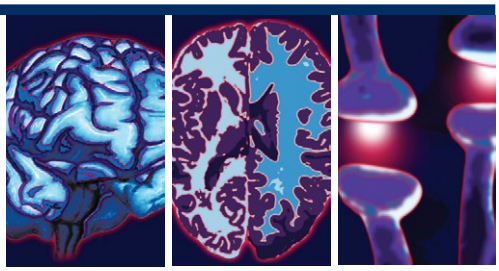
Review



Identification of Risk-Conferring Genes of Schizophrenia Using Endophenotypes

Hongyan Ren[1,3,4], Mingli Li[1,2], Qiang Wang[1,2,†], Tao Li[1,2]

Abstract

Schizophrenia is a complex disease predisposed by genes, environment and their interaction. Its diagnosis is mainly based on clinic observations with its treatment following a trial-and-error manner. Both linkage studies and following association studies have identified some genes and genomic regions which gain insight into pathophysiological foundation of the disease. However, low replication rate of detected variation between/within different populations prevented these findings from clinical application in the diagnosis and the precise treatment of disease. With introduction of endophenotypes and extended endpphenotypes, multiple disease-related traits such as neurocognitive deficits and neuroimaging alterations enable a further refinement of phenotypes used in genetic and genomic studies of schizophrenia. Here, the authors discuss the several endophenotypes emerging from their previous studies and the methods which could incorporate both the dichotomous variable of diagnosis and quantitative traits into genetic/genomic studies of schizophrenia. Furthermore, the authors demonstrated some alternative methodologies utilizing the big data generated from recent multi-sites cohort studies.

Keywords:

Schizophrenia, GWAS, Endophenotype, Extended-endophenotypes, Multivariates

Introduction

Schizophrenia is a chronically disabling illness affecting 1% of the global population. And it is identified as a complex illness affected by multiple genetic variants of weak effect and gene-environment interactions (G*E) including early childhood and maternal stress as well as viral infection [1,2]. Schizophrenia, similar to other complex diseases, is subject to two hypothesis: "Common Disease, Common Variant (CDCV)", where the most genetic risk for common, complex diseases is due to genetic loci with a common population frequency (>0.01) [3], and "Common Disease, Rare Variant (CDRV)", which hypothesizes that the specific genetic variations causing a disease with a common prevalence (a prevalence greater than 1–5%) are not necessarily found to be common in the population as suggested by the CDCV but rather are comprised of a multiplicity of risk alleles, each of which is individually rare in the population [4,5].

With advances in genotyping and sequencing technologies and decreasing of the price for genome-wide scan, the genome-wide scan or whole-genome sequencing is currently conducted

[1]Mental Health Center and Psychiatric Laboratory, State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu, Sichuan, PR China

[2]Brain Research Center, West China Hospital, Sichuan University, Chengdu, Sichuan, PR China

[3]Psychiatry and Medical Genetics, University of Alberta, Edmonton, AB, Canada

[4]Affiliated Brain Hospital of Guangzhou Medical University, Guangzhou Brain Hospital, Guangzhou, Guangdong, PR China

[†]Author for correspondence: Dr. Qiang Wang, Psychiatric Laboratory and Mental Health Center, State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu, Sichuan, PR China. Phone: +86 18980605803, email: wangqiang130@scu.edu.cn

in an increasingly large sample size. For the past decade, genome-wide association studies (GWAS) of schizophrenia have been carried out with large sample sizes in different populations, shedding novel light on the pathogenesis of the disease and potential treatment agents. For example, 108 loci and genomic regions were recently identified by PGC in a GWAS involving 37,000 cases and 113,000 controls [6]. These new findings provide additional evidence that schizophrenia, instead of oligogenic, is a polygenic disease putatively involving gene networks or pathways related to immunology and calcium signaling.

Conundrum of phenotype

Nevertheless, the larger does not necessarily lead to the better. Replication is still the key challenge facing the GWAS(s) of schizophrenia, and it is exacerbated by the fact that the overlapping of results from different GWAS(s) was relatively low and there is significant genetic correlation between schizophrenia, bipolar disorder and major depressive disorder (Consortium 2013). These facts should not be surprising. First, brain is more complex an organ than was anticipated with number of neuronal interconnections and their permutations estimated as ~2×10^{10} [7,8]. Second, schizophrenia is a highly heterogeneous illness with high between-individual variability in clinical presentation. Although the hallmark of schizophrenia is psychosis, patients with schizophrenia can display severe affective symptoms and neurological soft signs. Hitherto, a laboratory-based test for schizophrenia is absent, and hence, the diagnosis relies heavily on clinical observations and self-report questionnaires [9]. Notably, the majority of GWAS(s) in the past decade and current studies prefer dichotomous clinical diagnosis as the phenotype, ignoring the subtypes or subgroups (negative symptoms and younger age of onset) in the cohort, which might lead to inconsistent conclusions and challenges while translating these results into clinical applications.

The endophenotypes of schizophrenia

From genetic underpinnings to clinical and behavioral manifestations recognized by disease classification system, the gap is filled by multiple structural, functional, and cognitive changes found to be both heritable and impervious to treatment [10,11]. Since revisiting the concept proposed by Gottesman, *et al.* in 2003, after he introduced it to the psychopathology literature of schizophrenia in 1970's, the endophenotype of schizophrenia has become the subject of heated debates and studies. A plethora of endophenotypes has been identified based on the criteria proposed by Gottesman, *et al.* especially in the dimension of cognition and neuroimaging [12,13]. Moreover, in a recent systematic review, the cognitive traits of schizophrenia were substantiated to exist under a strong genetic influence, which is not confounded by environmental and illness-related factors such as medication [14]. Wang, *et al.* (2007) recruited 112 first-episode, drug-naïve schizophrenic patients, 296 non-psychotic first-degree relatives and 452 normal controls for the comparison of sustained attention using continuous performance test (CPT) and found that when compared to normal controls, both probands of schizophrenia and their relatives showed lower performance in CPT, especially in "hit reaction time" which catalogs psychomotor processing speed of the correct response [15]. In another study, a battery of neurocognitive tests designed to assess five different neurocognitive domains (attention and speed of information processing, memory and learning, verbal function, visuoconstructive abilities, and executive function), was used to evaluate the first-episode patients with schizophrenia and their first-degree relatives as well as normal controls. Of the three groups, patients with schizophrenia performed poorest at all the neuropsychological tests, suggesting a broad range of neurocognitive deficits. The first-degree relatives of these patients showed a similar but a less severe pattern of poor performance. The findings demonstrated that these selected neurocognitive deficits found in the families of patients with schizophrenia might represent "endophenotypes" denoting varying degrees of vulnerability to schizophrenia. These features might be valuable in molecular genetic studies of the disease in the first-episode schizophrenic patients and their first-degree relatives [16].

