Topographic diversity of structural connectivity in schizophrenia

Hongtao Ruan[1,+], Qiang Luo[2,3,+,*], PhD, Lena Palaniyappan[4,5,+], MBBS, PhD, Chu-Chung Huang[6,7], PhD, Chun-Yi Zac Lo[3], PhD, Albert C Yang[8,9], MD, PhD, Mu-En Liu[8,9], MD, PhD, Shih-Jen Tsai[8,9] MD, Ching-Po Lin[6,7], PhD, Jianfeng Feng[1,3,10,11,12], PhD

[1] School of Mathematical Sciences, Fudan University, Shanghai, China

[2] School of Life Sciences, Fudan University, Shanghai, China

[3] Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China

[4] Departments of Psychiatry &Medical Biophysics, University of Western Ontario, London, Canada

[5] Robarts Research Institute & Lawson Health Research Institute, London, Canada.

[6] Institute of Neuroscience, National Yang-Ming University, Taipei, Taiwan

[7] Brain Research Center, National Yang-Ming University, Taipei, Taiwan

[8] Department of Psychiatry, Taipei Veterans General Hospital, Taipei, Taiwan

[9] Division of Psychiatry, School of Medicine, National Yang-Ming University, Taipei, Taiwan

[10] Shanghai Center for Mathematical Sciences, Shanghai, China

[11] Department of Computer Science, University of Warwick, Coventry, UK

[12] Collaborative Innovation Center for Brain Science, Fudan University, Shanghai, China

+ these authors contributed equally to this work

*** Correspondence should be addressed to the following addresses:**

Associated Principal Investigator Qiang Luo, Ph.D.

School of Life Science, Fudan University, Shanghai 200433, PR China

Tel. +86-21-65648454

Email: qluo@fudan.edu.cn

ABSTRACT

The neurobiological heterogeneity of schizophrenia is widely accepted, but it is unclear how mechanistic differences converge to produce the observed phenotype. Establishing a pathophysiological model that accounts for both heterogeneity and phenotypic similarity is essential to inform stratified treatment approaches. In this cross-sectional diffusion tensor imaging (DTI) study, we recruited 77 healthy controls (HC), and 71 patients with DSM-IV diagnosis of schizophrenia (SCZ), and reconstructed the structural connectivity of 90 brain regions covering entire cerebral cortex. We first confirmed the heterogeneity in structural connectivity by showing a reduced inter-individual similarity in SCZ compared with HC. Moreover, we found it was not possible to cluster patients into subgroups with shared patterns of dysconnectivity, indicating a high degree of mechanistic divergence in schizophrenia. Instead of the strength of connectivity between any particular brain regions, we investigated the diversity (or statistically, the variance) of the topographic distribution of the strength was reduced. HC had higher topographic diversity in whole brain structural connectivity compared to the patient group ($P = 2 \times 10^{-6}, T = 4.96, Cohen's\ d = 0.87$). In 62 of the 90 brain regions, the topographic diversity was significantly reduced in patients compared to controls after FDR correction (<0.05). When topographic diversity was used as a discriminant feature for classification between patients and controls, we significantly ($P = 4.29 \times 10^{-24}$) improved the classification accuracy to 79.6% (sensitivity 78.3%, specificity 81.3%). This finding suggests highly individualized pattern of structural dysconnectivity underlying the heterogeneity of schizophrenia converges to a convergent common pathway as reduced topographic diversity for the clinical construct of the disease.

INTRODUCTION

Schizophrenia is often considered as an illness where aberration in brain connectivity is a core pathophysiological mechanism. Since Buchsbaum et al.'s initial diffusion imaging study in patients with schizophrenia [1], several reports implicating abnormal white matter integrity in schizophrenia have been published [2-5]. Nevertheless, the spatial localisations of white matter (WM) abnormalities have been diffuse [3, 4, 6] (including corpus callosum, cingulate bundle, and uncinate fasciculus, fibres from superior frontal gyrus, superior temporal gyrus, thalamus, the internal capsule, cerebellum, hippocampus, the parietal lobe, and the occipital lobe), with tracts serving richly connected hub regions being disproportionately affected [7]. Studies involving very large sample of patients indicate that regional abnormalities are indeed present, but they are of moderate magnitude even in the most affected tracts (Cohen's $d < 0.42$) [6]. This lack of consistency has often been ascribed to neurobiological heterogeneity of schizophrenia where region-specific pathology results in distinct sub-groups of patients [3, 8]. While the inter-individual variations in the severity of symptom dimensions and cognitive ability appears to be linked to differential WM integrity [9-12], it is important to note that most studies to date have linked distributed WM deficits to the broad clinical phenotype of schizophrenia, rather than hitherto unknown neurobiological subgroups. If aberrant brain connectivity is indeed a neurobiological substrate of the schizophrenia phenotype, then it is likely that these spatially distinct WM changes converge to deviate certain emergent features of WM architecture. In other words, we can expect to see a large effect-size deficit in WM architecture despite a low degree of inter-individual similarity in spatial distribution of WM abnormalities in schizophrenia.

With respect to the topography of WM tracts, direct structural connectivity among brain regions appears to be generally sparse, with connection density being as low as 3.0% when measured using highly resolved diffusion imaging [13]. Thus, in a healthy brain, a given region connects well to some regions, that might be functionally relevant, while connects sparsely to those that are not as relevant. Such

intra-individual spatial variation in the connectivity of one brain region to the rest forms the basis of distributed brain networks with physiological specificity [14]. Interestingly, tracer studies in monkeys indicate that the diverse connectional topographies of individual brain regions show remarkable inter-individual similarity [15, 16]. These observations suggest that 'topographic diversity in structural connectivity' is likely to be an important feature underpinning domain-specific neural functions. In this paper, we use the term topographic diversity to refer to within-individual variability across space, and contrast this from within-individual variability across time (e.g. training effects [17]).

