Specific Aims

A major research priority in medicine and psychiatry is the search for biomarkers to diagnose neuropsychiatric illnesses, characterize symptom clusters across diagnoses and predict treatment response (Insel and Cuthbert, 2009). Understanding interactions involving variables such as age, gender, clinical status, drug tolerance, and treatment, *i.e.*, biomarkers involving patterns of statistical dependence among multiple variables, can be critical for patients' outcome and survival. For k=2,3,..., define "k^{th}-order dependence" to be that which can be discerned in groups of k variables, but no fewer. Principal components analysis and many other standard multivariate statistical procedures only describe dependence of order 2. But dependence of order greater than 2 ("high-order" dependence) can be important.

Statistical dependence can be studied with regression-type or structural equation modeling, but the biological relationships that underlie these models are not always understood. Another approach is to conduct "agnostic" analyses that make few assumptions and treat all variables the same *a priori*. Regression analysis, for example, is not agnostic because it requires designating some variables as responses and others as predictors. A challenge for extracting meaningful biomarkers is to agnostically describe high-order dependence among many variables in an interpretable fashion.

We have developed a new statistical method, "concurrence topology", that can be used to succinctly describe high-order dependence in dozens of variables in an agnostic fashion. Concurrence topology represents multivariate data as a series of abstract shapes and describes the topology of those shapes. We have applied it in a preliminary analysis (article under review) to find differences in high-order functional connectivity in samples of 25 attention deficit hyperactivity disorder (ADHD) subjects and 41 controls. Using concurrence topology, we found numerous differences in high-order dependence structure in the two groups, including a robust difference in 6*th*-order dependence in individual subject's fMRI data, and even evidence of a difference in 7*th*-order dependence. This feat is difficult, if not impossible, with other statistical methods because a naive, but agnostic, approach would likely attempt to summarize all 3,365,856 different groups of seven brain regions among 32 regions. Instead of looking at the 3,365,856 trees, concurrence topology finds patterns in the forest. But for limits in computing speed, we could have investigated even higher orders of dependence (Aim 3).

We are proposing to use concurrence topology to obtain descriptions of high-order dependence to help identify novel biomarkers, by pursuing the following aims:

Aim 1. To uncover high-order functional connectivity biomarkers in *individual* subject data, we will first confirm and extend our brain imaging findings in the much larger ADHD-200 fMRI data set (776 subjects), and second, apply our method to the ABIDE autism spectrum disorder (ASD) resting-state fMRI data set (1,110 subjects). We will use concurrence topology to classify individuals by diagnosis and even subtype. The large sample sizes will allow cross-validation to reduce the false positive rate.

Aim 2. To find biomarkers involving non-time series and non-imaging data based on *group*-level patterns of high-order dependence, we will first refine our method for doing this and then apply it to dozens of variables from different domains. These variables include clinical, demographic, behavioral, questionnaire, genotype, *etc.* in pre-existing data to look for diagnostic biomarkers for ADHD, ASD, suicidal behavior, and treatment response in major depressive disorder. Again, large sample sizes will permit cross-validation.

Aim 3. To apply concurrence topology with more variables (regions, in fMRI applications), more fMRI subjects, and to examine higher levels of dependence, we will, concurrently with and in support of Aims 1 and 2, rewrite our software to make it faster and easier to read and maintain for wider adoption by the scientific community.

The above Aims permit us to further develop and apply our new statistical approach to extract biomarkers as patterns of high-order dependence in neuropsychiatric data. Our long-term goal is to uncover the biological relationships that underlie the patterns of dependence we find with concurrence topology to improve our understanding of the pathophysiology in these and other neuropsychiatric illnesses.

Research Strategy

1. Significance

1.1. Meeting the challenge of biomarker discovery

Diagnosis of mental disorders suffers from a dearth of reliable biomarkers [1]. The importance of identifying biomarkers for mental disorders is reflected by its inclusion in the National Institute of Mental Health's Strategic Objectives (Strategy 1.3): "Currently, very few biomarkers have been identified for mental disorders due in part to their complexity and an incomplete understanding of the neurobiological basis of mental disorders..." Examples of disorders in need of biomarkers are attention deficit hyperactivity disorder (ADHD), autism spectrum disorder (ASD), major depressive disorder (MDD), and suicidal behavior.

