Schizophrenia is a common mental disorder with high heritability and strong genetic heterogeneity. Common disease-common variants hypothesis predicts that schizophrenia is attributable in part to common genetic variants. However, recent studies have clearly demonstrated that copy number variations (CNVs) also play pivotal roles in schizophrenia susceptibility and explain a proportion of missing heritability. Though numerous CNVs have been identified, many of the regions affected by CNVs show poor overlapping among different studies, and it is not known whether the genes disrupted by CNVs contribute to the risk of schizophrenia. By using cumulative scoring, we systematically prioritized the genes affected by CNVs in schizophrenia. We identified 8 top genes that are frequently disrupted by CNVs, including NRXN1, CHRNA7, BCL9, CYFIP1, GJA5, NDE1, SNAP29, and GJA5. Integration of genes affected by CNVs with known schizophrenia susceptibility genes (from previous genetic linkage and association studies) reveals that many genes disrupted by CNVs are also associated with schizophrenia. Further protein-protein interaction (PPI) analysis indicates that protein products of genes affected by CNVs frequently interact with known schizophrenia-associated proteins. Finally, systematic integration of CNVs prioritization data with genetic association and PPI data identifies key schizophrenia candidate genes. Our results provide a global overview of genes impacted by CNVs in schizophrenia and reveal a densely interconnected molecular network of de novo CNVs in schizophrenia. Though the prioritized top genes represent promising schizophrenia risk genes, further work with different prioritization methods and independent samples is needed to confirm these findings. Nevertheless, the identified key candidate genes may have important roles in the pathogenesis of schizophrenia, and further functional characterization of these genes may provide pivotal targets for future therapeutics and diagnostics.

Key words: schizophrenia/copy number variation/prioritization/integrative analysis/NRXN1/CHRNA7

Introduction

Schizophrenia is a common psychiatric disorder with high heritability and strong genetic heterogeneity. Genetic epidemiology studies reveal that about 0.5%–1% of world population was affected by schizophrenia. Though it is well established that schizophrenia occurs at high frequency and has a strong genetic component (with heritability estimates about 0.80), the genetic architecture of schizophrenia is largely unknown. As the predominant working hypothesis, the common disease-common variants (CD-CV) hypothesis suggests that schizophrenia is associated primarily with common genetic variants. Based on the CD-CV hypothesis, most of the genetic association studies of schizophrenia have focused on common genetic variation and numerous risk variants have been identified. The CD-CV hypothesis also constitutes the rationale of genome-wide association studies (GWAS), in which up to millions of single-nucleotide polymorphisms (SNPs) were assessed in thousands of individuals. Because both the power and throughput have greatly increased in GWAS, it has been widely expected...
that the genetic basis of schizophrenia can be resolved by large-scale GWAS. However, despite large-scale genome-wide meta-analysis of linkage and association studies of schizophrenia have been conducted and multiple promising candidate variants have been identified,15-19 most identified variants individually or in combination confer relatively small contribution to schizophrenia risk and only explain a small proportion of heritability, leading most of the heritability unaccounted for. For example, it is estimated that less than 2% of the 80% heritability of major psychiatric diseases is attributable to genes identified by linkage and association studies.20 Therefore, finding the missing heritability is a major challenge in genetic study of schizophrenia.

Since accumulating evidence indicates that common variants (individually or in combination) only account for a small proportion of schizophrenia heritability, it is likely that rare variants (including rare SNPs and copy number variations [CNVs]) may also contribute to schizophrenia susceptibility.21 Consistent with this, recent studies have clearly shown that CNVs have pivotal roles in schizophrenia.22-27 Complementary to the CD-CV hypothesis, the common disease-rare variants hypothesis21 suggests that schizophrenia is in part attributable to rare de novo CNVs with relatively high penetrance. In fact, it is likely that both common and rare variants contribute to risk of schizophrenia and lead to the strong genetic heterogeneity of schizophrenia.28 CNVs are structural genomic variants that mainly consist of insertions, deletions, duplications, and translocations. The size of CNVs ranges from 1 kilobase (kb) to several megabase (Mb) pairs. Among the structural variations identified in schizophrenia, the 22q11 deletion is one of the most representative CNVs.29 A microdeletion in chromosome 22 (22q11.2 deletion syndrome) causes velocardiofacial syndrome (also known as DiGeorge syndrome),30 a severe disorder with variable phenotypes. The size of this deletion usually ranges from 1.5 to 3 Mb, with about 35–60 known genes being deleted.31 It is well established that the 22q11 deletion is strongly associated with schizophrenia, and about 20%–30% adults with this CNV have schizophrenia.29,31 Of note, the catechol-O-methyltransferase (COMT), one of the best characterized and promising schizophrenia candidate genes, is also located in 22q11 region. Multiple genetic, functional, and animal studies have shown that COMT is significantly associated with schizophrenia.32,33 In addition to COMT, other promising candidate genes such as DISC1 and PED4B also have been identified by cytogenetic analyses.34-36

Though many large-scale genome-wide scans have been performed and numerous CNVs have been identified, many of the regions affected by CNVs showed poor overlapping among different studies. So far, only a small number of specific CNVs are found to be enriched in schizophrenia cases, including duplication at 16p11 and deletions at 1q21, 15q11, 15q13, and 22q11.23,37-42 It is likely that these identified CNVs (deletions and duplications) disrupt the normal function of adjacent genes, which subsequently increases risk of schizophrenia. Thus, it is important to investigate whether the genes disrupted by CNVs contribute to risk of schizophrenia. Nevertheless, because there is no systematic and comprehensive overview of all genes disrupted by CNVs in schizophrenia, it is difficult to assess whether and how the genes affected by CNVs confer risk of schizophrenia. To better understand the role of CNVs in the etiology of schizophrenia, 2 basic questions need to be addressed. The first is to generate the global profile of genes affected by CNVs in schizophrenia. The second is to investigate the role of the genes affected by CNVs in schizophrenia. To this aim, we systematically prioritized all of the genes affected by CNVs. We identified 8 top genes that are frequently disrupted by CNVs in schizophrenia cases in different studies, including NRXN1, CHRNA7, BCL9, CYFIP1, GJA8, NDE1, SNAP29, and GIA5. Integration of prioritization data with genetic association and protein-protein interaction (PPI) data further support the importance of these genes in schizophrenia susceptibility. Our results provide a global overview of genes affected by CNVs in schizophrenia and identify promising candidate genes for schizophrenia. Further functional characterization of the identified key genes may provide new insight into the etiology of schizophrenia and potential therapeutic targets.