In this study, we first test if the inter-individual similarity in regional WM connectivity is reduced in a group of 71 patients with schizophrenia and 77 matched healthy controls. Second, using WM pathways that are relevant to variations in symptom severity among patients, we test whether the heterogeneity of WM connectivity in patients can be understood by a number of subgroups. Third, we test whether topographic diversity in patients deviates sufficiently (i.e. effect size of 1 or greater) contributing to statistical discrimination of schizophrenia phenotype from healthy controls. We also tested whether the disrupted WM architecture emerged from aberrations in strong, rather than weakly connected WM pathways. Our overall aim was to establish a quantifiable convergent mechanism that is proximal to the syndromic expression of schizophrenia with an effect size that is comparable to that reported for observable cognitive deficits in this illness. There is an active pursuit to alter WM structure (especially myelin content) as a means to impact cognitive deficits in schizophrenia. To this end, identifying a potential physiological end-point that is not only proximal to disease expression but also relevant to substantial number, if not all, patients with schizophrenia is a critical step to make further progress.

METHODS

Subjects and Image Acquisition

In the current sample (table 1, table s1), we had 71 patients with schizophrenia (36 Male, 35 Female, age 42.2 ± 9.3) and 77 age matched healthy controls (33 Male, 44 Female, age 42.1 ± 9.1). All participants were recruited from the Taipei Veterans General Hospital, and this study was approved by the Institutional Review Board of Taipei Veterans General Hospital. Written informed consent was obtained from all participants. The age of all participants varied from 20 to 55, and the education years varied from 9 to 17. They are all right-handed. The total score of the Positive and Negative Syndrome Scale (PANSS) was over 35 in all patients at the time of scanning. The PANSS scores were low because of the long-term treatment (table s1). All images were acquired using a 3T MR system (Siemens Magnetom Tim Trio, Erlangen, German) at National Yang-Ming University, equipped with a high-resolution 12-channel head array coil. To minimize the head motion during the scan, each subject's head was immobilized with cushions inside the coil during the scan. A high-resolution anatomical T1-weighted image was acquired with sagittal 3D magnetization-prepared rapid gradient echo (MPRAGE) sequence: repetition time (TR) = 3500 ms, echo time (TE) = 3.5 ms, inversion time = 1100 ms, flip angle = 7°, field of view (FOV) = 256 mm $\times$ 256 mm, 192 slices, slice thickness = 1 mm, voxel size =1.0 mm $\times$ 1.0 mm $\times$ 1.0 mm. The diffusion images gradient encoding schemes include 30 non-collinear directions (according to the minimal energy arrangement of electron distribution) (http://www2.research.att.com/~njas/electrons/dim3) with b-value 1000 s/mm^2 and 3 non-diffusion weighted image volume. With the consideration of total brain coverage, each volume consisted of 70 contiguous axial slices (thickness: 2 mm) acquired using a single shot spin-echo planar imaging (EPI) sequence (TR: 11000 ms, TE: 104 ms, NEX: 6, Matrix size: 128 $\times$ 128, voxel size: 2 mm $\times$ 2 mm $\times$ 2 mm, matrix size: 128 $\times$ 128).

Image Preprocessing and Construction of Connectivity Matrix

Brain was extracted from all images using FSL's brain extraction tool [18, 19]. The diffusion-weighted volumes were corrected for eddy-current induced distortions and movements by affine registration to the no diffusion-weighted (B0) image using the diffusion-specific FSL's EDDY tool [20].

Fiber-tracking was performed by the MRtrix0.2.12 package (http://www.nitrc.org/projects/mrtrix/, [21]). We first extracted the b=0 image and constructed the brain mask. By the MRtrix, we acquired the diffusion tensor parameters and fractional anisotropy (FA) map. Next, the response function of each voxel above FA threshold of 0.7 was estimated from the DWI, and this was used as the input of the constrained spherical deconvolution computation [22]. Finally, we performed deterministic fiber-tracking within whole brain using the orientations provided by constrained spherical deconvolution, with an FA cutoff of 0.15, step size of 0.2 mm, minimum length of 10 mm, maximum length of 200 mm, until 2,000,000 fibers were reconstructed for each individual. The AAL template [23] (without parcellations in cerebellum) was registered from MNI standard space into the space of each subject's diffusion data by the program Statistical Parametric Mapping (SPM8, http://www.fil.ion.ucl.ac.uk/spm/, [24]). For each subject, each fiber was labelled according to its termination to the AAL template in native space. From these data, we constructed a 90×90 structural connectivity matrix A, whose element A_{ij} was computed as the mean of the FA values of all included fibers that formed the connection between regions i and j and incorporated the information on the integrity of the interregional white matter connection. The diagonal elements A_{ii} of A was 0 because we only considered the inter-regional connections here.

If a common pathological process resulted in the reduced variance (diversity), a consistent spatial distribution of changes in FA can be expected. We estimated the between-subject similarity by calculating the correlations of the whole brain connectivity (i.e., the FA-weighted structural connectivity matrix A) between subjects of the same group. This resulted in a similarity matrix for each group. This similarity distribution was then compared between healthy subjects and patients with

schizophrenia by t-test to examine whether the similarity was significantly different in group level.

