

## ARTICLE OPEN



# The genetic determinants of language network dysconnectivity in drug-naïve early stage schizophrenia

Jingnan Du<sup>1,2,3,11</sup>, Lena Palaniyappan<sup>4,5,11</sup>, Zhaowen Liu<sup>6,7,11</sup>, Wei Cheng<sup>1,3</sup>, Weikang Gong<sup>8</sup>, Mengmeng Zhu<sup>9</sup>, Jijun Wang<sup>2,11</sup>, Jie Zhang<sup>1,2,3</sup> and Jianfeng Feng<sup>1,3,10</sup>

Schizophrenia is a neurocognitive illness of synaptic and brain network-level dysconnectivity that often reaches a persistent chronic stage in many patients. Subtle language deficits are a core feature even in the early stages of schizophrenia. However, the primacy of language network dysconnectivity and language-related genetic variants in the observed phenotype in early stages of illness remains unclear. This study used two independent schizophrenia dataset consisting of 138 and 53 drug-naïve first-episode schizophrenia (FES) patients, and 112 and 56 healthy controls, respectively. A brain-wide voxel-level functional connectivity analysis was conducted to investigate functional dysconnectivity and its relationship with illness duration. We also explored the association between critical language-related genetic (such as FOXP2) mutations and the altered functional connectivity in patients. We found elevated functional connectivity involving Broca's area, thalamus and temporal cortex that were replicated in two FES datasets. In particular, Broca's area - anterior cingulate cortex dysconnectivity was more pronounced for patients with shorter illness duration, while thalamic dysconnectivity was predominant in those with longer illness duration. Polygenic risk scores obtained from FOXP2-related genes were strongly associated with functional dysconnectivity identified in patients with shorter illness duration. Our results highlight the criticality of language network dysconnectivity, involving the Broca's area in early stages of schizophrenia, and the role of language-related genes in this aberration, providing both imaging and genetic evidence for the association between schizophrenia and the determinants of language.

*npj Schizophrenia* (2021)7:18; <https://doi.org/10.1038/s41537-021-00141-8>

## INTRODUCTION

Schizophrenia is one of the most costly and disabling neuropsychiatric disorders. Despite the large number of neuroimaging studies investigating this illness, the origins and the mechanism of schizophrenia has remained largely elusive. Mechanistic delineation is impeded by confounding effects of treatment, as well as the dynamic progression of the disease through several stages in some subgroups of affected patients. While progressive nature of brain changes has been demonstrated in this illness<sup>1–3</sup>, especially in patients with poor prognosis<sup>4,5</sup>, a number of patients recover after the first psychotic episode, and cognitive functioning does not appear to deteriorate over time<sup>6</sup>. Importantly, the early symptomatic years form the most crucial stage of the illness, as long term trajectory is mostly determined during this stage<sup>3,7</sup>. It has been suggested that prognostically critical abnormalities in functional connectivity are evident during the first two years of illness<sup>8–10</sup>, highlighting the critical importance of studying patients with drug-naïve first-episode schizophrenia (FES). Furthermore, the brain regions critical for the variable prognostic trajectories are likely to be captured by studying the early stages of the illness<sup>2,11,12</sup>.

While familial history is the largest known risk factor for schizophrenia<sup>13</sup>, the brain networks through which the genetic risk operates in schizophrenia remain unknown. Given that

functional dysconnectivity patterns are modified with illness progression and treatment<sup>9,14</sup>, it is likely that the early, rather than later stage dysconnectivity, relates more to genetic factors. Patients with schizophrenia have marked language dysfunction and it has been suggested that schizophrenia has a close aetiological relationship with language<sup>15,16</sup>. However, it is yet unknown if language-related genetic markers influence the early stage dysconnectivity in drug-naïve first episode schizophrenia (FES) patients.

In this study we recruited two independent datasets consisting of altogether 191 drug-naïve FES patients and 168 healthy subjects. These unique datasets involved patients with illness duration ranging from 1 to 100 weeks, allowing for a dynamic analysis of disease progression. We adopted voxel-level whole-brain functional connectivity analysis to investigate the progressive functional connectivity changes in these patients that are divided into short-duration and long-duration groups based on median of illness duration. We furthermore tested whether the variations in schizophrenia-related network-level dysconnectivity is related to the genetic mutations in language-readiness genes. Given that compensatory (i.e., secondary) brain changes occur when the illness persists over time<sup>17</sup>, we expected genetic determinants to play a crucial role in the dysconnectivity patterns in early, but not the later stages of the illness. To this end we

<sup>1</sup>Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China. <sup>2</sup>Shanghai Key Laboratory of Psychotic Disorders, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China. <sup>3</sup>Key Laboratory of Computational Neuroscience and Brain Inspired Intelligence (Fudan University), Ministry of Education, Shanghai, China. <sup>4</sup>Department of Psychiatry and Robarts Research Institute, University of Western Ontario, London, ON, Canada. <sup>5</sup>Lawson Health Research Institute, London, ON, Canada. <sup>6</sup>Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. <sup>7</sup>Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. <sup>8</sup>Centre for Functional MRI of the Brain (FMRIB), Nuffield Department of Clinical Neurosciences, Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford, UK. <sup>9</sup>State Key Laboratory of Cognitive Neuroscience and Learning and IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China. <sup>10</sup>Department of Computer Science, University of Warwick, Coventry, UK. <sup>11</sup>These authors contributed equally: Jingnan Du, Lena Palaniyappan, Zhaowen Liu, Jijun Wang. ✉email: [jzhang080@gmail.com](mailto:jzhang080@gmail.com); [jffeng@fudan.edu.cn](mailto:jffeng@fudan.edu.cn)

obtained pathway-specific polygenic risk scores (PRS) from language-related genes. Selecting a subset of genes that are grouped into the same pathway, or co-expressed with given candidate genes for calculating PRS, termed 'pathway-specific PRS' or biologically informed PRS<sup>18–20</sup>, is widely used in image genetics of psychiatric disorders including schizophrenia<sup>21,22</sup>. The goal of this work is to bridge the mechanistic gap between genetic variations and neuroimaging-based measure of brain-wide dysconnectivity in schizophrenia. To this end, we investigate if the dysconnectivity in language-related regions, (i.e., Broca's area) and corresponding language-related genetic variations are related to the early stages of the disease.

## RESULTS

### Cross-validated thalamic and IFG dysconnectivity

For the primary dataset, a total of 5925 voxel-level functional connectivities had  $p$ -values smaller than the CDT  $p = 2 \times 10^{-8}$  ( $z = 5.5$ ) were identified, forming 327 FC clusters with spatial contiguity. The 33 largest FC clusters consisting of 4365 of the 5925 functional connectivities had cluster-size FWER-corrected  $p < 0.05$ , see Supplementary Table 2. Other small-sized FC clusters with cluster-size FWER-corrected  $p > 0.05$  are identified as false positives and are filtered out. To show the 33 significant FC clusters clearly, each one of them was mapped to the corresponding region in AAL2<sup>23</sup> atlas (i.e., the region comprising of the majority of the voxels), as shown in Supplementary Table 2. The anatomical location of the brain areas relating to the 33 FC clusters was shown in Fig. 1a and Supplementary Fig. 1. Meanwhile, to display the distribution of altered voxel-level functional connectivities within each AAL2 regions, the Manhattan plot of voxel-based BWA results with voxels grouped in accordance with the AAL2 atlas labels was given as Supplementary Fig. 2.

The 33 significant FC clusters primarily involved thalamus, temporal cortex, inferior frontal cortex, precuneus, and subcortical structures like pallidum, see Fig. 1b. Among them, ten altered FC clusters involved bilateral thalamus, with the top 2 largest FC clusters connecting bilateral thalamus and middle/ superior temporal lobe. There were 5 FC clusters which connected the left inferior frontal gyrus (triangular and orbital part, including Broca's area), to anterior cingulate cortex, left medial superior frontal gyrus, and subcortical regions (putamen and pallidum). In addition, there were 4 FC clusters which connected precuneus to right middle/superior temporal gyrus, hippocampus, and left calcarine, and 6 FC clusters connected pallidum to middle/superior temporal gyrus, the left orbital part of inferior frontal gyrus (IFG).

