Strong synaptic transmission impact by copy number variations in schizophrenia


Schizophrenia is a psychiatric disorder with onset in late adolescence and unclear etiology characterized by both positive and negative symptoms, as well as cognitive deficits. To identify copy number variations (CNVs) that increase the risk of schizophrenia, we performed a whole-genome CNV analysis on a cohort of 977 schizophrenia cases and 2,000 healthy adults of European ancestry who were genotyped with 1.7 million probes. Positive findings were evaluated in an independent cohort of 758 schizophrenia cases and 1,485 controls. The Gene Ontology synaptic transmission family of genes was notably enriched for CNVs in the cases (P = 1.5 × 10^{-7}). Among these, CACNA1B and DOC2A, both calcium-signaling genes responsible for neuronal excitation, were deleted in 16 cases and duplicated in 10 cases, respectively. In addition, RET and RIT2, both ras-related genes important for neural crest development, were significantly affected by CNVs. RET deletion was exclusive to seven cases, and RIT2 deletions were overrepresented common variants CNVs in the schizophrenia cases. Our results suggest that novel variations involving the processes of synaptic transmission contribute to the genetic susceptibility of schizophrenia.

S
chizophrenia is a devastating mental disorder characterized by reality distortion. Common features are positive symptoms of hallucinations, delusions, disorganized speech, and abnormal thought process; negative symptoms of social deficit, lack of motivation, anhedonia, and impaired emotion processing; and cognitive deficits with occupational dysfunction. Onset of symptoms typically occurs in late adolescence or early adulthood, with ∼1.5% of the population affected (1).

Previous studies have associated various copy number variations (CNVs) with schizophrenia including deletions of 22q11.2 (2), NRXN1 (3), APBA2 (3), and CNTNAP2 (4). These CNVs are rare, however, and they account for a relatively small proportion of the overall genetic risk in schizophrenia. Large, rare CNVs affecting many different genes enriched in neurodevelopmental pathways have been reported as well (5–7), with novel deletions and duplications of genes observed in 15% of cases versus 5% of controls in one study (P = 0.0008) (5). A study of CNVs in Chinese schizophrenia patients detected no significant difference in rare CNVs between cases and controls, however (8). Another study of 1,013 cases and 1,084 controls of European ancestry also failed to find more rare CNVs >100 kb in cases or enrichment for neurodevelopmental pathways (9). Specific loci exhibiting runs of homozygosity (ROHs) also have been associated with schizophrenia (10), and association of de novo CNVs (P = 7.8 × 10^{-4}) was recently reported in sporadic schizophrenia cases compared with controls (11). Comparison of genomic findings in schizophrenia and autism has suggested a diametric etiology (12). CNVs have been shown to contribute to the complex etiology underlying various psychiatric and neurodevelopmental disorders (13, 14). Whereas rare recurrent CNVs have been reported in patients with schizophrenia, these explain only a small fraction of the genetic risk of this common complex disease (15, 16). Accordingly, we have applied approaches with the objective of discovering variations and biological pathways contributing to the pathobiology of schizophrenia.

Our study cohort included 1,206 schizophrenia cases and 1,378 neurologically normal controls from the Genetic Association Information Network (GAIN), genotyped on the Affymetrix 6.0 array (17). We downloaded the data files from the database of Genotype and Phenotype (dbGaP) (ncbi.nlm.nih.gov/gap; study phs000021.v2.p1) and analyzed them for CNV associations. This GAIN project, also known as Molecular Genetics of Schizophrenia (MGS), previously reported linkage to 8p23.3-p21.2 and 11p13.1-q14.1 (18) and association of FGFR2 and ZNF804A in a genome-wide association study (GWAS) (19, 20), but failed to associate previously reported candidate genes (21) and found novel associations of common genotype variants on 6p22.1 (22). In addition, 351 schizophrenia cases and 2,107 control subjects from the University of Pennsylvania (UPenn) were included, along with 178 schizophrenia cases from Mount Sinai School of Medicine and Sheba Medical Center. Both cohorts were genotyped on the Affymetrix 6.0 array at The Children’s Hospital of Philadelphia (CHOP).

