

# Exploring different impaired speed of genetic-related brain function and structures in schizophrenic progress using multimodal analysis\*

Na Luo<sup>#</sup>, Lin Tian<sup>#</sup>, Vince D. Calhoun, *Fellow IEEE*, Jiayu Chen, Dongdong Lin, Victor M. Vergara, Shuquan Rao, Fuquan Zhang<sup>\*</sup> and Jing Sui<sup>\*</sup>, *senior Member IEEE*

**Abstract**—Schizophrenia (SZ) is a highly heritable disease exhibiting substantial structural and functional brain impairments. The duration of illness and medication use may cause different presentations of impairments in patients. To understand the progressive variations of the disease, most recent studies have reported brain functional or structural abnormalities associated with illness duration, but a comprehensive study of pathology underlying brain structure, function and illness duration is still limited. In this work, we employed a three-way parallel independent component analysis (pICA) algorithm to jointly analyze grey matter volume (GM), functional connectivity (FC) and single nucleotide polymorphisms (SNPs) from drug-naïve first-episode [FESZ], chronic schizophrenia [CSZ] and healthy controls [HC], aiming to identify the linked alterations in SNP-GM-FC components, and evaluate the impairment speed of imaging measures associated with SZ-susceptible genetic variants in different disease stages (FESZ and CSZ). Results demonstrated significant group differences on GM and FC in hippocampus, temporal gyrus and cerebellum between SZ and HC, which are also significantly correlated with SNPs residing in genes like *GABBR2*, *SATB2*, *CACNA1C*, *PDE4B*, involved in pathways of cell junction, synapse and neuron projection. Moreover, two-sample t-tests showed that GM volume and FC strength presented similar trends of progressive decrease with the increase of the illness duration (HC>FESZ>CSZ). Besides that genetic-related GM and FC components both showed significant associations with illness duration, FC indicates the higher impairment speed than GM, suggesting that functional connectivity may serve as a more sensitive measure to detect the disruptions in SZ at the very early stage.

## I. INTRODUCTION

Schizophrenia (SZ) is a highly heritable complex disease manifested with structural and functional brain abnormalities. Advanced magnetic resonance imaging (MRI) technique and high-throughput genome wide association studies (GWAS)

make it feasible to characterize the alterations in SZ compared to healthy controls (HC) from both neuroimaging and genetic views. Abnormalities in gray matter (GM) volume, especially in the prefrontal cortex and temporal gyrus, have been well documented [1, 2]. Functional connectivity (FC), a factor reflecting the interactions of brain activities among different regions is also reported with significant alternations mainly involved in prefrontal, hippocampus and temporal gyrus [3, 4] in SZ. A landmark study [5] from the Psychiatric Genomics Consortium (PGC) has reported a schizophrenia Genome wide association studies (GWAS) study of 36,989 cases and 113,075 controls, highlighting multiple risk genes *i.e.*, *CACNA1C*, *SATB2* and *DRD2*.

The stage of illness and medication use may also substantially influence brain structure and function such as progressive changes between drug-naïve first-episode schizophrenia (FESZ) and chronic schizophrenic patients (CSZ). As reviewed in [6], SZ showed progressive gray matter volume loss mostly in the frontal and temporal, and increase in the lateral ventricle. A recent resting functional MRI study compared the FC between first-episode and chronic stages of schizophrenia, revealing that 90% of the FC changes were involved the inferior frontal gyrus in first-episode patients. While for the chronic patients, FC differences were demonstrated in broader brain areas, including reduced thalamo-frontal connectivity, and increased thalamo-temporal and thalamo-sensorimotor connectivity [7].

Nevertheless, most of current works were only focused on single imaging modality while the interconnections among the alternations in different imaging modalities as well as their relations to illness duration are not clear. In this study, we applied a multivariate model, three-way parallel independent component analysis (pICA) [8, 9], to combine GM volume, FC and SNP together in a joint analysis, aiming to identify the linked alterations in SNP-GM-FC components, and to evaluate the impairment speed of brain imaging measures associated with SZ-susceptible genetic variants in different disease stages (FESZ vs. CSZ). After pICA, we conducted post hoc analysis to examine the group difference in each modality among 3 groups, as well as their associations with illness duration.