ADHD "affects at least 5-10% of school-age children and is associated with substantial lifelong impairment, with annual direct costs exceeding $36 billion/year in the US. Despite a voluminous empirical literature, the scientific community remains without a comprehensive model of the pathophysiology of ADHD. Further, the clinical community remains without objective biological tools capable of informing the diagnosis of ADHD for an individual or guiding clinicians in their decision-making regarding treatment." (http://fcon_1000.projects.nitrc.org/indi/adhd200/)

ASD "are now recognized to occur in more than 1% of children, causing immense suffering to individuals and their families." (http://fcon_1000.projects.nitrc.org/indi/abide/)

MDD has an overwhelming impact on the health of Americans, as noted by the NIMH (http://www.nimh.nih.gov/health/publications/the-numbers-count-mental-disorders-in-america/index.shtml). It is the leading cause of disability in the U.S. for ages 15-44 and affects approximately 14.8 million American adults, or about 6.7 percent of the U.S. population age 18 and older in a given year (http://www.who.int/healthinfo/global_burden_disease/GBD_report_2004update_AnnexA.pdf, http://www.census.gov/popest/national/asrh/).

Suicide is the eleventh leading cause of death in the United States with over 30,000 individuals committing suicide per year [2], making suicide a significant public health concern. A recent National Institute of Mental Health (NIMH) initiative underscores the need for further research on how to reduce the suicide rate [3]. There are currently no biological markers that are being used to identify those at risk.

We attribute the elusiveness of biomarkers partly to the fact that traditional methods used to analyze neuropsychiatric data do not adequately reflect their complexity. The Research Domain Criteria project (RDoC, http://www.nimh.nih.gov/research-funding/rdoc/index.shtml) of the NIMH encourages investigation of "functioning ... across multiple units of analysis, from genes to neural circuits to behaviors, cutting across disorders as traditionally defined." This is facilitated by statistical analysis of variables across several domains at once. This cannot be done with just a handful of variables.

In the proposed research we will further develop and apply a new statistical method to capture this complexity in functional connectivity data [4] (**Aim 1**) and data from multiple non-time series data sources (**Aim 2**) to help identify biomarkers of neuropsychiatric illnesses, specifically related to ADHD, ASD, and suicidal behavior. We will also make our software fast and in accord with modern open source, distributed software practices for ease of use by the scientific community (**Aim 3**).

1.2. Taking advantage of large data sources for extracting biomarkers

Advances in neuroimaging brain activity have opened up tremendous stores of rich data from which biomarkers may be drawn (see sections **3.5.1** and **3.5.2** for descriptions of the data we will use). An important aspect of brain activity is the interaction among brain regions. This interaction is reflected in "functional connectivity" in neuroimaging time series, which is "*statistical dependencies between spatially segregated neuronal events*" [5]. For concreteness we discuss functional connectivity in the context of blood oxygenated level-dependent (BOLD) functional magnetic resonance imaging (fMRI, [6,4]) with the understanding that our methods apply to general multivariate time series. Moreover, we consider the problem of describing or summarizing the functional connectivity in the fMRI *of an individual person*. This is the level of statistics with which **Aim 1** is concerned.

Functional neuroimaging data alone tell only a part of the story about the health of an individual. Phenotypic data (and non-time series imaging data) of many kinds are routinely acquired from patients and should be included in a search for biomarkers of complex neuropsychiatric illnesses. **Aim 2**, consistent with RDoC, is concerned with describing or summarizing data from multiple sources (**3.5.2**), not simply functional connectivity times series brain imaging data. (However, summaries from analyses performed under **Aim 1** can serve as inputs to the sorts of analyses of **Aim 2**.) The method with which **Aim 2** is concerned starts with *group*-level statistics and translates them into the level of the individual.

