

## ORIGINAL ARTICLE

## Rare damaging variants in DNA repair and cell cycle pathways are associated with hippocampal and cognitive dysfunction: a combined genetic imaging study in first-episode treatment-naive patients with schizophrenia

Z Yang<sup>1,6</sup>, M Li<sup>1,2,6</sup>, X Hu<sup>3</sup>, B Xiang<sup>1</sup>, W Deng<sup>1,2</sup>, Q Wang<sup>1,2</sup>, Y Wang<sup>1</sup>, L Zhao<sup>1</sup>, X Ma<sup>1,2</sup>, PC Sham<sup>4</sup>, G Northoff<sup>5</sup> and T Li<sup>1,2</sup>

Schizophrenia is a complex neurodevelopmental disorder where changes in both hippocampus and memory-related cognitive functions are central. However, the exact relationship between neurodevelopmental-genetic factors and hippocampal-cognitive dysfunction remains unclear. The general aim of our study is to link the occurrence of rare damaging mutations involved in susceptibility gene pathways to the structure and function of hippocampus in order to define genetically and phenotypically based subgroups in schizophrenia. In the present study, by analyzing the exome sequencing and magnetic resonance imaging data in 94 first-episode treatment-naive schizophrenia patients and 134 normal controls, we identified that a cluster of rare damaging variants (RDVs) enriched in DNA repair and cell cycle pathways was present only in a subgroup including 39 schizophrenic patients. Furthermore, we found that schizophrenic patients with this RDVs show increased resting-state functional connectivity (rsFC) between left hippocampus (especially for left dentate gyrus) and left inferior parietal cortex, as well as decreased rsFC between left hippocampus and cerebellum. Moreover, abnormal rsFC was related to the deficits of spatial working memory (SWM; that is known to recruit the hippocampus) in patients with the RDVs. Taken together, our data demonstrate for the first time, to our knowledge, that damaging rare variants of genes in DNA repair and cell cycle pathways are associated with aberrant hippocampal rsFC, which was further relative to cognitive deficits in first-episode treatment-naive schizophrenia. Therefore, our data provide some evidence for the occurrence of phenotypic alterations in hippocampal and SWM function in a genetically defined subgroup of schizophrenia.

*Translational Psychiatry* (2017) **7**, e1028; doi:10.1038/tp.2016.291; published online 14 February 2017

## INTRODUCTION

Schizophrenia, which affects ~1% of the population, is a devastating mental disorder characterized by a range of symptoms including abnormal perception, thinking, cognitive, emotion and social behavior and so on. It is widely accepted that a complex interplay of genetic and environmental factors contributes to the etiology of schizophrenia,<sup>1</sup> which therefore is regarded a neurodevelopmental disorder.<sup>2,3</sup> During the last decades, researchers have made significant strides in teasing apart schizophrenia's convoluted genetic vulnerabilities. These included initial epidemiology studies within family, twin and adoption,<sup>4,5</sup> and then linkage studies mainly using microsatellite markers<sup>6</sup> to candidate genes and genome-wide association studies.<sup>1</sup>

Recently, genome-wide association studies identified that some common variants with small to moderate effects as well as rare but gene-disrupting copy number variants are significantly associated with schizophrenia.<sup>1–3,7</sup> Studies using the whole-exome sequencing technologies have also revealed that rare damaging nonsense and loss-of-function variants and/or *de novo* variants conferred higher risk to schizophrenia. However, to validate these variants in population remains challenging because

of the low frequency and effect sizes.<sup>8–13</sup> Converging evidence from studies of common and rare variants identified broad classes of genes and pathways involved in calcium ion channel, synaptic plasticity and neurodevelopment processes,<sup>9,12–14</sup> chromatin remodeling and synaptic network.<sup>12,15</sup> Although these results show significant leads for identifying susceptibility genes of schizophrenia, the exact and specific genetic factors contributing to the abnormal neurodevelopment processes in schizophrenia remain inconclusive.

A key brain region in schizophrenia is the hippocampus<sup>16,17</sup> where reduced volume in especially left (rather than right) hippocampus<sup>18</sup> as well as resting-state functional connectivity (rsFC) differences to other regions<sup>19,20</sup> have been observed. This corresponds to the observation of neuropsychological deficits, especially dysfunctional spatial working memory (SWM) and logical memory in patients with schizophrenia.<sup>21–23</sup> However, the exact linkage between hippocampus and memory deficits on the one hand and underlying genetic-neurodevelopmental changes on the other hand remain unclear.

The general aim of our study was to conduct multilevel genetic imaging cognitive investigation in a unique large sample of first-

<sup>1</sup>The State Key Laboratory of Biotherapy, Psychiatric Laboratory, West China Hospital, Sichuan University, Chengdu, Sichuan, China; <sup>2</sup>The Mental Health Center, West China Hospital, Sichuan University, Chengdu, Sichuan, China; <sup>3</sup>Biobank, West China Hospital, Sichuan University, Chengdu, Sichuan, China; <sup>4</sup>Centre for Genomic Sciences and Department of Psychiatry, University of Hong Kong, Pokfulam, China and <sup>5</sup>Institute of Mental Health Research, University of Ottawa, Ottawa, ON, Canada. Correspondence: Dr T Li, The Mental Health Center, West China Hospital, Sichuan University, Chengdu 610041, Sichuan, China.

E-mail: litaohx@scu.edu.cn

<sup>6</sup>These authors contributed equally to this work.

Received 2 November 2016; revised 13 November 2016; accepted 27 November 2016

episode treatment-naïve schizophrenic patients. For that purpose, we combined genetic analysis of neurodevelopmentally related genes with structural and functional imaging of the hippocampus and related cognitive measures. Specifically, genetic investigation in schizophrenic (and healthy) subjects served to determine specific genetic neurodevelopmentally relevant subtypes, which were then further characterized phenotypically using structural and functional hippocampal imaging as well as related cognitive, that is, memory measures.

The first specific aim was to unravel unique rare damaging variants (RDVs) in a unique sample of first-episode treatment-naïve schizophrenia patients relative to healthy controls by using whole-exome sequencing approach, and then apply Gene Ontology (GO) enrichment strategy to identify gene pathways. Consequently, we focused on the RDVs in DNA repair and cell cycle pathways for the following reasons. First, the majority of these genes in DNA repair and cell cycle pathways are most highly expressed in various brain regions, such as hippocampus;<sup>8–13</sup> previous studies suggested indeed that genes involved in the regulation of cell cycle and DNA repair significantly influence the hippocampus function in schizophrenia.<sup>24–28</sup> However, how the expression profile of these genes in cell cycle and DNA repair pathways map spatially and temporally during the critical neurodevelopmental stage in specifically the hippocampus in schizophrenia remains unclear.<sup>8–13</sup> We therefore hypothesized that we could distinguish two genetic subtypes in schizophrenia, one with RDVs and one without RDVs, as related to hippocampal developmental function.