Control subjects from UPenn were originally recruited in relation with cross-sectional case-control studies on HDL cholesterol, coronary angiography, and heart transplantation outcomes at UPenn. The average age of the control cohort was 62 years, and no subjects had any major psychoses or other psychological symptoms. Samples from GAIN and UPenn were subsequently randomly divided into a discovery cohort of 977 schizophrenia cases and 2,000 controls and an independent replication cohort of 758 schizophrenia cases and 1,485 controls, including samples from Mount Sinai School of Medicine/Sheba Medical Center. Bias of contribution to specific loci among these sample sources was monitored.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission. J.L. is a guest editor invited by the Editorial Board.

1To whom correspondence should be addressed. E-mail: hakonarson@chop.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1000274107/-/DCSupplemental.
All subjects were diagnosed with schizophrenia based on criteria in the Diagnostic and Statistical Manual of Mental Disorders IV (DSM-IV) (23). Subjects had at least 6 months’ duration of the “A” criteria for schizophrenia, were at least 18 years old at the time of the interview, and were known by their informant for at least 2 years. Additional demographic data for the study subjects are presented in Tables S1 and S2.

Various array technologies, including aCGH, Affymetrix GeneChip and Illumina BeadChip, have been used to identify CNVs in healthy subjects. Previous studies have revealed significant common variations in the general healthy population (24). Various algorithms are also being used to call CNVs, most of which use the hidden Markov model (HMM) as implemented in PennCNV (25). Clustering of all Affymetrix data in one run with Affymetrix Power Tools (APT), which implements BirdSeed, is essential to minimize the stratification resulting from clustering bias. Indeed, the genotypes provided by dbGaP in matrix format had significant clustering bias among three apparent runs of APT on sample subsets based on Eigenstrat analysis. We ran APT with all Affymetrix 6.0 samples in a single run, yielding a typical result showing few African and Asian admixed samples without the three modes from clustering bias. Any possible batch effect from sample collection or processing was similarly excluded, and no overall processing variation was detected. Significant differential CNV detection bias is introduced by array batch effects; unusually given that large case and control sets were genotyped at the two sites (Broad Institute and CHOP).

### Results

We analyzed a total of 1,735 schizophrenia samples, genotyped on the Affymetrix 6.0 array, that met strictly established data quality thresholds for CNV, from which 977 cases were used in the discovery phase and 758 cases were used in the replication phase (Methods). An average of 45.4 CNV calls were made for each individual using the PennCNV software. We called four different copy number (CN) states, including 9,059 homozygous deletions (CN 0), 21,526 hemizygous deletions (CN 1), 9,750 duplications (CN 3), and 4,024 biallelic duplications/monoaallelic triplications (CN 4). The CNV calls ranged from 3 to 3,253 probes, with an average of 48 probes per CNV call, and their sizes ranged from 6 bp to 8.1 Mb, with an average size of 88.4 kb.

The CNV calls from the schizophrenia cases were compared with those from 3,485 healthy subjects. The control individuals had comparable CNV frequency to that of the cases. An average of 45.1 CNV calls were made for each control individual using the PennCNV software. Among these, we identified 29,257 homoygous deletions (CN 0), 70,052 hemizygous deletions (CN 1), 32,906 duplications (CN 3), and 14,217 biallelic duplications/monoaallelic triplications (CN 4). The CNV calls spanned from 3 to 9,258 probes, with an average of 48.6 probes per CNV call, and their sizes ranged from 4 bp to 12.7 Mb, with an average size of 87.9 kb.

In an attempt to replicate and better classify the reported abundance of rare CNVs in schizophrenia cases, we determined CNV case and control frequencies, applying different CNV association conditions: 100 kb CNV size, 100 kb CNV size, and not present in the Database of Genomic Variants (DGV); 10+ probe CNV size, 10+ probe CNV size, and not present in DGV; and samples with multiple novel genes impacted by CNVs. The 100-kb CNV size inclusion threshold excludes many CNVs that are informative and thereby affects many of the loci presented as novel to cases. For example, using the 100-kb threshold would have excluded 77% of the CNV calls in our discovery cohort. In general, CNVs called with 10 probes demonstrate a very low false-positive rate based on experimental validation and results in the exclusion of only 6% of our called CNVs. Under all conditions, we were unable to replicate the previously reported significant over-representation of rare CNVs affecting many genes in schizophrenia versus controls (Table S3) (5–7). But such broad conclusions might overlook bias introduced by certain samples with large numbers of CNVs called (typically as a result of sample quality), sample age, cell line artifacts, probe resolution, GC content, lack of genotyping information, and subphenotype characterization.