The results on the primary dataset were cross-validated on an independent dataset. 16 of the 33 FC clusters were significantly altered, see Fig. 2b and Supplementary Table 5, mostly involving left inferior frontal triangular gyrus, bilateral thalamus, left middle temporal lobe, and right superior temporal lobe.

### Thalamic and IFG connectivities are correlated with negative symptoms

We found that FC clusters between thalamus and superior/middle temporal gyrus, and between left inferior frontal gyrus (i.e., Broca's area) and right anterior cingulate cortex were significantly correlated with PANSS negative total score, see Table 1. A further analysis of score of each item revealed that the connectivity between left inferior frontal gyrus (triangular and orbital part) with anterior cingulate cortex was correlated with conceptual disorganization (positive symptom), emotional withdrawal, passive/apathetic social withdrawal and difficulty in abstract thinking (negative symptom), which are related to language impairments, see Supplementary Table 6 and Supplementary Table 7.

### IFG and thalamic dysconnectivity respectively in short and long duration FES

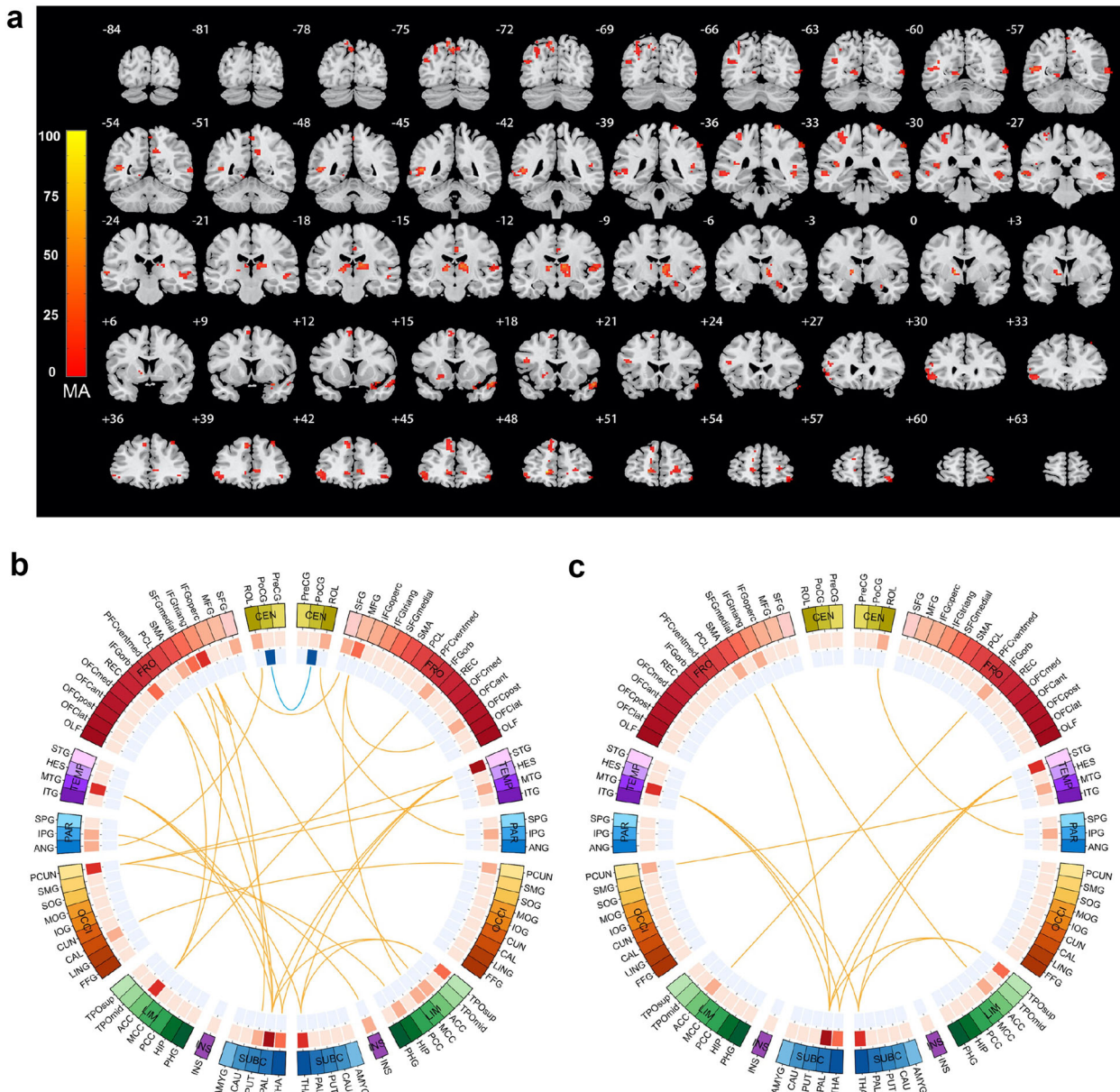
After identifying changes in FES patients overall, we then explored alterations specific to different stages, i.e., long (69 patients with a mean of 48.6 weeks) and short (69 patients with a mean of 11.2 weeks) duration group. We applied the cluster-level inference approach to identify significantly altered FC clusters (CDT  $p = 3 \times 10^{-7}$  ( $z = 5$ ) and cluster-size FWER-corrected  $p = 0.05$ ) in long and short duration group. The results were shown in Supplementary Tables 3 and 4.

There were 4 FC clusters (Fig. 2, Supplementary Table 3) that can survive the cluster-size FWER-corrected  $p < 0.05$  threshold in short-duration group including 3 FC clusters linking the left triangular inferior frontal gyrus, which is part of Broca's area, the right orbital and triangular part of the inferior frontal gyrus to precuneus and anterior cingulate cortex in the opposite hemisphere. In the long-duration group, 29 FC clusters (Fig. 2, Supplementary Table 4) had cluster-size FWER-corrected  $p < 0.05$  and most clusters linked thalamus bilaterally, middle temporal lobe and superior temporal lobe and triangular part of inferior frontal gyrus. There were 5 FC clusters involving thalamus in the top 10 clusters in long duration group.

### Language genes play more important roles in short-duration FES

To explore the genetic mechanism underlying the altered functional connectivity in FES patients, and specifically, the role of language-related genes regarding the language hypothesis, we performed genetic association analysis using pathway-specific polygenic risk scores obtained from the gene cluster involving FOXP2 (6 genes), which are implicated in both language and schizophrenia, and are associated with development, see Fig. 3a. First, we found that 20 SNPs in these genes are associated with disease ( $p < 0.05$ ), see Fig. 3b. Then we calculated the correlation coefficients between the polygenic risk score from these 20 SNPs and the altered FC clusters identified in short-duration FES patients (4 FC clusters, see Supplementary Table 3) and long-duration FES patients (29 FC clusters, see Supplementary Table 4), respectively, and found that polygenic risk scores had stronger correlation coefficients ( $t(33) = 1.87$ ,  $p = 0.036$ ) with altered FC clusters identified in short-duration group (between left inferior frontal gyrus and right anterior cingulate cortex, and between left angular gyrus and left rolandic operculum), see Fig. 3c. This suggested that the 6 language genes together may be responsible for the FC changes in short-duration FES patients. In contrast, polygenic risk scores did not show significant correlation with any altered FC clusters identified in the longer duration group. To verify this we furthermore performed a comparison between the strength of correlation (a single correlation coefficient obtained by plotting the mean of all altered voxel-level FCs against polygenic risk score in each subject) in short duration (1375 voxel-level FCs) vs. long duration patient groups (13,233 voxel-level FCs). By non-parametric permutation ( $n = 10,000$ ) we found a significant correlation between PRS and the mean FCs in the short-duration group ( $p = 0.049$ ) but not the long-duration group ( $p = 0.716$ ).