In addition to neurocognitive measures, both functional and structural neuroimaging have made much progress in the understanding of schizophrenia; for example, the decreased gray matter (GM) volume in bilateral insula, cingulate cortex, parahippocampal gyrus, and the left middle frontal gyrus and the implication of brain areas related to "self" in pathophysiology of schizophrenia [17]. Despite the inconsistency

in the neuroimaging findings in schizophrenia, an increasing number of studies indicated that it is a brain disease with gross dysconnectivity. Accompanied by the advent of new analytical methods such as graphic theories and dynamic causal modeling [18,19], the emphasis has been gradually shifted from focal areas related to single brain function to network-based topological configurations of the brain and the dynamic interaction among different regions of the brain in response to various environmental stimuli. Furthermore, some other studies combined the different neuroimaging techniques, i.e., multimodal neuroimaging for an enhanced understanding of both functional and structural changes in schizophrenia. For example, Wei, *et al.*, (2015) combined different modules of neuroimaging study (Voxel-based morphometry and Diffusion Tensor Imaging) to detect the alternations in the white matter (WM) of the brains of schizophrenic patients with or without deficit symptoms and their first-degree relatives. Compared to the controls and patients without deficit symptoms, both patients with deficit symptoms and their first-degree relatives showed significantly lower WM volumes and microstructural integrity, especially in the right extra-nuclear regions [20]. Such a multimodal strategy could comprehensively highlight the aberrant neural network associated with schizophrenia. Besides, the machine learning approach such as pattern recognition analysis has also been used to identify the neuroimaging biomarkers that could distinguish patients with schizophrenia from healthy controls [21,22]. One of main advantages to include endophenotype in the studies of complex disease like schizophrenia is that it further stratifies patients into more homogeneous groups based on their deficits in certain dimensions of the diseases, avoiding the false positive/negative rate arisen from the heterogeneity within the patient group. Besides, most animal models of schizophrenia are built to simulate one or two dimensions of the illness features. For example, prepulse inhibition/ acoustic startle reflex model parallels to attentional/information processing deficits; maze based delayed non-match to sample task, *etc.* [23,24]. Using endophenotypes which could be mapped to the parallel animal model makes it possible to further validate and improve the animal model of schizophrenia.

The literature review by Prasad, *et al.* first proposed the term "extended endophenotype," a network of endophenotypes linked on the same putative or documented functional basis. Constructing such "extended endophenotypes" might improve the chances of delineating the pathway from the genetic variations to the behavioral phenotype. This could support the deconstruction of the schizophrenia phenotype into physiologically meaningful clinical phenotypes that may be amenable to developing rational pharmacotherapy [25]. For instance, one study identified the co-expression of verbal memory and the reduced gray volume in the left hippocampus in the relatives of schizophrenic patients as compared to the controls; no significant difference was observed in the co-expression pattern between the relatives and the patients [26]. In addition to neurocognition-neuroimaging co-expression, the co-segregation of schizophrenia with personality traits (schizotypy) is well-recognized, which could be correlated with neurocognitive deficits and brain structural/functional aberrations. Based on the results by Prasad et al., we found a correlation amongst the FA in the right cerebral frontal lobar subgyral WM) performance IQ, and negative syndromes in schizophrenia, which could be one of the candidate extended endophenotypes of schizophrenia [27].

Genetic study of schizophrenia incorporating endophenotypes

Another advantage of the identification of endophenotypes in schizophrenia is its potential to boost the power of GWAS for detecting the genetic variants conferring the risk of disease. Although the large sample size is essential for gene-mapping of schizophrenia using a case-control design, some studies added illness-related endophenotypes and their interaction effect with diagnosis to the statistical model of association in order to increase the power of the study, the results proved that such an innovative strategy is a valuable alternative to the simplified case-control design [28-31] . For example, in a group of 74 first-episode treatment-naïve patients with schizophrenia and 51 healthy controls, Wang, *et al.* (2013) found a significant difference in gray matter volumes in three brain areas including left hOC3v in the collateral sulcus of visual cortex (hOC3vL), left cerebellar vermis lobule 10 (vermisL 10), and right cerebellar vermis lobule 10 (vermisR 10). Consequently, the study carried out the genome-wide association of the gray matter volume in these three brain regions as one of the endophenotypes for schizophrenia. The results identified SNPs in three genes (*TBXAS1*,

PIK3C2G, and *HS3ST5*) as the top signals ($p < 10^{-6}$) of association [32]. Given the sample size in the study by Wang et al., the power of association analysis was increased by an order of magnitude. Studies of genetic and genomic association of other illness-related endophenotypes in recent ten years were summarised in Table 1. As it has illustrated, although incorporating the endophenotypes in genetic and genomic studies of schizophrenia provides more insights into the genetic architecture and biologic mechanism of the illness, result inconsistency and limited ability to account for disease prevalence are still the main issues lurking behind current studies, which is likely due to difference in sample size and studied population. Moreover, majority of the genetic studies, especially ones targeting variants in candidate genes, lacked the description of effect size in their manuscripts, which makes it difficult for any further inferential explorations. As the consortium-based studies with large sample size, such as ENIGMA, proceed, more genetic markers associated with different neuroimaging phenotypes or cognitive phenotypes would be uncovered. Using polygenic risk score (PRS) benchmarked against the effect size generated from these studies might be another useful alternative approach to detect any potential effect of gene, gene × disease and gene × environment on endophenotypes.