Methods

Systematic Literature Search

To systematically evaluate and prioritize the genes affected by CNVs in schizophrenia, we first extracted all of the CNV studies for schizophrenia from the PubMed (http://www.ncbi.nlm.nih.gov/pubmed) with the search terms “schizophrenia” and “copy number variation.” CNV studies published before April 25, 2013 were included in this analysis. Original CNV studies were included if they met one of the following criteria: (1) GWAS of CNV performed in schizophrenia cases and controls; (2) CNV studies on candidate regions performed in individuals with schizophrenia and normal subjects; and (3) Structural variation studies conducted in schizophrenia cases and normal controls. The full text of each original study was read carefully to ensure eligibility. In total, 32 original studies were included in this study (supplementary table S1).

Extraction of Genes Affected by CNVs

For each included study, we first obtained all of the CNVs identified in schizophrenia cases. The genes affected by the identified CNVs were then manually extracted. Because we focus on genes that may confer risk of schizophrenia, only CNVs identified in individuals with schizophrenia were considered.
**Cumulative Scoring**

After extracting all of the genes affected by CNVs from all eligible studies, we used a cumulative scoring method to prioritize genes affected by CNVs. The rationale of our cumulative scoring approach is that if the disruption of a specific gene is associated with schizophrenia, it may be detected in several independent studies. Based on the frequency of their disruption (ie, impacted by CNVs) in schizophrenia cases in 32 eligible studies, these genes were given different scores. To avoid bias, we weighted genes identified by genome-wide CNV studies and regional studies differently. In brief, genes (ie, affected by CNVs) identified by genome-wide studies were given higher weight (2 points), whereas genes from regional studies were assigned relatively lower weight (1 point). For example, if the disruption of a specific gene by CNVs was observed in individuals of schizophrenia in 10 genome-wide CNV studies, the probability that this gene is associated with schizophrenia is high. Therefore, this gene is assigned to 20. If disruption of a gene is detected only once in a regional study of 32 included studies, the likelihood that this gene is significant associated with schizophrenia is relatively low. Thus, this gene is assigned to 1.

In addition, we also adjusted the weights of the included studies based to their total sample size (cases plus controls). In brief, we categorized the included studies into 5 groups. Each group receives a different score based on their sample size. The detailed classification of the 5 categories is as follows: (1) sample size < 500, 0.1 point; (2) sample size 500–1000, 0.2 point; (3) sample size 1000–5000, 0.3 point; (4) sample size 5000–10 000, 0.4 point; and (5) sample size > 10 000, 0.5 point. In this scoring system, larger number represents higher likelihood that the gene under consideration is associated with schizophrenia.

**Compiling of a Comprehensive List of Known Schizophrenia Susceptibility Genes**

Integration of the genes affected by CNVs with known schizophrenia susceptibility genes will provide important information for gene prioritization and identification of disease-associated genes. To this aim, a comprehensive list of schizophrenia candidate genes is needed. We, therefore, systematically combined different data sets that contained well-characterized schizophrenia susceptibility genes. The first data set is from the SZGene database (http://www.szgene.org/) (supplementary table S2). The second data set is from the Schizophrenia Gene Resource (SZGR) database (http://bioinfo.mc.vanderbilt.edu/SZGR/) (supplementary tables S3–S6). The third data set is from recent work of Ayalew et al. (supplementary table S7). The fourth data set consists of manually curated top schizophrenia candidate genes that reached genome-wide significance level in recent GWAS of schizophrenia (supplementary table S8). Because many of genes from above databases are overlapping, we generated a comprehensive and nonoverlapping list of schizophrenia candidate genes by combining these data sets (supplementary table S9). It should be noted that none of these schizophrenia data sets used CNV data to prioritize the schizophrenia candidate genes. More detailed information about included schizophrenia susceptibility genes can be found in supplementary data.

**Integration of Prioritization Evidence With Known Schizophrenia Susceptibility Genes**

To further evaluate whether the prioritized genes confer risk of schizophrenia, we investigated the overlap between genes affected by CNVs and known schizophrenia candidate genes. We adjusted the weight of linkage and association studies and GWAS appropriately. If a gene was disrupted by CNV in schizophrenia cases and previous genetic linkage or association studies also support that this gene is associated with schizophrenia, the total score of this gene rises by 1 point. If a CNV affected gene is significantly associated with schizophrenia in GWAS of schizophrenia, this gene represents a high-confidence schizophrenia susceptibility gene, therefore, the total score of this gene rises by 3 points.

**Integration of Prioritization Data With GWAS Findings From Schizophrenia Psychiatric Genomics Consortium**

To explore whether the prioritized genes showed suggestive association with schizophrenia at the gene level, we also tested the association between these prioritized genes and schizophrenia in the Schizophrenia Psychiatric Genomics Consortium (PGC) sample (9394 cases and 12 462 controls). The association data were downloaded from the Schizophrenia PGC Web site, and gene-based P-values were calculated by using VEGAS. We applied a similar weight scheme used by Sun et al. Briefly, the prioritized genes were weighted based on their gene-based association significance level. If the gene-based P-value of a prioritized gene is greater than .01 and less than .05 (.01 < P < .05), then the total score of this gene rises by 1 point. If the gene-based P-value of a prioritized gene is less than .01 or .001, then the total score of this gene rises by 2 or 3 points.