To establish of the specificity of the FOXP2 cluster genes in the functional connectivity abnormalities seen in our untreated schizophrenia samples, we furthermore reanalyzed genetic association using SNPs from all 20 genes that are not filtered by co-expression and FOXP2-linked clustering. The mean of the absolute correlation between the unfiltered PRS (from the 20 genes) and the altered FCs was lower than using the PRS obtained from the FOXP2 cluster genes, and there was no significant correlation between the unfiltered PRS and FCs in both the early and later stages of schizophrenia (mean correlation coefficient in patients of short and long illness duration being 0.069 and 0.062, respectively, without significant difference). This highlighted the



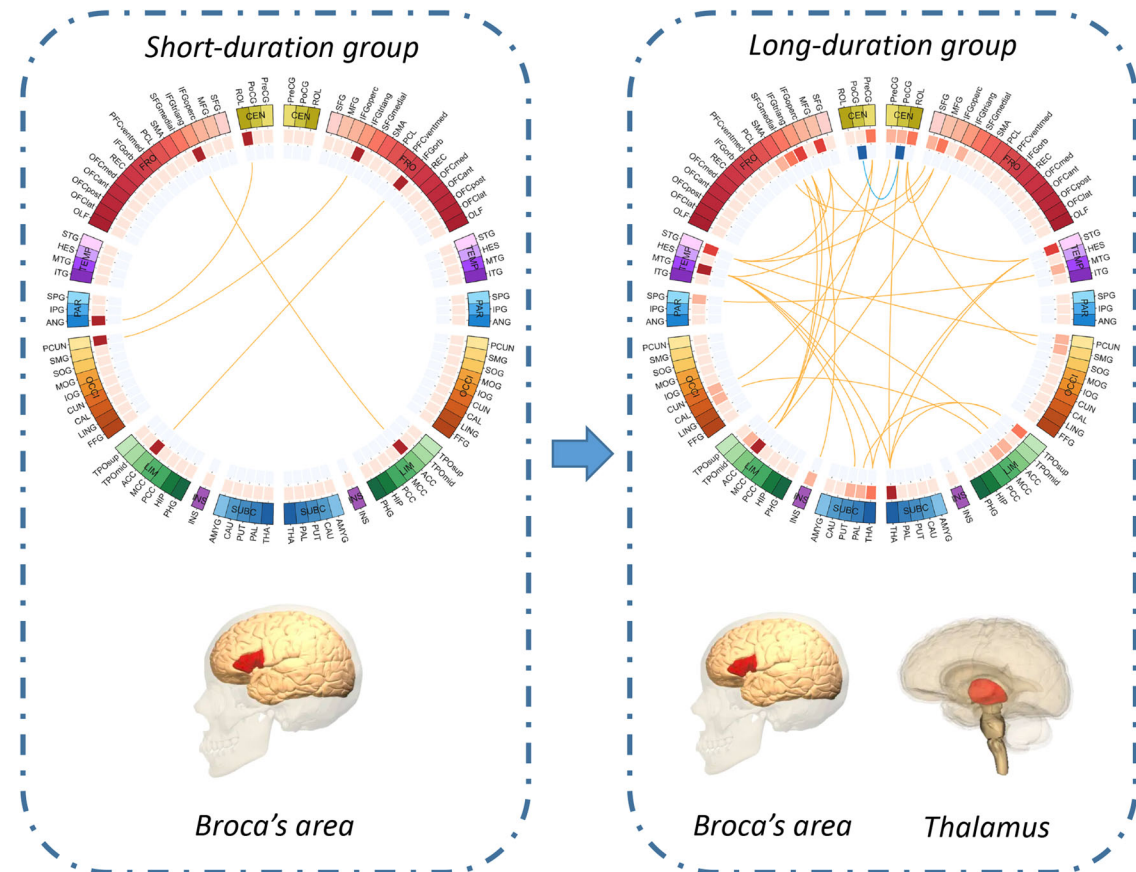
**Fig. 1** Significantly altered functional connectivity clusters (FC clusters) identified in the first-episode schizophrenia (FES) patients compared to healthy controls in the primary and validation dataset. **a** Voxels showing the largest number of voxel-level functional connectivity differences in FES patients in the primary dataset. The color bar represents the measure of association (MA) which is the number of significantly altered functional connectivities relating to each voxel. **b** Significantly altered FC clusters identified in FES patients in the primary dataset (33 FC clusters in total, details see Supplementary Table 2). The color of the 3 circles (from inside to outside) denotes: the number of deceased links; the number of increased links; the 94 different automatic anatomical labeling (AAL2) regions. Red lines mean that patients have a higher mean functional connectivity than controls, and blue lines indicate the opposite. The anatomical abbreviations in the figure are shown in the Supplementary Table 1. **c** 16 of the 33 significantly altered FC clusters were replicated in an independent validation dataset, involving the left inferior frontal triangular gyrus, bilateral thalamus, the left middle temporal lobe and the right superior temporal lobe.

specificity of FOXP2 gene assemblies with conjoint spatial expression in functional dysconnectivity seen in FES.

**DISCUSSION**

In a large sample of drug-naïve FES patients with a broad range of illness duration (1 to 100 weeks), we report that increased resting-state functional connectivity involving thalamus and Broca’s area as the predominant aberration in functional connectivity. Notably, for the first time we establish that irrespective of illness duration, untreated patients consistently show altered connectivity in

inferior frontal gyrus (site of the Broca’s area) and superior/middle temporal gyrus, cortical regions critical for various levels of human language processing. For patients in the earliest stages of illness (shorter duration), mutations in the genetic determinants of language (FOXP2 related genes) play a critical role in the observed functional dysconnectivity between Broca’s area and anterior cingulate cortex. Finally, we also note that the degree of anterior cingulate cortex dysconnectivity with language-related areas (inferior frontal and superior temporal regions) is correlated with the severity of core negative symptoms of schizophrenia (emotional withdrawal, passive/apathetic social withdrawal and

