

# Pathogenic rare copy number variants in community-based schizophrenia suggest a potential role for clinical microarrays

Gregory Costain<sup>1,†</sup>, Anath C. Lionel<sup>2,3,†</sup>, Daniele Merico<sup>2</sup>, Pamela Forsythe<sup>4</sup>, Kathryn Russell<sup>1</sup>, Chelsea Lowther<sup>1</sup>, Tracy Yuen<sup>1</sup>, Janice Husted<sup>5</sup>, Dimitri J. Stavropoulos<sup>6,7</sup>, Marsha Speevak<sup>8</sup>, Eva W.C. Chow<sup>1,9</sup>, Christian R. Marshall<sup>2,3</sup>, Stephen W. Scherer<sup>2,3</sup> and Anne S. Bassett<sup>1,9,\*</sup>

<sup>1</sup>Clinical Genetics Research Program, Centre for Addiction and Mental Health, Toronto, ON, Canada M5S 2S1, <sup>2</sup>The Centre for Applied Genomics and Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada M5G 1L7, <sup>3</sup>Department of Molecular Genetics and McLaughlin Centre, University of Toronto, Toronto, ON, Canada M5G 1L7, <sup>4</sup>Horizon Health and Dalhousie University, Saint John, NB, Canada E2L 4L4, <sup>5</sup>Health Studies, University of Waterloo, Waterloo, ON, Canada N2L 3G1, <sup>6</sup>Cytogenetics Laboratory, Department of Pediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada M5G 1X8, <sup>7</sup>Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada M5S 1A1, <sup>8</sup>Department of Genetics, Credit Valley Hospital, Mississauga, ON, Canada L5M 2N1 and <sup>9</sup>Department of Psychiatry, University of Toronto, Toronto, ON, Canada M5T 1R8

Received March 7, 2013; Revised May 17, 2013; Accepted June 20, 2013

**Individually rare, large copy number variants (CNVs) contribute to genetic vulnerability for schizophrenia. Unresolved questions remain, however, regarding the anticipated yield of clinical microarray testing in schizophrenia. Using high-resolution genome-wide microarrays and rigorous methods, we investigated rare CNVs in a prospectively recruited community-based cohort of 459 unrelated adults with schizophrenia and estimated the minimum prevalence of clinically significant CNVs that would be detectable on a clinical microarray. A blinded review by two independent clinical cytogenetic laboratory directors of all large (>500 kb) rare CNVs in cases and well-matched controls showed that those deemed to be clinically significant were highly enriched in schizophrenia (16.4-fold increase,  $P < 0.0001$ ). In a single community catchment area, the prevalence of individuals with these CNVs was 8.1%. Rare 1.7 Mb CNVs at 2q13 were found to be significantly associated with schizophrenia for the first time, compared with the prevalence in 23 838 population-based controls (42.9-fold increase,  $P = 0.0002$ ). Additional novel findings that will facilitate the future clinical interpretation of smaller CNVs in schizophrenia include: (i) a greater proportion of individuals with two or more rare exonic CNVs >10 kb in size (1.5-fold increase,  $P = 0.0109$ ) in schizophrenia; (ii) the systematic discovery of new candidate genes for schizophrenia; and, (iii) functional gene enrichment mapping highlighting a differential impact in schizophrenia of rare exonic deletions involving diverse functions, including neurodevelopmental and synaptic processes (4.7-fold increase,  $P = 0.0060$ ). These findings suggest consideration of a potential role for clinical microarray testing in schizophrenia, as is now the suggested standard of care for related developmental disorders like autism.**

## INTRODUCTION

Schizophrenia is a complex neuropsychiatric disease that affects up to 1% of the general population and shows evidence for

neurodevelopmental origins and genetic heterogeneity (1). There is compelling evidence that individually rare, large copy number variants (CNVs), especially 22q11.2 microdeletions, play a role in the genetic causation of schizophrenia (1,2). As

\*To whom correspondence should be addressed at: Anne S. Bassett, 33 Russell Street, Room 1100, Toronto, ON, Canada M5S 2S1. Tel: +1 4165358501; x2734; Fax: +1 4165357199; Email: anne.bassett@utoronto.ca

<sup>†</sup>The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

for other major neurodevelopmental disorders, such as autism and intellectual disability (3–6), findings from recent genome-wide case–control studies of CNVs (1,3,7–11) support a multiple rare variant model of causation in schizophrenia (7,12). In general, rare and large (e.g., >500 kb) CNVs, often overlapping multiple genes, are more likely to have a clinically relevant phenotype (3). In schizophrenia, unresolved questions remain regarding the anticipated yield today on clinical microarray testing of large rare CNVs of potential clinical relevance (4,7). Most case–control sampling schemes to date have not been epidemiologic in nature, and therefore cannot provide reliable estimates (7). This uncertainty about yield may contribute to the low rate of uptake of clinical genetic testing in schizophrenia: only 5 of 38 779 samples processed at Signature Genomics were from patients with this common and serious mental illness (13).

Smaller rare CNVs may implicate individual risk genes for schizophrenia (14,15), but the clinical interpretation of these variants is more challenging. A practical consideration for an adult-onset condition like schizophrenia is the typical absence of parental samples to determine *de novo*/inheritance status. High-resolution genome-wide data and systematic approaches (both quantitative and descriptive/qualitative) to identify putative candidate genes are needed to inform a more sophisticated approach to the clinical assessment of CNVs in schizophrenia. The existence of overarching genetic networks, whose disruption may contribute to expression of schizophrenia, is a particularly tantalizing possibility (8,16).

We prospectively ascertained and systematically assessed a community-based sample of Canadian patients with schizophrenia to address the above issues. All subjects underwent direct clinical screening assessments for potential syndromic features using a standardized protocol that included a review of available lifetime medical records and assessment of physical features (2,17,18). Cases were compared with an unrelated epidemiologic control sample of comparable (European) ancestry (Ontario Population Genomics Platform; OPGP) (18). We used a high-resolution genome-wide microarray and proven analytic methods for CNV detection and evaluation (5,6,16,18–20). An independent, blinded review process identified large rare CNVs deemed to be ‘Pathogenic’ or of ‘Uncertain clinical significance; likely pathogenic’ in cases and controls, based on the well-established American College of Medical Genetics (ACMG) guidelines for CNV interpretation (21). A representative epidemiologic subsample from a single catchment area (2) allowed us to estimate for the first time the minimum collective prevalence of individuals with large, clinically significant CNVs in schizophrenia (1,7). To shed light on the contribution of smaller CNVs to the genetic architecture of schizophrenia, we investigated the total burden of such CNVs and the potential support for a ‘multiple hit’ model of causation. We also developed a systematic approach to the discovery of new candidate genes for schizophrenia, which included functional gene enrichment mapping to identify relevant genetic networks.