On the basis of these considerations, the second specific aim consisted of measuring especially rsFC (as controlled for by structural volume) of the hippocampus in the same patients using brain magnetic resonance imaging (MRI), and to explore the relationship among genes, rsFC of the hippocampus and related neurocognitive functions. We hypothesized different hippocampal rsFC patterns in those schizophrenic patients with RDVs when compared with the ones without RDVs. To further underline hippocampal involvement, we included cognitive measures like spatial and logical memory that have been shown to be associated with specifically hippocampal function.<sup>29</sup> We hypothesized that schizophrenic patients with RDVs may have different memory-related cognitive functions relative to schizophrenic patients without RDVs.

There are several unique advantages in the present study. First, this is the first investigation, to our knowledge, to study the contribution of RDVs to the imaging and neurocognitive phenotype of schizophrenia. Second, we recruited first-episode treatment-naïve patients with schizophrenia to rule out the confounding factors of chronicity of the illness and treatment effects on neuroimaging and neurocognitive assessments. Third, extending limits on the power to identifying specific loci responsible for the disorder,<sup>30,31</sup> we improve the definition of the phenotype and/or reducing the phenotypic complexity of schizophrenia for genetic studies by combining the latter with cognitive (for example, neuropsychological) and neural (for example, neuroimaging measures) that are more closely related to the phenotype.

## MATERIALS AND METHODS

### Samples

There were total 234 participants including 97 first-episode treatment-naïve patients with schizophrenia and 137 healthy controls. Patients with schizophrenia were recruited at the Mental Health Centre of the West China Hospital, Sichuan University, China. Healthy volunteers were recruited from the community. The study was approved by the ethical committee in West China Hospital of Sichuan University. All participants were Han Chinese and provided written informed consent for their participation in this study.

### Clinical and memory assessments

All patients were interviewed by a trained psychiatrist using the Structured Clinical Interview for the DSM-IV (SCID).<sup>32</sup> DSM-IV criteria for schizophrenia were used for diagnosis. Those who were initially diagnosed with schizophrenia from psychosis due to the illness duration (less than 6 months) were followed up for at least 6 months to meet the DSM-IV criteria for schizophrenia. Psychopathology associated with schizophrenia was evaluated using the positive and negative syndrome scale.<sup>33</sup> Healthy controls were screened with the SCID-P non-patient version for the lifetime absence of psychiatric illnesses. Subjects with the existence of organic brain disorders, alcohol or drug abuse, pregnancy or any severe physical illness, such as brain tumor or epilepsy, were excluded from the study.

Schizophrenic patients and healthy controls completed SWM test in the Cambridge Neuropsychological Test Automated Battery (CANTAB; <http://www.cantab.com>) and immediate and delayed (30 min) logical memory subtest of the Wechsler Memory Scale,<sup>34</sup> respectively.

### MRI data acquisition

Overall, 74 healthy controls and 74 patients with schizophrenia underwent structural MRI scans on a Signa 3.0-T scanner (General Electric, Medical Systems, Milwaukee, WI, USA), whereas 65 healthy controls and 55 patients were scanned to obtain brain resting-state functional MRI in the Department of Radiology at West China Hospital. Detailed procedures of scanning are presented in Supplementary Information, sections 7 and 8.

### Sequencing and variant calling

All samples were sequenced using the TruSeq Exome Enrichment Kit (San Diego, CA, USA) optimized for IlluminaHiSeq2000 sequencing. The pipeline of raw data processing and variants calling is present in Supplementary Figure 2, which includes using Burrows-Wheeler Alignment tool<sup>35</sup> alignment reads to the reference human genome (hg19); Picard tools (<http://picard.sourceforge.net/>) to collect quality statistics and fix read group problem; GATK<sup>36</sup> for IN/DEL alignment; Samtools (<http://samtools.sourceforge.net/>) to filter out low-quality reads; and GATK to perform SNP and INDEL calling. Validation of selected variants was conducted by Sanger sequencing.

### Data analysis

We use PLINK<sup>37</sup> and KGGSeq (<http://statgenpro.psychiatry.hku.hk/limx/kggseq/doc/UserManual.html>)<sup>38</sup> to perform individual and variants' quality control. We use software KGGSeq to integrate databases to annotate the minor allele frequency of variants, including Hapmap and 1000 Genome. The cutoff frequency was set at 0.1% for rare variance. Methods implemented in PLINK/Seq (<http://atgu.mgh.harvard.edu/plinkseq/>) were employed for single site and gene-based association analysis. After filtering, we performed GO enrichment analysis in 2895 mutations within 2442 genes in cases and 4484 mutations within 3481 genes in controls using GeneMANIA (<http://www.genemania.org/>). The results of enrichment indicated that two GO (GO: 0006281 and GO:0007049) only present in cases. We identified more mutations in controls (that is, 2895 mutations within 2442 genes in schizophrenic patients and 4484 mutations within 3481 genes in healthy controls) mainly because, in present study, the sample size in controls (134) is bigger than cases (94). The bigger sample size in controls provides the higher chance to identify mutations. The weighted gene coexpression network analysis (WGCNA)<sup>39</sup> was used for expression profile, all of which was carried out by the WGCNA R package.<sup>40</sup> All of Brain expression data from different points in life were acquired from a published study (The Human Brain Transcriptome).<sup>41</sup> Sample size and statistical power were conducted by the statistical software, Exome Power Calculation (<http://darth.ssg.uab.edu:8080/epc/>).

rsFC of bilateral hippocampus and six hippocampal subregions was calculated by seed-to-voxel analysis using Statistical Parametric Mapping (SPM8, <http://www.fil.ion.ucl.ac.uk/spm>) and data-processing assistant for resting-state functional MRI from resting-state functional MRI (Supplementary material). The time courses averaged over all voxels of each hippocampus and six hippocampal subregions were extracted. Pearson's correlation coefficients ( $r$ ) between time courses of left/right hippocampus and all other voxels were calculated and transformed to Fisher's  $z$ -scores to derive rsFC maps. Thus, 67 349 pairs of functional connectivity were calculated between right hippocampus and the other voxels of the brain, whereas 67 370 pairs of functional connectivity were computed between left hippocampus and the other voxels of the brain. Statistical tests on the functional connectivity maps of hippocampus

between patients with schizophrenia and healthy controls were performed using analysis of covariance with sex, age and education years, and volume of hippocampus as covariates in SPSS. The significant threshold was set at  $P < 0.05$ , corrected for multiple comparisons based on Monte Carlo simulations. Subsequently, the mean Z-value of each cluster with a significant functional connectivity (FC) difference was extracted and were compared by analysis of variance followed by *post hoc* test (least significant difference) in SPSS 18.0, across patients with RDVs, patients without RDVs and healthy controls (SPSS, Chicago, IL, USA), significant level of P values were set at less than 0.05. Analysis of covariance followed by *post hoc* test were applied to test the differences of functional connectivity and memory functions among patients with RDVs, patients without RDVs and healthy controls. Partial correlation analysis was used to analyze the relationship between the functional connectivities of bilateral hippocampus (and six hippocampal subregions) and cognitive, that is, memory functions, with age, sex and education years (and structural hippocampal volume) as covariance (Supplementary material). Furthermore, Fisher's Z-transformation was used to transform the correlation coefficients (r value) to Z-score, and Z-test was used for testing the differences of correlations between groups. Corrections for multiple comparisons were applied as appropriate.