De novo mutation and sporadic schizophrenia

Although the findings from GWAS(s) contribute remarkably towards the physiological understanding of schizophrenia, the strongest signal of associations found to date does not go beyond the odds ratio of 1.2. The common variants generated from GWAS(s) only explained the variance of disease liability partially. To account for the "missing heritability," the role of rare variants is focused with respect to the risk of disease.

Nevertheless, the first type of mutations identified to be associated with schizophrenia were copy-number variants (CNV), especially the deletions on chromosome 22q11.2 (22qDS) [33]. Approximately 1% of patients with schizophrenia can be accounted for by 22qDS [34-36]. Although schizophrenic patients with 22qDS have similar clinical presentations of core symptoms, treatment response, and neuroimaging examination to the patients without such mutations, patients with such a form of schizophrenia are likely to exhibit a low IQ, congenital anomalies, and distinguishable physical features, further providing the evidence for the genetic heterogeneity of schizophrenia [37]. Later, with the advances of technologies in genotyping in comprehensive samples, associated CNVs have been identified including *NRXN1*, *VIPR2*, 1q21, 7q11, 15q11,15q13, 16p13, and 17q12. Recently, Schizophrenia Working Groups of the Psychiatric Genomics Consortium in collaboration with Psychosis Endophenotypes International Consortium conducted the largest genome-wide study of CNVs encompassing 21,094 cases and 20,227 controls. The results not only echoed some of the well-established loci found before, such as *NRXN1*, 22q11.2, and 15q13.3 but also indicated that patients with schizophrenia carry a heavier global burden of CNV, especially in the genes enriched in synaptic function and neurobehavioral phenotypes [38,39]. Furthermore, the population prevalence of schizophrenia remains stable even with a low fecundity of patients; the genetic variations arisen *de novo* were hypothesized to replenish the population risk and counteract the reduced selective fitness of the disease [40,41]. Although increased *de novo* rates have been reported in other neurodevelopmental disorders such as autism and intellectual disability (ID) [42-44], the largest study did not display any evidence of increased nonsynonymous or loss-of-function (LOF) *de novo* mutations in schizophrenia [45,46]. However, *de novo* LOF mutations are found to be enriched in the subgroup of schizophrenic patients with low education, suggesting their role in neurodevelopmental processes across diagnostic boundaries. Wang, *et al.* tested this hypothesis by collecting T1-weighted MR and diffusion-tensor (DT) imaging from 68 patients with first-episode, drug-naïve schizophrenia and 100 healthy controls in order to compare the WM integrity indexed by fractional anisotropy (FA). The patient group demonstrated lower FA in brain areas including the left temporal lobe and right corpus callosum. Furthermore, Wang et al. stratified the same group of patients into one group with a family history and one without and observed that the group without a family history showed more severe FA deficit than the group with a family history [47]. However, this finding still needs further confirmation in independent samples.

In addition to many advantages shown by the association study of rare variants, the foremost limitation is that the approach is sample size-

Table 1: Summary of genetic-/genomic-studies of schizophrenia-related endophenotypes in patients with schizophrenia and healthy subjects.