**Prioritization of Genes Affected by CNVs by Using Endeavour**

We also utilized gene prioritization tool (Endeavour) to prioritize and rank genes affected by CNVs. The rationale for Endeavour prioritization is based on functional similarity to training gene list, ie, how similar a candidate gene is to a profile derived from genes already known to be involved in the processes (training genes). More detailed information about Endeavour prioritization can be found in supplementary methods.
Gene Ontology Analysis of Genes Affected by CNVs

To investigate whether genes affected by CNVs were enriched for specific functional categories, we performed Gene Ontology (GO) analysis. The “DAVID Bioinformatics Resources 6.7” (http://david.abcc.ncifcrf.gov) was used to conduct the primary GO analysis. GO terms—biological processes (GO_BP), cellular components (GO_CC), and molecular functions (GO_MF)—were used. The Benjamini-Hochberg procedure was used to correct the P-values of the over-represented GO terms.

PPI Analysis and Assessment of the Significance of the PPI Network

To further examine whether the prioritized genes were involved in schizophrenia, we performed PPI analysis. If a gene is implicated in schizophrenia, it may participate in the molecular network formed by schizophrenia susceptibility genes. The PPI networks (including direct and indirect) among genes affected by CNVs and known schizophrenia risk genes were extracted from InWeb, a well-characterized PPI database developed by Lage et al. To evaluate whether genes affected by CNVs are significantly connected via PPIs, permutation test was used to assess the significance of networks built from PPI data. For more details about PPI network construction, classification, and significance assessment, please refer to supplementary methods and our previous work.

Network Analysis

The significance analysis of PPI network was performed by Disease Association Protein-Protein Link Evaluator (DAPPLE). In addition, CytoScape and different plugins implemented in CytoScape were used to perform network analysis. More detailed information about network topological analysis can be found in supplementary methods.

Systematic Integration of Prioritization Data With Genetic Association and PPI Data

To systematically prioritize and rank genes affected by CNVs, we integrated data from different sources, including data from CNV studies, genetic association information, and PPI data. Evidence from these different resources contributes to the overall score of the prioritized genes. More detailed information can be found in the supplementary methods.

Expression and Co-expression Analysis of the Top Prioritized Genes and Genes in the PPI Network

To investigate whether the top prioritized genes impacted by CNVs are expressed in the human brain, we studied the expression patterns of these genes in diverse human tissues by using the Gene Enrichment Profiler. The Gene Enrichment Profiler contains expression profiles for about 12 000 genes across 126 normal primary human tissues.

For the top prioritized genes and gene pairs in the PPI network, we also performed co-expression analysis by using RNA-sequencing-based expression data from the BrainSpan: Atlas of the Developing Human Brain (http://www.brainspan.org/). More detailed information about the co-expression analysis can be found in the supplementary data and study of Gulsuner et al.

Results

Extracting CNVs and Identification of Genes Affected by CNVs

To systematically evaluate and prioritize the genes affected by CNVs in schizophrenia, we first extracted all of the CNV studies for schizophrenia from the PubMed with the search terms “schizophrenia” and “copy number variation.” A total of 204 matched English publications were retrieved as of April 25, 2013. To ensure eligibility, we systematically evaluated these publications. The abstracts and full text of these studies were carefully read and evaluated. In total, 32 eligible original CNV studies were included in our analysis (supplementary table 1) and most of them were genome-wide CNV investigations (78%, 25 of 32). We carefully read and scrutinized each included study. The reported CNVs were then manually retrieved, which were mainly extracted from the text and supplementary data of the original publications. For each CNV identified in schizophrenia cases, the genes affected by this CNV were manually screened and extracted. In total, 1303 genes were mapped to the CNVs identified in schizophrenia cases. Among the 1303 identified genes, 1025 of them represent unique (nonoverlapping) genes (supplementary table 1).

Cumulative Scoring Reveals Top Genes Affected by CNVs

We first prioritized the 1303 genes affected by CNVs using cumulative scoring approach. The CNV affected genes were scored based on their disruption in schizophrenia cases in the 32 selected studies. In addition, we adjusted the weights by accounting for study sample size and study type (ie, GWAS or regional CNV studies; for details, please see “Methods” section). According to this scoring method, all of genes affected by CNVs were systematically prioritized and ranked based on their cumulative score. We found that NRXN1 ranked first among all of genes affected by CNVs (table 1). Ten independent studies found that NRXN1 was impacted by CNVs in schizophrenia cases. The cumulative score of NRXN1 is 18.4 points. CHRNA7 ranked second (table 1), with 6 independent studies revealing the disruption of...
Table 1. Systematic Prioritization and Integration of Genes Affected by CNVs Identify Top Candidate Genes for Schizophrenia