Our longterm goal is to use these functional connectivity and multiple-source biomarkers to gain insight into the pathophysiology of mental illnesses for prediction of disease course and treatment response in individual patients to better personalize treatment accordingly.

1.3. Overcoming problems with current approaches for extracting biomarkers

Individual variables might serve as biomarkers. Many statistical methods can help identify such variables. But patterns of the joint distribution of variables (*i.e.*, the form of statistical dependence) might also serve as biomarkers. One way to study these patterns is to specify a regression, classifier, or structural equation model involving only a few variables. (We lump these methods together under the generic term "model".)

Dependence structure often has an "order": If a feature of the joint distribution of variables can be detected by looking at k variables at a time, but not by looking only at $k-1$ variables at a time, then that feature reflects "k^{th}-order dependence" among the variables. For example, traditional cluster analysis of variables, (Pearson, Kendall, and Spearman) correlation, factor analysis (including principal components analysis), and linear discriminant analysis are measures of 2nd-order dependence because those analyses can all be carried out by looking at the variables two at a time. "High-order" dependence is dependence of order at least three. A model including interactions registers high-order dependence.

However, one might not wish to posit such a specific model, but instead proceed "agnostically". Data analysis is "agnostic" if, *a priori*, all variables are treated the same (for $k = 1, 2, 3,...$ all groups of k variables are treated the same) and few *a priori* assumptions are made about the nature of the joint distribution. The aforementioned cluster, correlation, *etc.* analysis methods are all agnostic.

If there are many variables it is difficult to capture high-order dependence agnostically using a model. In general, in agnostically analyzing increasingly high orders of dependence in large numbers of variables one must overcome a "combinatorial explosion". Agnostically studying k^{th}-order dependence means looking for a pattern in all "k-tuples" of variables, *i.e.*, all groups of k variables, of which there can be very many. In our preliminary fMRI analyses (**3.4.1**) we examined 2nd- (2,701 pairs), 3rd- (64,824 triplets), and 4th-order dependence (1,150,626 quadruplets) in 74 variables. We also looked at 7th-order dependence in 32 variables (Example 1). That analysis involves 3,365,856 septuplets. So on the face of it if one wished to study 7th-order dependence among 32 variables one would have to make sense of 3,365,856 numbers.

Statistical methods such as independent components analysis [7], generalizations of factor analysis [8], and "boosting" and "kernel-based" methods [9] in machine learning can capture high-order dependence in an implicit manner, but for purposes of interpretation or explanation a more explicit presentation is desired. Current methods that can be more explicit are limited in how high an order of dependence they can describe. Examples are latent variable methods [10] and perhaps the method of [11]. The "lasso" [9] is a regression method that can accommodate a large number of terms in the model and at the same time selects terms, but will not scale to tens of thousands or millions of possible interaction terms.

1.4. Concurrence topology

In this proposal, we discuss a new method, "concurrence topology", that makes remarkable headway in overcoming the combinatorial explosion without the drawbacks of the above-mentioned statistical methods. The paper by Ellis and Klein [12] (under review by a statistics journal) gives an introductory account. Concurrence topology makes use of ideas from the mathematical field of algebraic topology [13]. (Topologist Prof. Steven Ferry at Rutgers University has agreed to serve as an unpaid consultant providing support in the area of topological theory. See letter of support.) Our work was inspired by [14], which applied topological methods to analyze the firing of simulated rat hippocampal place cells.

Concurrence topology is used on dichotomized data. It works by describing k^{th}-order dependence, not directly in terms of k-tuples, but in terms of larger structures called "homology classes". Concurrence topology begins by representing multivariate data as a series of abstract shapes. Homology classes are just holes in the shapes and represent what might be thought of as inhibitory relationships. Cluster analysis can used to study the dependence among variables. Viewing cluster analysis as a method for finding, not clusters,

but *gaps* between clusters, concurrence topology can be thought of as "cluster analysis on steroids." One advantage of concurrence topology is that it radically reduces the size of the selection problem from selecting from thousands or millions of *k*-tuples to selecting from a few dozen (usually fewer) homology classes. This gives one a reasonable chance at managing the multiple comparisons problem.

