On using structural network patterns for prediction of genetic risks in Schizophrenia

Madhura A Ingalhalikar[1], Luke Bloy[1], Drew Parker[1], Raquel Gur[2], Ruben Gur[2], and Ragini Verma[1]

[1]*Section of Biomedical Image Analysis, University of Pennsylvania, Philadelphia, Pennsylvania, United States,* [2]*Brain Behavior Laboratory, University of Pennsylvania, Philadelphia, Pennsylvania, United States*

Background and Objective: Differences in brain structure patterns in asymptomatic family members (FM) of schizophrenia patients (SCZ) have demonstrated an effect of the genetic predisposition to the disease [1]. Earlier studies have mainly concentrated on analyzing structural (T1) [2] images and the scalars from diffusion tensor images (DTI) [3]. Since SCZ is characterized by abnormalities in the communication between spatially disparate brain structures, it becomes vital to analyze the connectivity patterns in the asymptomatic biological relatives who are at risk of the disease. This work focuses on effectively quantifying the extent of pathological abnormality associated with each FM. We utilize connectivity matrices computed from the DTI data, and harness the potential of network properties (global and local) in combination with high dimensional pattern classifiers.

Method: Our data consisted of 70 males (31 SCZ, 29 controls (CNT) and 10 FM) and 82 females (26 SCZ, 33 CNT and 23 FM). Images were acquired on Siemens 3T Trio™ scanner using a 12 channel head coil. DTI was performed using a single shot spin-echo, echo-planar sequence with the following parameters: TR/TE=6400/97 ms, b-value of 1000 s/mm^2 and 64 gradient directions. Cortical parcellation and sub-cortical segmentation was obtained using Freesurfer [4] on structural T1 images and a total of 78 ROI's were extracted to represent the nodes of the structural network that include 68 cortical regions and 10 subcortical structures (Fig 1 (left)). These labels were then transferred to the diffusion space via intrasubject affine transformation. Probabilistic tractography [5] was performed on all the subjects and the connectivity matrices (Fig 1(left)) as well as the network measures (global and local) were computed. The global features included density, shortest path, global efficiency, modularity, assortativity and transitivity of the network, while the local features involved degree, centrality, vulnerability, local efficiency and strengths of each node [6]. The resulting feature vector (with 396 features) was then utilized in a classification framework.

In the next step, feature selection was performed using s2n feature ranking. This method ranks features with the ratio of the absolute difference of the class means over the average class standard deviation and is known to work efficiently for heterogeneous datasets [7]. The selected features were then used to train a non-linear support vector machine (SVM) classifier using a Gaussian kernel. The classifier model was validated using the 15-fold method. In the 15-fold cross validation (CV), one fifteenth samples were chosen for testing, while other samples were used for feature selection and classifier modeling. This type of validation was run multiple times and the resulting accuracy and the probabilistic scores were averaged over the runs. The output probabilistic score referred to as the distributed network connectivity score (DNCS) for each subject represented the level of abnormality of that subject. Finally, after validation, a classifier model over all the controls and patients was applied to the FM and a DNCS, was computed for each FM. The DNCS has a range of -1 to +1 describing the extent to which the brain patterns of the FM match the patterns of patients or controls (where 0 to -1 is the patient spectrum and 0 to +1 is the control spectrum).