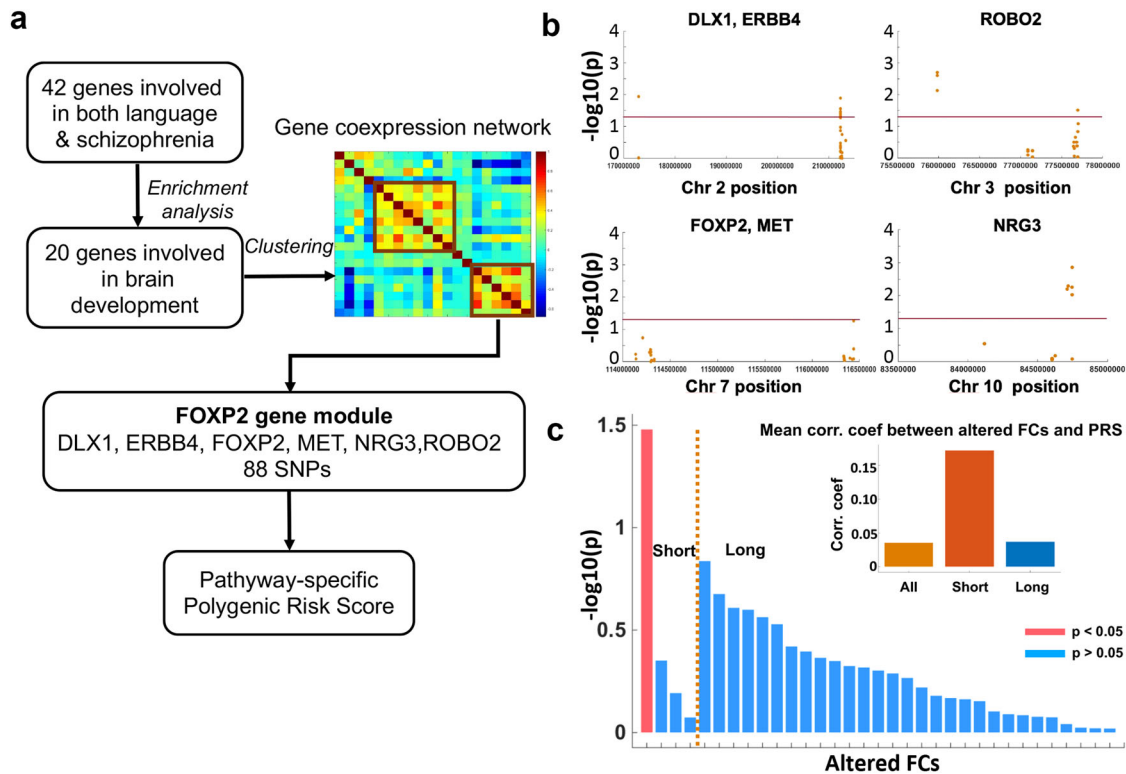


**Fig. 2** Significantly altered functional connectivity clusters (FC clusters) identified in both the short-duration group ( $11.2 \pm 6.2$  weeks) and the long-duration group ( $48.6 \pm 22.8$  weeks) first-episode schizophrenia (FES) patients. The color of the 3 circles (from inside to outside) denotes: the number of decreased functional connectivity; the number of increased functional connectivity; the 94 brain regions in automatic anatomical labeling (AAL2) atlas. Red lines mean that FES patients have a higher mean functional connectivity than healthy controls, and blue lines indicate the opposite. The anatomical abbreviations in the figure are shown in the Supplementary Table 1. Short-duration group had functional dysconnectivity mainly in the Broca's area (the left inferior frontal gyrus), while long-duration group had widespread functional dysconnectivity involving both the left inferior frontal gyrus and the thalamus. The brain images in this figure are snapshots of 3D brain model of the Society for Neuroscience (2017) retrieved from <https://www.brainfacts.org/3DBrain>.

**Table 1.** Correlations coefficient between the functional connectivity links and PANSS negative score.

Functional connectivity clusters					
Region 1	Region 2	<i>r</i>	<i>p</i>	FDR <i>p</i>	
Left thalamus	Right superior temporal gyrus	0.3043	0.0005	0.0165	
Right thalamus	Left middle temporal gyrus	0.2374	0.0060	0.0469	
Right thalamus	Right superior temporal gyrus	0.2306	0.0063	0.0469	
Left inferior frontal gyrus, triangular part	Right anterior cingulate cortex	0.2256	0.0068	0.0469	
Right thalamus	Right superior temporal gyrus	0.2254	0.0071	0.0469	
Left pallidum	Left middle temporal gyrus	0.2138	0.0129	0.0693	
Right thalamus	Left middle temporal gyrus	0.2112	0.0149	0.0693	
Right thalamus	Left middle temporal gyrus	0.2101	0.0168	0.0693	
Left pallidum	Right superior temporal gyrus	0.2016	0.0191	0.0700	
Left thalamus	Left middle temporal gyrus	0.1995	0.0235	0.0775	
Left precuneus	Right superior temporal gyrus	0.1849	0.0324	0.0894	
Left thalamus	Left middle temporal gyrus	0.1900	0.0325	0.0894	

<sup>a</sup>FDR *p* False Discovery Rate corrected *p* value.



**Fig. 3 Genetic association analysis between altered functional connectivities (FCs) identified in FES patients and language-related genes.** **a** Workflow to choose candidate genes in genetic association study. We identified one gene cluster (with 6 genes) containing FOXP2, which is related to neurodevelopment, language, and schizophrenia. **b** Association analysis between the 88 SNPs in the 6 genes in the FOXP2 cluster and schizophrenia. **c** Correlation coefficients between the pathway-specific polygenic risk scores (PRS) obtained from the above 6 genes and each of the altered FCs identified in both the short-duration patient group and the long-duration patient group. The bars represented the log  $P$  values of the correlation between the PRS and 4 altered FCs identified in short-duration group and the 29 FCs identified in the long-duration group (the red bar indicate  $p < 0.05$ ). The upper right part of this figure showed the mean value of the correlation coefficients (between altered FCs and PRS) in different patients groups (orange for all FES patients, red for short-duration FES group, and blue for long-duration FES group).

difficulty in abstract thinking), reflecting the primacy of language dysfunction in its pathophysiology.

Our results complement our previous findings of Broca's area dysconnectivity in first-episode schizophrenia and thalamo-cortical dysconnectivity in chronic schizophrenia<sup>8</sup>. In contrast to our previous multi-site work, we used drug-naïve FES patients with wide distribution of illness duration (1 to 100 weeks) in the current study, allowing us to study the effect of chronicity in a drug-naïve state. The short-duration group in the current study had FC changes (mainly Broca's area) similar to the first-episode group in our previous sample, while the long-duration group had FC changes (mainly thalamus and temporal lobe) similar to the chronic illness group in the previous sample. These complementary observations strongly suggest that Broca's area (IFG) may be the 'site of origin' of FC changes in schizophrenia (see Fig. 2). This result is supported by recent findings of elevated PFC connectivity in FES patients<sup>9</sup>, and increased functional connectivity involving Broca's area that could have prognostic and therapeutic implications<sup>24</sup>.

Broca's area plays a key role in both syntactical and semantic processing of language. Structured language production plays an important role in multi-step thinking<sup>25</sup>, with aberrant connectivity of Broca's region likely contributing to difficulties in negative symptoms and disorganization, often seen in early stages of psychosis, even before the onset of reality distortion<sup>26</sup>. Such symptoms are also of prognostic importance in the presence of psychosis<sup>27</sup>. In contrast, for patients with longer illness duration (23.3 weeks–100 weeks), the most significant changes involve the superior/middle temporal cortex and thalamus. The involvement

of superior temporal cortex is suggestive of language-related processing deficits<sup>28,29</sup>, and can be viewed as an index of chronicity of the untreated illness. It is also possible that patients with predominant Broca's area dysconnectivity present earlier due to higher illness severity or acuity. Longitudinal fMRI studies in untreated samples are required to test these speculations.

The changes in inferior frontal gyrus functional connectivity in short-duration patient group and its correlation with negative symptoms indicates that dysfunction in Broca's area may be related to schizophrenia, in line with the hypothesis of Crow<sup>15,16</sup>. Crow, suggested that schizophrenia is the breakdown of the normal left-sided brain specialization for language and a "side effect" of human linguistic evolution<sup>15,16</sup>; this was supported by recent polygenic analysis of GWAS data<sup>30</sup>. On the basis of 42 candidate genes implicated in both language readiness and schizophrenia<sup>31–33</sup>, and clustered using their whole-brain gene expression profiles (Allen Brain Atlas), we identified a group of genes including FOXP2, which were closely associated with language, schizophrenia, and brain development. Association analysis between the 88 SNPs in the FOXP2 genes cluster and schizophrenia revealed IFG hyper-connectivity being closely related with polygenetic scores of the FOXP2 cluster in patients at an early-stage illness, while other genetic/non-genetic factors may influence the dysconnectivity patterns in those with longer duration of illness. This provides the first direct neuroimaging and genetic evidence for Crow's linguistic primacy hypothesis. Our results also highlight the importance of teasing apart the effect of medication exposure and illness duration in imaging genetics studies of psychosis. We speculate that the genetic determinants

of the initial functional dysconnectivity critical for symptom emergence are likely to be different from the determinants of illness persistence and chronicity. This is consistent with the observation that not all FES patients develop a persisting illness<sup>6</sup>, contributing to the heterogeneous trajectory of this condition. Compensatory brain changes may also occur after early psychotic symptoms<sup>17</sup>, contributing to structural or functional reorganization<sup>34</sup>, thus weakening the link between genetic risk factors and observed dysconnectivity phenotype.

Schizophrenia is considered by some authors as a ‘language-related human specific disease’ or ‘logopathy’, which promises a better understanding of the disease pathophysiology<sup>35</sup>. FOXP2 gene plays an important role in the development of the neural systems that mediate speech and language<sup>36</sup>. FOXP2 polymorphisms have been involved in neuropsychiatric and developmental disorders like schizophrenia<sup>37,38</sup>, autism spectrum disorders, and language disorder like dyslexia<sup>39</sup>. Importantly, FOXP2 polymorphisms have been shown to be associated with the severity of a core language symptom, i.e., formal thought disorder in schizophrenia<sup>40</sup>, and the inferior frontal gyrus is a site of pathology that has been most consistently associated with FOXP2 disruption in multiple studies<sup>41</sup>. In addition to speech and language disorders, FOXP2 has also been implicated in adult ADHD<sup>42,43</sup>. A FOXP2 knockout models implicate a role for this gene in regulating dopamine in subcortical circuits<sup>44</sup>. Taken together, our results have the potential to tie up the linguistic basis of schizophrenia with its neurochemical basis in dopamine.

