Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

Review Article

Complex biomarker discovery in neuroimaging data: Finding a needle in a haystack☆

Gowtham Atluri [a], Kanchana Padmanabhan [b], Gang Fang [c], Michael Steinbach [a], Jeffrey R. Petrella [d], Kelvin Lim [e], Angus MacDonald III [f], Nagiza F. Samatova [b], P. Murali Doraiswamy [g], Vipin Kumar [a,]*

[a] *Department of Computer Science and Engineering, University of Minnesota — Twin Cities, USA*
[b] *Department of Computer Science, North Carolina State University, USA*
[c] *Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, USA*
[d] *Department of Radiology, Duke University Medical Center, USA*
[e] *Department of Psychiatry, University of Minnesota — Twin Cities, USA*
[f] *Department of Psychology, University of Minnesota — Twin Cities, USA*
[g] *Department of Psychiatry and the Duke Institute for Brain Sciences, Duke University, USA*

ARTICLE INFO

Article history:
Received 8 March 2013
Received in revised form 27 June 2013
Accepted 16 July 2013
Available online xxxx

ABSTRACT

Neuropsychiatric disorders such as schizophrenia, bipolar disorder and Alzheimer's disease are major public health problems. However, despite decades of research, we currently have no validated prognostic or diagnostic tests that can be applied at an individual patient level. Many neuropsychiatric diseases are due to a combination of alterations that occur in a human brain rather than the result of localized lesions. While there is hope that newer imaging technologies such as functional and anatomic connectivity MRI or molecular imaging may offer breakthroughs, the single biomarkers that are discovered using these datasets are limited by their inability to capture the heterogeneity and complexity of most multifactorial brain disorders. Recently, complex biomarkers have been explored to address this limitation using neuroimaging data. In this manuscript we consider the nature of complex biomarkers being investigated in the recent literature and present techniques to find such biomarkers that have been developed in related areas of data mining, statistics, machine learning and bioinformatics.

Contents

1. Introduction

Public health consequences of neurological and mental disorders, such as Alzheimer's disease (AD), bipolar disorder, or schizophrenia (SZ) are enormous. Yet, critical needs for reliable biomarkers for early detection and prognostic prediction of such disorders are still unmet (Kubicki et al., 2007; MacDonald and Schulz, 2009; Pettersson-Yeo et al., 2011). The purpose of this article is to review different data mining, machine learning, and statistical techniques that can help unearth

* Corresponding author. Tel.: +1 612 624 8023; fax: +1 612 625 0572.
E-mail address: kumar@cs.umn.edu (V. Kumar).

2213-1582/$ – see front matter © 2013 The Authors. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.nicl.2013.07.004

neurological-disease relevant biomarkers using data from imaging studies. Since, the neuroimaging community has also been contributing to the development of informatics tools for biomarker discovery, the techniques reviewed include those that the neuroimaging community already uses as well. All these techniques could further improve the complex biomarker discovery process with eventual use in clinical setting.

Neuroimaging technologies such as volumetric MRI, functional MRI (fMRI) and diffusion tensor imaging (DTI) are in wide use to indirectly estimate altered cortical tissue, functional and physical connections in neuropsychiatric disease states (Honey et al., 2009; Park et al., 2008). Volumetric MRI measures the cortical thickness of a region, whereas fMRI and DTI allow one to construct a brain network for a subject where each defined region (e.g. dorsolateral prefrontal cortex, CA1 region in hippocampal) in the brain is termed as a "node" and a functional/physical connection (e.g. frontal-hippocampal connectivity) is termed as an "edge" (Bullmore and Bassett, 2011; Sporns, 2011). The volumetric features or the edges measured from fMRI or DTI, referred to as 'features' henceforth, provide an opportunity to study the altered properties underlying neuropsychiatric diseases. These features can be binary (representing a healthy volume of a region, or the presence or absence of a connection) or weighted (indicating volume or strength of a connection). Features of the brain measured from multiple subjects are then used to predict a phenotype of interest (Ragland et al., 2007). Phenotypes can be symptoms such as cognition, depression or mania, or a disease diagnosis such as SZ (Drevets and Todd, 2005). A set of features that show different properties in different subgroups of the phenotype is referred to as a "biomarker" in the rest of this paper. In the case of a binary biomarker, the set of features could be (mostly) present in subjects of one group and not present in the subjects of the other group, and in the case of a continuous biomarker they could have high values in one group and low values in the other group.

Research in neuroimaging data has focused on exploring the hypothesis that mental disorders manifest due to the loss of cortical tissue or altered connectivity in the brain, i.e., reduction in the temporal lobe volume, aberrant connectivity within the default network or attention network that in turn disrupts cognitive functions (Fornito and Harrison, 2012; Stephan et al., 2009). A vast majority of these studies (Jafri et al., 2008; Li et al., 2010; Liang et al., 2006; Luck et al., 2011) focus on discovering the features that individually show a different degree of volume or neural connectivity in disease subjects when compared with healthy subjects.

While insightful, this direction of research has not yet yielded any conclusive causal factors for major mental disorders. This is likely due to several well-known challenges. First, the large number of individual factors, such as thousands of edges, makes it difficult to find statistically significant single markers without sufficiently large study samples. In particular, multiple hypothesis testing resulting from the enormous number of potential hypotheses increases the chances of statistical errors, i.e., mistaking spurious patterns for real ones. Second, the complexity of the diseases being considered makes it unlikely that meaningful predictive patterns can be found by only looking at individual factors and largely ignoring their interrelations. Third, many diseases are heterogeneous by nature, i.e., patients with a particular disease may form different subgroups, and biomarkers appropriate for one subgroup may not apply to another. Given the inability of many commonly used analytic techniques to handle these challenges (statistical significance, disease complexity, and disease heterogeneity), it is no surprise that even when statistically significant biomarkers are found by one group in one study, they are rarely reproduced in follow-up studies by other groups or sometimes by even the same group (Kubicki et al., 2007; Pettersson-Yeo et al., 2011).

Research in biomarker discovery from neuroimaging data is at a crucial juncture where the field is beginning to acknowledge the need for complex multivariate analysis based techniques instead of currently used univariate analysis to capture the complex mechanisms underlying disease. Existing clinical studies demonstrate that there is an increase in predictive power for models built using a combination of imaging features when compared to that of single (Bressler and Menon, 2010; Westman et al., 2013; Wolz et al., 2011). Existing studies also show that although SZ is widely treated as a single phenotype, there exist two different subgroups of subjects (those with good outcome and those with poor outcome) that exhibit different structural properties in the brain (Mitelman et al., 2003; Nenadic et al., 2012). For example, subjects with poor outcome had significantly smaller temporal and occipital lobe gray matter volumes (Mitelman et al., 2003). These observations in early clinical studies (Bressler and Menon, 2010; Westman et al., 2013; Wolz et al., 2011) show the need to design computational methods that can be used to mine complex biomarkers.

There are several ways of defining complex biomarkers that are relevant to a neuropsychiatric disease. For example, a simultaneous reduction in volumes of multiple regions, or loss of a set of edges (e.g., left frontal–hippocampal connectivity plus right frontal–hippocampal connectivity) could result in a disease, even though a reduction in volume of one region or a loss of one edge (e.g., left frontal–hippocampal connectivity alone) does not result in a disease. In fact, a few recent studies including Westman et al. (2013) and Wolz et al. (2011)) have shown that models built using a combination of features result in more predictive power than univariate approaches. In contrast, it is possible that an fMRI study of a disease might find hundreds of edges altered when compared to controls, of which only the loss of a specific subset of edges might cause changes to the functional network structure that result in disease. Likewise, it is also possible that only the loss of a set of edges that belong to a specific functional group (e.g., "executive network") may result in loss of executive functioning in a disease such as SZ or geriatric depression. We refer to the edge sets belonging to a functional group as brain 'pathways'. Exploring such complex types of alterations in biomarker data could potentially improve the reproducibility and statistical power of imaging studies. This paper presents a set of techniques that attempt to identify complex biomarkers that may manifest in any of the above scenarios.