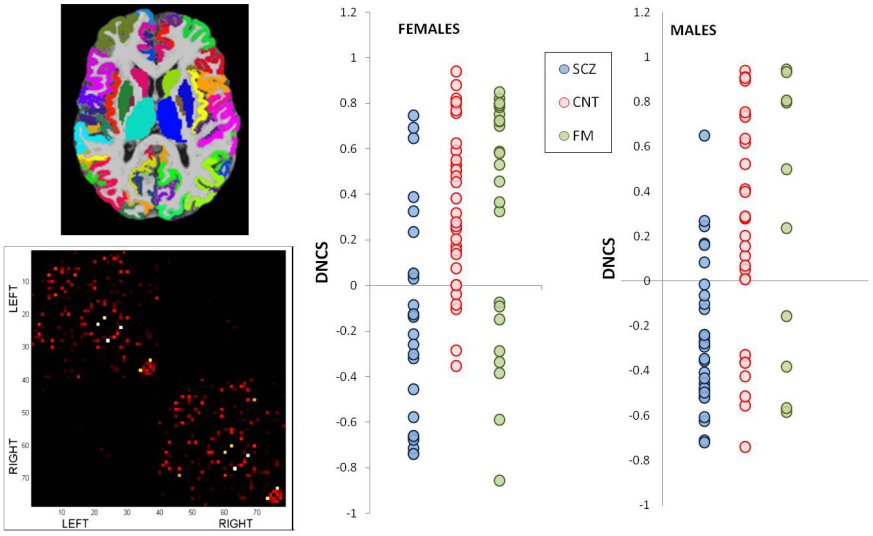

Figure 1: (left) Cortical parcellations (68) and the sub-cortical segmentations (10) employed as ROI's in computing the network shown below. (right) Plot of DNCS from the classifier.

Results: Fig.1 (right) shows DNCS computed for each subject in training as well as testing. For females, the average training accuracy was 72.5 ± 4% (specificity of ~84% and sensitivity of ~ 66%) while in males, the average CV accuracy was 73.9± 2.8% (specificity of ~77% and sensitivity of ~83%). The FM test scores are plotted in green (Fig.1 (right)). Out of these, 4/10 males and 9/23 females were classified as being closer to patients. The table below displays the network properties of the particular ROI's that were selected frequently in the 15 fold CV. In males, the global changes were prominent, as the density of the network and the characteristic path length played a vital role in discriminating the groups. While in females only nodal properties of superior frontal cortex, lingual and paracentral gyrus were selected frequently. In both genders, the degree/strength from superior frontal cortex was important in classifying the groups.

FEATURES	FEMALES	MALES
Global	---	Density, Characteristic path length
Nodal features		
1. Degree	Caudate (L)	Superior frontal cortex (R,L) Rostral middle frontal cortex (R,L) Pallidum (R)
2. Strength	Superior frontal cortex (L) Superior parietal cortex (L) Cuneus (L) Lingual cortex (R) Paracentral (R)	Superior frontal cortex (R,L)
3. Local Efficiency	Paracentral (L)	Lingual (R)
4. Centrality	Pallidum (R)	
5. Vulnerability	Caudate (L) Caudal middle frontal (R) Paracentral (L)	Lingual (R)

Discussion: This study investigated the degree of connectivity changes exhibited by asymptomatic family members of the patients with SCZ. The novel approach combined the power of graphical models and extracted network features with pattern classification to identify connection abnormalities. Choosing network features that describe the macroscopic behavior of connectivity as well the attributes defining regional connectivity properties aided in understanding the global as well as nodal changes. Results suggest that FM demonstrate an endophenotypic trait since in both cases around 40% of FM were classified as having patient-like characteristics based on their DNCS. In females, even though the classifier was highly specific (chance of classifying a test subject as control was high), ~40% of the FMs were classified closer to patients suggesting increased genetic vulnerability to SCZ than males. Furthermore, the ranking of the features demonstrated hierarchical changes in these network properties owing to SCZ. In males, global connectivity changes were more obvious than in females. Finally, the DNCS has the potential to locate the asymptomatic FM over the patient-control spectrum based on the connectivity pattern changes. In conclusion, we provide evidence that high-dimensional pattern classification can identify complex and subtle connectivity based endophenotypes that are shared by SCZ and unaffected FM.

References: [1] Cannon et al..Arch Gen Psychiatry. 55(12) 1998, [2] Fan et al. Biological Psychiatry 63(1) 2008 [3] Hoptman et al. Schizo. Res. 106(203) 2008 [4] Fischl, B et al. Neuroimage 9(2), 1999 [5] Behrens T. et al. MRM 50(5), 2003 [6] Rubinov et al. Neuroimage 52(3), 2010 [7] Golub et al. Science 286(5439),1999.

Acknowledgement: This work was supported by NIH grants MH060722, MH042191 and MH064045.