## RESULTS

**Genetic investigation of single nucleotide variants and In/Dels**  
In the current study, 234 subjects (97 patients with first-episode treatment-naive schizophrenia and 137 healthy controls) were collected for exome sequencing to identify those alleles, genes or gene networks that harbor rare coding variants of moderate or large effect on risk for schizophrenia (Supplementary Information, sections 1). Six samples were removed with low quality along with likely contamination after quality control and variant calling (Supplementary Information, sections 2 and 3). Quality control of all variants was conducted by in-house software KGGSeq.<sup>38</sup> The final data set comprised 94 schizophrenic cases and 134 controls. Detailed demographic and clinical information, as well as summary for sequencing quality, was shown in Table 1. There were no significant differences in age, sex distribution and years of education (Table 1). On average, we obtained 7.67 Gb of mappable sequence data per individual after exome enrichment, targeting ~62 Mb from exons and their flanking regions. In all, 99.75% of the reads were

properly aligned to the reference genome. Our median read depth is  $\sim 45\times$ , which is higher than the estimated average depth ( $33\times$ ) required for highly accurate downstream heterozygous variant detection. In addition, 88.60% of the captured target exons were covered by high-quality genotype calls at least 10 times to ensure good detection sensitivity.<sup>42</sup> Technical sequencing metrics, including total coverage, proportion of deeply covered targets and initial mapping reads, indicated no difference between schizophrenic patients and healthy controls (Table 2).

Importantly, we found that schizophrenic patients enriched significantly more nonsynonymous and coding variants than controls (Table 2,  $P < 0.001$ ). Allele counts between schizophrenic cases and healthy subjects did not show significant difference after adjustment (Supplementary Information, section 4). We performed two series of gene-based tests: a one-sided burden test of an increased rare allele rate in cases (<https://atgu.mgh.harvard.edu/plinkseq/>) and the CALPHA test.<sup>43</sup> Both tests indicated that gene NRK (Nik Related Kinase, NM\_198465, chrX: 105132399...105199499) had increased rare allele rate in cases ( $P \leq 1.0 \times 10^{-6}$ ). However, even according to the most stringent Bonferroni correction ( $0.05/20000 = 2.5 \times 10^{-6}$ , 20 000 are the total gene number; the results of NRK gene still present significant statistical difference.<sup>44</sup> Nine rare variants within NRK present in 13 schizophrenic patients, but not in healthy controls (Supplementary Table 1).

NRK is a protein-coding gene, which is associated with hypermobility syndrome, hyperinsulinemic and hypoglycemia. GO annotations related to this gene include protein serine/threonine kinase activity and small GTPase regulator activity. Previous studies did not denote any association between NRK and schizophrenia; however, an important paralog of this gene is TAOK2, which is essential for dendrite morphogenesis and has been associated with autism spectrum disorder.<sup>45</sup>

Next, we adopted an alternative strategy in which we studied RDVs. Previous reports have shown that RDVs have a higher likelihood of having a role in schizophrenia (Supplementary Information, section 5). After applying several filters (Supplementary Figure 1), 2895 mutations within 2442 genes in schizophrenic patients and 4484 mutations within 3481 genes in healthy controls were investigated with GO enrichment at the gene-based level using GeneMANIA. There were 505 and 952 pathway (or GO) terms enriched in cases and controls separately, but 56 RDVs in 42 genes within two pathways of DNA repair (GO: 0006281, 31 RDVs in 20 genes) and cell cycle (GO: 0007049, 25 RDVs in 22 genes) were present only in 39 schizophrenic patients (Supplementary Table 2 and Table 3), which also indicated a significant *q*-value. There are 60 control-only categories that were not statistically significant. In the present study, we focused on pathways damaged in cases only. We identified the developmental

**Table 1.** Statistic of demographic data for all samples

	Case (N = 94)	Control (N = 134)	P-value
Gender (male/female)	46/48	71/63	0.64
Age (mean $\pm$ s.d)	23.94 $\pm$ 6.92	24.39 $\pm$ 10.95	0.37
Education (mean $\pm$ s.d)	12.79 $\pm$ 3.15	13.50 $\pm$ 2.89	0.07

**Table 2.** Summary for sequencing quality

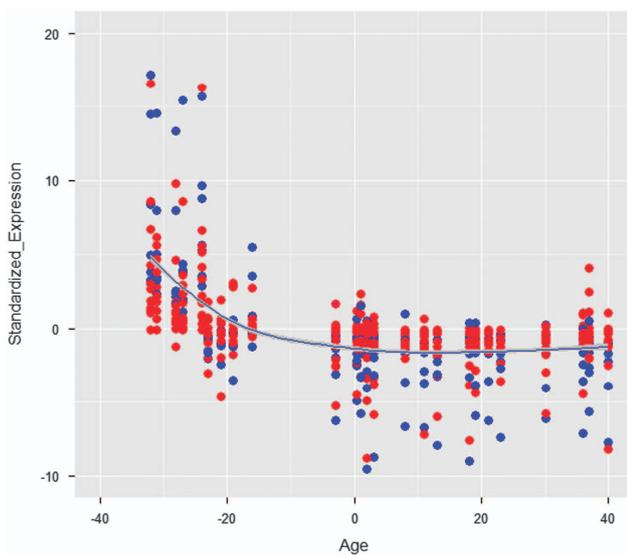
	Total (N = 228)	Case (N = 94)	Control (N = 134)	P-value
Total reads	75 918 287	76 078 777	75 805 704	0.45
Total yield (bp)	7 667 746 978	7 683 956 451	7 656 376 154	0.45
Initial mappable reads (%)	75 731 958 (99.75%)	75 897 153 (99.76%)	75 616 076 (99.75%)	0.44
1 $\times$ , Target genotypes (%)	58 570 040 (94.34%)	58 500 027 (94.23%)	58 619 153 (94.42%)	0.08
10 $\times$ , Target genotypes (%)	55 003 180 (88.60%)	54 965 986 (88.53%)	55 029 271 (88.64%)	0.33
Mean target depth	44.95	46.54	43.84	<b>0.009</b>
SNPs	74 413	74501	74351	0.25
Coding SNPs	20 229	20157	20280	<b>&lt; 0.001</b>
Synonymous SNPs	10 491	10457	10516	<b>0.001</b>
Nonsynonymous SNPs	9220	9185	9245	<b>&lt; 0.001</b>
Indels	7645	7664	7632	0.19
Coding indels	383	378	386	<b>&lt; 0.001</b>

Abbreviation: SNP, single-nucleotide polymorphism. Bold values show significant statistic difference.