When using a threshold for CNV calls of 100 kb and larger, we replicated the 22q11.2 deletions robustly and detected CNV associations with GRID1, CNTNAP2, DISC1, and NRXN1, as reported previously (2–4). However, further review revealed multiple smaller CNVs (<100 kb) in these regions in both the cases and controls, suggesting that large CNVs in these regions and additional environmental and/or genetic factors might be required for expression of the schizophrenia phenotype. Deletions of 1q21.1 and 15q13.3 (6) were detected, but were not significantly associated with schizophrenia (Figs. S1 and S2).

We next carried out a genome-wide single SNP association analysis. We found no loci that were significant genome-wide; however, we detected nominally significant associations to several genes that are essential for brain development and function and of potential relevance to schizophrenia, including, but not limited to, ASTN2, CNTN5, and GRIK2 (\(p = 2.29 \times 10^{-5}, 6.63 \times 10^{-5}\), and \(2.53 \times 10^{-3}\), respectively; Table 1). As demonstrated in reports associating genotypes within the MHC locus with schizophrenia (22), such nominal significance might exist in an analysis of large cohorts and result in genome-wide significance in a meta-analysis.

To identify CNV loci potentially contributing to schizophrenia, we applied a segment-based scoring approach that scans the genome for consecutive probes with more frequent CN changes in cases compared with controls (Fig. S3). The genomic span for these consecutive probes forms common CNV regions, or CNVRs. In the discovery cohort of 977 schizophrenia cases and 2,000 healthy subjects, we identified CNVRs that were either unique to or had significantly higher frequency in cases versus controls and were detected in the independent replication cohort as well (Table 2). We also uncovered several rare CNV loci that were overrepresented in the discovery cohort but not observed in the replication cohort, likely because of their low frequencies (Table S4).

To assess the reliability of our CNV detection method, we experimentally validated all of the significant CNVRs using two additional methods, Illumina Human Hap550 Beadchip and quantitative PCR (qPCR), which is widely used by the research community for independent validation of CNVs (Table S5). We also examined the CNV frequency of 4,000 healthy control subjects genotyped on the Illumina 550 array recruited by the Center for Applied Genomics at CHOP and established CNV frequency in those samples comparable to that observed in controls typed on Affymetrix 6.0. We validated all significant schizophrenia-associated CNVs detected by the Illumina 550 chip with qPCR for two-tiered validation. Thus, we applied experimental validation on all of the CNVRs to ensure positive confirmation of all final results reported. Although the false-negative rate is unknown, based on conservative quality thresholds, it would not be expected to differ significantly between case and control cohorts.

To replicate the significant findings, we examined a replication cohort of 758 schizophrenia cases and 1,485 controls. Of the 25 significant loci in the discovery cohort, 8 were enriched in the replication cohort as well, reaching nominal significance (Table 2). Among these eight loci, five were very rare in controls (<0.25%), whereas the other three presented common CNVs that were significantly overrepresented in the cases. The resulting combined \(P\) values ranged from \(2.87 \times 10^{-6}\) to \(5.25 \times 10^{-5}\) for all CNVs in Table 2, four of which survived correction for 21 tests for deletion CNVRs and 5 tests for duplication CNVRs. Notably, two genes (CACNA1B and DOC2A) belong to the calcium-signaling family and the other two (RET and RIT2) belong to the Ras-signaling gene family, both of which are involved in neuronal development and signaling.

Although some genes did not replicate in our independent set of cases and controls of relatively modest size, these genes play
Table 1. GWA of SNP genotypes from 1,067 schizophrenia cases and 1,304 controls