We define four types of complex biomarkers (and analytic techniques) based on different interesting combinations discussed above: (i) Linear biomarkers, (ii) Combinatorial biomarkers, (iii) Pathway biomarkers, and (iv) Network biomarkers. Several techniques developed in data mining, machine learning, and genomic data analysis communities can be helpful in discovering these four different types of biomarkers by overcoming some of the known challenges. A similar classification scheme for biomarkers has been used in genomic studies (Ayers and Cordell, 2010; Chuang et al., 2007; Fang et al., 2012a; Holden et al., 2008). A number of existing studies have analyzed neuroimaging data sets obtained from multiple technologies such as fMRI, DTI, and PET data collected on the same set of subjects to study group differences. However, in this manuscript we focus on only those studies that analyze one type of neuroimaging data set.

2. Linear biomarkers

Given a dataset of features (structural information, edges in anatomic or functional networks) obtained from the brain of several subjects and a continuous valued phenotype of interest for these subjects (e.g., cognition, psychosis ratings), a linear biomarker is a weighted sum of the features that is predictive of the phenotype. The computational problem here is the estimation of the weights such that the weighted sum is most predictive of the phenotype. A traditional approach to estimate these weights is to use a linear regression model (Friedman et al., 2001), where the features for a set of subjects are represented as matrix X, whose rows are subjects and columns are features obtained from neuroimaging techniques, as shown in Fig. 1(a). The phenotype is represented as a column vector, Y whose rows are subjects. The linear regression model then estimates a vector β, such that $Y = X\beta + \varepsilon$, where ε accounts for the error. The heart of the model $Y = X\beta$ is depicted in

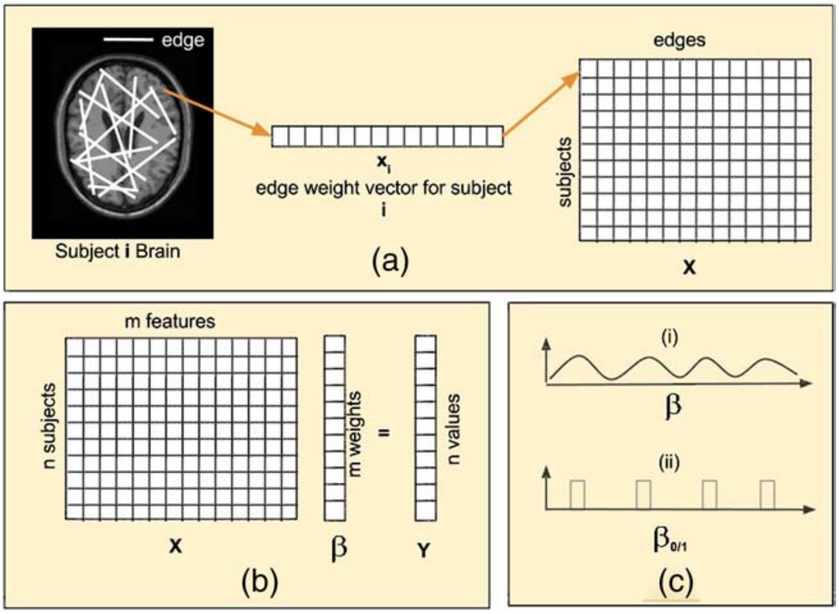

Fig. 1. Illustration of linear biomarker discovery: (a) matrix representation by treating edges in the brain as features, (b) linear regression setup where X represents the features (edges in brain networks or volumetric information) for all subjects, β represents the weights for features, and Y represents the phenotype value for each subject, and (c) resultant β from linear regression and LASSO.

Fig. 1(b). This model is solved such that the sum of the squared error is minimized, i.e., β assigns a weight to each feature in the dataset in such a way that the weighted sum of all features (Xβ) could approximate the phenotype (Y), with a minimal error (ε). The advantage of using linear regression based models lies in the availability of well-documented standard software. Using linear regression, Kubicki et al. (2011) showed that the gray matter volumes of Superior Temporal Gyrus and Inferior Frontal Gyrus, and the functional and anatomical connections between them were predictive of hallucinations in SZ. Note that individual correlations between each features and the phenotype did not yield any significant associations.

However there are several challenges that arise when linear regression is applied to neuroimaging datasets. One challenge is high dimensionality, i.e., the large number of features in the brain, e.g., a large number of edges as a result of the large number of nodes (voxels) in the brain that are of the order of 100,000. This leads to a computational challenge for a traditional linear regression scheme. Another challenge is that only a few features (e.g., a few functional edges in the brain out of the billions of edges) are expected to be associated with a given phenotype. A traditional linear regression model generally assigns weights over all the features in an effort to find the best association plausible. A typical weight assignment is similar to the one shown in Fig. 1(c). Although one could potentially select the features that are weighted highly by the model as relevant features for a given phenotype, it is often unclear what the right number of features is, and therefore such an approach could result in an erroneous discovery of associated features.

A variant of linear regression called LASSO (Least Absolute Shrinkage and Selection Operator) developed in the machine learning and statistics domains can address the two key challenges that arise for linear biomarker discovery (Friedman et al., 2001; Wu and Lange, 2008). LASSO introduces a penalty on β (in addition to the sum of squares of the elements) such that the absolute sum of all elements of β is small. When this model is used to estimate β using matrix X and the vector Y, it results in a β vector where most components are 0's and any non-zero element could be indicative of an edge relevant to the phenotype. Fig. 1(c) shows a typical vector that results from a LASSO type model, where only a few values in $\beta_{0/1}$ are nonzero. This model allows for automatic selection of relevant edges without having to choose a parameter for the number of features as in the case of a general linear regression model. Efficient approaches are available to handle the high dimensional nature of the datasets. LASSO type models were also found to be promising in genomic case–control data analysis, where there are tens to hundreds of samples and up to hundreds of thousands of genomic features like SNPs and gene expression (Ayers and Cordell, 2010; Beck et al., 2011; Ghosh and Chinnaiyan, 2005).

Linear biomarkers approaches have shown promise in discovering imaging features that could explain group differences in ADHD (Bohland et al., 2012), AD (Liu et al., 2012) and neuro-cognitive deficits (Bunea et al., 2011). Recently, Bohland et al. (2012)used a LASSO type approach to select relevant features from anatomical and functional network measures in combination with non-neuroimaging features to predict Attention deficit hyperactivity disorder (ADHD) in individual subjects among a group of mixed disease and controls. They noticed that the features selected from all three modalities resulted in the best performance on a test set. Liu et al. (2012) used a LASSO model with spatial constraints to find the set of imaging features (T1-weighted baseline MR brain images) that show increased prediction accuracy between AD and mild cognitive impairment. Bunea et al. (2011) demonstrated the use of penalized least squares regression approaches to predict neuro-cognitive deficits using a dataset that comprises of DTI and brain volumetric measures from HIV infected subjects. Logistic and linear regression models have been previously implemented in many statistical packages such as R, SAS, and Matlab and are easily available for use by the scientific community for analysis.

3. Combinatorial biomarkers

A key assumption underlying linear regression based techniques presented in the previous section is that the discovered biomarkers are valid across all the subjects in the study (i.e., disease is homogenous). However, this assumption does not always hold true, due to disease and population heterogeneity. Different subsets of patients tend to have different factors that drive the phenotype of interest. For example, about 50% of patients with Mild Cognitive Impairment (MCI) are amyloid scan positive but the other 50% are not, thus suggesting that MCI is a greatly heterogeneous condition. In this section, we focus on biomarkers that can capture the "subspace" scenarios. In particular, we focus on techniques from the data mining area of association analysis (Agrawal and Srikant, 1994; Han et al., 2007; Pang-Ning et al., 2006), which has well developed approaches for finding patterns (biomarkers) in data sets with binary features and binary outcomes (e.g., phenotype or disease label).

