Novel Biomarker Identification Approaches for Schizophrenia using fMRI and Retinal Electrophysiology

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE UNIVERSITY OF MINNESOTA

BY

PANTEA MOGHIMI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

KELVIN O. LIM

THEODEN I. NETOFF

NOVEMBER 2017

Acknowledgements

Abstract

Schizophrenia is a chronic mental illness. The exact cause if schizophrenia is not yet known. Extensive research has been done to identify robust biomarkers for the disease using non-invasive brain imaging techniques. A robust biomarker can be informative about pathophysiology of the disease and can guide clinicians into developing more effective interventions. The aim of this dissertation is two folds. First, we seek to identify robust biomarkers using resting state fMRI activity from a cohort of schizophrenic and healthy subjects in a purely data driven approach. We will calculate multivariate network measures and use them as features for classification of the subjects into healthy and diseased. The network measures will be calculated using nodes defined by the AAL anatomical atlas as well as a functional atlas constructed from the fMRI activity. Network measures with high classification rate may be used as potential biomarkers. We will employ double cross-validation to estimate generalizability of our results to a new population of subjects that were not used in biomarker identification. Second, we seek to identify biomarkers using electroretinogram (ERG). We will use a data driven approach to classify individuals based on the pattern of retinal activity they exhibit in response to visual stimulation. Characteristics of the ERG result in high classification rate are presented as potential biomarkers of schizophrenia.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Schizophrenia is a chronic debilitating mental disease that affects 1.1% of adult US population ("Schizophrenia" 2016a) and more than 21 million people worldwide ("Schizophrenia" 2016b). Typical onset age of schizophrenia is between the ages of 15 and 25 years. Schizophrenia is characterized by hallucination, delusion, as well as abnormalities in attention, working memory, and social and emotional behavior. The main risk factor for development of schizophrenia is genetic (Cardno et al. 1999), although several environmental factors have also been reported (McDonald and Murray 2000).

Systematically, symptoms of schizophrenia are grouped into three categories: positive symptoms, negative symptoms, and cognitive symptoms.

Positive symptoms are feelings or behaviors that are typically not present in healthy individuals and include four symptom groups (Andreasen et al. 1995):

1. Hallucinations in any sensory modality, e.g. hearing voices or smelling odors that are not present.
2. Delusions, which includes a broad range of beliefs, such as the delusion that patient's thoughts are broadcasted to everybody or that other people are inserting thoughts into patient's mind.
3. Bizarre behavior, which captures abnormal behavioral traits such as unusual appearance or aggressive behavior.
4. Positive formal thought disorder, which characterizes disorganized thinking as manifested in incoherent speech.

Negative symptoms are typical feelings and behaviors that are present in healthy individuals but absent in schizophrenic patients. Negative symptoms consist of five groups (Andreasen et al. 1995).

1. Affect flattening or blunting and refers to a lack of expression of emotions, e.g. absence of changes in facial expression.
2. Alogia, or poverty of speech, which refers to lack of content in produced speech.

3. Avolition-apathy, refers to lack of motivation or emotion which manifests itself in behaviors such as lack of hygiene or lack of persistence at work.
4. Anhedonia-asociality and is lack of interest in recreational or social activities.
5. Attention which characterizes social attentiveness and attention during a mental status task.

Positive and negative symptoms are quantified using two rating scales: Scale for the Assessment of Positive Symptoms (SAPS) (N. Andreasen 1984) and Scale for the Assessment of Negative Symptoms (SANS) (N. Andreasen 1983) which measure severity of positive and negative symptoms respectively.

Cognitive symptoms of schizophrenia encompass a wide range of cognitive inabilities in schizophrenic patients such as disruption in working memory (Goldman-Rakic 1994), declarative memory (Cirillo and Seidman 2003), attention (Cohen and Servan-Schreiber 1992), language deficits (Cohen and Servan-Schreiber 1992), and context processing (Cohen and Servan-Schreiber 1992; MacDonald 2008).

The most common treatment for schizophrenia currently is antipsychotic medication (Chien and Yip 2013). In recent years psychosocial intervention and therapy based methods have gained popularity (Chien et al. 2013). In cases with severe symptoms, surgical lobotomy might also be performed (Soares et al. 2013).

Exact pathophysiology of schizophrenia is not yet known, but has been explored using several imaging modalities. Several studies have compared morphology and size of anatomical areas in schizophrenic patients with that of healthy control individuals using Computerized Tomography (CT) and Magnetic Resonance Imaging (MRI) and changes in volume and morphology have been reported (Shenton et al. 2001). Structure of white matter fiber tracts, visualized using an imaging technique called Diffusion Tensor Imaging (DTI), has also been reported to be affected by schizophrenia (Foong et al. 2000; Minami et al. 2003; Ardekani et al. 2003; Davis et al. 2003; Camchong et al. 2011). Patterns of functional activity, measured with Blood Oxygen Level Dependent (BOLD) signal using Functional Magnetic Resonance Imaging (fMRI), have also been reported to be different in schizophrenic patients (Skudlarski et al. 2010). Collectively, studies using DTI and fMRI support the dysconnectivity hypothesis (Friston 1998) of schizophrenia. This hypothesis posits that symptoms of schizophrenia are caused by

failure of functional integration across the brain. More specifically, healthy pattern of interaction between different functional modules of the brain is disrupted in the disease population.

The mechanism by which these connectivity patterns are disrupted is not yet clear. Historically, the dominant hypothesis was that abnormal activation pattern of dopaminergic pathways in the brain is the cause of schizophrenia (Howes and Kapur 2009). This hypothesis, known as the dopamine hypothesis, was mainly based on the observation that antipsychotic drugs such as chlorpromazine were dopamine antagonists. However, limited progress in development of effective therapeutic interventions based on the dopamine hypothesis, led to development of the alternative N-methyl-D-aspartate receptor (NMDA receptor) hypofunction hypothesis (Snyder and Gao 2013). The NMDA receptor is a glutamatergic receptor playing a crucial rule in synaptic plasticity during early neurodevelopmental stages. The NMDA hypofunction hypothesis attributes schizophrenia to dysfunction of the NMDA receptors. Interestingly, activity of dopaminergic neurons are also regulated by NMDA receptors, which can explain abnormality in activity of dopaminergic neurons in schizophrenia. Given NMDA receptors' role in synaptic plasticity, the NMDA hypothesis is in agreement with the dysconnectivity hypothesis. More specifically, failure in integration of information across different functional modules in the brain can be explained by disruption in formation of neural circuitry caused by dysfunction of NMDA receptors.

Diagnosis of schizophrenia is currently done in a clinical setting and based on behavioral symptoms reported by patients, and not based on any measurable biomarkers. Identification of robust biomarkers, i.e. biomarkers with high replication rate across studies and subject populations, is important for several reasons. The first reason is that robust biomarkers can shed light on the physiological mechanism of the psychosis. Second, robust biomarkers can be used to develop novel more effective therapeutic interventions for schizophrenia. Moreover, robust biomarkers can potentially be used to identify the disease in its prodromal state for early intervention in the population with a genetic predisposition to develop schizophrenia. Third, robust biomarkers may be used to predict responsiveness of patients to interventions, which can be further utilized to develop more individualized treatment regimens for each patient.

Currently, there is no generally accepted biomarker for schizophrenia. Extensive attempts to identify patterns of disrupted connectivity using fMRI datasets have been made (Pettersson-Yeo et al. 2011; Orrù et al. 2012). However, different studies have implicated different areas of the brain. The general framework in fMRI biomarker identification studies is to use regions delineated by an anatomical atlas as functional modules. Activity of single regions, or pairwise similarity between activity of these regions are then compared across schizophrenic and healthy individuals. The premise is that disruption in integration of functional activity will result in abnormal activation or coactivation pattern between these regions. Anatomical atlases divide the brain into its major gyri based on anatomical landmarks and might not align with functional divisions of the brain. Therefore, using anatomical atlases to define functional modules might result in a reduction in discriminative power of resultant biomarkers. Furthermore, disrupted connectivity patterns might follow complex motifs that are not picked up by examining activity of single regions or pairwise coactivation patterns between them.

This dissertation is broken into three chapters that aims to identify more robust biomarkers for schizophrenia using a data driven machine learning approach.

Chapter 2 focuses on the need to identify functional divisions of the brain more accurately. In this chapter a parcellation of the brain that is based on functional activity of voxels is constructed and evaluated. A data driven method will be used to group voxels based on their activation pattern captured by fMRI. **We hypothesize that a functional parcellation captures functional divisions of the brain more precisely than an anatomical atlas.**

Chapter 3 will focus on identifying biomarkers using fMRI. The functional parcellation as well as an anatomical atlas will be used to generate connectivity maps from which network features will be extracted. Graph theoretic network measures were calculated to generate features characterizing the connectivity patterns. Using these measure, subjects were classified as healthy or schizophrenic in a data driven framework using machine learning techniques. **Our hypothesis is that graph theoretic measures, calculated using a functional parcellation of the brain, will result in more robust biomarkers**.

Chapter 4 focuses on the dysfunctionality of NMDA receptors in schizophrenic patients and whether it can be used to identify biomarkers. NMDA receptors are expressed in retinal cells and disruption in their functionality affects activation pattern of retinal cells. Retinal activity from a group of healthy and schizophrenic patients was used to classify each individual using machine learning techniques. **We hypothesize that differences in activity of retinal cells in healthy and schizophrenic individuals can be used as a robust biomarker for schizophrenia.**

Chapter 2: Construction and Evaluation of Parcellation Methods of the Brain from fMRI data

Introduction

Brain atlases are a ubiquitous tool used for analysis of functional magnetic resonance imaging (fMRI) datasets. Atlases group all the voxels into several dozens of contiguous regions. One of the main advantages of using atlases in analysis of fMRI datasets is that average of time series of voxels within each region are calculated and used instead of time series of each single voxel. This approach increases signal to noise ratio and reduces dimensionality of the dataset from thousand of voxels to dozens of regions.

Historically, anatomical atlases have been used for analysis of fMRI datasets. Anatomical atlases divide the brain into its major gyri using anatomical landmarks identified manually. Several anatomical atlases such the Automated Anatomical Labeling atlas (AAL) (Tzourio-Mazoyer et al. 2002), Desikan-Killiany atlas (Desikan et al. 2006), and Destrieux atlas (Fischl et al. 2004; Destrieux et al. 2010) have been produced and are commonly used.

However, anatomical divisions of the brain may not reflect its functional organization confounding two neighboring areas that have different functional purposes into a single area. Network analysis of functional activity may not be as accurate or as sensitive as analysis performed on functional atlases. Therefore, in recent years several attempts have been made to produce parcellations of the brain that are based on functional activity of voxels, rather than their spatial location (e.g. (Cohen et al. 2008; Power et al. 2011; Yeo et al. 2011; Craddock et al. 2012; X. Shen et al. 2013; Thirion et al. 2014; Honnorat et al. 2015)). These studies use data driven unsupervised methods, known as clustering, to group voxels with similar functional activity to form regions. Functional activity is typically collected in resting state, when subjects are asked to relax in the scanner and are not required to perform any tasks, although one study has used task evoked activity for parcellation (Thirion et al. 2014). The resultant parcellation, known as a functional brain atlas, can be used in place of anatomical atlases.

6

Functional parcellation methods have several limitations. Some methods do not produce contiguous regions (e.g. (Power et al. 2011; Yeo et al. 2011; X. Shen et al. 2013; Thirion et al. 2014) resulting in brain regions that are scattered across the brain which complicates interpretation of any results obtained from the atlas. Some methods, such as the K-means or spectral clustering algorithms, are biased towards regions of equal size (e.g. (Craddock et al. 2012; Thirion et al. 2014; Honnorat et al. 2015)). Some methods only parcelate the cortical surface neglecting subcortical structures (Cohen et al. 2008).

A test of a parcellation is to measure the similarity of activity from voxels within the same region to similarity between voxels from different regions. Similarity between voxels is typically calculated as the pairwise cross correlation between the BOLD signal of the voxels (Craddock et al. 2012; X. Shen et al. 2013; Thirion et al. 2014; Honnorat et al. 2015). However, the presence of strong autocorrelations within each time series can cause spuriously high correlation values (Christova et al. 2011). Removal of autocorrelation, a process known as prewhitening, may therefore change parcellation results.

To address these gaps, we used the hierarchical clustering algorithm to parcellate the brain into contiguous regions, using 6 minute resting state fMRI scans. Hierarchical clustering is not biased towards regions of equal size and can accommodate heterogeneity in size of the regions. We added a spatial constraint to enforce region contiguity. To explore the role of prewhitening, we constructed two functional atlases, one using raw time series, and another using pre-whitened time series. We evaluated the resultant atlases in terms of homogeneity of the regions, separation between regions, and reproducibility of results. We explored effect of scan duration on parcellation results. We also compared functional atlases constructed using combined datasets from our group of subjects to atlases constructed individual datasets.

Methods

Participants

A group of 88 (27 female, age: M = 33.4, SD = 11.9) subjects with no neurological disorders participated in this study. All participants gave informed consent and were compensated for their participation. All procedures were done in accordance with a University of Minnesota IRB approved protocol.

Image Acquisition

Each subject underwent a six minute resting state fMRI image acquisition, during which subjects were instructed to stay still and awake and keep their eyes closed for about 6 minutes. Images were acquired using a Siemens Trio 3T scanner (Erlangen, Germany) with the following sequence parameters: gradient-echo echo-planar imaging (EPI) 180 volumes, repetition time (TR) 2 seconds, echo time (TE) 30ms, flip angle 90°, 34 contiguous AC-PC aligned axial slices, voxel size 3.4 x 3.4 x4.0 mm, matrix 64 x 64 x 34 totalling 139,264 voxels.

In addition to functional activity, a T1-weighted anatomical image was acquired using a magnetization prepared rapid gradient-echo sequence. A field map was also acquired and used to correct for geometric distortions introduced by field inhomogeneities: TR = 300ms, TE = 1.91 ms/4.37 ms, flip angle = 55°, voxels size = 3.4 x 3.4 x 4.0 mm (Camchong et al. 2011; Atluri et al. 2015).

The raw fMRI data was preprocessed using FEAT and MELODIC from the FSL software package as follows. The first three volumes were removed from each subject scan to account for magnetization stabilization. This resulted in a 5.9 minute time series per voxel (177 time points). Each scan was motion corrected, B0 field map unwarped, and corrected for slice scan time. Non-brain portions of the images were removed and a spatial smoothing kernel was applied to the dataset (6mm full-width half-maximum). The images were then grand mean and intensity normalized and temporally filtered between 0.01 and 0.08Hz. All images were then registered to the MNI152 space. Using probabilistic independent component analysis (PICA), noise introduced by head motion, respiration, cardiac pulsation, and scanner artifacts was removed. Spatial and temporal

characteristics of noise components are described in MELODIC manual (https://fsl.fmrib.ox.ac.uk/fslcourse/lectures/melodic.pdf). The dataset was then resampled to 3 x 3 x 3mm, resulting in 47640 voxels.

Prewhitening

Prewhitening refers to removal of autocorrelation from a given time series so that similar to white noise, the resultant time series are decorrelated. Presence of autocorrelation in BOLD time series can lead to spurious high cross correlation values between different voxels (P. Christova et al. 2011). We prewhitened the time series from voxel i, $x_i(t)$ by calculating its Fourier transform $X_i(f)$ and dividing it by its power spectrum $|X_i(f)|$, to result in a flat power spectrum, similar to white noise. The resultant signal $X_i^W(f)$ was then inverse Fourier transformed into the time domain $x_i^W(t)$ (Equation 2.1).

$$
\begin{aligned}
x_i(t) &\leftrightarrow X_i(f) \\
X_i^W(f) &= X_i(f)/|X_i(f)| \\
x_i^W(t) &\leftrightarrow X_i^W(f)
\end{aligned}
\qquad\qquad \textit{Equation (2.1)}
$$

Parcellation

To parcellate the brain using fMRI data, voxels with similar time series are grouped together to form regions. This is typically done using data-driven clustering algorithms, where each cluster constitutes one region (e.g. (Craddock et al. 2012; Thirion et al. 2014; Honnorat et al. 2015)).

We chose the agglomerative hierarchical clustering algorithm with Ward's minimum variance as linkage criterion (Ward 1963; Tan, Steinbach, and Kumar 2006). Hierarchical clustering algorithm is not biased towards clusters of equal size like the K-means or spectral clustering algorithms (Tan, Steinbach, and Kumar 2006) and results in more reproducible parcellations (Thirion et al. 2014). The agglomerative hierarchical clustering algorithm starts with each datapoint (voxel) as a single cluster. It then merges the cluster pair that minimizes Ward's criterion to form a new cluster. Ward's criterion calculates total within-cluster variance resulting from merging each pair of clusters. The merging

process is iterated until all clusters are merged to form a single cluster containing all data points. Information about membership of each datapoint to each cluster at each stage of merging is stored in a structure called a dendrogram. Different parcellation scales, i.e. number of regions the brain is parcellated into, are constructed by cutting the dendrogram at the stage that contain the desired number of clusters. The resultant parcellations are then used as functional brain atlases. To obtain atlases with contiguous regions, we applied a spatial constraint so that only spatially adjacent clusters can be merged.

A clustering algorithm requires a distance measure between voxel pairs. We used the correlation distance for parcellation. Correlation distance between voxels i and j, $d_{i,j}$, is equal to $d_{i,j} = 1 - r_{i,j}$ where $r_{i,j}$ is the zero-lag crosscorrelation between the two voxels.

We constructed two types of functional atlas: i) atlas constructed using the original time series, referred to as <u>original functional atlas</u> in this manuscript; ii) atlas constructed using prewhitened time series, referred to as <u>white functional atlas</u> in this manuscript.

We constructed functional atlases at two different levels: i) Group level, where time series from the entire group of subjects were combined to construct a group dataset, which was then used for construction of the functional atlas. To combine individual datasets, for each voxel, time series from all subjects were concatenated to construct a single time series. ii)Individual level, where the functional atlas was constructed for each individual dataset. In this manuscript, unless stated otherwise, functional atlas refers to a group level functional atlas.

Evaluation

To evaluate the resultant functional atlases we employed three approaches: i) calculating homogeneity of the regions, which measures how similar voxels within a single region are to each other; ii) calculating separation between regions, which measures how dissimilar voxels in different regions are with respect to voxels within regions; iii) calculating reproducibility of parcellation results, which quantifies how reproducible the results are if a different group of subjects, ot different datasets from the same subject were used for parcellation. Each approach is explained here in detail.

Homogeneity

Several measures of homogeneity have been used to evaluate parcellation methods. We used the following 5 measures to quantify homogeneity of the resultant atlases:

1. Average pairwise correlation coefficient between voxels within each region (Craddock et al. 2012; Honnorat et al. 2015), referred to as *rt* in this manuscript.
2. Average pairwise correlation coefficient between functional connectivity maps between voxels within each region (Craddock et al. 2012), referred to as *rs* in this manuscript.
3. Percentage of variance explained by the first principal component (Bishop 2006) of time series of voxels within each region (Gordon et al. 2014), referred to as *pcat* in this manuscript.
4. Percentage of variance explained by the first principal component of functional connectivity maps of voxels within each region, referred to as *pcs* in this manuscript.
5. Kendall's coefficient of concordance (Kendall Maurice; Gibbons 1990) between voxels within each region (Shi et al. 2007), referred to as *KCC* in this manuscript.

Due to the spatial autocorrelation present in fMRI datasets, a contiguous random grouping of voxels is bound to result in regions with a certain degree of homogeneity. Therefore, to test our null hypothesis we compared distribution of homogeneity of the functional atlases to homogeneity of randomly constructed atlases with contiguous regions and similar size distributions. Homogeneity depends on size of regions. Smaller regions contain fewer voxels which results in less diversity among the voxels. The extreme case is a region that consists of a single voxel, which is perfectly homogeneous. Therefore, the random atlases must match the functional atlases in distribution of region sizes. We constructed random atlases that consisted of spatially contiguous regions with similar region size distribution to functional atlases, but assignment of voxels to regions was performed randomly. To construct a random atlas with M regions, we randomly picked M initial voxels as seeds, with each seed constituting a single region. Pairwise Euclidean distance between each of the voxels and the seed voxels was calculated and each voxel was assigned to the region with the closest seed. Since this algorithm does not guarantee the distribution of region sizes to match that of our atlases, 1000 random

atlases were generated and the mismatch of their size distribution to the functional atlases was calculated. Then the 10 random atlases with lowest mismatch were used as the final random atlases. Average mismatch for the 10 chosen atlases was less than 5%.

Cluster Separation

To quantify separation between regions, we used the Silhouette coefficient (Rousseeuw 1987) which has been used in several studies to evaluate parcellation algorithms (Kelly et al. 2010; Yeo et al. 2011; Craddock et al. 2012; Long et al. 2014; Parisot et al. 2016). The silhouette coefficient measures how similar each voxel is to voxels within its region compared to voxels in other regions. To calculate Silhouette coefficient for voxel i, first, average correlation distance, a_i, between voxel i and all other voxels assigned to the same region. Then, the lowest average correlation distance between that voxel and all other regions, b_i is calculated, where average distance between the voxel and each region is average distance between that voxel and all the voxel belonging to that region. The silhouette coefficient for voxel i is then calculated as $S_i = (b_i - a_i)/max(b_i, a_i)$. Silhouette coefficient takes up values between 1 and -1, where a value of 1 indicates that the region the voxel belongs to is well separated from other regions.

Similar to homogeneity values, Silhouette coefficient values from the functional atlases were compared to values from randomly constructed atlases.

Reproducibility

To calculate how reproducible the results of functional parcellation are across different groups, we divided our subjects into two groups, and constructed separate functional atlases from each group's raw dataset. We then compared the agreement between the two atlases. This comparison was done at four different parcellation scales, 90, 500, 1000, and 4000 regions. We quantified the agreement using the Adjusted Rand Index (ARI), a measure of comparing different groupings of the same dataset (Rand 1971; Vinh, Epps, and Bailey 2010). Rand index (RI) is a normalized measure that calculates agreement between two parcellations (Equation 2.2). ARI is a corrected form of RI that subtracts expected RI values that are to be observed due to chance. ARI can take up

values between 1 (total agreement between the two parcellations) and -1 (total disagreement between the two parcellations).

---

a: Total number of voxels pairs that are assigned to the same region in both parcellations

b: Total number of voxels pairs that are assigned to different regions in both parcellations

c: Total number of voxels pairs that are assigned to the same region in parcellation 1 and to different regions in parcellation 2

d: Total number of voxels pairs that are assigned to different regions in parcellation 1 and to the same region in parcellation 2

$$RI = \frac{a+b}{a+b+c+d}$$ *Equation 2.2*

Scan duration

To evaluate effect of scan duration on homogeneity and reproducibility of the functional atlas, we constructed functional atlases using a range of scan durations. A subset of our subjects (N=24), were scanned for a second and third time in six and nine months after the first scan. Each scan session lasted 5.9 minutes. Raw time series for each voxel were concatenated for each subject to construct a 17.7 minutes long time series per voxel.

To examine reproducibility of parcellation results as a function of scan duration, we divided the long time series into two halves. 360s and 600s from each half were taken and used to construct functional atlases with 90, 500, and 2000 regions.

To quantify the effect of scan duration on regional homogeneity, the 17.7 minute long time series were truncated at different time points, and truncated time series were used separately to construct functional brain atlases at two different scales, 90 and 500 regions. Time series were truncated after the first 360s (6 minutes), 600s (10 minutes), 840s (14 minutes) and 1062s (17.7 minutes). Homogeneity of the regions of each functional atlas was then calculated.

Level of Analysis

We constructed individual level functional atlases from raw time series of 7 of our subjects by parcellating each dataset separately. We quantified degree of agreement between each individual level atlas and our group level atlas using ARI at three different parcellation scales, 90, 500 and 2000 regions. We also calculated ARI between each pair of individual level atlases at those parcellation scales.

Results

We constructed two functional atlases using the agglomerative hierarchical clustering with the linkage method for merging criterion. The first functional atlas, referred to as the original functional atlas, was constructed using raw time series. The second functional atlas, referred to as the white functional atlas, was constructed using pre-whitened time series. Pre-whitening removes autocorrelation from each time series, resulting in elimination of spurious cross correlation between voxels (Christova et al. 2011). We observed that average pairwise cross correlation between all voxels was reduced after pre-whitening (Supplementary Figure 2.1).

To compare properties of the functional atlases against anatomical atlas, we constructed each functional atlas with 90 regions to compare it to a commonly used anatomical atlas, known as the AAL, which also consists of 90 regions. Schematics of the resultant functional atlases along with the AAL anatomical atlas (Tzourio-Mazoyer et al. 2002) are shown in Figure (2.1A, 2.1B, and 2.1C). We also tested other linkage criteria. However, the other linkage criteria resulted in atlases with several regions that constituted of single voxels and regions that were extremely large, encompassing entire lobes. Distribution of region sizes was most comparable to that of an anatomical atlas only when the linkage criterion was used (Supplementary Figure 2.2).

Figure 2.1D shows distribution of region sizes for the three atlases. As can be seen in Figure 2.1D left, the two functional atlases have similar size distributions compared to the AAL atlas. However the AAL atlas seems to have several regions with smaller sizes than the functional atlases. We also constructed ten random parcellations that matched each functional atlas in size distribution. Properties of random atlases were compared to

that of the functional atlases. As can be seen in Figure 2.1D, middle and right, these random parcellations match the functional atlases very closely.

We then calculated homogeneity of regions for each of the three atlases. Homogeneity is a measure of similarity between voxels within each region. Homogeneity was calculated as average pairwise cross correlation between voxels of each region (rt). To compare the original functional atlas with the AAL atlas, both atlases were applied to raw time series to calculate homogeneity. The two distributions were significantly different (two sample Kolmogorov-Smirnov test, p<0.001). To compare the white functional atlas with the AAL atlas, both atlases were applied to prewhitened time series to calculate homogeneity. The two distributions were significantly different (two sample Kolmogorov-Smirnov test, p<0.001). The resultant distributions are shown in Figure 2.1E, left. Due to presence of spatial autocorrelation in fMRI datasets, adjacent voxels are bound to have a certain degree of correlation. Therefore, even in random groupings of parcellations a certain degree of homogeneity will be observed. To confirm that our functional parcellation algorithm is having a superior performance to random parcellations, we compared distribution of homogeneity values for each functional atlas with that of ten random parcellations that matched the functional atlases in size distribution. The resultant distributions for the original functional atlas, and the white functional atlas are shown in Figure 2.1E, middle and Figure 2.1E, right. We found that the homogeneity distribution of each functional atlas with each of the ten random parcellations was significantly different (Kolmogorov-Smirnov test, Corrected for multiple comparisons, p<0.001). Therefore functional atlases result in more homogenous regions than the anatomical atlas and random parcellations. We also calculated homogeneity using several other homogeneity measures (see methods). Distribution of the homogeneity values for each measure is shown in Supplementary Figure 2.3. We observed that functional atlases compared to both AAL and random parcellations produce significantly more homogenous regions when rs and $pcat$ (see methods) are used as measures of homogeneity. However, when $pcas$ was used, we did not observe a significant difference between functional atlases and the anatomical atlas or random parcellations. When KCC was used as the measure of homogeneity, we observed a significant different between the functional atlases and the anatomical atlas, but no significant difference between the functional atlases and random parcellations.

Next, we looked at separation between regions. Separation between regions quantifies how dissimilar voxels in separate regions are compared to voxels within the same region. We quantified separation between regions using the Silhouette coefficient. Distribution of Silhouette values for both functional atlases, as well as the AAL atlas applied to both raw and pre-whitened time series is shown in Figure 2.1F, left. We observed no significant difference between values of the AAL atlas applied to raw time series compared to the original functional atlas (Kolmogorov-Smirnov test, p = 0.78). Similarly, no significant difference between values of the AAL atlas applied to pre-whitened time series and the white functional atlas was observed (Kolmogorov-Smirnov test, p = 0.08). When Silhouette values were compared across the original functional atlas and each random parcellation (Figure 2.1F, middle), only two out of ten distributions were significantly different (Kolmogorov-Smirnov test, corrected for multiple comparisons, p<0.005). Similarly, only five out of ten distributions were significantly different when comparing white atlas to each random parcellation (Kolmogorov-Smirnov test, corrected for multiple comparisons, p<0.005, Figure 2.1F, right).

Subsequently, we examined the degree of agreement between each pair of atlases using the adjusted rand Index (ARI) (see methods). The results are shown in Table 2.1. As can be seen, the functional atlases calculated on raw and pre-whitened data are in 89% agreement between each other. However, the AAL atlas and each of the functional atlases are in much lower agreement (about 25%).

Pair of Atlases	ARI
Original- vs. white functional atlas	0.89
AAL vs. original functional atlas	0.25
AAL. vs. white functional atlas	0.26

Table 2.1. Degree of Agreement between each pair of Atlases. Degree of agreement between each pair of atlases is quantified using the adjusted rand index (ARI).

Figure 2.1. Parcellation Properties

(A)   A schematic of the anatomical AAL atlas, from left to right: coronal, sagittal and horizontal views

(B)   A schematic of the original functional atlas, constructed using raw time series, from left to right: coronal, sagittal and horizontal views.

(C)   A schematic of the white functional atlas, constructed using prewhitened time series, from left to right: coronal, sagittal and horizontal views.

(D)   Size distributions. Regions were rank ordered based on their size. X-axis: Rank of each region, Y-axis: Number of voxels within the region in log scale.Left: Distribution of size of regions for the AAL, original functional, and white functional atlases. Middle: Distribution of size of regions for the original functional atlas constructed using raw time series as well as ten size matched random parcellations. Right: Distribution of size of regions for the white functional atlas constructed using prewhitened time series as well as ten size matched random parcellations.