<table>
<thead>
<tr>
<th>SNP</th>
<th>UNADJ P value</th>
<th>Chr</th>
<th>Position</th>
<th>Gene</th>
<th>A1</th>
<th>F_A</th>
<th>F_U</th>
<th>Distance</th>
<th>Count SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4697472</td>
<td>1.35 x 10^-6</td>
<td>4</td>
<td>24307401</td>
<td>SQD3</td>
<td>1</td>
<td>0.4675</td>
<td>0.3973</td>
<td>85,734</td>
<td>10</td>
</tr>
<tr>
<td>rs1587434</td>
<td>1.73 x 10^-6</td>
<td>6</td>
<td>66672076</td>
<td>EYS</td>
<td>2</td>
<td>0.0519</td>
<td>0.0257</td>
<td>191,994</td>
<td>4</td>
</tr>
<tr>
<td>rs11789407</td>
<td>2.29 x 10^-6</td>
<td>9</td>
<td>120399367</td>
<td>ASTN2</td>
<td>1</td>
<td>0.5203</td>
<td>0.4512</td>
<td>595,986</td>
<td>8</td>
</tr>
<tr>
<td>rs1555543</td>
<td>4.46 x 10^-6</td>
<td>1</td>
<td>96717385</td>
<td>PTBLP,PTBP2</td>
<td>1</td>
<td>0.4545</td>
<td>0.3884</td>
<td>296,084</td>
<td>4</td>
</tr>
<tr>
<td>rs35648</td>
<td>5.95 x 10^-6</td>
<td>10</td>
<td>80171865</td>
<td>AF086162</td>
<td>1</td>
<td>0.1242</td>
<td>0.1714</td>
<td>61,304</td>
<td>6</td>
</tr>
<tr>
<td>rs2155907</td>
<td>6.63 x 10^-6</td>
<td>11</td>
<td>97599883</td>
<td>CNUTS</td>
<td>2</td>
<td>0.4035</td>
<td>0.3393</td>
<td>778,692</td>
<td>5</td>
</tr>
<tr>
<td>rs2271293</td>
<td>9.96 x 10^-6</td>
<td>16</td>
<td>66595971</td>
<td>NUTF2</td>
<td>1</td>
<td>0.1425</td>
<td>0.1006</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>rs4981929</td>
<td>9.96 x 10^-6</td>
<td>14</td>
<td>31442403</td>
<td>NUBPL</td>
<td>1</td>
<td>0.527</td>
<td>0.462</td>
<td>2358</td>
<td>11</td>
</tr>
<tr>
<td>rs11713590</td>
<td>1.12 x 10^-5</td>
<td>3</td>
<td>5706142</td>
<td>EDEM</td>
<td>1</td>
<td>0.425</td>
<td>0.486</td>
<td>459,455</td>
<td>10</td>
</tr>
<tr>
<td>rs12140791</td>
<td>1.85 x 10^-5</td>
<td>1</td>
<td>16037908</td>
<td>ROSTAP</td>
<td>2</td>
<td>0.06232</td>
<td>0.03569</td>
<td>503</td>
<td>0</td>
</tr>
<tr>
<td>rs12538910</td>
<td>1.92 x 10^-5</td>
<td>7</td>
<td>57318107</td>
<td>DO579B2</td>
<td>1</td>
<td>0.423</td>
<td>0.3633</td>
<td>50,874</td>
<td>5</td>
</tr>
<tr>
<td>rs10499040</td>
<td>2.53 x 10^-5</td>
<td>6</td>
<td>104889038</td>
<td>GRK2</td>
<td>2</td>
<td>0.1345</td>
<td>0.09555</td>
<td>2,264,387</td>
<td>3</td>
</tr>
<tr>
<td>rs1357338</td>
<td>1.19 x 10^-4</td>
<td>1</td>
<td>174197509</td>
<td>RFWD2</td>
<td>1</td>
<td>0.0188</td>
<td>0.00652</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>rs4509495</td>
<td>1.33 x 10^-4</td>
<td>4</td>
<td>42018121</td>
<td>CASK</td>
<td>2</td>
<td>0.1495</td>
<td>0.2015</td>
<td>196,216</td>
<td>4</td>
</tr>
<tr>
<td>rs4813376</td>
<td>1.92 x 10^-4</td>
<td>20</td>
<td>19799455</td>
<td>RIN2</td>
<td>2</td>
<td>0.1856</td>
<td>0.1453</td>
<td>18,744</td>
<td>2</td>
</tr>
<tr>
<td>rs6560936</td>
<td>4.43 x 10^-4</td>
<td>13</td>
<td>113964074</td>
<td>RASA2</td>
<td>2</td>
<td>0.5009</td>
<td>0.4497</td>
<td>40,707</td>
<td>4</td>
</tr>
</tbody>
</table>

The most significant SNP is reported with neighboring SNPs within 10 kb and significance ranging within a power of 10 noted in the "Count SNPs" column. F_A indicates allele frequency affected; F_U, allele frequency unaffected. Genes associated with brain development and function are in bold type.