Clustering Patients with Schizophrenia

Followed the approach of detecting subgroups in patients [25], we tried to identify these subgroups, if any, by applying two different clustering methods to the connectivity pattern, i.e., the FA weighted structural connectivity between brain regions.

K-means clustering: We applied k-means clustering to the patients and controls, separately. The number of clusters (k value) in our analysis ranged from 1 to 6. The optimal number of clusters was determined by the GAP statistics [26], measuring the quality of the resulting clusters.

CCA-based hierarchical clustering: To avoid the possibility that the clustering result was affected by the inherent limitations of k-means, we also employed another clustering method which had been successfully applied recently to define neurophysiological subtypes of depression [27]. The structural connectivity was first selected according to their (Spearman's rank) correlations with the PANSS scales (Positive scale, or Negative scale, or General Psychopathology scale with $P < 0.005$, uncorrected) in patients. Secondly, by canonical correlation analysis (CCA), we identified two linear combinations of those symptom-associated structural connectivity, which had the most significant symptom-associations ($r > 0.9$). We applied a hierarchical clustering to identify subgroups of patients to these two linear-combined symptom-associated canonical components. As these features are selected for their relevance to symptom severity, we expected to see a few subgroups within the patient sample, but no subgroups when applied to healthy controls.

The Topographic Diversity of Structural Connectivity

We computed the variance of the brain structural connectivity for each subject as follows:

$$V_{brain} = \frac{1}{4005}\sum_{i<j}(A_{ij} - \mu)^2,$$

where μ was the mean of 4005 structural connections. A larger V_{brain} indicated a higher global topographic diversity of structural connectivity. Similarly, for each brain region defined by the AAL template, the variance of its regional connectivity pattern was calculated to represent the regional topographic diversity for every subject as follows:

$$V_{region}^{i} = \frac{1}{89}\sum_{j\neq i}(A_{ij} - \mu_i)^2,$$

where μ_i was the FA-weighted mean of 89 structural connections connected to region i. For group comparison, we implemented a linear regression to compare these variances between patients and controls with age, gender, education used as covariates.

Comparison of Top 10 or Last 10 Connections

For a fixed region i, we ranked its FA weighted connections to the other 89 brain region in a descending order. To explore the origins of the group difference in the regional diversity, we calculated the mean FA of the top 10 connections and that of the bottom 10 connections. For each region, the mean value of top 10 (bottom 10) was linearly regressed on age, gender, education and diagnostic group identity (schizophrenia: 0; healthy control: 1) for all participants. After FDR correction for multiple comparisons, this allowed us to test whether the disrupted topographic diversity emerged from aberrations in strong (top 10), rather than weakly connected (bottom 10) WM pathways. The effects of different number of top (bottom) connections range from 5 to 25 were also calculated and we found the effects were similar.

Classification of Patients and Controls

To study if the topographic diversity provided any incremental value in distinguishing patients from healthy controls, we used a random forest classification approach [28], that can correct for overfitting and provide ranking of the discriminatory

features. Two models were built on the basis of: (Model 1) 4005 structural connectivity among 90 brain regions; (Model 2) variance of each region, mean FA of the top 10 connections in the regional connectivity pattern, combined with structural connectivity.

Firstly, we ran the random forest algorithm with 1000 trees for 100 times on the predictors to obtain an average importance score for each predictor. The highest few predictors were used in the final model to classify the subjects into patients and controls. Secondly, using five-fold cross-validation with 100 repeats, we compared the classification accuracies of these two models by a paired-sample t-test， and also estimated an average accuracy of classification for each model. The results were validated by exploring different numbers of selected predictors for these models.

RESULTS

Inter-individual similarity of structural connectivity decreased in patients

Healthy controls had greater overall similarity in structural connectivity ($P = 2.76 \times 10^{-74}, T = 18.52, Cohen's\ d = 0.51$), indicating notable within-group heterogeneity in structural connectivity patterns among patients (figure 1A).

Topographic diversity of structural connectivity reduced in patients

Compared with patients, HC had greater diversity (variance) in the whole brain structural connectivity compared ($P = 2 \times 10^{-6}, T = 4.96,\ Cohen's\ d = 0.87$), indicating that the difference between the strong connections and the weak connections were larger in healthy controls (figure 1B).

After FDR correction (<0.05), we found that the diversity of regional connectivity pattern in 62 of 90 brain regions were reduced in patients compared with controls (figure 1C). Areas showing reduced diversity were mainly located in frontal lobe, prefrontal lobe, parietal lobe, occipital lobe and subcortical areas. The top 5 most abnormal regions were SMA.L ($P = 2.14 \times 10^{-8}, T = 5.93, Cohen's\ d = 1.04$), SMA.R ($P = 5.9 \times 10^{-6}, T = 4.71,\ Cohen's\ d = 0.83$), THA.L ($P = 1.52 \times 10^{-5}, T = 4.48, Cohen's\ d = 0.79$), ORBinf.L ($P = 1.8 \times 10^{-5}, T = 4.44, Cohen's\ d = 0.78$) and PoCG.L ($P = 2.38 \times 10^{-5}, T = 4.37, Cohen's\ d = 0.77$).

Furthermore, we found for 61 out of 62 brain regions identified above, the connectivity strength reduced in the top 10 strongest connections ($P < 0.05$ FDR corrected) (figure1D, table s2), but no significant changes in the lowest 10 weak connections (figure 1D, table s2).