## RESULTS

### Clinically significant and novel large rare structural variants in schizophrenia

Two clinical cytogenetic laboratory directors conducted independent blinded examinations of all rare (present in <0.1% of 2357

population controls) CNVs 500 kb–6.5 Mb in size in schizophrenia cases of European ancestry and OPGP controls. This revealed a significant enrichment in schizophrenia of individuals with ‘Pathogenic’ (21) or ‘Uncertain clinical significance; likely pathogenic’ (21) (collectively termed ‘clinically significant’) large rare autosomal CNVs [16 of 420 versus 1 of 416;  $P < 0.0001$ , odds ratios (OR) 16.44 (95% CI 2.17–124.51)]. In the schizophrenia case sample, there were 16 CNVs considered clinically significant at eight loci (Table 1; Fig. 1, Supplementary Material, Fig. S2) (1,4,7,8,10,22–25). All were very rare (i.e., present in none of 2357 population controls used to adjudicate both the schizophrenia case and OPGP control samples), including six CNVs at four loci novel to schizophrenia: 2q13, 3q13.31, 5p15.33-p15.32 and 10q11.22-q11.23 (Fig. 1) (4,22–25). None of the corresponding individuals with schizophrenia had a pediatric diagnosis of moderate or severe intellectual disability, multiple major congenital anomalies, epilepsy or autism (2). Notably, five typical 2.6 Mb 22q11.2 deletions in cases (none in controls) were a priori excluded (see Materials and Methods) to demonstrate that these findings are not being driven by those established variants. The single large rare CNV in OPGP control individuals deemed to be clinically significant was a typical 2.6 Mb 22q11.2 duplication (4). There were no significant differences in demographics, years of education, age at onset, family history of schizophrenia, self-reported history of special education or syndromic designation (see Materials and Methods) in these cases with large, rare, clinically significant CNVs when compared with subjects with no large rare CNVs 500 kb–6.5 Mb in size. There were also eight very large (>6.5 Mb) anomalies in schizophrenia cases (none in OPGP controls) that were considered clinically significant (Table 1), including a 10 Mb loss (deletion) overlapping *FOXP2* (Supplementary Material, Fig. S3) (26).

A novel finding was the detection in three of 420 schizophrenia cases of European ancestry of very rare, recurrent 1.7 Mb CNVs at 2q13 (4,22) (Table 1; Fig. 1), a significant enrichment compared with the prevalence ( $n = 4$ ) in 23 838 population-based controls (Supplementary Material, Table S3) [ $P = 0.0002$ ; OR 42.87 (95% CI 9.56–192.14)]. To our knowledge, this CNV has not been previously reported in other genotyped schizophrenia cohorts [e.g., the International Schizophrenia Consortium dataset (9)]. None of the three individuals in our case sample with these 2q13 CNVs had any of the pediatric diagnoses [developmental delay, multiple congenital anomalies and/or autism spectrum disorder (ASD)] previously reported for samples assessed at clinical laboratories (4,22). Two of the three individuals reported a history of learning difficulties in school, but all had IQs within the average range. For the one schizophrenia proband with a positive family history, we were able to confirm co-segregation of the 2q13 duplication with schizophrenia. Of the 10 genes overlapped by these 2q13 CNVs, three are promising candidates for schizophrenia: the neurodevelopmental facilitator *ANAPC1* (27), the neuronal apoptosis regulator *BCL2L11* (28), and the TAM receptor component and multiple sclerosis risk gene *MERTK* (29) (Table 1; Fig. 1).

### Estimated prevalence of all clinically significant CNVs in a community sample of schizophrenia

We used our schizophrenia catchment sample (2) to estimate the minimum prevalence of large (500 kb–6.5 Mb) rare clinically

**Table 1.** Very rare, clinically significant<sup>a</sup> confirmed CNVs discovered in unrelated probands with schizophrenia

Subject Case #	Catchment area	CNV characteristics Cytoband	Start	Size (bp)	CN	Flanking segmental duplications	# of genes	Candidate gene(s)
Large, very rare CNVs <6.5 Mb in size, excluding 22q11.2 deletions <sup>b</sup> ( <i>n</i> = 16 subjects)								
2	•	1q21.1	144 472 163	1 839 252	Gain <sup>c</sup>	•	16	<i>BCL9, GJA5<sup>d</sup>, GJA8<sup>d</sup>, PDZK1<sup>e,f</sup>, PRKAB2</i>
3	•	1q21.1	144 643 825	1 653 983	Gain <sup>c</sup>	•	14	
8	•	2q13	111 105 101	1 727 363	Loss	•	10	<i>ANAPC1, BCL2L11, MERTK</i>
7 <sup>g</sup>	•	2q13	111 105 101	1 727 363	Gain	•	10	
9	•	2q13	111 105 101	1 727 363	Gain	•	10	
13 <sup>g</sup>	•	3q13.31	115 308 450	2 062 410	Loss		7	<i>DRD3, GAP43<sup>d</sup>, LSAMP, ZBTB20</i>
17	•	5p15.33-p15.32	1 864 574	3 687 431	Loss		9	IrxA cluster ( <i>IRX1, IRX2, IRX4, NDUFS6</i> )
40 <sup>g</sup>	•	10q11.22-q11.23	45 905 767	5 423 684	Gain	•	52	<i>CHAT, ERCC6, GDF2, GPRIN2<sup>d</sup>, MAPK8, SLC18A3</i>
48	•	15q11-q13	20 224 763	6 498 447	Gain <sup>h</sup>	•	115	GABA receptor gene cluster <sup>f</sup> ( <i>GABRB3, GABRA5, GABRG3, MAGEL2, NDN, UBE3A</i> )
49	•	15q11-q13	21 192 955	5 680 224	Gain <sup>h</sup>	•	106	
50	•	15q11-q13	21 192 955	5 015 049	Gain <sup>h</sup>	•	101	
52	•	15q13.2-q13.3	28 608 929	1 690 584	Loss	•	10	<i>CHRNA7<sup>f</sup>, TRPM1<sup>f</sup></i>
55 <sup>g</sup>	•	16p11.2	29 474 810	624 599	Gain	•	28	<i>DOC2A, MAPK3, PRRT2<sup>d</sup>, QPRT, SEZ6L2, TBX6</i>
57 <sup>g</sup>	•	16p11.2	29 474 810	624 599	Gain	•	28	
56	•	16p11.2	29 474 810	659 635	Gain	•	38	
58	•	16p11.2	29 474 810	659 635	Gain	•	38	
Very rare CNVs > 6.5 Mb in size ( <i>n</i> = 8 subjects; one subject also appears above)								
247 <sup>g</sup>		6p25.3-p25.1	94 661	6 836 704	Loss <sup>c</sup>		41	<i>FOXC1, GMDS, NRN1, TUBB2B</i>
206	•	7q22.2-q31.1	105 304 955	10 037 597	Loss <sup>i</sup>	•	37	<i>COG5, DOCK4, FOXP2, GPR85, IMMP2L, LAMB1, NRCAM, PIK3CG<sup>d</sup>, PNPLA8</i>
115 <sup>g</sup>	•	8p23.3-p23.1	148 062	6 828 275	Loss <sup>c</sup>	•	25	<i>ANGPT2<sup>d</sup>, CLN8<sup>d</sup>, CSMD1<sup>f</sup>, DLGAP2, MCPHI<sup>d</sup></i>
277 <sup>g</sup>	•	19p13.3-p13.2	2 705 548	9 595 341	Gain <sup>c</sup>		280	Various, including <i>DNMT1, DOCK6<sup>d</sup></i>
173	•	X chr (47, XXX)	–	–	Gain	–	975	Various, including <i>ILIRAPLI<sup>f</sup>, SYN1</i>
278	•	X chr (47, XXX)	–	–	Gain	–	975	
57 <sup>g</sup>	•	X chr (47, XXY)	–	–	Gain	–	975	