(E)   Distribution of homogeneity values for each atlas. Homogeneity was measured as average pairwise correlation coefficients between voxels in each region (rt) across all regions and all subjects. Left: Distribution of homogeneity values for the AAL atlas applied to both raw and prewhitened time series, as well as the original functional atlas constructed using raw time series applied to raw time series and white functional atlas constructed using prewhitened time series to prewhitened time series. Middle: Distribution of homogeneity values for the original functional atlas constructed using raw time series and ten size matched random parcellations. Right: Distribution of homogeneity values for the white functional atlas constructed using prewhitened time series and ten size matched random parcellations.

(F)   Distribution of Silhouette Coefficients for each atlas across subjects. For each

subject, silhouette values across all regions were averaged. Left: Distribution of Silhouette values for the AAL atlas applied to both raw and prewhitened time series, as well as the original functional atlas constructed using raw time series applied to raw time series and white functional atlas constructed using prewhitened time series to prewhitened time series. Middle: Distribution of Silhouette values for original functional atlas constructed using raw time series and ten size matched random parcellations. Right: Distribution of Silhouette values for the white functional atlas constructed using prewhitened time series and ten size matched random parcellations.

We also examined reproducibility of our results across subjects by dividing our subject set into two groups of equal size and constructed a functional atlas using the raw time series from each group separately. We then calculated degree of agreement between the two resultant atlases at several parcellation scales, using ARI. The results are shown in Figure 2.2. As can be seen the degree of agreement between the two atlases is maximum at 90 regions and minimum at 1000 regions and then increases as parcellate into 4000 regions. However, the range in ARI is only 3% (from 33% at 90 regions to 30% at 500 regions) and these differences are probably negligible. Our conclusion from these findings is that reproducibility is not dependent on parcellation scale and that functional parcellation only moderately generalizes to new subjects.

We then examined how reproducibility changes as a function of scan duration. For a subset of our subjects, we had a second and third set of fMRI scans taken 6 and 9 months after the first scan, respectively. Each scan lasted 5.9 minutes. We constructed longer time series for each voxel by concatenating the time series from each scan, resulting in 1062s (17.7 minutes) of data. We then divided the long time series into two equal halves. We constructed several functional atlases by taking epochs of different durations from each half. Degree of agreement between the functional atlases constructed using different halves was quantified using ARI. The results are shown in Figure 2.3. As can be seen, increasing scan duration, increases reproducibility of results across datasets. This increase is more drastic at a parcellation scale of 90 regions (7% increase) compared to 500 and 2000 regions (~2% increase).
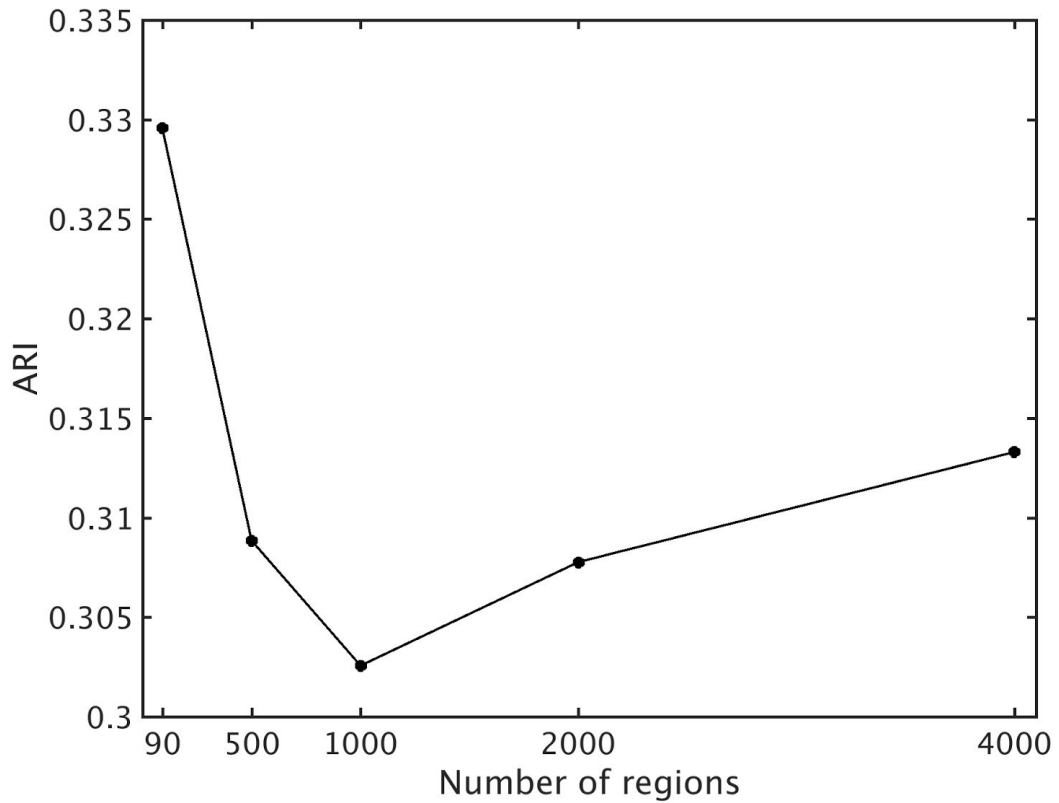
Figure 2.2. Reproducibility of the Functional Parcellation across Subjects

Agreement between two group level functional atlases at different parcellation scales. Functional atlases were constructed using the raw time series. X-axis shows scale of parcellation, i.e. number of regions. Y-axis is level of agreement between the two atlases quantified using ARI.
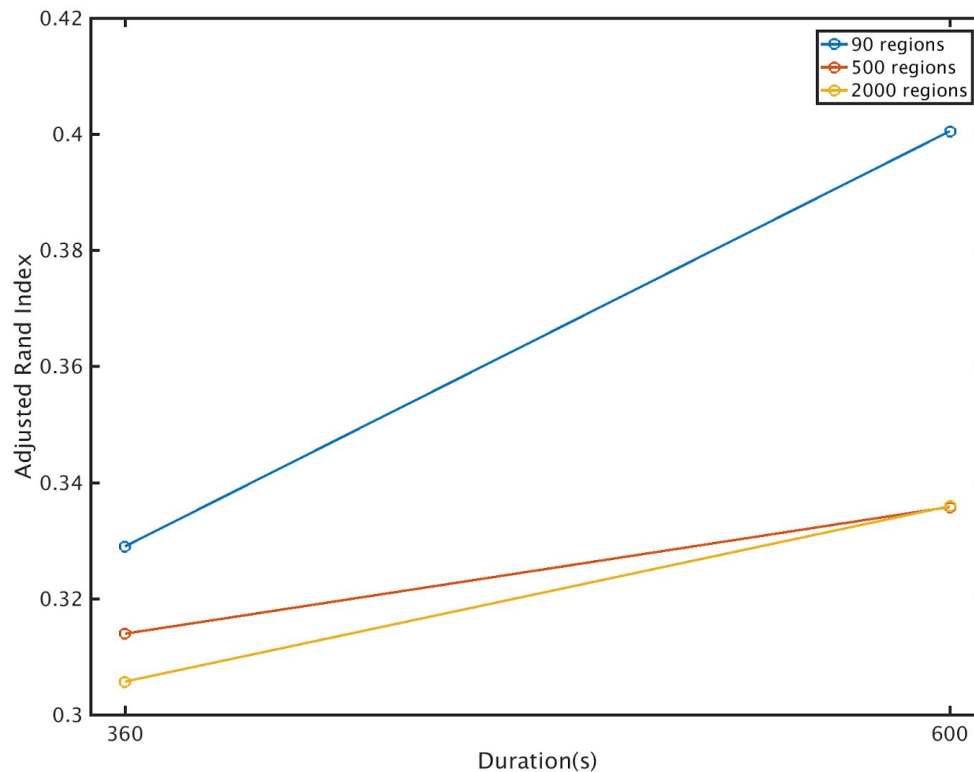
Figure 2.3. Effect of Scan Duration on Reproducibility

Agreement between functional atlases constructed using different datasets of equal duration, at different durations and parcellation scales.

In addition, we examined homogeneity of parcellation results as a function of scan duration used for construction of the parcellation atlas. The results are shown in Figure 2.4. We are showing the results when average pairwise correlation between voxels within each region (*rt*) is used as the measure of homogeneity. A similar pattern was observed with other measures (data not shown). As can be seen in this Figure, scan duration does not affect homogeneity of the regions. However, parcellation at a finer scale (500 regions) results in more homogeneous regions than coarser scale (90 regions).

21
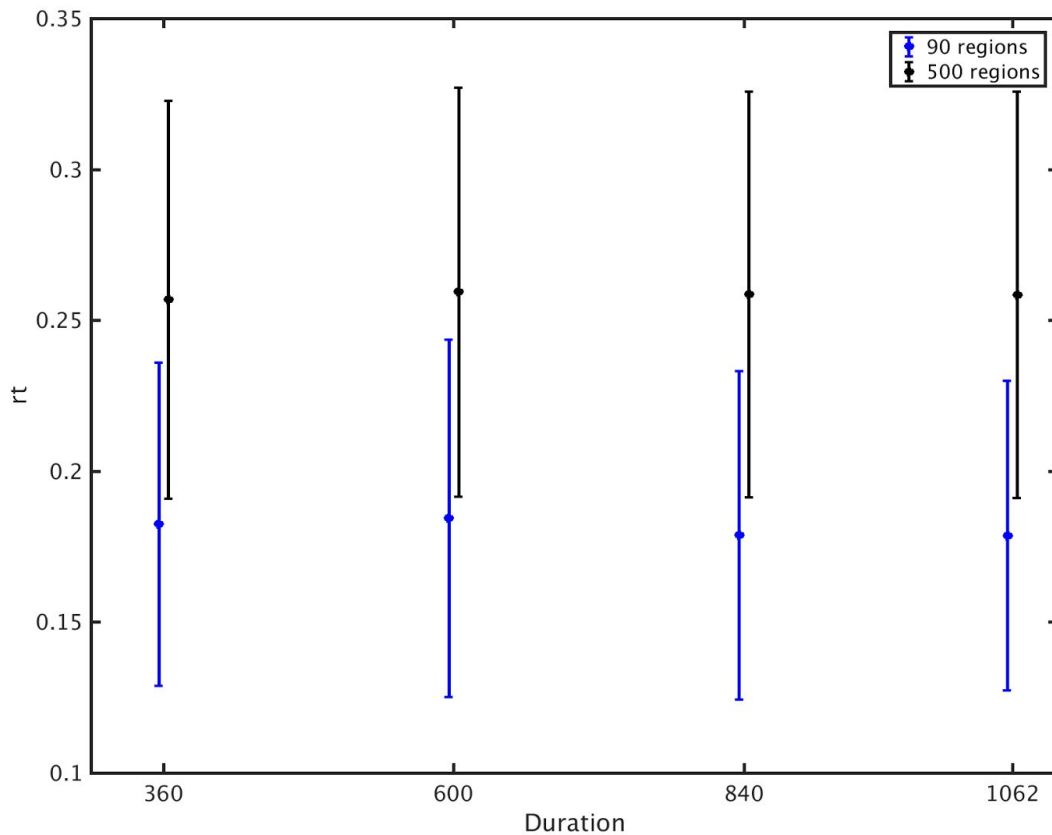
Figure 2.4. Effect of Scan Duration on Homogeneity

Average and standard deviation of homogeneity across across all regions vs. scan duration (second) used for construction of the atlas.

Lastly, we constructed individual level atlases of 7 of our subjects, using raw time series. To probe how much individual atlases differ from the group level, and how much they differ from each other, we calculated ARI between each individual atlas and the group level atlas, as well as degree of agreement between each pair of individual level atlases. Results are shown in Figure 2.5. As can be seen degree of agreement between the individual and group level atlases are slightly higher than that of any individual level atlas to other individual level atlases.
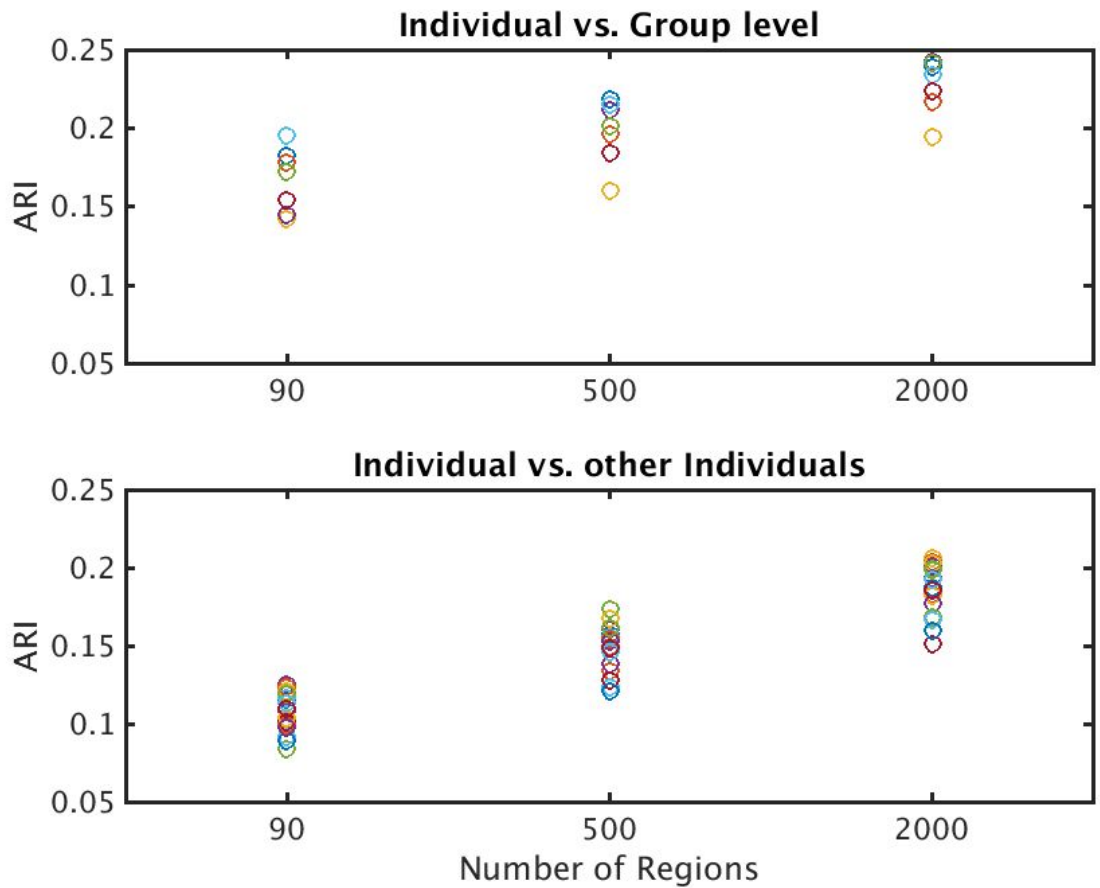
Figure 2.5. Individual vs. Group Level Atlases

Top: Degree of agreement between the seven individual level atlases and the group level atlas at different parcellation scales

Bottom: Degree of agreement between each pair of the seven individual atlases (21 pairs total) at different parcellation scales.

Discussion

We constructed and evaluated group level functional atlases using both raw and pre-whitened time series, using the hierarchical clustering algorithm. We compared the

resultant atlases with anatomical atlases. We also evaluated the resultant atlases by quantifying their homogeneity, separation between between regions, and reproducibility. Lastly, we characterized effect of scan duration on homogeneity and reproducibility, and compared group and individual level atlases.

Our functional atlases, resembled the anatomical atlas in terms of distribution of size of the regions. However compared to the anatomical atlas, or random parcellations, both functional atlases resulted in significantly more homogeneous regions. This demonstrates the effectiveness of the functional atlas to group voxels with similar functional activity together. This significant difference was observed using three out of the five homogeneity measures we used. These measures capture different aspects of the dataset. Since these measures capture different aspects of degree of homogeneity between regions, these results demonstrate that studies evaluating their results using different measures, might not necessarily be comparable, as one measure might show significance, while others would not.

Separation between the regions, quantified using the Silhouette coefficient, showed no difference between the functional atlases and the anatomical atlas or random parcellations. Although we observed that the difference between distribution of Silhouette values between the white functional atlas and the AAL atlas was close to significance (p=0.08).

We observed that pre-whitening results in regions with lower homogeneity and Silhouette coefficient values. This is expected, since pre-whitening removes spuriously high correlations. However, the white functional atlas was in high agreement with the original functional atlas (89%). This shows that even though pre-whitening shifts distribution of pairwise correlations between voxels to the left, overall it does not affect pattern of pairwise cross correlations among voxels. Degree of agreement between the two functional atlases and the anatomical atlas was drastically lower (~25%), demonstrating that regions delineated by functional activity do not align very well with anatomically marked regions.

Reproducibility of the results across different subjects was moderate, as degree of agreement between functional atlases constructed using different groups of subjects was slightly above 30%. Reproducibility did not depend on parcellation scale. The

reason is interindividual differences in where the functional borders between regions are. A group level atlas reflects the commonalities between all the individuals in the group, which is presumably very similar to that of a different group of subjects, provided that these groups are large enough. Our results show that larger groups (at least larger than 22 subjects) are required to construct group level functional atlases that are highly reproducible across groups.

Reproducibility of the functional atlases across datasets was calculated as degree of agreement between functional atlases constructed from different datasets of the same subjects, and was comparable to reproducibility across subjects. Although reproducibility across datasets showed improvement as scan duration increased. This trend is expected, as longer scan durations result in more robust estimates of the cross correlation between different voxels. A recent study has shown that reproducibility across datasets increase as scan duration increases and approaches maximum value at about 30 minutes of data (Laumann et al. 2015).

Duration of scan used for construction of the functional atlas, resulted in higher reproducibility across datasets, but did not affect homogeneity of the regions. Although counterintuitive, it is possible that improvements to homogeneity will happen at longer durations. In other words, even our longest duration (17.7 minutes), is not long enough to produce any changes in homogeneity of the regions.

We observed that individual level atlases have a higher degree of agreement with the group level functional atlas compared to agreement with other individual level atlases. As pointed out before, individual level atlases capture the unique functional organization of each individual, when a group level atlas captures commonalities among the individuals it contains. So the group level atlas can be thought of an "average" functional atlas.

Overall, spatially constrained hierarchical clustering algorithm seems to be a promising method for construction of the functional atlases. However, it seems that to evaluate the method more rigorously, datasets with longer scan duration and lower repetition time (TR) are required.

Supplementary Material
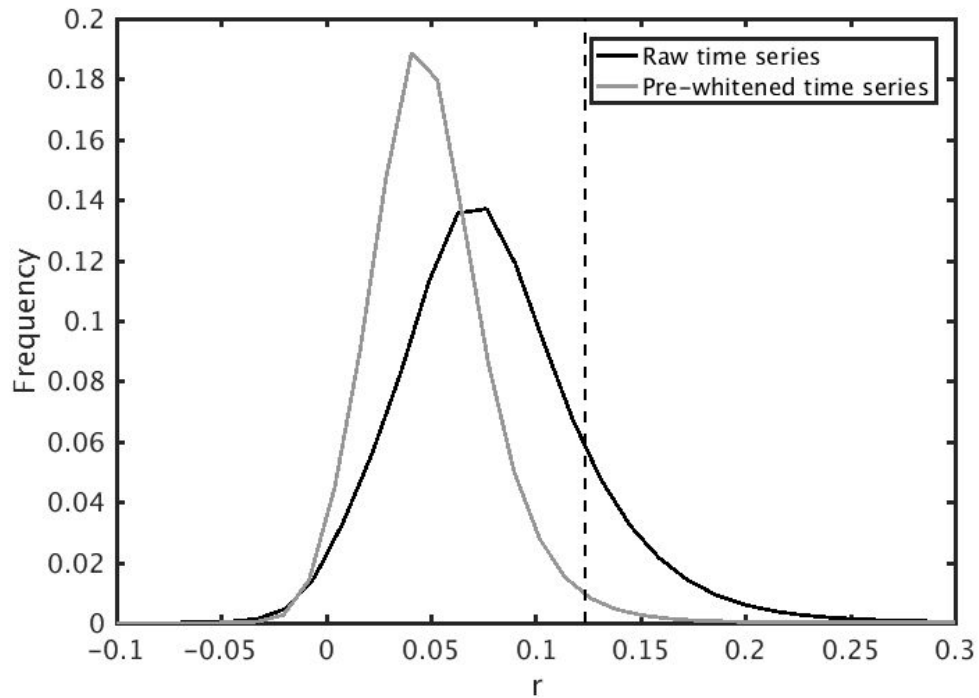
Supplementary Figure 2.1. Distribution of Pairwise Cross Correlations

Distribution of pairwise cross correlation between voxels for raw and pre-whitened time series. Vertical dash line marks significance threshold at p<0.05.
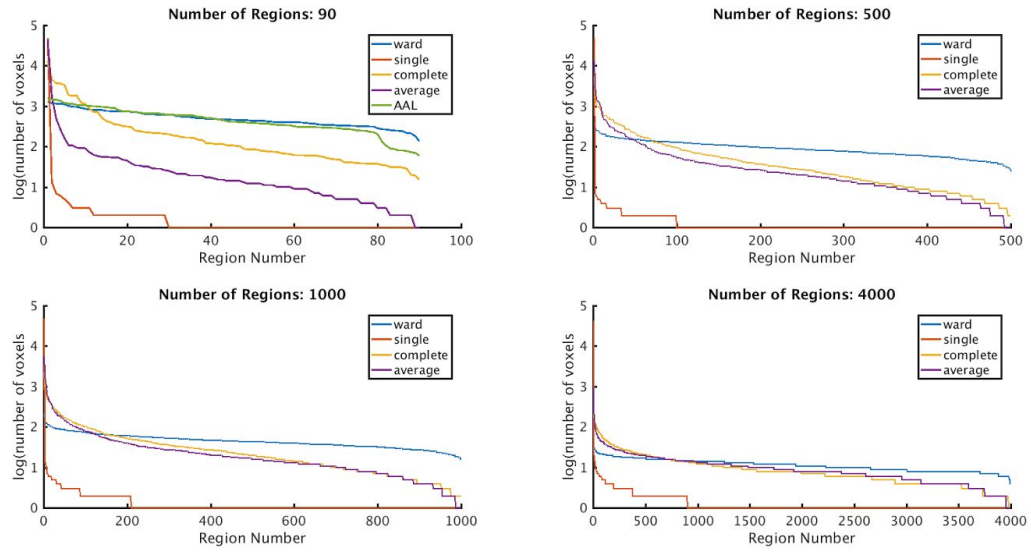
Supplementary Figure 2.2. Distribution of Region Sizes for Different Linkage Methods

Distribution of size of regions for functional atlases constructed using different linkage criteria at different scales. Size of the regions are sorted from highest to lowest.

X-axis: rank of the region, Y-axis: number of voxels in the region in log scale. Size distribution of the AAL anatomical atlas is also shown at a scale of 90 regions. Original time series were used for construction of these atlases.
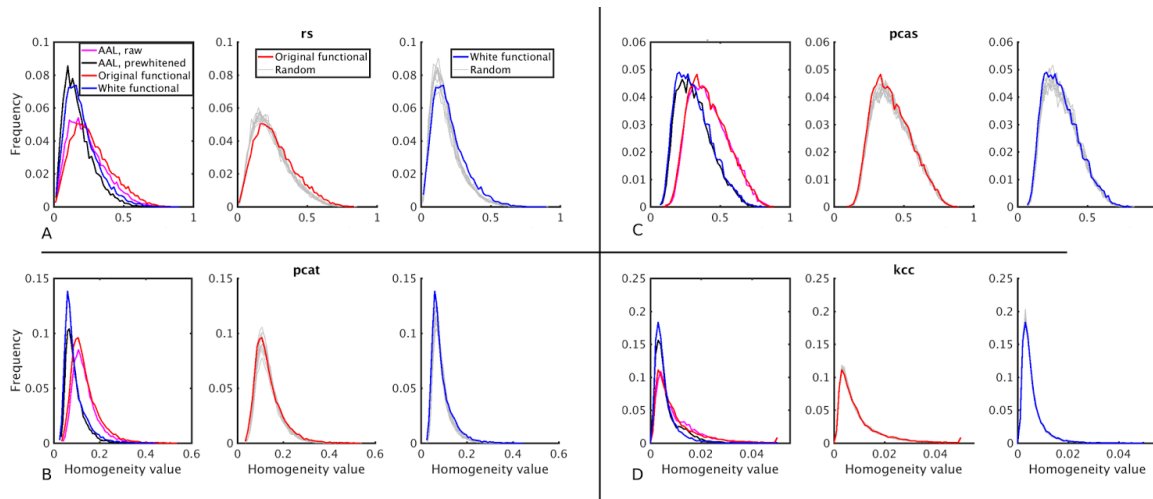
Supplementary Figure 2.3. Distribution of Different Homogeneity Measures across Different Atlases

A) Distribution of pairwise correlation between functional connectivity maps of voxels within each region (*rs*). <u>left:</u> Distribution of values for the two functional atlases, as well as the AAL anatomical atlas applied to raw and pre-whitened time series. Distribution of values was significantly different between the original functional atlas and AAL atlas applied to raw data (Kolmogorov-Smirnov test, p<0.01). Distribution of values was significantly different between the white functional atlas and AAL atlas applied to prewhitened data (Kolmogorov-Smirnov test, p<0.01). <u>Middle:</u> Distribution of homogeneity values for the original functional atlas and ten size matched random parcellations. Distribution of the values was significantly different between the functional atlas and each fo the ten random parcellations (Kolmogorov-Smirnov test, Corrected for multiple comparisons, p<0.001). <u>Right:</u> Distribution of homogeneity values for the white functional atlas and ten size matched random parcellations. Distribution of the values was significantly different between the functional atlas and each fo the ten random parcellations (Kolmogorov-Smirnov test, Corrected for multiple comparisons, p<0.001).

B) Distribution of percent of variance explained by the first principal component of the time series of voxels within each region (*pcat*). Arrangement is the same as A.

Distribution of values was significantly different between the original functional atlas and AAL atlas applied to raw data (Kolmogorov-Smirnov test, p<0.01). Distribution of values was significantly different between the white functional atlas and AAL atlas applied to prewhitened data (Kolmogorov-Smirnov test, p<0.01). Distribution of values was significantly different between each functional atlas and each of the ten random parcellations (Kolmogorov-Smirnov test, Corrected for multiple comparisons, p<0.001).

C) Distribution of percent of variance explained by the first principal component of the functional connectivity map of voxels within each region (*pcas*). Arrangement is the same as A. Distribution of values was not significantly different between the original functional atlas and AAL atlas applied to raw data (Kolmogorov-Smirnov test, p=17). However, distribution of values was significantly different between the white functional atlas and AAL atlas applied to prewhitened data (Kolmogorov-Smirnov test, p<0.01). When the original functional atlas was compared to each of the ten random parcellations, only four out of ten distributions were significantly different (Kolmogorov-Smirnov test, Corrected for multiple comparisons, p<0.001). When the white functional atlas was compared to each of the ten random parcellations, only five out of ten distributions were significantly different (Kolmogorov-Smirnov test, Corrected for multiple comparisons, p<0.001).

D) Distribution of Kendall's coefficient of concordance (*KCC*). Arrangement is the same as A. Distribution of values was significantly different between the original functional atlas and AAL atlas applied to raw data (Kolmogorov-Smirnov test, p<0.01). Distribution of values was significantly different between the white functional atlas and AAL atlas applied to prewhitened data (Kolmogorov-Smirnov test, p<0.01). When the original functional atlas was compared to each of the ten random parcellations, only five out of ten distributions were significantly different (Kolmogorov-Smirnov test, Corrected for multiple comparisons, p<0.001). When the white functional atlas was compared to each of the ten random parcellations, only five out of ten distributions were significantly different (Kolmogorov-Smirnov test, Corrected for multiple comparisons, p<0.001).

Chapter 3: Classification of Schizophrenia Patients using Network Measures: A Data Driven Approach

Introduction

Schizophrenia is a debilitating disease that affects about 1.1% of the adult US population according National Institute of Mental Health . One hypothesis about cause of schizophrenia, known as the "disconnectivity hypothesis" (Friston 1998) posits that normal pattern of functional connectivity between distinct regions of the brain is disrupted. This hypothesis has been studied extensively over the past decade using Functional Magnetic Resonance Imaging (fMRI). FMRI provides a unique means to study schizophrenia because it is non-invasive, can be used to identify prodromal state of schizophrenia, and unlike Electroencephalography (EEG) can image deep brain structures, such as thalamus. Many studies have searched for a common neural biomarker of the disease. A reliable biomarker could help clinicians with diagnosis. Biomarkers may also provide insight into the mechanism of the disease and could guide researchers to developing novel therapeutic interventions.

Finding biomarkers that are replicable across different studies and patient populations has been challenging (Pettersson-Yeo et al. 2011). Different studies have found abnormal patterns of connectivity in different regions of the brain. For example, the thalamus is often implicated, but in some studies they find a hyperconnectivity (Wolf et al. 2009; Skudlarski et al. 2010; Zhang et al. 2012; Atluri et al. 2015) of the thalamus and other studies hypoconnectivity (Andreasen et al. 1996; Zhou et al. 2007; Tu et al. 2010).