Supporting functional roles in schizophrenia and might replicate with further studies of larger sample sizes. Additional associated Ras-related cell cycle regulation family genes include PTPRRG, RAB23, TM2D3, SCH2, and RAPGEF2. In addition, PTLBP, RIN2, and RASA4 are Ras genes supported by our genotype GWA presented in Table 1. Additional associated calcium-signaling family genes include CAMK2D and KCNB4. Also associated are PARK2, RFW2, and PTPRB, which we previously associated with autism (13), the latter interacting with the contactin gene family. Because these nominally significant loci might seem unimpressive individually based on their respective P values, we sought to identify the perturbed gene pathways based on CNVs of different loci and to determine whether genes in any particular signaling or biological pathway were enriched for CNVs. Using this approach, we nominally associated an additional five Ras neural crest development genes and two calcium regulatory signaling genes, for a total of seven Ras genes and four Ras-linked calcium-dependent signaling genes impacted with CNVs associated with schizophrenia. Taken together, the Gene Ontology (GO) class synaptic transmission genes (CACAIAIC, PARK2, KCNB4, GID2, DOC2A, COMT, RIT2, and ATXN1) was significantly enriched in the cases (P = 1.5 x 10^-7).

Discussion

The genes impacted by or proximal to significant CNVs encode proteins with intriguing function. PDPR (pyruvate dehydrogenase phosphatase regulatory gene) is involved in glycine catabolism, and the International Schizophrenia Consortium (ISC) data show six novel deletions in cases and one such deletion in controls. As shown in Fig. 1, this locus replicated in 30 independent cases and directly affected PDPR, using the UCSC Genome Browser (26) with Build 36. Pyruvate dehydrogenase is of fundamental importance to glycolysis, and the brain requires half of the total body glucose utilization (27). DOC2A is expressed mainly in brain and is involved in Ca(2+)-dependent neurotransmitter release. This large recurrent duplication (and deletion) of 16p11.2 was observed in autism patients as well (13, 28, 29). Within 22q11.21, COMT catalyzes the transfer of a methyl group from S-adenosylmethionine to catecholamines, including the neurotransmitters dopamine, epinephrine, and norepinephrine. The 22q11.21 deletion locus was previously reported in 11 cases and no controls by the ISC (6), an association to schizophrenia that has been previously reported and well supported (2). CACAIAIC is a N-type calcium channel that controls neurotransmitter release from neurons and was found to be associated with bipolar disorder and schizophrenia via GWAS by the Wellcome Trust Case Control Consortium (30). A gene with sequence homology, CACNAIC, has been robustly associated with bipolar disorder based on genotypes of 4,387 cases and 6,209 controls (31). One deletion and four duplications were found in cases, whereas one control duplication was found over the span of CACNAIC. RET is a receptor tyrosine kinase, a cell surface molecule that transduces signals for cell growth and differentiation, which plays a crucial role in neural crest development (32). RET loss of function is associated with Hirschsprung disease, whereas gain of function is associated with cancer development. WDR1 is involved in actin formation and sensory perception of sound. Studies using shotgun mass spectrometry found that WDR1 was differentially expressed in the dorsolateral prefrontal cortex of schizophrenia patients (33). The linkage of bipolar disorder to chromosome 4p and haplotype analysis have implicated the WDR1 locus (34). RIT2 is a Ras-like protein expressed in neurons. PIK3C3 has been shown to harbor a promoter mutation that increases the risk of schizophrenia and bipolar disorder (35). We previously reported SUMF1 (UNQ3037) deletion in 11 unrelated cases in association with autism (13). SUMF1 catalyzes the hydrolysis of sulfate esters, such as glycosaminoglycans, sulfolipids, and steroid sulfates.

Although Ras has been the focus of numerous cancer studies as a pivotal tumor suppressor, less emphasis has been given to its native biological roles in neuronal survival, differentiation, and plasticity. Calcium-mediated pathways of Ras activation might be a critical mechanism to couple rapid and transient neuronal electrical activity with long-term changes in nervous system development and function (36–40). The small GTPase protein Ras couples calcium influx to many forms of synaptic plasticity, including rapid synaptic potentiation and new synapse formation. Consistent with many essential roles of Ras signaling in neuronal plasticity, mutations in the Ras-signaling pathway are associated with other diseases causing cognitive impairments and learning deficits, such as autism, X-linked mental retardation, and neurofibromatosis 1 (41–43). Indeed, we have identified rare, highly penetrant CNVs in ubiquitin genes and common CNVs that were overrepresented in neuronal development in autism (13). Furthermore, based on genotype association, a common variant on 5p14.1 between CDH10 and CDH9 encoding neuronal cell-adhesion molecules also has been associated with autism (44).