## II. MATERIALS AND METHODS

### A. Materials

In this study, 159 subjects were recruited from Wuxi Mental Health Center, including 87 HCs, 20 FESZs and 52 CSZs. Demographic data were provided in Table 1. The FESZs were defined as having illness duration  $\leq 2$  years,

\*Research supported by the Chinese National Natural Science Foundation and Chinese Academy of Sciences (CAS), NIH and the Primary Research & Development Plan of Jiangsu.

<sup>#</sup>These authors contributed equally to this work.

Na Luo and Jing Sui are with Brainnetome Center and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences and University of Chinese Academy of Sciences, Beijing 100190, China (\*corresponding author to provide phone: 86-10-82544518; e-mail: [jing.sui@nlpr.ia.ac.cn](mailto:jing.sui@nlpr.ia.ac.cn), [luona2015@ia.ac.cn](mailto:luona2015@ia.ac.cn)).

Lin Tian and Fuquan Zhang are with Wuxi Mental Health Center, Wuxi 214000, China (\*corresponding author: [zhangfq@njmu.edu.cn](mailto:zhangfq@njmu.edu.cn)).

Vince D. Calhoun, Jing Sui, Jiayu Chen, Dongdong Lin and Victor M. Vergara are with The Mind Research Network. Dr. Calhoun is also with Dept. of ECE, University of New Mexico, Albuquerque, NM, 87106, USA ([vcalhoun@unm.edu](mailto:vcalhoun@unm.edu)).

Shuquan Rao is with School of Life Science and Engineering, Southwest Jiaotong University, Chengdu, 610031, China.

while the CSZs were defined as having illness duration >2 years. MRI scans were obtained on a 3.0-Tesla Magnetom Trio Tim (Siemens Medical System) at the Department of Medical Imaging, Wuxi People's Hospital, Nanjing Medical University. Written informed consent was obtained from all study participants under protocols approved by the Institutional Review Boards. Resting-state functional MRI (fMRI) was collected with a single-shot gradient-echo echo-planar-imaging sequence sensitive to blood oxygenation level-dependent contrast [TR=2000 ms; TE=30 ms; field of view (FOV) =220×220 mm<sup>2</sup>; matrix=64×64; flip angle=90°; voxel size=3.44×3.44×4.60 mm<sup>3</sup>]. For the structural MRI (sMRI) data, a multi-echo MPRAGE sequence was used with the following parameters: TR/TE/TI=2530/3.44/1100ms, flip angle=7°, FOV=256×256 mm, matrix size=256×256×192, Voxel size=1×1×1mm. DNA was extracted from peripheral blood cell according to the standard protocol by protease K digestion, phenol-chloroform extraction and ethanol precipitation. The whole-genome genotyping was performed on Illumina human PsychArray-24 spanning 571,054 loci.

TABLE I. DEMOGRAPHIC CHARACTERISTICS

Demographic	HC	FESZ	CSZ
Number	87	20	52
Gender (M/F)	47/40	10/10	30/22
Age (y) (Mean±SD)	39.91±14.84	33.60±10.80	46.12±11.38
Education (Mean±SD)	12.67±4.28	10.65±4.65	10.75±2.69
Duration of Illness (y) (Mean±SD)	NA	0.96±0.85	20.63±9.84
Medication (Mean±SD)	NA	Drug-naïve	17.26±9.44

Note: The Medication is Chlorpromazine equivalent (standardized current dose of antipsychotic medication).