There are many approaches to identifying biomarkers using fMRI: 1) Univariate approach, which quantifies characteristics of each brain region separately and compares that across the patient and healthy groups (Shi et al. 2007; Bassett et al. 2012). 2) Bivariate approach, which quantifies interaction between each pair of regions in the brain and searches for region pairs that are discriminative across patient and healthy groups (Lynall et al. 2010; Bassett et al. 2012; Tang et al. 2012; Su et al. 2013; Guo et al. 2013;

Kim et al. 2016; Arbabshirani, Castro, and Calhoun 2014). 3) Multivariate approach, which quantifies more complex interactions between all regions, typically using graph theoretic measures (Lynall et al. 2010; Bassett et al. 2012; van den Heuvel et al. 2010, 2013; Fekete et al. 2013; Anderson and Cohen 2013; Singh and Bagler 2016). Recent studies have demonstrated that the multivariate approach has more discriminating power (Venkataraman et al. 2010; Atluri et al. 2013).

Discriminative power of calculated measures is quantified either using statistical tests (e.g. (Lynall et al. 2010; van den Heuvel et al. 2010, 2013)) or classifiers such as support vector machines (SVM) (e.g. (Castro et al. 2011; Bassett et al. 2012; Anderson and Cohen 2013; Arbabshirani et al. 2013; Singh and Bagler 2016)). Statistical tests compare the means and depend on number of subjects. On the other hand, classification performance of a measure is useful because the measure indicates how the accuracy of the classifier if it were to be used in clinical setting. To be more specific, significant statistical difference between two groups does not automatically translate into high classification performance. For example, if the class distributions have different means, but high standard deviation and therefore overlap, significant statistical difference could occur with low classification performance.

Diversity of results on schizophrenia biomarkers is attributable to multiple factors. First, typical sample size are low (average 38 subjects based on sample sizes reported in (Pettersson-Yeo et al. 2011)). Second, different studies use different pre-processing steps. Third, tests that identify biomarkers that are statistically significantly different vs. those that use classification performance put different values on the performance of the features. Fourth, how biomarkers are validated in the studies are very different and some methods may result in more robust generalization than others.

In this study we examined predictive power of several multivariate biomarkers, but also explored how such predictive power can be affected by pre-processing steps and the biomarker discovery method. This work extends previous work in four important ways. First, we used a large sample (170 subjects). Second, we used a purely data driven approach to finding biomarkers. We used a comprehensive set of graph theoretic measures and used data driven measures to find which ones have more discriminative power. Third, we used an alternative pre-processing pipeline where functional activity

was pre-whitened to remove spurious cross correlation between voxels. Moreover, definition of brain regions was extracted based on functional activity of the subjects themselves, as opposed to using a conventional anatomical atlas. And fourth, we used a rigorous double cross-validation method to discover biomarkers and report their prediction accuracies that we expect will be more robust across different patient populations.

In this study, we used 6 minute resting state fMRI data from patients with schizophrenia and healthy control subjects. All time series were first pre-whitened, then used to construct a functional atlas. Regions of the functional atlas were used to construct unweighted graphs for each subject, and several multivariate graph theoretic measures were calculated. The measures were then used to classify patients from healthy controls using support vector machines. Most informative features were identified and used to report classification performance in a double cross-validation scheme where separate sets of subjects were used for feature selection and classification respectively.

Methods

Participants

A total of 170 subjects participated in this study: 52 chronic (17 female, age: M = 37.0, SD = 10.8) and 30 first episode (8 female, age: M = 25.7, SD = 7.1) schizophrenic patients as well as two groups of healthy control subjects matched to each patient group in demography: 55 control subjects (18 female, age: M = 38.0, SD = 11.9) to match the chronic group and 33 control subjects (9 female, age: M = 25.5, SD = 6.9) to match the first episode group (see Table 3.1 for detailed information on participants). All participants gave informed consent and were compensated for their participation. Schizophrenia patients were assessed for negative and positive symptoms using the Scale for Assessment of Negative Symptoms (SANS) and Scale for Assessment of Positive Symptoms (SAPS) (Andreasen and Olsen 1982). All procedures were done in accordance with a University of Minnesota IRB approved protocol.

	Age	SANS	SAPS	Medication
Chronic Schizophrenia Patients (N =52)	37.0(10.8)	33.0(14.3)	23.5(17.3)	1 Typical 38 Atypical 5 Both 4 No mdes 4 N/A
Chronic Healthy Controls (N = 55)	38.0(11.9)	N/A	N/A	N/A
First Episode Schizophrenia Patients (N = 30)	25.7(7.1)	30.1(17.4)	25.3(16.9)	0 Typical 21 Atypical 0 Both 3 No mdes 6 N/A
First Episode Healthy Controls (N = 33)	25.5(6.9)	N/A	N/A	N/A

Table 3.1. Summary of Characteristics of Study Participants. Mean (and SD) Demographic and Diagnostic Characteristics of participants. *Note:* SANS, Scale for Assessment of Negative Symptoms; SAPS, Scale for Assessment of Positive Symptoms.

Imaging Data Acquisition and Preprocessing

Resting state fMRI scan for a duration of six minutes was collected from each participant as detailed in (Camchong et al. 2011; Atluri et al. 2015). Participants were instructed to remain as still as possible, stay awake and keep their eyes closed. Images were acquired using a Siemens Trio 3T scanner (Erlangen, Germany). Sequence parameters used in this study are as follows: gradient-echo echo-planar imaging (EPI) 180 volumes, repetition time (TR) 2 seconds, echo time (TE) 30ms, flip angle 90°, 34 contiguous AC-PC aligned axial slices, voxel size 3.4 x 3.4 x4.0 mm, matrix 64 x 64 x 34 totalling 139,264 voxels.

Participants were asked at the end of the scan whether or not they stayed awake during the scan and for the one patient that fell asleep during the scan the scan was repeated

under awake conditions. Also, a T1-weighted anatomical image was acquired using a magnetization prepared rapid gradient-echo sequence. In addition, a field map was acquired and used to correct for geometric distortions introduced by field inhomogeneities: TR = 300ms, TE = 1.91 ms/4.37 ms, flip angle = 55°, voxels size = 3.4 x 3.4 x 4.0 mm.

To remove recording artifacts and noise, register the data and downsample to a manageable size, the raw fMRI data was preprocessed using FEAT and MELODIC from the FSL software package as follows. First, the first three volumes were excluded from each subject scan to account for magnetization stabilization. The subsequent scans were then motion corrected, B0 field map unwarped, and corrected for slice scan time. Non-brain portions of the images were removed and a spatial smoothing kernel was applied to the dataset (6mm full-width half-maximum). The images were then grand mean and intensity normalized and temporally filtered between 0.01 and 0.08Hz. All images were then registered to the MNI152 space. To remove noise introduced by head motion, respiration, cardiac pulsation, and scanner artifacts, probabilistic independent component analysis (PICA) was used. Spatial and temporal characteristics of noise components are described in MELODIC manual (https://fsl.fmrib.ox.ac.uk/fslcourse/lectures/melodic.pdf). The dataset was then resampled to 3 x 3 x 3mm, resulting in 47640 voxels.

Functional parcellation

Functional parcellation is the process of grouping voxels with similar functional activity together to form regions using data-driven algorithms. Functional parcellation uses cross-correlation between the voxel time series to identify voxels with similar functional activity. Ideally, cross-correlation is calculated on time series that are stationary and have no auto-correlation, characteristics of white noise. However, BOLD time series are typically non-stationary and are highly autocorrelated leading to spuriously high cross-correlations (Christova et al. 2011). For accurate functional maps it is important to remove these factors that lead to spuriously high correlation values. One approach to removing non-stationary and autocorrelated trends from the time series has been coined "pre-whitening" (Christova et al. 2011). This approach has previously been used to find biomarkers for Post Traumatic Stress Disorder (PTSD) and shown to enhance

classification performance of the biomarkers (Peka Christova et al. 2015). Therefore, all time series were prewhitened prior to constructing the functional atlas.To prewhiten the time series from voxel i $x_i(t)$, the Fourier transform of the time series, $X_i(f)$, was calculated and divided by the absolute value of the spectrum (Equation 3.1), so that similar to white noise, the amount of power in each frequency band was equal, and the autocorrelation of the time series became an impulse.

$$x_i(t) \leftrightarrow X_i(f)$$

$$X_i^W(f) \; = \; X_i(f) \, / \, \left| X_i(f) \right| \leftrightarrow x_i^W(t) \hspace{4cm} \textit{Equation (3.1)}$$

The resultant spectrum, $X_i^W(f)$, was then transformed back into the time domain $x_i^W(t)$ to make a prewhitened data set. While this approach uses an a-causal approach to pre-whitening, unlike like fitting an ARMA model, it is highly efficient and if only the zero time lag correlation is used, then this approach can be used for undirected similarity measures.

The functional atlases was constructed at group level by combining scans from the control subjects. To combine individual scans, we concatenated time series from all the subjects, to obtain a single time series per voxel (Figure 3.1A). We then calculated the correlation adjacency matrix between the voxels $C = (diag(\Sigma))^{-\frac{1}{2}}.\Sigma.diag(\Sigma))^{-\frac{1}{2}}$ where $\Sigma = X_W^T.X_W$ is the covariance matrix and X_W is matrix, where each column corresponds to one time series. For a dataset consisting of N voxels, the correlation adjacency matrix is an NxN symmetric matrix where value of the ith row and j^{th} column is the Pearson correlation coefficient (Altman 2006) between time series of voxels i and j (Figure 3.1B). Pairwise correlation values were then used to calculate pairwise correlation distance between voxels, which is equal to 1 minus the correlation coefficient between the pair, and ranges from 0 to 2.

To construct the atlas, we used the agglomerative hierarchical clustering algorithm, with Ward's minimum variance method as the linkage criterion (Ward 1963; Tan, Steinbach, and Kumar 2006). Using this algorithm, first each voxel is treated as a single region or

cluster. Then, the pair of clusters with minimum distance variance among all the pairs are grouped together to form a single cluster. This process is repeated until all the voxels are merged into a single cluster. The algorithm keeps track of which voxels belong to which cluster in every step. This information is stored in a structure called a dendrogram (Figure 3.1C). The final cluster assignments for each data point is then obtained by 'cutting' the dendrogram at the desired scale. For example to construct a functional atlas with 90 regions, information about voxel memberships is extracted from the dendrogram at the level where voxels are clustered into 90 groups. To obtain contiguous regions, a spatial constraint was enforced when constructing the dendrogram that allowed two clusters to be merged only if they contained spatially neighboring voxels, and therefore their merger would result in a contiguous region. Our choice of the ward's linkage method was based on an exploratory analysis of different parcellation methods described in detail in [Chapter 2].
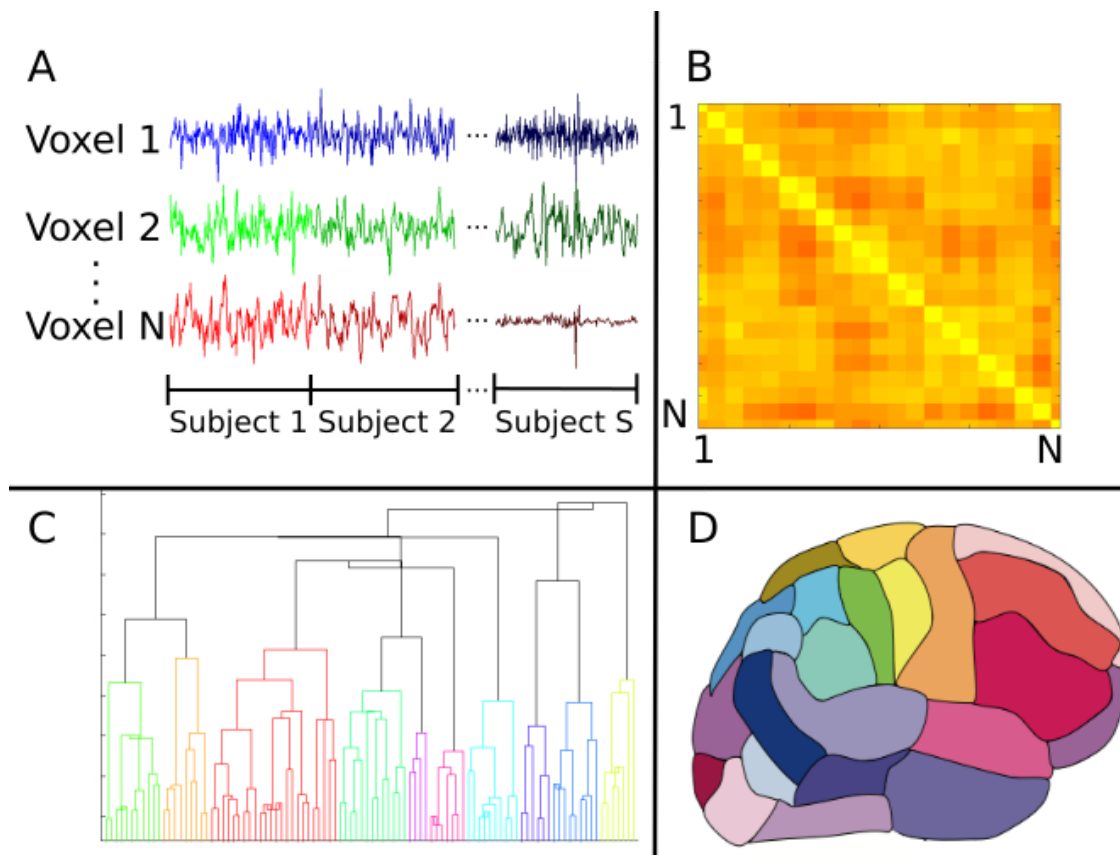
Figure 3.1. Construction of the Group Functional Atlas

(A)   Combine Datasets. In order to combine datasets from individual subjects, time series from all control subjects were concatenated for each voxel. Each individual dataset consisted of 47640 voxels and 177 time points.

(B)   Adjacency Matrix. Pairwise correlation coefficient between the combined time series was calculated and used to construct the correlation adjacency matrix. For our dataset, this was a 47640 x 47640 symmetrix matrix which was used to calculate pairwise correlation distance between voxels.

(C)   Dendrogram. A dendrogram contains all the information about membership of each datapoint at each stage of hierarchical clustering. At the bottom of the dendrogram, each single data point constitutes a single cluster. At each stage of the hierarchy, the pair of clusters that are most similar as evaluated by the linkage criterion are merged to form bigger clusters. Eventually, at the top of the hierarchy all data points are merged to form a single cluster.

(D)   After cutting the dendrogram at the proper scale, i.e. the desired number of clusters or regions, a parcellation of the brain is produced based on which voxel belongs to which cluster, which is used as the atlas for the rest of our analyses.

To compare classification accuracy of the with an anatomical atlas, we constructed a functional group atlas with 90 regions to compare to the commonly used Automated Anatomical Labeling (AAL) atlas, which also has 90 regions excluding cerebellum (Tzourio-Mazoyer et al. 2002).

Network Model

After constructing the functional atlas, a graph model of the brain was constructed for each subject by first applying the atlas to the individual datasets. Time series of all

voxels within a single region were averaged to obtain a single time series per region. Pairwise Pearson correlation coefficient between the regions was then calculated and used to construct a weighted undirected graph, where each region constituted one node and the links were weighted by the correlation coefficient value between nodes (Figure 3.2). Calculation of the network measures requires all the weights to be non-negative, so negative weights were set to zero. There is currently no general consensus over the cause of negative correlation coefficients (Chen et al. 2011). We observed that only 2±3% of all cross correlations were negative. Several measures are specific to binary graphs (Supplementary Table 3.1). In order to construct a binary graph, weights that were below a threshold were set to zero and weights above the threshold were set to one. The threshold was chosen to obtain a binary graph with 30% connection density where 30 percent of all links were set to have a weight of one (Lynall et al. 2010). In addition, some measures required the graph to be divided into communities (Supplementary Table 3.1) and the information about the community membership was required for their calculation. To divide the graph into communities, the Louvain method for community detection (Reichardt and Bornholdt 2006; Ronhovde and Nussinov 2009) was used. After constructing the weighted and binary graphs, several graph theoretic measures (Bullmore and Sporns 2009; Rubinov and Sporns 2010b) were calculated (see Figure 3.2 and Supplementary Table 3.1 for a list of the measures), using the Brain Connectivity Toolbox (Rubinov and Sporns 2010b, [a] 2010). Some measures that required specification of extra parameters, as summarised in Supplementary Table 3.1.

Graph theoretic measures typically capture characteristics of each node (producing one value per node), each pair of nodes (producing one value per node-pair), or the entire network as a whole (producing one value per network). In addition, for each measure, its average and standard deviation across all regions were also used as separate measures. For a graph with 90 regions therefore, 19000+ network measures total were produced.

Network Measures

Node Degree
Strength
Rich Club Curve
Assortativity
Transitivity
Reachability
Distance
Distance for Weighted Netwroks
Global Efficiency
Local Efficiency
Characteristic Path length
Overlap Amongst Neighbors
Generalized Topological Overlap
Matching Index
Clustering Coefficient
Participation Coefficient
Within Module Degree Centrality
Eigenvector Centrality
Node Betweenness Centrality
Shannon Entropy Based Diversity
Radius of Graph
Eccentricity
Diameter of Graph
Number of Communities
Average Community Size

Figure 3.2. Network Level Model of the Brain

After applying the functional atlas to each individual dataset, time series of voxels within each region were averaged resulting in a single time series per region. These regions were used as nodes of the graph, where the link between each pair of nodes was weighted by the correlation coefficient between the nodes. A set of measures that capture network characteristics were then calculated for each subject.

Classification

To classify control subjects from schizophrenic patients we used support vector machines (SVM) (Vapnik 1995; Bishop 2006). SVMs are robust to presence of noisy data points, because they maximize the classification margin (Figure 3.3A). There are two free parameters for an SVM that need to be set by the experimenter: box constraint (C value) and kernel. We used a C value equal to 1, and a linear kernel.

The19000+ graph theoretic measures were used as features for classification of 170 subjects into either control or schizophrenic. Using this feature set for classification of a data set poses two challenges on classification. The first challenge is that feature set is orders of magnitude larger than the number of subjects (a problem called 'the curse of dimensionality') (Jain, Duin, and Mao 2000). This forces the classifier to pick up patterns that are specific to the subjects that are used for its training and therefore are not generalizable to other subjects, a phenomenon called 'overfitting' (Clarke et al. 2008). The second challenge is that not all features are equally informative to the classifier (Guyon and Elisseeff 2003). We need to know which features are contributing more to the classification process in order to extract effective biomarkers. Therefore, we need to reduce dimensionality of the data by selecting an optimal or sub-optimal subset of features for classification. Here we used SVM for both feature selection and classification, so to ensure the optimized feature set is generalizable across subjects, we used a double cross validation scheme (Filzmoser, Liebmann, and Varmuza 2009; Sundermann et al. 2014).

To perform double cross validation, the subject set was randomly partitioned into three separate subsets: train, validation, and test subsets (illustrated in Figure 3.3B). The train subset was used to train an SVM model. The model performance was then validated on the validation subset. The training and validation subsets were used to iteratively optimize feature selection and SVM parameters. This ensures that the final performance is not influenced by the optimization, and reflects performance of the features more robustly than a single cross validation scheme, which uses the optimal SVM model to classify the validation subset itself.

The subject set (170 subjects) was divided into five randomly chosen subsets of equal size (34 subject each) and used in a 5-fold cross validation (Efron and Gong 1983; Efron 1983). One fold was left out to be used as the test subset and the rest were used for the SVM model optimization. For the feature selection, the training and validation folds were shuffled 4 times and used iteratively to select features that performed best on the validation set (Figure 3.3B). This performance is reported as the single cross validation performance. Once the features were selected, the SVM was trained using both the train and validation folds and then applied to the test fold to obtain the final classification

accuracy ($\frac{TP + TN}{T}$), where TP is number of true positives, i.e. patients classified correctly, TN is number of true negatives, i.e. controls classified correctly, T is the total number of subjects. This approach is similar to the leave one out cross validation (LOOCV) scheme, except that instead of leaving out a single subject, we leave out a single fold. In addition to classification accuracy, sensitivity ($\frac{TP}{P}$) and specificity ($\frac{TN}{N}$) (Fawcett 2006) were also reported (P is the total number of patients and N is the total number of controls).

The random partitioning into the 5 folds was performed 10 times resulting in 50 optimized feature sets, of 30 features each, and 50 prediction accuracies. To determine if inclusion of any feature in the feature set occurs more often than expected by chance, we calculated the probability of each feature appearing n times out of 50. The probability of each feature appearing once in each selected set is equal to the probability of drawing 30 random samples from a set of F items without replacement, which can be calculated with the hypergeometric distribution $p_{select,S} = h(1 | F, S, P)$ where F is the total number of features used (F=1618), S is the number of samples (S=30), and P is the number with the desired property (P=1). Given the probability of sampling each feature at random, we can then calculate the number of times that feature is expected to appear with N_{fold} independent draws using a binomial distribution $Pr(n) = B(N_{fold}, p_{select,S})$. Features that appeared more frequently than what was predicted by chance were further analyzed.

With 10 random partitions (N_{perm}), each subject is used in the test subset 10 times. We calculated the proportion of times each subject was misclassified (M), a measure we call "misclassification rate" in this manuscript (equation 2). Correlation between misclassification rate (MR) and severity of symptoms for the patient population was calculated:

$$MR = \frac{M}{N_{perm}}$$

In order to see if choice of classification algorithm affects the classification performance, we compared performance of SVM to SVM with Adaboost (short for Adaptive Boosting) (Freund and Schapire 1997; Yoav Freund 1999) with a linear SVM (C value equal to 1), and 10 weak classifiers were trained.

All analyses were implemented in MATLAB 2016b.
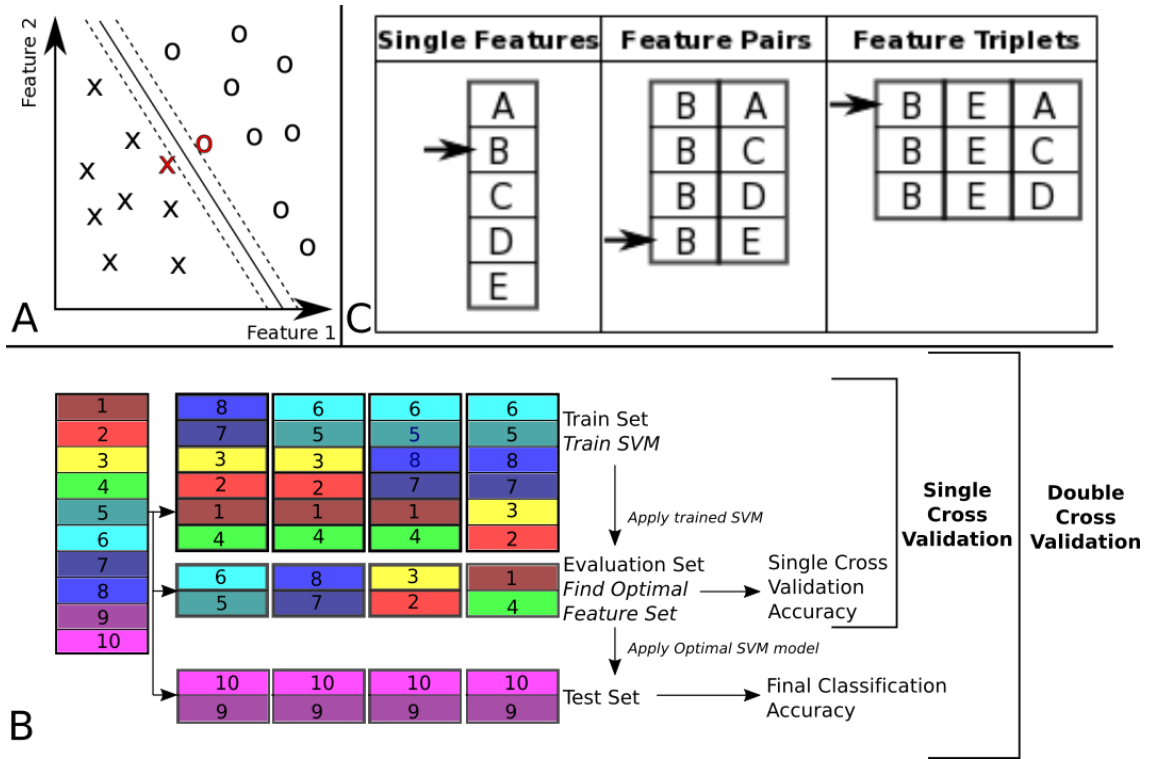
Figure 3.3. Classification

(A)   Support vector machines are a type of supervised classifier that maximizes the margin between the separating hyperplane (continuous black line) and data points. Data points closest to the hyperplane (red data points) are the support vectors. In this toy example, the data consists of two features. Our dataset consists of 38000+ features, i.e. classification happens in a 19000+ dimensional space. The hyperplane is characterized by a set of weights (W) and constant (b) and projects the dataset onto a single dimensions.

(B)   Double cross validation procedure divided the subjects into three subsets. The training and validation subsets were used for optimizing the feature set, and the test subset is used for classification performance. Different subsets are determined in a 5-fold cross validation division scheme, where the entire dataset is divided into 5 equal size subsets and each subset is used once as the test subset. In this toy example, subjects 9 and 10 are used as the test set,

while the rest of the subjects are used for feature selection. For feature selection a 4 fold cross validation scheme is used, where the subjects are partitioned into 4 equal size groups and each group is used once as the validation subset. 4 SVM models are trained for each train subset, and performance of each set of features is averaged across the 4 classifiers. Features that perform better on average are then chosen to be tested on the test set.

(C) Sequential Forward Selection algorithm is demonstrated in a toy example. First, performance of every single feature is calculated by training a SVM using that feature only on the train subset and applying the weights on the validation subset. The single feature with the highest performance is picked (feature B in this example, left column). Subsequently, performance of combination of feature B with all the remaining features is calculated, by training an SVM using each feature pair separately on the train subset and applying the weights to the validation subset. Feature pair with the highest performance is then selected (features (B, E) in this example, middle column). The selected feature pair is then combined with all the remaining features to form feature triplets, performance of each is then calculated through the same cross validation procedure. The feature triplet with the highest performance is then picked (features (B, E, A) in this example, right column).

Feature Selection

With 19000+ features, only a fraction are informative for classification and the others dilute classification power by causing the classifier to overfit. Therefore, it is beneficial to choose the subset of features that are the most informative.

These were determined using a data driven greedy search procedure, called sequential forward selection (SFS) (Guyon and Elisseeff 2003) (Figure 3.3C). First, the classification accuracy of each single feature alone was measured using SVM, cross validating across the train and validation subsets. Only features with prediction accuracy above 60% were used for the subsequent stages of the optimization. Then, the feature

with the top performance was progressively combined with other features, selecting the combinations with highest accuracies, until a set of 30 features were selected. This method of feature selection is computationally expensive but it is more robust than simply selecting 30 features that independently have the highest performance. Many of the top features alone may have redundant information. This algorithm accounts for the combinatory effect of features. Moreover, while a feature might have low classification performance on its own, in combination with other features it can improve performance (Guyon and Elisseeff 2003). The sequential forward selection algorithm is not guaranteed to find the globally optimal set of features that would maximize classification accuracy, but it is guaranteed to find a local optimum (Liu and Motoda 2007).

To see if our feature selection algorithm improves classification accuracy for our dataset, we compared the accuracy achieved by the sequential forward selection algorithm to that of the best 40 features and top 40 features selected using Fisher's linear discriminant analysis (LDA) (Fisher 1936; Bishop 2006). Fisher's linear discriminant analysis transforms the data into a space where the the separation between the two classes is maximized. Calculation of weights for the linear transformation involves matrix inversion, which is not possible if the within-class scatter matrix is singular, depending on structure of the dataset. Therefore, Moore-Penrose pseudo-inverse of the matrix was calculated (Campbell and Meyer 2008).

Results

Two atlases, the anatomical AAL atlas and a functional atlas constructed using time series from the control group, were used as region definitions for construction of brain networks. The functional activity in each region of the atlas was averaged and zero-lag cross correlation between the regions were used to construct undirected weighted graphs for each subject. Several graph theoretic measures were then calculated for each network and used as features for classification. This resulted in 19,000 features. To reduce the number of features, we selected features whose classification accuracy using linear SVM achieved greater than 60% accuracy. These features were then used in optimizing combinations of features. A distribution of the single feature classification accuracy using the functional atlas and the AAL atlas is shown in Figure 3.4A. For both the functional and the AAL atlas 1618 features on average reached the 60% cutoff.

Using these top performing features, we then optimized for the combination of features that provided the best classification. Classification accuracy was calculated for single cross validation using features selected with the training and validation subsets. The final reported calculated classification accuracy was calculated by double cross validation where the optimized SVM model and features were then applied to a final test set not used in the feature selection optimization. Comparison of the single cross validation and the double cross validation performances for the functional atlas are shown in Figure 3.4B. The single cross validation results correlate with previously reported classification rates with accuracy maximizing at 87% using 14 features, which is significantly above chance level (p<0.001, two sample t-test). The double cross validation maximum accuracy was significantly lower, dropping to 64% at 4 features, which is still significantly above chance (p<0.001, two sample t-test), but about 20% lower than the single cross validation performance. However, this is probably a more accurate rate that would generalize to prospective studies. The high classification rate reported by the single cross validation can be accounted for by the overfitting using the feature selection optimization step. Results are also reported using the AAL atlas in Figure 3.4C. The single cross validation performance was 85% using 18 features, similar to the functional atlas. Double cross validation results however was 73% using a single feature. When we used the Adaboost classifier instead of SVM, accuracy of the functional atlas did not change (Table 3.2, Supplementary Figure 3.1). On the contrary performance of the AAL atlas decreased by 10% compared to the SVM classifier.