We also observed enrichment of several mitochondria-linked genes affected by CNVs in the schizophrenia cases. Parkin is es-
Table 2. CNVRs statistically overrepresented in schizophrenia cases and replicated in an independent case-control cohort

<table>
<thead>
<tr>
<th>CNVR Probes</th>
<th>Cases discovery</th>
<th>Controls discovery</th>
<th>Discovery P value</th>
<th>Cases replication</th>
<th>Controls replication</th>
<th>Combined P value</th>
<th>OR</th>
<th>Gene</th>
<th>Distance from gene</th>
<th>Type</th>
<th>Replication ISC</th>
<th>Canary</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr1:68743639–68770545</td>
<td>9</td>
<td>19</td>
<td>10</td>
<td>1.17 × 10⁻³</td>
<td>16</td>
<td>9</td>
<td>2.87 × 10⁻⁶</td>
<td>3.756</td>
<td>PDR</td>
<td>Del</td>
<td>6:1 ISC; P = 0.13</td>
<td>N</td>
</tr>
<tr>
<td>chr16:29425212–30134444</td>
<td>217</td>
<td>5</td>
<td>0</td>
<td>3.78 × 10⁻³</td>
<td>8</td>
<td>1</td>
<td>5.69 × 10⁻⁶</td>
<td>26.30</td>
<td>52 genes; QPRT, DOCA, TBX6</td>
<td>0</td>
<td>6:3 ISC; P = 0.51</td>
<td>Y</td>
</tr>
<tr>
<td>chr22:17404806–19941349</td>
<td>1,529</td>
<td>8</td>
<td>0</td>
<td>1.32 × 10⁻⁴</td>
<td>2</td>
<td>0</td>
<td>1.62 × 10⁻³</td>
<td>NA</td>
<td>75 genes; COMT</td>
<td>0</td>
<td>11:0 ISC; P = 0.001</td>
<td>Y</td>
</tr>
<tr>
<td>chr9:140145139–140152969</td>
<td>7</td>
<td>12</td>
<td>4</td>
<td>6.54 × 10⁻⁴</td>
<td>4</td>
<td>4</td>
<td>8.68 × 10⁻⁴</td>
<td>4.045</td>
<td>CACNA1B</td>
<td>8.69 kb</td>
<td>Del</td>
<td>—</td>
</tr>
<tr>
<td>chr10:42932615–42934254</td>
<td>17</td>
<td>5</td>
<td>0</td>
<td>3.78 × 10⁻³</td>
<td>2</td>
<td>2</td>
<td>8.09 × 10⁻³</td>
<td>7.055</td>
<td>RET</td>
<td>0</td>
<td>Del</td>
<td>—</td>
</tr>
<tr>
<td>chr4:3881886–9884092</td>
<td>11</td>
<td>6</td>
<td>2</td>
<td>1.80 × 10⁻²</td>
<td>5</td>
<td>6</td>
<td>2.81 × 10⁻²</td>
<td>2.773</td>
<td>WDR1</td>
<td>154 kb</td>
<td>Del</td>
<td>—</td>
</tr>
<tr>
<td>chr18:38310567–38311765</td>
<td>25</td>
<td>115</td>
<td>163</td>
<td>1.98 × 10⁻³</td>
<td>67</td>
<td>137</td>
<td>2.91 × 10⁻²</td>
<td>1.244</td>
<td>RIT2, PIK3C3</td>
<td>265 kb, 395 kb</td>
<td>Del</td>
<td>—</td>
</tr>
<tr>
<td>chr3:4063809–4074877</td>
<td>30</td>
<td>14</td>
<td>13</td>
<td>4.01 × 10⁻²</td>
<td>7</td>
<td>10</td>
<td>5.25 × 10⁻²</td>
<td>1.844</td>
<td>SUMF1</td>
<td>0</td>
<td>Del</td>
<td>—</td>
</tr>
</tbody>
</table>