Given a dataset of neuroimaging features such as volumetric information or functional connections (edges), this information can be translated into a set of binary features, where each feature records the presence of a characteristic of interest with a 1, e.g., a volume being high or low, or an area being active or inactive. Presence of a feature is indicated by 1. A phenotype of interest (SZ and healthy) is also represented as a binary variable, typically with a 1 indicating presence of a disease and a 0 indicating absence (a control). A combinatorial biomarker is a subset of features that are present mostly in one group of subjects. Note that the combinatorial biomarker is only relevant for those subjects in which the subset of features are all present. Consider the example shown in Fig. 2(a), where a dataset X whose rows are subjects, columns are features, with values 1 (shown in black) are indicative of the presence of the features, while values 0 (shown in white) are indicative of a features absence. The grouping of subjects is represented by a column vector Y, where black color indicates SZ and white color indicates healthy. The submatrix A in X represents two features that are all present in a subset of four subjects and they belong to the SZ group. Therefore, A is a combinatorial biomarker that is associated with SZ. Note that there can be many combinatorial biomarkers in a given dataset. In this example, submatrix B is associated with SZ and submatrices C and D are associated with healthy subjects. One could argue that the features could be discovered by individual testing and then grouped together to recover the submatrices A, B, C, and D. However, there could be scenarios where the individual features themselves are not informative about the phenotype but together they have more information about the phenotype. Individually, the columns representing submatrix A in Fig. 2(b) are equally frequent in healthy and SZ groups, however the columns in A together are present only in the SZ group. Therefore, such biomarkers cannot be discovered using traditional linear regression type techniques or by univariate testing.

Combinatorial biomarkers are substantially different from linear biomarkers in that each combinatorial biomarker potentially explains a subset of subjects, whereas a linear biomarker is expected to cover all the subjects in the study. This gives combinatorial biomarkers more flexibility to capture the heterogeneous nature of the subjects and their associated signals in the data. For example, submatrices A and B cover different subsets of subjects that have the phenotype in Fig. 2(a). This strength however leads to additional challenges: computational complexity and statistical significance assessment. Approaches for discovering combinatorial biomarkers have to explore the space of all possible combinations of edges in the brain to discover these biomarkers exhaustively (Fang et al., 2012b). For a set of n edges the number of all possible combinations is of the order $2^n - 1$. This further leads to an additional challenge of statistical significance due to multiple hypothesis testing. When $2^n - 1$ hypotheses are tested, there is a much bigger chance for some of them to turn out to be true just by chance. Therefore, the statistical significance values have to be adjusted to account for this occurrence.

Efficient approaches to discover combinatorial biomarkers, referred to as pattern mining techniques, have been developed in the field of data mining in the last decade (Agrawal and Srikant, 1994; Han et al., 2007; Pang-Ning et al., 2006). These approaches were first designed to

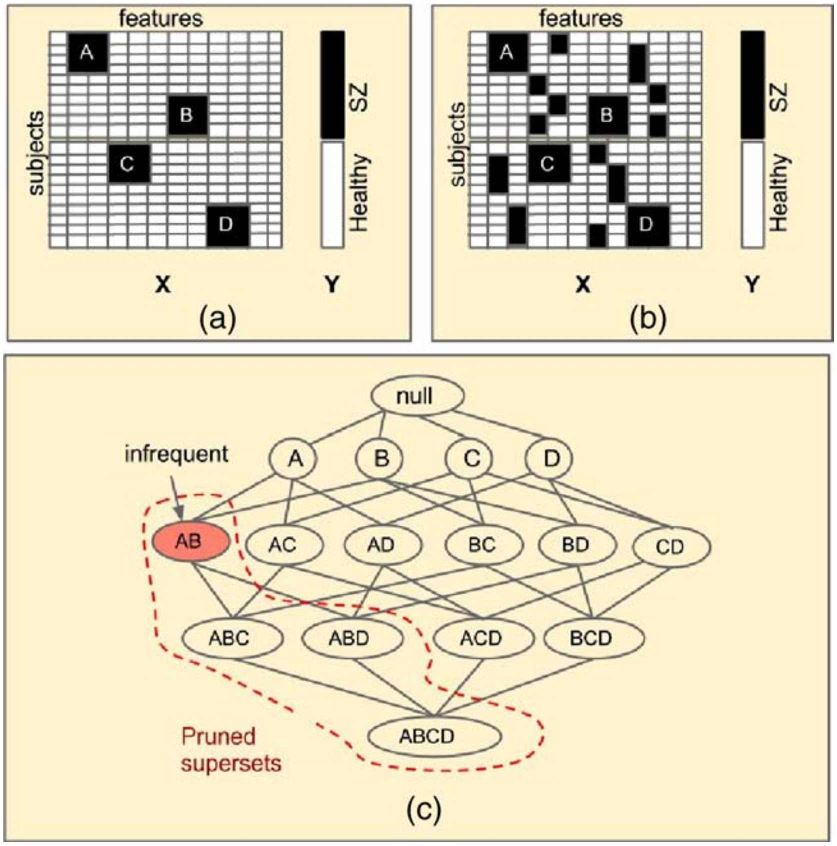

Fig. 2. Illustration of combinatorial biomarker discovery: (a and b) X is a hypothetical data matrix where columns represent features derived from neuroimaging data and rows represent subjects. The subjects belong to two groups Healthy and schizophrenia (SZ) as indicated by the column vector Y. In matrix X, an element (row, column) with black color indicates that the feature is present for a given subject. A, B, C, and D are interesting submatrices in X that have information about Y. The columns representing these submatrices in (a) are individually associated with Y, but those in (b) are not associated. (c) Efficient search space pruning: The Apriori principles allows pruning of supersets when a set is not interesting.

discover the combinations of items that are purchased together in large market basket datasets where each record (transaction) has a list of items that are purchased by one customer (Agrawal and Srikant, 1994; Agrawal et al., 1993). The pattern mining techniques draw their efficiency from the anti-monotonicity property which guarantees that if a combination of items is not frequently purchased together then a combination that includes these items is not frequently purchased too (Agrawal and Srikant, 1994). This property is also referred to as the "Apriori principle." Fig. 2(c) shows the set of all possible combinations of items (A, B, C, D) (Xiong et al., 2006) in the form of a lattice, where each node represents one combination of items. Frequent pattern mining techniques typically search this lattice depicting all possible combinations. Once a combination is found to be infrequent then all combinations that are extensions of the infrequent combinations are not enumerated. In Fig. 2(c), when the combination AB is found to be infrequent all its supersets are excluded from being enumerated and tested. Since the early 1990's, several efficient algorithms to explore the search space have been developed (Coatney and Parthasarathy, 2005; Han et al., 2000; Zaki, 2000; Zaki and Hsiao, 1999). Some of these algorithms have also been found to be promising in bioinformatics problems involving gene expression and protein interaction network datasets (Atluri et al., 2000; Atluri et al., 2009; Bellay et al., 2011; Gupta et al., 2011; Pandey et al., 2009).

Pattern mining techniques for discovering combinatorial biomarkers have been proposed for gene expression datasets (Fang et al., 2012a, 2012b), where the goal is to find combinations of genes that are all highly expressed in subjects with cancer and not expressed together in healthy subjects. These techniques have shown promise in discovering biomarkers from genomic lung cancer data sets. These techniques have also been extended to work with continuous valued gene expression datasets (Fang et al., 2010).

The challenge of statistical power posed by the large search space of combinatorial biomarkers can be overcome by providing a False Discovery Rate (FDR) and retaining only those biomarkers that are robust to multiple hypothesis testing. An approach to compute FDR for biomarkers is to first use randomized datasets to discover combinatorial biomarkers and then to compare the association strength of real biomarkers with those discovered from a randomized dataset. Note that this approach takes into account the multiple hypothesis testing as the combinatorial biomarkers are discovered from real and randomized datasets by exploring an exponentially large search space.

