Neurosci Bull February 1, 2015, 31(1): 87–98. http://www.neurosci.cn
DOI: 10.1007/s12264-014-1494-4


·Review·

# Genetic studies of schizophrenia: an update

Jingchun Chen[1], Fei Cao[2], Lanfen Liu[3], Lina Wang[3], Xiangning Chen[1]

[1]*Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA 23298, USA*

[2]*Department of Neurology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China*

[3]*Department of Psychiatry, Shandong Mental Health Center, Jinan 250014, China*

Corresponding author: Jingchun Chen. E-mail: jchen@vcu.edu



Schizophrenia (SCZ) is a complex and heterogeneous mental disorder that affects about 1% of global population. In recent years, considerable progress has been made in genetic studies of SCZ. A number of common variants with small effects and rare variants with relatively larger effects have been identified. These variants include risk loci identified by genome-wide association studies, rare copy-number variants identified by comparative genomic analyses, and *de novo* mutations identified by high-throughput DNA sequencing. Collectively, they contribute to the heterogeneity of the disease. In this review, we update recent discoveries in the field of SCZ genetics, and outline the perspectives of future directions.

**Keywords:** schizophrenia; genome-wide association study; copy-number variant; *de novo* mutation; sequencing; genetics

## Introduction

Schizophrenia (SCZ) is a severe neuropsychiatric disorder with a lifetime prevalence of ~1% worldwide[1]. Clinically, SCZ is characterized by symptoms traditionally classified as positive (delusions and hallucinations), negative (flattened emotions and social withdrawal), and impairments of cognitive functions[2-4]. The age at onset is typically in late adolescence or early adulthood[5]. This disorder has a considerable impact not only on patients' health and well-being, but also on society and health services worldwide[6].

SCZ is a complex genetic disorder with a relatively high heritability, exceeding 60% in two national family studies[7,8] and 80% in twin studies[9]. After decades of frustration, genetic studies of SCZ have made significant progress in recent years with the application of genome-wide association studies (GWASs) and next-generation DNA sequencing technologies. In these studies, a number of genes have been identified with common risk alleles, rare copy-number variations (CNVs), and *de novo* mutations (DNMs). From these studies, we have learned that the genetic risk for SCZ is highly polygenic: many genes contribute to the development of the disorder but the contribution from individual genes is relatively small. These studies also reveal the complexity of the genetic architecture that includes structural variations (CNVs), common and rare single-nucleotide variations (SNVs), and DNMs. In this brief review, we update recent progress on genetic studies of SCZ, focusing on the common SNVs discovered by GWASs, as well as CNVs and rare SNVs and DNMs discovered by high-throughput DNA sequencing.

## Common Variants Contributing to SCZ

SCZ is now established as a heritable disorder by family and twin studies[7-9]. In the early years, the search for genes involved in SCZ by linkage and candidate gene studies did not produce replicable and consistent results[10,11]. From 2009 onwards, a number of creditable candidates were identified, largely by GWAS, a linkage disequilibrium-based technique designed to find links between genetic variations and diseases in a homogeneous population without *a*

*priori* knowledge of the disease. The variants discovered by GWASs are common variants, conventionally defined as those with allele frequencies ≥1%. This is largely due to the design of genotype chips used in GWAS, a design intended to test the "common diseases – common variants" hypothesis. Table 1 summarizes the top findings from GWASs in recent years. Of the loci identified, the region encompassing the major histocompatibility complex (MHC) on chromosome 6p[12] is the most significant and consistent. Many markers in this region reach genome-wide significance ($P \leq 5.0 \times 10^{-8}$). The association signals cover an interval of ~6 million base-pairs, including all three classes of MHC regions that encode for >50 genes[13,14]. Due to the high linkage disequilibrium and complex genomic structure, it is difficult to determine whether one or multiple genes in this region are involved in the disorder. It has long been speculated that the immune system is involved in SCZ, so the finding that the genetic effects are enriched in the MHC region or even in regions outside the MHC that are also involved in acquired immunity is consistent with this hypothesis. After the first GWAS reporting on the MHC region, a total of 30 loci across the whole genome were reported to be associated with SCZ by 2013[15-25], including the genes for transcription factor 4 (TCF4), neurogranin (NRGN), and DPYD/MIR317 that are known to play crucial roles in brain development. Of these GWASs, only two used Chinese subjects, one used Japanese subjects and the rest used Caucasian subjects. Most recently, the Psychiatric Genomics Consortium (PGC) SCZ group published the largest SCZ GWAS, identifying 108 independent loci across the genome, including all but 5 loci reported before[26].

GWASs, as noted above, are designed to discover associations between common variants and diseases. The successful discovery of many risk loci for SCZ provides evidence that its genetic architecture is polygenic by nature, and that individual genes have limited effects on its development. The common variants identified so far have a low genotypic relative risk individually (odds ratios (ORs) 1.1- to 1.5-fold). But collectively, these variants account for >50% of the heritability[27]. As we examine these variants closely, it is clear that most do not have known functions, and many of them are not located in protein-coding genes,

**Table 1. The 25 most common variants identified by GWASs**

| Index SNP | Chr | Position (hg19) | Allele | OR (95% CI)[a] | *P*-value[a] | Genes | References |
|---|---|---|---|---|---|---|---|
| rs115329265 | 6 | 27143833-30174131 | AG | 1.21 (1.17-1.24) | $3.48\times10^{-31}$ | MHC class II including HIST1H2BJ, PRSS16, NKAPL, and TRIM26 | [12,15,16, 19-21,25,26] |
| rs1702294 | 1 | 97792625-98559084 | TC | 0.89 (0.87-0.91) | $3.36\times10^{-19}$ | DPYD, MIR137 (micro-RNA), | [19,25,26, 108] |
| rs11191419 | 10 | 104423800-105165583 | AT | 0.91 (0.89-0.93) | $6.2\times10^{-19}$ | ARL3, AS3MT, C10orf32, CNNM2, CYP17A1, INA, NT5C2, PCGF6, PDCD11, SFXN2, TAF5, TRIM8, USMG5, and WBP1L | [19,25,26, 108,109] |
| rs2007044 | 12 | 2321860-2523731 | AG | 0.91 (0.89-0.93) | $3.22\times10^{-18}$ | CACNA1C | [24-26,108] |
| rs4129585 | 8 | 143309503-143330533 | AC | 1.09 (1.07-1.11) | $1.74\times10^{-15}$ | TSNARE1 | [25,26] |
| chr7_2025096_I | 7 | 1896096-2190096 | DI3 | 0.92 (0.90-0.94) | $8.2\times10^{-15}$ | MAD1L1 | [25,26,110] |
| rs4391122 | 5 | 60499143-60843543 | AG | 0.92 (0.90-0.94) | $1.1\times10^{-14}$ | ZSWIM6 | [25,26,110] |