Significant CNVRs based on a combined discovery and replication cohort of 1,735 schizophrenia cases and 3,485 healthy controls of European ancestry. Replication ISC: Samples from different sample sources must have reasonable contributing frequency. Canary: A CNV calling algorithm run on the sample set in addition to PennCNV-Affy to establish independent calling positive replication (Y) or lack of replication (N). † indicates more samples with Canary calls. Del, deletion. Dup, duplication. CNVRs that survive multiple testing with Bonferroni adjustment in the discovery phase (P < 0.05 after correction for 20 tests in case of deletion and 5 tests in case of duplications), survived replication, and experimental validation are listed in bold. The CNVR is the CNV region shared between cases. Probes give the number of SNP and CN probes present on the Affymetrix 6.0 array in the respective CNVR from which signal was indicative of a CNV. The discovery and combined P values are based on a Fisher’s exact test of the respective sample cohorts. The count of samples in each subgroup of cases and controls in discovery and replication is provided. The nearest gene and proximal distance are provided for potential functional impact and a means to compare other sample sets that might find CNVs in the region. The “Replication ISC” column shows the frequency of cases and controls in the International Schizophrenia Consortium CNV calls of 3,391 cases and 3,181 controls. The “Canary” column shows if the analysis of the log2 ratio of intensity through the Canary CNV calling algorithm replicates the CNV call from PennCNV-Affy. Key functional genes are provided for brevity. The gene count for the two largest CNVs includes hypothetical genes.
eleventy mitochondrial inner membrane.

In other genes (53, 54), schizophrenia cases and controls. Two of the affected genes (55) enriched for CNVs in schizophreniacases compared with controls, their CNV call boundaries shown by red lines. The schizophrenia cases of our control cohort of 3,485 showing CNV overrepresentation in cases. ISC samples in which the GC wave factor of LRR was between 0.35 and replicated many of these using an independent data set of their CNV call boundaries shown by red lines. The schizophrenia cases of our control cohort of 3,485 showing CNV overrepresentation in cases. ISC 1,735 cases population run on Affymetrix 6.0 are shown in comparison with 1,735 cases population run on Affymetrix 6.0 are shown in comparison with enrichment of CNVs in schizophreniacases compared with controls, and all other samples were excluded. Wave artifacts roughly correlating with GC content result from high from hybridization bias of low full-length DNA quantity are known to interfere with accurate inference of CNVs (55); thus, only samples in which the GC wave factor of LRR was between 0.02 and 0.02 were accepted. If the count of CNV calls made by PennCNV exceeds 80, then the DNA quality is usually poor; thus, only samples with a CNV call count <80 were included. Any duplicate samples (such as monozygotic twins) had one sample excluded.

In conclusion, using a genome-wide approach for high-resolution CNV detection, we have identified candidate genomic loci with enrichment of CNVs in schizophrenia cases compared with controls, and replicated many of these using an independent data set of schizophrenia cases and controls. Two of the affected genes (CA2N41B and DOC2A) encode calcium-signaling molecules, and two other genes (RET and RIT2) belong to the Ras-signaling gene family, both of which are involved in neuronal development and signaling. Together, these genes show significant enrichment in the gene family of synaptic transmission molecules based on GO (P = 1.5 × 10⁻⁷).

The enrichment of genes within this molecular system suggests susceptibility mechanisms for schizophrenia and should spur identification of additional variations, including structural variations and single-base changes, in candidates within these gene networks. In addition, our results call for functional expression assays to assess the biological effects of CNVs in these candidate genes in brain tissue.

Methods

**Affymetrix 6.0 Assay for CNV Discovery.** High-throughput, genome-wide SNV and CN genotyping was performed, using the Affymetrix 6.0 technology, at the Center for Applied Genomics at CHOP. dbGaP samples were genotyped on the same platform at the Broad Institute. The genotype data content, together with the intensity data provided by the SNP probes on the genotyping array, provide high confidence for CNV calls. Importantly, the simultaneous analysis of intensity data and genotype data in the same experimental setting establishes a highly accurate definition for normal diploid states and any deviation thereof. To call CNVs, we used the PennCNV-Affy algorithm, which combines multiple sources of information, including log R range (LRR) and B allele frequency (BAF) at each SNP marker, an HMM specifically trained on Affymetrix 6.0 data, and SNP spacing and population frequency of the B allele to generate CNV calls.