Continued

Table 1. Continued

Subject Case #	Catchment area	CNV characteristics		Size (bp)	CN	Flanking segmental duplications	# of genes	Candidate gene(s)
		Cytoband	Start					
279 <sup>g</sup>	•	Y chr (47, XYY)	–	–	Gain <sup>c</sup>	–	34	<i>IL9R, SPRY3, SRY, TSPY, UTY, VAMP7</i>

Case #, case number for subjects ( $n = 454$ ) with schizophrenia ( $n = 2$  were not of European ancestry: cases 278, 279); Catchment area, subjects originating from the only community mental health clinic in a specific catchment area of ~150 000 people (•), see the text for details; Cytoband, cytogenetic location of CNV; CNV start, hg18 (NCBI Build 36.1, March 2006); CNV size, in base pairs; CN, type of copy number aberration; Flanking segmental duplications, known flanking segmental duplications (from the UCSC Genome Browser hg18 version) that cover at least 20% of the CNV length (•); # of genes, number of known genes overlapped by CNV as annotated in the Database of Genomic Variants (<http://projects.tcag.ca/variation/>; September 2011); Candidate gene(s), selected based on reported neuropsychiatric/neurodevelopmental phenotype identified from systematic searches of human (e.g. Online Mendelian Inheritance in Man; <http://www.omim.org/>) and model organism (e.g. Mouse Genome Informatics; <http://www.informatics.jax.org/>) databases (for CNVs overlapping >250 genes, all genes were not systematically searched).

<sup>a</sup>CNVs assessed independently by two clinical cytogenetic laboratory directors to be 'Pathogenic' or of 'Uncertain clinical significance; likely pathogenic' (see Materials and Methods) (21).

<sup>b</sup>Recurrent 1.5–3 Mb 22q11.2 deletions ( $n = 5$ , all meeting syndromic criteria) are not shown; of these, two are from the community catchment population and one is not of European ancestry. Bold cytoband indicates a novel association with schizophrenia; see Fig. 1. The other four loci (1q21.1, 15q11-q13, 15q13.2-q13.3, 16p11.2) are displayed in Supplementary Material, Fig. S2.

<sup>c</sup>Previously reported by our group: 1q21 gains (46), 6p25 loss (58), 8p23 loss (2), 19p13 gain (2), XYY (2).

<sup>d</sup>Gene implicated in a recent next-generation sequencing study of schizophrenia (34).

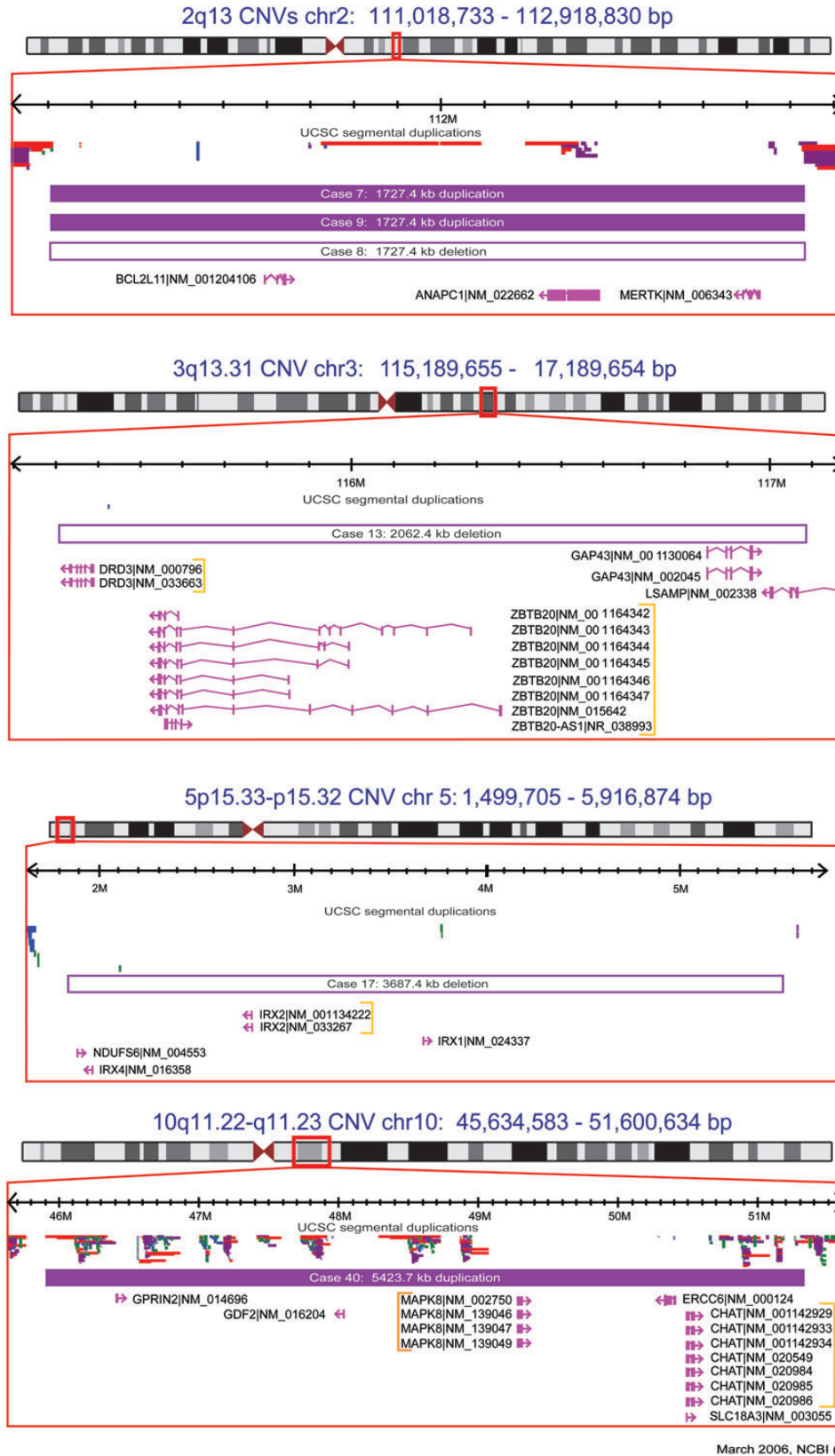
<sup>e</sup>Gene overlapped by 1.8 Mb gain in Case 2 only.