BENCH (Biclustering-driven ENsemble of Classifiers), developed by Padmanabhan et al. (2012), is another combinatorial biomarker discovery approach specifically designed for highly underdetermined problems (i.e., the number of features is much higher than the number of subjects/patients). The method is specifically tailored for the cases that exhibit different discriminatory signatures between subgroups of samples without any *prior* knowledge about subgroupings. These combinatorial techniques would represent a novel approach to discovery of large-scale connectivity biomarkers in neuroimaging data. Because these approaches were primarily designed for binarized data, potential loss of information has dissuaded its use in clinical studies.

To the best of our knowledge, combinatorial biomarker based approaches have not been used in neuroimaging literature. One reason for this is the lack of strategies to transform continuous features obtained from neuroimaging technologies to binary features that most combinatorial techniques work with. This gap needs to be addressed before new studies could reap the benefits of these approaches to explore combinations of features effectively as well as their ability to discover subgroups in disease subjects.

4. Pathway biomarkers

The functionality of the brain is known to be a coordinated effort of multiple regions. For example, the brain processes sensory information with the help of a salience network that encompasses functional connections between bilateral insula and anterior cingulate cortex. Some known brain subnetworks are: (i) default mode network (DMN), (ii) salience network (SN), and (iii) central executive network (CEN) (Lee et al., 2012). Exploring the association of these brain subnetworks with disease will enable researchers to study the relationship between these subnetworks and their role in mental disorders. Motivated by the progress of finding associations between known biological pathways and common complex diseases in genomic data analysis, we refer to these type of biomarkers as 'pathway' biomarkers (Holden et al., 2008; Medina et al., 2009; Subramanian et al., 2005; Vandin et al., 2011; Wang et al., 2009; Zhang et al., 2010). The fact that such biomarkers conform to existing knowledge allows investigators to interpret their role in disease. In fact, a few neuroimaging studies (Calhoun et al., 2008; Kim et al., 2010; Öngür et al., 2010; Palaniyappan et al., 2010; Sun et al., 2009; White et al., 2010; Woodward et al., 2011) have investigated the association of known subsystems with a disease and found promising results. For example, Woodward et al. (2011) found association of functional connections within the DMN and CEN network with SZ.

The most common connectivity biomarker tested in AD is the DMN, which has shown direct correlations with edges of the network and cognition (Hedden et al., 2009). There have been multiple attempts to use measures of DMN function as a biomarker for early diagnosis (e.g., Greicius et al. (Greicius, 2008)); however, intersubject variability is currently too high for use in individual subjects. Most studies in this category typically choose a subnetwork of interest and investigate its association with the disease; this may result in spurious findings, as the subnetworks not considered in the study could be more relevant to AD. Therefore, these subnetworks should be studied in comparison with each other and not in isolation. As such, data-driven approaches to exploring multiple functional networks, such as the twelve resting state networks identified by Greicius (2008), have the potential for enhanced accuracy.

One simplistic approach to discover associations of brain pathways with a phenotype is to compute the significance score of association for every edge in a brain network and then test the statistical significance of association of a brain pathway based on the scores of its constituent edges. This framework is shown in Fig. 3, where each edge in the brain network is referred to as a feature, and groups of features that are in known brain pathways are referred to as functional groups. The significance of association for a brain pathway, generally referred to as enrichment score, is obtained by comparing the distribution of association scores of its constituent edges with that of association scores from random selection of edges. A related approach to discover brain pathway associations is to first rank all edges based on their association with the phenotype and test each brain pathway if their constituent edges are all at the highly associated end of the ranking. Permutation

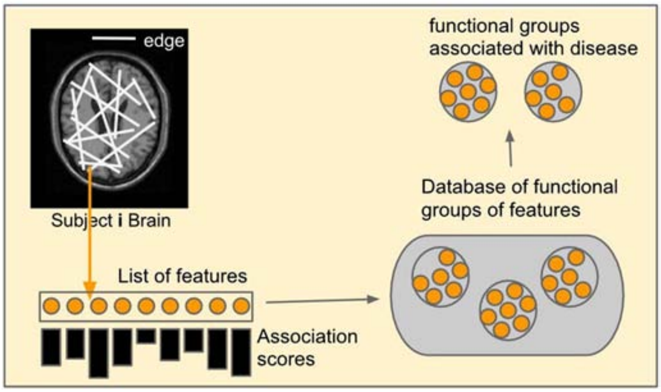
edge
Subject i Brain
List of features
Association scores
functional groups associated with disease
Database of functional groups of features

Fig. 3. Illustration of a 'pathway' based biomarker discovery approach. The features (often edges in the brain networks) are evaluated individually and then the functional groups (resting state networks) are evaluated for enrichment with highly significant features (edges).

based testing can also be used to quantify the significance of the brain pathway associations. A similar approach has been pursued in genetic association studies, referred to as Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005). Several extensions of GSEA and other related approaches have been proposed in the literature (Holden et al., 2008; Medina et al., 2009; Wang et al., 2009). (A good survey of these approaches is available in Wang et al. (2010).) These variations include the choice of scoring a genomic feature for association (e.g., p-values from *t*-test or chi-squared test), determining the enrichment score for a pathway as the minimum of p-values of the features contained in the pathway (Medina et al., 2009) and choice in the approach for estimating the statistical significance for pathway enrichment (e.g., phenotype based permutation of feature set permutation). Note that the success of this class of approaches is limited by the strength of association of individual features.

Variants of linear and combinatorial biomarkers have the potential to address this limitation. A variant of LASSO, group-LASSO, can select a set of the edges in the dataset that are known to be part of brain subnetwork and are associated with the phenotype. Group-LASSO techniques select all or none of the edges from a given group when they estimate β. This approach generally discovers the best brain subnetwork that is associated with the phenotype in question. Moreover, it has the potential to discover combinations that can be formed by features that may not be individually associated with the phenotype in question. Discriminative pattern mining techniques can also be used to discover pathway biomarkers by constraining the search space of the patterns to only those that fall under known brain pathways. This reduces the computational complexity of the pattern mining technique and can also improve the statistical significance as the number of hypotheses generated is restricted to those groups of edges that fall within known pathways.

The DENSE (Dense and ENriched Subgraph Enumeration) method developed by Hendrix et al. (2011) is a fast and theoretically guaranteed method that could take in as input a *prior* knowledge defined as a set of query nodes from a brain network and enumerate all the dense subnetworks in the brain network that contain user-defined percentage of the query nodes. While this method may not be directly applicable to identifying biomarkers common to a group of subjects as it works on one network at a time, it is, however, very useful to refine biomarkers identified. If the nodes of a brain pathway can be provided as input, then a particular subject's network could be analyzed to identify some of the peripheral nodes and edges that are associated with the biomarker that can offer more information about the subject under analysis.

There are several clinical applications of pathway biomarker type approaches in the context of investigating markers for SZ (Mamah et al., 2013; Orliac et al., 2013; Tu et al., 2013) and bipolar disorder (Mamah et al., 2013). Mamah et al. (2013), studied evaluated the role of mean connectivity (obtained from resting state fMRI data) of within five known neural subnetworks (default mode, fronto-parietal, cingulo-opercular, cerebellar, and salience networks) in SZ and bipolar disorder. They found that the decrease in within-connectivity in cingulo-opercular subnetwork is large in degree in SZ than in bipolar disorder. Orliac et al. (2013) studied the functional connectivity within DMN and SN in SZ, while Tu et al. (2013) studied disconnectivity in fronto-parietal network in SZ. While these studies show the usefulness of discovering pathway biomarkers, the methodologies discussed above will provide a systematic way to discover them.

5. Network biomarkers

Neuroimaging data obtained using fMRI or DTI is naturally represented in the form of a network, where nodes are brain regions and edges represent connections (physical or functional) (Bullmore and Sporns, 2009). In this context, we define network biomarkers as features of the network that could explain group differences between healthy and disease subjects. These features could be topological characteristics of nodes, or subnetworks that have significantly different topological properties in the two groups.