We also looked at sensitivity and specificity of our classification algorithm (Table 3.2 and Supplementary Figure 3.2). Sensitivity and specificity for the functional atlas were 65%. These values were higher for the AAL atlas with average sensitivity equal to 77% and specificity equal to 68%. Using Adaboost as the classifier increased specificity and decrease sensitivity in both atlases. Maximum specificity obtained using the Adaboost classifier was 85% and 80% for functional and AAL atlases respectively.

Next, we looked at different methods of dimensionality reduction (Supplementary Figure 3.3). We compared accuracy performance of the forward sequential selection (SFS) algorithm with that of the Linear Discriminant Analysis (LDA) and choosing top features with the highest performance when used independently. Choice of the dimensionality

45

reduction method did not affect performance accuracy (Table 3.2). However, both LDA and independent methods achieved the same performance at higher number of features compared to the SFS algorithm.

We then looked at effect of prewhitening on performance accuracy (Table 3.2 and Supplementary Figure 3.4). We constructed two functional atlases, one with prewhitened time series, and another with raw time series. They were applied to prewhitened and raw time series respectively to construct the graph and compare their respective classification accuracies. We repeated the same procedure with the AAL atlas, when the atlas was applied to prewhitened and raw time series to construct the graph used for classification. We did not observe any difference between performance of the two functional atlases. However, classification performance when the AAL atlas was applied to prewhitened time series was 9% higher than that of raw time series.

Dataset	Pre-whitened			
Classifier	SVM			Adaboost
Dimensionality Reduction	SFS	LDA	Independent	SFS
Functional	**Acc.=64%** Sen.=65% Spe.=65%	Acc.=65%	Acc.=65%	Acc.=63% Sen.=44% **Spe.=85%**
AAL	**Acc.=73%** Sen.=77% Spe.=68%	Acc.=67%	Acc.=73%	Acc.=62% Sen.=52% **Spe.=80%**
Dataset	Raw			
Classifier	SVM			Adaboost
Dimensionality Reduction	SFS	LDA	Independent	SFS
Functional	Acc.=64%	Acc.=63%		Acc.=63%
AAL	Acc.=64%	Acc.=65%		Acc.=61%

Table 3.2. Performance Summary. Classification accuracy, sensitivity and specificity for different datasets, classifiers, dimensionality reduction methods, and atlases. SFS:

Sequential Forward Selection; LDA: Linear Discriminant Analysis; Independent: Selecting top features based on their independent performance.
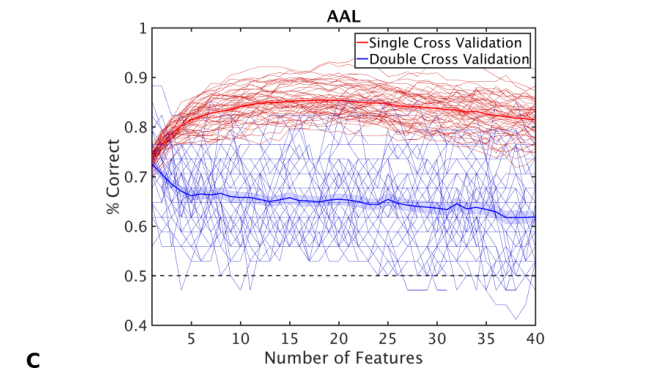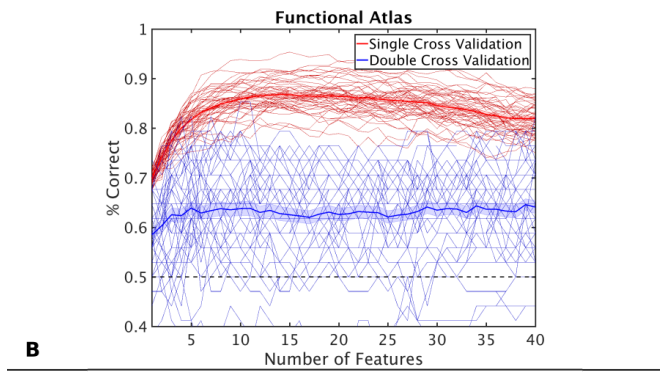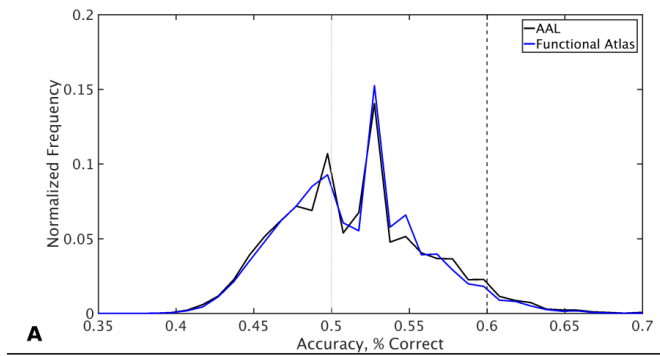
Figure 3.4. Classification Performance

(A)   Distribution of classification accuracy of single features, when used independently for classification, for the functional and AAL atlases. Dashed vertical lines mark the 60% accuracy threshold. Features with less than 60% accuracy were not considered in the feature selection process.

(B)   Classification performance as a function of number of features, using the functional atlas to define nodes of the network.

(C)   Classification performance as a function of number of features, using AAL atlas to define nodes of the network.

Through the feature selection process we identified the top 40 most informative features, which was repeated through n-fold cross validation 50 times. Therefore, each feature could appear in the selected feature set from 0 to 50 times. To identify those features that were selected more often than would be expected if selected randomly, we calculated the probability of a feature being selected n times due to chance, with n ranging from 1 to 50 times. The number of features that were selected n times, as well as the expected number, for the functional atlas is shown in Figure 3.5A. The probability of a feature appearing ten times or more due to chance is very small. Therefore, we further analyzed all features that were selected 10 or more times resulting in four features. Each of these features are listed in Table 3.3 with a description of the anatomical regions of the nodes involved and the network measure. These functional areas do not necessarily align with anatomical areas, therefore we report the names of the areas from the AAL atlas that had the highest overlap and percentage of the region overlap with that anatomical region. The single cross validation single feature classification rate of each feature is also reported. The best single features achieved about 70% classification accuracy, by combining them together the single cross validation performance was enhanced to more than 80% (as seen in Figure 3.5B). We then looked at frequency of each functional region showing up in the top 4 features. These four features were comprised of five functional regions. The location of the five nodes that had the highest frequency of appearing in the top four features are shown in

Figure 3.5B. These anatomical regions (in order of highest frequency to lowest) were located in the left temporal lobe, right occipital lobe, central portion of bilateral thalami, and left frontal/parietal lobes. These four features included three networks measures: distance, generalized topological overlap, and matching index. Distribution of values of the most informative four features is shown in Supplementary Figure 3.5A.

Since the AAL atlas showed maximal accuracy with only a single feature, and adding more features was detrimental to classification performance, we focused on that single feature. This feature appeared in the top feature set 47 times out of 50 times, which was significantly above chance. This top feature was the matching index between left postcentral gyrus and left thalamus (Table 3.3). Location of these two regions is shown in Figure 3.5C. Distribution of values for this features is shown in Supplementary Figure 3.5B.
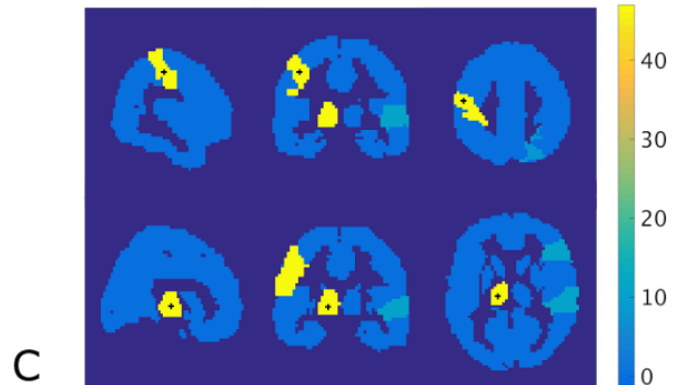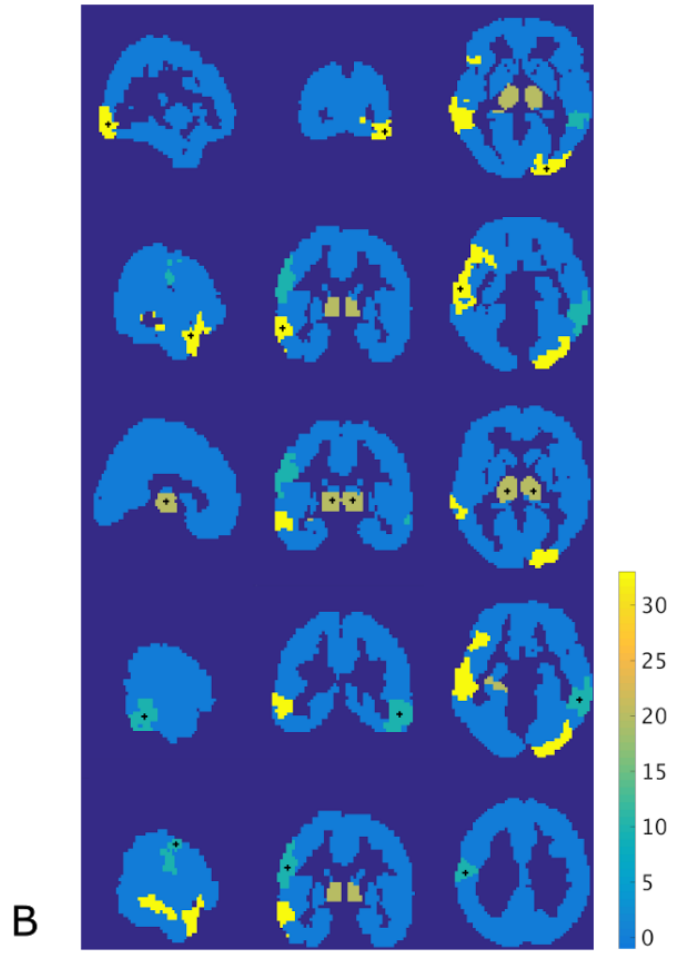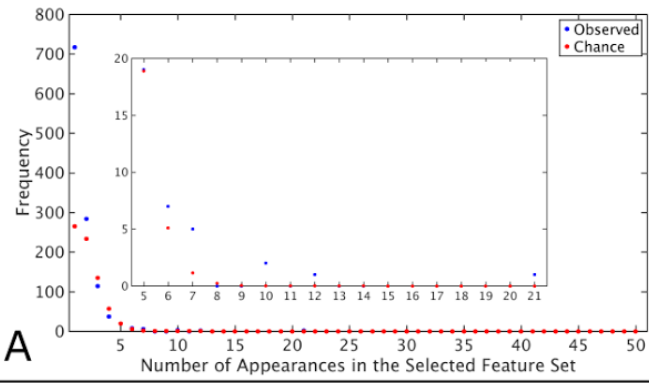
Figure 3.5. Most Informative Features

A) Number of features (y-axis) vs frequency of single features appearing in the 50 selected feature sets (x-axis). The internal figure is a zoomed in version that shows how the most informative features were selected. For example, one feature has appeared 21 times in the selected feature set.

B) Spatial maps showing where the most informative regions are for the functional atlas. The + marks center of the region. Colormap shows number of appearances of the region as a part of one the top 4 features. Left: Sagittal view, Middle: Coronal view, Left: Horizontal view

C) Spatial maps showing where the most informative regions are for the most informative feature of the AAL atlas. Format is the same as B.

Rank	Frequency of Appearance (out of 50)	Name of network measure	First Anatomical Region with the highest overlap (% overlap)	Second Anatomical Region with the highest overlap (% overlap)	d' (p-value)	Single feature classification
			Functional atlas			
1	21	Distance	Left Middle Temporal Gyrus(53), Left Superior temporal Pole(19)	Right Inferior Occipital Cortex(39), Right Lingual Gyrus(17)	0.87 (<0.001)	69%
2	12	Generalized Topological Overlap	Left Middle Temporal Gyrus(53), Left Superior Temporal Pole(19)	Right Inferior Occipital Cortex(39), Right Lingual Gyrus(17)	-0.88 (<0.001)	69%
3	10	Matching Index	Right Inferior Temporal Gyrus(50), Right Middle Temporal Gyrus(48)	Left Thalamus(49), Right Thalamus(43)	0.74 (<0.001)	67%
4	10	Matching Index	Left Postcentral Gyrus(72), Left Precentral Gyrus(10)	Left Thalamus(49), Right Thalamus(43)	0.72 (<0.001)	68%
			AAL			
1	47	Matching Index	Left Postcentral Gyrus	Left Thalamus	0.94 (<0.001)	73%

Table 3.3. List of the most informative features. From left, column 1: Rank of the feature in terms of frequency of appearing in the selected feature set; column 2: Number of appearances in the selected feature set; column 3: Name of network measure, all measures characterize relationship between pairs of nodes (regions); column 4: In order to get an idea about where the most informative regions are, two anatomical regions from the AAL atlas that had the highest overlap with the region are listed; column 5: Same as column 2 for the second region; 6th column: Sensitivity index, also known as d' (equation 3.2) and 2 sample t-test p-value comparing distribution of each feature across the control and schizophrenic population. A positive d' value indicates that feature values are higher for the patients compared to control subjects; column 7: Performance of the feature on its own.

$$d' = \frac{\mu_S - \mu_C}{\frac{1}{2}\sqrt{(\sigma_S^2 + \sigma_C^2)}}$$ *Equation 3.2*

Where μ_S and μ_C are mean of the schizophrenic and control groups respectively, σ_S

and σ_C are standard deviation of the schizophrenic and control groups.

We also looked at the pattern of misclassification across subjects. We calculated the misclassification rate for each subject for both functional and AAL atlases (Figure 3.6A). Misclassification rate for the functional atlas was fairly uniform. However, misclassification rate of the subjects using AAL atlas was bimodal. Some subjects were misclassified correctly more than 60% of the times, while others were misclassified less than 20% of the time. To understand the characteristics of the classifiable and unclassifiable group, we looked at percentage of control and schizophrenic subjects in each group (Figure 3.6B). We further looked at percentage of chronic and first episode schizophrenic patients in each group. We did not observe any trend in the classifiable and unclassifiable subjects.

One possible reason for misclassification of the schizophrenic subjects as controls is the mild severity of their symptoms, which causes their functional activity to resemble that of healthy subjects. To investigate this possibility we plotted composite SANS and SAPS score of the patients vs. misclassification rate (Supplementary Figure 3.6). Further, we plotted each category of the SANS and SAPS scores vs. misclassification rate. We observed no significant correlation between the scores and misclassification rates.
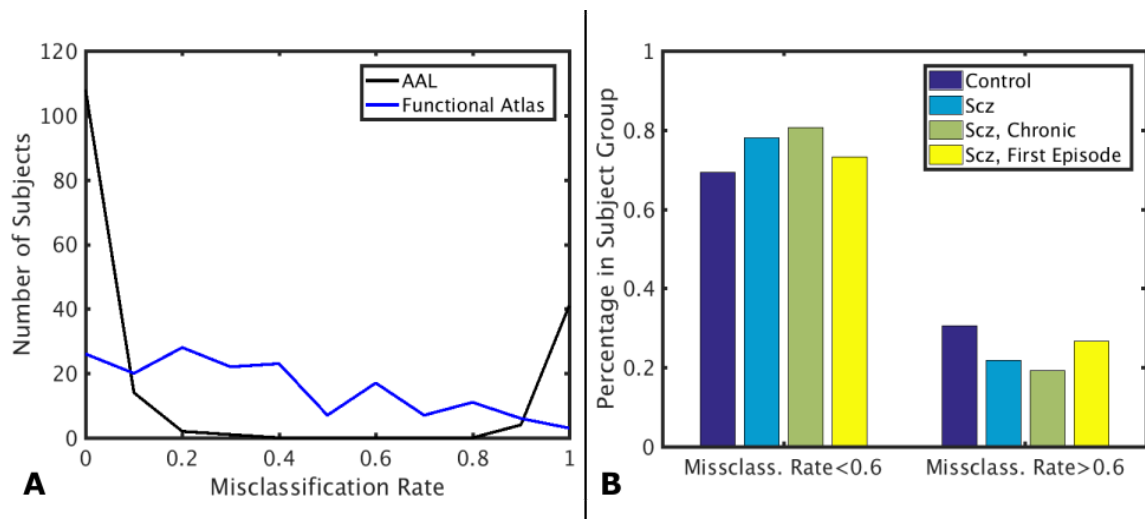
Figure 3.6. Misclassification Rate

A) Distribution of misclassification rate across subjects for the AAL and functional atlases

B) Percentage of subjects in each subject group with low and high misclassification rate using the AAL atlas. atlas. X-axis: Subjects that are highly classifiable (misclassification rate < 0.6) and unclassifiable (misclassification rate > 0.6). Y-axis: Percentage of subjects in each subject group.

One possible reason for the difference in performance of the single and double cross validation performances is that the selected features do not generalize to another group of subjects. However, another possibility is the stochastic nature of the functional activity of the brain, and consequently the structure of the functional networks that are constructed using the functional activity. To tease apart between these two possibilities, we took 42 of our subjects (24 control, 14 schizophrenic) for which we had two scans, taken six months apart. We used the dataset from the first scan for feature selection and classifier training. We then used the resultant classifier and feature set to classify the second dataset from the same subjects. This procedure performs the second cross validation across datasets instead of across subjects. Performance of the double cross validation across datasets is shown in Figure 3.7. The results show that stochasticity present in the functional activity is causing the poor generalizability of our classification process.
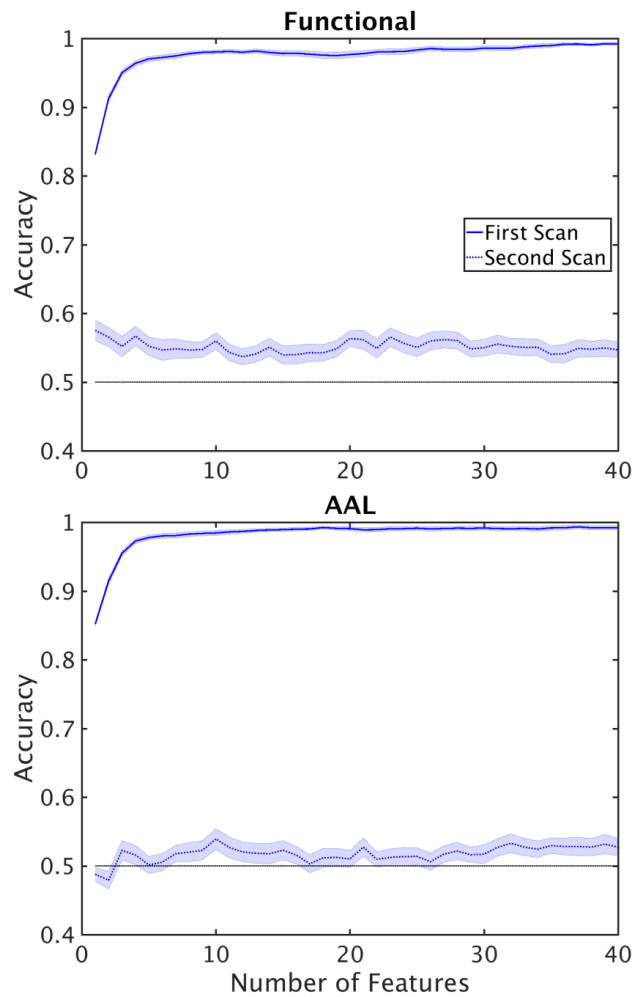
Figure 3.7. Performance of the Classifier on Second Scans

Performance of the classifier when classifier optimization is optimized using the first scan and tested on second scans for functional (top) and AAL (bottom) atlases. Horizontal lines mark chance level.

Discussion

In this study we developed and tested a classification pipeline to discriminate schizophrenic patients and healthy controls. We used pre-whitened BOLD time series to

construct a network model of the brain, using both the AAL anatomical atlas and our functional atlas to define nodes of the network. We extracted multivariate graph theoretic measures and used them as features for classification of the subjects using linear SVM. Measures that were most informative about state of the disease were identified as biomarkers for schizophrenia. We adopted a double cross validation scheme to identify the most informative features. The highest classification accuracy was 72% using the AAL atlas with a single feature: the matching index between left postcentral gyrus. Adding any other features decreased accuracy. Moreover, pre-whitening of the time series significantly improves classification performance in double cross validation. A subset of the subjects, including both healthy and schizophrenic subjects, were misclassified more than 80% of the time. However, no significant correlation was found between misclassification rate of the patients and the severity of their symptoms. Classification accuracy did not improve using a functional atlas. Presumably, this is because duration of resting state activity of was not long enough to robustly capture functional structure of the brain [Chapter 2]. We also observed a significant decrease in performance from single cross validation to double cross validation, except for the single most informative feature from the AAL atlas. Even though we obtained a significantly above chance accuracy, the fairly high false positive and false negative rates means this method does not approach the necessary performance to be useful clinically, particularly if the goal is to identify prodromal state in at risk patient population.

Using machine learning techniques for biomarker identification using fMRI datasets has been extensively explored (see (Zarogianni, Moorhead, and Lawrie 2013; Sundermann et al. 2014; Kambeitz et al. 2015)). A summary of previous work in this area is provided in Supplementary Table 3.2. Our study builds upon existing work in the following aspects: i) we used a large cohort of subjects, ii) we performed double cross validation, iii) we pre-whitened the time series prior to construction of the network.

Machine learning techniques have also been used on other imaging modalities to identify biomarkers for schizophrenia (see (Kambeitz et al. 2015; Zarogianni, Moorhead, and Lawrie 2013) for a review). Several studies have used structural T1 weighted MR images (e.g. (Iwabuchi, Liddle, and Palaniyappan 2013), 77% accuracy, single cross validation, (Nieuwenhuis et al. 2012), 70.4%, double cross validation) and Diffusion

Tensor Imaging (DTI) (e.g. (Ingalhalikar et al. 2010) , 90.6% accuracy, using single cross validation).

Typical datasets used to identify schizophrenia biomarkers have used small populations of several dozen subjects that are to be classified in a high dimensional space several orders of magnitude higher than number of subjects. This phenomenon is called the curse of dimensionality (Jain, Duin, and Mao 2000). In the high dimensional space, the classifier picks up on subtle variations that are specific to the subject set used for training the classifier, which generalizes poorly to unseen data. For robustness it is necessary to reduce dimensionality of the dataset before classification. Two common approaches to dimensionality reduction are commonly used. The first approach is to combine the existing features to construct a smaller set of new features, such as with Principal Component Analysis (PCA), and Linear Discriminant Analysis (LDA) (Bishop 2006). These methods generate a set of linear weights to existing features to construct new features. The second approach is to select a subset of features that carry more information pertaining to classification. A common method in this category is to test the performance of each feature independently and select the subset of feature whose performance is best. The second approach is typically more time consuming than the first approach, but the selected features are directly mapped onto the calculated features, unlike the features made from linear combinations of all the features, as is done with PCA and LDA.

Here we used the forward sequential selection algorithm (Guyon and Elisseeff 2003) to reduce dimensionality of the dataset. We compared the results to two other dimensionality reduction methods, LDA and selection of features with top classification accuracy on their own (Supplementary Figure 3.3). Both methods underperformed with respect to the forward sequential selection method. The forward sequential selection algorithm is more computationally expensive than the other two methods, but its major advantage is that it reduces the redundancy present in the dataset. More specifically, a fair level of correlation has been observed with network level characteristics of the brain (Lynall et al. 2010). Features that have correlation with each other, carry the same information, and are bound to have similar performances when used independently. But their combination does not result in higher performance because due to the correlation.

Therefore, use of more complex feature selection methods, despite their computational expense, is beneficial for classification studies.

The considerable degradation of performance from single to double cross validation demonstrates the importance of testing the final performance with double cross validation when feature selection is an important aspect of the optimization. Reported results based on single cross validation are overly optimistic for out of sample data (Sundermann et al. 2014). In fact, simulations have shown that that even when two classes of data points are generated from the same distribution (i.e. there is no meaningful difference between the two classes), single cross validation is biased towards above chance classification accuracy (Simon et al. 2003). Double cross validation prevents the classifier from overfitting to the dataset that is used for biomarker discovery. Double cross validation has been employed in several studies for example study of schizophrenia using fMRI (see Supplementary Table 3.2), schizophrenia using T1 weighted structural images (Koutsouleris et al. 2015; Nieuwenhuis et al. 2012), depressive disorder using fMRI (Rosa et al. 2015), and autism using fMRI (J. S. Anderson et al. 2011). However, the majority of studies have not performed double cross validation (Supplementary Table 3.2), presumably due to limited sample size (Sundermann et al. 2014). Our results, directly comparing single and double classification performances, supports our hypothesis that single cross validation reports overly inflated accuracy rates. As previously suggested by others in brain imaging (Sundermann et al. 2014) and genetic (Simon et al. 2003) biomarker identification fields, we propose adoption of double cross validation as a standard paradigm for biomarker discovery using brain imaging datasets.

We also performed double cross validation across datasets and not subjects, but did not observe any significant improvement. This shows that inherent stochasticity in fMRI datasets that can be caused by cognitive state of the subject during the scan poses a serious challenge in generalizability of the results. An important remedy to this problem is longer scan durations, scatter across several sessions. Typical scan duration for classification studies of schizophrenia has been between 5 to 10 minutes (Supplementary Table 3.2).

We also explored the effect of the atlas used for defining nodes in the classification performance by comparing the AAL anatomical atlas with a functional atlas constructed from our dataset. The classification accuracy was higher using the AAL atlas. This observation does not necessarily mean that functional anatomical atlases are superior to functional atlases. Extensive evaluation of our functional parcellation algorithm [Chapter 2] concluded that our dataset was not long enough for construction of a robust functional atlas. Previous studies have concluded that minimum duration of resting state activity required for construction of a functional atlas that is replicable across different datasets from the same group of subjects is approximately 27 minutes (Laumann et al. 2015), which is more than four times the duration we used (6 minutes) for construction of the functional atlas. While usage of anatomical atlases for classification studies remains the norm, a few studies have used atlases constructed using data driven algorithms and DTI datasets (Hu et al. 2013; Wang et al. 2016). While we observed 9% decrease in maximum double cross validation accuracy when using a functional atlas instead of an anatomical atlas, (Wang et al. 2016) observed a ~10% increase in single cross validation accuracy when using their data driven atlas instead of an anatomical atlas. Several studies have used Independent Component Analysis (ICA) to produce parcellations (Supplementary Table 3.2). However, ICA does not produce contiguous regions, rather functional networks, comprising of multiple regions. A parcellation with contiguous regions is more straightforward to interpret. Moreover, a parcellation with contiguous regions makes it easier to localize the biomarker to a brain region that is impacted by schizophrenia. If a single region within a functional network is implicated in the disease, the entire network will be implicated using a network based parcellation, which includes regions that are not affected by the disease.

We used multivariate network level measures as classification features in this study, including a mixture of global measures as well as measures that characterize single regions or pairwise statistics. Type of features extracted from resting state fMRI datasets and used for classification varies across studies. One of the most common features is the pairwise correlation coefficient between average time series from different brain regions (Shen et al. 2010; Venkataraman et al. 2012; Tang et al. 2012; Guo et al. 2013; Yu et al. 2013; Su et al. 2013; Kim et al. 2016). This bivariate feature however fails to pick up on more sophisticated motifs in the functional structure of the brain. Network

measures, being multivariate, are capable of identifying more complex patterns in group differences and have been used in several classification studies (Bassett et al. 2012; Fekete et al. 2013; A. Anderson and Cohen 2013; Singh and Bagler 2016). However these studies either use global networks measures (Bassett et al. 2012; Fekete et al. 2013; Anderson and Cohen 2013), or use average and standard deviation of local measures (Singh and Bagler 2016), which eliminates spatial information about the most discriminating regions. Our data driven greedy feature selection method preserves this information.

We observed that a single feature produced maximum classification accuracy, using the AAL atlas. As more features were used for classification, single cross validation accuracy increased but double cross validation accuracy decreased. This indicates that the added features did not generalize well across subjects. Their addition to the feature set caused the classifier to put some weight on other features, diluting useful information. Using the functional atlas, we found four features whose appearance in the selected feature set was statistically meaningful. The reduction form a 19000+ feature space to a few features, reveals the tremendous redundancy inherent to the dataset. Similar to our results, (Fan et al. 2011) obtained a 85% double cross validation accuracy using seven features. In another study, (Tang et al. 2012) got a 93.2% double cross validation accuracy using 550 features.

We observed that prewhitening of the time series increased classification performance. Similar observation was made when fMRI datasets were used to classify Post traumatic Stress Disorder (PTSD) patients from controls (Peka Christova et al. 2015). In contrast to our results, (Arbabshirani et al. 2014) did not observe any difference between discriminability of prewhitened and raw time series in a cohort of healthy and schizophrenic subjects. However, that study compared bivariate measures across the groups, whereas we used multivariate measures, which are capable of picking up on more complex differences between the groups.

Of the 25 different network measures used to generate features (Supplementary Table 3.1), all of the five most informative features came from three measures: distance, generalized topological overlap, and matching index. All of the top five features were

from pairwise network measures, none were from the entire network or nodal metrics. Each measure and the observed trends associated with it are discussed more in detail.