<sup>f</sup>Gene(s) overlapped by one or more rare CNVs in other unrelated probands with schizophrenia in our sample: *PDZKI*, 826 kb exonic gain (see Table 2 and Supplementary Material, Fig. S2 for details); GABA receptor gene cluster, 2.4 Mb exonic gain (see the text, Table 2, and Supplementary Material, Fig. S2 for details); *CHRNA7*, 54 kb exonic loss (see the text and Supplementary Material, Fig. S2 for details); *TRPM1*, 28 kb exonic gain [see the text and Supplementary Material, Fig. S2 for details; also implicated in a previous CNV study of schizophrenia (67)]; *CSMD1*, 65 kb exonic, and 13 kb, 17 kb, 92 kb and 117 kb intronic, losses [also implicated in previous CNV studies of schizophrenia (8,11)]; *ILIRAPL1*, 3.1 Mb exonic loss [see the text and Table 2 for details] and 60 kb intronic loss (also implicated in a previous CNV study of schizophrenia (11)).

<sup>g</sup>Met criteria for syndromic subgroup (2,17,18) (see Materials and Methods for details).

<sup>h</sup>Maternal origin of 15q11–q13 gain, per results from Multiplex Ligation-dependent Probe Amplification analysis to determine methylation status of the imprinted gene *SNRPN* (see Materials and Methods for details).

<sup>i</sup>Displayed in Supplementary Material, Fig. S3.



**Figure 1.** Large, rare, clinically significant CNVs, absent in 2773 (population-based + OPGP) controls, at loci novel to schizophrenia (4,22–25). Solid and open bars represent gains (duplications) and losses (deletions), respectively. Genomic parameters are from NCBI Build 36. Only selected genes are shown (see Table 1 for details).

significant CNVs that would be detectable on a clinical microarray. In the 248 unrelated case subjects in this catchment area with CNV data, the number of individuals with these clinically significant CNVs, including two with 22q11.2 deletions, was 15 (6.0%, 95% CI 3.7–9.8%) (Table 1). Nine (60.0%) of these 15 individuals were in the non-syndromic subgroup (i.e., were a priori found to not meet our established criteria for syndromic features; see Materials and Methods). If the five anomalies >6.5 Mb were included (Table 1), the estimated prevalence would be 8.1% (95% CI 5.2–12.2%). Notably, none of these 20 variants had been detected prior to study participation. We estimate that there are a total of 370 unrelated adults with schizophrenia in this catchment area (i.e.,  $n = 122$  not included in this study, in addition to the  $n = 248$  with CNV data; see Materials and Methods). A conservative estimate of the minimum prevalence of clinically significant CNVs in this catchment area, based on the assumption of no clinically significant variants in the estimated  $n = 122$  unrelated individuals unavailable for study, would thus be 20 in 370 or 5.4% (95% CI 3.5–8.3%).

### Large rare CNVs of unknown significance in schizophrenia

Table 2 shows the remaining large rare CNVs in schizophrenia cases deemed to be of ‘Uncertain clinical significance (no sub-classification)’ (21) (termed ‘variants of unknown significance’; VUS), or of ‘Uncertain clinical significance; likely benign’ or ‘Benign’ (21) (collectively termed ‘benign’). Of these, VUS [33 of 420 versus 12 of 416;  $P = 0.0012$ , OR 2.87 (95% CI 1.46–5.64)], but not benign CNVs, were significantly enriched in schizophrenia cases of European ancestry compared to OPGP controls. There were three novel loci where there were overlapping large rare CNVs in unrelated schizophrenia cases: 6q11.1 (two gains), 7p21.3 (one gain, one loss) and 12q21.31 (one gain, one loss) (Table 2; Supplementary Material, Fig. S4). All were VUS, except for the smaller of the 6q11.1 gains (Table 2), making the latter a less likely candidate region for schizophrenia. The 12q21.31 CNVs overlap *LIN7A*, an intriguing candidate gene (Fig. 2) that interacts with *DLG1*, *DLG2* and *GRIN2B*, and is implicated in postsynaptic density functions (8). There was also a 2.4 Mb gain at 15q12–q13.1 (a VUS), nested within the typical 15q11–q13 duplication region, that overlapped just five genes, including three gamma-aminobutyric acid (GABA) receptor component genes (Table 2; Fig. 2, Supplementary Material, Fig. S2) (30). Other large rare CNVs in single case subjects overlapped CNVs previously reported in schizophrenia or other neuropsychiatric disorders, involving genes of interest including *RYR2*, *MYTIL*, *LRP1B*, *IL1RAPL1* and *ZNF804A* (Table 2) (8,11,31–33). Interestingly, several genes we had annotated as candidates were subsequently implicated in recent next-generation sequencing studies of schizophrenia (34,35) (Tables 1, 2).

### Genome-wide CNV burden in schizophrenia

The proportion of subjects with one or more large (500 kb–6.5 Mb), rare (present in <0.1% of population controls) autosomal CNVs was significantly greater in the schizophrenia cases than in the OPGP controls [62 of 420 versus 21 of 416;

$P < 0.0001$ , OR 3.26 (95% CI 1.95–5.45)] (Table 3). As described above, this finding was driven by clinically significant CNVs and VUS, and not benign variants. We a priori excluded 22q11.2 deletions and variants >6.5 Mb from these analyses (see Materials and Methods) to demonstrate that this enrichment is robust to the exclusion of variants that might reasonably be discovered without the use of chromosomal microarray technology. Case–control differences were most marked for large rare exonic gains [38 of 420 versus 11 of 416;  $P < 0.0001$ , OR 3.66 (95% CI 1.85–7.27)] (Table 3). Within the schizophrenia sample, the subgroup a priori designated as ‘syndromic’ using previously published clinical screening criteria for adults (17,18) was significantly enriched for subjects with one or more large rare CNVs only when restricting to exonic losses [7 of 73 syndromic cases versus 12 of 347 non-syndromic cases;  $P = 0.0220$ , OR 2.96 (95% CI 1.12–7.80)] (Table 3).