**Table 3.** Aberrant functional connectivity of hippocampus and hippocampal subregions in patients group

	Brain region	Peak MNI (x y z)	Peak Z-score	Cluster size
Hippocampus_L	Inferior parietal lobule_L	-36 -56 43	6.8463	42
	Cerebellar posterior lobe	9 -75 -45	7.3691	40
<i>Hippocampal subregions</i>				
DG_L	Inferior parietal lobule_L	-36 -54 42	10.7382	77
DG_R	Inferior parietal lobule_L	-48 -54 48	8.2654	46
	Posterior cingulate cortex	-6 -33 30	7.0922	33
CA_L	Calcarine_R	24 -93 0	11.1746	68
CA_R	Calcarine_R	21 -93 3	13.0346	97
	Fusiform extent to lingual_L	-27 -54 -3	9.205	32
SUB_L	Medial frontal cortex	3 30 -12	7.1874	37

Abbreviations: CA, cornuammonis; DG, dentate gyrus; L, left; MNI, Montreal Neurological Institute; SC, subicular complex; R, right. Note: The statistical significance threshold for main effect for diagnosis of functional connectivity of each hippocampal subregions was set at  $P < 0.05$ , corrected for multiple comparisons based on Monte Carlo simulations.



**Figure 1.** The developmental expression profile of 22 genes in the cell cycle pathway and 20 genes in the DNA repair pathway in hippocampus. Following colors for different Gene Ontology (GO) categories were used: red for DNA repair and blue for cell cycle, and a smooth curve with confidence interval (gray range). The expression profiles show high expression during the stage of brain development before birth, with a sharp decrease in expression after birth.

expression profile of 22 genes in cell cycle pathway and 20 genes in DNA repair pathway (Supplementary Information, section 6), which show high expression during the stage of brain development before birth, with a sharp decrease in expression after birth in hippocampus. The expression profile obtained by WGCNA is presented in Figure 1.

Clinical characterization and hippocampal characterization in schizophrenic patients with or without RDVs

**Clinical characterization.** Comparisons here focus on schizophrenic and healthy subjects as well as on the two genetic-based subgroups with schizophrenia, that is, patients with RDVs and those without RDVs. There were no significant differences in age and sex among patients with RDVs, patients without RDVs and healthy controls (Table 4); however, the education years in patients without RDVs are significantly lower than those in patients with RDVs and controls. In addition, there were no significant differences in the

duration of illnesses and severity of clinical symptoms between the two genetic-based patient subgroups.

*rsFC of hippocampus*

**Functional connectivity of bilateral hippocampus:** Using analysis of covariance with sex, age, education and the hippocampal volume as covariates, we found that left hippocampus showed significant difference in rsFC with left inferior parietal cortex (Montreal Neurological Institute (MNI) atlas coordinates:  $x = -36$ ,  $y = -56$ ,  $z = 43$ ; voxels = 42), and the right cerebellar posterior lobe (MNI coordinate:  $x = 9$ ,  $y = -75$ ,  $z = -45$ ; voxels = 40) between the three groups at  $P < 0.05$  (corrected for multiple comparisons based on Monte Carlo simulations; Figure 2a). *Post hoc* test indicated that, compared with healthy controls and patients without RDVs, patients with RDVs demonstrated increased rsFC between left hippocampus and left inferior parietal cortex, as well as decreased rsFC between left hippocampus and right cerebellum posterior lobe. In contrast, there was no significantly different rsFC between healthy controls and patients without RDVs (Figure 2b). Unlike in the left hippocampus, there were no significant rsFC differences in the right hippocampus with other brain regions among the three groups.

**rsFC of the six hippocampal subregions:** Analysis of covariance showed significant differences in rsFC between the left dentate gyrus (DG) and left inferior parietal cortex (MNI coordinate:  $x = -36$ ,  $y = -54$ ,  $z = 42$ ; voxels = 77), between right DG and left inferior parietal cortex (MNI coordinate:  $x = -48$ ,  $y = -54$ ,  $z = 48$ ; voxels = 46) as well as posterior cingulate cortex (MNI coordinate:  $x = -6$ ,  $y = -33$ ,  $z = 30$ ; voxels = 33), between left CA and right calcarine (MNI coordinate:  $x = 24$ ,  $y = -93$ ,  $z = 0$ ; voxels = 68), between right CA and right calcarine (MNI coordinate:  $x = 21$ ,  $y = -93$ ,  $z = 3$ ; voxels = 68) as well as left fusiform (MNI coordinate:  $x = -27$ ,  $y = -54$ ,  $z = -3$ ; voxels = 32), between left SC and orbital medial frontal cortex (MNI coordinate:  $x = 3$ ,  $y = 30$ ,  $z = -12$ ; voxels = 37) among the three groups at  $P < 0.05$  (corrected for multiple comparisons based on Monte Carlo simulations).

*Post hoc* tests indicated that increased rsFC between left DG and left inferior parietal cortex was only found in patients with RDVs, whereas increased rsFC between right DG and left inferior parietal cortex as well as posterior cingulum was detected in both patients group. rsFC of CA was significantly decreased in patients without RDVs and a trend decreased in patients with RDVs. Decreased rsFC of left SC and orbital medial frontal cortex was found only in patients without RDVs (Table 3, Figure 3).

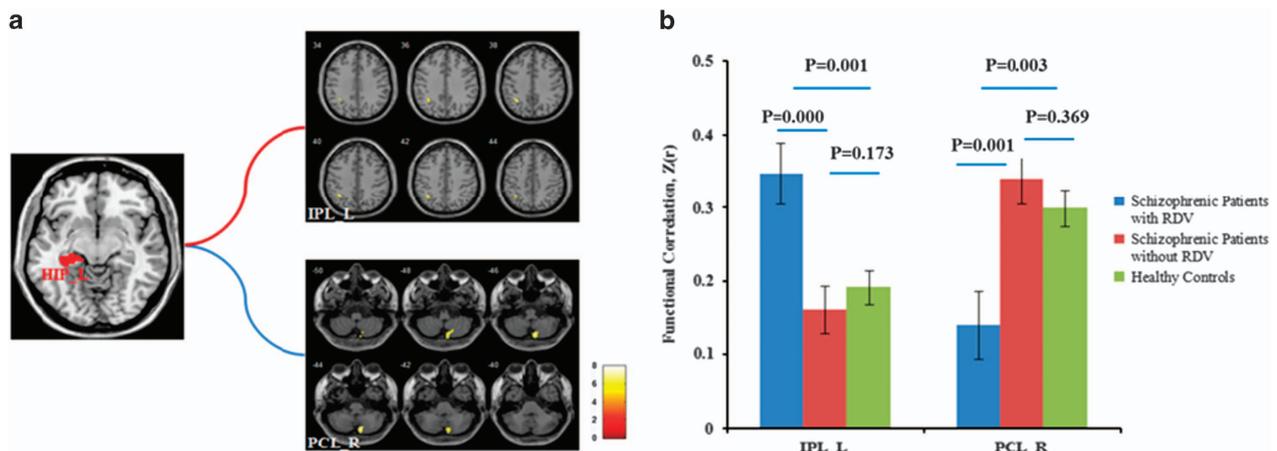
Relationships between hippocampal rsFC and memory

Compared with healthy controls, both patient groups showed significant impairments in SWM as well as in immediate and