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Cumulative Scoring Evidencea</th>
<th>Genetic Linkage and Association Evidenceb</th>
<th>Gene-based Evidence (P-values From PGC GWAS)c</th>
<th>Physical Interaction Evidence (Interactor)d</th>
<th>Endeavour P-valuee</th>
<th>Total Scoref</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRXN1</td>
<td>18.4</td>
<td>0</td>
<td>1 (.042)</td>
<td>0</td>
<td>N/A</td>
<td>19.4</td>
</tr>
<tr>
<td>CHRNA7</td>
<td>13.9</td>
<td>1</td>
<td>0 (.28)</td>
<td>0</td>
<td>1.53 × 10⁻⁴</td>
<td>14.9</td>
</tr>
<tr>
<td>BCL9</td>
<td>9.5</td>
<td>3g</td>
<td>0 (.92)</td>
<td>1 (TCF4)</td>
<td>4.16 × 10⁻³</td>
<td>13.5</td>
</tr>
<tr>
<td>CYFIP1</td>
<td>11.8</td>
<td>0</td>
<td>0 (.27)</td>
<td>1 (SYN2)</td>
<td>3.28 × 10⁻¹</td>
<td>12.8</td>
</tr>
<tr>
<td>GJA8</td>
<td>11.8</td>
<td>1</td>
<td>0 (.73)</td>
<td>0</td>
<td>6.70 × 10⁻³</td>
<td>12.8</td>
</tr>
<tr>
<td>NDE1</td>
<td>10.7</td>
<td>0</td>
<td>1 (.038)</td>
<td>1 (KIF2A)</td>
<td>4.64 × 10⁻¹</td>
<td>12.7</td>
</tr>
<tr>
<td>SNAP29</td>
<td>10.1</td>
<td>1</td>
<td>0 (.70)</td>
<td>1 (STX1A)</td>
<td>N/A</td>
<td>12.1</td>
</tr>
<tr>
<td>GJA5</td>
<td>11.7</td>
<td>0</td>
<td>0 (.35)</td>
<td>0</td>
<td>2.03 × 10⁻¹</td>
<td>11.7</td>
</tr>
<tr>
<td>NTAN1</td>
<td>10.7</td>
<td>0</td>
<td>0 (.18)</td>
<td>0</td>
<td>4.70 × 10⁻¹</td>
<td>10.7</td>
</tr>
<tr>
<td>PRKAB2</td>
<td>9.5</td>
<td>0</td>
<td>0 (.94)</td>
<td>1 (NQO2)</td>
<td>7.38 × 10⁻¹</td>
<td>10.5</td>
</tr>
<tr>
<td>VIPR2</td>
<td>9.4</td>
<td>0</td>
<td>0 (.90)</td>
<td>1 (ADCYAP1)</td>
<td>1.35 × 10⁻²</td>
<td>10.4</td>
</tr>
<tr>
<td>PARK2</td>
<td>9.3</td>
<td>0</td>
<td>0 (.54)</td>
<td>1 (SYT11, GRIN2B, YWHAH)</td>
<td>2.55 × 10⁻¹</td>
<td>10.3</td>
</tr>
<tr>
<td>COMT</td>
<td>9.2</td>
<td>1</td>
<td>0 (.10)</td>
<td>1 (GRIN2B, GRIN1, NOS1, ERBB4, GRID1)</td>
<td>1.23 × 10⁻³</td>
<td>10.2</td>
</tr>
<tr>
<td>DLG2</td>
<td>6.7</td>
<td>0</td>
<td>2 (.00717)</td>
<td>1 (GRIN2B, GRIN1, NOS1, ERBB4, GRID1)</td>
<td>4.48 × 10⁻²</td>
<td>9.7</td>
</tr>
<tr>
<td>FMO5</td>
<td>9.5</td>
<td>0</td>
<td>0 (.89)</td>
<td>0</td>
<td>3.35 × 10⁻¹</td>
<td>9.5</td>
</tr>
<tr>
<td>NIPA1</td>
<td>9.5</td>
<td>0</td>
<td>0 (.58)</td>
<td>0</td>
<td>4.66 × 10⁻²</td>
<td>9.5</td>
</tr>
<tr>
<td>ACP6</td>
<td>9.5</td>
<td>0</td>
<td>0 (.83)</td>
<td>0</td>
<td>1.15 × 10⁻¹</td>
<td>9.5</td>
</tr>
<tr>
<td>TUBGCP5</td>
<td>9.5</td>
<td>0</td>
<td>0 (.23)</td>
<td>0</td>
<td>4.73 × 10⁻¹</td>
<td>9.5</td>
</tr>
<tr>
<td>MYH11</td>
<td>8.4</td>
<td>0</td>
<td>0 (.089)</td>
<td>1 (SLC1A2, SYNGR1)</td>
<td>1.14 × 10⁻¹</td>
<td>9.4</td>
</tr>
<tr>
<td>A2BP1</td>
<td>9.0</td>
<td>0</td>
<td>0 (.071)</td>
<td>0</td>
<td>N/A</td>
<td>9.0</td>
</tr>
<tr>
<td>KLHL22</td>
<td>8.8</td>
<td>0</td>
<td>0 (.36)</td>
<td>0</td>
<td>2.78 × 10⁻¹</td>
<td>8.8</td>
</tr>
<tr>
<td>RTN4R</td>
<td>6.6</td>
<td>1</td>
<td>0 (.11)</td>
<td>1 (RTN4)</td>
<td>3.71 × 10⁻³</td>
<td>8.6</td>
</tr>
<tr>
<td>ABCC1</td>
<td>8.4</td>
<td>0</td>
<td>0 (.43)</td>
<td>0</td>
<td>3.74 × 10⁻¹</td>
<td>8.4</td>
</tr>
<tr>
<td>PRODH</td>
<td>6.9</td>
<td>1</td>
<td>0 (.74)</td>
<td>0</td>
<td>3.10 × 10⁻³</td>
<td>7.9</td>
</tr>
<tr>
<td>ZDHHC8</td>
<td>6.6</td>
<td>1</td>
<td>0 (.96)</td>
<td>0</td>
<td>1.24 × 10⁻⁴</td>
<td>7.6</td>
</tr>
<tr>
<td>CRKL</td>
<td>6.6</td>
<td>0</td>
<td>0 (.16)</td>
<td>1 (ERBB3, ERBB4, NPAS3, ANK3)</td>
<td>4.38 × 10⁻²</td>
<td>7.6</td>
</tr>
<tr>
<td>THAP7</td>
<td>6.6</td>
<td>0</td>
<td>1 (.033)</td>
<td>0</td>
<td>8.51 × 10⁻²</td>
<td>7.6</td>
</tr>
<tr>
<td>DGC6L</td>
<td>6.6</td>
<td>0</td>
<td>0 (.25)</td>
<td>1 (DTNB1)</td>
<td>6.79 × 10⁻¹</td>
<td>7.6</td>
</tr>
</tbody>
</table>

Note: CNVs, copy number variations; GWAS, genome-wide association studies; NA, not available; PGC, Psychiatric Genomics Consortium.