Clusters based on WM connectivity patterns

According to the quality of clusters given by the GAP statistic [26], we found that the optimal number of clusters given by k-means clustering was 1 in both patients and

controls (figure 2A). Similarly, the CCA-based hierarchical clustering algorithm gave no evidence for any subgroup in patients (figure 2B, tables s3 and s4).

Topographic diversity discriminates patients from controls

Six connections, predominantly frontal (table 2), were selected by the random forest and the final model with these variables yielded a mean accuracy rate of 76.1% (sensitivity 73.3%, specificity 78.6%). Introducing the topographic diversity to the model, i.e. the variance and mean FA of top 10 connections from each region, the mean accuracy rate increased to 79.6% (sensitivity 78.3%, specificity 81.3%), which was significantly better than the model based on structural connectivity alone ($P = 4.29 \times 10^{-24}$ by a paired-sample t-test [29], figure 3A), and such improvement was not affected by the number of predictors selected (figure 3B). No correlation (table s5) has been identified between either the regionally or the globally diversity and the medical treatment (the CPZ dosage in table s1).

DISCUSSION

To our knowledge, this is the first study to systematically examine if the individually distinct WM changes seen in patients with schizophrenia lead to a convergent effect that is relevant to the expressed clinical phenotype. We report 4 major findings from the current study. 1) The inter-individual spatial similarity in WM connectivity is greatly reduced in schizophrenia compared to healthy controls. 2) Despite the spatial dissimilarity, patients with schizophrenia, as a group, show large effect-sized reductions in topographic diversity, which when used in conjunction with the observed pattern of regional changes in each patient, results in successful discrimination of the patients from controls. 3) The reduced topographic diversity relates almost exclusively to disruptions in strongly connected WM pathways. 4) We could not identify subgroups of patients with clustering of WM connectivity patterns.

Our findings add an important clarification to the divergent results of DTI studies in schizophrenia. In patients with schizophrenia, subtle disruptions in WM

connectivity occur in a person-specific (idiosyncratic) fashion. The patterns of resulting deficits, when combined, are sufficiently discriminative of the phenotype, but this is not due to a generalised non-specific deficit affecting all WM pathways. Our results suggest that the WM deficit in schizophrenia is likely to be specific to the strongly connected pathways that are more liable to have direct axonal connections. This inference reconciles the spatial divergence [30] with the discriminatory accuracy[31-34] (generally 70-90%) when using multivariate approaches in the DTI studies.

We lacked neuropsychological data to examine cognitive specificity of regional WM aberrations. Interestingly, there was no discernible association between topographic diversity and symptom burden, but we observed a notable canonical relationship between symptom severity and regional WM aberrations. This supports the notion that the highly variable, small magnitude, lower-level mechanistic markers may operate to produce individual symptoms, while the system-level indicators are more relevant to the emergence if the broader clinical phenotype of schizophrenia. In this sense, topographic diversity may represent the elusive neural integrative deficit [35] that is proximal to clinical phenotype but relatively distal to etiological factors [36].

There is an enthusiastic acceptance that schizophrenia is a neurobiologically heterogeneous illness; but the evidence demonstrating such heterogeneity is not yet available, or at best, weak [37]. In most studies, the absence of clear demarcation of a neurobiological feature between patients and controls (i.e. small effect size) or the demonstration of increased variance [38] among patients is taken to be supportive of the underlying heterogeneity. While this is certainly a viable explanation, we report that convergent neurobiological abnormalities at a systemic level can stem from diverse changes. Thus, the neurobiological heterogeneity inferred at a lower level, may not be observable at the ‘systems’ level. It is important to note that the lack of clustering solutions does not negate the neurobiological heterogeneity of schizophrenia. On the contrary, these results are expected if there is a very high level of heterogeneity. But the large effect-size reduction in systems-level property (diversity), reminds us that it may be premature to dismiss the possibility that schizophrenia, as observed by

contemporary clinicians, could be a system-level deficit in WM architecture that may vary only in degree across individuals (figure s1 summarizes the possible mechanistic nature of schizophrenia).

In summary, the convergence of highly individualized (idiosyncratic) pattern of structural dysconnectivity resulting in reduced topographic diversity in schizophrenia validates the century-old concept of dysconnectivity or integration deficit as a binding feature that brings together seemingly disparate dimensions of symptoms in patients with schizophrenia. When parsing the heterogeneity of psychosis, it is critical to accommodate emergent system-level properties that can capture the gestalt of the currently used clinical construct of schizophrenia.

Limitation

It is important to consider 2 important caveats in this study. We recruited a sample of medicated patients with an established history of schizophrenia. While this reduced the heterogeneity that is typically seen in early stage, drug-naïve samples with relatively weaker diagnostic stability, a selection-bias towards a single group of patients with relatively more severe illness and poor prognosis cannot be excluded. It is rather surprising that this bias has not improved the inter-individual similarity in WM aberrations. Secondly, compared to recent consortium-based neuroimaging studies [6], our sample size was relatively modest. This might explain the lack of clustering solutions within the patient group. Nevertheless, it is acknowledged that the heterogeneity in psychiatric disorders are expected to be diffuse [39]; as a result, an 'optimal' clustering solution may not be achieved even in large samples [40].