The first feature selected by the functional atlas was the distance between two regions in the left temporal and right occipital lobes. Distance between two nodes is the shortest path between them in a binary graph (Rubinov and Sporns 2010b). As reported in Table 3.2 and Supplementary Figure 3.5, the distance between two regions in the right occipital lobe and left temporal lobe is lower in the control group compared to schizophrenic patients. Interestingly, the distance between these two nodes in majority of control subjects is 1, meaning that the two regions are connected to each other directly via a single link. Distance between these same two nodes is 2 between majority of schizophrenic subjects, which means the direct link between the two nodes is absent in patient group, showing a hypoconnectivity between these two regions. Changes in volume of the left middle temporal gyrus in schizophrenic patients has previously been reported (Onitsuka et al. 2004; M. Hu et al. 2013). Moreover, the middle temporal gyrus has been implicated in other fMRI classification studies (Castro et al. 2011; Yang et al. 2010), albeit bilaterally. Disruption in functional activity of the right inferior occipital gyrus has been reported in another study (Castro et al. 2011).

The second most informative feature generated using the functional atlas was the generalized topological overlap between the same two regions (Table 3.2). Generalized topological overlap quantifies the extent to which a pair of nodes have similar m-th step neighbors in binary graphs (Rubinov and Sporns 2010b). The m-th step neighbors of a node are all the nodes in the binary graph that are reachable through a maximum of m steps. We observed that generalized topological overlap between regions in the right occipital lobe and left temporal lobe is higher in the control group compared to patients (Supplementary Figure 3.5). This shows that functional connectivity pattern between these two regions diverges from each other in the patient group.

The third most informative feature using the functional atlas was the matching index between two regions in the right temporal lobe and the thalamai (Table 3.2). Matching index between two nodes quantifies the similarity between their functional connectivity profiles based on the number of common neighbors between the two nodes and is applicable to binary graphs (Rubinov and Sporns 2010b). We observed increased

matching index between regions in the right temporal gyrus and bilateral thalamai in the schizophrenic group compared to controls (Table 3.2 and Supplementary Figure 3.5). The first region overlapped with the right inferior and middle temporal gyri. The other region overlapped with the ventral portion of bilateral thalamai. As discussed before, middle temporal gyrus has been implicated in schizophrenia in other classification studies (Castro et al. 2011; Yang et al. 2010). Disruption of functional connectivity of thalamus has also been found in several other studies (Skudlarski et al. 2010; Atluri et al. 2015; Kim et al. 2016).

Increased matching index in the schizophrenic group was also observed between another pair of regions, the fourth most informative measure using functional atlas (Table 3.2, Supplementary Figure 3.5). The first region overlapped with both postcentral and precentral gyri, and the second region overlapped with the ventral portion of bilateral thalamai. Postcentral gyrus has been implicated in schizophrenia in several other studies (Yang et al. 2010; Castro et al. 2011; Rashid et al. 2016). Interestingly, another study reported that functional connectivity between the left postcentral gyrus and right thalamus was different across the healthy and schizophrenic group (Kim et al. 2016).

The single feature that produced maximum classification accuracy using the AAL atlas also indicated an increase in matching index between the postcentral gyrus and left thalamus (Table 3.2 and Supplementary Figure 3.5). Matching index between the left thalamus and left postcentral gyrus was lower in the control group than the Schizophrenic group, consistent with the fourth most informative feature using the functional atlas.

Future Directions

Several improvements can be done to increase classification performance to approach clinically useful values. First, acquiring longer durations of functional activity results in more robust functional networks that can enhance performance, especially decreasing the gap between single and double cross validation results. The second improvement may be to use more robust brain atlases. Anatomical atlases are based on physical landmarks of the brain, while our functional atlas was constructed using 6 minutes of resting state activity, which might not be enough to capture functional organization of the

brain. Recent effort to construct brain atlases using multi-modal datasets such as combining fMRI with myelin maps (Glasser et al. 2016) are promising. Third, more robust biomarkers can be developed by using of multi-modal feature sets, by combining feature extracted from different modalities such as T1 weighted images, fMRI, and DTI (e.g. (Silva et al. 2014)). The feature set can further be supplemented with non-brain related datasets such as genetic biomarkers (e.g. (Yang et al. 2010)). Fourth, medication load could be a confounding factor that we could not adequately account for. Unfortunately, the study of unmedicated schizophrenic patients is not practical, except for patients that are at the onset of the disease and are antipsychotic naive. More recordings from schizophrenic patient population at the onset of their disease could provide valuable insight biomarkers.

Conclusion

We used a relatively large fMRI dataset to classify schizophrenic patients from healthy subjects using network measures as features fed into an SVM classifier, implementing double cross validation to validate the classifier. We compared classification accuracy of the results when a functional atlas and an anatomical atlas were used to calculate the network measures. We observed that the AAL atlas had a higher performance than the functional atlas, although both atlases produced above chance performance. We also observed that prewhitening of the fMRI time series improves classification results. Yet another important observation was the significant difference between result of single and double cross validation.

Supplementary Material

Supplementary Tables

Measure Number	Measure Name	Graph type	Measure type	Extra Parameters
1	Number of Communities	Weighted	Global	We used the Louvain community detection algorithm.
2	Average Community Size	Weighted	Global	
3	Transitivity	Weighted	Global	
4	Assortativity	Weighted	Global	
5	Rich Club Curve	Binary	Global	Calculated for a range of degrees, from 1 to the average node degree of the network .
6	Characteristic Path Length	Binary	Global	
7	Global Efficiency	Weighted	Global	
8	Radius of Graph	Binary	Global	
9	Diameter of graph	Binary	Global	
10	Node Betweenness Centrality	Weighted	Node	
11	Eigenvector Centrality	Weighted	Node	
12	Shannon Entropy	Weighted	Node	

13	Within Module Degree Centrality	Weighted	Node	
14	Participation Coefficient	Weighted	Node	
15	Clustering Coefficient	Weighted	Node	
16	Node Degree	Weighted	Node	
17	Strength	Weighted	Node	
18	Eccentricity	Binary	Node	
19	Local Efficiency	Weighted	Node	
20	Reachability	Binary	Pair	
21	Distance	Binary	Pair	
22	Weighted Distance	Weighted	Pair	
23	Overlap Amongst Neighbors	Weighted	Pair	
24	Generalized Topological Overlap	Binary	Pair	Calculated up to 3rd step neighbors.
25	Matching Index	Binary	Pair	

Supplementary Table 3.1. List of graph theoretic measures used for classification.
From left to right: column 1, measure number; 2, measure name (Bullmore and Sporns 2009; Rubinov and Sporns 2010b); column 3, type of graph used for calculating the measure. Some measures are specific to binary graphs. To calculate these measures the weighted graph was thresholded and converted to binary; column 4, type of measure. For a graph with N nodes, global measures characterize the entire network and produce one value. Measure that characterize nodes produce N values. Measure that characterize measure pairs produce $N.(N-1)/2$ values; column 5, Some measure require extra parameters to be calculated. Value of the parameters are specified in this column.

Dataset	Number of Subjects	Features	Dimensionality Reduction Method(s)	Classifier	Double Cross Validation	Accuracy	Ref.
fMRI, Oddball Auditory Task	68, Scz = 34, H = 34	Activation Map	LDA + mRMR	SVM	No	88%	(Juneja, Rana, and Agrawal 2014)
fMRI, Category Exemplar Word Pair Task	25, Scz = 15, H = 10	Activation Map	PCA	LDA	No	>80%	(Ford et al. 2003)
fMRI, AX-Continuous performance test	102, Scz = 51, H = 51	Behavioral Performance	None	LDA	No	58%	(Yoon et al. 2012)
		Activation contrast between A and B cues in DLPFC				62%	
		Activation contrast between A and B cues in the entire brain				59%	
fMRI, resting state, 6 minutes	83, Scz = 48, H = 35	Regional Homogeneity, regions defined anatomically	PCA	LDA	No	80%	(Shi et al. 2007)
		Voxelwise homogeneity				74%	
fMRI, Resting State, 6 minutes	52, Scz = 32, H = 20	Pairwise Correlation Coefficient between anatomically defined regions	Kendall tau rank correlation coefficient feature selection + locally linear embedding manifold learning	K-means clustering	No	87%	(H. Shen et al. 2010)
fMRI, Auditory Oddball task and Genetic data	40, Scz = 20, H = 20	Genetic data, Activation map and ICA map	Forward Sequential feature selection method	SVM	No	87%	(Yang et al. 2010)

|							
fMRI, verbal fluency task	104, Scz = 32, H = 40, Bipolar = 32	Activation map	None	SVM, three way classification	No	92%, Scz vs non-Scz	(Costafr eda et al. 2011)
fMRI, resting state, 6 minutes	58, Scz = 29, H = 29	Size of the largest connected component	None	SVM	No	75%	(Basset t et al. 2012)
fMRI, resting state	68, Svz = 40, H = 28	Lattice Auto-Associat ive Memories	None	k-NN	No	~83%	(Chyzh yk and Graña 2015)
fMRI, resting state, >10 minutes	36, Scz = 18, H = 18	Pairwise regional functional connectivity between anatomically defined regions	Decision tree feature selection	Decision tree	No	75%	(Venkat araman et al. 2012)
fMRI, resting state, 6 minutes	146, Scz = 72, H = 74	Voxel-wise regional homogeneity, amplitude of low fluctuations, and functional homotopy	Feature selection based on class distance	SVM and random forests	No	80%	(Savio and Graña 2015)
fMRI, resting state, 10 minutes	20, Scz = 10, H = 10	Fine Granularity Functional Interaction between subnetworks	Separability threshold + PCA	SVM	No	77.5%	(X. Hu et al. 2013)
		Functional connectivity				77.5%	
		Fine Granularity Functional Interaction between subnetworks + Functional connectivity				95%	
fMRI, resting state, 6 minutes	71, Scz = 24, H = 22, Healthy Siblings of Scz subjects = 25	Pairwise Correlation Coefficient between anatomically defined regions	PCA	SVM, 3 way classificati on between schizophr enic patients, their healthy siblings, and	No	62%	(Yu et al. 2013)

					healthy controls		
fMRI, resting state, >5 minutes	370, Scz = 195, H = 175	Pairwise Correlation Coefficient between networks defined by ICA	mRMR	SVM	Yes	84%	(Arbabs hirani, Castro, and Calhou n 2014)
		Auto connectivity of network time series				80%	
		Pairwise Correlation Coefficient between networks defined by ICA +auto connectivity of network time series				88%	
fMRI, 0- and 2-back memory task	37, Scz = 17, H = 20	Activation Map	Searchlight Based feature Extraction	SVM	Yes	91	(Bleich-Cohen et al. 2014)
fMRI, resting state, 6 minutes	64, Scz = 32, H = 32	Pairwise Pearson Correlation Coefficient between anatomically defined regions	Thresholded Kendall-Tau coefficient	SVM, linear	No	81.2%	(Su et al. 2013)
		Maximal Information Coefficient				76.6%	
		Extended Maximal Information Coefficient				82.8%	
fMRI, resting state, 6 minutes	49, Scz = 24, H = 25	Pairwise Correlation Coefficient Over Time between Regions of interest determined by Task Activity	Threshold based on t-test	SVM, Polynomi al Kernel	Yes	81.3%	(H. Shen et al. 2014)
				SVM, Linear		73.5%	
				SVM, Gaussian		79.6%	
fMRI, resting state, 5	100, Scz = 50, H = 50	Pairwise Correlation Coefficient	None	Deep Neural Network	Yes	85%	(Kim et al. 2016)

minutes		between anatomically defined Regions		SVM, Linear		77%	
fMRI, resting state, 10 minutes	18, Scz = 8, H = 10	Network Level Measures using nodes defined by the AAL atlas, constructed multiple graphs	Threshold based on t-test + Recursive Feature Elimination	SVM, Block Diagonal Optimizati on with Spherical Kernels	Yes	96%	(Fekete et al. 2013)
fMRI, resting state, 5 minutes	56, Scz = 28, H = 28	Z-map of spatial components identified using ICA	T-test + PCA + LDA	Nearest Neighbors	No	93%	(Du et al. 2012)
fMRI, Auditory Oddball task						98%	
fMRI, resting state, 6 minutes	146, Scz = 72, H = 74	Network measures based on networks identified by ICA	None	SVM, Radial basis kernel	No	65%	(A. Anders on and Cohen 2013)
fMRI, resting state, 5 minutes	56, Scz = 28, H = 28	Functional connectivity between networks identified by ICA	None	SVM, linear	No	83%	(Arbabs hirani et al. 2013)
				SVM, Radial basis and polynomia l kernels		96%	
				Decision trees		96%	
				K nearest neighbor classifier		96%	
fMRI, resting state, >5 minutes	159, Scz = 60, H = 61, Bipolar = 38	Functional connectivity between networks identified by ICA	Double input symmetric relevance (DISR)	Three-wa y, SVM, linear	Yes	59%	(Rashid et al. 2016)
		Functional connectivity between networks identified by ICA over time				84%	
		Combination				89%	

		of the above feature sets					
fMRI, Auditory oddball task	52, Scz = 31, H = 21	Gramian between regions identified using ICA, using magnitude and phase information separately	Recursive Feature Elimination	Multiple Kernel Learning	Yes	85%	(Castro et al. 2011)
fMRI, Monetary Incentive Delay Task	98, Scz = 44, H = 54	Activation Map	None	SVM	No	85%	(Koch et al. 2015)
fMRI, resting state, 6 minutes	146, Scz = 72, H = 74	Binary and weighted network measures based on regions identified by ICA	All possible single, pair, triads and tetrads of features	SVM	No	65%	(Singh and Bagler 2016)
fMRI, Auditory odd ball task	70, Scz = 34, H = 36	Activation map	ICA + PCA	Projection pursuit algorithm	No	91%	(Demirci, Clark, and Calhoun 2008)
fMRI, resting state, 6 minutes	62, Scz = 31, H = 31	Functional connectivity pattern identified by ICA	Sequential forward selection feature selection	SVM	Yes	85.5%	(Fan et al. 2011)
fMRI, resting state, 6 minutes	44, Scz = 22, H = 22	Pairwise correlation coefficient between anatomically defined regions	Kendall tau rank correlation coefficient + PCA	SVM	Yes	93.2%	(Tang et al. 2012)
fMRI, resting state, 6 minutes	274, Scz = 152, H = 122	Z-scored pairwise correlation coefficient between specific anatomically delineated regions	None	SVM, Radial basis function kernel	No	73.4%	(Guo et al. 2013)

Supplementary Table 3.2. Summary of Previous Work. A summary of other classification studies of schizophrenia using fMRI datasets. This list is not exhaustive, but a limited survey of literature.
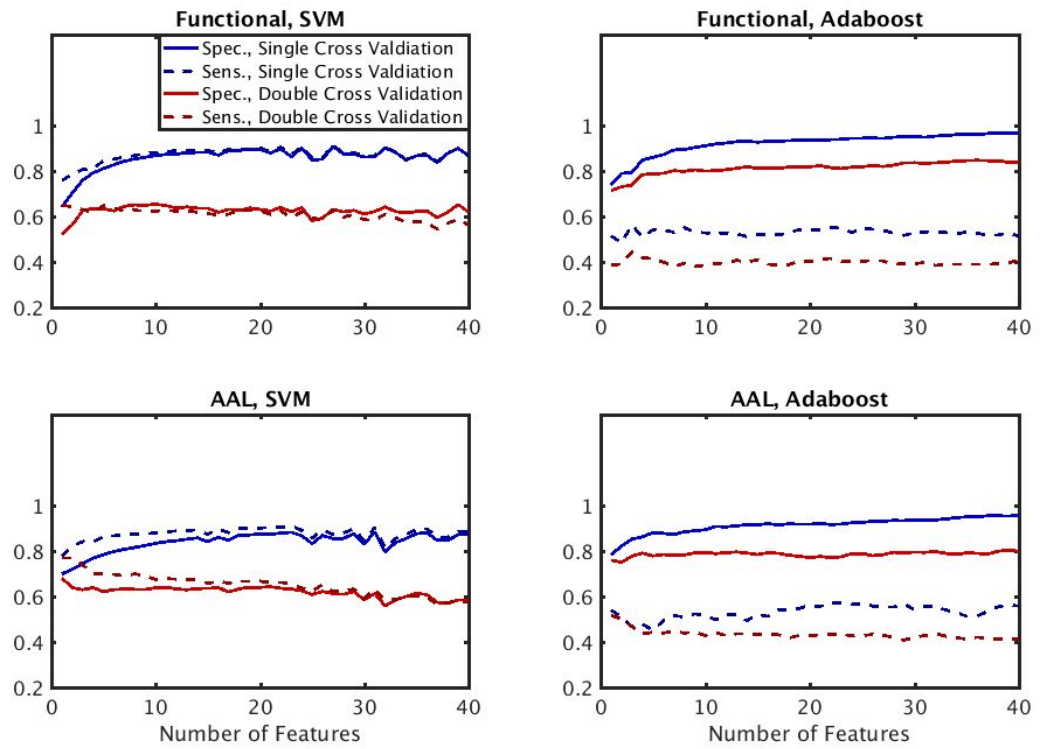
Supplementary Figures

Functional Atlas, Adaboost

AAL, Adaboost

A

B
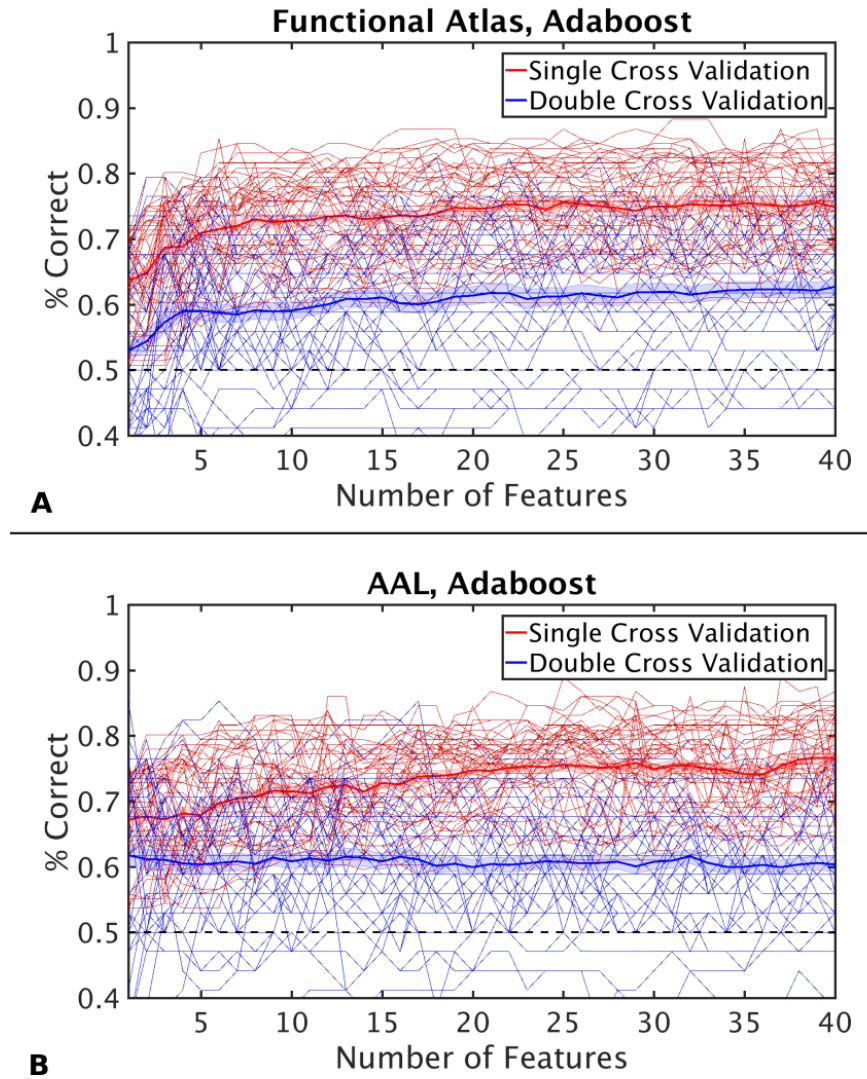
Supplementary Figure 3.1. Performance of Different Classifiers

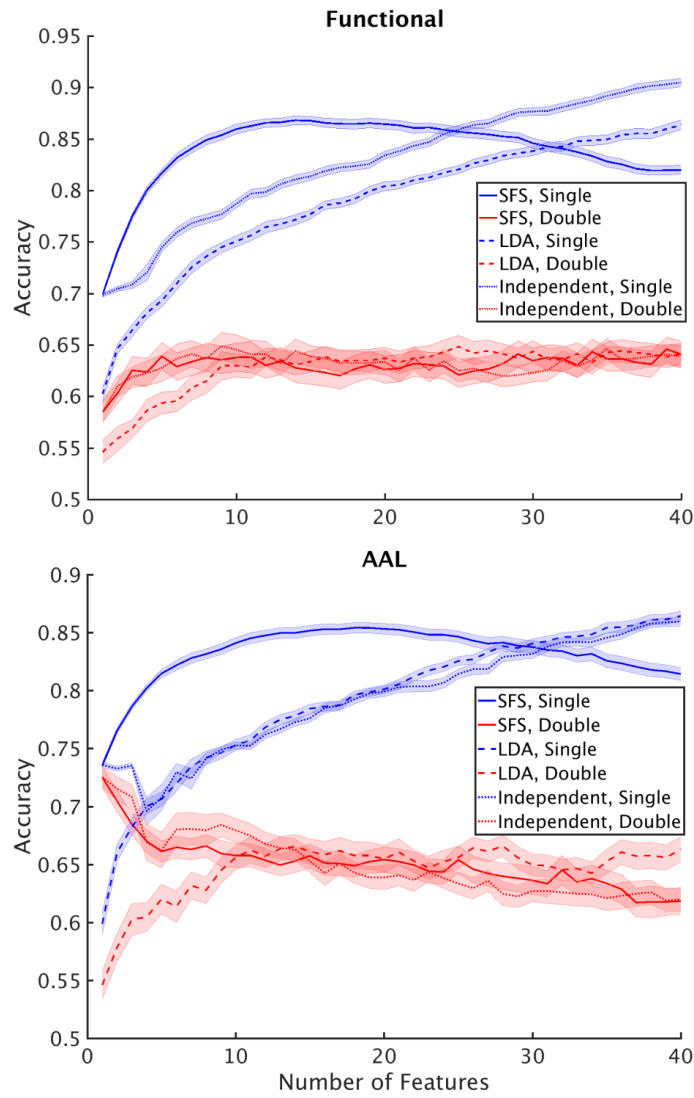Performance as a function of number of features for the adaptive boost (Adaboost) classifiers.

A) Performance of the Adaboost classifier when functional atlas was used to construct the network and extract features.

B) Performance of the Adaboost classifier when the AAL atlas was used to construct the network and extract features. Dotted line marks chance level.

Supplementary Figure 3.2. Sensitivity and Specificity

Sensitivity and Specificity as a function of number of features for functional and AAL atlases and different classifiers.
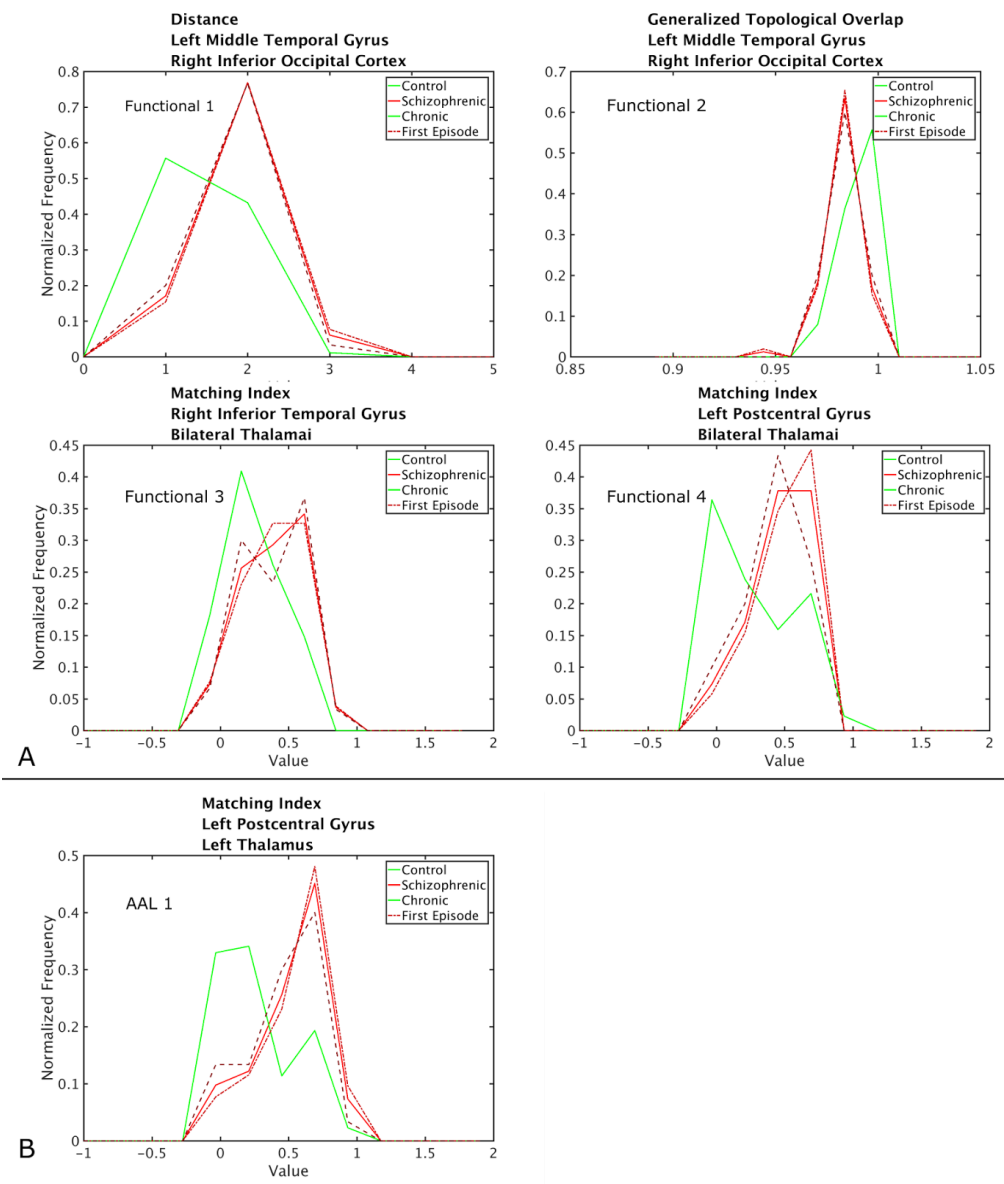
Supplementary Figure 3.3. Feature Selection Method

Classification performance vs. number of features for different dimensional reduction methods for both functional (top) and AAL (bottom) atlases.
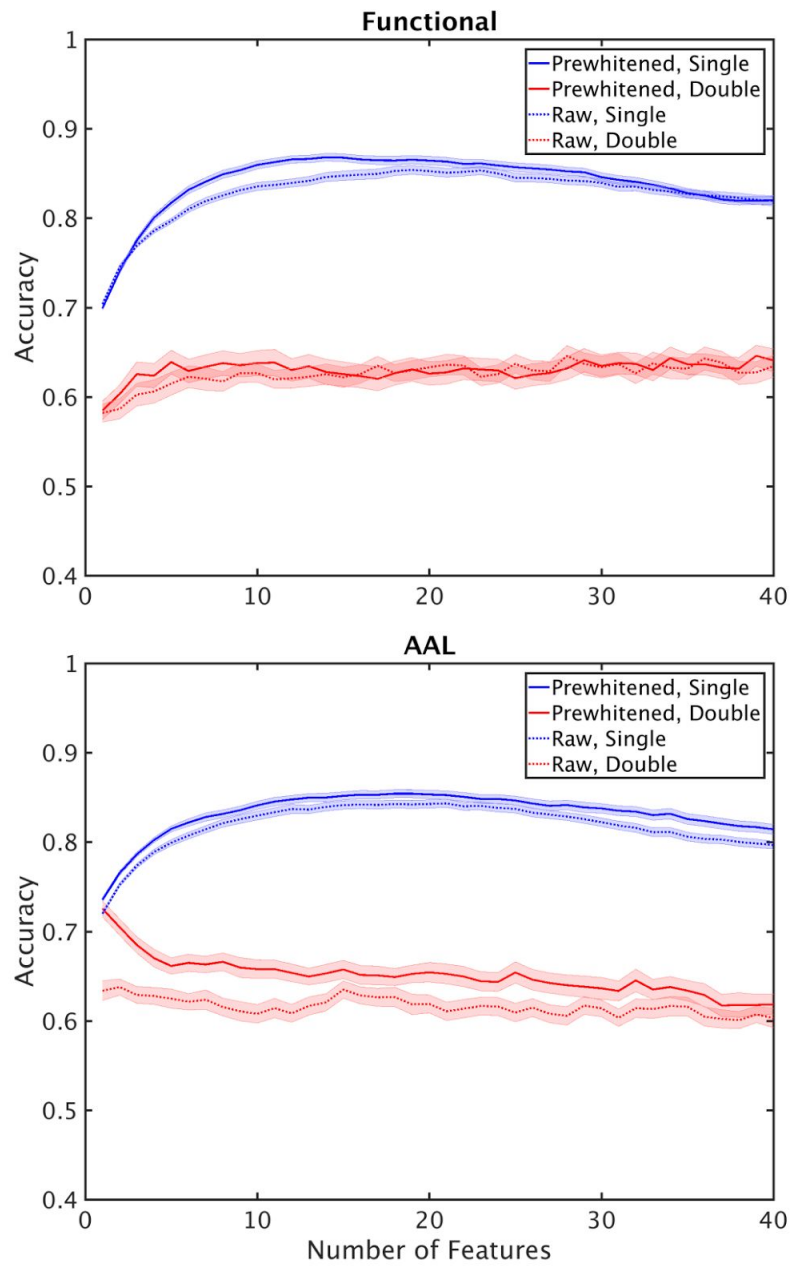
Supplementary Figure 3.4. Raw vs Prewhitened

Classification accuracy vs. number of features when raw and prewhitened time series are used for construction of the network when the functional (top) and AAL (bottom) atlases were used for node definition

Supplementary Figure 3.5. Distribution of Top Features

Distribution of the top features across different subject groups for A) functional and B) AAL atlases.
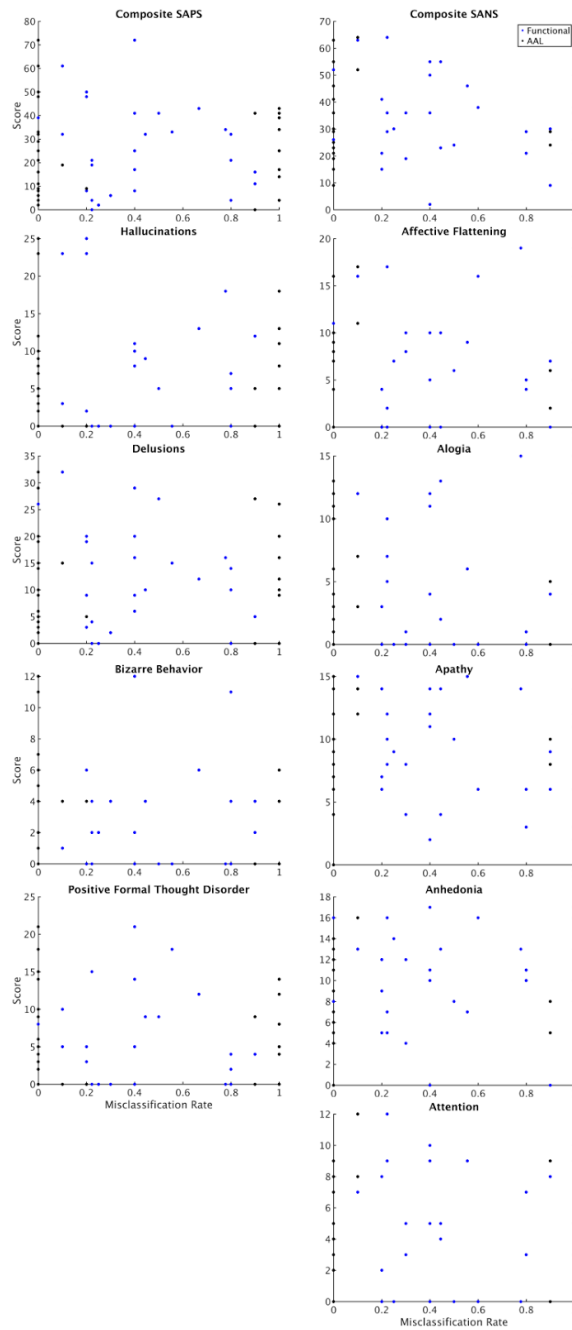
Supplementary Figure 3.6. Misclassification Rate vs. Symptom Severity

SANS and SAPS scores (y-axis) vs. Misclassification rate (x-axis) for the patient group for both atlases.