### B. Data Processing

The collected fMRI data was preprocessed by using Brant software package (<http://brant.brainnetome.org/en/latest/>). The first 10 volumes of each functional time series were discarded for the magnetization equilibrium. Slice timing was performed with the middle slice as the reference frame. Then images were realigned using INRIalign and spatially normalized into Montreal Neurological Institute (MNI) space, resliced to 3×3×3 mm. Denoising was further performed to regress out motion parameters, white matter (WM), and cerebrospinal fluid (CSF) and smoothed with an 8 mm Gaussian model. Leveraging the 116×116 automatic Gaussian labelling (AAL) atlas [10], we constructed FC from Pearson cross-correlations between all pairs of regions based on the mean BOLD signals, resulting in 6786 edges.

The T1-weighted sMRI data were preprocessed using voxel based morphometry (VBM) in SPM8 ([www.fil.ion.ucl.ac.uk/spm/software/spm8](http://www.fil.ion.ucl.ac.uk/spm/software/spm8)), with a unified model of image registration, bias correction, tissue classification and spatial normalization to the standard MNI space. Modulated normalized parameters were used to segment the brain into WM, GM and CSF probabilistic maps. The resulting GM images consisted of voxelwise GM volume were resliced to 3×3×3mm, resulting in 53×63×46 voxels and smoothed with an 8 mm Gaussian model.

The SNP data were preprocessed by a series of standard quality control (QC) procedures using PLINK [11, 12]. Poorly genotyped individuals were removed if having

discordant sex, high missing rate (>3%), unusual heterozygosity (>3 SD from mean) and unusual similarity or relatedness. SNPs were further removed for missing rate (>5%), rare variants with minor allele frequency <0.01, Hard Weinberg equilibrium ( $p < 0.00001$ ) and case-control genotype call rate difference ( $p < 0.00001$ ). After QC, the remaining 289402 SNPs were then mapped with PGC's SCZ2 database [5] ( $p < 0.01$ ), resulting in 4937 SNPs for subsequent analysis.

### C. Three-way parallel ICA

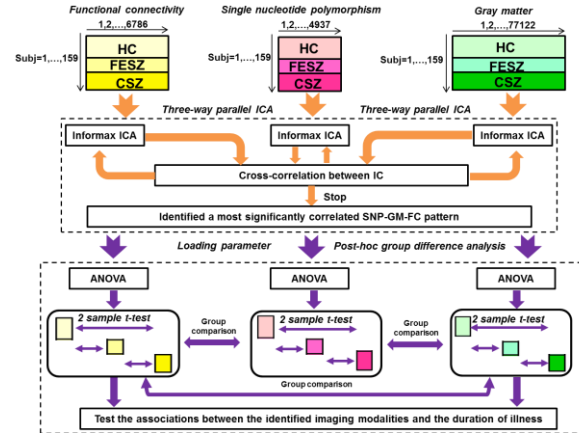


Figure 1. Schematic illustration of the whole analysis.

Before our subsequent analysis, medication use has been regressed for the chronic patients. Then the formed FC, GM and SNP matrices were jointly analyzed using the three-way pICA method [8, 9] as Fig. 1, which extracted independent components (ICs) using independent component analysis (ICA) on the genetic and brain phenotype data and jointly evaluates the association among the three data modalities. The code is available for public use through the Fusion ICA Toolbox (FIT, <http://mialab.mrn.org/software/fit>). The component extraction was based upon the Infomax algorithm[13]. Infomax extracts independent components through maximization of entropy which measures uncertainty associated with a random variable. As illustrated in Eq. (1), the datasets are first decomposed into a linear combination of underlying components.  $X_1, X_2, X_3$  are observation matrices composed of measurements of FC, GM and SNP (three  $N_{\text{subject}} \times N_{\text{voxel}} \times FC/SNP$  matrices prepared above);  $S_1, S_2, S_3$  represent the unknown sources for each phenotype with each row representing an independent component;  $A_1, A_2, A_3$  are linear mixing matrices/loadings with each column representing a loading parameter associated with a specific component in  $S_1, S_2, S_3$ ;  $W_1, W_2, W_3$  are unmixing matrices. Based on the decomposition, the Infomax algorithm attempts to find the  $W$  matrix resulting in independence through maximizing an entropy function as defined in Eq. (2), where  $f_y(Y)$  is the probability density function of  $Y$ ;  $E$  is the expected value;  $H$  is the entropy function;  $W_0$  is the bias vector.