Author-year	Endophenotype	Study design	Study group	Significant Results	Effect size
Roussos *et al.* (2008)	Prepulse Inhibition of Startle	Candidate gene (COMT)	Caucasian patients with schizophrenia	Val/Val < Val/Met < Met/Met	0.25
Baker *et al.* (2005)	Mismatch negativity event-related potentials MMN	Candidate gene (COMT)	Caucasian patients with 22DS	Val < Met	N/A
Decoster *et al.* (2012)	P300 event-related potential	Candidate genes	Caussian patients with schizophrenia	rs1045642(ABCB1)	-1.822
Lu *et al. (2007)*	P50 event-related potential	Candidate gene (COMT)	Caussian patients with schizophrenia	Met/Val < Met/Met < Val/Val	0.13
Greenwood *et al.*(2013)	P50 event-related potential	Genome-wide linkage analysis	Caussian patients with schizophrenia	Not significant	N/A
Demily et al. (2016)	P50 event-related potential	Candidate gene (COMT)	Caucasian patients with schizophrenia	Not significant	N/A
Haraldsson *et al.*(2010)	Antisaccade task for eye movements	Candidate gene (COMT)	Caussian patients with schizophrenia	Val < Met	-0.11
Schmechtig *et al.*(2010)	Antisaccade task for eye movements	Cnadidate gene (NRG1)	Caussian healthy subjects	G < A	N/A
Kattoulas *et al.* (2012)	Antisaccade task for eye movements	Candidate gene (RGS4)	Caussian healthy subjects	G < A	0.057
Vaidyanathan *et al.*(2014)	Antisaccade task for eye movements	GWAS	Caucasian healthy subjects	rs4973397(B3GNT7)	0.029
Donohoe *et al.* (2007)	Continuous Performance Test	Candidate gene (Dysbindin)	Caucasian patients with schizophrenia	Not significant	N/A
Greenwood *et al.*(2011)	Continuous Performance Test	Candidate genes	Caucasian patients with schizophrenia	Val265Ile(TAAR6)	N/A
Greenwood *et al.*(2013)	Continuous Performance Test	Genome-wide linkage analysis	Caucasian patients with schizophrenia	10q26	LOD=2.4
Walters *et al.*(2010)	Continuous Performance Test	Candidate gene (ZNF804)	Caucasian patients with schizophrenia	Not significant	N/A
Ucok *et al.*(2007)	Continuous Performance Test	Candidate gene (HTR2A)	Turkish patients with schizophrenia	CC+CT < TC	N/A
Liao *et al.*(2008)	Continuous Performance Test	Candidate gene (COMT)	Chinese Han patients with schizophrenia	Val < Met	N/A
Greenwood *et al.*(2013)	California Verbal Learning Test	Genome-wide linkage analysis	Caucasian patients with schizophrenia	8q24	2.4
Greenwood *et al.*(2011)	California Verbal Learning Test	Candidate genes	Caucasian patients with schizophrenia	Gly884Glu (GRM1)	N/A
Wedenoja *et al.*(2008)	California Verbal Learning Test	Candidate genes	Caucasian patients with schizophrenia	RELN	N/A
Roffman *et al.*(2007)	California Verbal Learning Test	Candidate gene (MTHFR)	Caucasian patients with schizophrenia	Not significant	N/A
Wirgenes *et al.*(2010)	California Verbal Learning Test	Candidate gene (COMT)	Caucasian patients with schizophrenia	Not significant	N/A
Greenwood *et al.* (2011)	Letter-Number Sequencing test	Candidate Genes	Caucasian patients with schizophrenia	CTNNA2	N/A
Aguilera *et al.*(2008)	Letter-Number Sequencing test	Candidate gene (COMT)	Caucasian patients with schizophrenia	Met < Val	0.58
Walters *et al.*(2013)	Letter-Number Sequencing test	Candidate gene region	Caucasian patients with schizophrenia	Not significant	N/A
Greenwood *et al.*(2011)	Letter-Number Sequencing test	Candidate genes	Caucasian patients with schizophrenia	HTR2A(Ser34Ser)	N/A
Almasy *et al.*(2008)	Abstraction and Mental Flexibility	Genome-wide linkage analysis	Caucasian patients with schizophrenia	Chr5q	lod = 3.4
Greenwood *et al.*(2013)	Face memory	GENome-wide linkage analysis	Caucasian patients with schizophrenia	10q26	LOD=2.4
John *et al.*(2016)	Face memory	Candidate genes	Indian patients with schizophrenia	rs10734041(PIP4K2A)	N/A
Greenwood *et al.*(2013)	Emotion recognition	Genome-wide linkage analysis	Caucasian patients with schizophrenia	1p36	LOD=3.4
Greenwood *et al.*(2011)	Emotion recognition	Candidate genes	Caucasian patients with schizophrenia	NOS1AP	N/A
Guan *et al.*(2016)	Wisconsin Card Sorting Test	Candidate genes (HTR1A and HTR5A)	Chinese patients with schizophrenia	rs1800883 (C < G)	N/A
Barnett *et al.* (2007)	Wisconsin Card Sorting Test	Candidate gene (COMT)	Meta-analysis	Not significant in patients	N/A
Diaz-Asper *et al.*(2008)	Wisconsin Card Sorting Test	Cnadidate gene (COMT)	Caussian patients with schizophrenia	Not significant	N/A

Barnett *et al.*(2008)	Wisconsin Card Sorting Test	Candidate gene (COMT)	Caucasian patients with schizophrenia	Not significant	N/A
Rodriguez-Jimenez *et al.*(2007)	Wisconsin Card Sorting Test	Candidate gene (DRD2)	Caucasian healthy subjects	CT/TT < CC	N/A
Tao *et al.*(2008)	Wisconsin Card Sorting Test	Candidate gene (PRODH)	Chinese Han patients with schizophrenia	1945T/C-1852G/A	N/A
Liao *et al.*(2008)	Wisconsin Card Sorting Test	Candidate gene (COMT)	Chinese Han patients with schizophrenia	Val < Met	N/A
Nkam *et al.*(2017)	Wisconsin Card Sorting Test	Candidate genes	Caucasian patients with schizophrenia	rs6275(DRD2, C < T)	NA
Greenwood *et al.*(2013)	Sensorimotor Dexterity	Genome-wide linkage analysis	Caucasian patients with schizophrenia	2q24, 2q32	LOD=2.7
Greenwood *et al.* (2011)	Sensorimotor Dexterity	Candidate genes	Caucasian patients with schizophrenia	5q32 (HTR4)	N/A
Yokley et al. (2012)	Sensorimotor Dexterity	Candidate gene (NRG1)	Caucasian patients with schizophrenia	T < C	N/A
Donohoe *et al.*(2007)	Spatial Memory	Candidate gene (Dysbindin)	Caucasian patients with schizophrenia	C–A–T for rs2619539, rs3213207 and rs2619538	N/A
Zhang *et al.* (2012)	Spatial Memory	Candidate gene (CACNA1C)	Chinese Han patients with schizophrenia	G<A	N/A
Greenwood *et al.* (2011)	Spatial Memory	Candidate genes	Caucasian patients with schizophrenia	2q34(ERBB4)	N/A
Ren *et al.* (2015)	Spatial memory	GWAS	Chinese Han patients with schizophrenia	rs1411832(YWHAZP5)	N/A
Greenwood *et al.* (2011)	Spatial processing speed	Candidate genes	Caucasian patients with schizophrenia	8p12(NRG1)	N/A
Greenwood *et al.* (2013)	Spatial processing speed	Genome-wide linkage analysis	Caucasian patients with schizophrenia	16q23	LOD = 2.5
Burdick *et al.* (2007)	Wide Range Achievement Test	Candidate gene (DTNBP1)	Caucasian patients with schizophrenia	CTCTAC risk haplotype	0.02
Opgen-Rhein *et al.* (2008)	Wide Range Achievement Test	Candidate gene (DAOA)	Caucasian patients with schizophrenia	Not significant	N/A
Tao *et al.* (2008)	Trail-making test	Candidate gene (PRODH)	Caucasian patients with schizophrenia	G < A	N/A
Stein *et al.* (2012)	Intracranial volume	GWAS	Meta-analysis	rs10784502(HMGA2)	0.28
Ikram *et al.* (2012)	Intracranial volume	GWAS	Caucasian community-dwelling elderly	rs4273712 and rs9915547(KANSL1)	12.5 (rs4273712) −14.9 (rs9915547)
Donohoe *et al.* (2011)	Whole brain volume	Candidate gene (ZNF804)	Caucasian patients with schizophrenia	Not significant	N/A
Walters *et al.* (2013)	Total gray matter volume	Candidate genes	Caucasian patients with schizophrenia	rs6904071(HIST1H2BJ)	-5.93
Li *et al.* (2012)	Total gray matter volume	Candidate gene (VRK2)	Chinese healthy subjects	T < C	N/A
Wang *et al.* (2013)	Total gray matter volume	GWAS	Chinese Han patients with schizophrenia	Chr7q34(TBXAS1)	-0.04315
Huang *et al.* (2015)	Total gray matter volume	Candidate gene (CACNA1C)	Chinese healthy subjects	G < A	N/A
Kempton *et al.*(2009)	Total gray matter volume	Candidate gene (CACNA1C)	Caussian healthy subjects	GG < GA < AA	0.95
Wang *et al.* (2011)	Total gray matter volume	Candidate gene (CACNA1C)	Caucasian healthy subjects	GG < GA + AA	N/A
ENIGMA (2015):	Total gray matter volume	GWAS	Caussian population	rs945270 (KTN1)	48.89
Ho *et al.* (2011)	Total white matter	Candidate gene (CNR1)	Caussian patients with schizophrenia	rs7766029 rs9450898 rs12720071	N/A
Donohoe *et al.* (2011)	Total white matter	Candidate gene (ZNF804)	Caucasian patients with schizophrenia	Not significant	N/A
Addington *et al.* (2007)	Total white matter	Candidate gene (NRG1)	Caucasian patients with schizophrenia	420M9-1395	N/A
Szeszko *et al.* (2007)	Frontal volume	Candidate gene (DISC1)	Caucasian patients with schizophrenia	phe < leu	N/A
Molina *et al.* (2011)	Frontal volume	Candidate gene (TP53)	Caucasian patients with schizophrenia	Pro/Arg < Pro/Pro	N/A
Katarina *et al.* (2008)	Frontal volume	Candidate gene (BDNF)	Caussian patients with schizophrenia	T < A	N/A