Concurrence topology can reveal complex relationships within dozens (but not 100s or 1,000s) of variables. Hence, it can draw connections among multiple domains (*e.g.*, genes, circuits, behavior). Another advantage is that, in principle, concurrence topology can be used to investigate dependence of *any* order. (The obstacle is computational, addressed in **Aim 3**.) Concurrence topology is also explicit: homology classes are associated with specific orders of dependence and can be "localized" (**3.3**) to reveal specific groups of variables most closely associated with them. Since concurrence topology can be computed from time series data from just one subject (**Aim 1**), summaries of patient concurrence topology could potentially be used for diagnostic purposes. In the sort of analyses proposed under **Aim 2**, group-level descriptions of dependence are translated into individual-level interaction variables that can again be used for diagnosis.

2. Innovation

This proposal is innovative with regard to the methods we are developing, the domain to which we are applying them, and to the results we are seeking.

Methods: Concurrence topology is a new approach to multivariate data analysis that applies algebraic topology to statistics and is radically different from other currently used statistical methods. Not only does it more succinctly summarize high-order dependence than most conventional agnostic statistical methods do, but concurrence topology captures very different aspects of high-order dependence and therefore complements other approaches. (But we apply conventional statistical methods to concurrence topology output; **3.4.**)

Domain and results: We apply topology for the first time to resting state fMRI (functional connectivity) data and to non-imaging phenotypic neuropsychiatric variables and use topology for the first time to extract biomarkers of neuropsychiatric illnesses.

3. Approach

3.1. MRI data processing

PI Dr. Klein is involved in the development of state-of-the-art software workflows for MRI data processing (article under preparation; https://github.com/INCF/BrainImagingPipelines) and is the main developer of the Mindboggle software package (http://mindboggle.info) that automates cortical labeling [15] (**Fig.1**), feature extraction [16], and shape analysis [17,18], (articles in preparation). For the label definitions used by Mindboggle, Dr. Klein helped to create the Mindboggle-101 dataset (http://mindboggle.info/data), the largest and most complete set of free, publicly accessible, manually labeled human brain images in the world, labeled with the new "Desikan-Killiany-Tourville" (DKT) cortical labeling protocol that improves the accuracy of labeling cortical areas [19]. Further, he constructed a single "Gaussian classifier atlas" from 40 of the manually labeled brains (DTK40 atlas) and found that use of this atlas led to higher automated labeling accuracy than combining the registration results from the same number of individual atlases [19]. We will use the MRI and fMRI workflows and the Mindboggle's region labeling (with the DKT40 atlas) to preprocess our structural and functional data for use by our concurrence topology software.

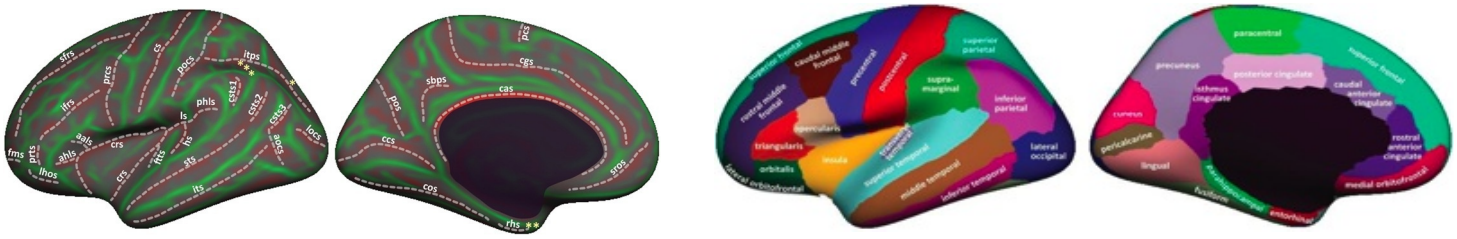

Fig.1. Brain labeling protocol (left, on an inflated cortex) and cortical labels (right) used by Mindboggle for extracting regions analyzed by our concurrence topology software.