Chapter 4: Is Schizophrenia in the Eyes of the Beholder?

Introduction

Schizophrenia is a debilitating chronic mental disorder. Schizophrenia is currently diagnosed in based on interviews with the patients with no clinically approved biomarker for the disease. Identification of non-invasive biomarkers for schizophrenia can be very beneficial. Biomarkers can shed light on mechanisms behind the disease and guide clinicians to develop more effective therapeutic interventions. Moreover, such biomarkers can be used to identify prodromal state that could be used used for early interventions before onset of the disease.

The exact cause of schizophrenia is not clear. It has previously been observed that N-methyl-D-Aspartate (NMDA) glutamate antagonist drugs such as phencyclidine or ketamine cause schizophrenia like symptoms in healthy subjects (Javitt and Zukin 1991; Driesen et al. 2013). This has led to development of the NMDA hypofunction hypothesis (also known as glutamate dysfunction hypothesis) of schizophrenia (Olney and Farber 1995; Coyle 1996; Moghaddam and Javitt 2012; Snyder and Gao 2013) which posits that schizophrenia is caused by a disruption in function of NMDA receptors.

NMDA receptors are expressed abundantly by retinal cells (Y. Shen, Liu, and Yang 2006). Therefore, we hypothesize that NMDA dysfunction might manifest as distortions in response of retinal cells to visual stimulation captured by electroretinograms (ERG). ERG is a non-invasive low cost test that is used to test functionality of retina in clinical settings and is typically used for detection of glaucoma (Colotto et al. 2000; Machida 2012). ERG waveform captures activity of rod and cone photoreceptors as well as bipolar and possibly Muller cells in response to a flash of light (Frishman 2012). ERG can also be measured after the subject has been light or dark adapted. Photopic response is the ERG collected after subject has been light adapted and captures response of the cones. Scotopic response reflects activity of the rods and is the ERG response after the subject has been dark adapted. A variation of ERG, called the pattern electroretinogram (pERG) is also used in clinical settings (Preiser et al. 2013). PERG is

invoked activity by a flickering checkerboard or grating pattern and captures activity of the retinal ganglion cells (Miura et al. 2009).

We collected ERG, Photopic, Scotopic, and pERG waveforms from a cohort of schizophrenic individuals and a group of healthy controls (71 subject total). They were used to identify abnormal patterns of activity in schizophrenic subjects that can potentially be used as biomarkers. We employed a data driven approach, where these waveforms were used to classify patients into schizophrenic or control. We analyzed classification power of these waveforms both in time and frequency domains. This resulted in a high dimensional dataset (several hundred dimensions) that was used to classify 71 subjects. Therefore, we used principal component analysis (PCA) (Bishop 2006) in conjunction with the sequential forward selection (Guyon and Elisseeff 2003) to reduce dimensionality of the dataset. The low dimensional dataset was then used for classification using the support vector machine (SVM) classifier (Vapnik 1995; Bishop 2006), since SVM is robust to outliers. To test reliability of the resultant biomarkers, we adopted a double cross validation scheme, where a subset of the subjects were used for biomarker identification. The identified biomarkers were then used to classify the a subset of subjects that did not participate in biomarker identification discovery. Performance of our classification paradigm and the proposed biomarkers are discussed.

Methods

Data collection

A group of 35 healthy control (17 female, age:M = 40, SD = 13) and 36 schizophrenic (17 female, age:M = 41, SD = 11) participants were recruited for this study. Six of these subject were brought back for a second data collection section. The second dataset from those subjects were treated as separate subjects. All participants gave informed consent and were compensated for their participation. All procedures were done in accordance with a University of Minnesota IRB approved protocol.

We used the Diagnosysllc, Boston, MA system to collect four types of signals at a sampling frequency of 1.2Hz: Electroretinogram (ERG), Photopic, Scotopic, and pattern Electroretinogram (pERG). ERG signal was collected using a full field Ganzfeld light flash covering the entire retina at at three different light intensities listed in Table 4.1.

Photopic response was collected after the subject was light adapted for ten minutes. Scotopic response was collected in ten minute dark adapted condition. Two types of pERG were collected, where the stimulus was a checkerboard. First set of pERG were collected at a single stimulus intensity presented on a Crt monitor, referred to as pERG Crt signal in this manuscript. For a subset of subjects, pERG was collected using a checkerboard stimulus presented at a variety of intensities and contrasts on an LED monitor, referred to as pERG LED signal in this manuscript. Since the type of monitor and stimulus intensity and contrast are different for pERG Crt and pERG LED, they were treated as two different signal types. A summary of the stimuli used for collection of each signal is provided in Table 4.1.

The ERG, pERG Crt, and pERG LED were collected by presenting the stimulus 150 times. Scotopic and Photopic responses were collected only for single trials. Each signal type was recorded for each eye separately. The waveforms from the two eyes were concatenated to produce a single waveform per trial per subject (Figure 4.1).

Outlier Detection and Processing

Excluding photopic and scotopic responses that included a single trial), other datasets contained a number of noisy trials with abnormal waveforms. We identified noisy waveforms for each subject by calculating mean and standard deviation of the waveform at each time point across all trials from that subject and excluded waveforms that deviated more than one standard deviation from the mean in O% of data points (Figure 4.1). Choice of O depended on the signal type and was chosen based on trial and error, where M was changed manually and the results were visually inspected until acceptable results were obtained. Choice of M for different signals is summarised in Table 4.1.

If more than 50% of trials from a subject are excluded, that subject is excluded from the study. Waveforms from the remaining subjects were then examined manually. Subject with abnormal waveforms were excluded from further analysis. Number of subjects included for each signal type are also reported in Table 4.1. All the acceptable trials from each subject were averaged to produce a single mean time series per subject. Average of baseline values were then subtracted from each waveform.

Signal type	Number of trials	Stimulus	*O*(%)	Number of included subjects
ERG	150	Step 1: Flash of light at 1 cd.s/m^2 intensity	50	62, Scz=33, H=29
		Step 2: Flash of light at 5 cd.s/m^2 intensity		60, Scz=30, H=30
		Step 3: Flash of light at 7 cd.s/m^2 intensity		58, Scz=29, H=29
pERG Crt	150	Checkerboard pattern at 100 cd.s/m^2 intensity	40	71, Scz=36, H=35
pERG LED	150	Step 1: Checkerboard pattern at 999 cd.s/m^2 intensity, 100% contrast	40	46, Scz=24, H=22
		Step 2: Checkerboard pattern at 470 cd.s/m^2 intensity, 100% contrast		46, Scz=24, H=22
		Step 3: Checkerboard pattern at 275 cd.s/m^2 intensity, 100% contrast		46, Scz=24, H=22
		Step 4: Checkerboard pattern at 999 cd.s/m^2 intensity, 100% contrast		41, Scz=22, H=19
		Step 5: Checkerboard pattern at 470 cd.s/m^2 intensity, 40% contrast		41, Scz=21, H=20
		Step 6: Checkerboard pattern at 275 cd.s/m^2 intensity, 30% contrast		40, Scz=22, H=18
		Step 7: Checkerboard pattern at 999 cd.s/m^2 intensity, 25% contrast		36, Scz=19, H=17
		Step 8: Checkerboard pattern at 470 cd.s/m^2 intensity, 20% contrast		35, Scz=20, H=15
		Step 9: Checkerboard pattern at 275 cd.s/m^2 intensity, 11% contrast		36, Scz=19, H=17

photopic	1	Flash of light at 1 cd.s/m² intensity	100	69, Scz=35, H=34
Scotopic	1	Flash of light at 1 cd.s/m² intensity	100	58, Scz=30, H=28
Combination				16, Scz=7, H=9

Table 4.1. Summary of Dataset. Summary of signal types and corresponding methods of collection and analysis. From left to right, column 1: Signal type; column 2: Number of trials collected; column 3: visual stimulus; column 4: criteria for detection of outlier trials for each signal type. A trial is excluded if it deviates from the average trace more than one standard deviation in O% of time points; column 5: Number o f subjects included in the study after outlier exclusion, Scz denotes subjects diagnosed with schizophrenia, H denotes healthy control subjects.

Feature extraction

In order to classify subjects based on the shape of their retinal response waveforms, distinguishing features must first be extracted. Such features can capture different aspects of the waveform. For this analysis, we extracted features both in time and frequency domain (Figure 4.1). Time domain features were simply amplitude of the waveform at each time point. For the frequency domain analysis we calculated the short time Fourier transform of each waveform in 30s windows with 50% overlap. Total power and phase in each frequency and time bin, were used as features.

Each feature constitutes one dimension in the dataset. Total number of dimensions is one order of magnitude higher than number of subject that are to be classified. To reduce dimensionality of the dataset we performed Principal Component Analysis (PCA). PCA applies a linear transformation to the dataset and transforms it into a lower dimensional space where each dimension is aligned with direction of maximum variance in the dataset (Figure 4.1). The principal components (PCs) extracted using PCA were then used for classification of the subjects.

Classification

We used a double cross validation algorithm to choose the most informative PCs and use them for classification of the subjects into healthy and Schizophrenic. We randomly divided the subjects into 3 subsets: training, validation, and testing using 5 fold cross validation partitioning. The subject set was partitioned into 5 equal sized subsets or folds. Three folds were used as training, one fold for validation, and one fold for testing.

Training and validation subsets were used for choosing the most informative PCs using the sequential forward selection algorithm (Guyon and Elisseeff 2003) (Figure 4.1). First, a linear support vector machine (SVM) classifier was trained on the training subset using each PC independently. The trained SVM was then used to classify the validation subset and classification accuracy was calculated. The PC with highest classification accuracy was paired with all the remaining PCs. Classification accuracy of each pair was similarly calculated by training a linear SVM on the training subset and using it to classify the validation subset. Pair of PCs with the highest classification accuracy was then combined with the remaining PCs to form triplets. This process was progressively performed until a set of 40 PCs was selected.

After determining the most informative 40 PCs, they were used to train a linear SVM on the combination of training and validation subsets. The resultant classifier was then used to classify the testing subset and calculating classification accuracy. The box constraint or C value used for training the SVM was equal to 10.

This process was repeated five times, so that each fold was once used once as the testing set. The random partitioning was also performed 10 times. This resulted in 50 performance accuracies and 50 sets of 40 most informative PCs.

We repeated this process for each waveform separately. We also concatenated all the waveforms to construct a combination waveform and used this long waveform for classification.
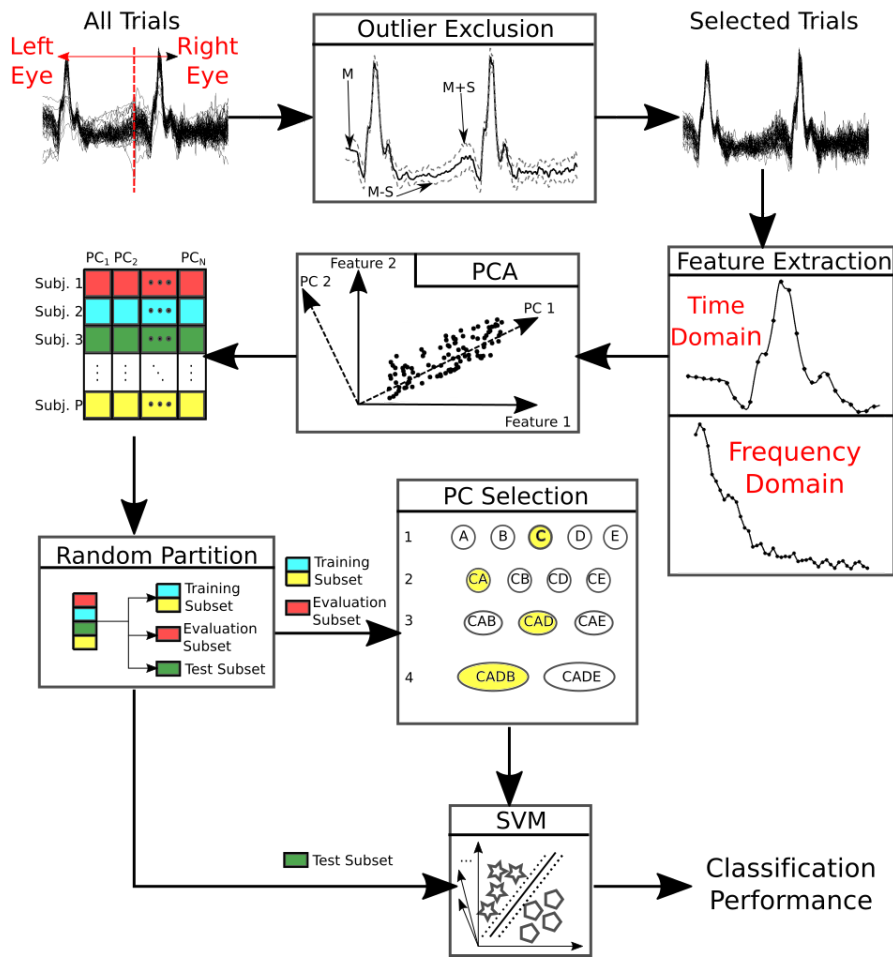
Figure 4.1. Classification Pipeline

Trials from left and right eyes are concatenated. For each subject, Trials that are not within one standard deviation of the mean for a certain percentage of time points, are excluded. Remaining trials are averaged for each subject. Time or frequency domain features are extracted from the average. Principal component analysis (PCA) is then performed to reduce dimensionality of the data. Entire subject set is then randomly partitioned into three groups. Training and evaluation subsets are used to select the most informative principal components (PCs), using the forward sequential selection algorithm. Selected PCs are then used to classify the test subjects. Classification performance is reported.

Results

Four types of retinal response with varying visual stimuli were collected from a cohort of schizophrenic and healthy subjects. Example traces of each signal type are shown in Figure 4.2.

Two types of features were extracted from each signal type, time domain and frequency domain. Each feature type was used for classification separately. We also combined all the waveforms across different response types and used the combination for classification. Classification results are reported in Table 4.2. Highest classification accuracy was at 70% and was obtained using the pattern electroretinogram response (pERG) collected using an LED monitor with a bar as visual stimulus (Table 4.1). Second best accuracy was obtained using electroretinogram at the highest stimulus light intensity, which resulted in an accuracy of 66%. Both performances were significantly above chance level (one sample t-test, $p<0.001$) and were obtained using 21 and 23 PCs for the pERG and ERG responses respectively. With both responses, accuracy increased initially as more PCs were used for classification (see Supplementary Figure 4.1). However, after reaching maximum performance, adding more PCs had a detrimental effect on accuracy.

Classification performance of these two responses were further examined. Classification using the pERG response resulted in more than 60% sensitivity and specificity values (Figure 4.3). Sensitivity of the classification using ERG was also above 60%, but its specificity was at 50% (Figure 4.3).

We then sought to look at the most informative PCs for each response type. Performance peaked at 21 PCs with the pERG LED response. Out of the 50 selected feature sets, we then calculated frequency of each PC appearing in the top 21 PCs. frequency of appearance for each PC is shown in Supplementary Figure 4.2. The 22nd, 7th and 15th PCs appear in 92%, 88% and 84% of the selected PC sets. These three PCs explain 0.2%, 1.8% and 0.5% of the variance respectively. Even though these PCs are not the top PCs in terms of the amount of variance they capture, they seem to contain more discriminative information than other PCs. Each PC is constructed by

applying a set of linear weights to each feature, which in our case is a sample of the ERG in time. In Figure 4.4A we plot the weights for these three PCs, which indicates the phase of the response the PC extracts..

Performance using the ERG response peaked at 23 PCs. Four PCs had a high frequency of appearing in the 50 sets of top 23 PCs (Supplementary Figure 4.2). These were the 40th, 6th, 15th and 1st PCs which appear in the top 23 PC set 78%, 76%, 76% and 72% times and explain 0.008%, 3.4%, 0.2% and 41.6% of the variance respectively.
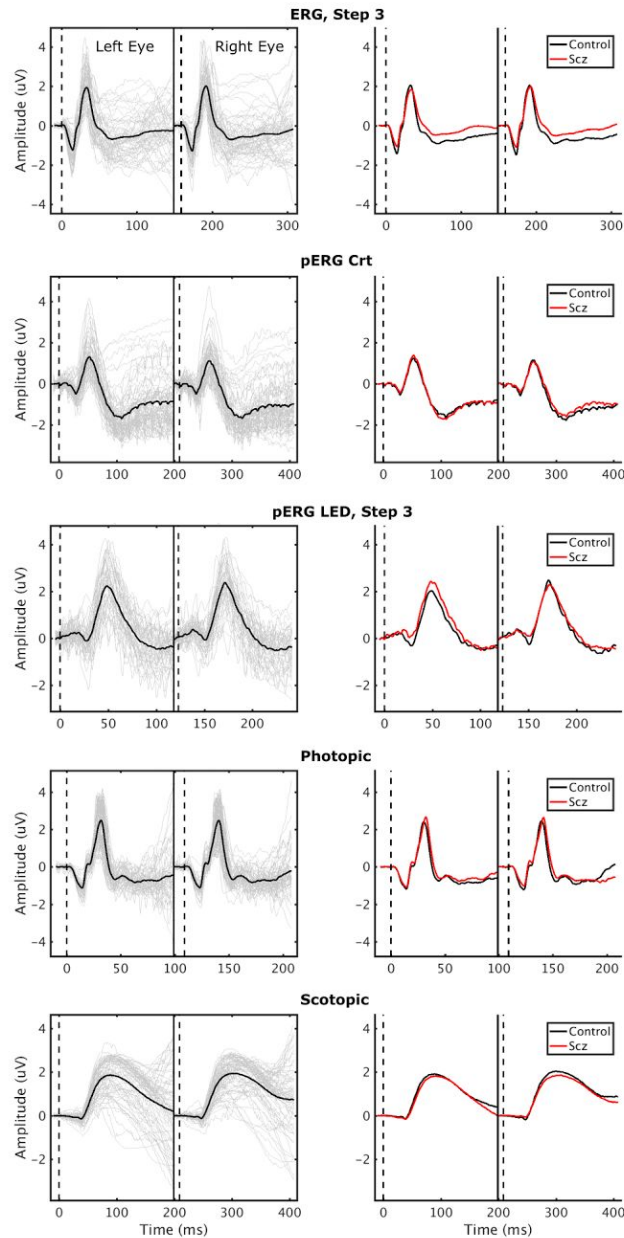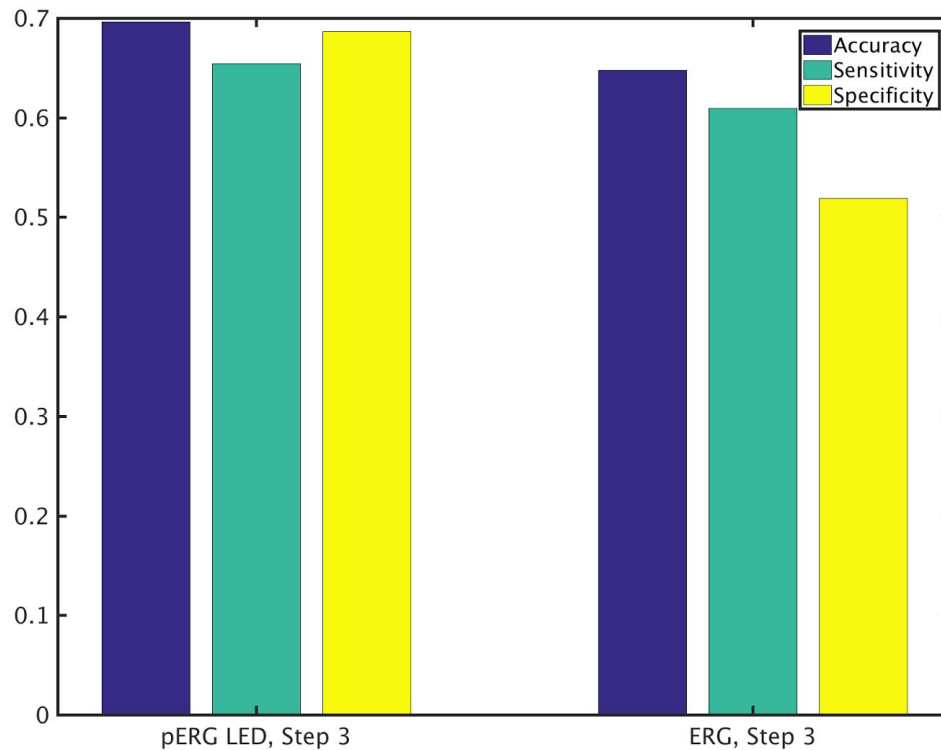
Figure 4.3. Performance Details. Classification accuracy, sensitivity and specificity for the two most informative responses.

Response	Accuracy, Time Domain Features	Accuracy, Frequency Domain Features
ERG, Step 1	58 ± 2	57 ± 2
ERG, Step 2	62 ± 2	58 ± 1
ERG, Step 3	**66 ± 2**	nan ±
pERG Crt	46 ± 2	59 ± 2
pERG LED, Step 1	62 ± 2	59 ± 2
pERG LED, Step 2	53 ± 2	54 ± 2
pERG LED, Step 3	**70 ± 2**	48 ± 2
pERG LED, Step 4	49 ± 2	49 ± 2
pERG LED, Step 5	53 ± 2	58 ± 2
pERG LED, Step 6	55 ± 3	60 ± 2
pERG LED, Step 7	64 ± 2	59 ± 2
pERG LED, Step 8	63 ± 2	55 ± 2
pERG LED, Step 9	53 ± 3	54 ± 2
Photopic	59 ± 2	52 ± 2
Scotopic	57 ± 2	49 ± 2
Combination	60 ± 4	43 ± 4

Table 4.2: Performance Summary. Classification accuracy for each signal, using both time and frequency domain features. Mean and standard error of classification accuracy across all 50 permutations is reported.
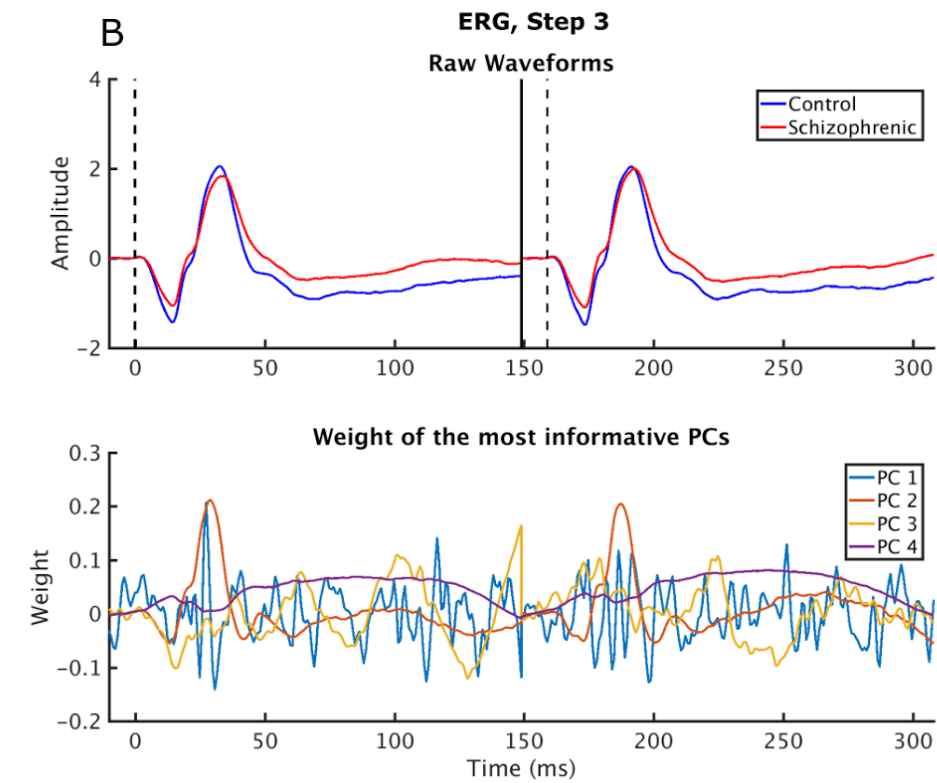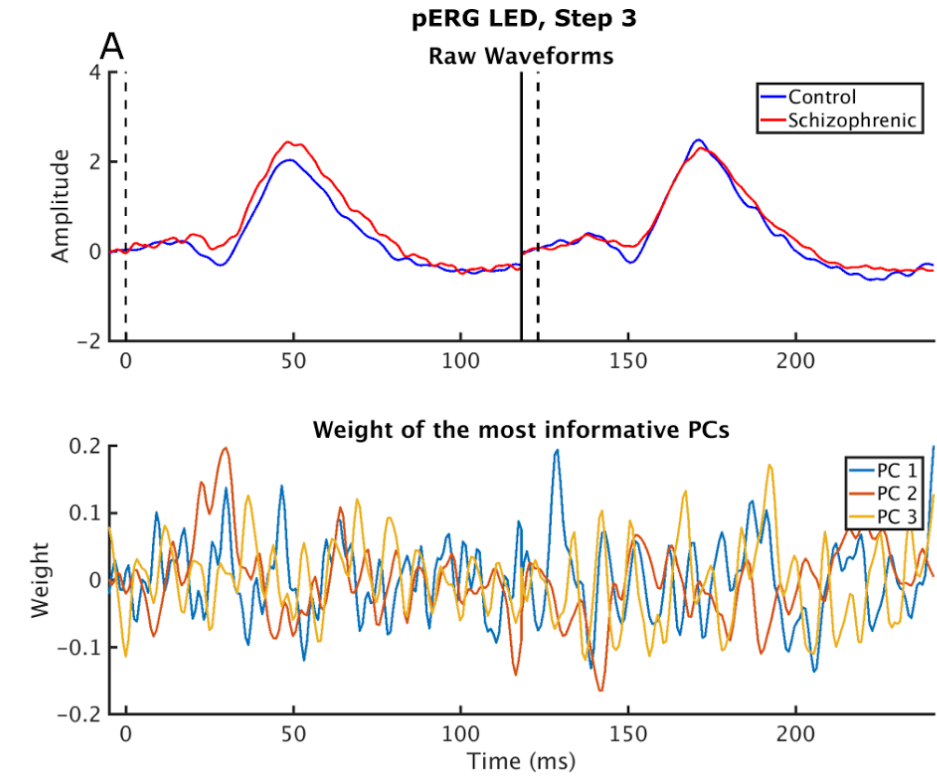
Figure 4.4. Most informative PCs

A) Raw waveform (top) and weight of the most informative PCs (bottom) for pERG LED, step 3 response. PCs are ranked based on the number of times they appeared in the 50 selected PC sets. PC 1 explains 0.2% of total variance and appears in the selected PC set 92% of the time. PC 2 explains 1.8% of total variance and appears in the selected PC set 88% of the time. PC 3 explains 0.5% of total variance and appears in the selected PC set 84% of the time.