Acknowledgement

This work was supported by grants (V105C-008, V105E17-002-MY2-1) from the Taipei Veterans General Hospital. This research was partially supported by the grants from the National Natural Science Foundation of China (No. 11101429, 11471081) and Natural Science Foundation of Shanghai (No. 17ZR1444400). C.P.L was

supported in part by funding from Ministry of Science and Technology (MOST 104-2633-B-400-001, MOST 104-2218-E-010-007- MY3, MOST 104-2221-E-010-013), National Health Research Institutes (NHRI-EX- 10310EI) and Academia Sinica (AS-104-TP- B10) of Taiwan. A.C.Y was supported by Taiwan (MOST 101-2314-B-075-041-MY3, 104-2314-B-075-078- MY2). S.J.T. was supported by Taiwan (MOST 103-2314-B-075-067-MY3, MOST 104-2745- B-075-002). JF is a Royal Society Wolfson Research Merit Award holder. JF is also supported by the National High Technology Research and Development Program of China (No. 2015AA020507) , the Key Program of National Natural Science Foundation of China (No. 91630314), International (Regional) Collaborative and Exchange Program of National Natural Science (No. 71661167002) the key project of Shanghai Science & Technology Innovation Plan (No. 15JC1400101), and the Shanghai Soft Science Research Program (No. 15692106604), and the National Centre for Mathematics and Interdisciplinary Sciences (NCMIS) of the Chinese Academy of Sciences. LP is supported by the Academic Medical Organization of Southwest Ontario; Bucke Family Fund; Chrysalis Foundation and Canadian Institute of Health Research Foundation Grant.

Funding and Disclosure

LP received speaker fees from Otsuka in 2017; LP received unrestricted educational grants from Otsuka Canada and Janssen Canada in 2017. All other authors declare no conflict of interest.

Figure legends

Figure 1. Topographic diversity of brain structural connectivity reduced in patients. A) Inter-individual similarity matrices of two groups, and group comparison of the between-subject similarity; B) Comparison of the within-subject variance in the whole brain structural connectivity; C) Spatial distribution of the 62 brain regions with significantly reduced variances in their regional connectivity patterns in patients compared with controls; D) Comparison of the mean FA of both the top 10 and bottom 10 connections in the regional connectivity pattern of these 62 brain regions between patients and controls.

Figure 2. Clustering by brain structural connectivity identified no subgroups in patients. A) The quality of the resulting clusters measured by the GAP statistics for k-means clustering with k = 1, 2, 3, 4, 5 and 6; B) Hierarchical clustering for schizophrenia and healthy control separately using two symptom-associated canonical components of the structural connectivity.

Figure 3. **Topographic diversity as a system-level feature improves the classification accuracy between patients and controls.** A) Two models with different sets of features and the accuracies; B) The accuracies of two different models with the number of selected features ranging from 5 to 15. The vertical lines on black nodes imply the standard error.

REFERENCES

1. Buchsbaum MS, Tang CY, Peled S, et al. MRI white matter diffusion anisotropy and PET metabolic rate in schizophrenia. *Neuroreport* Feb 16 1998;9(3):425-430.

2. Karlsgodt KH. Diffusion imaging of white matter in schizophrenia: Progress and future directions. *Biological psychiatry: cognitive neuroscience and neuroimaging* 2016;1(3):209-217.

3. Wheeler AL, Voineskos AN. A review of structural neuroimaging in schizophrenia: from connectivity to connectomics. *Frontiers in human neuroscience* 2014;8.

4. Kubicki M, McCarley R, Westin C-F, Park H-J, Maier S, Kikinis R, Jolesz FA, Shenton ME. A review of diffusion tensor imaging studies in schizophrenia. *Journal of psychiatric research* 2007;41(1):15-30.

5. Melonakos ED, Shenton ME, Rathi Y, Terry DP, Bouix S, Kubicki M. Voxel-based morphometry (VBM) studies in schizophrenia—can white matter changes be reliably detected with VBM? *Psychiatry Research: Neuroimaging* 2011;193(2):65-70.

6. Kelly S, Jahanshad N, Zalesky A, et al. Widespread white matter microstructural differences in schizophrenia across 4322 individuals: results from the ENIGMA Schizophrenia DTI Working Group. *Molecular Psychiatry* 2017.

7. Klauser P, Baker ST, Cropley VL, et al. White matter disruptions in schizophrenia are spatially widespread and topologically converge on brain network hubs. *Schizophrenia bulletin* 2017;43(2):425-435.

8. Sun H, Lui S, Yao L, et al. Two patterns of white matter abnormalities in medication-naive patients with first-episode schizophrenia revealed by diffusion tensor imaging and cluster analysis. *JAMA psychiatry* 2015;72(7):678-686.

9. Voineskos AN, Felsky D, Kovacevic N, et al. Oligodendrocyte genes, white matter tract integrity, and cognition in schizophrenia. *Cerebral cortex* 2012;23(9):2044-2057.

10. Palaniyappan L, Al-Radaideh A, Mougin O, Gowland P, Liddle PF. Combined white matter imaging suggests myelination defects in visual processing regions in schizophrenia. *Neuropsychopharmacology* 2013;38(9):1808-1815.

11. Karbasforoushan H, Duffy B, Blackford JU, Woodward ND. Processing speed impairment in schizophrenia is mediated by white matter integrity. *Psychological medicine* 2015;45(1):109-120.

12. Rowland LM, Spieker EA, Francis A, Barker PB, Carpenter WT, Buchanan RW. White matter alterations in deficit schizophrenia. *Neuropsychopharmacology* 2009;34(6):1514-1522.

13. Hagmann P, Cammoun L, Gigandet X, Meuli R, Honey CJ, Wedeen VJ, Sporns O. Mapping the structural core of human cerebral cortex. *PLoS biology* 2008;6(7):e159.