3.2. Persistent homology

Concurrence topology describes high-order dependence in terms of homology classes. Selection of homology classes is based on their "persistence" [20]. Persistence assigns to homology classes times of "birth" and "death". The difference between birth and death is the "lifespan" of the homology class. The longer the lifespan of a class, the more likely to be "real", *i.e.*, reproducible in other data sets, rather than being merely the product of sampling fluctuations. We can create a "persistence plot" of homology classes of a given order of dependence for a data set by plotting death vs. birth for all the homology classes of that order in the series of shapes. In a persistence plot each point represents a persistent class and the distance from the point up to the diagonal $x=y$ is the lifespan of the class. **Fig.2** is an example that portrays third- and higher-order dependence.

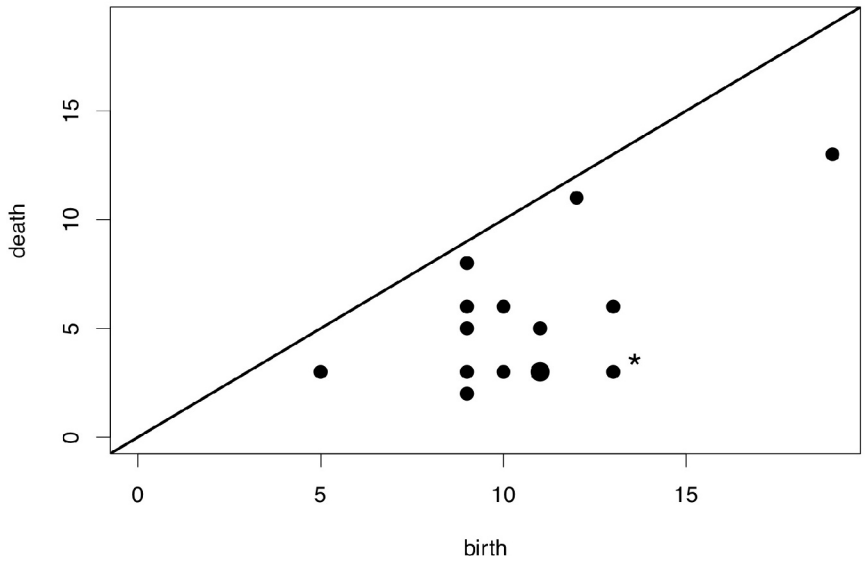

Fig.2. Persistence plot showing third- and higher-order dependence in the fMRI BOLD data from an individual control subject (as in **Aim 1**). The larger disk indicates two coinciding points. The point with the asterisk is discussed in **3.4.1**, Example 2.

3.3. Localization

Often researchers will want detailed information concerning high-order dependence in order to gain understanding of a disease process. Concurrence topology can tell more about a persistent homology class than just what is revealed in a persistence plot. A homology class involves all variables, but some variables are involved more directly than others. For purposes of interpretation (and for finding interactions, **3.4.2**) it is important to find the variables directly involved in persistent classes, at least for those with long lifespans. These can be found as members of "short cycles". A "cycle" is a structure in the data that wraps around the hole (or holes) represented by the homology class. A cycle belonging to a homology class representing k^{th}- and higher-order dependence can involve no fewer than than k variables. Those involving only k variables are "short cycles." Such short cycles are those most intimately associated with the corresponding hole and capture dependence of order exactly k. The process of finding short cycles is "localization". See **3.4.1** (Example 2) and **3.4.2** for applications of localization.

3.4. Preliminary results

3.4.1. fMRI data analyses

Most of our effort in using concurrence topology has been in applying it to resting state fMRI blood oxygenated-level dependent (BOLD) data [12]. The fMRI data set was generated at New York University and distributed as part of the 1000 Functional Connectomes project (http://fcon1000.projects.nitrc.org/). This data set includes 41 healthy controls and 25 adults diagnosed with ADHD. We computed BOLD values for 92 regions (whole brain), including 40 in the "default mode network" (DMN,[21]). We dropped regions that exhibited little variability, leaving 74 whole-brain and 32 DMN regions. fMRI BOLD data are multivariate time series, with one component per region.