**Table 4.** Sample summary and test scores for imaging

Category (mean $\pm$ s.d.)	Cases with RDVs (n = 26)	Cases without RDVs (n = 48)	Control (n = 74)	df	F/t/ $\lambda$ /value	P-value
Age (years)	25.69 $\pm$ 7.2	23.00 $\pm$ 6.92	23.23 $\pm$ 6.52	2	1.54	0.22
Gender (male/female)	13/13	24/24	41/33	2	0.43	0.81
Years of education	13.42 $\pm$ 3.19	11.90 $\pm$ 3.08	13.44 $\pm$ 3.24	2	3.72	0.03
Duration of disease	3.5	2		72	0.57	0.57
PANSS total	87.08 $\pm$ 17.16	88.06 $\pm$ 12.55		71	-0.28	0.78
PANSS positive	23.65 $\pm$ 8.27	24.34 $\pm$ 5.55		37.72	-0.38	0.71
PANSS negative	18.69 $\pm$ 8.50	18.87 $\pm$ 6.75		42.68	-0.09	0.93
PANSS general psychopathology	44.73 $\pm$ 8.84	44.85 $\pm$ 7.47		71	-0.06	0.95
Logical memory immediate	6.16 $\pm$ 3.31	7.84 $\pm$ 4.62	12.94 $\pm$ 3.89	2	45.47	< 0.001
Logical memory delay	4.09 $\pm$ 3.19	5.65 $\pm$ 4.39	11.02 $\pm$ 4.12	2	44.51	< 0.001
Hippocampus L milliliter	2534.70 $\pm$ 291.31	2512.42 $\pm$ 239.22	2722.55 $\pm$ 301.57	2	7.42	0.001
Hippocampus R milliliter	2573.12 $\pm$ 305.71	2559.78 $\pm$ 262.83	2741.44 $\pm$ 289.61	2	5.63	0.004

Abbreviations: L, left; PANSS, Positive and Negative Syndrome Scale; R, right; RDV, rare damage variant. Note: values are mean (s.d.).



**Figure 2.** Aberrant resting-state functional connectivity (rsFC) of the left hippocampus among three groups. (a) Left inferior parietal cortex and right cerebellum posterior lobe showed significantly different rsFC with left hippocampus among three groups. The statistical significance threshold was set at  $P < 0.05$  and corrected for multiple comparisons based on Monte Carlo simulations, for main effect for diagnosis of functional connectivity (FC) of left hippocampus. (b) ANCOVA with sex, age, education and the hippocampal volume as covariates. rsFC between left hippocampus and left inferior parietal cortex is increased, whereas rsFC between left hippocampus and cerebellum is decreased only in schizophrenic patients with rare damage variants (RDVs). The statistical significance threshold was set at  $P < 0.05$ , *post hoc* test by least significant difference. ANCOVA, analysis of covariance; HIP, hippocampus; IPL, inferior parietal cortex; L, left; PCL, cerebellum posterior lobe; R, right. Notes: s.e. for all figures.

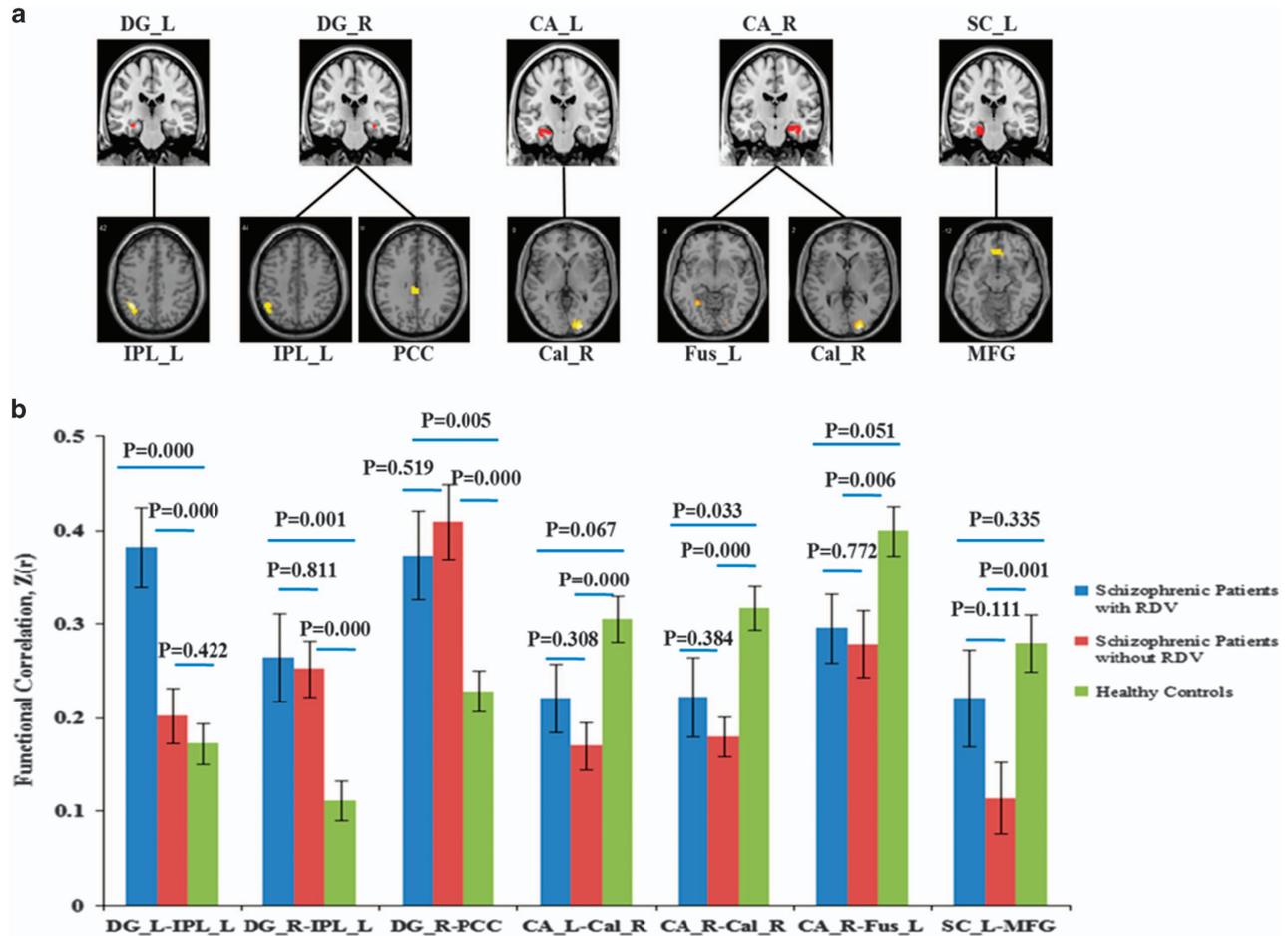
delayed logical memory (Supplementary Figure 4). No significant differences in these measures were observed between the two genetic-based subgroups in schizophrenic patients, that is, with and without RDVs (Supplementary Figure 4). Despite the lack of memory-related differences between the two genetic subgroups, we nevertheless observed different correlation patterns between rsFC and cognitive deficits. Errors in SWM were positively correlated to rsFC between left DG and left inferior parietal cortex in schizophrenic patients with RDVs ( $r = 0.670$ , uncorrected  $P = 0.034$ ), whereas they were negatively correlated to rsFC between left DG and left inferior parietal cortex in healthy controls ( $r = -0.373$ , uncorrected  $P = 0.032$ ; Figure 4). However, after correction for multiple tests, none of correlations above remains significant statistically, probably because of the small sample size. No correlation was observed in schizophrenic patients without RDV ( $r = -0.102$ , uncorrected  $P = 0.739$ ).