14. Greicius MD, Supekar K, Menon V, Dougherty RF. Resting-State Functional Connectivity Reflects Structural Connectivity in the Default Mode Network. *Cerebral Cortex* 2009;19(1):72-78.

15. Saleem KS, Miller B, Price JL. Subdivisions and connectional networks of the lateral prefrontal cortex in the macaque monkey. *Journal of Comparative Neurology* 2014;522(7):1641-1690.

16. Webster MJ, Bachevalier J, Ungerleider LG. Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. *Cerebral cortex* 1994;4(5):470-483.

17. Scholz J, Klein MC, Behrens TEJ, Johansen-Berg H. Training induces changes in white-matter

architecture. *Nat Neurosci* 11//print 2009;12(11):1370-1371.

18. Smith SM. Fast robust automated brain extraction. *Human brain mapping* 2002;17(3):143-155.

19. Jenkinson M, Pechaud M, Smith S. BET2: MR-based estimation of brain, skull and scalp surfaces. Paper presented at: Eleventh annual meeting of the organization for human brain mapping, 2005.

20. Andersson JL, Sotiropoulos SN. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage* 2016;125:1063-1078.

21. Tournier JD, Calamante F, Connelly A. MRtrix: Diffusion tractography in crossing fiber regions. *International Journal of Imaging Systems and Technology* 2012;22(1):53-66.

22. Tournier JD, Calamante F, Connelly A. Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. *Neuroimage* May 01 2007;35(4):1459-1472.

23. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 2002;15(1):273-289.

24. Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE. *Statistical parametric mapping: the analysis of functional brain images*: Academic press; 2011.

25. Hahamy A, Behrmann M, Malach R. The idiosyncratic brain: distortion of spontaneous connectivity patterns in autism spectrum disorder. *Nature neuroscience* 2015;18(2):302-309.

26. Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2001;63(2):411-423.

27. Drysdale AT, Grosenick L, Downar J, et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine* 2016.

28. Friedman J, Hastie T, Tibshirani R. *The elements of statistical learning.* Vol 1: Springer series in statistics New York; 2001.

29. Pereira F, Mitchell T, Botvinick M. Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage* 2009;45(1):S199-S209.

30. Melonakos E, Shenton M, Rathi Y, Terry D, Bouix S, Kubicki M. Voxel-based morphometry (VBM) studies in schizophrenia—can white matter changes be reliably detected with VBM? *Psychiatry Research* 2011;193(2):65-70.

31. Karageorgiou E, Schulz SC, Gollub RL, et al. Neuropsychological testing and structural magnetic resonance imaging as diagnostic biomarkers early in the course of schizophrenia and related psychoses. *Neuroinformatics* 2011;9(4):321-333.

32. Sun D, van Erp TG, Thompson PM, et al. Elucidating a magnetic resonance imaging-based neuroanatomic biomarker for psychosis: classification analysis using probabilistic brain atlas and machine learning algorithms. *Biological psychiatry* 2009;66(11):1055-1060.

33. Ingalhalikar M, Kanterakis S, Gur R, Roberts T, Verma R. DTI based diagnostic prediction of a disease via pattern classification. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010* 2010:558-565.

34. Yang H, Liu J, Sui J, Pearlson G, Calhoun VD. A hybrid machine learning method for fusing fMRI and genetic data: combining both improves classification of schizophrenia. *Frontiers in*

human neuroscience 2010;4.

35. Meehl PE. Schizotaxia revisited. *Archives of general psychiatry* 1989;46(10):935-944.

36. Tsuang MT, Lyons MJ, Faraone SV. Heterogeneity of schizophrenia. Conceptual models and analytic strategies. *The British Journal of Psychiatry* 1990;156(1):17-26.

37. Marquand AF, Wolfers T, Mennes M, Buitelaar J, Beckmann CF. Beyond lumping and splitting: a review of computational approaches for stratifying psychiatric disorders. *Biological psychiatry: cognitive neuroscience and neuroimaging* 2016;1(5):433-447.

38. Brugger SP, Howes OD. Heterogeneity and homogeneity of regional brain structure in schizophrenia: a meta-analysis. *JAMA psychiatry* 2017.

39. Schnack HG. Improving individual predictions: Machine learning approaches for detecting and attacking heterogeneity in schizophrenia (and other psychiatric diseases). *Schizophrenia Research* 2017.

40. Schnack HG, Kahn RS. Detecting neuroimaging biomarkers for psychiatric disorders: sample size matters. *Frontiers in psychiatry* 2016;7.

Figure 1.