*aCumulative scoring evidence is based on the frequency of their disruption in 32 eligible studies. In addition, study type (GWAS or regional CNV study) and sample size were also taken into account and weighted accordingly.

*bGenetic linkage and association evidence indicate whether the prioritized genes was reported to be associated with schizophrenia in previous genetic linkage or association studies.

*cPGC GWAS evidence is based on gene-based P-value calculated from Schizophrenia PGC data.

*dPhysical interaction evidence indicates whether proteins encoded by the prioritized genes interact with proteins encoded by known schizophrenia risk genes.

*eP-value is obtained by Endeavour prioritization tool.

*fTotal score shows the overall points of the prioritized genes. More work is needed to validate these results as a simple cumulative scoring method was used in this article.

*BCl9 is significantly associated with schizophrenia in a GWAS study. Statistically significant P-values (P < .05) are shown in bold.

CHRNA7 in schizophrenia cases. Thus, the cumulative score of CHRNA7 is 13.9 points. CYFIP1 and GJA8 ranked third (table 1). Five different studies showed that CYFIP1 and GJA8 were affected by CNVs in individuals with schizophrenia. And 22 other genes (GJA5, NDE1, NTAN1, SNAP29, BCL9, PARK2, FMO5, NIPA1, ACP6, TUBGCP5, VIPR2, A2BP1, KLHL22, MYH11, ABCC1, PRODH, DLG2, RTN4R, ZDHHC8, CRKL, THAP7, and DGC6L) ranked from 4th to 17th among the affected genes (table 1). Taken together, our cumulative scoring prioritization identified key promising candidate genes affected by CNVs in schizophrenia, including NRXN1, CHRNA7, CYFIP1, GJA8, NDE1, SNAP29, GJA5, BCL9, NTAN1, PRKAB2, VIPR2, PARK2, COMT, and DLG2 (table 1). These results strongly suggest that these identified top genes may have important roles in schizophrenia as they are frequently disrupted by CNVs in schizophrenia cases.
Integration of Prioritization Evidence With Known Schizophrenia Susceptibility Genes

Our cumulative scoring results revealed promising candidate genes for schizophrenia (table 1), eg, NRXN1, CHRNA7, COMT, BCL9, and VIPR2. These genes are frequently disrupted by CNVs in schizophrenia cases in different studies. Thus, they may represent authentic schizophrenia risk genes. To further test whether these genes have a role in schizophrenia, we integrated prioritization results from cumulative scoring with potential schizophrenia candidate genes. As our goal is to identify and prioritize the most promising candidate genes, only genes with high rankings (ie, genes with at least 6 points from cumulative scoring prioritization) were considered for further investigation.

We first generated a comprehensive list of well-characterized schizophrenia candidate genes based on previous human genetic studies, including linkage, association, convergent functional genomics, and recent GWAS of schizophrenia (supplementary table S9). This list was then used to match the genes affected by CNVs. If a CNV disrupted gene is found on the list of potential schizophrenia susceptibility genes, suggesting previous human genetic studies support the association between this gene and schizophrenia, the total score of this gene is increased by 1 point (for linkage and association studies) or 3 points (for GWAS). Through integrating cumulative prioritization data with known schizophrenia susceptibility genes, 8 genes were found in the comprehensive list of known schizophrenia susceptibility genes, including CHRNA7, GJA8, SNAP29, BCL9, COMT, RTN4R, PRODH, and ZDHHC8 (table 1). Of note, a recent study also found that BCL9 is significantly associated with schizophrenia at genome-wide significance level. Collectively, these results imply that these 8 genes may represent promising candidate genes for schizophrenia as 2 independent lines of evidence (ie, evidence from CNV and genetic association or linkage studies) support implication of these genes in schizophrenia.

Association Between Top Prioritized Genes and Schizophrenia in Schizophrenia PGC Sample

For genes with a cumulative prioritization score of 6 and more, we further tested the suggestive association between these genes and schizophrenia by utilizing gene-based P-values. Genetic association data from the Schizophrenia PGC (9394 cases and 12 462 controls) were used, and gene-based P-values were calculated. As CNV data have shown the potential roles of these genes in schizophrenia, therefore, this association may provide additional evidence for the implication of these genes in schizophrenia. Based on gene-based P-values, 4 genes (NRXN1, NDE1, DLG2, and THAP7) showed suggestive association with schizophrenia in the Schizophrenia PGC sample at gene level (table 1). Interestingly, NRXN1 was frequently disrupted by CNVs in schizophrenia cases in different studies. In fact, NRXN1 ranked no first in all of genes affected by CNVs (table 1) in individuals with schizophrenia. Taken together, these results strongly suggest that NRXN1 is an authentic schizophrenia susceptibility gene.

Integration of PPI Evidence With Prioritization Data

Recent studies have shown that schizophrenia susceptibility genes encode a highly interconnected PPI network. As a matter of fact, many basic biological functions are executed by protein complex. That is, many proteins execute their biological functions through PPIs, and disruption of any member of the protein complex may lead to similar functional consequences. In fact, we have shown that proteins encoded by schizophrenia susceptibility genes are significantly physically interacted and may perturb common biological networks. Therefore, investigating the physical interaction between proteins encoded by genes affected by CNVs and known schizophrenia susceptibility genes will provide pivotal information. Accumulating evidence has clearly indicated that PPI plays crucial roles in the identification of schizophrenia candidate genes. Of the best examples is the physical interaction between DISC1 and PDE4B. DISC1 and PDE4B were identified by cytogenetic analyses, and both of them are well-characterized candidate genes for schizophrenia. Intriguingly, previous study found that the physical interaction between DISC1 and PDE4B is important for schizophrenia susceptibility.