In our work, the increased functional connectivity of left IFG (see Supplementary Table 3), specifically with the opposite hemisphere, may be related to disrupted lateralization of schizophrenia patients, which may be associated with the language gene cluster that involves FOXP2, previously shown to be responsible for language lateralization in both healthy controls and FES patients<sup>45,46</sup>. Other genes in this FOXP2 gene cluster include NRG3, ERBB4, DLX1, ROBO2, and ROBO2, all are susceptibility gene in schizophrenia and is related to development<sup>47</sup>. Among them, the NRG3N, one in the large family of growth factors, and its receptors ERBB4 genes are part of the NRG-ERBB signaling network that has been implicated intensively in schizophrenia<sup>48</sup>. This pathway plays important roles in neural development including circuitry generation, axon ensheathment, neurotransmission, and synaptic plasticity. The variants of these genes have been shown to enhance synchronized oscillations of prefrontal cortex neurons via inhibitory synapses, possibly responsible for the increased functional connectivity related to IFG in first-episode patients<sup>49</sup>.

One limitation of our work is the lack of symptom scores or task parameters pertaining to linguistic deficits per se. So, we are not able to assess how language network dysconnectivity operates to influence the core feature of conceptual disorganization in schizophrenia. We cannot interpret the dysconnectivity observed here as a substrate of the thought, language, and communication disorder in schizophrenia. We also caution readers against generalizing the reported observations to verbal performance and abilities in patients, both of which were not examined here.

A polygenic cluster related to language as well as schizophrenia, comprised of FOXP2, and NRG-ERBB signaling pathways, influences the connectivity of Broca’s area, which in turn influences the clinical expression of core symptoms in untreated early stages of schizophrenia. Our results, reported from a discovery as well as a validation sample of drug-naïve subjects, bridges the mechanistic gap between genetic variations and clinical symptoms of schizophrenia. It also raises the intriguing question of the role that linguistic interventions can play in reducing the pathophysiological processes that define early stages of schizophrenia.

## METHODS

### Subjects

Resting-state fMRI and whole-exome sequencing data of all participants were collected at Shanghai Mental Health Center from 2016 to 2018. The study was approved by the Shanghai Mental Health Center Institutional Review Board, and all participants provided written informed consent. The discovery (primary) dataset included 138 patients and 112 controls (107 patients and 71 healthy controls with whole-exome sequencing data) and the validation dataset included 53 patients and 56 controls after excluding the subjects with mean framewise displacement exceeding 0.5 mm, see Table 2 for demographic and clinical data for details. All the subjects were Mandarin-speaking Han Chinese individuals from Shanghai metropolitan area. The FES patients were identified according to *DSM-IV* criteria<sup>50</sup> by qualified psychiatrists using all available clinical information including a diagnostic interview of patients and their family, clinical case notes, and clinician’s observations. Symptom severity was measured using the standardized positive and negative syndrome scale (PANSS) assessment. FES patients were defined as having illness duration less than 2 years, as this is the most common definition used in the literature<sup>51</sup>. The duration of untreated illness [DUP] ranged from 1 to 100 weeks. There were 25 patients (18.1%) with illness duration longer than 52 weeks. It is important to note that this broad range of DUP (i.e., illness duration at the time of scanning) is representative of FES samples across the world, and in line with the findings that on average the DUP of lower and middle income countries is twice as long as the DUP in high-income countries<sup>52</sup>. All healthy subjects were assessed in accordance with *DSM-IV* criteria as being free of schizophrenia and other axis I disorders, and none had neurological

**Table 2.** Demographic and clinical characteristics of first-episode schizophrenia patients and matched healthy controls in the primary and validation dataset.

Characteristic	Primary dataset				Validation dataset					
	FES	HC	Test statistic	<i>P</i> value	FES short group	FES long group	FES	HC	Test statistic	<i>P</i> value
Number of subjects	138	112	NA	NA	69	69	53	56	NA	NA
Age, mean (SD), years	24.0 (8.1)	24.5 (6.3)	−0.48	0.63	22.8 (7.8)	25.3(8.3)	25.5 (6.6)	25.8 (6.2)	−0.12	0.9
Gender (M/F)	79/ 59	50/ 62	3.92	0.05	45/ 24	34/ 35	31/ 22	30/ 26	0.41	0.68
Education level, mean (SD), years	11.6 (2.8)	12.8 (3.0)	−3.19	0.002	11.3 (2.8)	11.9 (2.8)	12.6 (3.2)	13.0 (2.7)	−0.08	0.78
Duration, weeks	29.9 (25.1)	NA	NA	NA	11.2 (6.2)	48.6 (22.8)	25.4 (18.1)	NA	NA	NA
PANSS positive score	20.5 (5.7)	NA	NA	NA	20.6 (5.8)	20.4 (5.6)	15.6 (4.4)	NA	NA	NA
PANSS negative score	14.9 (6.9)	NA	NA	NA	14.2 (6.5)	15.6 (7.2)	7.7 (2.9)	NA	NA	NA

<sup>a</sup>PANSS Positive and Negative Syndrome Scale.

<sup>b</sup>FES First Episode Schizophrenia group.

<sup>c</sup>HC Healthy Control group.

diseases, head trauma, or substance abuse. All participants were between the ages of 14 and 45 years; were right-handed; had no history of substance abuse or suicidal ideation; and had no MRI contraindications.

### Image acquisition and preprocessing

Resting-state fMRI data were acquired using a 3 T MRI scanner (Siemens) in an 8-min period in which the participants were awake in the scanner. A total of 240 volumes of images were obtained (TR/ TE: 2000/30 ms, Flip angle 77° matrix size: 64 × 64, voxel size: 3 × 3 × 3 mm<sup>3</sup>; FOV = 220 × 220 mm<sup>2</sup>, slices 50).

fMRI data were preprocessed using FSL<sup>53</sup> and AFNI<sup>54</sup>. Images of each subject were corrected for slice timing (FSL) and motion with 24 head movement parameters<sup>55</sup>. We registered functional images to 3 mm standard Montreal Neurological Institute (MNI) space by firstly aligning functional images to the subject's T1 structural images and then transformed them to the standard space (FSL). The functional volumes were spatially smoothed with a Gaussian kernel (6 mm<sup>3</sup> full-width half-maximum (FWHM)). Wavelet despiking was applied to denoise the time series of voxels<sup>56</sup>, and we band-pass filtered the time series between 0.01 and 0.1 Hz using AFNI. White matter signal, cerebrospinal fluid signal, and global mean signal were also regressed out (AFNI) in line with our previous work<sup>8,57</sup>. Subjects with mean framewise displacement exceeding 0.5 mm were excluded from the analysis.

### Whole exome sequencing data preprocessing

Library preparation – The 300 ng genomic DNA were sheared with Covaris LE220 Sonicator (Covaris) to target of 150–200 bp average size. DNA libraries were prepared using SureSelectXT Human All Exon V6 + UTR (Agilent). The fragments were repaired the 3' and 5' overhangs using End repair mix (component of SureselectXT) and purified using Agencourt AMPure XP Beads (Beckman). Finally, the pre-capture libraries containing exome sequences were captured using SureSelect capture library kit (Agilent).

Illumina sequencing DNA concentration of the enriched sequencing libraries was measured with the Qubit 2.0 fluorometer dsDNA HS Assay (Thermo Fisher Scientific). Paired-end sequencing is performed using an Illumina NovaSeq6000 system following Illumina-provided protocols for 2 × 150 paired-end sequencing.

Variant Detection Gene variant refer to changes in the base of genome, including point mutations caused by single base changes, or deletions, duplications, or insertions of multiple bases. In this analysis, the workflow of Sentieon<sup>58</sup> was carried out to detect SNV (Single Nucleotide Variants) and InDel (small insertions and deletions). The variant detection of analysis processes are as follows: 1. PF reads are aligned to the human reference genome (hg19) and sorted. 2. Mark duplicated PCR reads in bam files (Dedup.bam). 3. Reads in the region of InDel were realigned (InDel realign). 4. Base quality score recalibration (BQSR). 5. Detection of SNV and InDel. 6. Variant quality score recalibrate.