We applied concurrence topology to each subject's data separately and computed summaries of the results. Then we used standard statistical methods to compare the subject-wise concurrence topology summaries between the two groups. Concurrence topology can be applied to multivariate time series either in the time or Fourier domains [22,12]. Our analyses were exploratory but using concurrence topology we found numerous statistically significant differences (not adjusted for multiple comparisons) between the groups.

Example 1. Fewer ADHD subjects (64.0%) had any holes corresponding to dependence of order six or higher in the time domain in the DMN than did controls (92.6%). Thus, the two groups differ in their pattern of 6th-order dependence. There was a less robust finding of the same sort in 7th-order dependence (the "7th-order dependence in 32 variables" referred to in **1.3**). We also found differences in the whole brain in the Fourier domain in dependence orders 3 and 4.

Example 2. The persistent class marked with an asterisk in **Fig.2** has a long lifespan. It contains 16 short cycles and more than half the subjects in the data set have at least one of them. Thus, most of the subjects share essentially the same homology class. These similar classes in the data must reflect some population homology class and so warrant further study. In fact, 76% of the ADHD subjects have at least one of the 16 short cycles, but only 44% of the controls have any. Thus, this class seems related to the ADHD condition. Being able to agnostically pinpoint specific triples of regions that discriminate the two groups would be difficult with any other statistical method.

3.4.2. Interactions in logistic regression for identifying suicide attempters

We used concurrence topology for interaction identification in Hamilton Depression Rating Scale (HDRS) data on 700 subjects randomly drawn from a clinical database in the M.I.N.D. division of the New York State Psychiatric Institute (NYSPI), where PI Dr. Ellis works. Our goal was to use the HDRS items (excluding suicide item) to discriminate subjects with and without a history of suicide attempt. We further split the 700 subjects into a "training" and "testing" sample. Using concurrence topology we found in the training sample persistent homology classes that distinguished the two groups. Examination of these led to one 3rd-order interaction and three 2nd-order interactions. We fitted lasso [9] and step-wise logistic regression models on the training sample including the dichotomized items as main effects and also the items plus the three interactions. The model that included the interactions we found using concurrence topology did statistically significantly better in the test sample in recognizing subjects who had a history of suicide attempt than did the models without the interactions.

3.5. Proposed work

3.5.1. Individual functional connectivity data (Aim 1)

We will first confirm and extend the our brain imaging findings in [12] in the much larger ADHD-200 data set (http://fcon_1000.projects.nitrc.org/indi/adhd200/), and second, apply our method to the Autism Brain Imaging Data Exchange (ABIDE) (http://fcon_1000.projects.nitrc.org/indi/abide/) data set. The large sizes of these data sets will allow a critical cross-validation test of the discriminatory power of the concurrence topology method, in particular for discriminating diagnostic subtypes (ADHD-combined vs. inattentive, and autism vs. Aspergers vs. pervasive developmental disorder). The ADHD-200 Sample website includes the description: "776 resting-state fMRI and anatomical datasets aggregated across 8 independent imaging sites, 491 of which were obtained from typically developing individuals and 285 in children and adolescents with ADHD (ages: 7-21 years old). Accompanying phenotypic information includes: diagnostic status, dimensional ADHD symptom measures, age, sex, intelligence quotient (IQ) and lifetime medication status." The ABIDE data include previously collected resting state functional magnetic resonance imaging data sets and phenotypic information from 539 individuals with ASD and 573 typical controls from 16 international sites.

We also have functional connectivity brain imaging data from patients with MDD to which we can apply our method to attempt to predict antidepressant treatment response. PI Dr. Klein is a Co-Investigator on the "Biosignature Discovery for Personalized Treatment of Depression" (1U01MH092250-01), a large, multi-site project acquiring multimodal imaging data from 400 individuals with MDD, specifically designed to make data available to find biomarkers for MDD.