$$\begin{aligned}
 X_1 &= A_1 \cdot S_1 \xrightarrow{W_1 = A_1^{-1}} S_1 = W_1 \cdot X_1 \\
 X_2 &= A_2 \cdot S_2 \xrightarrow{W_2 = A_2^{-1}} S_2 = W_2 \cdot X_2 \\
 X_3 &= A_3 \cdot S_3 \xrightarrow{W_3 = A_3^{-1}} S_3 = W_3 \cdot X_3
 \end{aligned} \tag{1}$$

$$\max \{H(Y)\} = -E\{\ln f_y(Y)\}$$

$$Y = \frac{1}{1 + e^{-U}}; U = WX + W_0 \quad (2)$$

To enhance the inter-modality association, three-way pICA attempts to maximize the correlation calculated between the columns of the loadings matrices  $A_1, A_2, A_3$ . Thus, three entropy terms and an additional correlation term comprise the objective function of three-way pICA, as shown in Eq. (3), where  $f_{i,j,k}$  is the additional aggregation function measuring the combined connection strength of the columns  $i, j, k$  from the corresponding matrices  $A_1, A_2, A_3$ , respectively. This objective function will then be optimized through the gradient descent method.

$$\arg \max_{\substack{A_1, A_2, A_3 \\ i, j, k}} \{H(Y_1) + H(Y_2) + H(Y_3) + f_{i,j,k}\} \quad (3)$$

#### D. Correlation and group difference

Pair-wise correlations for all component pairs were evaluated, and significance levels were adjusted using Bonferroni multiple comparison correction for the number of combinations. Ten-fold subsample and permutation tests were then conducted to verify the stability and significance of the linked SNP-GM-FC components as used in our previous studies [14, 15]. Once a most significantly linked SNP-GM-FC pattern was identified, loading parameters of each modality were used to examine group differences by analysis of variance (ANOVA) across three groups and by two-sample t-test for pairwise comparison (Fig. 1). We further test the associations between the identified modalities and the duration of illness to evaluate the impairment speed in imaging modalities.

The identified components were converted to Z scores to select top contributing brain regions or SNPs for subsequent analyses. Top contributing SNPs ( $|Z\text{-scores}| > 2$ ) from the identified SNP component were annotated to genes which were further applied Gene Ontology (GO) analysis by an online database WebGestalt [16].

### III. RESULTS

#### A. SNP-GM-FC Multivariate Linkage

We identified a significantly correlated SNP-GM-FC link (FC-GM:  $r = 0.58$ ,  $p < 10^{-12}$ ; SNP-FC:  $r = 0.31$ ,  $p = 8.83 \times 10^{-5}$ ; SNP-GM:  $r = 0.32$ ,  $p = 3.39 \times 10^{-5}$ ), surviving the Bonferroni-correction. The correlations were consistent in 10-fold stability analysis with the average correlations of  $0.58 \pm 0.08$ ,  $0.28 \pm 0.04$ ,  $0.27 \pm 0.06$  for FC-GM, SNP-FC and SNP-GM respectively. Permutation tests showed the significance of the link with empirical P-values of 0.001 (FC-GM), 0.03 (SNP-FC) and 0.01 (SNP-GM). After controlling for the diagnosis, age, education and gender, partial correlations among the linked components were still significant (FC-GM:  $r = 0.19$ ,  $p = 0.015$ ; SNP-FC:  $r = 0.21$ ,  $p = 0.0074$ ; SNP-GM:  $r = 0.27$ ,  $p = 6.27 \times 10^{-4}$ , passing Bonferroni-correction), suggesting that the identified associations between FC, GM and SNP are general rather than group-biased. The FC, GM and SNP component demonstrated similar changing patterns: less minor/reference allele counts

on SNPs with positive component scores induce reduction of FC and GM.