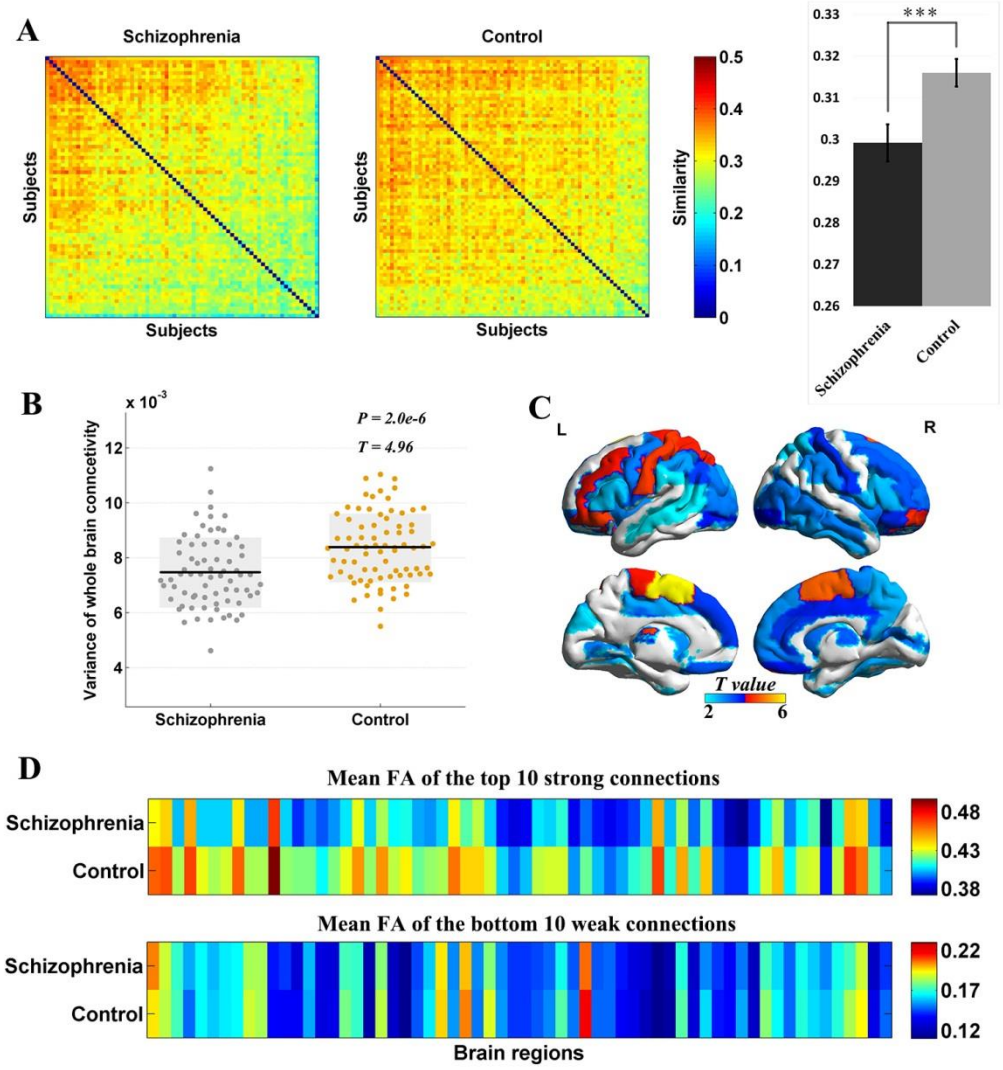
A
Schizophrenia
Control
Subjects
Subjects
Subjects
Subjects
Similarity
0.5
0.4
0.3
0.2
0.1
0
***
0.33
0.32
0.31
0.3
0.29
0.28
0.27
0.26
Schizophrenia
Control
B
x 10^-3
P = 2.0e-6
T = 4.96
Variance of whole brain connectivity
12
10
8
6
4
Schizophrenia
Control
C
L
R
T value
2
6
D
Mean FA of the top 10 strong connections
Schizophrenia
Control
0.48
0.43
0.38
Mean FA of the bottom 10 weak connections
Schizophrenia
Control
0.22
0.17
0.12
Brain regions

Figure 2.

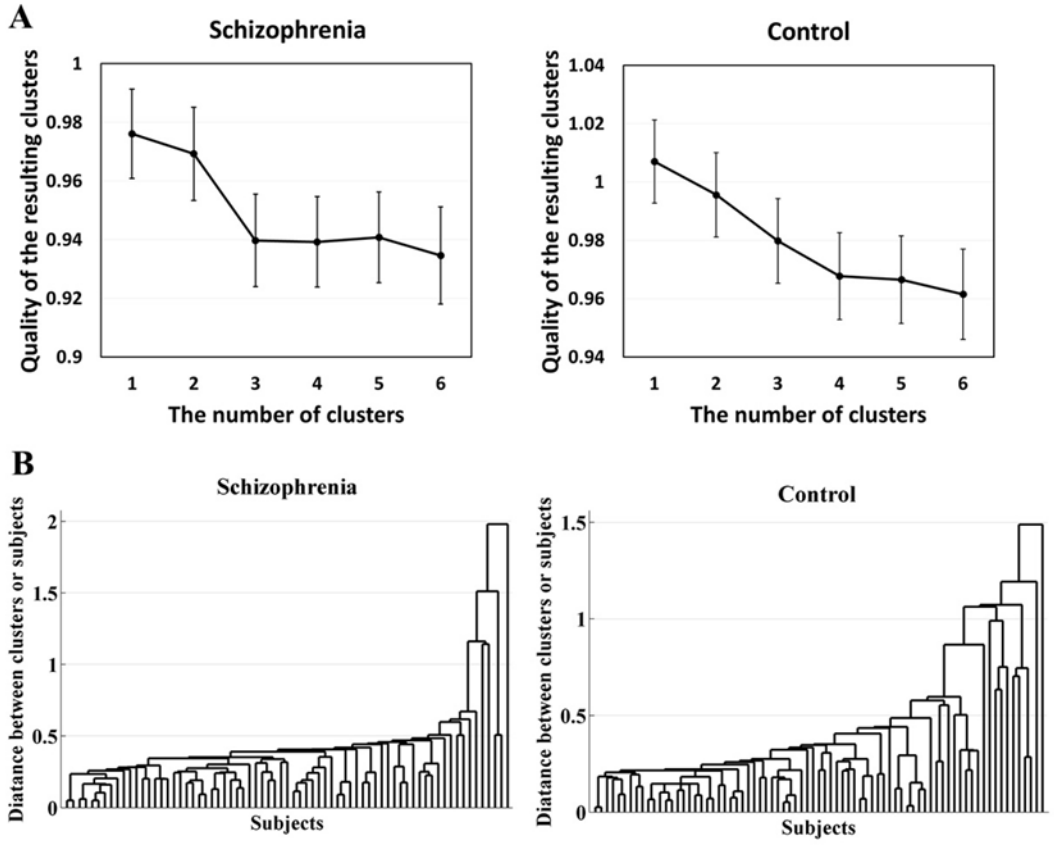
A
Schizophrenia
Control
Quality of the resulting clusters
The number of clusters
B
Schizophrenia
Control
Diatance between clusters or subjects
Subjects

Figure 3.