The schizophrenia group was also significantly enriched for subjects with two or more rare CNVs >10 kb in size that overlapped exons [122 of 420 versus 89 of 416;  $P = 0.0109$ , OR 1.50 (95% CI 1.10–2.06)] (Table 3). Further restricting to very rare CNVs (i.e., those found in none of the 2357 adjudication controls) strengthened these results [78 of 420 versus 44 of 416;  $P = 0.0012$ , OR 1.93 (95% CI 1.30–2.87)]. In contrast, as expected (6,18,19), the overall CNV profile (unrestricted, e.g., by rarity or size) was similar for schizophrenia cases and OPGP controls (Supplementary Material, Tables S1 and S2). The results were also nonsignificant for subjects with one or more, or with two or more, rare CNVs unrestricted by exonic status (Table 3). The results for all analyses were similar if subjects of non-European ancestry were included.

### Very rare, smaller CNVs identifying genes of interest for schizophrenia

Table 4 shows 15 candidate genes identified, independent of gene enrichment mapping, using a strategy to determine very rare CNVs that overlapped the same gene in two or more schizophrenia cases and in no controls. Five unrelated schizophrenia cases had loss CNVs overlapping the *RBF0X1* gene (previously *A2BP1*) (11) and four had intronic loss CNVs overlapping the *SOX5* gene (36). Arguably, the most persuasive of candidates may involve genes of smaller genomic extent (<200 kb) (20), where CNVs overlap exons; five genes met these conservative criteria: *DNM1L*, *HIST3H3*, *JAK2*, *LIMS1* and *PPP3CC*. Singleton very rare CNVs (Table 4) implicated other potential candidates, including *DISC1*, *GRK4*, *GRM4*, *PIK3C3* and *RELN* with overlapped exons, *GRIK1*, *GRIN2A*, and *GRM7*, with introns overlapped only, and individual genes within large CNVs (Tables 1, 2), e.g., *IL1RAPL1*, *TRPM1* (11,32). Several of these candidates (*DISC1*, *IL1RAPL1*, *JAK2*, *PIK3C3*, *RELN*, *TRPM1*) converged with functional gene enrichment mapping results using loss CNVs (Fig. 2; Supplementary Material, Fig. S6). There was also convergence with multiple genes implicated in two recent next-generation sequencing studies of schizophrenia (34,35), including 10 genes (*CACNA2D1*, *GRIN2A*, *GRM7*, *JAK2*, *NRXN1*, *PIK3C3*, *PTGER3*, *PTPRT*, *SLC1A1* and *USH2A*) implicated by multiple rare single

**Table 2.** Large (>500 kb) rare CNVs of uncertain pathogenicity discovered in unrelated probands with schizophrenia

Subject Case #	Catchment area	CNV characteristics			CN	Very rare	Flanking segmental duplications	# of genes	VUS	Candidate gene(s)
		Cytoband	Start	Size (bp)						
1		1q21.1 <sup>b</sup>	143 677 127	826 295	Gain		•	21	•	<i>PDE4DIP<sup>g</sup>, PDZK1<sup>c</sup>, PEX11B</i>
4		1q43	235 349 594	726 228	Gain	•		1	•	<i>RYR2<sup>c</sup></i>
5	•	2p25.3	1 377 419	745 146	Gain	•		3	•	<i>MYT1L<sup>c</sup></i>
6		2q11.1	94 691 614	735 772	Gain	•	•	8	•	<i>KCNIP3</i>
10	•	2q22.1	139 898 997	1 322 193	Loss	•		1	•	<i>LRP1B<sup>c,g</sup></i>
11		2q32.1	185 043 646	790 869	Loss	•		1	•	<i>ZNF804A<sup>c</sup></i>
12		3p12.1	83 940 054	1 038 151	Loss			1	•	–
14	•	4q25-q26	113 634 704	783 442	Gain	•		10	•	<i>ANK2<sup>g</sup>, NEUROG2</i>
15	•	4q33-q34.1	171 556 613	2 693 129	Loss	•		1	•	–
16	•	4q34.3-q35.1	179 878 817	4 343 697	Gain	•		6	•	<i>ODZ3</i>
18	•	5p15.1	15 134 588	628 993	Gain	•		1	•	–
19		5p13.2	37 294 917	520 350	Gain			2	•	–
20		5q13.3-q14.1	74 706 210	2 764 437	Gain	•		20	•	<i>OTP, PDE8B, SV2C<sup>g</sup></i>
22	•	6q11.1 <sup>d</sup>	61 944 399	1 031 545	Gain	•		2	•	<i>KHDRBS2, MTRNR2L9</i>
23		6q11.1 <sup>d</sup>	61 944 399	502 986	Gain			1		<i>MTRNR2L9</i>
24	•	6q14.3	86 804 893	577 156	Gain	•		0		–
25	•	6q16.2-q16.3	98 969 135	1 408 033	Gain	•		9	•	<i>POU3F2</i>
26	•	7p21.3 <sup>d</sup>	9 580 685	819 384	Loss	•		1	•	<i>PER4</i>
27	•	7p21.3 <sup>d</sup>	9 573 572	1 080 196	Gain	•		1	•	
28		7q31.31	118 515 093	1 204 095	Gain			1	•	<i>KCND2</i>
29	•	8p23.3	439 294	733 309	Gain	•	•	2	•	–
30		8p11.21-p11.1	43 132 979	772 981	Gain			2		<i>HGSNAT</i>
31		8q12.1	60 901 679	766 789	Gain	•		2	•	<i>CA8, RAB2A</i>
32	•	8q21.3	89 566 045	1 073 381	Gain	•		0		–
33		9p21.1	30 791 693	549 623	Loss	•		0		–
34		9p13.3-p13.2	36 173 229	862 436	Loss	•		5	•	<i>CLTA, GNE, PAX5<sup>g</sup></i>
35		9p12	41 440 153	535 035	Gain	•	•	6		–
36	•	9p11.2	44 936 301	570 215	Gain	•	•	0		–
37		9q12	65 370 766	902 773	Loss	•	•	1		–
38	•	9q21.11-q21.12	72 107 555	964 557	Loss	•		4	•	<i>KLF9, TRPM3<sup>g</sup></i>
39	•	10p11.22	33 636 052	651 356	Gain	•		1	•	<i>NRP1</i>
42	•	11q22.1	97 224 044	626 127	Loss	•		0		–
43	•	12p11.1	34 206 228	539 538	Gain			0		–
39	•	12q21.31 <sup>d</sup>	79 255 527	568 008	Gain	•		6	•	<i>LIN7A</i>
44	•	12q21.31 <sup>d</sup>	79 679 461	1 397 367	Loss	•		5	•	
45		13q13.1-q13.2	32 108 632	2 434 026	Loss	•		5	•	<i>KL<sup>g</sup>, NBEA<sup>g</sup>, PDS5B</i>
46	•	13q21.32-q21.33	66 742 848	993 434	Gain	•		0		–
47		14q21.1-q21.3	40 497 844	2 771 925	Gain	•		1	•	<i>LRFN5</i>
51	•	15q12-q13.1 <sup>b</sup>	23 766 255	2 437 700	Gain	•	•	5	•	<i>GABRB3<sup>c</sup>, GABRA5<sup>c</sup>, GABRG3<sup>c</sup></i>
36	•	15q13.2 <sup>d</sup>	28 153 539	564 254	Gain	•	•	7		<i>CHRFAM7A</i>
53		16p13.11	14 805 302	1 498 712	Gain		•	24	•	<i>NDE1, NOMO3, NTANI</i>
54		16p13.11	15 032 942	1 388 198	Gain		•	20	•	
59	•	16p11.1-p11.2	34 324 072	816 656	Gain			4		–