By using the well-defined PPI data from InWeb and CytoScape, we systematically investigated the PPI between proteins encoded by genes affected by CNVs and known schizophrenia susceptibility genes. Based on the PPI information, the candidate genes were prioritized by the “guilt by association” principle, ie, disease-associated genes tend to interact (locate closer) with each other than random proteins in the PPI network. If the protein product of a candidate gene is physically interacted with proteins encoded by known schizophrenia susceptibility genes, the total score of this gene rises by 1 point. Among 28 genes with cumulative prioritization score of 6 and more, we found that protein products of 12 genes showed physical interaction with proteins encoded by known schizophrenia susceptibility genes (table 1). The PPI results revealed several interesting interactions, eg, the interaction between BCL9 and TCF4 (both of BCL9 and TCF4 reached genome-wide significance level in recent GWAS of schizophrenia), the interaction between VIPR2 and ADCYAPI, and the interaction between DLG2 and ERBB4. These PPI results indicate that protein products encoded by genes affected by CNVs and known schizophrenia susceptibility genes tend to form protein complexes, which may play an important role in the pathogenesis of schizophrenia. Collectively, our PPI results provide further evidence...
for the involvement of these CNVs affected genes in schizophrenia.

**Genes Affected by CNVs Encode an Interconnected Molecular Network**

A total of 1025 unique (nonoverlapping) genes were identified to be disrupted by CNVs in schizophrenia cases (supplementary table S1). However, the physical interaction between protein products encoded by these genes is not known. Our recent work has shown that proteins encoded by top schizophrenia susceptibility genes formed a highly significant interconnected network. To examine whether protein products encoded by genes affected by CNVs are physically interacted, we investigated the PPI between proteins encoded by genes affected by CNVs. As described in our previous studies, PPI network was generated by using the DAPPLE (http://www.broadinstitute.org/mpg/dapple/dapple.php) and CytoScape. We found that protein products encoded by genes affected by CNVs form an interconnected network (figure 1). We further calculated the P-values of genes in the PPI network based on the probability that by chance the seed protein would be as directly connected to other seed proteins as is observed. We found that many proteins in the PPI network were significantly connected to other proteins (supplementary table S10), including PRKAB2 ($P = .024$), GJA8 ($P = .008$), and DLG2 ($P = .001$). This result indicates that genes disrupted by CNVs encode an interconnected molecular network, suggesting these CNVs perturb common molecular networks that modulate schizophrenia risk.

To further identify the protein complex encoded by genes that were disrupted by CNVs, we analyzed the PPI network. The whole PPI network was clustered into distinct modular clusters and multiple highly interconnected subnetworks were identified (figure 2). Intriguingly, several top candidate genes for schizophrenia also participate in these densely interconnected subnetworks, e.g., NRXN1,
NDE1, GJA8, PARK2, and DLG2 (figure 2). These results reveal that protein products of genes affected by CNVs tend to form protein complexes. In addition, these results also indicate genes affected by CNVs encode an interconnected molecular network.

Synaptic Transmission-Related Genes Were Enriched in Genes Impacted by CNVs in Schizophrenia

To examine whether genes impacted by CNVs in schizophrenia were enriched for specific functional categories, we performed GO analysis to identify the over-represented GO terms in the gene set (1025 unique genes) impacted by CNVs in schizophrenia. By using biological process as GO term, we identified 8 functional categories that were enriched among CNVs in schizophrenia (corrected P-value < .05; table 2). Interestingly, we found the most significantly enriched functional term was synaptic transmission (corrected P-value = 2.10 × 10^{-6}; table 2), suggesting synaptic transmission-related genes were frequently impacted by CNVs in schizophrenia. In addition to synaptic transmission, several other GO terms were also found to be over-represented in genes impacted by CNVs in schizophrenia, eg, cell-cell signaling (corrected P-value = 3.03 × 10^{-4}) and regulation of neurotransmitter levels (corrected P-value = .045). Our results further support previous observations that synaptic transmission was strongly impacted by CNVs in schizophrenia.26,66

Prioritization of Genes Impacted by CNVs in Schizophrenia by Endeavour

Because our above prioritization of genes impacted by CNVs in schizophrenia was mainly based on the disruption...
of these genes in different CNV studies and overlapping with known schizophrenia susceptibility genes and PPI data, we further prioritized genes impacted by CNVs in schizophrenia by using Endeavour, a well-characterized candidate gene prioritization tool. Endeavour uses information from the training genes (known schizophrenia susceptibility genes, supplementary table S9) to prioritize and rank seed genes (genes need to be prioritized, supplementary table S1) based on functional similarity. We found most of the top candidate genes captured by our cumulative prioritization were also identified and prioritized by Endeavour (table 1). These results strongly support that the genes prioritized by our cumulative prioritization may represent high-confidence candidate genes for schizophrenia.

Comprehensive Prioritization and Systematic Integration Identify Key Candidate Genes for Schizophrenia

A total of 1025 unique genes were impacted (due to deletion or duplication) by CNVs in individuals with schizophrenia. In order to prioritize and identify the most promising genes, we systematically integrated the comprehensive information from different sources. In this systematic prioritization process, evidence from different sources was used to calculate the overall score of a candidate gene. As our primary aim is to identify the most promising schizophrenia risk genes impacted by CNVs, evidence from CNVs studies in schizophrenia was given higher weights and play central roles in the prioritization (up to 18.4 points from CNVs studies; table 1). While evidences from previous human genetic studies and PPI were used to corroborate the prioritized genes, therefore, each of them contributes 1 point (GWAS supported genes contribute 3 points) to the overall score. In addition, GWAS supported suggestive association was also assigned relatively high scores (up to 3 points) as GWAS is a robust method to identify the potential schizophrenia genes.