### Cluster-based brain-wide functional connectivity study

In the present study, we adopted a cluster-based brain-wide association study (BWAS) method<sup>59</sup> to identify altered functional connectivity clusters in FES patients in the primary dataset. Each resting-state brain image included 47619 voxels (3 × 3 × 3 mm<sup>3</sup>) within the cerebrum (based on AAL2 atlas<sup>23</sup>, which was also used for labeling the clusters). For each participant, Pearson's correlation coefficients between the fMRI time series at each brain voxel were computed followed by Fisher's z-transformation. This resulted in a 47619 × 47619 functional connectivity matrix for each subject. A general linear model was used to compare functional connectivity between FES patients and healthy controls, with age, gender, education level, mean framewise displacement as covariates. A z-statistic was obtained which reflected the degree of change in patients for each voxel by voxel connectivity (i.e., each cell of the 47619 × 47619 matrix). A predefined cluster-defining threshold (CDT) of  $p = 2 \times 10^{-8}$  ( $z = 5.5$ ) was applied to each voxel-level test. Then the voxel–voxel connections with  $p$  value less than the CDT with edges linking a given pair of the same two distinct voxel clusters were identified as a functional connectivity cluster (FC cluster)<sup>59</sup>. To correct for multiple comparisons, the cluster-level inference method based on random field theory was used, which estimated the cluster-size family-wise error rate (FWER)-corrected  $p$  value using the size of each FC cluster, i.e., the number of voxel-level functional connectivities within it<sup>59</sup>. Large-sized FC clusters (FWER-corrected  $p < 0.05$ ) were identified as significantly altered FC clusters in patients.

We then calculated the partial correlation coefficients between the strength of significantly altered FC clusters (FWER-corrected  $p < 0.05$ ), specifically, the mean of all the voxel-level functional connectivities within it, with the positive and negative syndrome scale (PANSS) scores, with the same covariates as mentioned above. Here, 10,000 random permutations were implemented to estimate the null distribution of the correlation coefficients between functional connectivities and PANSS scores and also the statistical  $p$  values. A false discovery rate (FDR) procedure<sup>60</sup> was used to correct for multiple hypothesis testing.

### Stage-specific changes of functional connectivity with illness duration

In order to evaluate the dynamic changes of functional connectivity with illness duration, we split the FES patients in the primary dataset into two groups based on the median of duration (23.3 weeks): short (69 patients with a mean of 11.2 weeks (1–22.3 weeks)) duration group and long (69 patients with a mean of 48.6 weeks (23.3–100 weeks)) duration group (Table 2), and compared with all the healthy controls (112 subjects), respectively, adopting cluster-based brain-wide functional connectivity analysis at the voxel-level. Due to the relatively smaller sample size, we adopted a correction threshold (CDT  $p = 3 \times 10^{-7}$  ( $z = 5$ ) and cluster-size FWER-corrected  $p = 0.05$ ) in the cluster-level inference in the two groups.

### Cross-validation on an independent dataset

We validated our observations of the primary dataset in an independent validation dataset consisting of 109 participants (53 FES; mean age, 25.7 years; mean illness duration, 25.4 weeks). Specifically, we calculated the mean value of all the voxel-level functional connectivities within each FC cluster identified in the primary dataset in the independent validation dataset and then test whether these mean values were changed in the FES patients compared with healthy controls.

### Language dysfunction and negative symptoms

The dimension of disorganization and formal thought disorder correlates strongly with negative symptoms, as first shown by McGorry et al., in a factor analysis study<sup>61</sup> (with predictive value of this combined dimension confirmed by Dominguez et al.<sup>26</sup> in a 10-year longitudinal study). Furthermore a strong relationship exists between impoverishment (negative symptoms) and communication dysfunction in schizophrenia<sup>62</sup>, and this association occurs several years before positive psychotic symptoms appear<sup>63</sup>. Language dysfunction in schizophrenia is traditionally considered to include a positive dimension of incoherence, distractibility, and tangential speech and a negative dimension of reduced speech, poverty, and perseveration of ideas. Of these, the positive dimension is more episodic, seen during acute phases and share variance with other positive symptoms (delusions, hallucinations), while the negative dimension is more pervasive, trait like, and stable, making a suitable substrate for genetic studies<sup>64</sup>. Considering that the negative symptoms (Alogia, Autism, Ambivalence, and Affect Blunting) and disorganization (Loosening of Associations) are the fundamental features of this illness<sup>65</sup>, we consider the negative symptom dimension as an appropriate, stable proxy for the clinical expression of impoverished thought, language, and communication in schizophrenia.

### Genetic association analysis

To test the hypothesis if there is a relation between the altered functional connectivity in language-related areas and language-related gene mutations, we carried out a genetic association analysis using the pathway-based polygenic risk scores<sup>18,19</sup>, as gene sets centering on putative core function, or belonging to specific biologically relevant pathways, may make a larger-than-expected contribution to polygenic risk<sup>21,66</sup>. In particular, we chose the candidate genes that are related to schizophrenia, development, and language by combing knowledge from multiple sources, including existing literature, gene ontology analysis, and brain gene expression data. These genes involved the Forkhead box P2 (FOXP2) gene that is well-known to be associated with language which is shown to be a promising candidate in language dysfunction in schizophrenia<sup>47</sup>.

The workflow of the genetic analysis was shown in Fig. 3a. We started from the 42 genes that are shown to be related to both language readiness<sup>31–33</sup> and schizophrenia listed by Murphy & Benitez-Burraco in an extensive recent review<sup>47</sup>. We performed functional enrichment analysis (by TopGene Suite<sup>67</sup>) and selected 20 genes that had an established

ontological role in the human brain development, in line with the neurodevelopmental hypothesis of schizophrenia<sup>68</sup>. Next, we clustered these 20 genes using their whole-brain gene expression profiles (obtained from six adult brains in the Allen Brain Atlas<sup>69</sup>) by K-means algorithm, and we identified the genes that are co-expressed with FOXP2, which we termed as the FOXP2 cluster genes<sup>70</sup>. We chose to focus on FOXP2 gene cluster given the extant evidence that this gene bridges neural plasticity and spoken language<sup>71,72</sup>, and its expression is enriched in language areas and it plays a role in their specialization<sup>73</sup>. The clustering step ensured that we identified functional gene assemblies that require a co-regulated transcriptional profile and conjoint spatial expression to influence language function<sup>74</sup>. Thus, we identified a distinct module of language, brain development and schizophrenia related, co-expressed ensemble of genes not only important for brain maturation in anatomically modern humans, but also relevant to the expression of the clinical and connectivity phenotype of schizophrenia<sup>47</sup>.

We finally calculated the polygenic risk score using the SNPs in this FOXP2 gene cluster ( $p < 0.05$ ), i.e., weighted sum of these SNPs with log (odds ratio) for each SNP being the weight. The strength of the altered FC clusters that are identified in the whole FES group, long-duration patients group, and short-duration patients group were respectively extracted in all 107 patients with both genetic and imaging data in the whole FES group. Then we explored the relationship between this pathway-specific polygenic risk score and strength of the FC clusters in all 107 patients. Our general goal was to identify the correlation between schizophrenia-related network-level dysconnectivity and genetic mutations in language-readiness genes, rather than to predict disease state using all genotype data as traditional GWAS study does. Therefore we used the SNP variants called from exome sequencing data (specifically the FOXP2 gene cluster exomic loci) for the schizophrenia samples, following the principle of pathway-specific PRS. Note that exome data has been frequently used in imaging-genetic studies in psychiatric disorders<sup>75,76</sup>.

## Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## DATA AVAILABILITY

The datasets generated during the current study are not publicly available due to ethical codes for this study but are available from the corresponding author on reasonable request with the approval of The Research Ethics Committee of the Shanghai Jiao Tong University School of Medicine.

## CODE AVAILABILITY

The BWAS: Voxel-level connectome-wide association studies code is publicly available at <https://github.com/weikangong/BWAS>. The code used for these analyses is publicly available at the following url: <https://osf.io/2zaqv/>.