Wang *et al.* (2016)	Frontal functional connectivty	GWAS	Chinese Han patients with schizophrenia	rs6800381 (CHRM3)	0.28
Li *et al.* (2016)	Intracranial volume	Candidate genes	Chinese healthy subjects	Not significant	N/A
Li *et al.* (2016)	total brain volume	Candidate genes	Chinese healthy subjects	Not significant	N/A
Wassink *et al.* (2012)	Frontal volume	Candidate gene (ZNF804)	Caussian patients with schizophrenia	C < A	N/A
Benedetti *et al.* (2010)	Temporal volume	Candidate gene (GSK3-β)	Caucasian patients with schizophrenia	C < T	N/A
Voineskos *et al.* (2015)	Temporal volume	Candidate genes	Caucasian patients with schizophrenia	Not significnt	N/A
Adams *et al.* (2016)	Intracranial volume	GWAS	Caucasian population	rs199525 (17q21) rs11759026(6q22) rs2022464(6q21) rs11191683(10q24) rs9811910(3q28) rs138074335(12q14) rs2195243(12q23)	0.102 0.095 0.063 0.059 0.096 0.051 0.059
Vázquez-Bourgon *et al.* (2016)	Temporal volume	Candidate gene (DISC1)	Caucasian patients with schizophrenia	Leu < Phe	N/A
Mata *et al.* (2008)	Parietal volume	Candidate gene (NRG1)	Caussian patients with schizophrenia	Not significant	N/A
Takahashia *et al.* (2009)	Parietal volume	Candidate gene (DISC1)	Japanese patients with schizophrenia	Cys < Ser	-2.84
Ho *et al.* (2011)	Parietal volume	Candidate gene (CB1/ CNR1)	Caussian patients with schizophrenia	rs7766029 (C < T)	N/A
Addington *et al.* (2007)	Parietal volume	Candidate gene (NRG1)	Caucasian patients with schizophrenia	420M9-1395	N/A
Addington *et al.* (2007)	Temporal volume	Candidate gene (NRG1)	Caucasian patients with schizophrenia	420M9-1395	N/A
Addington *et al.* (2007)	Frontal volume	Candidate gene (NRG1)	Caucasian patients with schizophrenia	420M9-1395	N/A
Kuswanto *et al.* (2012)	Parietal volume	Candidate gene (ZNF804)	Chinese Han patients with schizophrenia	T < G	N/A
Trost *et al.* (2013)	temporal volume	Candidate gene (DISC1)	Caussian human subjects	T < A	N/A
Seshadri *et al.* (2007)	Parietal volume	Candidate genes	Caussian patients with schizophrenia	rs719435 (CCDC129)	N/A
Lencz *et al.* (2010)	Total white matter volume	Candidate gene (ZNF804)	Caucasian patients with schizophrenia	G < T	N/A
Hibar *et al.* (2017)	Hippocampal volume	GWAS	Caucasian population	rs77956314(HRK) rs61921502(MSRB3) rs11979341(SHH) rs7020341(ASTN2) rs2268894(DPP4) rs2289881(MAST4)	10.418 9.017 -6.755 6.645 -6.546 5.558
ENIGMA (2015):	Hippocampal volume	GWAS	Caussian population	rs77956314(HRK) rs61921502(MSRB3)	-55.18 39.9
Harrisberger *et al.* (2015)	Hippocampal volume	Meta-analysis	Caucasian/Japanese sample / mixed ethnicity	Not significant	N/A
Zhang *et al.* (2015)	Hippocampal volume	Candidate gene (BIN1)	Chinese healthy subjects	A < G	N/A
McIntosh *et al.* (2008)	White matter integrity	Candidate gene (NRG1)	N/A	T < C	N/A
Lett *et al.* (2013)	White matter integrity	Candidate gene (MIR137)	Mixed patients with schizophrenia	T < G	N/A
Wei *et al.* (2013)	White matter integrity	Candidate gene (ZNF804)	Chinese Han patients with schizophrenia	Not significant	N/A
Kuswanto *et al.* (2015)	White matter integrity	Candidate gene (MIR137)	Chinese patients with schizophrenia	T < G	N/A
Sprooten *et al.* (2012)	White matter integrity	Candidate gene (ZNF804)	Caussian healthy subjects	Not significant	N/A
Lopez *et al.* (2012)	White matter integrity	GWAS	Caucasian healthy subjects	rs7192208 (ADAMTS18)	-0.48
Wei *et al.* (2012)	White matter integrity	Candidate gene (ZNF804)	Chinese Han patients with schizophrenia	G < T	N/A
Zuliani *et al.* (2011)	White matter integrity	Candidate gene ErbB4	Healthy subjects with unknown ethnicity	G < A	N/A