We then compared the correlation coefficient between each of the two groups by using Z-test. We found that the difference in correlations was statistically significant between schizophrenic patients with RDVs and healthy controls ( $Z = 3.331$ ,  $P = 8.65 \times 10^{-4}$ ), as well as between patients with RDVs and those without RDVs ( $Z = 2.171$ ,  $P = 0.030$ ) after Bonferroni correction. No

significant difference was found between patients without RDVs and healthy controls ( $Z = 0.884$ ,  $P = 0.377$ ).

## DISCUSSION

Investigating a unique large sample of first-episode treatment-naive schizophrenic patients, we first identified that 56 RDVs in 42 genes within two pathways implicated in DNA repair and cell cycle were present only in a subgroup including 39 schizophrenic patients, whereas the remaining schizophrenic patients (that is, those without RDVs) as well as healthy subjects did not show these RDVs. Further analysis revealed that these genes are highly expressed during the stage of brain development before birth, with a sharp decreased expression after birth in specifically the hippocampus. Second, we identified a unique hippocampus rsFC pattern in schizophrenic patients with RDVs, that is, increased rsFC between left hippocampus (especially for left dentate gyrus) and left inferior parietal cortex, as well as decreased rsFC between the left hippocampus and cerebellum. Subsequently, we found a tentative significant correlation of altered hippocampal rsFC with spatial memory deficits in schizophrenic patients with RDVs.



**Figure 3.** Aberrant resting-state functional connectivity (rsFC) of hippocampal subregions. (a) The brain map of aberrant rsFC with hippocampal subregions among three groups. The statistical significance threshold was set at  $P < 0.05$ , corrected for multiple comparisons based on Monte Carlo simulations, for main effect for diagnosis of functional connectivity (FC) of the different hippocampal subregions. (b) rsFC between left DG and left IPL increased only in schizophrenic patients with rare damage variants, whereas decreased rsFC of left SC with MFG was only found in schizophrenic patients without rare damage variants. The statistical significance threshold was set at  $P < 0.05$ , *post hoc* test by least significant difference. CA, cornuammonis; Cal, calcarine; DG, dentate gyrus; Fus, fusiform; IPL, inferior parietal cortex; L, left; MFG, medial frontal cortex; PCC, posterior cingulate cortex; R, right; RDV, rare damaging variant; SC, subicular complex. Notes: s.e. for all figures.

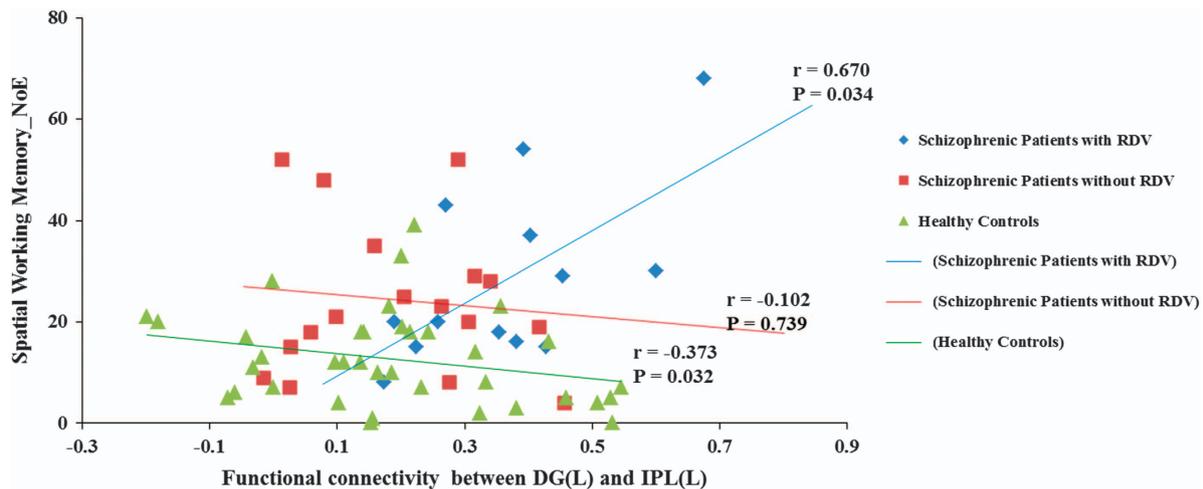
Schizophrenic patients harbor more than threefolds rare DNA mutations, especially in those with early onset of illness.<sup>14</sup> More *de novo* mutations were found recently at genomic hotspots, including chromosomes 1q21.1, 15q13.3, 16p13.1 and 22q11.2.<sup>46</sup> Studies in larger sample sets demonstrated a polygenic burden that increases the risk for schizophrenia, those genes primarily comprising many ultrarare nonsense mutations distributed across many genes, which mainly involved in neurodevelopment pathways.<sup>13</sup> Considering the low frequency and larger effect sizes of rare DNA mutations in current studies, we focused here on the case-unique RDVs, and found that these variants enriched in two pathways, for example, cell cycle regulation and DNA repair. Our finding in current study supported that disturbances of cell cycle regulation and DNA repair in post-mitotic neurons have been implicated in development of psychotic disorders.<sup>25,47-49</sup>

Importantly, most of the genes enrichment in these two pathways was detected to be highly expressed in the hippocampus in the fetal stage, with the expression level sharply decreasing after birth. On the basis of this finding, we could hypothesize that individuals possessing mutations in these two particular pathways may be more prone to develop schizophrenia later on because of the critical gene expressions during their fetal development. More interestingly, such hypothesis genetically specifies and conforms well with the neurodevelopmental model that emphasizes insults

as early as late-first or early-second trimester as central for pathological activation of neural circuits in schizophrenia.<sup>14,50,51</sup>

Our finding of the RDVs in these two pathways hints upon alterations in cell cycle regulation and the DNA repair in schizophrenic patients during embryogenesis, although the underlying mechanisms are still in explicit. Study from Katsel *et al.* suggested abnormal patterns of cell cycle gene and protein expression in schizophrenia, which may contribute to the oligodendroglial deficits observed in schizophrenia.<sup>52</sup> This could lead to changes in brain development and make the brain more susceptible to environmental risk factors, with downstream effects on neural progenitor proliferation and differentiation.<sup>53</sup>

In the present study, we detected aberrant functional connectivity of left hippocampus with inferior parietal lobe, posterior cingulate cortex, visual cortex and medial frontal gyrus in schizophrenic patients with RDVs. In contrast, schizophrenic patients without RDVs did not show such pattern. This suggests that the aberrant resting-state functional connections of left hippocampus in schizophrenia might likely be of neurodevelopmental origin (rather than of neurodegenerative origin or affected by antipsychotics with the latter being excluded here anyway because of the fact that our sample of treatment-naïve). Most interesting, we found that increased functional connectivity of left hippocampus (especially for the dentate gyrus) with left inferior