Received: 9 July 2020; Accepted: 12 January 2021;

Published online: 03 March 2021

## REFERENCES

- Van Haren, N., Cahn, W., Hulshoff Pol, H. & Kahn, R. The course of brain abnormalities in schizophrenia: can we slow the progression? *J. Psychopharmacol.* **26**, 8–14 (2012).
- Ellison-Wright, I., Glahn, D. C., Laird, A. R., Thelen, S. M. & Bullmore, E. The anatomy of first-episode and chronic schizophrenia: an anatomical likelihood estimation meta-analysis. *Am. J. Psychiatry* **165**, 1015–1023 (2008).
- Fenton, W. S. & McGlashan, T. H. Natural history of schizophrenia subtypes: II. Positive and negative symptoms and long-term course. *Arch. Gen. Psychiatry* **48**, 978–986 (1991).
- Lieberman, J. et al. Longitudinal study of brain morphology in first episode schizophrenia. *Biol. Psychiatry* **49**, 487–499 (2001).
- Mathalon, D. H., Sullivan, E. V., Lim, K. O. & Pfefferbaum, A. Progressive brain volume changes and the clinical course of schizophrenia in men: a longitudinal magnetic resonance imaging study. *Arch. Gen. Psychiatry* **58**, 148–157 (2001).
- Zipursky, R. B., Reilly, T. J. & Murray, R. M. The myth of schizophrenia as a progressive brain disease. *Schizophr. Bull.* **39**, 1363–1372 (2013).
- Cahn, W. et al. Brain volume changes in first-episode schizophrenia: a 1-year follow-up study. *Arch. Gen. Psychiatry* **59**, 1002–1010 (2002).
- Li, T. et al. Brain-wide analysis of functional connectivity in first-episode and chronic stages of schizophrenia. *Schizophr. Bull.* **43**, 436–448 (2017).
- Anticevic, A. et al. Early-course unmedicated schizophrenia patients exhibit elevated prefrontal connectivity associated with longitudinal change. *J. Neurosci.* **35**, 267–286 (2015).
- B Nejad, A., H Ebdrup, B., Y Glenthøj, B. & R Siebner, H. J. Cn Brain connectivity studies in schizophrenia: unravelling the effects of antipsychotics. *Curr. Neuropharmacol.* **10**, 219–230 (2012).
- Torres, U. S. et al. Patterns of regional gray matter loss at different stages of schizophrenia: a multisite, cross-sectional VBM study in first-episode and chronic illness. *NeuroImage* **12**, 1–15 (2016).
- Hulshoff Pol, H. E. & Kahn, R. S. What happens after the first episode? A review of progressive brain changes in chronically ill patients with schizophrenia. *Schizophr. Bull.* **34**, 354–366 (2008).
- Laursen, T. M. et al. Family history of psychiatric illness as a risk factor for schizoaffective disorder: a Danish register-based cohort study. *Arch. Gen. Psychiatry* **62**, 841–848 (2005).
- Guo, S. et al. Brain-wide functional dysconnectivity in schizophrenia: parsing diathesis, resilience, and the effects of clinical expression. *Can. J. Psychiatry* **65**, 21–29 (2020).
- Crow, T. J. Is schizophrenia the price that *Homo sapiens* pays for language? *Schizophr. Res.* **28**, 127–141 (1997).
- Crow, T. J. Schizophrenia as the price that *Homo sapiens* pays for language: a resolution of the central paradox in the origin of the species. *Brain Res. Rev.* **31**, 118–129 (2000).
- Palaniyappan, L. Inefficient neural system stabilization: a theory of spontaneous resolutions and recurrent relapses in psychosis. *J. Psychiatry Neurosci.* **44**, 367 (2019).
- Caspers, S. et al. Pathway-specific genetic risk for Alzheimer's disease differentiates regional patterns of cortical atrophy in older adults. *Cereb. Cortex* **30**, 801–811 (2020).
- Fazio, L. et al. Transcriptomic context of DRD1 is associated with prefrontal activity and behavior during working memory. *Proc. Natl Acad. Sci.* **115**, 5582–5587 (2018).
- Darst, B. F. et al. Pathway-specific polygenic risk scores as predictors of amyloid- $\beta$  deposition and cognitive function in a sample at increased risk for Alzheimer's disease. *J. Alzheimer's Dis.* **55**, 473–484 (2017).
- Ramos, A., Gonzalez, L. A. N., Weinberger, D. R., Mitchell, K. J. & Nicodemus, K. K. The role of polygenic risk score gene-set analysis in the context of the omnigenic model of schizophrenia. *Neuropsychopharmacology* **44**, 1562–1569 (2019).
- Pergola, G. et al. DRD2 co-expression network and a related polygenic index predict imaging, behavioral and clinical phenotypes linked to schizophrenia. *Transl. Psychiatry* **7**, e1006 (2017).
- Rolls, E. T., Joliot, M. & Tzourio-Mazoyer, N. Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *NeuroImage* **122**, 1–5 (2015).
- Cui, L.-B. et al. Putamen-related regional and network functional deficits in first-episode schizophrenia with auditory verbal hallucinations. *Schizophr. Res.* **173**, 13–22 (2016).
- Rolls, E. T. *Emotion and Decision Making Explained*. (Oxford University Press, 2013).
- Dominguez, M.-d.-G., Saka, M. C., Lieb, R., Wittchen, H.-U. & van Os, J. Early expression of negative/disorganized symptoms predicting psychotic experiences and subsequent clinical psychosis: a 10-year study. *Am. J. Psychiatry* **167**, 1075–1082 (2010).
- Demjaha, A., Valmaggia, L., Stahl, D., Byrne, M. & McGuire, P. J. Sb Disorganization/cognitive and negative symptom dimensions in the at-risk mental state predict subsequent transition to psychosis. *Schizophr. Bull.* **38**, 351–359 (2010).
- DeWitt, I. & Rauschecker, J. P. Phoneme and word recognition in the auditory ventral stream. *Proc. Natl Acad. Sci. USA* **109**, E505–E514 (2012).
- MacGregor, L. J., Pulvermüller, F., van Casteren, M. & Shtyrov, Y. Ultra-rapid access to words in the brain. *Nat. Commun.* **3**, 1–7 (2012).
- Srinivasan, S. et al. Genetic markers of human evolution are enriched in schizophrenia. *Biol. Psychiatry* **80**, 284–292 (2016).
- Boeckx, C. & Benítez-Burraco, A. Globularity and language-readiness: generating new predictions by expanding the set of genes of interest. *Front. Psychol.* **5**, 1324 (2014).
- Boeckx, C. A. & Benítez-Burraco, A. The shape of the human language-ready brain. *Front. Psychol.* **5**, 282 (2014).
- Benítez-Burraco, A. & Boeckx, C. Possible functional links among brain-and skull-related genes selected in modern humans. *Front. Psychol.* **6**, 794 (2015).
- Palaniyappan, L. Progressive cortical reorganisation: a framework for investigating structural changes in schizophrenia. *Neurosci. Biobehav. Rev.* **79**, 1–13 (2017).