Continued

Table 2. Continued

Subject Case #	Catchment area	CNV characteristics		Size (bp)	CN	Very rare	Flanking segmental duplications	# of genes	VUS	Candidate gene(s)
		Cytoband	Start							
60		17p13.1	10 552 154	607 279	Gain	•		5	•	–
61	•	19q12	32 648 879	2 916 074	Loss	•		10	•	<i>C19orf12<sup>g</sup></i> , <i>ZNF536</i>
62		20p12.1	14 305 685	1 108 941	Loss	•		1	•	<i>MACROD2<sup>h,c</sup></i>
63	•	22q12.1	25 696 811	1 304 432	Loss	•		6	•	<i>PITPNB</i>
64 <sup>f</sup>	•	Xp21.3-p21.2	27 619 068	3 129 173	Loss	•	•	11	•	<i>ILIRAPLI<sup>c,e</sup></i>
65		Xq11.1	63 283 656	721 753	Gain	•	•	3	•	–
66	•	Xq21.31	86 334 886	839 884	Gain	•		1	•	–

Case #, case number for subjects ( $n = 454$ ) with schizophrenia ( $n = 6$  were not of European ancestry: Cases 35, 37, 43, 60, 278, 279); Catchment area, subjects originating from the only community mental health clinic in a specific catchment area of ~150 000 people (•), see the text for details; Cytoband, cytogenetic location of CNV; CNV start, hg18 (NCBI Build 36.1, March 2006); CNV size, in base pairs; CN, type of copy number aberration; Very rare, not found in 2357 population controls and absent in OPGP controls (•), see the text for details; Flanking segmental duplications, known flanking segmental duplications (from the UCSC Genome Browser hg18 version) that cover at least 20% of the CNV length (•); # of genes, number of known genes overlapped by CNV as annotated in the Database of Genomic Variants (<http://projects.tcag.ca/variation/>; September 2011); VUS, variant of unknown significance (•) as assessed independently by two clinical cytogenetic laboratory directors (variants not so annotated were considered likely benign and not reportable clinically); Candidate gene(s), selected based on the reported neuropsychiatric/neurodevelopmental phenotype identified from systematic searches of human (e.g. Online Mendelian Inheritance in Man; <http://www.omim.org/>) and model organism (e.g. Mouse Genome Informatics; <http://www.informatics.jax.org/>) databases. There were no such candidate genes identified for 11 genic CNVs. <sup>g</sup>Five large CNVs in schizophrenia cases of European ancestry that were present in the OPGP controls, using a 50% reciprocal overlap criterion (19), are not shown (Supplementary Material, Table S4): 3p14.2 gain in Case 263, 9q12 loss in Case 172, 11q11 gain in Case 41, and two 14q32.33 losses in Cases 156 and 280.

<sup>h</sup>Displayed in Supplementary Material, Fig. S2.

<sup>c</sup>Gene implicated in a previous CNV study of schizophrenia: *RYR2* (8), *MYT1L* (31), *LRP1B* (32), *ZNF804A* (33), *MACROD2* (8), *ILIRAPLI* (11).

<sup>e</sup>Displayed in Supplementary Material, Fig. S4.

<sup>f</sup>Gene(s) overlapped by one or more additional rare CNVs in unrelated probands with schizophrenia in our sample: *PDZK1*, 1.8 Mb exonic gain (see Table 1 and Supplementary Material, Fig. S2 for details); GABA receptor gene cluster (*GABRB3*, *GABRA5*, *GABRG3*), 5.0 Mb, 5.7 Mb and 6.5 Mb exonic gains (see the text, Table 1, and Supplementary Material, Fig. S2 for details); *MACROD2*, 14 kb gain and 15 kb, 25 kb and 63 kb losses; *ILIRAPLI*, X chromosome aneuploidies (see the text and Table 1 for details) and 60 kb intronic loss.

<sup>f</sup>Female subject, therefore judged to be a VUS instead of 'Uncertain clinical significance; likely pathogenic' (21).

<sup>g</sup>Gene implicated in a previous next-generation sequencing study of schizophrenia: *PDE4DIP* (35), *LRP1B* (34), *ANK2* (34), *SV2C* (34), *PAX5* (34), *TRPM3* (34), *KL* (34), *NBEA* (34), *C19orf12* (34).

nucleotide variants (SNVs) (34), and one gene (*PTPRM*) implicated by a *de novo* rare SNV (35) (Table 4).

also novel clusters of gene sets involved in cell motility, signal transduction and transcription (Supplementary Material, Fig. S7).