To be continued

Continued from previous page

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| rs2851447 | 12 | 123448113-123909113 | CG | 0.92 (0.89-0.94) | $1.86\times10^{-14}$ | ABCB9, ARL6IP4, C12orf65, CDK2AP1, MPHOSPH9, OGFOD2, PITPNM2, RILPL2, SBNO1, SETD8 | [25,26] |
| chr2_200825237_I | 2 | 200715237-200848037 | I2D | 0.91 (0.89-0.93) | $5.65\times10^{-14}$ | AC073043.2, C2orf47, C2orf69, and TYW5 | [25,26] |
| rs10791097 | 11 | 130714610-130749330 | TG | 1.08 (1.06-1.10) | $1.09\times10^{-12}$ | SNX19 | [25,26] |
| rs11693094 | 2 | 185601420-185785420 | TC | 0.93 (0.91-0.95) | $1.53\times10^{-12}$ | ZNF804A | [26,85] |
| rs7893279 | 10 | 18681005-18770105 | TG | 1.13 (1.09-1.16) | $1.97\times10^{-12}$ | CACNB2 | [25,26] |
| rs12129573 | 1 | 73766426-73991366 | AC | 1.08 (1.06-1.10) | $2.03\times10^{-12}$ | LRRIQ3 | [25,26] |
| rs6704768 | 2 | 233559301-233753501 | AG | 0.93(0.91-0.95) | $2.32\times10^{-12}$ | C2orf82, EFHD1, GIGYF2, KCNJ13, and NGEF | [25,26] |
| rs55661361 | 11 | 124610007-124620147 | AG | 0.93 (0.91-0.95) | $2.8\times10^{-12}$ | ESAM, MSANTD2, NRGN, and VSIG2 | [16,26] |
| rs9636107 | 18 | 52747686-53200117 | AG | 0.93 (0.91-0.95) | $3.34\times10^{-12}$ | TCF4 | [16,17,19,25, 26,108] |
| chr11_46350213_D | 11 | 46342943-46751213 | I2D | 0.91 (0.88-0.93) | $1.26\times10^{-11}$ | AMBRA1, ARHGAP1, ATG13, CHRM4, CKAP5, CREB3L1, DGKZ, F2, HARBI1, MDK, and ZNF408 | [26,111] |
| rs11682175 | 2 | 57943593-58502192 | TC | 0.93 (0.91-0.95) | $1.47\times10^{-11}$ | FANCL and VRK2 | [17,26] |
| rs2535627 | 3 | 52541105-52903405 | TC | 1.07 (1.05-1.09) | $4.26\times10^{-11}$ | GLT8D1, GNL3, ITIH1, and ITIH3 | [24,26] |
| rs111294930 | 5 | 151941104-152797656 | AG | 1.09 (1.06-1.12) | $1.06\times10^{-10}$ | GRIA1 | [25,26] |
| rs2905426 | 19 | 19374022-19658022 | TG | 0.93 (0.91-0.95) | $3.63\times10^{-10}$ | CILP2, GATAD2A, HAPLN4, MAU2, NCAN, NDUFA13, PBX4, SUGP1, TM6SF2, and TSSK6 | [25,26] |
| rs77149735 | 1 | 243503719-244002945 | AG | 1.32 (1.20-1.44) | $3.73\times10^{-9}$ | AKT3 and SDCCAG8 | [24,26] |
| rs59979824 | 2 | 193848340-194028340 | AC | 0.94 (0.92-0.96) | $8.41\times10^{-9}$ | PCGEM1 | [19,26] |
| rs10503253 | 8 | 4177794-4192544 | AC | 1.07 (1.05-1.10) | $1.06\times10^{-8}$ | CSMD1 | [19,26,108] |
| rs7819570 | 8 | 89340626-89753626 | TG | 1.08 (1.05-1.11) | $1.22\times10^{-8}$ | MMP16 | [19,26] |

Chr, chromosome; CI, confidence interval; OR, odds ratio; SNP, single-nucleotide polymorphism. [a]All ORs and *P*-values are from the most recent PGC study[26].

suggesting that most of the identified variants are not causal. Therefore, even though GWASs have identified >100 variants associated with SCZ, we still cannot be certain to what extent common variants contribute to the etiology of SCZ. Another lesson learned from GWASs is that a stringent threshold requires a huge sample size to reliably identify risk genes. From the incremental PGC studies[19,25,26], it is clear that when the sample size reaches a critical value, every addition of 1 000 samples leads to the identification of 2–3 more new loci. The same trend is observed in studies of other complex diseases and traits, such as obesity, height, and blood lipids[28-30]. This trend demands collaborative work from many investigators to ensure the successful discovery of risk genes for complex diseases. In the case of SCZ studies, the PGC includes almost all existing samples of Caucasian ancestry worldwide.

Since we have good reasons to believe that there are other unidentified risk genes for SCZ, and that continuing to increase the sample size is not sustainable in the long run, other approaches must be considered. Based on this rationale, we and others have explored alternative approaches to discover risk genes for SCZ. With a two-stage design, by integrating data-mining and functional analyses of a selected number of candidates from GWAS datasets, we found that two markers in the CMYA5 gene are associated with SCZ[31]; one of the markers (rs10043986) changes a proline to leucine in the protein sequence. Using a convergent functional genomics approach, which utilizes poly-evidence scoring and pathway analyses, Ayalew *et al.* identified several genes involved in SCZ, including *DISC1* and *TCF4*[32]. By combining gene expression profiling and GWAS data, Luo *et al.* showed that *CAMKK2* is differentially expressed in SCZ and controls, and a marker in the gene, rs1063843, is associated with the expression and diagnosis of SCZ in a large GWAS dataset[33].