3.5.2. Non-temporal group data (Aim 2)

We will further develop our method to first detect high-order dependencies in multiple sources of non-temporal, non-imaging data at the group level and then translate them into individual level interaction variables. These interaction variables can be included in classification modes such as logistic regression as described in **3.4.2**. We will apply the method to classify clinical subtypes of ADHD and of ASD by accompanying phenotypic information (symptom measures, lifetime medication status, behavioral measures, IQ scores, *etc.*) in the ADHD-200 and ABIDE data sets.

The M.I.N.D. division at NYSPI has large collections of clinical (over 1,500 subjects, over 4,000 variables), neuropsychological (over 500 subjects), genotype (almost 500 genotyped, over 500 with DNA gene chips), postmortem (autoradiograms of brains of over 250 subjects), and structural MRI (over 400) data pertaining to suicide and MDD. With the guidance of Co-Investigator Dr. Mann we will use clinical and/or biological criteria to select samples of subjects and up to 100 or so variables across several domains (as envisioned by RDoC) for concurrence topology analysis. The database manager (TBH) will then put together analytic data sets from the NYSPI databases.

3.5.3. Software development (Aim 3)

PI Dr. Ellis has written a package of programs in the statistical programming R language (http://www.r-project.org) for doing concurrence topology analysis. Currently, the amount of computer time needed by concurrence topology varies greatly from data set to data set and in extreme cases can require more than one week to analyze one data set. The probability of having a long computation increases the more variables there are and especially with order of dependence being analyzed. To increase the computational capabilities of our software, we propose to make the software run much faster. To do this, Dr. Ellis, assisted by the programmer (TBH), will rewrite the most computationally intensive portions of the software in a compiled language, such as C. Increasing computational speed will be one of the first undertakings of the project because of its importance for further development of the algorithms and their application to the large data sets described above. Some data structures recur many times in our software. In order to improve maintainability of the code, we will employ object-oriented programming [23] to represent these data structures as standardized objects. Dr. Ellis will rewrite the code in stages.

PI Dr. Klein has considerable programming experience, having developed the Mindboggle software described above (**3.1**), and will be involved in the software engineering challenges of this proposal, ensuring that the project adopts modern practices of test-driven development and distributed version control (hosted by http://github.com). In addition to working from the current code base, he has some experience with and will further explore the capabilities of the Dionysus software (http://www.mrzv.org/software/dionysus/), a C++ library with Python bindings for computing persistent homology developed by Dmitriy Morozov (see letter of support).

In addition to Dr. Morozov, Prof. Konstantin Mischaikow at Rutgers University, an expert on computational topology, will serve as an unpaid consultant for the project. (See letter of support.)

3.6. Timeline

Aim 1. We will spend the first year and a half seeing if our ADHD brain imaging findings will replicate in the much larger ADHD-200 data set with 776 subjects, and applying our method to find individual functional connectivity biomarkers in the ABIDE ASD data set with 1,110 subjects, as well as the MDD data set with 400 subjects. We will try to discriminate between different subtypes of the disease using concurrence topology, and evaluate using cross-validation.

Aim 2. We will run analyses for identifying non-time series group data biomarkers to diagnose subtypes of ADHD and ASD and predict treatment response in MDD and suicidal behavior (five months for each condition, including time to write, submit, and revise publications). In Year 2, we will include in the ADHD and ASD analyses subject-level summaries of high-order dependence computed under **Aim 1**.

Aim 3. We will spend the first six months making the concurrence topology software faster, and in the following year, we will restructure the code to follow an object-oriented framework that will help make the code base more concise and easier to maintain. We will then write a paper describing and publicizing the software. During this period, we will also determine which portions of the Dionysus computational homology software could be used to advance our concurrence topology software, and if so, extend the Dionysus code base to do concurrence topology in case we find it faster and more appropriate to develop with this code base.

References

[1] Thomas R Insel and Bruce N Cuthbert. Endophenotypes: bridging genomic complexity and disorder heterogeneity. *Biological psychiatry,* 66(11):988–9, December 2009. (PMID: 19900610)

[2] Centers for Disease Control and Prevention. *National Center for Injury Prevention and Control, Web-based Injury Statistics Query and Reporting System (WISQARS).* 2010; Available from: www.cdc.gov/injury/wisqars/index.html.