**Figure 4.** Relationships between resting-state functional connectivity (rsFC) of hippocampus and spatial working memory. Increased rsFC between left dentate gyrus and left inferior parietal cortex was positively related to number of errors of spatial working memory in schizophrenic patients with rare damaging variants and negatively related to number of errors of spatial working memory in healthy controls, uncorrected  $P < 0.05$ . The difference in the degree of correlation was statistically significant between schizophrenic patients with RDVs and healthy controls ( $Z = 3.331$ ,  $P = 8.65 \times 10^{-4}$ ) after the Bonferroni correction, and a trend toward significance between patients with RDVs and patients without RDVs ( $Z = 2.171$ ,  $P = 0.030$  before Bonferroni correction), but not significant between patients without RDVs and healthy controls ( $Z = 0.884$ ,  $P = 0.377$ ). DG, dentate gyrus; IPL, inferior parietal cortex; NoE, number of error; L, left; RDV, rare damaging variant. Note: partial correlation analysis was used to analyze the relationship between rsFC of hippocampus and spatial working memory with age, sex and educations and structural hippocampal volume as covariates.

parietal cortex and decreased FC between left hippocampus and cerebellum were only found in schizophrenic patients with RDVs. This further underlines the neurodevelopmental impact of RDV on phenotypic makers like the left hippocampal rsFC and its high relevance for schizophrenia. The exact mechanisms mediating the transition from early prenatal RDVs' occurrence to hippocampal rsFC abnormalities during the later outbreak of schizophrenia remain unclear though.

Increased functional connectivity between left dentate gyrus and left inferior parietal cortex was correlated with more errors of SWM in schizophrenic patients with RDVs, whereas no such correlation was found in schizophrenic patients without RDVs (and opposite, that is, negative) and in healthy subjects. The dentate gyrus has been regarded as one of a few brain structures owning high rates of adult neurogenesis and has a critical role in resolving new memories and spatial memory.<sup>54–56</sup> A recent study demonstrated that the inferior parietal cortex is involved in spatial perception and spatial orientation in particular and spatial functions in general (the 'where').<sup>52</sup> In addition, the inferior parietal cortex is also the target of output from hippocampus.<sup>57</sup> The findings in the present study thus provided direct evidence that aberrant increased FC of left dentate gyrus with inferior parietal cortex might contribute to the severity of SWM deficits in schizophrenic patients with RDVs. Above all, our results provide a link among RDVs, dentate gyrus dysfunctional resting-state connectivity and SWM deficits in a genetic-based neurodevelopmental subgroup of patients with schizophrenia.

However, several limitations must be born in mind when interpreting our results. First, the sample size of the present study is relatively small for genetic study, which might be lack of the power to detect significant association signal between cases and controls. However, we adopt an alternative strategy for filtering rare variants and then enrich to find two special pathways. Second, not all the subjects underwent the memory test in both the two-patient groups, which could have affected the results of correlation analysis between imaging and cognitive data. Fortunately, there were no significant differences in clinical profiles between patients with RDVs who underwent cognitive tests and those who did not within both groups. Third, considering the

small sample size of patients who underwent the cognitive tasks, the correlation analysis between neuroimaging and neurocognition remained uncorrected for multiple comparisons, which might lead to false-positive results.

In sum, to our knowledge, our findings demonstrate for the first time the presence of genes implicated in DNA repair and cell cycle pathways and related specifically to hippocampal development in a subgroup of first-episode treatment-naive schizophrenic patients. Importantly, the presence of these genes directly had an impact on or modulated phenotypic expression as hippocampal rsFC and related spatial memory function in this subgroup. Hence, our findings bridge the gap from genes regulating hippocampal development before birth to corresponding phenotypical markers like rsFC of hippocampus and associated cognitive, that is, SWM function at the outbreak of first-episode schizophrenia.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGMENTS

This work was partly funded by National Nature Science Foundation of China Key Project 81130024 and 81630030 (to TL); National Natural Science Foundation of China/Research Grants Council of Hong Kong Joint Research Scheme 81461168029 (to TL and Pak C Sham).

## REFERENCES

- 1 Tsuang MT, Stone WS, Faraone SV. Genes, environment and schizophrenia. *Br J Psychiatry* 2001; **178**: s18–s24.
- 2 Weinberger DR. Implications of normal brain development for the pathogenesis of schizophrenia. *Arch Gen Psychiatry* 1987; **44**: 660–669.
- 3 Murray RM, Lewis SW. Is schizophrenia a neurodevelopmental disorder? *Br Med J* 1987; **295**: 681–682.
- 4 Coryell W, Zimmerman M. The heritability of schizophrenia and schizoaffective disorder: a family study. *Arch Gen Psychiatry* 1988; **45**: 323–327.
- 5 Kendler KS, McGuire M, Gruenberg AM, O'Hare A, Spellman M, Walsh D. The Roscommon family study: I. Methods, diagnosis of probands, and risk of schizophrenia in relatives. *Arch Gen Psychiatry* 1993; **50**: 527–540.