## Copy-Number Variations and SCZ

Changes of DNA copy involving insertion/deletion (indel) or duplication (dup) are known as CNVs. Indels can be as small as a few hundred base-pairs or as large as an entire chromosome, and about one-quarter of the human genome harbors CNVs[34]. As a result, CNVs can change the dosage of one or more genes in the regions covered by CNVs, and therefore exert a profound effect on the expression of the genes. Most studies of CNVs in SCZ are based on the analyses of signal intensity from GWAS microarrays. The assumption is that most parts of the genome (except for the X and Y chromosomes) have two copies, and the signal intensity of markers along the chromosome is approximately the same when smoothed across a reasonable number of markers or genomic distance. When the signal intensity changes consistently and continuously for a reasonable genomic interval (hundreds of kilobase-pairs or more), a copy-number change can be inferred. Typically, these inferred CNVs need to be verified experimentally *via* real-time quantitative PCR or other techniques.

Over the years, many CNVs have been found to increase the risk for SCZ. Most of these involve multiple genes, while some involve a single gene or do not have known genes in the interval. For example, 2p16.3 del only affects the NRXN1 gene[35-42], and 7p36.3 dup only affects the VIPR2 gene[38,43,44]. CNVs that alter the expression of multiple genes include 1q21.1 del/dup (34 genes)[36-39,41,45-47], 3q29del/dup (21 genes)[38-40,44,48,49], 15q13.3del (12 genes)[38,39,44,46,47,49], 16p13.1dup (11 genes)[40,41,50,51], and 22q11.2del/dup (53 genes)[38,39,44,46,47,51-56], *etc*. (see Table 2 for more details). For the involved genes, the effect is more pronounced[34,57-59]. The large 22q11.21 locus (3 Mb), also known as DiGeorge and velo-cardio-facial syndrome critical region, was first reported to be associated with SCZ in the 1990s[60], and this was verified in many later studies[38,39,44,46,47,51-55]. Most of the studies show that deletion of this region increases the risk of SCZ. A recent study indicates that a duplication of this same region is protective against SCZ[56], demonstrating a dosage effect on the development of SCZ. As this region contains >50 genes, it is still not clear which are involved and how they contribute to SCZ[61,62]. There is evidence that the total burden of rare CNVs (numbers of CNVs per individual in combination with the number of genes per CNV) is increased in SCZ patients at both the whole-genome and specific loci levels[37,46,47].

In general, CNVs have relatively low frequencies (typically less than 1 in 1 000 individuals), but account for a substantially higher risk (OR 2.7 to infinity)[63]. Although it is uncertain which gene(s) are responsible for the effects, as multiple genes are involved in most CNVs, evidence shows that the risky CNVs most likely affect genes involved

**Table 2. Top CNVs in SCZ**

| CNV | Position (Mb) | Size (Mb) | Types | CNV frequency | | OR (95% CI) | *P*-value | Genes | References[a] |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Cases | Controls | | | | |
| 1q21.1 | 144.6–148.0 | 3.4 | Del/dup | 0.00176 (20/11372) | 0.00021 (10/47311) | 8.3 (3.7–19.9) | $2.2 \times 10^{-8}$ | 34 | [36,37,**38**,39,41,45-47] |
| 2p16.3 | 49.9–51.5 | 0.02–0.42 | Exon del | 0.00182 (23/12627) | 0.00022 (10/45284) | 8.2 (3.8–19.4) | $5.5 \times 10^{-9}$ | 1 (NRXN1) | [35-37,**38**,39-42] |
| 3q29 | 197.2–198.8 | 0.84–1.6 | Del | 0.00080 (6/7539) | 0.00003 (1/39747) | 17.0 (1.4–1198.4) | $9.7 \times 10^{-3}$ | 19 | [38-40,**48**,49] |
| 3q29 | 196.8–196.9 | 0.05 | Dup | 0.00121 (10/8280) | <0.00013 (0/7431) | Inf (1.6–Inf) | $1.0 \times 10^{-2}$ | 2 | [39,**44**] |
| 7q36.3 | 158.5–158.8 | 0.12–0.36 | Exon dup | 0.00191 (14/7322) | 0.00047 (7/14814) | 4.0 (1.5–11.9) | $2.0 \times 10^{-3}$ | 1 (VIPR2) | [**38**,43,44] |
| 15q11.2 | 20.3–20.8 | 0.5 | Del | 0.00551 (26/4692) | 0.00192 (79/41115) | 2.7 (1.5–4.9) | $6.0 \times 10^{-4}$ | 4 | [40,**47**,49,51] |
| 15q11.2-13.1 | 20.3–26.4 | 4.1–9.0 | Mat dup | 0.00053 (4/7578) | 0.00007 (3/41367) | 7.3 (1.2–50.0) | $1.0 \times 10^{-2}$ | 13–24 | [49,**112**] |
| 15q13.3 | 28.7–30.3 | 1.5 | Del | 0.00193 (21/10866) | 0.00020 (9/45913) | 9.9 (4.3–24.4) | $2.0 \times 10^{-9}$ | 12 | [**38**,39,44,46,47,49] |
| 16p11.2 | 29.4–30.1 | 0.7 | Dup/Del | 0.00313 (31/9859) | 0.00027 (8/29589) | 11.6 (5.6–29.3) | $1.5 \times 10^{-12}$ | 29 | [37,**38**,39,44,49,55, 113-115] |
| 16p13.1 | 14.6–18.7 | 1.16 | Dup | 0.00299 (13/4332) | 0.00091 (32/35047) | 3.3 (1.3–7.9) | $7.1 \times 10^{-3}$ | 11 | [40,41,**50**,51] |
| 17p12 | 14.1–15.4 | 0.93–1.31 | Del | 0.00151 (8/5292) | 0.00015 (6/39213) | 9.9 (3.4–28.5) | $5.0 \times 10^{-5}$ | 15 | [40,**51**] |
| 17q12 | 34.8-36.2 | 1.4 | del/Dup | 0.00073 (5/6882) | 0.00018 (2/11255) | 4.2 (1.3-Inf) | $1.8 \times 10^{-2}$ | 18 | [**39**,116] |
| 22q11.2 | 17.1–20.2 | 1.4–2.5 | Del | 0.00307 (35/11 365) | <0.00002 (0/45 361) | Inf (35.9–Inf) | $<1.0 \times 10^{-16}$ | 53 | [**38**,39,44,46,47,51-55] |
| 22q11.2 | 18.9-21.9 | 1.5–3.0 | Dup | 0.014% (3/21 138) | 0.085% (22/25 867) | 0.2 (0.1–0.6) | $8.6 \times 10^{-4}$ | 31 | [**39**,56] |

[a]For each CNV case-control frequency, ORs and *P*-values are from the bolded study. CI, confidence interval; CNV, copy-number variant; Del, deletion; Exon del, exonic deletion; Dup, duplication; Exon dup, exonic duplication; Inf, infinite; Mat dup, maternally-derived duplication; OR, odds ratio.

in specific brain functions[64]. In addition, most CNVs are not SCZ-specific; in contrast, many have effects in multiple neurodevelopmental disorders[65].