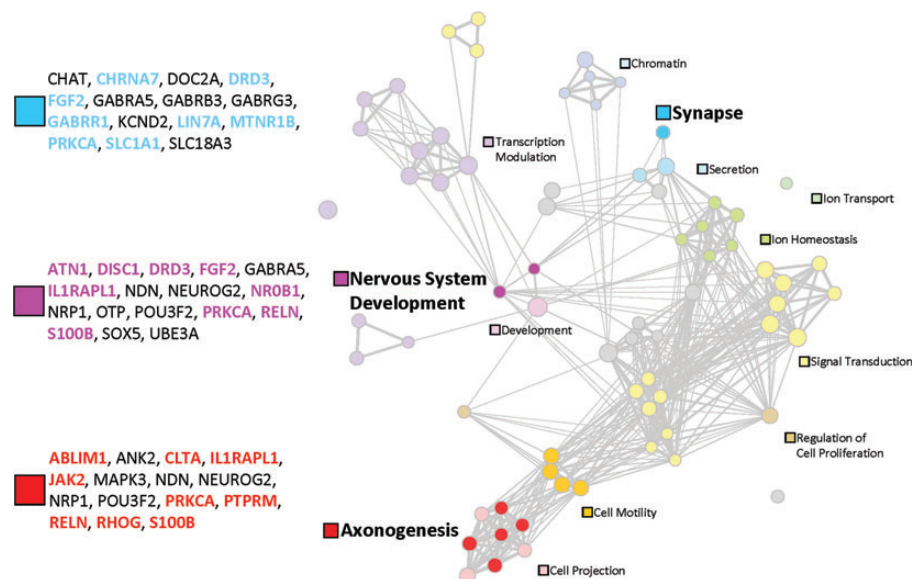
### Network of gene sets revealed by functional gene enrichment mapping

Figure 2 summarizes functional gene enrichment mapping results (6,18,37). For 70 (2.9%) of 2456 gene sets tested, we found a significant enrichment of very rare, exonic loss CNVs in schizophrenia cases compared with OPGP controls (OR 1.9, 95% CI 1.0–3.6), involving 73 support genes (Supplementary Material, Tables S5 and S6, Figs S5 and S6; Fig. 2). Considering just the three functional clusters comprising eight gene sets most clearly related to neurodevelopmental and synaptic processes (Supplementary Material, Table S7; Fig. 2), the effect was greater (OR 4.7, 95% CI: 1.3–26.0), involving 19 support genes for schizophrenia (Fig. 2). The other clusters having a less obvious functional scope may suggest more novel aspects of a gene network for schizophrenia (Fig. 2; Supplementary Material, Figs S5 and S6). Notably, there was a significant degree of functional cluster overlap in a formal cross-disease comparison of this schizophrenia network with known genes for ASD (Supplementary Material, Fig. S7). This included not only the eight gene sets annotated as neurodevelopmental and synaptic, but

### DISCUSSION

Derived from a clinically informed and rigorous approach, the results of this study provide new information on the genomic architecture of schizophrenia and on the potential role of clinical microarray testing. Adding novel data to those available from the component studies of existing consortia reports, this study involved (i) unparalleled systematic recruitment efforts in a community setting (2), (ii) consistent phenotyping of all cases using previously published methods (17,18), (iii) rigorous CNV methodology (18), (iv) exclusion of already known effects of 22q11.2 deletions, (v) blinded adjudication of rare CNVs by two independent clinical cytogenetic laboratory directors and (vi) clinical confirmation and return of results to all patients with clinically significant CNVs and their families. The results strengthen etiologic and mechanistic ties of schizophrenia to developmental disorders like autism (3,4,6), where clinical microarray testing is now indicated (38), and suggest a yield of reportable findings approaching that seen in these disorders (38,39). Rare, recurrent structural rearrangements at 2q13





**Figure 2.** A functional map of schizophrenia using results of the gene-set association analysis, displayed as a network of 70 gene sets (circles) related by mutual overlap (lines) using the Enrichment Map Cytoscape plugin (37). Circle size is proportional to the total number of ‘support’ genes in each gene set and line thickness represents the number of genes in common between two gene sets. Each support gene harbors a rare exonic loss CNV in one or more schizophrenia subjects but in no OPGP controls; only CNVs overlapping 15 or fewer genes were tested for gene-set association. Groups of functionally related gene sets are represented by the filled circles and respective labels; grey circles indicate gene sets not assigned to these 13 clusters. Three major functional groups are further highlighted by darker filled circles (blue, Synapse; pink, Nervous system development; red, Axonogenesis) with the respective 19 support genes for schizophrenia shown in color. We also list (in black) other genes in these three functional clusters that were overlapped by very rare loss or gain CNVs in subjects with schizophrenia (i.e., not in 2773 controls), and that have additional supportive data (see the text and Tables 1, 2 and 4). See Materials and Methods for more details.

were implicated for the first time as significant risk factors for schizophrenia. We provide strong evidence supporting a quantitatively greater, and qualitatively different, genome-wide burden of rare CNVs in schizophrenia compared with controls. New candidate genes and networks relevant to schizophrenia identified in this study will facilitate the future clinical interpretation of smaller CNVs.

### Clinically significant findings

The results of this study indicate that ~1 in 13 unrelated patients with schizophrenia may harbor a clinically significant large rare structural variant that would be detectable using methods available in most clinical laboratories. This yield is similar to that seen in a 2010 clinically driven study of genetic testing in ASD (39), a related neurodevelopmental condition in which clinical microarray testing is now considered the first-tier diagnostic test (38). None of the patients in the schizophrenia sample studied had autism or multiple congenital anomalies and few would meet the existing criteria for clinical microarray testing, e.g., on the basis of mild intellectual disability (38). Our data also suggest that most individuals with schizophrenia who have large rare CNVs would not have obvious developmental (syndromal) features, with the exception of some with large exonic losses where phenotypes tend to be more complex, as in 22q11.2 deletions (2–4,17,18). Thus, to discover most of the clinically significant variants identified in this study a genome-wide microarray would be necessary. Introduction of such genetic testing into clinical practice for an adult neuropsychiatric disease would have paradigm-shifting implications for psychiatry and for the still pediatric-focused field of medical genetics. As for ASD and developmental delay/intellectual disability (13),

the availability of clinical CNV data for schizophrenia, coupled with an improved ability to interpret these data for the benefit of patients and families, could have far-reaching scientific and clinical implications. Further studies are needed to demonstrate the worth of such clinical genetic testing in schizophrenia.

What are the potential benefits of identifying a ‘clinically significant’ CNV? The impetus to offer clinical microarray testing to individuals with schizophrenia in the hope of arriving at a partial etiologic explanation is similar to that described for other neurodevelopmental conditions (38–44). Uncertainty about penetrance and expressivity of rare variants is a common reality of practice in medical genetics, and remains compatible with the provision of helpful genetic counseling. With the exception of 22q11.2 deletions (45), most clinically significant variants in individuals with schizophrenia today would only be discovered on a research basis. More needs to be known about lifetime, including adult, expression (1,7,46) of these CNVs to better inform medical and/or psychiatric management. Nevertheless, all of these large, rare, clinically significant variants already have implications for genetic counseling that include, but are not limited to, reproductive implications for siblings and patients themselves (47), and are of potential benefit to patients and families (40,41,44). Our own efforts with this translation to clinical genetic counseling are ongoing, with an overall positive response to date, comparable with our experience with 22q11.2 deletion syndrome (40) and schizophrenia in general (48,49). There may be, therefore, a growing obligation to offer the return of selected, clinically relevant and clinically confirmed individual research results to patients and their families in large-scale genome-wide studies of schizophrenia such as this (50). To our knowledge, this is currently a rare practice, even for 22q11.2 deletions.