Based on above prioritization strategy, we generated a landscape of top candidate genes affected by CNVs in schizophrenia (table 1). In total, 28 promising candidate genes were identified (table 1). NRXN1 has the highest overall score (19.4 points), therefore, ranked first among all of genes. CHRNA7 ranked second. BCL9 ranked third. CYFIP1 and GJA8 ranked fourth. And NDE1, SNAP29, GJA5, NTAN1, PRKAB2, VIPR2, PARK2, COMT, and Dlg2 ranked from 5th to 13. Of note, up to 10 independent CNV studies found that NRXN1 was affected by CNVs in schizophrenia cases. And 6 different studies reported that CHRNA7 and SNAP29 were impacted by CNVs in schizophrenia cases. Considering the nature of low frequency of CNV, the probability that these genes were impacted by CNVs in multiple independent studies is extremely low. Thus, these results strongly suggest that these 3 genes are the most promising candidates for schizophrenia.

We assigned genes that reached genome-wide significance level 3 points and candidate genes 1 point in our study (column 3 in table 1). To further investigate the impact of different weight schemes on the final results, we tested another 2 weight schemes. First, we reduced the weight of GWAS genes (ie, genes reached genome-wide significance level were assigned to 2 points and candidate genes were assigned 1 point). Compared with the results listed in table 1, we found that only 3 out of 28 genes show slightly different ranking (BCL9, CYFIP1, and GJA8) by this weight scheme (table 1 and supplementary table S1). Second, we increased the weight of GWAS genes (ie, genes reached genome-wide significance level were assigned 5 points and candidate genes were assigned 1 point). Again, we found that only the ranking of 2 genes (CHRNA7 and BCL9) were slightly changed (table 1 and supplementary table S12). Therefore, adjusting the weight of GWAS and candidate genes will not affect the final prioritization results significantly.

In addition, because genes from genome-wide and regional CNV studies were assigned different scores (ie, genes from genome-wide CNV studies were given higher...
weights [2 points] and genes from regional CNV studies were given lower weight [1 point]), bias may be introduced into the prioritization results. To validate our weight methods, we performed additional analyses through using different weight schemes (It should be noted that only the weight scores of genome-wide and regional CNV studies changed, all of other conditions remain unaltered.). First, we reanalyzed the data by using the simplest weight scheme, i.e., genes from genome-wide CNV studies and regional studies were given the same weight (1 point). We found that the results obtained from this weight scheme showed high degree of similarity with the results listed in table 1 (supplementary table S13). As a matter of fact, we found that among the top 7 prioritized genes, 6 genes were same when the 2 weight schemes were used (table 1 and supplementary table S13).

Second, we also tested another weight scheme. Briefly, genes (i.e., affected by CNVs) identified by genome-wide studies were given higher weight (3 points), whereas genes from regional studies were assigned relatively lower weight (1 point). Again, the results obtained from this weight scheme showed high degree of similarity with the results listed in table 1 (supplementary table S13).

Finally, we excluded the regional studies and performed the prioritization using genes from genome-wide CNV studies. Once again, we found that prioritization results from genome-wide CNV studies showed high degree of similarity with the results reported in table 1. In fact, compared with the results listed in table 1, we found that 93% (26 out of 28) of top prioritized genes were same when regional CNV studies were excluded. Taken together, these results strongly suggest that most of the top prioritized genes can be ranked similarly regardless of the weight schemes (table 1 and supplementary table S15). It should be noted that a large proportion of these top genes were also found to be associated with schizophrenia in previous studies. For example, multiple previous investigations have reported the association between NRXN1, CHRNA7, SNAP29, COMT, BCL9, ERBB4, and schizophrenia.7,32,67–80 Collectively, these convergent lines of evidence support these prioritized top genes may have pivotal roles in schizophrenia susceptibility.

**Top Prioritized CNV Affected Genes Are Preferentially Expressed in Human Brain**

To further explore the potential biological function of the top prioritized genes (from prioritization of genes affected by CNVs), we investigated the expression pattern of these prioritized genes in diverse human tissues. We found that most of the top prioritized genes are preferentially expressed in central nervous system (figure 3), suggesting these genes may play a role in brain function. Interestingly, we also noticed that a small subset of the top prioritized genes is abundantly expressed in immune tissues (supplementary figure S1). These expression results further support that these top prioritized genes may have important roles in the pathogenesis of schizophrenia. In addition, these results also suggest the pivotal roles of immune-related genes in schizophrenia, which is consistent with previous observations.56,81–86

We further investigated whether top prioritized genes and gene pairs in PPI network are co-expressed in human brain tissues. If 1 gene pair showed significant (Pearson correlation P-value < .05) co-expression in any tissue at any developmental stage, this gene pair was considered significant co-expression. We found many top prioritized genes are co-expressed in human brain (supplementary table S16). For each gene pair in the PPI network, we also performed co-expression analysis. Intriguingly, we found most of the gene pairs in the PPI network showed significant co-expression in human brain tissues (supplementary tables S17–S19). The chance that so many gene pairs showed significant co-expression is very low (P < 1 × 10^{-5}, Chi-square test), suggesting gene pairs in PPI network tend to co-express in human brain. In summary, our results prioritized and ranked the promising schizophrenia susceptibility genes impacted by CNVs, which provide a starting point for further functional studies of these genes in schizophrenia.