## *De Novo* Mutations in SCZ

It is known that SCZ patients have a reduced reproductive rate[66,67], which is a negative selection pressure in evolution. However, the incidence rate of SCZ remains stable worldwide at ~1%. Therefore, there must be a genetic mechanism to supply causal factors to balance the negative selection. DNMs, defined as mutations arising sporadically either in the germ-line of the parents or at an early stage of embryonic development so that the mutations are only detected in affected individual but not in the parents, was proposed many years ago as a mechanism to offset the

negative selection[68]. This proposition was based on the epidemiological observation that paternal age is associated with an increased risk of SCZ[69-71]. This hypothesis did not receive much attention in genetic studies until recently, when high-throughput DNA sequencing of SCZ families provided direct evidence that affected offspring have excess DNMs across the genome[72,73]. Further studies of DNMs in individual genes[74,75], a particular set of genes[76,77], or the exome[78-80], also provided evidence that DNMs are enriched in SCZ patients. Direct measurement also indicates that SCZ patients have a higher mutation rate[76].

It should be pointed out that DNMs can be CNVs[48,81], but the overwhelming majority are SNVs. Due to their low frequencies, DNMs are mostly classified as rare mutations. The difference between DNMs and other rare mutations is that DNMs occur only in the offspring, not in the parents. Since the frequencies of DNMs are extremely low, most tests of association are performed at the level of selected genes (gene sets or pathways). Tests can also be conducted on the basis of functional classification of the DNMs: coding *versus* non-coding sequences, *versus* non-synonymous, and neutral *versus* deleterious. Detailed analyses of DNMs in affected offspring reveal an excess of missense and disrupting mutations in protein-coding sequences, especially those involved in synaptic functions[77,79]. While most DNMs are unique events, some are recurrent[79]. Analyses of rare mutations reach the same conclusion that the polygenic burden of rare disruptive mutations is excessive in SCZ patients[80]. The converging results from DNMs, CNVs and SNVs support the notion that these mutations are likely enriched in the same pathways and thus play similar pathological roles in the etiology of SCZ.

## Conclusions and Future Directions

Some important conclusions can be drawn from the above description. First, SCZ is a polygenic and heterogeneous disorder and its genetic basis involves defects in many genes. These defects can include common SNVs, common and rare CNVs, and rare and recurrent DNMs. Overall, more than one hundred common variants and many more rare variants (including both CNVs and DNMs) are associated with SCZ. The discovery of these loci confirms that a substantial number of genetic defects may be required for the manifestation of the disease and each individual gene has a limited effect. Currently, it is not clear what proportion of these variants is common and what proportion is rare. Given that the number of genes involved is likely to be more than hundreds, the genes responsible in individual patients may or may not overlap. This genetic heterogeneity not only imposes great challenges to the discovery of risk variants, but also demands individualized treatment for optimal effect. This explains why the commonly-used antipsychotics have very different effects on different patients.

Second, many of the common variants, CNVs, and DNMs identified in recent years are not specific to SCZ. Polygenic scores calculated from risk variants for SCZ can predict bipolar disorder[12], suggesting some sharing of genetic risks between these disorders. Further examination of bipolar disorder, major depression, autism, and attention deficit and hyperactivity disorder indicates broader sharing of genetic liabilities among these disorders[82-84]. For example, ZNF804A and TCF4, loci first identified by SCZ GWASs, are associated with bipolar disorder as well[83,85]. MIR137 targets multiple genes involved in SCZ, bipolar disorder, and autism[86-88]. While the extent of sharing between these disorders may differ, the pleiotropic effects seem to extend beyond these traditional psychiatric disorders[89-92]. The extent and the identity of the variants specific to SCZ remain unknown.

Third, most variants identified so far are non-functional and non-causal. In the largest SCZ GWAS that reported 108 independent loci, 15 loci had no known genes nearby, and 36 loci had >3 genes[26]. For each of these loci, tens to hundreds of SNVs are involved. This implies that most of the variants showing association signals are most likely not causal. The functional variants at these loci remain unknown and much effort is needed to discover their mechanisms, and thus to improve our understanding of the biological mechanisms involved in disease etiology[93]. Rare variants (including CNVs and DNMs) may be pathologically causal. Unfortunately, due to the extremely low frequencies of these rare variants, only a few are reported in cases[36,78-80]. For example, the largest exome sequencing project reported a significantly higher rate of rare (frequency <0.1%), disruptive mutations in cases compared to controls among gene sets that had previously been associated with SCZ[80]. Those variants found only or overwhelmingly

in affected individuals present a realistic opportunity to establish causal relationships in functional and animal model studies.

Fourth, although we cannot be sure of their relative proportions, both common and rare variants contribute to the development of SCZ. Given that SCZ is negatively selected in evolution but maintains a stable incidence rate, we would argue that rare mutations are more likely the driving force in SCZ, and DNMs are the main counter-event to balance negative selection[94-96]. This note is consistent with the polygenic nature and heterogeneity of SCZ.

Based on what we have learned from these recent studies, future studies require extensive collaboration among investigators and across disciplines. Collaborations and consortia of investigators are necessary to assemble the sample sizes required to detect common variants with small effects and rare variants with low frequencies. SCZ GWASs organized by the PGC are successful examples of the discovery of common variants and rare CNVs associated with the disease[23,25,26,46,97]. The same model has been adapted for exome sequencing to discover rare variants[80]. Multi-disciplinary collaboration is also a current trend. Geneticists need to work more closely with clinicians, statisticians, informaticians, and computer programmers to improve the processing, integration, and analysis of large genetic, phenotypical/clinical, and genomic datasets. This is because more and more studies use the systems biology approach to collect data, and more and more studies incorporate data and information from different fields. These studies produce ever-increasingly large datasets that require specialized techniques and expertise to process and analyze. To some extent, the success of a study depends on the capability and efficiency of data processing, integration, curation, and analysis.