#### B. Group difference

ANOVA test on the loading parameters of each modality among three groups revealed significant group differences ( $p = 6.09 \times 10^{-49}$ ,  $2.82 \times 10^{-15}$ , 0.0024 for FC, GM and SNP respectively) as shown in Fig. 2. Further pairwise comparison showed that: 1) FC presented an obvious decrease for each pair (HC-FESZ:  $p = 2.26 \times 10^{-6}$ , HC-CSZ:  $p < 1 \times 10^{-10}$ , FESZ-CSZ:  $p = 2.05 \times 10^{-8}$ ), especially between HC and FESZ; 2) For GM, significant group differences were only observed in HC vs. CSZ ( $p < 1 \times 10^{-10}$ ) and FESZ vs. CSZ ( $p = 5.34 \times 10^{-7}$ ); 3) For SNP component, group differences were significant in both HC vs. FESZ ( $p = 0.0023$ ) and HC vs. CSZ ( $p = 0.0083$ ), whereas FESZ vs. CSZ did not differ much, suggesting comparable genetic risks between first episode and chronic patients. Moreover, Fig. 3 described the relationship between GM, FC, and SNP loadings and the duration of illness, where FC exhibited a sharper slope ( $k = -0.0036$ ) than GM ( $k = -0.0006$ ). These suggested that FC may be impaired faster compared to GM.

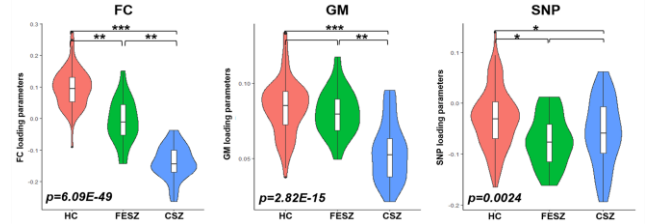


Figure 2. ANOVA analysis results and pairwise two sample t-test results of each modality for the identified SNP-GM-FC pattern. Note: \*represents  $0.00001 < p < 0.05$ , \*\* represents  $1e-5 < p < 1e-10$ , \*\*\* represents  $p < 1e-10$ .

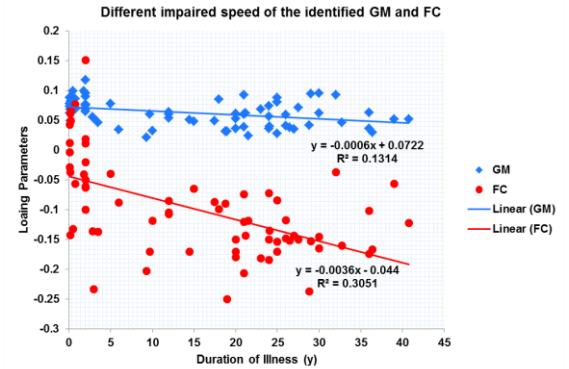
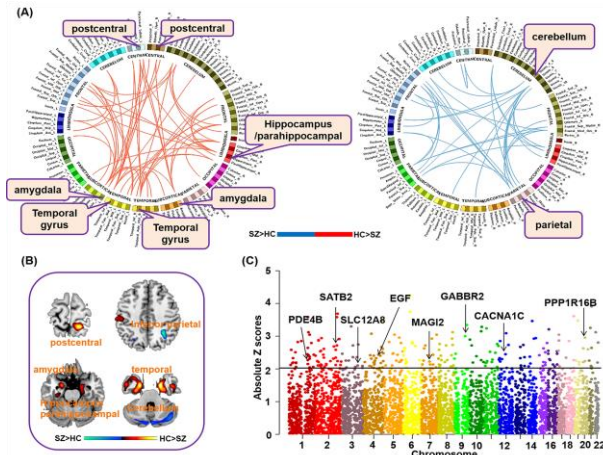


Figure 3. The relationship between the genetic-related imaging components (GM and FC) and illness duration.