- 6 McGue M, Gottesman II. The genetic epidemiology of schizophrenia and the design of linkage studies. *Eur Arch Psychiatr Clin Neurosci* 1991; **240**: 174–181.
- 7 Ripke S, Neale BM, Corvin A, Walters JT, Farh K-H, Holmans PA et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**: 421.
- 8 Gulsuner S, Walsh T, Watts AC, Lee MK, Thornton AM, Casadei S et al. Spatial and temporal mapping of *de novo* mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* 2013; **154**: 518–529.
- 9 Kirov G, Pocklington A, Holmans P, Ivanov D, Ikeda M, Ruderfer D et al. *De novo* CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol Psychiatry* 2011; **17**: 142–153.
- 10 Malhotra D, McCarthy S, Michaelson JJ, Vacic V, Burdick KE, Yoon S et al. High frequencies of *de novo* CNVs in bipolar disorder and schizophrenia. *Neuron* 2011; **72**: 951–963.
- 11 Xu B, Ionita-Laza I, Roos JL, Boone B, Woodrick S, Sun Y et al. *De novo* gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat Genet* 2012; **44**: 1365–1369.
- 12 Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P et al. *De novo* mutations in schizophrenia implicate synaptic networks. *Nature* 2014; **506**: 179–184.
- 13 Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P et al. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* 2014; **506**: 185–190.
- 14 Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 2008; **320**: 539–543.
- 15 McCarthy S, Gillis J, Kramer M, Lihm J, Yoon S, Berstein Y et al. *De novo* mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. *Mol Psychiatry* 2014; **19**: 652–658.
- 16 Tamminga CA, Stan AD, Wagner AD. The hippocampal formation in schizophrenia. *Am J Psychiatry* 2010; **167**: 1178–1193.
- 17 Heckers S. Neuroimaging studies of the hippocampus in schizophrenia. *Hippocampus* 2001; **11**: 520–528.
- 18 Harrison PJ. The hippocampus in schizophrenia: a review of the neuropathological evidence and its pathophysiological implications. *Psychopharmacology* 2004; **174**: 151–162.
- 19 Zhou Y, Shu N, Liu Y, Song M, Hao Y, Liu H et al. Altered resting-state functional connectivity and anatomical connectivity of hippocampus in schizophrenia. *Schizophr Res* 2008; **100**: 120–132.
- 20 Benetti S, Mechelli A, Picchioni M, Broome M, Williams S, McGuire P. Functional integration between the posterior hippocampus and prefrontal cortex is impaired in both first episode schizophrenia and the at risk mental state. *Brain* 2009; **132**(Pt 9): 2426–2436.
- 21 Li Z, Li Y, Chen L, Chen P, Hu Y. Prevalence of depression in patients with hypertension: a systematic review and meta-analysis. *Medicine* 2015; **94**: e1317.
- 22 Ma X, Wang Q, Sham PC, Liu X, Rabe-Hesketh S, Sun X et al. Neurocognitive deficits in first-episode schizophrenic patients and their first-degree relatives. *Am J Med Genet B Neuropsychiatr* 2007; **144B**: 407–416.
- 23 Wang Q, Vassos E, Deng W, Ma X, Hu X, Murray RM et al. Factor structures of the neurocognitive assessments and familial analysis in first-episode schizophrenia patients, their relatives and controls. *Aust N Z J Psychiatry* 2010; **44**: 109–119.
- 24 Magin GK, Robison SH, Breslin N, Wyatt RJ, Alexander RC. DNA repair and mutant frequency in schizophrenia. *Mutat Res* 1991; **255**: 241–246.
- 25 Saadat M, Pakyari N, Farrashbandi H. Genetic polymorphism in the DNA repair gene XRCC1 and susceptibility to schizophrenia. *Psychiatry Res* 2008; **157**: 241–245.
- 26 Katsel P, Davis KL, Li C, Tan W, Greenstein E, Kleiner Hoffman LB et al. Abnormal indices of cell cycle activity in schizophrenia and their potential association with oligodendrocytes. *Neuropsychopharmacology* 2008; **33**: 2993–3009.
- 27 Engmann O, Hortobagyi T, Pidsley R, Troakes C, Bernstein HG, Kreutz MR et al. Schizophrenia is associated with dysregulation of a Cdk5 activator that regulates synaptic protein expression and cognition. *Brain* 2011; **134**(Pt 8): 2408–2421.
- 28 Nagy Z. Cell cycle regulatory failure in neurones: causes and consequences. *Neurobiol Aging* 2000; **21**: 761–769.
- 29 Burgess N, Maguire EA, O'Keefe J. The human hippocampus and spatial and episodic memory. *Neuron* 2002; **35**: 625–641.
- 30 Belmaker R, Agam G. Major depressive disorder. *N Engl J Med* 2008; **358**: 55–68.
- 31 Mata DA, Ramos MA, Bansal N, Khan R, Guille C, Di Angelantonio E et al. Prevalence of depression and depressive symptoms among resident physicians: a systematic review and meta-analysis. *JAMA* 2015; **314**: 2373–2383.
- 32 First MB, Spitzer RL, Gibbon M, Williams JBW. *Structured Clinical Interview for DSM-IV Axis I Disorders, Clinician Version (SCID-CV)*. Amer Psychiatric Press Inc: Washington, DC, 1996.
- 33 Kay SR FA, Opler LA. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophr Bull* 1987; **13**: 261–276.
- 34 Gong Y. *Wechsler Memory Scale-Revised in China*. Hunan Medical School Press: Hunan, 1987.
- 35 Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009; **25**: 1754–1760.
- 36 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297–1303.
- 37 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
- 38 Li M-X, Gui H-S, Kwan JS, Bao S-Y, Sham PC. A comprehensive framework for prioritizing variants in exome sequencing studies of Mendelian diseases. *Nucleic Acids Res* 2012; **40**: e53–e53.
- 39 Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 2005; **4**: 1128.
- 40 Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; **9**: 559.
- 41 Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M et al. Spatio-temporal transcriptome of the human brain. *Nature* 2011; **478**: 483–489.
- 42 Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 2008; **456**: 53–59.
- 43 Neale BM, Rivas MA, Voight BF, Altshuler D, Devlin B, Orho-Melander M et al. Testing for an unusual distribution of rare variants. *PLoS Genet* 2011; **7**: e1001322.
- 44 Liu JZ, Mcrae AF, Nyholt DR, Medland SE, Wray NR, Brown KM et al. A versatile gene-based test for genome-wide association studies. *Am J Hum Genet* 2010; **87**: 139–145.
- 45 de Anda FC, Rosario AL, Durak O, Tran T, Gräff J, Meletis K et al. Autism spectrum disorder susceptibility gene TAOX2 affects basal dendrite formation in the neocortex. *Nat Neurosci* 2012; **15**: 1022–1031.
- 46 Stefansson H, Rujescu D, Cichon S, Pietiläinen OP, Ingason A, Steinberg S et al. Large recurrent microdeletions associated with schizophrenia. *Nature* 2008; **455**: 232–236.
- 47 Benes FM. Relationship of GAD(67) regulation to cell cycle and DNA repair in GABA neurons in the adult hippocampus: bipolar disorder versus schizophrenia. *Cell Cycle* 2010; **9**: 625–627.
- 48 Benes FM. Regulation of cell cycle and DNA repair in post-mitotic GABA neurons in psychotic disorders. *Neuropharmacology* 2011; **60**: 1232–1242.
- 49 Benes FM, Lim B, Subburaju S. Site-specific regulation of cell cycle and DNA repair in post-mitotic GABA cells in schizophrenic versus bipolars. *Proc Natl Acad Sci USA* 2009; **106**: 11731–11736.
- 50 Rappoport JL, Addington AM, Frangou S, Psych M. The neurodevelopmental model of schizophrenia: update 2005. *Mol Psychiatry* 2005; **10**: 434–449.
- 51 Owen MJ, O'Donovan MC, Thapar A, Craddock N. Neurodevelopmental hypothesis of schizophrenia. *Br J Psychiatry* 2011; **198**: 173–175.
- 52 Andersen RA. Inferior parietal lobule function in spatial perception and visuo-motor integration. *Comprehensive Physiol* 2011, Supplement 5: Handbook of Physiology, The Nervous System, Higher Functions of the Brain, pp 483–518.
- 53 Katsel P, Davis KL, Li C, Tan W, Greenstein E, Hoffman LBK et al. Abnormal indices of cell cycle activity in schizophrenia and their potential association with oligodendrocytes. *Neuropsychopharmacology* 2008; **33**: 2993–3009.
- 54 Cameron HA, McKay RD. Adult neurogenesis produces a large pool of new granule cells in the dentate gyrus. *J Comp Neurol* 2001; **435**: 406–417.
- 55 Aimone JB, Deng W, Gage FH. Resolving new memories: a critical look at the dentate gyrus, adult neurogenesis, and pattern separation. *Neuron* 2011; **70**: 589–596.
- 56 Snyder JS, Cameron HA. Could adult hippocampal neurogenesis be relevant for human behavior? *Behav Brain Res* 2012; **227**: 384–390.
- 57 Clower DM, West RA, Lynch JC, Strick PL. The inferior parietal lobule is the target of output from the superior colliculus, hippocampus, and cerebellum. *J Neurosci* 2001; **21**: 6283–6291.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017

Supplemental Information accompanies the paper on the Translational Psychiatry website (<http://www.nature.com/tp>)