Looking forward, to understand the genetic mechanism of SCZ, we should focus on the following areas. The first is functional studies aiming at the discovery and understanding of causal variants at identified loci. As described above, many loci have been identified by GWASs, most of which should contain genuine variants contributing to the development of SCZ. Therefore, the time is ripe to pursue functional studies to understand the mechanism. Since most of the loci discovered by GWASs encompass large genomic intervals and contain multiple genes, deep sequencing of a substantial number of subjects is a necessary first step to discover the causal variants at these loci. Deep sequencing can provide a catalog of variants with potentially deleterious functions. By combining functional genomics analyses with molecular, cellular, and animal model studies, we hope to demonstrate that a variant, or a group of variants, causes functional changes of a gene, leading to changes of the properties of neurons such as migration, communication, and differentiation. While it may be difficult to prove the causality of the variants in SCZ, we can reasonably interpret the effects of the variants if they have functional consequences at the level of neurons or lead to behavioral and cognitive changes in animal models similar to SCZ patients. Since a single variant may not be sufficient to cause observable changes resembling SCZ, cellular and animal models accommodating multiple variants should be explored.

The second direction is to continue the search for new risk loci. There is a good reason to believe that more loci await discovery. Given the small effect sizes of common variants and modest effects at best even for rare variants or DNMs, further searching for new loci requires a significant increase of power, to which there are several approaches. The simple approach is to increase sample sizes. Since most of the well-studied Caucasian samples have been included in PGC studies, there is not much room to expand the sample size, and the collection of new samples takes time, so this is not sustainable in the long run. For samples from other understudied ethnicities, organizing consortia and collaboration will be the most effective approach to discover novel risk loci. Integration of information from independent sources, including genome-wide functional genomic data such as biological pathways, gene expression, and DNA/chromatin modification, is another approach to improving power. These functional data can be used to exclude unlikely genes/loci, effectively reducing the number of tests needed for unbiased searches across the genome, thus improving the power to discover novel loci. For rare variants, pathway- and network-based analyses are essential. Furthermore, refinement of phenotype and the use of endophenotype and multiple related phenotypes to purify samples provide another approach to improving power. As discussed above, SCZ is genetically heterogeneous, so if we combine clinical information and functional endophenotypes (such as cognitive functions and immunological responses) to screen samples and define

biologically-based disease subgroups, we can reduce the heterogeneity within the group and thus improve power. A benefit of having a biologically-defined phenotype is that it can provide insights into the underlying mechanism and provide options for treatment of the disease.

A third direction is to study gene-environment interactions. It has long been speculated that the immune system is involved in SCZ. As the immune system is the primary defense against environmental pathogens, infection with pathogens leads to the activation/dysfunction of immune responses, which can modulate the risk of developing SCZ. This is consistent with increased incidence in individuals who are migrants, have an urban upbringing[98], are exposed to maternal infection during pregnancy[99,100], and suffer childhood infection[101]. While the specifics of the interaction remain largely unknown, some studies have implicated specific genes that can serve as a model for studying gene-environment interactions[102,103]. Another well-known fact is that there is a small but consistent difference of incidence between males and females. This difference can be seen as a special case of gene-environment interaction where sex hormones are the most likely mediators. The study of gene-sex interaction can follow the models of other sexually dimorphic diseases[104]. In SCZ, the RELN gene has been shown to have sexually different effects[105-107]. Other environmental factors, such as stressful/traumatic life events and substance use may also alter the risk of developing SCZ. Systematic examination of the interactions between these environmental factors and genetic variants would provide insights into how environmental factors modulate and mediate the risk of SCZ, thus improving our understanding of the pathology of this disorder.

## ACKNOWLEDGEMENTS

This review was supported by the National Institutes of Health, USA (MH101054).

Received date: 2014-09-19; Accepted date: 2014-12-16

## REFERENCES

[1] Leucht S, Burkard T, Henderson J, Maj M, Sartorius N. Physical illness and schizophrenia: a review of the literature. Acta Psychiatr Scand 2007, 116: 317–333.

[2] Crow TJ. The two-syndrome concept: origins and current status. Schizophr Bull 1985, 11: 471–486.

[3] Andreasen NC. A unitary model of schizophrenia: Bleuler's "fragmented phrene" as schizencephaly. Arch Gen Psychiatry 1999, 56: 781–787.

[4] Sass LA, Parnas J. Schizophrenia, consciousness, and the self. Schizophr Bull 2003, 29: 427–444.

[5] Hafner H, Maurer K, Loffler W, Riecher-Rossler A. The influence of age and sex on the onset and early course of schizophrenia. Br J Psychiatry 1993, 162: 80–86.

[6] Millier A, Schmidt U, Angermeyer MC, Chauhan D, Murthy V, Toumi M, *et al.* Humanistic burden in schizophrenia: a literature review. J Psychiatr Res 2014, 54: 85–93.

[7] Lichtenstein P, Yip BH, Bjork C, Pawitan Y, Cannon TD, Sullivan PF, *et al.* Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. Lancet 2009, 373: 234–239.

[8] Wray NR, Gottesman II. Using summary data from the danish national registers to estimate heritabilities for schizophrenia, bipolar disorder, and major depressive disorder. Front Genet 2012, 3: 118.

[9] Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch Gen Psychiatry 2003, 60: 1187–1192.

[10] Voisey J, Swagell CD, Hughes IP, Lawford BR, Young RM, Morris CP. Analysis of HapMap tag-SNPs in dysbindin (DTNBP1) reveals evidence of consistent association with schizophrenia. Eur Psychiatry 2010, 25: 314–319.

[11] Shi J, Gershon ES, Liu C. Genetic associations with schizophrenia: meta-analyses of 12 candidate genes. Schizophr Res 2008, 104: 96–107.

[12] Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 2009, 460: 748–752.

[13] Debnath M, Cannon DM, Venkatasubramanian G. Variation in the major histocompatibility complex [MHC] gene family in schizophrenia: associations and functional implications. Prog Neuropsychopharmacol Biol Psychiatry 2013, 42: 49–62.

[14] Corvin A, Morris DW. Genome-wide association studies: findings at the major histocompatibility complex locus in psychosis. Biol Psychiatry 2014, 75: 276–283.

[15] Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I, *et al.* Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature 2009, 460: 753–757.

[16] Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, *et al.* Common variants conferring risk of schizophrenia. Nature 2009, 460: 744–747.

[17] Steinberg S, de JS, Andreassen OA, Werge T, Borglum AD, Mors O, *et al.* Common variants at VRK2 and TCF4 conferring risk of schizophrenia. Hum Mol Genet 2011, 20:

4076–4081.