**Discussion**

Though intensive studies have been performed and significant progress has been made in past decades, the high heritability and strong genetic heterogeneity still post as a major challenge to the genetic dissection of schizophrenia. GWAS of schizophrenia have identified multiple promising candidate genes. However, they only explain a small fraction of heritability. Therefore, more work is needed to explore the missing heritability. Recent investigations have clearly demonstrated the involvement of CNVs in schizophrenia, strongly suggesting that CNVs may account for some of the unexplained heritability of schizophrenia. As one major source of genetic variation, CNVs play important roles in phenotypic differences and disease susceptibility. Compared with SNPs, CNVs have lower frequency and higher penetrance. The large effect sizes of CNVs provide good opportunities to identify the promising candidate genes for schizophrenia. Though multiple studies of CNVs in schizophrenia have been performed in different populations and many potential candidates have been identified, a large proportion of identified CNVs were unique events with little overlap between studies. Therefore, a systematic and global overview of all CNVs involved in the etiology of schizophrenia is needed. In this study, we systematically prioritized all of genes impacted by CNV in schizophrenia. Our results provide a global view of genes affected by CNVs in schizophrenia.

The identified top genes represent the most promising candidate genes as they are repeatedly impacted by CNVs
in individuals with schizophrenia in different studies. In addition, human genetic studies and PPI data also support the implication of these top genes in schizophrenia. Expression analysis further reveals that most of these top genes were highly expressed in human brain. These convergent lines of evidence strongly suggest that the prioritized genes are high-confidence candidates for schizophrenia. Further function investigation (eg, knockout study in mice) of these genes will provide pivotal information on the pathophysiology of schizophrenia and potential therapeutic targets.

One of the most interesting genes is NRXN1, which ranked highest among all of genes affected by CNVs in schizophrenia. Deletion of NRXN1 was frequently observed in schizophrenia cases in multiple different studies, suggesting NRXN1 is likely an authentic schizophrenia risk gene. NRXN1 belongs to neurexin, an important gene family that functions in the vertebrate nervous system as cell adhesion molecules and receptors. Neurexins are localized to presynaptic terminals and interact with postsynaptic cell adhesion molecules such as neuroligins (encoded by NLGN family), therefore, connect presynaptic and postsynaptic neurons at synapses. Previous investigations have revealed that neurexins are involved in synapse formation and mediate signaling across the synapse. In addition, neurexins also play an important...
role in the release of neurotransmitters from presynaptic vesicles and shape the properties of neural networks by specifying synaptic functions. Intriguingly, genetic association study also found that \textit{NRXN1} is significantly associated with schizophrenia,\(^6\) further supporting the potential role of \textit{NRXN1} in schizophrenia susceptibility.

The \textit{CHRNA7} gene encodes the neuronal acetylcholine receptor subunit \(\alpha\)-7. The nicotinic acetylcholine receptors are ligand-gated ion channels that mediate fast signal transmission at synapses. Multiple genetic association studies have found that genetic variations in \textit{CHRNA7} gene are significantly associated with schizophrenia.\(^6\)-\(^8\) The \textit{BCL9} gene is involved in Wnt signaling pathway\(^9\) and \textit{CYFIP1} gene plays an important role in synapse function, suggesting schizophrenia is likely caused by synaptic dysfunction. In fact, our GO analysis also revealed that the synaptic transmission genes are significantly enriched among genes affected by CNVs. Consistent with our results, multiple previous investigations also documented the dysfunction of synapse-related genes in schizophrenia.\(^6\)-\(^6\)

Our previous studies revealed that top schizophrenia susceptibility genes encode a highly interconnected PPI network\(^6\)-\(^7\), suggesting perturbations of common underlying molecular processes or pathways that modulate risk to schizophrenia. We further investigated the PPI among genes impacted by CNVs in this study. Again, we found that genes affected by CNVs in schizophrenia encode an interconnected molecular network, further corroborating our previous results. The finding that schizophrenia risk genes encode a densely interconnected molecular network may provide both theoretical and practical significance. First, it may help to elucidate the molecular mechanisms of schizophrenia. As many basic biological functions are executed by protein complexes, studying the schizophrenia risk genes in the context of biological process or molecular pathway may provide important information on the pathogenesis of schizophrenia. Second, it may provide new perspectives for design of antipsychotic drugs, eg, to design drugs to target pathways rather than a specific gene. Finally, it may provide a novel explanation for the strong genetic heterogeneity of schizophrenia, ie, dysfunction of any gene in this molecular network will lead to same functional consequences that eventually contribute to risk of schizophrenia.

In summary, we generated the first landscape of genes impacted by CNVs in schizophrenia. Our results provide a comprehensive and global view of genes affected by CNVs in schizophrenia. Importantly, we prioritized all of the genes impacted by CNVs by comprehensive integration of different evidence from different sources. The identification of these promising candidate genes provides a starting point for further functional studies of these genes in schizophrenia.

### Supplementary Material

**Supplementary material** is available at [http://schizophreniabulletin.oxfordjournals.org](http://schizophreniabulletin.oxfordjournals.org).

### Funding

National Natural Science Foundation of China (81271006); Hangzhou City Health Science Foundation (20120633B01); and Zhejiang Provincial Science Foundation (2010GZY0089).

### Acknowledgments

Schizophrenia PGC data were obtained from Ricopili ([http://www.broadinstitute.org/mpg/ricopili/](http://www.broadinstitute.org/mpg/ricopili/)). We thank members of Schizophrenia PGC and Stephan Ripke, who developed the Ricopili. We thank Jimmy Z. Liu for his help in gene-based association test. We are grateful to the editor and 2 anonymous reviewers for their constructive comments. The authors have declared that there are no conflicts of interest in relation to the subject of this study.

### References

A Landscape of Copy Number Variations in Schizophrenia