Topological properties of brain networks have the potential to offer insights into the functionality of the brain (Rubinov and Sporns, 2010). An extensive number of studies have pursued the goal of studying how topological properties differentiate in healthy and subjects from those with a brain disorder (Camchong et al., 2011; Liu et al., 2008; Lynall et al., 2010; Rotarska-Jagiela et al., 2010; Stam et al., 2007). Table 1 presents a representative sample of these studies listing different topological properties, including degree of a node, clustering coefficient, robustness and efficiency, considered in each of these studies. Graph-theoretic approaches have been applied to AD Supekar et al. (2008), demonstrating a loss of small-world properties of whole-brain networks, and modest correlation with cognitive status. In addition, this approach has been used to look at the impact of different lesion patterns (e.g., diffuse vs. hub-targeted attacks) on global metrics.

Given the complex nature of mental disorders such as AD and SZ, subgraph approaches that focus only on portions of the network may yield more accurate correlations with the disease in question. For example, a subnetwork in the brain can show different topological properties in healthy and disease groups that cannot be reflected in the individual properties of a node or an edge. For this reason, a set of nodes in a network that exhibits different topological properties in disease and healthy groups of subjects can also be treated as a network biomarker. One example is a group of nodes that are densely connected in one group of subjects compared to the other group (as shown in Fig. 4). Another example is a subset of nodes whose diameter in the induced subgraph is different between the two groups. Yet another example is a subset of nodes that play a critical role in the connectivity (in effect

Table 1
A selective sample of studies that use network topological properties to explain group differences in brain networks. rsfMRI: resting state fMRI data, MEG: Magnetoencephalography, EEG: Electroencephalography MRI: Magnetic Resonance Imaging.
Network topological properties: D: degree, CC: clustering coefficient, CPL: characteristic path length, LE: local efficiency, GE: global efficiency, H: hubs, M: modularity, SW: small worldness, R: robustness, CS: connection strength, CV: connectivity variance, CD: connection distance, LCC: largest connected component, C: centrality. Rubinov and Sporns (2010) provides a definition for all these properties and discusses their usefulness in interpreting brain networks.

Citation	Phenotype	Neuro-imaging Data	Network Properties
Alexander-Bloch et al. (2010)	Childhood-onset schizophrenia	rsfMRI	LE, CC, M, SW, R
Yu et al. (2011a)	Schizophrenia	rsfMRI	CS, CC, H
Bassett et al. (2012)	Schizophrenia	rsfMRI	CS, CV, LCC
Alexander-Bloch et al. (2013)	Childhood-onset schizophrenia	rsfMRI	GE, CC, SW, M, CD
Yu et al. (2011b)	Schizophrenia	rsfMRI	D, CS, LE, GE, CPL, CC, SW
Liu et al. (2013)	Alzheimer's	rsfMRI	CD, CC, GE
Zhang et al. (2011)	Major depressive disorder	rsfMRI	SW, GE, C
Alexander-Bloch et al. (2012)	Childhood-onset schizophrenia	rsfMRI	M
Lynall et al. (2010)	Schizophrenia	rsfMRI	CC, SW, R
Wang et al. (2012)	Amnestic mild cognitive impairment	rsfMRI	CC, CPL, M, CS
Stam et al. (2009)	Alzheimer's	MEG	CC, CPL, R
Supekar et al. (2008)	Alzheimer's	rsfMRI	CC, CPL
Buckner et al. (2009)	Alzheimer's	rsfMRI	H
Chen et al. (2011)	Aging	MRI	M
Bassett et al. (2008)	Schizophrenia	MRI	D, CPL, CC, SW
Wu et al. (2012)	Aging	MRI	SW, M
Jalili and Knyazeva (2011)	Schizophrenia	EEG	SW, R, M
Cole et al. (2012)	Cognitive control and intelligence	rsfMRI	CS
Wu et al. (2012)	Schizophrenia	rsfMRI	D, CS, CC, CPL, GE, LE
Liu et al. (2008)	Schizophrenia	rsfMRI	D, CC, CPL, GE, LE, SW

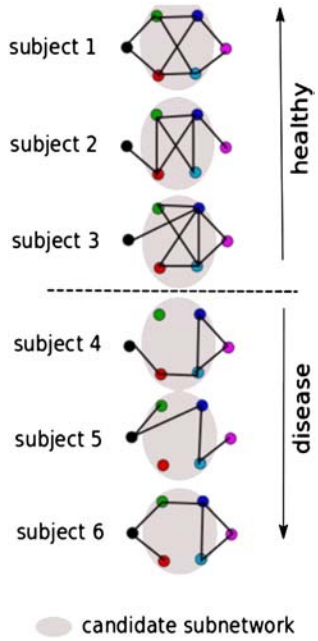

Fig. 4. Illustration of a subgraph discriminating between three healthy subjects and three disease subjects. The figure shows 6 networks from 3 healthy and 3 disease subjects. The shaded region in these networks covers nodes that are densely connected in healthy subjects and sparsely connected in disease subjects. Discovering such novel sets of nodes or subnetworks is essential.

functionality) of the entire system (network), i.e., removing those set of edges could affect the connectivity in one group of subjects more than the other group. These examples illustrate how network structure allows one to measure the impact of a selected set of nodes on the system (brain) as a whole and to understand the nature of connectivity within the subset of nodes to study the relationship between subgraph connectivity and the disease in question.

The key difference between the above examples of network biomarkers and the pathway biomarkers is that pathway biomarkers work with known subnetworks and test for their hypothesis-driven association with disease, whereas network biomarkers find subnetworks that are associated with the disease in an unbiased manner. Thus, conceptually all pathway biomarkers can be considered to be a subtype of network biomarkers. The advantage of hypothesis-driven focused pathway biomarker analyses is that the findings are bound to comply with well-studied subnetworks in the brain, are easy to interpret, and less subject to spurious findings. The disadvantage of using pathway biomarkers is that with this approach it may be impossible to find novel subnetworks that are hidden in the data but truly associated with a disease. Further, many of our a priori assumptions may be wrong and in that case, searching only for known pathways may result in a confirmation bias and limit our ability to find true causes of disease. Network biomarkers are appropriate in these scenarios, where a global unbiased search of the data is required. However, they are computationally more intensive given the size of the search space of all possible subnetworks.

One approach to derive subgraphs in the brain network whose dysfunction resulted in the manifestation of a phenotype is to first find the edges that are associated with the phenotype individually; construct a network of these associated edges, and discover significantly densely connected regions in this network. Zalesky et al's Network Based Statistic (NBS) approach (Zalesky et al., 2010) works in a similar fashion and it discovers the largest connected component in the network of significantly associated edges. Similar approaches have also been employed in genomic case–control data analysis to identify protein networks that are associated in cancer (Chuang et al., 2007; Ideker and Sharan, 2008; Karni et al., 2009; Vandin et al., 2011; Wang et al., 2011). The advantage of these approaches is that even when the strength of the strongly associated edges is not statistically significant, the subnetworks discovered can be statistically significant if they form a connected structure. A drawback of NBS approaches is that they cannot discover subnetworks when the individual edges are not associated with the phenotype, but when they are collectively associated. For example, consider a scenario where two edges (frontal–caudate and frontal–amygdala) connect three brain regions (e.g., caudate, amygdala, frontal lobe) that interactively accomplish a task (mood); each edge by itself cannot capture this synergy of all the three relevant regions and so the above mentioned approach will not find the individual edges to be associated with the task and hence the synergistic system will be missed.

A suite of network biomarker discovery techniques (Chen et al., 2012; Padmanabhan et al., 2012; Schmidt and Samatova, 2009; Schmidt et al., 2012) proposed in genomic data analysis can be potentially used in the context of neuroimaging datasets. Schmidt and Samatova (2009) α,β-motif finder algorithm is designed to discover cliques (a subgraph where every node is connected to every other node) in an underlying network that is significantly associated with a phenotype. In order to discover general network biomarkers, beyond cliques, Padmanabhan et al. (2012) proposed an approach to find connected subgraph biomarkers. SPICE (System Phenotype-related Interplaying Components Enumerator) (Chen et al., 2012) was proposed to discover subgraphs that explain only subsets of subjects. These techniques can significantly improve the state-of-the-art in network biomarker discovery from brain networks.