The loading plots in Fig. 2 were adjusted as  $HC > patients$  for all modalities on the mean of loading parameters, so that the positive Z-values (red regions) indicate higher contribution in HC than patients and the negative Z-values (blue regions) indicate higher contribution in patients than HC. The connectogram maps of the FC component thresholded at  $|Z| > 2.5$  is shown in Fig. 4A, which indicated that patients have lower FC mainly in postcentral, temporal, amygdala, hippocampus/ parahippocampus, and increased FC mainly in parietal and cerebellum related connectivity. Similarly, by thresholding the GM component with  $|Z| > 2.5$ , the spatial map as shown in Fig. 4B indicated that patients

showed GM decrease mainly in parahippocampal gyrus, temporal gyrus, postcentral gyrus, amygdala and increase in cerebellum and inferior parietal lobule.

Fig. 4C presents a Manhattan plot of all 4937 SNPs, corresponding to highly reported SZ-susceptible genes, *i.e.*, *PDE4B*, *SATB2*, *EGF*, *GABBR2* and *CACNA1C*. 258 SNPs were selected as top-contributing SNPs with  $|Z| > 2$ . After annotating these SNPs, 101 genes were fed to the WebGestalt website for pathway analysis. Primarily significant indicated a significant enrichment of cell junction ( $p=8.26 \times 10^{-7}$ ), synapse ( $p=7.48 \times 10^{-6}$ ) and neuron projection ( $p=1.71 \times 10^{-5}$ ) (FDR corrected  $< 0.05$ ).



**Figure 4. The joint SNP-GM-FC component identified by three-way pICA.** (A) Connectogram plot of the identified FC component with  $|Z\text{-scores}| > 2.5$ . Red connecting lines represent FC decrease in schizophrenia. Blue connecting lines represent FC increase in schizophrenia. (B) Spatial map of the identified GM component with  $|Z\text{-scores}| > 2.5$ . (C) Manhattan plot of the identified SNP component.

#### IV. CONCLUSION

This study firstly combined FC, GM and SNP features to explore the co-altered imaging-genetic pattern at different illness stages of schizophrenia. Results suggested that: 1) both GM and FC components show significant impairments in hippocampus, temporal gyrus and cerebellum in SZ, which are also significantly correlated with genetic component including genes *GABBR2*, *SATB2*, *CACNA1C*, *PDE4B*, involved in pathways like cell junction, synapse and neuron projection; 2) GM and FC presented similar progressively impairment trend, while SNP component indicated no significant group difference between FESZ and CSZ; 3) both genetic related GM and FC components showed significant correlations with illness duration, FC exhibit higher impaired speed than GM, suggesting that functional connectivity may serve as a more sensitive measure to detect the disruptions in SZ at the very early stage. These results increase our knowledge that different imaging features may present different impaired speed with the progress of schizophrenia, which awaits further verification and replication.

#### ACKNOWLEDGMENT

This work was supported by the Chinese National Science Foundation (No. 81471367, 61773380, 81301148), the National High-Tech Development Plan (863, No.

2015AA020513), "100 Talents Plan" of Chinese Academy of Sciences, the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDB02060005), the Primary Research & Development Plan of Jiangsu Province (BE2016630), and NIH Grant R01EB005846, 1R01MH094524 & P20GM103472.