**CNV Quality Control.** We calculated quality control measures on our Affymetrix 6.0 and HumanHap550 GWAS data based on statistical distributions to exclude poor-quality DNA samples and false-positive CNVs. The first threshold is the percentage of attempted SNPs that were successfully genotyped. Only samples with a call rate >96% were included. The genome-wide intensity signal must have as little noise as possible. Only samples with an SD of normalized intensity (LRR) <0.35 were included. All samples must have Caucasian ethnicity based on hierarchical clustering of AIMs genotypes and all other samples were excluded. Wave artifacts roughly correlating with GC content result from high from hybridization bias of low full-length DNA quantity are known to interfere with accurate inference of CNVs (55); thus, only samples in which the GC wave factor of LRR was between 0.02 and 0.02 were accepted. If the count of CNV calls made by PennCNV exceeds 80, then the DNA quality is usually poor; thus, only samples with a CNV call count <80 were included. Any duplicate samples (such as monozygotic twins) had one sample excluded.

**Statistical Analysis of CNVs.** We evaluated CNV frequency between cases and controls at each SNP using Fisher’s exact test. We only considered loci that were significant between cases and controls (P < 0.05), where cases in the MGS/Gur discovery cohort had the same variation replicated in MGS/Gur or were not observed in any of the control subjects, and were validated with an independent method. We report statistical local minimums to narrow the association in reference to a region of nominal significance including SNPs residing within 1 Mb of one another. Resulting significant CNVRs were excluded if they met any of the following criteria: (i) residing on telomere or centromere proximal cytobands, (ii) arising in a “peninsula” of common CNV arising from variation in boundary truncation of CNV calling, (iii) genomic regions with extremes in GC content that produce hybridization bias, or (iv) samples contributing to multiple CNVRs. We used DAVID (56) to assess the significance of functional annotation clusters of CNVRs and CNVR results into functional categories. To adjust for the number of tests performed, we made correction of 21 deletion and 5 duplication CNVRs, based on significance in the discovery cohort.

**PennCNV-Affy.** The Affymetrix 6.0 provided 848,415 SNP markers and 888,023 CN markers, which were analyzed to construct canonical clustering positions using the PennCNV-Affy workflow, which normalizes the Cartesian coordinates provided by Affymetrix. PennCNV-Affy uses called genotypes and normalizes intensity from APT to create reference cluster positions in polar coordinates to compute relative differences in the signal from each sample in the form of BAF and LRR. BAF, LRR, population BAF, interprobe distance, and HMM model files were then analyzed by PennCNV to make CNV calls for each sample. The HMM was trained on Affymetrix 6.0 data. We used PennCNV-Adapt because they use the LRR rather than the log 2 ratio. We reviewed the log 2 ratio values in the visualization tools Affymetrix Genotyping Console Heat Map and Browser (Fig. S4). The log 2 ratio is based on quantile normalization, the sum of signal intensity for the A allele and B allele for each sample and the median across all samples; for a given sample, A + B allele intensity is divided by the median value and the logarithm base 2 is taken. In contrast, the LRR is based on defined signal intensity clusters of AA, AB, and BB genotypes across a large group of samples. Given this expected intensity value, the observed A + B signal intensity is divided by this expected value, and the logarithm is taken. The number of CNVs called per individual by PennCNV-Affy was lower than that called by Birdseye.

**ACKNOWLEDGMENTS.** We gratefully thank all of the schizophrenia patients who participated in this study and all of the control subjects who donated blood samples to CHOP and the UPenn for genetic research purposes. We thank the technical staff at CHOP’s Center for Applied Genomics for generating the...
17. Manolio TA, et al., GAIN Collaborative Research Group, Collaborative Association Study contributions. We thank the MGS consortium investigators for making the genome data available. Support for MGS case-control was provided by funding from the National Institute of Mental Health and the National Alliance for Research on Schizophrenia and Depression. Genotyping of part of the sample was supported by the Genetic Association Information Network, and the Paul Michael Donovan Charitable Foundation. The genotyping of samples was provided through the National Center for Research Resources and performed by the Center for Genotyping and Analysis at the Broad Institute of Harvard and MIT. The dataset used for the analyses described in this manuscript was obtained for research purposes by the DNA Sequencing Center, which was supported by the National Institutes of Health through the Human Genome Mapping Program. The authors wish to thank the investigators who contribute to the NIMH-funded U10 Research Centers in Genomic Medicine Collaborative Study and the DCRI, and thank T.P.C. and K.B.M.)]. The collection and genotyping of samples from Mount Sinai and Sheba Hospital was partially supported by the National Institute of Mental Health, Conte Center for Neuroscience of Mental Disorders (Grant MH066392 to J.D.B.).  