Simpler versions of network biomarkers have been used in analyzing neuroimaging data in the past, as shown in Table 1. However, the above discussed network biomarker approaches are yet to be applied to this data to discover complex variants of network biomarkers.

6. Concluding remarks

In this manuscript we considered the nature of complex biomarkers being investigated in the recent literature and presented techniques that are designed in related areas of data mining, statistics, machine learning and bioinformatics. Most of the techniques presented here have been refined for over a decade since their inception and so they can be directly applied to study the hypotheses being considered in neuroimaging studies. Thus there is significant potential for advancing the state of the art in complex biomarker discovery for neuroimaging data.

Specifically, the current state of the art provides neuroimaging based biomarkers that are typically based on single features and are good indicators of the mental disorder after the disorder has begun for some disorders, e.g., AD (Linden, 2012). However, complex neuroimaging biomarkers hold out the promise helping predict high risk subjects before a disease, can also better help understand the differences between various mental disorders, e.g., schizophrenia and bipolar, and can provide insights where several subgroups exists, e.g., schizophrenia. The techniques covered in this manuscript have demonstrated their ability to find complex biomarkers in other domains and offer a new and promising set of new tools for neuroimaging investigators. Indeed, since many of these techniques have already been applied to genetic and clinical data, there is a real possibility of finding complex biomarkers spanning multimodal data, thus further enhancing the breadth and depth of our understanding of neurological disorders.

Acknowledgments

PMD has received research grants and advisory fees from several pharmaceutical and imaging companies for other studies. He owns stock in Sonexa and Clarimedix whose products are not discussed here. JRP has received research grants and advisory fees from pharmaceutical companies.

Lim has received research grants Schizophrenia R01 MH0606621 and NIDA CSIA P20DA024196. Kumar and Steinbach have received NSF research grant IIS-0916439.

References

Agrawal, R., Srikant, R., 1994. Fast algorithms for mining association rules. Proc 20th Int Conf Very Large Data Bases, VLDB. , vol. 1215, pp. 487–499.

Agrawal, R., Imieliński, T., Swami, A., 1993. Mining association rules between sets of items in large databases. ACM SIGMOD Record. , vol. 22. ACM, pp. 207–216.

Alexander-Bloch, A.F., Gogtay, N., Meunier, D., Birn, R., Clasen, L., Lalonde, F., Lenroot, R., Giedd, J., Bullmore, E.T., 2010. Disrupted modularity and local connectivity of brain functional networks in childhood-onset schizophrenia. Frontiers in Systems Neuroscience 4.

Alexander-Bloch, A., Lambiotte, R., Roberts, B., Giedd, J., Gogtay, N., Bullmore, E., 2012. The discovery of population differences in network community structure: new methods and applications to brain functional networks in schizophrenia. NeuroImage 59, 3889–3900.

Alexander-Bloch, A.F., Vértes, P.E., Stidd, R., Lalonde, F., Clasen, L., Rapoport, J., Giedd, J., Bullmore, E.T., Gogtay, N., 2013. The anatomical distance of functional connections predicts brain network topology in health and schizophrenia. Cerebral Cortex 23, 127–138.

Atluri, G., Gupta, R., Fang, G., Pandey, G., Steinbach, M., Kumar, V., 2009. Association analysis techniques for bioinformatics problems. Bioinformatics and Computational Biology 1–13.

Atluri, G., Bellay, J., Pandey, G., Myers, C., Kumar, V., 2000. Discovering coherent value bicliques in genetic interaction data. In Proceedings of 9th International Workshop on Data Mining in Bioinformatics (BIOKDD'10).

Ayers, K.L., Cordell, H.J., 2010. SNP Selection in genome-wide and candidate gene studies via penalized logistic regression. Genetic Epidemiology 34, 879–891.

Bassett, D.S., Bullmore, E., Verchinski, B.A., Mattay, V.S., Weinberger, D.R., Meyer-Lindenberg, A., 2008. Hierarchical organization of human cortical networks in health and schizophrenia. The Journal of Neuroscience 28, 9239–9248.

Bassett, D.S., Nelson, B.G., Mueller, B.A., Camchong, J., Lim, K.O., 2012. Altered resting state complexity in schizophrenia. NeuroImage 59, 2196–2207.

Beck, A.H., Sangoi, A.R., Leung, S., Marinelli, R.J., Nielsen, T.O., van de Vijver, M.J., West, R.B., van de Rijn, M., Koller, D., 2011. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. Science Translational Medicine 3, 108–113.

Bellay, J., Atluri, G., Sing, T.L., Toufighi, K., Costanzo, M., Ribeiro, P.S.M., Pandey, G., Baller, J., VanderSluis, B., Michaut, M., 2011. Putting genetic interactions in context through a global modular decomposition. Genome Research 21, 1375–1387.

Bohland, J.W., Saperstein, S., Pereira, F., Rapin, J., Grady, L., 2012. Network, anatomical, and non-imaging measures for the prediction of ADHD diagnosis in individual subjects. Frontiers in Systems Neuroscience 6.

Bressler, S.L., Menon, V., 2010. Large-scale brain networks in cognition: emerging methods and principles. Trends in Cognitive Sciences 14, 277–290.

Buckner, R.L., Sepulcre, J., Talukdar, T., Krienen, F.M., Liu, H., Hedden, T., Andrews-Hanna, J.R., Sperling, R.A., Johnson, K.A., 2009. Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. The Journal of Neuroscience 29, 1860–1873.

Bullmore, E.T., Bassett, D.S., 2011. Brain graphs: graphical models of the human brain connectome. Annual Review of Clinical Psychology 7, 113–140.

Bullmore, E., Sporns, O., 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience 10, 186–198.

Bunea, F., She, Y., Ombao, H., Gongvatana, A., Devlin, K., Cohen, R., 2011. Penalized least squares regression methods and applications to neuroimaging. NeuroImage 55, 1519–1527.

Calhoun, V.D., Maciejewski, P.K., Pearlson, G.D., Kiehl, K.A., 2008. Temporal lobe and "default" hemodynamic brain modes discriminate between schizophrenia and bipolar disorder. Human Brain Mapping 29, 1265–1275.

Camchong, J., MacDonald, A.W., Bell, C., Mueller, B.A., Lim, K.O., 2011. Altered functional and anatomical connectivity in schizophrenia. Schizophrenia Bulletin 37, 640–650.

Chen, Z.J., He, Y., Rosa-Neto, P., Gong, G., Evans, A.C., 2011. Age-related alterations in the modular organization of structural cortical network by using cortical thickness from MRI. NeuroImage 56, 235.

Chen, Z., Padmanabhan, K., Rocha, A.M., Shpanskaya, Y., Mihelcic, J., Scott, K., Samatova, N.F., 2012. SPICE: discovery of phenotype-determining component interplays. BMC Systems Biology 6, 40.

Chuang, H.-Y., Lee, E., Liu, Y.-T., Lee, D., Ideker, T., 2007. Network-based classification of breast cancer metastasis. Molecular Systems Biology 3.

Coatney, M., Parthasarathy, S., 2005. MotifMiner: efficient discovery of common substructures in biochemical molecules. Knowledge and Information Systems 7, 202–223.

Cole, M.W., Yarkoni, T., Repovš, G., Anticevic, A., Braver, T.S., 2012. Global connectivity of prefrontal cortex predicts cognitive control and intelligence. The Journal of Neuroscience 32, 8988–8999.

Drevets, W.C., Todd, R.D., 2005. Depression, Mania, and Related Disorders.

Fang, G., Kuang, R., Pandey, G., Steinbach, M., Myers, C.L., Kumar, V., 2010. Subspace differential coexpression analysis: problem definition and a general approach. Proceedings of the 15th Pacific Symposium on Biocomputing (PSB). vol. 15. Citeseer, pp. 145–156.

Fang, G., Haznadar, M., Wang, W., Yu, H., Steinbach, M., Church, T.R., Oetting, W.S., Van Ness, B., Kumar, V., 2012a. High-order snp combinations associated with complex diseases: efficient discovery, statistical power and functional interactions. PloS One 7, e33531.