dependent. However, deep whole-genome sequencing (WGS) of individuals in large numbers is not cost-effective and likely to remain the same in the near future. One alternative preferential strategy is to select samples carrying extreme phenotypes [48]. For quantitative traits, the number of individuals needed to be sequenced to reach a given power can be reduced by half if they are chosen from the upper and lower 10% tails of the phenotype distribution

[49,50]. For example, a case-control study can sample affected individuals with an early-onset of disease, family history, and poor response to treatment as compared to controls who are late-onset, without a family history, and with good response to treatment.

Phenome-wide association study (PheWAS) and related methodology

Hitherto, the majority of genomic studies adopt the principle of phenotype-to-genotype, revealing the associated variants after matching the phenotype of interest with that of the controls. However, several studies identified a minimum of 10% of genetic variants in the human genome that display a pleiotropic effect on human traits and diseases [51]. For example, the MHC region was implicated in increasing the vulnerability to both autoimmune diseases and psychiatric disorders [51,52]. These findings emphasize that only the "tip of iceberg" will be seen in any genomic study of a single candidate phenotype. Empowered by the rapid expansion of phenotypic data generated from health record information and epidemiological investigations, some studies embarked on the exploration of genotype-to-phenotype association by choosing a reverse pathway, mapping the genotypes to multiple phenotypes simultaneously (PheWAS), detailing the phenome supported by genetic variants with a pleiotropic effect [53]. Nevertheless, the neuropsychiatric PheWAS is still in its infancy with the existing studies still restricted to single gene/SNP scanning [54]. Moreover, the issues such as the lack of effective methodology incorporating the genome-wide and phenomic data continue to impede its application to a broad dataset recently divulged from large cohort studies such as US and UK biobanks. Notwithstanding, a few nascent methodologies might ameliorate such difficulties. (1) The raw genotypic data-based approach is in contrast with univariate association analysis commonly used in GWAS, and such an approach simultaneously includes multiple phenotypes in the reversed regression model of phenotype-to-genotype. For example, MultiPhen, fully taking into account the linear combination of single phenotypes included in the model, was found to boost the power of discovering the independent associated loci missed in univariate GWAS; parallel independent component analysis (parallel ICA), capable of combining genome-wide and whole-brain structural or functional neuroimaging data, could identify the risk-associated genetic variants in linear correlation with multiple neuroimaging traits [55]; And Machine learning is also an optimal choice to take full advantage of genetic and neuroimaging phenome to translate fundamental research to clinical classification [56]. (2) Summary statistics-based approach; with a rapid accumulation and easy accessibility of results from GWAS for different kind of univariate phenotypes in large sample sizes, some methodology has been developed to fully embrace these univariate summary statistics for all trait-related phenotypes to arrive at a global-trait P value. TATES, Trait-based Association Test that uses Extended Simes procedure, weights the p values for univariate phenotypes by the effective numbers of p-value and the numbers of top p-value for each phenotypes and such an algorithm was found to be more powerful and less computationally intensive than raw-data based approach [57]. Furthermore, if consider the regularized relationship within phenotypic variables for some traits, such as how BMI is the function of both height and weight (BMI= weight/height^2), treating these variables with same weight in association analysis might either lose the power or increase type-1 errors. Based on this rationale, a genome-wide inference study (GWIS), exploiting the effective size and standard errors from GWAS summary statistics, takes into account the inner function deriving trait from its phenotypic components. Nivard, *et al.* (2016) compared GWAS of BMI with GWIS from summary statistics of GWAS on height and weight and indicated that there is substantial overlapping in top hits from two analyses and the effect size maintained the same with no notable inflation of type I errors even when a significant difference in sample size existed within constituent phenotypes. The particular areas which GWIS could be applied with greater accuracy and efficacy include (bio) chemical reactions involving metabolites of which the concentrations have been analyzed in a GWAS or the traits with equations describing (active) membrane transport of proteins or metabolites given that GWAS summary statistics are available for their concentrations on both sides of the barrier. Another application of GWIS is increasing the effective sample size for the GWAS of a complex function. If not all constituent phenotypes have been measured in genotyped cohorts, these cohorts are excluded from the GWAS but can still contribute to a GWIS [58].

Future directions

The last decade witnessed the rapid rise in the identification of the associated loci of schizophrenia from GWAS of different sample size in different populations. The ongoing projects such as 1000 Genomes Project and UK Biobank will facilitate the efforts in providing broader panels of genomic benchmarks for whole-genome sequencing studies, thereby capturing the full picture of genomic variations underlying this debilitating complex disease. However, there are a few caveats that the future studies should fully recognize while designing the study, analyzing the data, and interpreting the results. First, the whole-genome sequencing will likely remain expensive in the intermediate future, especially for the sample size large enough to identify the loci covering the entire spectrum of allele frequency. Identifying the individuals distributed at the extremes of disease liability (extreme phenotype) might reduce such a financial burden and the cost of data-collection and computation. Further, the high heterogeneity of clinical manifestations in patients with schizophrenia might dissolute the signals of association using a case-control design. Therefore, stratifying the patients under the category of "illness" according to their endophenotypic/extended endophenotypic profiles might increase the power of association; similarly, selecting the quantitative endophenotypes could evade the misspecification of cohort and reduce the potential ascertainment errors. Still, the effect environmental factors mediated by potential epigenetic mechanism should be taken into consideration when designing and interpreting the studies, such as including GE interaction in the statistical model and downstream genetic expression analysis [59]. In addition, new methodology including multivariate association analysis that incorporates multiple phenotypes into the association study of phenome such as GWIS, could increase the chance of discovering the loci or genomic region, which predispose the individuals to schizophrenia by interacting with each other and with environmental factors at the varied level of the pathogenesis of schizophrenia.