B) Raw waveform (top) and weight of the most informative PCs (bottom) for ERG, step 3 response. PCs are ranked based on the number of times they appeared in the 50 selected PC sets. PC 1 explains 0.0.008% of total variance and appears in the selected PC set 78% of the time. PC 2 explains 3.4% of total variance and appears in the selected PC set 76% of the time. PC 3 explains 0.2% of total variance and appears in the selected PC set 76% of the time. PC 4 explains 41.6% of total variance and appears in the selected PC set 72% of the time.

We compared performance of the double cross validation paradigm with single cross validation. In the single cross validation paradigm, the set of selected PCs are used for classification of the same subjects that were used for their selection. The results are shown in Supplementary Figure 4.3. As can be seen, single cross validation accuracy is higher than double cross validation for almost every response type. The difference between classification accuracy between the two paradigms can be as high as ~30%. These results demonstrate the bias of the single cross validation paradigm to overestimate classification accuracy.

Discussion

In this study we used different types of retinal response from a group of healthy and schizophrenic patients to identify biomarkers for the disease. We used a data driven approach to identify traits of the retinal response that were indicating of the state of the disease. Both time and frequency domain features were tested for classification. We used PCA and sequential forward selection algorithms to identify the most discriminative traits. We then used SVM to classify the subjects into healthy and schizophrenic. Highest classification accuracies we obtained were 70% and 66% using pERG and ERG activity respectively. Moreover, classification accuracy was dependent on intensity of the visual stimulus used to invoke the response. Highest classification accuracy was obtained using time domain features. Frequency domain features resulted in lower classification accuracy.

We observed that the type of visual stimulus used to activate retina might be an important factor in the amount of discriminability power across the two groups. PERg activity was recorded using both checkerboard and bar visual stimuli at different intensity level and contrasts. We observed that the bar stimulus at the highest contrast and lowest intensity levels invoked the most discriminative response with 70% accuracy. We also observed that the ERG response was most discriminative at the highest light intensity used.

We observed that combination of the signals together did not produce higher accuracies. The reason is that addition of uninformative signals will dilute the amount of information in the dataset, which reduces generalizability of the results. A similar trend was observed in classification accuracy vs. number of PCs used for classification (Supplementary Figure 4.1). Accuracy peaked at optimal number of PCs and started to decrease as more and more PCs were added.
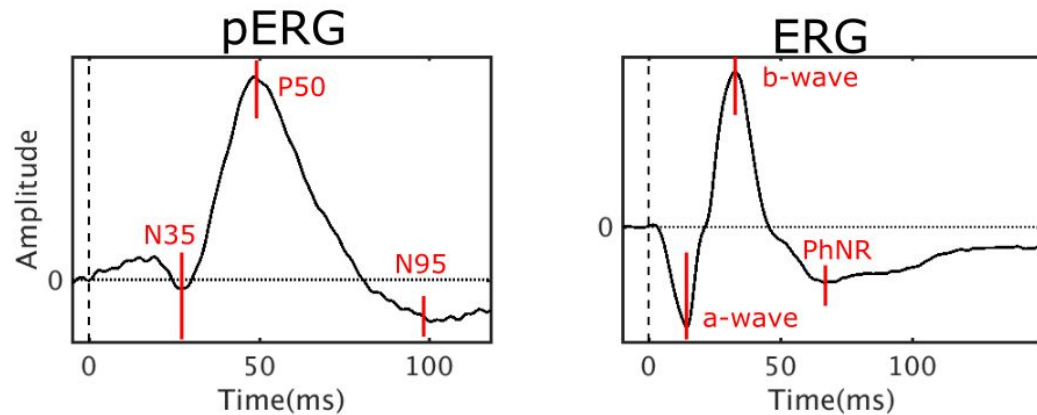
Figure 4.5. Different components of pERG and ERG responses.

Left: Three components of the pERG response: N35, a negative component at 35ms; P50: A positive component at 50ms; N95, a negative component at 95ms. Right: Three components of the ERG response: a-wave, b-wave, and Photopic negative response PhNR

Both pERG and ERG responses have three main components shown in Figure 4.5. The three main components of the pERG are N35, P50, and N95. Variations in latency and amplitude of the P50 and N95 are used for diagnosis of damage to retina (Holder 2001) and can be indicative of a reduction in the population of retinal ganglion cells (Weinstein et al. 1988). Invasive experiments in non-human primates have shown that the N95 components reflect activity of the retinal ganglion cells (Luo and Frishman 2011). The three components of the ERG response are a-wave, b-wave and photopic negative response (PhNR). The a-wave components has been shown to mainly reflect activity of the cone photoreceptors. The b-wave is attributed to activity of bipolar cells (Miura et al. 2009). The PhNR component reflects the activity of retinal ganglion cells (Colotto et al. 2000; Miura et al. 2009; Luo and Frishman 2011; Machida 2012).

The three most discriminative PCs identified by our classification approach using the pERG response do not emphasize any specific part of the waveform. The first PC seems to have peaks right before the a-wave and right after the b-wave. The second PC has peaks at the a-wave components. The third PC does not seem to emphasize any specific part of the waveform.

The four most discriminative PCs identified for the ERG response reveal interesting traits. The most discriminative PC has a distinct peak aligned with the b-wave component. The second most discriminative PC puts higher weight on the portion of the response between the a-wave and b-wave components. The third most informative PC puts higher weight on the a-wave component as well as the slow transient portion of the waveform following the PhNR component. Finally, the fourth puts higher weight on the PhNR response and the transient activity following it.

Collectively, our results point to abnormal activity of the retinal ganglion cells as well as cone photoreceptors and bipolar cells. Since retinal cells express NMDA receptors, our findings support the NMDA hypofunction hypothesis.

Conclusion

This study showed that activity of retinal cells captured using non-invasive measurements is different across healthy and schizophrenic groups. We used a data driven machine learning approach to extract differential patterns of activation across the two groups and observed that response of cones, bipolar cells, and retinal ganglion cells have abnormal activity patterns in schizophrenia. These findings support the NMDA hypofunction hypothesis. They also demonstrate potential of retinal response doe identification of schizophrenia in prodromal state in the high risk population.

Supplementary Figures
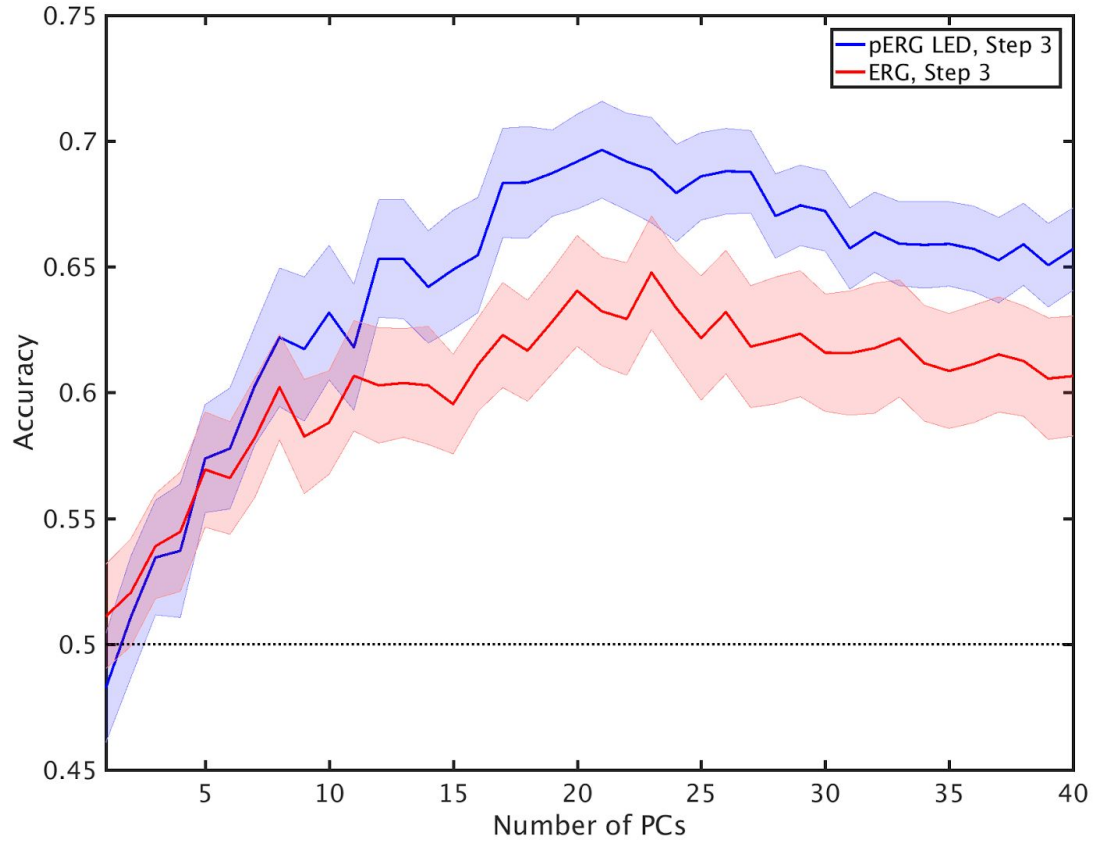
Supplementary Figure 4.1. Performance vs. Number of PCs

Classification accuracy vs. number of PCs for the two most informative responses.

Black dashed line marks chance level.
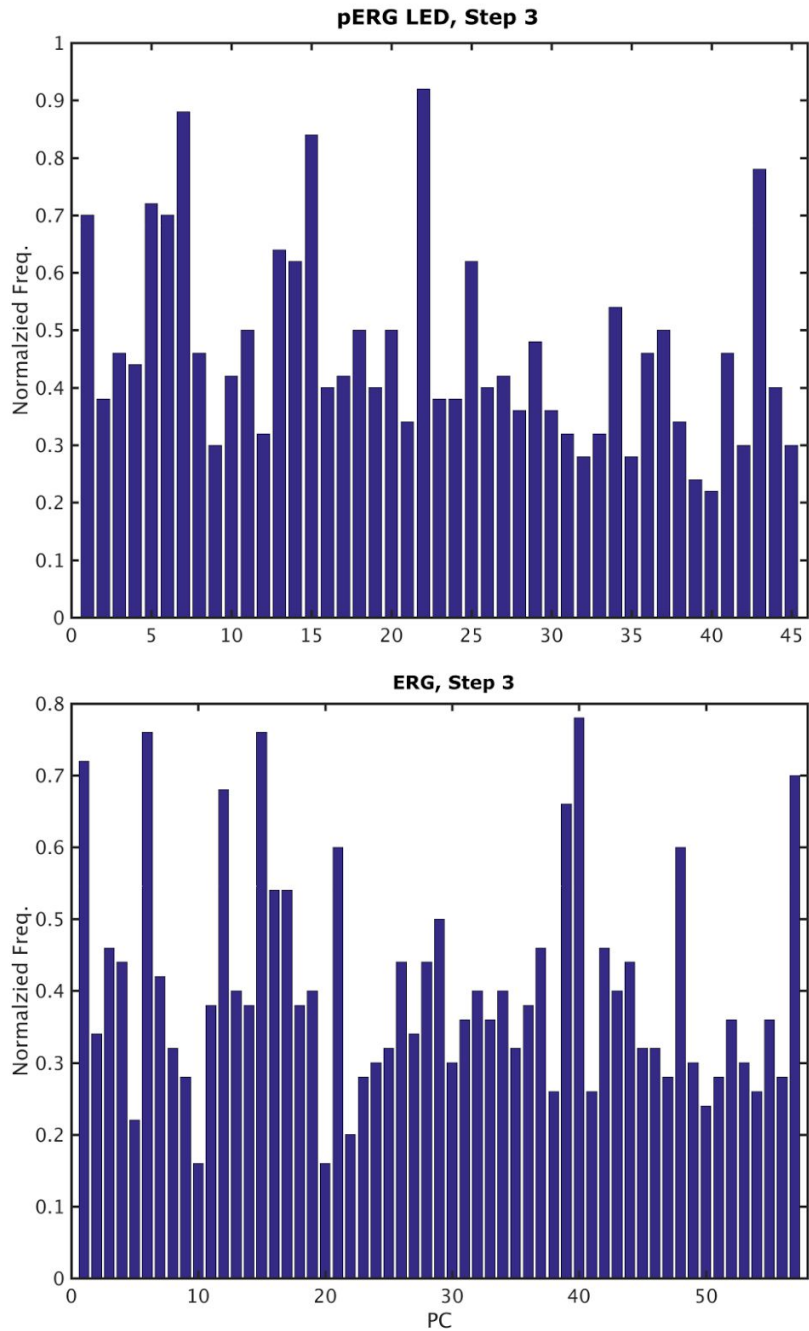
Supplementary Figure 4.2. PC Frequency

Frequency of each PC appearing in the 50 selected PC set for pERG LED, step 3 (top) and ERG, step 3 (bottom). PCs are ranked based on the amount of variance they explain.
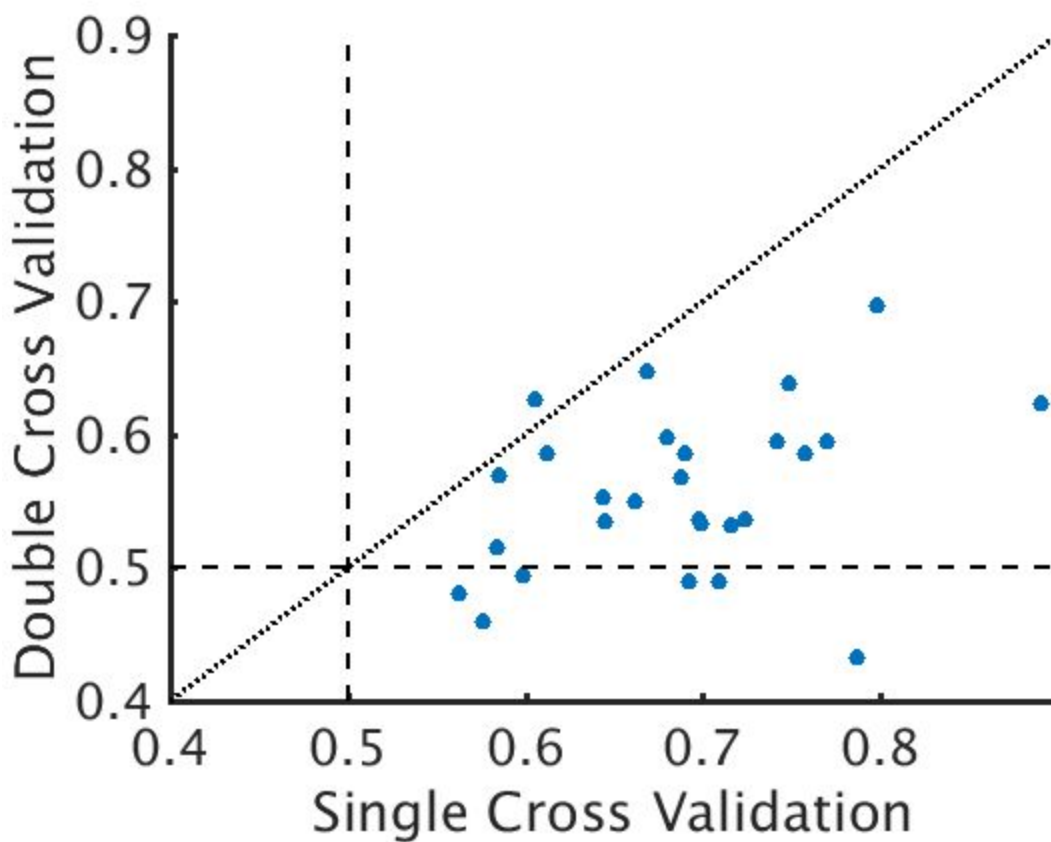
Supplementary Figure 4.3. Single vs. Double Cross Validation

Classification accuracy using double cross validation (y-axis) vs. single cross validation (x-axis). Each dot is performance of one response type in the time or frequency domain. Dotted black line is the identity line. Vertical and horizontal dashed lines mark chance level.

Chapter 5: Conclusion

In this thesis, we developed and used a machine learning classification paradigm for biomarker identification for schizophrenia, using fMRI and retinal electrophysiology datasets. Our approach was purely data driven and used a double cross validation scheme to test robustness of the results. Therefore, our results provide a robust estimate of classification power of the identified biomarkers.

Our fMRI dataset comprised of 6 minutes of resting state activity from 170 subjects, including healthy and schizophrenic individuals. We employed a data driven algorithm to construct an atlas of the brain using the fMRI data itself. We used our functional atlas long with a commonly used anatomical atlas to construct network level models of the brain for each subject. We extracted several graph theoretic measures from these networks and used for classification of the subjects. We observed above chance classification accuracy using both atlases. We also observed that our classification performance was higher using the anatomical atlas. Another interesting finding was that classification performance was higher when prewhitened time series were used for construction of the networks.

Our retinal electrophysiology dataset consisted of several types of retinal activity invoked by different visual stimuli, acquired from a cohort of healthy and schizophrenic subjects. These responses were used for classification, using both time domain and frequency domain features extracted from the dataset. We obtained a 70% classification accuracy using pERG response at a low light intensity with high contrast. We also obtained a 66% classification accuracy using the ERG response at a high light intensity. Both accuracies were observed using time domain features. Frequency domain features of the same signals resulted in lower classification accuracies. Responses to other visual stimuli did not produce high classification accuracies. Our results support the NMDA hypofunction hypothesis of schizophrenia.

Future Directions

We did not observe an improvement in classification accuracy using our functional atlas. However, the lack of evidence that functional atlases improve classification cannot be used to conclude that functional atlases are not useful for biomarker identification. First, our 6 minute resting state activity might not be long enough to capture intrinsic functional connectivity of the brain, an observation supported by our results and others (Laumann et al. 2015). Longer scans therefore might result in more robust parcellations of the brain. Second, other data driven parcellation algorithms such as ICA could improve classification accuracy. Another interesting direction would be to use multimodal atlases developed in recent years (e.g. (Glasser et al. 2016)) for biomarker identification. These atlases combine several data sets and are more robust in capturing the borders between functional modules of the brain.

While we limited our feature set to common graph theoretic measures, other complex network levels measures have been proposed which could reveal more complex abnormal patterns in the diseased brain. One of such measures is controllability (Gu et al. 2014). Controllability refers to ability to control global state of the brain networks by manipulating relevant local interactions. It has been suggested that controllability of the brain might be reduced in schizophrenic patients (Gu et al. 2014).

Our network measures were constructed using a static model of the brain, i.e. assuming that network level organization of the brain remains constant during the 6 minutes scan duration. This assumption might is certainly not valid, cognitive states change rapidly and the cognitive state of the subject may affect the network measures we have used as features. Using a dynamic model, where the entire scan is divided into shorter time windows and separate network measures are extracted from each window independently may result in higher classification accuracies. This approach has been shown to boost classification accuracy using bivariate features (Rashid et al. 2016).

Classification accuracy could be improved by using different feature selection methods or classifiers. Feature selection methods such as minimum redundancy maximum relevance (mRMR) (Peng, Long, and Ding 2005) or norm minimization methods (Nie et al. 2010) could pick up on features that might not be picked up by our current feature

selection method. Decisions Trees and Random Forests are powerful classification algorithms that are capable of picking up on sophisticated decision boundaries that might not be captured by parametric kernels used by SVM classifiers. Further exploration using these classifiers could enhance classification.

Even though we observed above chance accuracy, a 70% classification accuracy is not useful in clinical settings for identification of prodromal state of schizophrenia. Even though using better analytic tools, more sophisticated features, and other classifiers might improve classification accuracy, it is most likely that we are limited by the amount of data we have available at this point.

References

Altman, Douglas G. 2006. *Practical Statistics for Medical Research*. Chapman & Hall/CRC.

Anderson, Ariana, and Mark S. Cohen. 2013. "Decreased Small-World Functional Network Connectivity and Clustering across Resting State Networks in Schizophrenia: An fMRI Classification Tutorial." *Frontiers in Human Neuroscience* 7 (September). books.google.com: 520.

Anderson, Jeffrey S., Jared A. Nielsen, Alyson L. Froehlich, Molly B. DuBray, T. Jason Druzgal, Annahir N. Cariello, Jason R. Cooperrider, et al. 2011. "Functional Connectivity Magnetic Resonance Imaging Classification of Autism." *Brain: A Journal of Neurology* 134 (Pt 12): 3742–54.

Andreasen, N. 1983. "The Scale for the Assessment of Negative Symptoms (SANS)."

———. 1984. "The Scale for the Assessment of Positive Symptoms (SAPS)."

Andreasen, N. C., S. Arndt, R. Alliger, D. Miller, and M. Flaum. 1995. "Symptoms of Schizophrenia. Methods, Meanings, and Mechanisms." *Archives of General Psychiatry* 52 (5): 341–51.

Andreasen, N. C., D. S. O'Leary, T. Cizadlo, S. Arndt, K. Rezai, L. L. Ponto, G. L. Watkins, and R. D. Hichwa. 1996. "Schizophrenia and Cognitive Dysmetria: A Positron-Emission Tomography Study of Dysfunctional Prefrontal-Thalamic-Cerebellar Circuitry." *Proceedings of the National Academy of Sciences of the United States of America* 93 (18): 9985–90.

Andreasen, N. C., and S. Olsen. 1982. "Negative v Positive Schizophrenia. Definition and Validation." *Archives of General Psychiatry* 39 (7): 789–94.

Arbabshirani, Mohammad R., Eduardo Castro, and Vince D. Calhoun. 2014. "Accurate Classification of Schizophrenia Patients Based on Novel Resting-State fMRI Features." *Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference* 2014. ieeexplore.ieee.org: 6691–94.

Arbabshirani, Mohammad R., Eswar Damaraju, Ronald Phlypo, Sergey Plis, Elena Allen, Sai Ma, Daniel Mathalon, et al. 2014. "Impact of Autocorrelation on Functional Connectivity." *NeuroImage* 102 Pt 2 (November): 294–308.

Arbabshirani, Mohammad R., Kent A. Kiehl, Godfrey D. Pearlson, and Vince D. Calhoun. 2013. "Classification of Schizophrenia Patients Based on Resting-State Functional Network Connectivity." *Frontiers in Neuroscience* 7 (July). researchgate.net: 133.

Ardekani, Babak A., Jay Nierenberg, Matthew J. Hoptman, Daniel C. Javitt, and Kelvin O. Lim. 2003. "MRI Study of White Matter Diffusion Anisotropy in Schizophrenia." *Neuroreport* 14 (16): 2025–29.

Atluri, Gowtham, Kanchana Padmanabhan, Gang Fang, Michael Steinbach, Jeffrey R. Petrella, Kelvin Lim, Angus Macdonald, Nagiza F. Samatova, P. Murali Doraiswamy, and Vipin Kumar. 2013. "Complex Biomarker Discovery in Neuroimaging Data: Finding a Needle in a Haystack." *NeuroImage. Clinical* 3 (January): 123–31.

Atluri, Gowtham, Michael Steinbach, Kelvin O. Lim, Vipin Kumar, and Angus MacDonald. 2015. "Connectivity Cluster Analysis for Discovering Discriminative Subnetworks in Schizophrenia." *Human Brain Mapping* 36 (2): 756–67.

Bassett, Danielle S., Brent G. Nelson, Bryon A. Mueller, Jazmin Camchong, and Kelvin O. Lim. 2012. "Altered Resting State Complexity in Schizophrenia." *NeuroImage* 59 (3): 2196–2207.

Bishop, Christopher M. 2006. "Pattern Recognition and Machine Learning (Information Science and Statistics)," August. Springer-Verlag New York, Inc. http://dl.acm.org/citation.cfm?id=1162264.

Bleich-Cohen, Maya, Shahar Jamshy, Haggai Sharon, Ronit Weizman, Nathan Intrator, Michael Poyurovsky, and Talma Hendler. 2014. "Machine Learning fMRI Classifier Delineates Subgroups of Schizophrenia Patients." *Schizophrenia Research* 160 (1-3). Elsevier: 196–200.

Bullmore, Ed, and Olaf Sporns. 2009. "Complex Brain Networks: Graph Theoretical Analysis of Structural and Functional Systems." *Nature Reviews. Neuroscience* 10 (3). Nature Publishing Group: 186–98.

Camchong, Jazmin, Angus W. MacDonald, Christopher Bell, Bryon A. Mueller, and Kelvin O. Lim. 2011. "Altered Functional and Anatomical Connectivity in Schizophrenia." *Schizophrenia Bulletin* 37 (3): 640–50.

Campbell, Stephen L., and Carl D. Meyer. 2008. *Generalized Inverses of Linear Transformations (Classics in Applied Mathematics)*. Society for Industrial and Applied Mathematics.

Cardno, A. G., E. J. Marshall, B. Coid, A. M. Macdonald, T. R. Ribchester, N. J. Davies, P. Venturi, et al. 1999. "Heritability Estimates for Psychotic Disorders: The Maudsley Twin Psychosis Series." *Archives of General Psychiatry* 56 (2): 162–68.

Castro, Eduardo, Manel Martínez-Ramón, Godfrey Pearlson, Jing Sui, and Vince D. Calhoun. 2011. "Characterization of Groups Using Composite Kernels and Multi-Source fMRI Analysis Data: Application to Schizophrenia." *NeuroImage* 58 (2): 526–36.

Chen, Guangyu, Gang Chen, Chunming Xie, and Shi-Jiang Li. 2011. "Negative Functional Connectivity and Its Dependence on the Shortest Path Length of Positive Network in the Resting-State Human Brain." *Brain Connectivity* 1 (3): 195–206.

Chien, Wai Tong, Sau Fong Leung, Frederick Kk Yeung, and Wai Kit Wong. 2013. "Current Approaches to Treatments for Schizophrenia Spectrum Disorders, Part II: Psychosocial Interventions and Patient-Focused Perspectives in Psychiatric Care." *Neuropsychiatric Disease and Treatment* 9 (September): 1463–81.

Chien, Wai Tong, and Annie Lk Yip. 2013. "Current Approaches to Treatments for Schizophrenia Spectrum Disorders, Part I: An Overview and Medical Treatments." *Neuropsychiatric Disease and Treatment* 9 (September): 1311–32.

Christova, Peka, Lisa M. James, Brian E. Engdahl, Scott M. Lewis, and Apostolos P. Georgopoulos. 2015. "Diagnosis of Posttraumatic Stress Disorder (PTSD) Based on Correlations of Prewhitened fMRI Data: Outcomes and Areas Involved." *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale* 233 (9): 2695–2705.

Christova, P., S. M. Lewis, T. A. Jerde, J. K. Lynch, and A. P. Georgopoulos. 2011. "True Associations between Resting fMRI Time Series Based on Innovations." *Journal of Neural Engineering* 8 (4): 046025.

Chyzhyk, Darya, and Manuel Graña. 2015. "Classification of Schizophrenia Patients on Lattice Computing Resting-State fMRI Features." *Neurocomputing* 151, Part 1 (March). Elsevier: 151–60.

Cirillo, Michael A., and Larry J. Seidman. 2003. "Verbal Declarative Memory Dysfunction

in Schizophrenia: From Clinical Assessment to Genetics and Brain Mechanisms." *Neuropsychology Review* 13 (2): 43–77.

Clarke, Robert, Habtom W. Ressom, Antai Wang, Jianhua Xuan, Minetta C. Liu, Edmund A. Gehan, and Yue Wang. 2008. "The Properties of High-Dimensional Data Spaces: Implications for Exploring Gene and Protein Expression Data." *Nature Reviews. Cancer* 8 (1): 37–49.

Cohen, Alexander L., Damien A. Fair, Nico U. F. Dosenbach, Francis M. Miezin, Donna Dierker, David C. Van Essen, Bradley L. Schlaggar, and Steven E. Petersen. 2008. "Defining Functional Areas in Individual Human Brains Using Resting Functional Connectivity MRI." *NeuroImage* 41 (1): 45–57.

Cohen, J. D., and D. Servan-Schreiber. 1992. "Context, Cortex, and Dopamine: A Connectionist Approach to Behavior and Biology in Schizophrenia." *Psychological Review* 99 (1): 45–77.

Colotto, A., B. Falsini, T. Salgarello, G. Iarossi, M. E. Galan, and L. Scullica. 2000. "Photopic Negative Response of the Human ERG: Losses Associated with Glaucomatous Damage." *Investigative Ophthalmology & Visual Science* 41 (8). iovs.arvojournals.org: 2205–11.

Costafreda, Sergi G., Cynthia H. Y. Fu, Marco Picchioni, Timothea Toulopoulou, Colm McDonald, Eugenia Kravariti, Muriel Walshe, Diana Prata, Robin M. Murray, and Philip K. McGuire. 2011. "Pattern of Neural Responses to Verbal Fluency Shows Diagnostic Specificity for Schizophrenia and Bipolar Disorder." *BMC Psychiatry* 11 (1): 18.

Coyle, J. T. 1996. "The Glutamatergic Dysfunction Hypothesis for Schizophrenia." *Harvard Review of Psychiatry* 3 (5): 241–53.

Craddock, R. Cameron, G. Andrew James, Paul E. Holtzheimer, Xiaoping P. Hu, and Helen S. Mayberg. 2012. "A Whole Brain fMRI Atlas Generated via Spatially Constrained Spectral Clustering." *Human Brain Mapping* 33 (8): 1914–28.