Fang, G., Pandey, G., Wang, W., Gupta, M., Steinbach, M., Kumar, V., 2012b. Mining low-support discriminative patterns from dense and high-dimensional data. IEEE Transactions on Knowledge and Data Engineering 24, 279–294.

Fornito, A., Harrison, B.J., 2012. Brain connectivity and mental illness. Frontiers in Psychiatry 3.

Friedman, J., Hastie, T., Tibshirani, R., 2001. The Elements of Statistical Learning: Springer Series in Statistics.

Ghosh, D., Chinnaiyan, A.M., 2005. Classification and selection of biomarkers in genomic data using LASSO. Journal of Biomedicine and Biotechnology 2005, 147–154.

Greicius, M., 2008. Resting-state functional connectivity in neuropsychiatric disorders. Current Opinion in Neurology 21, 424–430.

Gupta, R., Rao, N., Kumar, V., 2011. Discovery of error-tolerant biclusters from noisy gene expression data. BMC Bioinformatics 12, S1.

Han, J., Pei, J., Yin, Y., 2000. Mining frequent patterns without candidate generation. ACM SIGMOD Record. , vol. 29. ACM, pp. 1–12.

Han, J., Cheng, H., Xin, D., Yan, X., 2007. Frequent pattern mining: current status and future directions. Data Mining and Knowledge Discovery 15, 55–86.

Hedden, T., Van Dijk, K.R., Becker, J.A., Mehta, A., Sperling, R.A., Johnson, K.A., Buckner, R.L., 2009. Disruption of functional connectivity in clinically normal older adults harboring amyloid burden. The Journal of Neuroscience 29, 12686–12694.

Hendrix, W., Rocha, A.M., Padmanabhan, K., Choudhary, A., Scott, K., Mihelcic, J.R., Samatova, N.F., 2011. DENSE: efficient and prior knowledge-driven discovery of phenotype-associated protein functional modules. BMC Systems Biology 5, 172.

Holden, M., Deng, S., Wojnowski, L., Kulle, B., 2008. GSEA-SNP: applying gene set enrichment analysis to SNP data from genome-wide association studies. Bioinformatics 24, 2784–2785.

Honey, C., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.-P., Meuli, R., Hagmann, P., 2009. Predicting human resting-state functional connectivity from structural connectivity. Proceedings of the National Academy of Sciences 106, 2035–2040.

Ideker, T., Sharan, R., 2008. Protein networks in disease. Genome Research 18, 644–652.

Jafri, M.J., Pearlson, G.D., Stevens, M., Calhoun, V.D., 2008. A method for functional network connectivity among spatially independent resting-state components in schizophrenia. NeuroImage 39, 1666.

Jalili, M., Knyazeva, M.G., 2011. EEG-based functional networks in schizophrenia. Computers in Biology and Medicine 41, 1178–1186.

Karni, S., Soreq, H., Sharan, R., 2009. A network-based method for predicting disease-causing genes. Journal of Computational Biology 16, 181–189.

Kim, D.I., Sui, J., Rachakonda, S., White, T., Manoach, D.S., Clark, V., Ho, B.-C., Schulz, S.C., Calhoun, V.D., 2010. Identification of imaging biomarkers in schizophrenia: a coefficient-constrained independent component analysis of the mind multi-site schizophrenia study. Neuroinformatics 8, 213–229.

Kubicki, M., McCarley, R., Westin, C.-F., Park, H.-J., Maier, S., Kikinis, R., Jolesz, F.A., Shenton, M.E., 2007. A review of diffusion tensor imaging studies in schizophrenia. Journal of Psychiatric Research 41, 15–30.

Kubicki, M., Alvarado, J.L., Westin, C.-F., Tate, D.F., Markant, D., Terry, D.P., Whitford, T.J., De Siebenthal, J., Bouix, S., McCarley, R.W., 2011. Stochastic tractography study of inferior frontal gyrus anatomical connectivity in schizophrenia. NeuroImage 55, 1657–1664.

Lee, M.H., Hacker, C.D., Snyder, A.Z., Corbetta, M., Zhang, D., Leuthardt, E.C., Shimony, J.S., 2012. Clustering of resting state networks. PLoS One 7, e40370.

Li, X., Branch, C.A., Nierenberg, J., DeLisi, L.E., 2010. Disturbed functional connectivity of cortical activation during semantic discrimination in patients with schizophrenia and subjects at genetic high-risk. Brain Imaging and Behavior 4, 109–120.

Liang, M., Zhou, Y., Jiang, T., Liu, Z., Tian, L., Liu, H., Hao, Y., 2006. Widespread functional disconnectivity in schizophrenia with resting-state functional magnetic resonance imaging. Neuroreport 17, 209–213.

Linden, D., 2012. The challenges and promise of neuroimaging in psychiatry. Neuron 73, 8.

Liu, Y., Liang, M., Zhou, Y., He, Y., Hao, Y., Song, M., Yu, C., Liu, H., Liu, Z., Jiang, T., 2008. Disrupted small-world networks in schizophrenia. Brain 131, 945–961.

Liu, M., Zhang, D., Yap, P.-T., Shen, D., 2012. Tree-guided sparse coding for brain disease classification. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012. Springer, pp. 239–247.

Liu, Y., Yu, C., Zhang, X., Liu, J., Duan, Y., Alexander-Bloch, A.F., Liu, B., Jiang, T., Bullmore, E., 2013. Impaired long distance functional connectivity and weighted network architecture in Alzheimer's disease. Cerebral Cortex (in press).

Luck, D., Buchy, L., Czechowska, Y., Bodnar, M., Pike, G.B., Campbell, J.S., Achim, A., Malla, A., Joober, R., Lepage, M., 2011. Fronto-temporal disconnectivity and clinical short-term outcome in first episode psychosis: a DTI-tractography study. Journal of Psychiatric Research 45, 369–377.

Lynall, M.-E., Bassett, D.S., Kerwin, R., McKenna, P.J., Kitzbichler, M., Muller, U., Bullmore, E., 2010. Functional connectivity and brain networks in schizophrenia. The Journal of Neuroscience 30, 9477–9487.

MacDonald, A.W., Schulz, S.C., 2009. What we know: findings that every theory of schizophrenia should explain. Schizophrenia Bulletin 35, 493–508.

Mamah, D., Barch, D.M., Repovš, G., 2013. Resting state functional connectivity of five neural networks in bipolar disorder and schizophrenia. Journal of Affective Disorders (in press).

Medina, I., Montaner, D., Bonifaci, N., Pujana, M.A., Carbonell, J., Tarraga, J., Al-Shahrour, F., Dopazo, J., 2009. Gene set-based analysis of polymorphisms: finding pathways or biological processes associated to traits in genome-wide association studies. Nucleic Acids Research 37, W340–W344.

Mitelman, S.A., Shihabuddin, L., Brickman, A.M., Hazlett, E.A., Buchsbaum, M.S., 2003. MRI assessment of gray and white matter distribution in Brodmann's areas of the cortex in patients with schizophrenia with good and poor outcomes. The American Journal of Psychiatry 160, 2154–2168.

Nenadic, I., Gaser, C., Sauer, H., 2012. Heterogeneity of brain structural variation and the structural imaging endophenotypes in schizophrenia. Neuropsychobiology 66, 44–49.

Öngür, D., Lundy, M., Greenhouse, I., Shinn, A.K., Menon, V., Cohen, B.M., Renshaw, P.F., 2010. Default mode network abnormalities in bipolar disorder and schizophrenia. Psychiatry Research: Neuroimaging 183, 59–68.

Orliac, F., Naveau, M., Joliot, M., Delcroix, N., Razafimandimby, A., Brazo, P., Dollfus, S., Delamillieure, P., 2013. Links among resting-state default-mode network, salience network, and symptomatology in schizophrenia. Schizophrenia Research 148, 74–80.