Acknowledgements

This work was partly funded by National Basic Research Program of China (973 Program h 2007CB512301) and National Key Research and Developmental Program of China (Grand No. 2016YFC 1307005).

References

1. Jablensky A. Epidemiology of schizophrenia: the global burden of disease and disability. *Eur. Arch. Psychiatry. Clin. Neurosci* 250(6), 274-285 (2000).
2. Børglum AD, Demontis D, Grove J, *et al.* Genome-wide study of association and interaction with maternal cytomegalovirus infection suggests new schizophrenia loci. *Mol. Psychiatry* 19(3), 325-333 (2014).
3. Pritchard JK, NJ Cox. The allelic architecture of human disease genes: common disease–common variant… or not? *Hum. Mol. Genet* 11(20), 2417-2423 (2002).
4. El-Fishawy P. Common Disease-Rare Variant Hypothesis, in Encyclopedia of Autism Spectrum Disorders, F.R. Volkmar, Editor. Springer New York: New York, NY 720-722 (2013).
5. Gibson G. Rare and common variants: twenty arguments. *Nat. Rev. Genet* 13(2), 135-145 (2012).
6. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511(7510), 421-427 (2014).
7. Drachman DA. Do we have brain to spare? *Neurology* 64(12), 2004-2005 (2005).
8. Sporns O, Tononi G, Kötter R. The human connectome: a structural description of the human brain. *PLoS. Comput. Biol* 1(4), e42 (2005).
9. Gejman PV, Sanders AR, Duan J. The role of genetics in the etiology of schizophrenia. *Psychiatr. Clin. North. Am* 33(1), 35-66 (2010).
10. Gottesman II, Shields J. Schizophrenia and genetics: A twin study vantage point (1972).
11. Gottesman II, Shields J. A critical review of recent adoption, twin, and family studies of schizophrenia: Behavioral genetics perspectives. *Schizophr. Bull* 2(3), 360 (1976).
12. Aydın E, Ülgen MCansu, Tabo A, *et al.* Executive function and genetic loading in nonpsychotic relatives of schizophrenia patients. *Psychiatry. Res* 248(1), 105-110 (2017).
13. Chahine G, Richter A, Wolter S, *et al.* Disruptions in the left frontoparietal network underlie resting state endophenotypic markers in schizophrenia. *Hum. Brain. Mapp* 38(4), 1741-1750 (2017).
14. Blokland GA, Mesholam-Gately RI Toulopoulou T, *et al.* Heritability of Neuropsychological Measures in Schizophrenia and Nonpsychiatric Populations: A Systematic Review and Meta-analysis. *Schizophr. Bull* sbw146 (2016).
15. Wang Q, Chan R, Sun J, *et al.* Reaction time of the Continuous Performance Test is an endophenotypic marker for schizophrenia: a study of first-episode neuroleptic-naive schizophrenia, their non-psychotic first-degree relatives and healthy population controls. *Schizophr. Res* 89(1), 293-298 (2007).
16. Ma X. Neurocognitive deficits in first-episode schizophrenic patients and their first-degree relatives. *Am. J. Med. Genet. Part B: Neuropsych Genetics* 144(4), 407-416 (2007).
17. Glahn DC, Laird AR, Ellison-Wright I, *et al.* Meta-analysis of gray matter anomalies in schizophrenia: application of anatomic likelihood estimation and network analysis. *Biol. Psychiatry* 64(9), 774-781 (2008).
18. Friston KJ, Harrison L, Penny W. Dynamic causal modelling. *Neuroimage* 19(4), 1273-1302 (2003).
19. Rubinov M, Sporns O. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* 52(3), 1059-1069 (2010).
20. Lei W, Li N, Deng W, *et al.* White matter alterations in first episode treatment-naïve patients with deficit schizophrenia: a combined VBM and DTI study. *Sci. Rep* 5(1), 12994 (2015).
21. Pinaya WH, Gadelha A, Doyle OM, *et al.* Using deep belief network modelling

to characterize differences in brain morphometry in schizophrenia. *Sci. Rep* 6(1), 38897 (2016).