[18] Shi Y, Li Z, Xu Q, Wang T, Li T, Shen J, *et al.* Common variants on 8p12 and 1q24.2 confer risk of schizophrenia. Nat Genet 2011, 43: 1224–1227.

[19] Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA, *et al.* Genome-wide association study identifies five new schizophrenia loci. Nat Genet 2011, 43: 969–976.

[20] Ikeda M, Aleksic B, Kinoshita Y, Okochi T, Kawashima K, Kushima I, *et al.* Genome-wide association study of schizophrenia in a Japanese population. Biol Psychiatry 2011, 69: 472–478.

[21] Yue WH, Wang HF, Sun LD, Tang FL, Liu ZH, Zhang HX, *et al.* Genome-wide association study identifies a susceptibility locus for schizophrenia in Han Chinese at 11p11.2. Nat Genet 2011, 43: 1228–1231.

[22] Rietschel M, Mattheisen M, Degenhardt F, Kahn RS, Linszen DH, Os J, *et al.* Association between genetic variation in a region on chromosome 11 and schizophrenia in large samples from Europe. Mol Psychiatry 2012, 17: 906–917.

[23] Irish Schizophrenia Genomics Consortium. Genome-wide association study implicates HLA-C*01:02 as a risk factor at the major histocompatibility complex locus in schizophrenia. Biol Psychiatry 2012, 72: 620–628.

[24] Hamshere ML, Walters JT, Smith R, Richards AL, Green E, Grozeva D, *et al.* Genome-wide significant associations in schizophrenia to ITIH3/4, CACNA1C and SDCCAG8, and extensive replication of associations reported by the Schizophrenia PGC. Mol Psychiatry 2013, 18: 708–712.

[25] Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kahler AK, Akterin S, *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat Genet 2013, 45: 1150–1159.

[26] Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature 2014, 511: 421–427.

[27] Lee SH, DeCandia TR, Ripke S, Yang J, Sullivan PF, Goddard ME, *et al.* Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. Nat Genet 2012, 44: 247–250.

[28] Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet 2010, 42: 937–948.

[29] Lango AH, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 2010, 467: 832–838.

[30] Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, *et al.* Discovery and refinement of loci associated with lipid levels. Nat Genet 2013, 45: 1274–1283.

[31] Chen X, Lee G, Maher BS, Fanous AH, Chen J, Zhao Z, *et al.* GWA study data mining and independent replication identify cardiomyopathy-associated 5 (CMYA5) as a risk gene for schizophrenia. Mol Psychiatry 2011, 16: 1117–1129.

[32] Ayalew M, Le-Niculescu H, Levey DF, Jain N, Changala B, Patel SD, *et al.* Convergent functional genomics of schizophrenia: from comprehensive understanding to genetic risk prediction. Mol Psychiatry 2012, 17: 887–905.

[33] Luo XJ, Li M, Huang L, Steinberg S, Mattheisen M, Liang G, *et al.* Convergent lines of evidence support CAMKK2 as a schizophrenia susceptibility gene. Mol Psychiatry 2014, 19: 774–783.

[34] Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell 2012, 148: 1223–1241.

[35] Rujescu D, Ingason A, Cichon S, Pietilainen OP, Barnes MR, Toulopoulou T, *et al.* Disruption of the neurexin 1 gene is associated with schizophrenia. Hum Mol Genet 2009, 18: 988–996.

[36] Need AC, Ge D, Weale ME, Maia J, Feng S, Heinzen EL, *et al.* A genome-wide investigation of SNPs and CNVs in schizophrenia. PLoS Genet 2009, 5: e1000373.

[37] 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.

[38] Levinson DF, Duan J, Oh S, Wang K, Sanders AR, Shi J, *et al.* Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. Am J Psychiatry 2011, 168: 302–316.

[39] Szatkiewicz JP, O'Dushlaine C, Chen G, Chambert K, Moran JL, Neale BM, *et al.* Copy number variation in schizophrenia in Sweden. Mol Psychiatry 2014, 19: 762–773.

[40] Magri C, Sacchetti E, Traversa M, Valsecchi P, Gardella R, Bonvicini C, *et al.* New copy number variations in schizophrenia. PLoS One 2010, 5: e13422.

[41] Ikeda M, Aleksic B, Kirov G, Kinoshita Y, Yamanouchi Y, Kitajima T, *et al.* Copy number variation in schizophrenia in the Japanese population. Biol Psychiatry 2010, 67: 283–286.

[42] Kirov G, Rujescu D, Ingason A, Collier DA, O'Donovan MC, Owen MJ. Neurexin 1 (NRXN1) deletions in schizophrenia. Schizophr Bull 2009, 35: 851–854.

[43] Yuan J, Jin C, Sha W, Zhou Z, Zhang F, Wang M, *et al.* A competitive PCR assay confirms the association of a copy number variation in the VIPR2 gene with schizophrenia in Han Chinese. Schizophr Res 2014, 156: 66–70.

[44] Vacic V, McCarthy S, Malhotra D, Murray F, Chou HH, Peoples A, *et al.* Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. Nature 2011, 471: 499–503.

[45] Mefford HC, Sharp AJ, Baker C, Itsara A, Jiang Z, Buysse K, *et al.* Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. N Engl J Med 2008, 359: 1685–1699.

[46] International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature 2008, 455: 237–241.

[47] Stefansson H, Rujescu D, Cichon S, Pietilainen OP, Ingason A, Steinberg S, *et al.* Large recurrent microdeletions associated with schizophrenia. Nature 2008, 455: 232–236.

[48] Mulle JG, Dodd AF, McGrath JA, Wolyniec PS, Mitchell AA, Shetty AC, *et al.* Microdeletions of 3q29 confer high risk for schizophrenia. Am J Hum Genet 2010, 87: 229–236.

[49] Kirov G, Pocklington AJ, 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 2012, 17: 142–153.

[50] Ingason A, Rujescu D, Cichon S, Sigurdsson E, Sigmundsson T, Pietilainen OP, *et al.* Copy number variations of chromosome 16p13.1 region associated with schizophrenia. Mol Psychiatry 2011, 16: 17–25.

[51] Kirov G, Grozeva D, Norton N, Ivanov D, Mantripragada KK, Holmans P, *et al.* Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. Hum Mol Genet 2009, 18: 1497–1503.

[52] Bassett AS, Marshall CR, Lionel AC, Chow EW, Scherer SW. Copy number variations and risk for schizophrenia in 22q11.2 deletion syndrome. Hum Mol Genet 2008, 17: 4045–4053.