Padmanabhan, K., Wilson, K., Rocha, A.M., Wang, K., Mihelcic, J.R., Samatova, N.F., 2012. In-silico identification of phenotype-biased functional modules. Proteome Science 10, S2.

Palaniyappan, L., Mallikarjun, P., Joseph, V., White, T., Liddle, P., 2010. Reality distortion is related to the structure of the salience network in schizophrenia. Psychological Medicine 13, 1–8.

Pandey, G., Atluri, G., Steinbach, M., Myers, C.L., Kumar, V., 2009. An association analysis approach to biclustering. Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 677–686.

Pang-Ning, T., Steinbach, M., Kumar, V., 2006. Introduction to Data Mining. WP Co.

Park, C.-h., Kim, S.Y., Kim, Y.-H., Kim, K., 2008. Comparison of the small-world topology between anatomical and functional connectivity in the human brain. Physica A: statistical mechanics and its applications 387, 5958–5962.

Pettersson-Yeo, W., Allen, P., Benetti, S., McGuire, P., Mechelli, A., 2011. Dysconnectivity in schizophrenia: where are we now? Neuroscience and Biobehavioral Reviews 35, 1110–1124.

Ragland, J., Yoon, J., Minzenberg, M., Carter, C., 2007. Neuroimaging of cognitive disability in schizophrenia: search for a pathophysiological mechanism. International Review of Psychiatry 19, 417–427.

Rotarska-Jagiela, A., van de Ven, V., Oertel-Knöchel, V., Uhlhaas, P.J., Vogeley, K., Linden, D.E., 2010. Resting-state functional network correlates of psychotic symptoms in schizophrenia. Schizophrenia Research 117, 21–30.

Rubinov, M., Sporns, O., 2010. Complex network measures of brain connectivity: uses and interpretations. NeuroImage 52, 1059–1069.

Schmidt, M.C., Samatova, N.F., 2009. An algorithm for the discovery of phenotype related metabolic pathways. Bioinformatics and Biomedicine, 2009 BIBM'09 IEEE International Conference on. IEEE, pp. 60–65.

Schmidt, M.C., Rocha, A.M., Padmanabhan, K., Shpanskaya, Y., Banfield, J., Scott, K., Mihelcic, J.R., Samatova, N.F., 2012. NIBBS-search for fast and accurate prediction of phenotype-biased metabolic systems. PLoS Computational Biology 8, e1002490.

Sporns, O., 2011. The human connectome: a complex network. Annals of the New York Academy of Sciences 1224, 109–125.

Stam, C., Jones, B., Nolte, G., Breakspear, M., Scheltens, P., 2007. Small-world networks and functional connectivity in Alzheimer's disease. Cerebral Cortex 17, 92–99.

Stam, C., De Haan, W., Daffertshofer, A., Jones, B., Manshanden, I., van Walsum, AvC, Montez, T., Verbunt, J., De Munck, J., Van Dijk, B., 2009. Graph theoretical analysis of magnetoencephalographic functional connectivity in Alzheimer's disease. Brain 132, 213–224.

Stephan, K.E., Friston, K.J., Frith, C.D., 2009. Dysconnection in schizophrenia: from abnormal synaptic plasticity to failures of self-monitoring. Schizophrenia Bulletin 35, 509–527.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America 102, 15545–15550.

Sun, D., van Erp, T.G., Thompson, P.M., Bearden, C.E., Daley, M., Kushan, L., Hardt, M.E., Nuechterlein, K.H., Toga, A.W., Cannon, T.D., 2009. Elucidating a magnetic resonance imaging-based neuroanatomic biomarker for psychosis: classification analysis using probabilistic brain atlas and machine learning algorithms. Biological Psychiatry 66, 1055.

Supekar, K., Menon, V., Rubin, D., Musen, M., Greicius, M.D., 2008. Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. PLoS Computational Biology 4, e1000100.

Tu, P.-C., Lee, Y.-C., Chen, Y.-S., Li, C.-T., Su, T.-P., 2013. Schizophrenia and the brain's control network: aberrant within-and between-network connectivity of the frontoparietal network in schizophrenia. Schizophrenia Research 147, 339–347.

Vandin, F., Upfal, E., Raphael, B.J., 2011. Algorithms for detecting significantly mutated pathways in cancer. Journal of Computational Biology 18, 507–522.

Wang, K., Zhang, H., Kugathasan, S., Annese, V., Bradfield, J.P., Russell, R.K., Sleiman, P.M., Imielinski, M., Glessner, J., Hou, C., 2009. Diverse genome-wide association studies associate the IL12/IL23 pathway with Crohn Disease. American Journal of Human Genetics 84, 399.

Wang, K., Li, M., Hakonarson, H., 2010. Analysing biological pathways in genome-wide association studies. Nature Reviews Genetics 11, 843–854.

Wang, X., Gulbahce, N., Yu, H., 2011. Network-based methods for human disease gene prediction. Briefings in Functional Genomics 10, 280–293.

Wang, J., Zuo, X., Dai, Z., Xia, M., Zhao, Z., Zhao, X., Jia, J., Han, Y., He, Y., 2012. Disrupted functional brain connectome in individuals at risk for Alzheimer's disease. Biological Psychiatry 73, 472–481.

Westman, E., Aguilar, C., Muehlboeck, J.-S., Simmons, A., 2013. Regional magnetic resonance imaging measures for multivariate analysis in Alzheimer's disease and mild cognitive impairment. Brain Topography 26, 9–23.

White, T.P., Joseph, V., Francis, S.T., Liddle, P.F., 2010. Aberrant salience network (bilateral insula and anterior cingulate cortex) connectivity during information processing in schizophrenia. Schizophrenia Research 123, 105–115.

Wolz, R., Julkunen, V., Koikkalainen, J., Niskanen, E., Zhang, D.P., Rueckert, D., Soininen, H., Lötjönen, J., 2011. Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease. PLoS One 6, e25446.

Woodward, N.D., Rogers, B., Heckers, S., 2011. Functional resting-state networks are differentially affected in schizophrenia. Schizophrenia Research 130, 86–93.

Wu, T.T., Lange, K., 2008. Coordinate descent algorithms for lasso penalized regression. The Annals of Applied Statistics 224–244.

Wu, K., Taki, Y., Sato, K., Kinomura, S., Goto, R., Okada, K., Kawashima, R., He, Y., Evans, A.C., Fukuda, H., 2012. Age-related changes in topological organization of structural brain networks in healthy individuals. Human Brain Mapping 33, 552–568.

Xiong, H., Tan, P.-N., Kumar, V., 2006. Hyperclique pattern discovery. Data Mining and Knowledge Discovery 13, 219–242.

Yu, Q., Plis, S.M., Erhardt, E.B., Allen, E.A., Sui, J., Kiehl, K.A., Pearlson, G., Calhoun, V.D., 2011a. Modular organization of functional network connectivity in healthy controls and patients with schizophrenia during the resting state. Frontiers in Systems Neuroscience 5.

Yu, Q., Sui, J., Rachakonda, S., He, H., Gruner, W., Pearlson, G., Kiehl, K.A., Calhoun, V.D., 2011b. Altered topological properties of functional network connectivity in schizophrenia during resting state: a small-world brain network study. PLoS One 6, e25423.

Zaki, M.J., 2000. Generating non-redundant association rules. Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 34–43.

Zaki, M.J., Hsiao, C.-J., 1999. Charm: An Efficient Algorithm for Closed Association Rule Mining. Citeseer.

Zalesky, A., Fornito, A., Bullmore, E.T., 2010. Network-based statistic: identifying differences in brain networks. NeuroImage 53, 1197–1207.

Zhang, K., Cui, S., Chang, S., Zhang, L., Wang, J., 2010. i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. Nucleic Acids Research 38, W90–W95.

Zhang, J., Wang, J., Wu, Q., Kuang, W., Huang, X., He, Y., Gong, Q., 2011. Disrupted brain connectivity networks in drug-naive, first-episode major depressive disorder. Biological Psychiatry 70, 334–342.