#### REFERENCES

- [1] R. E. Gur, P. E. Cowell, A. Latshaw, B. I. Turetsky, R. I. Grossman, S. E. Arnold, *et al.*, "Reduced dorsal and orbital prefrontal gray matter volumes in schizophrenia," *Arch Gen Psychiatry*, vol. 57, pp. 761-8, Aug 2000.
- [2] T. Onitsuka, M. E. Shenton, D. F. Salisbury, C. C. Dickey, K. Kasai, S. K. Toner, *et al.*, "Middle and inferior temporal gyrus gray matter volume abnormalities in chronic schizophrenia: an MRI study," *Am J Psychiatry*, vol. 161, pp. 1603-11, Sep 2004.
- [3] A. S. Meyer-Lindenberg, R. K. Olsen, P. D. Kohn, T. Brown, M. F. Egan, D. R. Weinberger, *et al.*, "Regionally specific disturbance of dorsolateral prefrontal-hippocampal functional connectivity in schizophrenia," *Arch Gen Psychiatry*, vol. 62, pp. 379-86, Apr 2005.
- [4] S. M. Lawrie, C. Buechel, H. C. Whalley, C. D. Frith, K. J. Friston, and E. C. Johnstone, "Reduced frontotemporal functional connectivity in schizophrenia associated with auditory hallucinations," *Biol Psychiatry*, vol. 51, pp. 1008-11, Jun 15 2002.
- [5] C. Schizophrenia Working Group of the Psychiatric Genomics, "Biological insights from 108 schizophrenia-associated genetic loci," *Nature*, vol. 511, pp. 421-7, Jul 24 2014.
- [6] R. S. K. Hilleke E. Hulshoff Pol, "What Happens After the First Episode? A Review of Progressive Brain Changes in Chronically Ill Patients With Schizophrenia," *Schizophrenia Bulletin*, vol. 34, pp. 354-366, 2008.
- [7] T. Li, Q. Wang, J. Zhang, E. T. Rolls, W. Yang, L. Palaniyappan, *et al.*, "Brain-Wide Analysis of Functional Connectivity in First-Episode and Chronic Stages of Schizophrenia," *Schizophr Bull*, vol. 43, pp. 436-448, Mar 1 2017.
- [8] G. D. Pearson, J. Liu, and V. D. Calhoun, "An introductory review of parallel independent component analysis (p-ICA) and a guide to applying p-ICA to genetic data and imaging phenotypes to identify disease-associated biological pathways and systems in common complex disorders," *Front Genet*, vol. 6, p. 276, 2015.
- [9] V. M. Vergara, A. Ulloa, V. D. Calhoun, D. Boutte, J. Y. Chen, and J. Y. Liu, "A three-way parallel ICA approach to analyze links among genetics, brain structure and brain function," *Neuroimage*, vol. 98, pp. 386-394, Sep 2014.
- [10] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, *et al.*, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *Neuroimage*, vol. 15, pp. 273-289, Jan 2002.
- [11] C. A. Anderson, F. H. Pettersson, G. M. Clarke, L. R. Cardon, A. P. Morris, and K. T. Zondervan, "Data quality control in genetic case-control association studies," *Nat Protoc*, vol. 5, pp. 1564-73, Sep 2010.
- [12] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, *et al.*, "PLINK: a tool set for whole-genome association and population-based linkage analyses," *Am J Hum Genet*, vol. 81, pp. 559-75, Sep 2007.
- [13] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput*, vol. 7, pp. 1129-59, Nov 1995.
- [14] J. Chen, V. D. Calhoun, G. D. Pearson, S. Ehrlich, J. A. Turner, B. C. Ho, *et al.*, "Multifaceted genomic risk for brain function in schizophrenia," *Neuroimage*, vol. 61, pp. 866-75, Jul 16 2012.
- [15] D. Lin, J. Chen, S. Ehrlich, J. R. Bustillo, N. Perrone-Bizzozero, E. Walton, *et al.*, "Cross-Tissue Exploration of Genetic and Epigenetic Effects on Brain Gray Matter in Schizophrenia," *Schizophr Bull*, May 17 2017.
- [16] J. Wang, D. Duncan, Z. Shi, and B. Zhang, "WEB-based GENE SeT Analysis Toolkit (WebGestalt): update 2013," *Nucleic Acids Research*, vol. 41, pp. W77-W83, Jul 2013.