22. Mikolas P, Melicher T, Skoch A, *et al.* Connectivity of the anterior insula differentiates participants with first-episode schizophrenia spectrum disorders from controls: a machine-learning study. *Psychol. Med* 46(13), 2695-2704 (2016).
23. Kellendonk C, Simpson EH, Kandel ER. Modeling cognitive endophenotypes of schizophrenia in mice. *Trends. Neurosci* 32(6), 347-358 (2009).
24. Braff DL, Geyer MA. Sensorimotor gating and schizophrenia: human and animal model studies. *Arch. Gen. Psychiatry* 47(2), 181-188 (1990).
25. Prasad KM, Keshavan MS. Structural cerebral variations as useful endophenotypes in schizophrenia: do they help construct "extended endophenotypes"? *Schizophr. Bull* 34(4), 774-790 (2008).
26. Seidman LJ, Faraone SV, Goldstein JM, *et al.* Left hippocampal volume as a vulnerability indicator for schizophrenia: a magnetic resonance imaging morphometric study of nonpsychotic first-degree relatives. *Arch. Gen. Psychiatry* 59(9), 839-849 (2002).
27. Wang Q, Cheung C, Deng W, *et al.* Fronto-parietal white matter microstructural deficits are linked to performance IQ in a first-episode schizophrenia Han Chinese sample. *Psychol. Med* 43(10), 2047-2056 (2013).
28. Terwisscha van Scheltinga AF, Bakker SC, van Haren NE, *et al.* Genetic schizophrenia risk variants jointly modulate total brain and white matter volume. *Biol. Psychiatry* 73(6), 525-531 (2013).
29. Roussos P, Giakoumaki SG, Zouraraki C, *et al.* The Relationship of Common Risk Variants and Polygenic Risk for Schizophrenia to Sensorimotor Gating. *Biol. Psychiatry* 79(12), 988-996 (2016).
30. Zai G, Robbins TW, Sahakian BJ, *et al.* A review of molecular genetic studies of neurocognitive deficits in schizophrenia. *Neurosci. Biobehav. Rev* 72(1), 50-67 (2017).
31. Nkam I, Ramoz N, Breton F, *et al.* Impact of DRD2/ANKK1 and COMT Polymorphisms on Attention and Cognitive Functions in Schizophrenia. *PLoS. one* 12(1), e0170147 (2017).
32. Wang Q, Xiang B, Deng W, *et al.* Genome-wide association analysis with gray matter volume as a quantitative phenotype in first-episode treatment-naive patients with schizophrenia. *PLoS. One* 8(9), e75083 (2013).
33. Karayiorgou M, Morris MA, Morrow B, *et al.* Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proc. Natl. Acad. Sci. U S A* 92(17), 7612-7616 (1995).
34. Pawlowska B, Tomankiewicz-Zawadzka A, Ilnicka A, *et al.* A study on the occurrence of the deletion 22q11.2 in patients affected with a psychiatric disease. *Psychiatr. Pol* 41(2), 251-60 (2000).
35. Horowitz A, Shifman S, Rivlin N, *et al.* A survey of the 22q11 microdeletion in a large cohort of schizophrenia patients. *Schizophr. Res* 73(2-3), 263-267 (2005).
36. Arinami T, Ohtsuki T, Takase K, *et al.* Screening for 22q11 deletions in a schizophrenia population. *Schizophr. Res* 52(3), 167-170 (2001).
37. Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell* 148(6), 1223-1241 (2012).
38. CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium; Psychosis Endophenotypes International Consortium. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat. Gen* (2017).
39. Marshall CR, Howrigan D, Merico D, *et al.* Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat. Gen* 49(1), 27-35 (2017).
40. Malaspina D. Paternal factors and schizophrenia risk: de novo mutations and imprinting. *Schizophr. Bull* 27(3), 379-393 (2001).
41. Kranz TM, Harroch S, Manor O, *et al.* De novo mutations from sporadic schizophrenia cases highlight important signaling genes in an independent sample. *Schizophr. Res* 166(1), 119-124 (2015).
42. Sanders SJ, Murtha MT, Gupta AR, *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 485(7397), 237-241 (2012).
43. Neale BM, Kou Y, Liu L, *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 485(7397), 242-245 (2012).
44. Rauch A, Wieczorek D, Graf E, *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* 380(9854), 1674-1682 (2012).
45. Fromer M, Pocklington AJ, Kavanagh DH, *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* 506(7487), 179-184 (2014).
46. Miller G, Rockstroh B. CHAPTER OUTLINE. The Neurobiology of Schizophrenia 17 (2016).
47. Wang Q, Deng W, Huang C, *et al.* Abnormalities in connectivity of white-matter tracts in patients with familial and non-familial schizophrenia. *Psychol. Med* 41(8), 1691-700 (2011).
48. Peloso GM, Rader DJ, Gabriel S, *et al.* Phenotypic extremes in rare variant study designs. *Eur. J. Hum. Genet* 24(6), 924-930 (2016).
49. Li D, Lewinger JP, Gauderman WJ, *et al.* Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies. *Genet. Epidemiol* 35(8), 790-799 (2011).
50. Barnett IJ, Lee S, Lin X. Detecting rare variant effects using extreme phenotype sampling in sequencing association studies. *Genet. Epidemiol* 37(2), 142-151 (2013).
51. Sivakumaran S, Agakov F, Theodoratou E, *et al.* Abundant pleiotropy in human complex diseases and traits. *Am. J. Hum. Genet* 89(5), 607-618 (2011).
52. Andreassen OA, Harbo HF, Wang Y, *et al.* Genetic pleiotropy between multiple sclerosis and schizophrenia but not bipolar disorder: differential involvement of immune-related gene loci. *Mol. Psychiatry* 20(2), 207-214 (2015).
53. Cotsapas C, Voight BF, Rossin E, *et al.* Pervasive sharing of genetic effects in autoimmune disease. *PLoS. Genet* 7(8), e1002254 (2011).
54. Mier D, Kirsch P, Meyer-Lindenberg A. Neural substrates of pleiotropic action of genetic variation in COMT: a meta-analysis. *Mol. Psychiatry* 15(9), 918-927 (2010).
55. Pearlson GD, Liu J, Calhoun VD. An introductory review of parallel independent component analysis (p-ICA) and a guide to applying p-ICA to genetic data and imaging phenotypes to identify disease-associated biological pathways and systems in common complex disorders. *Front. Genet* 6(1), 276 (2015).
56. Tandon N1, Nanda P, Padmanabhan JL, *et al.* Novel gene-brain structure relationships in psychotic disorder revealed using parallel independent component analyses. *Schizophr. Res* 182(1), 74-83 (2016).
57. van der Sluis S, Posthuma D, Dolan CV. TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies. *PLoS. Genet* 9(1), e1003235 (2013).
58. Nieuwboer HA, Pool R, Dolan CV, *et al.* GWIS: Genome-wide inferred statistics for functions of multiple phenotypes. *Am. J. Hum. Genet* 99(4), 917-927 (2016).
59. Schmitt A, Martins-de-Souza D, Akbarian S, *et al.* Consensus paper of the WFSBP Task Force on Biological Markers: Criteria for biomarkers and endophenotypes of schizophrenia, part III: Molecular mechanisms. *World. J. Biol. Psychiatry* 1-27 (2016).