[3] Insel, T. The under-recognized public health crisis of suicide. 2010 [cited 2010 September 10]; Director's Blog]. Available from: http://www.nimh.nih.gov/about/director/2010/the-under-recognized-public-health-crisis-of-suicide.shtml.

[4] Martijn P van den Heuvel and Hilleke E Hulshoff Pol, Exploring the brain network: A review on resting-state fMRI functional connectivity, *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology* 20 (2010), no. 8, 519–534 (PMID: 20471808)

[5] Klass Enno Stephan and Karl J. Friston, Analyzing effective connectivity with functional magnetic resonance imaging, *WIREs Cogn Sci,* 1:446--459, doi = {10.1002/wcs.58}, 2010 (PMID: 21209846).

[6] P. Jezzard, P.M. Matthews, and S.M. Smith, *Functional MRI: An Introduction to Methods,* Oxford University Press, Oxford, 2002.

[7] Aapo Hyva ¨rinen, Juha Karhunen, and Erkki Oja, *Independent Component Analysis,* Wiley, New York, 2001.

[8] Donald S. Burdick, An introduction to tensor products with applications to multiway data analysis, *Chemometrics and Intelligent Laboratory Systems* 28 (1995), 229–237.

[9] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction,* Springer, New York, 2001.

[10] David Bartholomew, Martin Knott, and Irini Moustaki, *Latent Variable Models and Factor Analysis: A Unified Approach,* Wiley, Chichester, 2011.

[11] David B. Dunson and Chuanhua Xing, Nonparametric Bayes modeling of multivariate categorical data, *J. Amer. Statist. Assoc.* 104 (2009), 1042–1051.

[12] Steven P. Ellis and Arno Klein, Describing high-order statistical dependence using "concurrence topology", with application to functional MRI brain data, (under review by a statistics journal and posted on arXiv, http://*arxiv.org/*abs/1212.1642), 2012.

[13] J. R. Munkres, *Elements of Algebraic Topology,* Benjamin/Cummings, Menlo Park, CA, 1984, Reprinted by Perseus Publishing, Cambridge.

[14] Carina Curto and Vladimir Itskov. Cell Groups Reveal Structure of Stimulus Space. *PLoS Computational Biology,* 4(10):e1000205, October 2008 (PMID: 18974826)

[15] Mindboggle -- A neuroinformatics framework for automated anatomical labeling and shape analysis of brain images Klein, A, Lee, N, Bao, F, Häme, F. 2011. *Society for Neuroscience 2011 Conference.*

[16] Automated extraction of nested sulcal features from human brain MRI data Klein, A, Bao, F, Lee, N, Im, K, Rivière, D. 2011. *Human Brain Mapping 2011 Conference.*

[17] Mindboggle -- An informatics framework for open research in quantifying the shape of the human brain Klein, A, Lee, N, Bao, F, Häme, F. 2011. *BioImage Informatics II* 2011 Conference at Janelia Farm.

[18] A graph-based database of hierarchical brain features Lee, N, Klein, A. 2011. *Neuroinformatics* 2011 Conference.

[19] 101 labeled brain images and a consistent human cortical labeling protocol, Klein, A, Tourville, J. 2012 *Frontiers in Brain Imaging Methods.* 6:171. DOI: 10.3389/fnins.2012.00171.

[20] Herbert Edelsbrunner and John L. Harer, *Computational Topology: An Introduction,* American Mathematical Society, Providence, 2010.

[21] Uddin, L.Q., Kelly, A.C., Biswal, B.B., Castellanos, F.X., and Milham, M.~P. (2009). Functional connectivity of default mode network components: Correlation, anticorrelation, and causality. *Human Brain Mapping,* 30:625 -- 637 (PMID: 18219617)

[22] David R. Brillinger, *Time Series: Data Analysis and Theory,* Society for Industrial and Applied Mathematics, Classics In Applied Mathematics, Philadelphia, 2001.

[23] H. Abelson, J. Sussman, and J. Sussman *Structure and Interpretation of Computer Programs,* MIT Press; ISBN 0-262-01077-1, 1984.