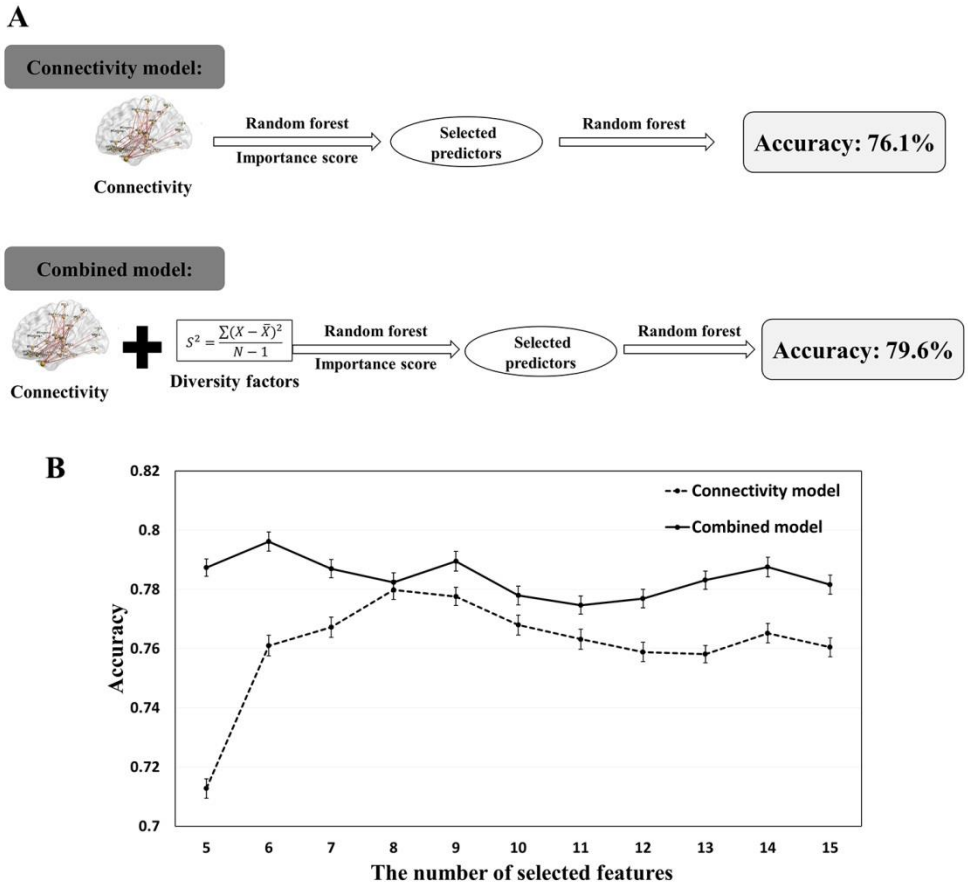
A
Connectivity model:
Connectivity
Random forest
Importance score
Selected predictors
Random forest
Accuracy: 76.1%
Combined model:
Connectivity
$S^2 = \frac{\sum(X-\bar{X})^2}{N-1}$
Diversity factors
Random forest
Importance score
Selected predictors
Random forest
Accuracy: 79.6%
B
Connectivity model
Combined model
Accuracy
0.82
0.8
0.78
0.76
0.74
0.72
0.7
5
6
7
8
9
10
11
12
13
14
15
The number of selected features

Table 1. Demographics and behavioral statistics for schizophrenia and control groups

		Schizophrenia	Control
Gender		35 F, 36 M	44 F, 33 M
Age		42.2 ± 9.3	42.1 ± 9.1
Education		12.8 ± 2.2	14.2 ± 2.1
Duration		15.6 ± 9.3	
PANSS	**Positive scale**	10.1 ± 2.9	
	Negative scale	9.9 ± 2.6	
	General Psychopathology scale	22.1 ± 3.5	

Table 2. Selected predictors in two different random forest models.

Model	**Selected predictors**	**Property of predictor**	**Importance score**
Connectivity model	SFGmed.L-ORBsupmed.L	Structural connectivity	0.3104
	MFG.R- SFGmed.L	Structural connectivity	0.29744
	ORBinf.R-SFGmed.L	Structural connectivity	0.2604
	SMA.L-LING.R	Structural connectivity	0.2487
	ORBsup.L-SOG.L	Structural connectivity	0.2381
	SMA.L-SFGmed.L	Structural connectivity	0.2219
Combined model	SMA.L	Variance	0.2407
	MFG.R- SFGmed.L	Structural connectivity	0.2201
	IOG.R	Variance	0.2101
	SFGmed.L-ORBsupmed.L	Structural connectivity	0.2034
	SMA.L-LING.R	Structural connectivity	0.1690
	ORBinf.R-SFGmed.L	Structural connectivity	0.1590