35. Radanovic, M., Sousa, R. T. D., Valiengo, L., Gattaz, W. F. & Forlenza, O. V. J. A. D. N.-P. Formal thought disorder and language impairment in schizophrenia. *Arquivos de neuro-psiquiatria* **71**, 55–60 (2013).
36. Liégeois, F. et al. Language fMRI abnormalities associated with FOXP2 gene mutation. *Nat. Neurosci.* **6**, 1230 (2003).
37. Li, T. et al. FoxP2 is significantly associated with schizophrenia and major depression in the Chinese Han population. *World J. Biol. Psychiatry* **14**, 146–150 (2013).
38. Sanjuán, J. et al. Association between FOXP2 polymorphisms and schizophrenia with auditory hallucinations. *Psychiatr. Genet.* **16**, 67–72 (2006).
39. Tolosa, A. et al. FOXP2 gene and language impairment in schizophrenia: association and epigenetic studies. *BMC Med. Genet.* **11**, 114 (2010).
40. Rossell, S. et al. Is language impairment in schizophrenia related to language genes? *Eur. Neuropsychopharmacol.* **27**, S459–S460 (2017).
41. Hoogman, M. et al. Assessing the effects of common variation in the FOXP2 gene on human brain structure. *Front. Hum. Neurosci.* **8**, 473 (2014).
42. Faraone, S. V. & Larsson, H. Genetics of attention deficit hyperactivity disorder. *Mol. Psychiatry* **24**, 562–575 (2018).
43. Lai, C. S. L., Gerrelli, D., Monaco, A. P., Fisher, S. E. & Copp, A. J. FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. *Brain* **126**, 2455–2462 (2003).
44. Enard, W. et al. A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell* **137**, 961–971 (2009).
45. Ocklenburg, S. et al. FOXP2 variation modulates functional hemispheric asymmetries for speech perception. *Brain Lang.* **126**, 279–284 (2013).
46. Pinel, P. et al. Genetic variants of FOXP2 and KIAA0319/TTRAP/THEM2 locus are associated with altered brain activation in distinct language-related regions. *J. Neurosci.* **32**, 817–825 (2012).
47. Murphy, E. & Benítez-Burraco, A. Bridging the gap between genes and language deficits in schizophrenia: an oscillopathic approach. *Front. Hum. Neurosci.* **10**, 422 (2016).
48. Mei, L. & Nave, K.-A. J. N. Neuregulin-ERBB signaling in the nervous system and neuropsychiatric diseases. *Neuron* **83**, 27–49 (2014).
49. Hou, X.-J., Ni, K.-M., Yang, J.-M. & Li, X.-M. J. N. Neuregulin 1/ErbB4 enhances synchronized oscillations of prefrontal cortex neurons via inhibitory synapses. *Neuroscience* **261**, 107–117 (2014).
50. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition. (APA Publishing, Washington DC, USA, 2000).
51. Newton, R. et al. Diverse definitions of the early course of schizophrenia—a targeted literature review. *npj Schizophr.* **4**, 1–10 (2018).
52. Large, M., Farooq, S., Nielsen, O. & Slade, T. Duration of untreated psychosis in low and middle income economies: the relationship between GDP and DUP. *Br. J. Psychiatry* **193**, 272–278 (2008).
53. Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W. & Smith, S. M. Fsl. *NeuroImage* **62**, 782–790 (2012).
54. Cox, R. W. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29**, 162–173 (1996).
55. Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. & Turner, R. Movement-related effects in fMRI time-series. *Magn. Reson. Med.* **35**, 346–355 (1996).
56. Patel, A. X. et al. A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series. *NeuroImage* **95**, 287–304 (2014).
57. Wang, Q. et al. “Brain Connectivity Deviates by Sex and Hemisphere in the First Episode of Schizophrenia”—a route to the genetic basis of language and psychosis? *Schizophr. Bull.* **45**, 484–494 (2018).
58. Weber, J. A., Aldana, R., Gallagher, B. D. & Edwards, J. S. Sentieon DNA pipeline for variant detection—Software-only solution, over 20x faster than GATK 3.3 with identical results. Report No. 2167-9843, (PeerJ PrePrints, 2016).
59. Gong, W. et al. Statistical testing and power analysis for brain-wide association study. *Med. Image Anal.* **47**, 15–30 (2018).
60. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).
61. McGorry, P., Bell, R., Dudgeon, P. & Jackson, H. The dimensional structure of first episode psychosis: an exploratory factor analysis. *Psychol. Med.* **28**, 935–947 (1998).
62. Liddle, P. F. The core deficit of classical schizophrenia: implications for predicting the functional outcome of psychotic illness and developing effective treatments. *Can. J. Psychiatry* **64**, 680–685 (2019).
63. Ott, S. L., Roberts, S., Rock, D., Allen, J. & Erlenmeyer-Kimling, L. Positive and negative thought disorder and psychopathology in childhood among subjects with adulthood schizophrenia. *Schizophr. Res.* **58**, 231–239 (2002).
64. Kircher, T., Bröhl, H., Meier, F. & Engelen, J. Formal thought disorders: from phenomenology to neurobiology. *Lancet Psychiatry* **5**, 515–526 (2018).
65. Bleuler, E. Dementia Praecox or the group of schizophrenias. (International Universities Press, New York, NY, USA, 1950).
66. Network, T. et al. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat. Neurosci.* **18**, 199 (2015).
67. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. J. Nar ToppGene suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37**, W305–W311 (2009).
68. Owen, M. J., O’Donovan, M. C., Thapar, A. & Craddock, N. Neurodevelopmental hypothesis of schizophrenia. *Br. J. Psychiatry* **198**, 173–175 (2011).
69. Shen, E. H., Overly, C. C. & Jones, A. R. J. Tin The allen human brain atlas: comprehensive gene expression mapping of the human brain. *Trends Neurosci.* **35**, 711–714 (2012).
70. Spath, H. *The cluster dissection and analysis theory fortran programs examples*. (Prentice-Hall, Inc., 1985).
71. Deriziotis, P. & Fisher, S. E. Speech and language: translating the genome. *Trends Genet.* **33**, 642–656 (2017).
72. Fisher, S. E. & Scharff, C. FOXP2 as a molecular window into speech and language. *Trends Genet.* **25**, 166–177 (2009).
73. Miller, J. A. et al. Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014).
74. Radulescu, E. et al. Identification and prioritization of gene sets associated with schizophrenia risk by co-expression network analysis in human brain. *Mol. Psychiatry* **25**, 791–804 (2020).
75. Zhang, Y. et al. A joint study of whole exome sequencing and structural MRI analysis in major depressive disorder. *Psychol. Med.* **50**, 384–395 (2020).
76. Zhang, Y. et al. The genetic risk underlying synapse plasticity and neuroinflammation in Major Depressive Disorder (MDD): a joint Whole Exome Sequencing (WES) and structural MRI study. *Eur. Neuropsychopharmacol.* **29**, S847 (2019).

## ACKNOWLEDGEMENTS

The neuroimaging and genetic data used here is from ZIB Consortium. This work was supported by Shanghai Municipal Science and Technology Major Project (No.2018SHZDZX01) and ZJ Lab (to Z.J.); 2016YFC0906402 (to Z.J.); NSFC 61973086 (to Z.J.); Shanghai Science & Technology Innovation Plan key project Grant Nos. 15JC1400101 and 16JC1420402 (to J.F.); National Natural Science Foundation of China Grant Nos. 71661167002 (to J.F.), 91630314 (to J.F.); Shanghai AI Platform for Diagnosis and Treatment of Brain Diseases Grant No. 2016–17 (to J.F.); Base for Introducing Talents of Discipline to Universities Grant No. B18015 (to J.F.); L.P. was partly supported by Canadian Institute of Health Research Foundation Grant (grant number 375104); Opportunities fund and Innovation fund for Academic Medical Organization of Southwest Ontario; Bucke Family Fund, The Chrysalis Foundation and The Arcangelo Rea Family Foundation (London, Ontario). L.P. acknowledges salary support from the Tanna Schulich Chair of Neuroscience and Mental Health. L.P. reports personal fees from Otsuka Canada, Janssen Canada, SPMM Course Limited, UK, Canadian Psychiatric Association; book royalties from Oxford University Press; investigator-initiated educational grants from Janssen Canada, Sunovion and Otsuka Canada outside the submitted work. W.C. is supported by grants from the National Natural Sciences Foundation of China (No.81701773, 11771010), Sponsored by Shanghai Sailing Program (No. 17YF1426200) and is also sponsored by Natural Science Foundation of Shanghai (No. 18ZR1404400). J.W. is supported by NSFC 81971251 and is also sponsored by Science and Technology Commission of Shanghai Municipality (19411950800). All authors report no relevant conflicts.

## AUTHOR CONTRIBUTIONS

J.D., J.Z., and J.F. proposed the study. J.D., Z.L., W.G., and W.C. analyzed the data. J.D., L.P., J.Z., M.Z., J.W., and J.Z. considered how to analyze the data, and wrote and approved the manuscript. All authors revised and optimized further versions of the manuscript. All the authors have contributed to and have approved the final manuscript.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41537-021-00141-8>.

**Correspondence** and requests for materials should be addressed to J.Z. or J.F.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative

Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021