[53] Monks S, Niarchou M, Davies AR, Walters JT, Williams N, Owen MJ, *et al.* Further evidence for high rates of schizophrenia in 22q11.2 deletion syndrome. Schizophr Res 2014, 153: 231–236.

[54] Grozeva D, Conrad DF, Barnes CP, Hurles M, Owen MJ, O'Donovan MC, *et al.* Independent estimation of the frequency of rare CNVs in the UK population confirms their role in schizophrenia. Schizophr Res 2012, 135: 1–7.

[55] Glessner JT, Reilly MP, Kim CE, Takahashi N, Albano A, Hou C, *et al.* Strong synaptic transmission impact by copy number variations in schizophrenia. Proc Natl Acad Sci U S A 2010, 107: 10584–10589.

[56] Rees E, Kirov G, Sanders A, Walters JT, Chambert KD, Shi J, *et al.* Evidence that duplications of 22q11.2 protect against schizophrenia. Mol Psychiatry 2014, 19: 37–40.

[57] Sullivan PF, Daly MJ, O'Donovan M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nat Rev Genet 2012, 13: 537–551.

[58] Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, Itsara A, *et al.* A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. Nat Genet 2010, 42: 203–209.

[59] 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.

[60] Karayiorgou M, Morris MA, Morrow B, Shprintzen RJ, Goldberg R, Borrow J, *et al.* Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. Proc Natl Acad Sci U S A 1995, 92: 7612–7616.

[61] Squarcione C, Torti MC, Di FF, Biondi M. 22q11 deletion syndrome: a review of the neuropsychiatric features and their neurobiological basis. Neuropsychiatr Dis Treat 2013, 9: 1873–1884.

[62] Williams NM. Molecular mechanisms in 22q11 deletion syndrome. Schizophr Bull 2011, 37: 882–889.

[63] Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet 2008, 40: 695–701.

[64] Raychaudhuri S, Korn JM, McCarroll SA, Altshuler D, Sklar P, Purcell S, *et al.* Accurately assessing the risk of schizophrenia conferred by rare copy-number variation affecting genes with brain function. PLoS Genet 2010, 6: e1001097.

[65] Grayton HM, Fernandes C, Rujescu D, Collier DA. Copy number variations in neurodevelopmental disorders. Prog Neurobiol 2012, 99: 81–91.

[66] Bassett AS, Bury A, Hodgkinson KA, Honer WG. Reproductive fitness in familial schizophrenia. Schizophr Res 1996, 21: 151–160.

[67] Laursen TM, Munk-Olsen T. Reproductive patterns in psychotic patients. Schizophr Res 2010, 121: 234–240.

[68] Malaspina D, Harlap S, Fennig S, Heiman D, Nahon D, Feldman D, *et al.* Advancing paternal age and the risk of schizophrenia. Arch Gen Psychiatry 2001, 58: 361–367.

[69] Byrne M, Agerbo E, Ewald H, Eaton WW, Mortensen PB. Parental age and risk of schizophrenia: a case-control study. Arch Gen Psychiatry 2003, 60: 673–678.

[70] Dalman C, Allebeck P. Paternal age and schizophrenia: further support for an association. Am J Psychiatry 2002, 159: 1591–1592.

[71] Malaspina D. Paternal factors and schizophrenia risk: de novo mutations and imprinting. Schizophr Bull 2001, 27: 379–393.

[72] Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of de novo copy number mutations with sporadic schizophrenia. Nat Genet 2008, 40: 880–885.

[73] Xu B, Roos JL, Dexheimer P, Boone B, Plummer B, Levy S, *et al.* Exome sequencing supports a de novo mutational paradigm for schizophrenia. Nat Genet 2011, 43: 864–868.

[74] Gauthier J, Champagne N, Lafreniere RG, Xiong L, Spiegelman D, Brustein E, *et al.* De novo mutations in the

gene encoding the synaptic scaffolding protein SHANK3 in patients ascertained for schizophrenia. Proc Natl Acad Sci U S A 2010, 107: 7863–7868.

[75] Tarabeux J, Champagne N, Brustein E, Hamdan FF, Gauthier J, Lapointe M, *et al.* De novo truncating mutation in Kinesin 17 associated with schizophrenia. Biol Psychiatry 2010, 68: 649–656.

[76] Awadalla P, Gauthier J, Myers RA, Casals F, Hamdan FF, Griffing AR, *et al.* Direct measure of the de novo mutation rate in autism and schizophrenia cohorts. Am J Hum Genet 2010, 87: 316–324.

[77] 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.

[78] Girard SL, Gauthier J, Noreau A, Xiong L, Zhou S, Jouan L, *et al.* Increased exonic de novo mutation rate in individuals with schizophrenia. Nat Genet 2011, 43: 860–863.

[79] 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.

[80] 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.

[81] Mulle JG, Pulver AE, McGrath JA, Wolyniec PS, Dodd AF, Cutler DJ, *et al.* Reciprocal duplication of the williams-beuren syndrome deletion on chromosome 7q11.23 is associated with schizophrenia. Biol Psychiatry 2014, 75: 371–377.

[82] Huang J, Perlis RH, Lee PH, Rush AJ, Fava M, Sachs GS, *et al.* Cross-disorder genomewide analysis of schizophrenia, bipolar disorder, and depression. Am J Psychiatry 2010, 167: 1254–1263.

[83] Williams HJ, Craddock N, Russo G, Hamshere ML, Moskvina V, Dwyer S, *et al.* Most genome-wide significant susceptibility loci for schizophrenia and bipolar disorder reported to date cross-traditional diagnostic boundaries. Hum Mol Genet 2011, 20: 387–391.

[84] Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet 2013, 45: 984–994.

[85] Williams HJ, Norton N, Dwyer S, Moskvina V, Nikolov I, Carroll L, *et al.* Fine mapping of ZNF804A and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder. Mol Psychiatry 2011, 16: 429–441.

[86] Whalley HC, Papmeyer M, Romaniuk L, Sprooten E, Johnstone EC, Hall J, *et al.* Impact of a microRNA MIR137 susceptibility variant on brain function in people at high genetic risk of schizophrenia or bipolar disorder. Neuropsychopharmacology 2012, 37: 2720–2729.

[87] Collins AL, Kim Y, Bloom RJ, Kelada SN, Sethupathy P, Sullivan PF. Transcriptional targets of the schizophrenia risk gene MIR137. Transl Psychiatry 2014, 4: e404.