Davis, Kenneth L., Daniel G. Stewart, Joseph I. Friedman, Monte Buchsbaum, Philip D. Harvey, Patrick R. Hof, Joseph Buxbaum, and Vahram Haroutunian. 2003. "White Matter Changes in Schizophrenia: Evidence for Myelin-Related Dysfunction." *Archives of General Psychiatry* 60 (5): 443–56.

Demirci, Oguz, Vincent P. Clark, and Vince D. Calhoun. 2008. "A Projection Pursuit Algorithm to Classify Individuals Using fMRI Data: Application to Schizophrenia." *NeuroImage* 39 (4). Elsevier: 1774–82.

Desikan, Rahul S., Florent Ségonne, Bruce Fischl, Brian T. Quinn, Bradford C. Dickerson, Deborah Blacker, Randy L. Buckner, et al. 2006. "An Automated Labeling System for Subdividing the Human Cerebral Cortex on MRI Scans into Gyral Based Regions of Interest." *NeuroImage* 31 (3): 968–80.

Destrieux, Christophe, Bruce Fischl, Anders Dale, and Eric Halgren. 2010. "Automatic Parcellation of Human Cortical Gyri and Sulci Using Standard Anatomical Nomenclature." *NeuroImage* 53 (1): 1–15.

Driesen, N. R., G. McCarthy, Z. Bhagwagar, M. Bloch, V. Calhoun, D. C. D'Souza, R. Gueorguieva, et al. 2013. "Relationship of Resting Brain Hyperconnectivity and Schizophrenia-like Symptoms Produced by the NMDA Receptor Antagonist Ketamine in Humans." *Molecular Psychiatry* 18 (11): 1199–1204.

Du, Wei, Vince D. Calhoun, Hualiang Li, Sai Ma, Tom Eichele, Kent A. Kiehl, Godfrey D. Pearlson, and Tülay Adali. 2012. "High Classification Accuracy for Schizophrenia with Rest and Task FMRI Data." *Frontiers in Human Neuroscience* 6 (June).

ncbi.nlm.nih.gov: 145.

Efron, Bradley. 1983. "Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation." *Journal of the American Statistical Association* 78 (382). Taylor & Francis: 316–31.

Efron, Bradley, and Gail Gong. 1983. "A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation." *The American Statistician* 37 (1). Taylor & Francis: 36–48.

Fan, Yong, Yong Liu, Hong Wu, Yihui Hao, Haihong Liu, Zhening Liu, and Tianzi Jiang. 2011. "Discriminant Analysis of Functional Connectivity Patterns on Grassmann Manifold." *NeuroImage* 56 (4): 2058–67.

Fawcett, Tom. 2006. "An Introduction to ROC Analysis." *Pattern Recognition Letters* 27 (8): 861–74.

Fekete, Tomer, Meytal Wilf, Denis Rubin, Shimon Edelman, Rafael Malach, and Lilianne R. Mujica-Parodi. 2013. "Combining Classification with fMRI-Derived Complex Network Measures for Potential Neurodiagnostics." *PloS One* 8 (5). journals.plos.org: e62867.

Filzmoser, Peter, Bettina Liebmann, and Kurt Varmuza. 2009. "Repeated Double Cross Validation." *Journal of Chemometrics* 23 (4). John Wiley & Sons, Ltd.: 160–71.

Fischl, Bruce, André van der Kouwe, Christophe Destrieux, Eric Halgren, Florent Ségonne, David H. Salat, Evelina Busa, et al. 2004. "Automatically Parcellating the Human Cerebral Cortex." *Cerebral Cortex* 14 (1): 11–22.

Fisher, R. A. 1936. "THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS." *Annals of Eugenics* 7 (2). Blackwell Publishing Ltd: 179–88.

Foong, J., M. Maier, C. A. Clark, G. J. Barker, D. H. Miller, and M. A. Ron. 2000. "Neuropathological Abnormalities of the Corpus Callosum in Schizophrenia: A Diffusion Tensor Imaging Study." *Journal of Neurology, Neurosurgery, and Psychiatry* 68 (2). BMJ Publishing Group Ltd: 242–44.

Ford, James, Hany Farid, Fillia Makedon, Laura A. Flashman, Thomas W. McAllister, Vasilis Megalooikonomou, and Andrew J. Saykin. 2003. "Patient Classification of fMRI Activation Maps." In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003*, 58–65. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg.

Freund, Yoav, and Robert E. Schapire. 1997. "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting." *Journal of Computer and System Sciences* 55 (1). Orlando, FL, USA: Academic Press, Inc.: 119–39.

Frishman, Laura J. 2012. "Electrogenesis of the Electroretinogram." In *Retina*, edited by Stephen J. Ryan. Saunders.

Friston, K. J. 1998. "The Disconnection Hypothesis." *Schizophrenia Research* 30 (2): 115–25.

Glasser, Matthew F., Timothy S. Coalson, Emma C. Robinson, Carl D. Hacker, John Harwell, Essa Yacoub, Kamil Ugurbil, et al. 2016. "A Multi-Modal Parcellation of Human Cerebral Cortex." *Nature* 536 (7615): 171–78.

Goldman-Rakic, P. S. 1994. "Working Memory Dysfunction in Schizophrenia." *The Journal of Neuropsychiatry and Clinical Neurosciences* 6 (4): 348–57.

Gordon, Evan M., Timothy O. Laumann, Babatunde Adeyemo, Jeremy F. Huckins, William M. Kelley, and Steven E. Petersen. 2014. "Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations." *Cerebral Cortex* 26 (1). Oxford University Press: 288–303.

Guo, Shuixia, Keith M. Kendrick, Jie Zhang, Matthew Broome, Rongjun Yu, Zhening Liu,

and Jianfeng Feng. 2013. "Brain-Wide Functional Inter-Hemispheric Disconnection Is a Potential Biomarker for Schizophrenia and Distinguishes It from Depression." *NeuroImage. Clinical* 2 (June): 818–26.

Gu, Shi, Fabio Pasqualetti, Matthew Cieslak, Scott T. Grafton, and Danielle S. Bassett. 2014. "Controllability of Brain Networks." *arXiv [q-bio.NC]*. arXiv. http://arxiv.org/abs/1406.5197.

Guyon, Isabelle, and André Elisseeff. 2003. "An Introduction to Variable and Feature Selection." *Journal of Machine Learning Research: JMLR* 3 (Mar). jmlr.org: 1157–82.

Heuvel, Martijn P. van den, René C. W. Mandl, Cornelis J. Stam, René S. Kahn, and Hilleke E. Hulshoff Pol. 2010. "Aberrant Frontal and Temporal Complex Network Structure in Schizophrenia: A Graph Theoretical Analysis." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 30 (47): 15915–26.

Heuvel, Martijn P. van den, Olaf Sporns, Guusje Collin, Thomas Scheewe, René C. W. Mandl, Wiepke Cahn, Joaquín Goñi, Hilleke E. Hulshoff Pol, and René S. Kahn. 2013. "Abnormal Rich Club Organization and Functional Brain Dynamics in Schizophrenia." *JAMA Psychiatry* 70 (8). American Medical Association: 783–92.

Holder, G. E. 2001. "Pattern Electroretinography (PERG) and an Integrated Approach to Visual Pathway Diagnosis." *Progress in Retinal and Eye Research* 20 (4): 531–61.

Honnorat, N., H. Eavani, T. D. Satterthwaite, R. E. Gur, R. C. Gur, and C. Davatzikos. 2015. "GraSP: Geodesic Graph-Based Segmentation with Shape Priors for the Functional Parcellation of the Cortex." *NeuroImage* 106 (February): 207–21.

Howes, Oliver D., and Shitij Kapur. 2009. "The Dopamine Hypothesis of Schizophrenia: Version III--the Final Common Pathway." *Schizophrenia Bulletin* 35 (3): 549–62.

Hu, Maorong, Jun Li, Lisa Eyler, Xiaofeng Guo, Qingling Wei, Jingsong Tang, Feng Liu, et al. 2013. "Decreased Left Middle Temporal Gyrus Volume in Antipsychotic Drug-Naive, First-Episode Schizophrenia Patients and Their Healthy Unaffected Siblings." *Schizophrenia Research* 144 (1-3). Elsevier: 37–42.

Hu, Xintao, Dajiang Zhu, Peili Lv, Kaiming Li, Junwei Han, Lihong Wang, Dinggang Shen, Lei Guo, and Tianming Liu. 2013. "Fine-Granularity Functional Interaction Signatures for Characterization of Brain Conditions." *Neuroinformatics* 11 (3): 301–17.

Ingalhalikar, Madhura, Stathis Kanterakis, Ruben Gur, Timothy P. L. Roberts, and Ragini Verma. 2010. "DTI Based Diagnostic Prediction of a Disease via Pattern Classification." *Medical Image Computing and Computer-Assisted Intervention: MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention* 13 (Pt 1): 558–65.

Iwabuchi, Sarina J., Peter F. Liddle, and Lena Palaniyappan. 2013. "Clinical Utility of Machine-Learning Approaches in Schizophrenia: Improving Diagnostic Confidence for Translational Neuroimaging." *Frontiers in Psychiatry / Frontiers Research Foundation* 4 (August). researchgate.net: 95.

Jain, A. K., R. P. W. Duin, and Jianchang Mao. 2000. "Statistical Pattern Recognition: A Review." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (1). ieeexplore.ieee.org: 4–37.

Javitt, D. C., and S. R. Zukin. 1991. "Recent Advances in the Phencyclidine Model of Schizophrenia." *The American Journal of Psychiatry* 148 (10): 1301–8.

Juneja, A., B. Rana, and R. K. Agrawal. 2014. "A Novel Approach for Classification of

Schizophrenia Patients and Healthy Subjects Using Auditory Oddball Functional MRI." In *Artificial Intelligence (MICAI), 2014 13th Mexican International Conference on*, 75–81. ieeexplore.ieee.org.

Kambeitz, Joseph, Lana Kambeitz-Ilankovic, Stefan Leucht, Stephen Wood, Christos Davatzikos, Berend Malchow, Peter Falkai, and Nikolaos Koutsouleris. 2015. "Detecting Neuroimaging Biomarkers for Schizophrenia: A Meta-Analysis of Multivariate Pattern Recognition Studies." *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology* 40 (7). nature.com: 1742–51.

Kelly, Clare, Lucina Q. Uddin, Zarrar Shehzad, Daniel S. Margulies, F. Xavier Castellanos, Michael P. Milham, and Michael Petrides. 2010. "Broca's Region: Linking Human Brain Functional Connectivity Data and Non-Human Primate Tracing Anatomy Studies." *The European Journal of Neuroscience* 32 (3): 383–98.

Kendall Maurice; Gibbons, Jean Dickinson. 1990. "Rank Correlation Methods." In , 5th ed. Oxford University Press.

Kim, Junghoe, Vince D. Calhoun, Eunsoo Shim, and Jong-Hwan Lee. 2016. "Deep Neural Network with Weight Sparsity Control and Pre-Training Extracts Hierarchical Features and Enhances Classification Performance: Evidence from Whole-Brain Resting-State Functional Connectivity Patterns of Schizophrenia." *NeuroImage* 124. Elsevier: 127–46.

Koch, Stefan P., Claudia Hägele, John-Dylan Haynes, Andreas Heinz, Florian Schlagenhauf, and Philipp Sterzer. 2015. "Diagnostic Classification of Schizophrenia Patients on the Basis of Regional Reward-Related FMRI Signal Patterns." *PloS One* 10 (3). journals.plos.org: e0119089.

Koutsouleris, Nikolaos, Eva M. Meisenzahl, Stefan Borgwardt, Anita Riecher-Rössler, Thomas Frodl, Joseph Kambeitz, Yanis Köhler, et al. 2015. "Individualized Differential Diagnosis of Schizophrenia and Mood Disorders Using Neuroanatomical Biomarkers." *Brain: A Journal of Neurology* 138 (Pt 7). Oxford Univ Press: 2059–73.

Laumann, Timothy O., Evan M. Gordon, Babatunde Adeyemo, Abraham Z. Snyder, Sung Jun Joo, Mei-Yen Chen, Adrian W. Gilmore, et al. 2015. "Functional System and Areal Organization of a Highly Sampled Individual Human Brain." *Neuron* 87 (3): 657–70.

Liu, Huan, and Hiroshi Motoda. 2007. *Computational Methods of Feature Selection (Chapman & Hall/Crc Data Mining and Knowledge Discovery Series)*. Chapman & Hall/CRC.

Long, Xiangyu, Dominique Goltz, Daniel S. Margulies, Till Nierhaus, and Arno Villringer. 2014. "Functional Connectivity-Based Parcellation of the Human Sensorimotor Cortex." *The European Journal of Neuroscience* 39 (8): 1332–42.

Luo, Xunda, and Laura J. Frishman. 2011. "Retinal Pathway Origins of the Pattern Electroretinogram (PERG)." *Investigative Ophthalmology & Visual Science* 52 (12). iovs.arvojournals.org: 8571–84.

Lynall, Mary-Ellen, Danielle S. Bassett, Robert Kerwin, Peter J. McKenna, Manfred Kitzbichler, Ulrich Muller, and Ed Bullmore. 2010. "Functional Connectivity and Brain Networks in Schizophrenia." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 30 (28): 9477–87.

MacDonald, Angus W., 3rd. 2008. "Building a Clinically Relevant Cognitive Task: Case Study of the AX Paradigm." *Schizophrenia Bulletin* 34 (4): 619–28.

Machida, Shigeki. 2012. "Clinical Applications of the Photopic Negative Response to

Optic Nerve and Retinal Diseases." *Journal of Ophthalmology* 2012 (October). Hindawi. doi:10.1155/2012/397178.

McDonald, Colm, and Robin M. Murray. 2000. "Early and Late Environmental Risk Factors for Schizophrenia." *Brain Research Reviews* 31 (2). Elsevier: 130–37.

Minami, T., K. Nobuhara, G. Okugawa, K. Takase, T. Yoshida, S. Sawada, S. Ha-Kawa, K. Ikeda, and T. Kinoshita. 2003. "Diffusion Tensor Magnetic Resonance Imaging of Disruption of Regional White Matter in Schizophrenia." *Neuropsychobiology* 47 (3): 141–45.

Miura, Gen, Minhua H. Wang, Kevin M. Ivers, and Laura J. Frishman. 2009. "Retinal Pathway Origins of the Pattern ERG of the Mouse." *Experimental Eye Research* 89 (1): 49–62.

Moghaddam, Bita, and Daniel Javitt. 2012. "From Revolution to Evolution: The Glutamate Hypothesis of Schizophrenia and Its Implication for Treatment." *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology* 37 (1): 4–15.

Nie, Feiping, Heng Huang, Xiao Cai, and Chris H. Ding. 2010. "Efficient and Robust Feature Selection via Joint \mathscrl2,1-Norms Minimization." In *Advances in Neural Information Processing Systems 23*, edited by J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, 1813–21. Curran Associates, Inc.

Nieuwenhuis, Mireille, Neeltje E. M. van Haren, Hilleke E. Hulshoff Pol, Wiepke Cahn, René S. Kahn, and Hugo G. Schnack. 2012. "Classification of Schizophrenia Patients and Healthy Controls from Structural MRI Scans in Two Large Independent Samples." *NeuroImage* 61 (3): 606–12.

Olney, J. W., and N. B. Farber. 1995. "Glutamate Receptor Dysfunction and Schizophrenia." *Archives of General Psychiatry* 52 (12): 998–1007.

Onitsuka, Toshiaki, Martha E. Shenton, Dean F. Salisbury, Chandlee C. Dickey, Kiyoto Kasai, Sarah K. Toner, Melissa Frumin, Ron Kikinis, Ferenc A. Jolesz, and Robert W. McCarley. 2004. "Middle and Inferior Temporal Gyrus Gray Matter Volume Abnormalities in Chronic Schizophrenia: An MRI Study." *The American Journal of Psychiatry* 161 (9). Am Psychiatric Assoc: 1603–11.

Orrù, Graziella, William Pettersson-Yeo, Andre F. Marquand, Giuseppe Sartori, and Andrea Mechelli. 2012. "Using Support Vector Machine to Identify Imaging Biomarkers of Neurological and Psychiatric Disease: A Critical Review." *Neuroscience and Biobehavioral Reviews* 36 (4): 1140–52.

Parisot, Sarah, Salim Arslan, Jonathan Passerat-Palmbach, William M. Wells 3rd, and Daniel Rueckert. 2016. "Group-Wise Parcellation of the Cortex through Multi-Scale Spectral Clustering." *NeuroImage*, May. Elsevier. doi:10.1016/j.neuroimage.2016.05.035.

Peng, Hanchuan, Fuhui Long, and Chris Ding. 2005. "Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (8): 1226–38.

Pettersson-Yeo, William, Paul Allen, Stefania Benetti, Philip McGuire, and Andrea Mechelli. 2011. "Dysconnectivity in Schizophrenia: Where Are We Now?" *Neuroscience and Biobehavioral Reviews* 35 (5): 1110–24.

Power, Jonathan D., Alexander L. Cohen, Steven M. Nelson, Gagan S. Wig, Kelly Anne Barnes, Jessica A. Church, Alecia C. Vogel, et al. 2011. "Functional Network Organization of the Human Brain." *Neuron* 72 (4). Elsevier: 665–78.

Preiser, Dunja, Wolf A. Lagrèze, Michael Bach, and Charlotte M. Poloschek. 2013. "Photopic Negative Response versus Pattern Electroretinogram in Early Glaucoma." *Investigative Ophthalmology & Visual Science* 54 (2): 1182–91.

Rand, William M. 1971. "Objective Criteria for the Evaluation of Clustering Methods." *Journal of the American Statistical Association* 66 (336). Taylor & Francis Group: 846–50.

Rashid, Barnaly, Mohammad R. Arbabshirani, Eswar Damaraju, Mustafa S. Cetin, Robyn Miller, Godfrey D. Pearlson, and Vince D. Calhoun. 2016. "Classification of Schizophrenia and Bipolar Patients Using Static and Dynamic Resting-State fMRI Brain Connectivity." *NeuroImage* 134 (July). Elsevier: 645–57.

Reichardt, Jörg, and Stefan Bornholdt. 2006. "Statistical Mechanics of Community Detection." *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics* 74 (1 Pt 2): 016110.

Ronhovde, Peter, and Zohar Nussinov. 2009. "Multiresolution Community Detection for Megascale Networks by Information-Based Replica Correlations." *Physical Review E* 80 (1). American Physical Society: 016109.

Rosa, Maria J., Liana Portugal, Tim Hahn, Andreas J. Fallgatter, Marta I. Garrido, John Shawe-Taylor, and Janaina Mourao-Miranda. 2015. "Sparse Network-Based Models for Patient Classification Using fMRI." *NeuroImage* 105 (January). Elsevier: 493–506.

Rousseeuw, Peter J. 1987. "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis." *Journal of Computational and Applied Mathematics* 20 (November): 53–65.

Rubinov, Mikail, and Olaf Sporns. 2010a. "Brain Connectivity Toolbox." https://sites.google.com/site/bctnet/.

———. 2010b. "Complex Network Measures of Brain Connectivity: Uses and Interpretations." *NeuroImage* 52 (3): 1059–69.

Savio, Alexandre, and Manuel Graña. 2015. "Local Activity Features for Computer Aided Diagnosis of Schizophrenia on Resting-State fMRI." *Neurocomputing* 164 (September). Elsevier: 154–61.

"Schizophrenia." 2016a. *National Institute of Mental Health*. February. https://www.nimh.nih.gov/health/statistics/prevalence/schizophrenia.shtml.

———. 2016b. *World Health Organization*. April. http://www.who.int/mediacentre/factsheets/fs397/en/.

Shen, Hui, Zhenfeng Li, Ling-Li Zeng, Lin Yuan, Fanglin Chen, Zhening Liu, and Dewen Hu. 2014. "Internetwork Dynamic Connectivity Effectively Differentiates Schizophrenic Patients from Healthy Controls." *Neuroreport* 25 (17). journals.lww.com: 1344–49.

Shen, Hui, Lubin Wang, Yadong Liu, and Dewen Hu. 2010. "Discriminative Analysis of Resting-State Functional Connectivity Patterns of Schizophrenia Using Low Dimensional Embedding of fMRI." *NeuroImage* 49 (4): 3110–21.

Shenton, M. E., C. C. Dickey, M. Frumin, and R. W. McCarley. 2001. "A Review of MRI Findings in Schizophrenia." *Schizophrenia Research* 49 (1-2): 1–52.

Shen, X., F. Tokoglu, X. Papademetris, and R. T. Constable. 2013. "Groupwise Whole-Brain Parcellation from Resting-State fMRI Data for Network Node Identification." *NeuroImage* 82 (November): 403–15.

Shen, Yin, Xiao-Ling Liu, and Xiong-Li Yang. 2006. "N-Methyl-D-Aspartate Receptors in the Retina." *Molecular Neurobiology* 34 (3): 163–79.

Shi, Feng, Yong Liu, Tianzi Jiang, Yuan Zhou, Wanlin Zhu, Jiefeng Jiang, Haihong Liu, and Zhening Liu. 2007. "Regional Homogeneity and Anatomical Parcellation for fMRI Image Classification: Application to Schizophrenia and Normal Controls." *Medical Image Computing and Computer-Assisted Intervention: MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention* 10 (Pt 2): 136–43.

Silva, R. F., E. Castro, C. N. Gupta, M. Cetin, M. Arbabshirani, V. K. Potluru, S. M. Plis, and V. D. Calhoun. 2014. "The Tenth Annual MLSP Competition: Schizophrenia Classification Challenge." In *2014 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 1–6.

Simon, Richard, Michael D. Radmacher, Kevin Dobbin, and Lisa M. McShane. 2003. "Pitfalls in the Use of DNA Microarray Data for Diagnostic and Prognostic Classification." *Journal of the National Cancer Institute* 95 (1): 14–18.

Singh, Megha, and Ganesh Bagler. 2016. "Network Biomarkers of Schizophrenia by Graph Theoretical Investigations of Brain Functional Networks." *arXiv [q-bio.QM]*. arXiv. http://arxiv.org/abs/1602.01191.

Skudlarski, Pawel, Kanchana Jagannathan, Karen Anderson, Michael C. Stevens, Vince D. Calhoun, Beata A. Skudlarska, and Godfrey Pearlson. 2010. "Brain Connectivity Is Not Only Lower but Different in Schizophrenia: A Combined Anatomical and Functional Approach." *Biological Psychiatry* 68 (1): 61–69.

Snyder, Melissa A., and Wen-Jun Gao. 2013. "NMDA Hypofunction as a Convergence Point for Progression and Symptoms of Schizophrenia." *Frontiers in Cellular Neuroscience* 7 (March): 31.

Soares, Matheus Schmidt, Wellingson Silva Paiva, Eda Z. Guertzenstein, Robson Luis Amorim, Luca Silveira Bernardo, Jose Francisco Pereira, Erich Talamoni Fonoff, and Manoel Jacobsen Teixeira. 2013. "Psychosurgery for Schizophrenia: History and Perspectives." *Neuropsychiatric Disease and Treatment* 9 (April): 509–15.

Su, Longfei, Lubin Wang, Hui Shen, Guiyu Feng, and Dewen Hu. 2013. "Discriminative Analysis of Non-Linear Brain Connectivity in Schizophrenia: An fMRI Study." *Frontiers in Human Neuroscience* 7 (October). frontiersin.org: 702.

Sundermann, B., D. Herr, W. Schwindt, and B. Pfleiderer. 2014. "Multivariate Classification of Blood Oxygen Level–Dependent fMRI Data with Diagnostic Intention: A Clinical Perspective." *AJNR. American Journal of Neuroradiology* 35 (5). Am Soc Neuroradiology: 848–55.

Tang, Yan, Lifeng Wang, Fang Cao, and Liwen Tan. 2012. "Identify Schizophrenia Using Resting-State Functional Connectivity: An Exploratory Research and Analysis." *Biomedical Engineering Online* 11 (August): 50.

Tan, Pang-Ning, Michael Steinbach, and Vipin Kumar. 2006. *Introduction to Data Mining*. 1st ed. Pearson Addison Wesley.

Thirion, Bertrand, Gaël Varoquaux, Elvis Dohmatob, and Jean-Baptiste Poline. 2014. "Which fMRI Clustering Gives Good Brain Parcellations?" *Frontiers in Neuroscience* 8 (January): 167.

Tu, Peichi, Randy L. Buckner, Lilla Zollei, Kara A. Dyckman, Donald C. Goff, and Dara S. Manoach. 2010. "Reduced Functional Connectivity in a Right-Hemisphere Network for Volitional Ocular Motor Control in Schizophrenia." *Brain: A Journal of Neurology* 133 (Pt 2): 625–37.

Tzourio-Mazoyer, N., B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. 2002. "Automated Anatomical Labeling of Activations in

SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain." *NeuroImage* 15 (1): 273–89.

Vapnik, Vladimir N. 1995. *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag New York, Inc.

Venkataraman, Archana, Marek Kubicki, Carl-Fredrik Westin, and Polina Golland. 2010. "Robust Feature Selection in Resting-State fMRI Connectivity Based on Population Studies." *Conference on Computer Vision and Pattern Recognition Workshops IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Workshops*, 63–70.

Venkataraman, Archana, Thomas J. Whitford, Carl-Fredrik Westin, Polina Golland, and Marek Kubicki. 2012. "Whole Brain Resting State Functional Connectivity Abnormalities in Schizophrenia." *Schizophrenia Research* 139 (1-3). Elsevier: 7–12.

Vinh, Nguyen Xuan, Julien Epps, and James Bailey. 2010. "Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance." *Journal of Machine Learning Research: JMLR* 11 (March). JMLR.org: 2837–54.

Wang, Qi, Rong Chen, Joseph JaJa, Yu Jin, L. Elliot Hong, and Edward H. Herskovits. 2016. "Connectivity-Based Brain Parcellation : A Connectivity-Based Atlas for Schizophrenia Research." *Neuroinformatics* 14 (1): 83–97.

Ward, Joe H., Jr. 1963. "Hierarchical Grouping to Optimize an Objective Function." *Journal of the American Statistical Association* 58 (301): 236–44.

Weinstein, G. W., G. B. Arden, R. A. Hitchings, S. Ryan, C. M. Calthorpe, and J. V. Odom. 1988. "The Pattern Electroretinogram (PERG) in Ocular Hypertension and Glaucoma." *Archives of Ophthalmology* 106 (7): 923–28.

Wolf, Robert Christian, Nenad Vasic, Fabio Sambataro, Annett Höse, Karel Frasch, Markus Schmid, and Henrik Walter. 2009. "Temporally Anticorrelated Brain Networks during Working Memory Performance Reveal Aberrant Prefrontal and Hippocampal Connectivity in Patients with Schizophrenia." *Progress in Neuro-Psychopharmacology & Biological Psychiatry* 33 (8): 1464–73.

Yang, Honghui, Jingyu Liu, Jing Sui, Godfrey Pearlson, and Vince D. Calhoun. 2010. "A Hybrid Machine Learning Method for Fusing fMRI and Genetic Data: Combining Both Improves Classification of Schizophrenia." *Frontiers in Human Neuroscience* 4 (October). frontiersin.org: 192.

Yeo, B. T. Thomas, Fenna M. Krienen, Jorge Sepulcre, Mert R. Sabuncu, Danial Lashkari, Marisa Hollinshead, Joshua L. Roffman, et al. 2011. "The Organization of the Human Cerebral Cortex Estimated by Intrinsic Functional Connectivity." *Journal of Neurophysiology* 106 (3): 1125–65.

Yoav Freund, Robert E. Schapire. 1999. "A Short Introduction to Boosting." *citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.107.3285*. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.107.3285.

Yoon, Jong H., Danh V. Nguyen, Lindsey M. McVay, Paul Deramo, Michael J. Minzenberg, J. Daniel Ragland, Tara Niendham, Marjorie Solomon, and Cameron S. Carter. 2012. "Automated Classification of fMRI during Cognitive Control Identifies More Severely Disorganized Subjects with Schizophrenia." *Schizophrenia Research* 135 (1-3). Elsevier: 28–33.

Yu, Yang, Hui Shen, Huiran Zhang, Ling-Li Zeng, Zhimin Xue, and Dewen Hu. 2013. "Functional Connectivity-Based Signatures of Schizophrenia Revealed by Multiclass Pattern Analysis of Resting-State fMRI from Schizophrenic Patients and Their

Healthy Siblings." *Biomedical Engineering Online* 12 (10). biomedcentral.com. http://www.biomedcentral.com/content/pdf/1475-925X-12-10.pdf.

Zarogianni, Eleni, Thomas W. J. Moorhead, and Stephen M. Lawrie. 2013. "Towards the Identification of Imaging Biomarkers in Schizophrenia, Using Multivariate Pattern Classification at a Single-Subject Level." *NeuroImage. Clinical* 3 (September). Elsevier: 279–89.

Zhang, Degang, Lei Guo, Xintao Hu, Kaiming Li, Qun Zhao, and Tianming Liu. 2012. "Increased Cortico-Subcortical Functional Connectivity in Schizophrenia." *Brain Imaging and Behavior* 6 (1): 27–35.

Zhou, Yuan, Meng Liang, Tianzi Jiang, Lixia Tian, Yong Liu, Zhening Liu, Haihong Liu, and Fan Kuang. 2007. "Functional Dysconnectivity of the Dorsolateral Prefrontal Cortex in First-Episode Schizophrenia Using Resting-State fMRI." *Neuroscience Letters* 417 (3): 297–302.