Colcombe, **R. C. Craddock**, A. Di Martino, Q. Li, X. N. Zuo, F. X. Castellanos, and M. P. Milham. A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. *NeuroImage*, 76:183–201, Aug 2013. [DOI:10.1016/j.neuroimage.2013.03.004] [PMID:23499792] [Open Access].

[21] Chao-Gan Yan, **R. Cameron Craddock**, Xi-Nian Zuo, Yue-Feng Zang, and Michael P. Milham. Standardizing the intrinsic brain: towards robust measurement of inter-individual variation in 1000 functional connectomes. *NeuroImage*, 80:246–262, Oct 2013. [DOI:10.1016/j.neuroimage.2013.04.081] [PMID:23631983] [Open Access].

[22] Chao-Gan Yan, **R. Cameron Craddock**, Yong He, and Michael P. Milham. Addressing head motion dependencies for small-world topologies in functional connectomics. *Frontiers in Human Neurosci*, 7:910, 2013. [DOI:10.3389/fnhum.2013.00910] [PMID:24421764] [Open Access].

[23] Daniel Clark, Christian Haselgrove, David Kennedy, Zhizhong Liu, Michael Milham, Petros Petrosyan, Carinna Torgerson, John Van Horn, and **R. Cameron Craddock**. Harnessing cloud

computing for high capacity analysis of neuroimaging data from NDAR. In *Proceedings Organization of Human Brain Mapping 21st Annual Meeting, Honolulu, Hawaii*, Honolulu, Hawaii, 2015.

[24] Zarrar Shehzad, Steven Giavasis, Qingyang Li, Yassine Benhajali, Chaogan Yan, Zhen Yang, Michael Milham, Pierre Bellec, and **R. Cameron Craddock**. The Preprocessed Connectomes Project Quality Assessment Protocol - a resource for measuring the quality of mri data. *Frontiers in Neuroscience*, (47), 2015.

[25] **R. Cameron Craddock**, Jonathan M. Lisinski, and Stephen LaConte. Online denoising strategies for real-time tracking default mode network activity. In *Proceedings Third Biennial International Conference on Resting-State Functional Connectivity*, Magdeburg, Germany, 2012.

[26] P. Kundu, B. E. Benson, K. L. Baldwin, D. Rosen, W. M. Luh, P. A. Bandettini, D. S. Pine, and M. Ernst. Robust resting state fMRI processing for studies on typical brain development based on multi-echo EPI acquisition. *Brain Imaging Behav*, 9(1):56–73, Mar 2015.