[88] Devanna P, Vernes SC. A direct molecular link between the autism candidate gene RORa and the schizophrenia candidate MIR137. Sci Rep 2014, 4: 3994.

[89] Andreassen OA, Djurovic S, Thompson WK, Schork AJ, Kendler KS, O'Donovan MC, *et al.* Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. Am J Hum Genet 2013, 92: 197–209.

[90] Sellgren C, Frisell T, Lichtenstein P, Landen M, Askling J. The association between schizophrenia and rheumatoid arthritis: a nationwide population-based Swedish study on intraindividual and familial risks. Schizophr Bull 2014, 40: 1552–1559.

[91] Zhao Z, Xu J, Chen J, Kim S, Reimers M, Bacanu SA, *et al.* Transcriptome sequencing and genome-wide association analyses reveal lysosomal function and actin cytoskeleton remodeling in schizophrenia and bipolar disorder. Mol Psychiatry 2014.

[92] Lencz T, Knowles E, Davies G, Guha S, Liewald DC, Starr JM, *et al.* Molecular genetic evidence for overlap between general cognitive ability and risk for schizophrenia: a report from the Cognitive Genomics consorTium (COGENT). Mol Psychiatry 2014, 19: 168–174.

[93] Mowry BJ, Gratten J. The emerging spectrum of allelic variation in schizophrenia: current evidence and strategies for the identification and functional characterization of common and rare variants. Mol Psychiatry 2013, 18: 38-52.

[94] McClellan JM, Susser E, King MC. Schizophrenia: a common disease caused by multiple rare alleles. Br J Psychiatry 2007, 190: 194–199.

[95] Veltman JA, Brunner HG. De novo mutations in human genetic disease. Nat Rev Genet 2012, 13: 565–575.

[96] Ku CS, Polychronakos C, Tan EK, Naidoo N, Pawitan Y, Roukos DH, *et al.* A new paradigm emerges from the study of de novo mutations in the context of neurodevelopmental disease. Mol Psychiatry 2013, 18: 141–153.

[97] Genome-wide association study identifies five new schizophrenia loci. Nat Genet 2011, 43: 969–976.

[98] Selten JP, van d, V, Rutten BP, Cantor-Graae E. The social defeat hypothesis of schizophrenia: an update. Schizophr Bull 2013, 39: 1180–1186.

[99] Betts KS, Williams GM, Najman JM, Scott J, Alati R. Maternal prenatal infection, early susceptibility to illness and adult psychotic experiences: a birth cohort study. Schizophr Res 2014, 156: 161–167.

[100] Khandaker GM, Zimbron J, Lewis G, Jones PB. Prenatal

maternal infection, neurodevelopment and adult schizophrenia: a systematic review of population-based studies. Psychol Med 2013, 43: 239–257.

[101] Nielsen PR, Laursen TM, Mortensen PB. Association between parental hospital-treated infection and the risk of schizophrenia in adolescence and early adulthood. Schizophr Bull 2013, 39: 230–237.

[102] Demontis D, Nyegaard M, Buttenschon HN, Hedemand A, Pedersen CB, Grove J, *et al.* Association of GRIN1 and GRIN2A-D with schizophrenia and genetic interaction with maternal herpes simplex virus-2 infection affecting disease risk. Am J Med Genet B Neuropsychiatr Genet 2011, 156B: 913–922.

[103] Borglum AD, Demontis D, Grove J, Pallesen J, Hollegaard MV, Pedersen CB, *et al.* Genome-wide study of association and interaction with maternal cytomegalovirus infection suggests new schizophrenia loci. Mol Psychiatry 2014, 19: 325–333.

[104] Ober C, Loisel DA, Gilad Y. Sex-specific genetic architecture of human disease. Nat Rev Genet 2008, 9: 911–922.

[105] Shifman S, Johannesson M, Bronstein M, Chen SX, Collier DA, Craddock NJ, *et al.* Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. PLoS Genet 2008, 4: e28.

[106] Liu Y, Chen PL, McGrath J, Wolyniec P, Fallin D, Nestadt G, *et al.* Replication of an association of a common variant in the Reelin gene (RELN) with schizophrenia in Ashkenazi Jewish women. Psychiatr Genet 2010, 20: 184–186.

[107] Li W, Song X, Zhang H, Yang Y, Jiang C, Xiao B, *et al.* Association study of RELN polymorphisms with schizophrenia in Han Chinese population. Prog Neuropsychopharmacol Biol Psychiatry 2011, 35: 1505–1511.

[108] Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet 2013, 381: 1371–1379.

[109] Moskvina V, Craddock N, Holmans P, Nikolov I, Pahwa JS, Green E, *et al.* Gene-wide analyses of genome-wide association data sets: evidence for multiple common risk alleles for schizophrenia and bipolar disorder and for overlap in genetic risk. Mol Psychiatry 2009, 14: 252–260.

[110] Bergen SE, O'Dushlaine CT, Ripke S, Lee PH, Ruderfer DM, Akterin S, *et al.* Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. Mol Psychiatry 2012, 17: 880–886.

[111] Rietschel M, Mattheisen M, Degenhardt F, Muhleisen TW, Kirsch P, Esslinger C, *et al.* Association between genetic variation in a region on chromosome 11 and schizophrenia in large samples from Europe. Mol Psychiatry 2012, 17: 906–917.

[112] Ingason A, Kirov G, Giegling I, Hansen T, Isles AR, Jakobsen KD, *et al.* Maternally derived microduplications at 15q11-q13: implication of imprinted genes in psychotic illness. Am J Psychiatry 2011, 168: 408–417.

[113] Georgieva L, Rees E, Moran JL, Chambert KD, Milanova V, Craddock N, *et al.* De novo CNVs in bipolar affective disorder and schizophrenia. Hum Mol Genet 2014, 23: 6677–6683.

[114] McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S, *et al.* Microduplications of 16p11.2 are associated with schizophrenia. Nat Genet 2009, 41: 1223–1227.

[115] Guha S, Rees E, Darvasi A, Ivanov D, Ikeda M, Bergen SE, *et al.* Implication of a rare deletion at distal 16p11.2 in schizophrenia. JAMA Psychiatry 2013, 70: 253–260.

[116] Moreno-De-Luca D, Mulle JG, Kaminsky EB, Sanders SJ, Myers SM, Adam MP, *et al.* Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. Am J Hum Genet 2010, 87: 618–630.