**Table 3.** Rare autosomal CNV burden in 420 unrelated adults of European ancestry with schizophrenia<sup>a</sup>

	Schizophrenia cases versus OPGP controls		All schizophrenia cases (n = 420)		Analysis		Schizophrenia cases		Analysis			
	OPGP controls (n = 416)	n (%)	All schizophrenia cases (n = 420)	n (%)	P	OR (95% CI)	Non-syndromic cases (n = 347)	n (%)	Syndromic cases (n = 73)	n (%)	P	OR (95% CI)
Large rare CNVs (> 500 kb)												
Loss or gain	21 (5.05)	62 (14.76)			<b>&lt;0.0001</b>	<b>3.26</b> (1.95) (5.45)	46 (13.26)	16 (21.92)		NS	1.84 (0.97) (3.47)	
Exonic loss or gain	18 (4.33)	57 (13.57)			<b>&lt;0.0001</b>	<b>3.47</b> (2.01) (6.01)	42 (12.10)	15 (20.55)		NS	1.88 (0.98) (3.61)	
Exonic loss	7 (1.68)	19 (4.52)			<b>0.0180</b>	<b>2.77</b> (1.15) (6.66)	12 (3.46)	7 (9.59)		<b>0.0220</b>	<b>2.96</b> (1.12) (7.80)	
Exonic gain	11 (2.64)	38 (9.05)			<b>&lt;0.0001</b>	<b>3.66</b> (1.85) (7.27)	30 (8.65)	8 (10.96)		NS	1.30 (0.57) (2.97)	
All rare CNVs (any size)												
Loss or gain <sup>b</sup>	372 (89.42)	376 (89.52)			NS	1.01 (0.65) (1.57)	310 (89.34)	66 (90.41)		NS	1.12 (0.48) (2.63)	
Exonic loss or gain	249 (59.86)	264 (62.86)			NS	1.14 (0.86) (1.50)	217 (62.54)	47 (64.38)		NS	1.08 (0.64) (1.83)	
All rare CNVs (any size)												
Loss or gain <sup>b</sup>	264 (63.46)	282 (67.14)			NS	1.18 (0.88) (1.56)	232 (66.86)	50 (68.49)		NS	1.08 (0.63) (1.85)	
Exonic loss or gain	89 (21.39)	122 (29.05)			<b>0.0109</b>	<b>1.50</b> (1.10) (2.06)	98 (28.24)	24 (32.88)		NS	1.24 (0.72) (2.14)	

NS = non-significant (P > 0.05). Significant results are shown in bold font.

<sup>a</sup>Rare = CNVs with <0.1% prevalence in 2357 population controls of European ancestry (used to adjudicate both schizophrenia cases and OPGP controls). Rare CNVs <10 kb or >6.5 Mb in size, and all sex chromosome CNVs, were excluded (see Materials and Methods). Five syndromic subjects of European ancestry in the schizophrenia group with typical 1.5–3 Mb 22q11.2 deletions (i.e., exonic losses) were not included in the study (see Materials and Methods). CNV burden was measured using the proportion of subjects with one or more, or with two or more, rare CNVs; measuring rare CNV burden (all sizes) for cases and controls using the total number of genes overlapped by CNVs or total genomic extent of CNVs showed similar results.

<sup>b</sup>Results were similarly nonsignificant for rare CNVs involving losses only and gains only, and for non-exonic losses and non-exonic gains

### Toward a more sophisticated approach to the clinical assessment of genome-wide CNVs in schizophrenia

The new candidate genes and network highlighted in this report will facilitate the clinical interpretation of genome-wide CNVs in schizophrenia, including CNVs that are neither large nor recurrent. We took a conservative approach to enrichment mapping, using only exonic losses in the analyses and excluding major multigenic CNVs. The results showed that the genes implicated by rare CNVs in schizophrenia are especially over-represented by those with synaptic and neurodevelopmental functions, consistent with other studies (8,36). Moreover, the findings extended attention from such well-recognized candidates to less obvious genes. Importantly, these were not restricted to genes with multiple functions (e.g., *DISC1* and *RELN*). The gene sets implicated in this study may assist with the interpretation of smaller genic CNVs in schizophrenia. This more complex gene network for schizophrenia was similar to that seen in autism (6). For the first time, a formal analysis has demonstrated that in schizophrenia there is significant over-representation of known genes for ASD (Supplementary Material, Fig. S7). This finding provides additional evidence in support of using data from ASD to assist in the clinical interpretation of CNVs in schizophrenia, and vice versa. The fact that none of the study participants had a history of autism or moderate to severe intellectual disability supports the likelihood that CNVs shared with these pediatric conditions exhibit true pleiotropy (51). Many of the implicated genes in this study show regional gene expression differences in mid-fetal development of the brain (e.g., *CHRNA7* and *LIN7A* in the thalamus) (52), a critical time for the pathogenesis of schizophrenia (53). We note, however, that the prolonged maturation of the human brain provides multiple opportunities for perturbation of gene expression through gene dosage changes (53).

Complementary to these results, we demonstrated enrichment in schizophrenia of not only individuals with large rare exonic CNVs, but also those with two or more rare CNVs >10 kb that overlapped exons. These results represent some of the first concrete evidence supporting a ‘multiple hit’ hypothesis for schizophrenia, outside of multigene large CNVs (7,10,12,36). There is also support for regulatory (e.g., large intronic CNVs) and epigenetic mechanisms (e.g., chromatin modification cluster in the enrichment map gene network) as important contributory factors to the etiology of schizophrenia. There remained such evidence after excluding the potentially overpowering influence of multigenic CNVs. The new candidate genes and network for schizophrenia identified in this report may also help guide the future interpretation of results from whole exome and whole genome sequencing studies (34). Indeed, there was a notable degree of overlap between the promising candidate genes overlapped by rare CNVs in our sample (Tables 1, 2, 4) and those genes containing rare SNVs in schizophrenia cases in recent next-generation sequencing studies (34,35).

### Advantages and limitations

This is the first study of genome-wide CNVs in schizophrenia to include clinical adjudication of identified rare CNVs and to use a systematic clinically informed genetic approach to investigate a well-characterized cohort of adults assessed a priori for certain

**Table 4.** Putative candidate genes for schizophrenia overlapped by very rare (<500 kb) CNVs

Candidate gene	Cytoband	Gene size (nt)	Case #	CNV start	CNV size (bp)	CN	Exonic
<i>Very rare CNVs overlapping the same CNS-related gene in two or more unrelated schizophrenia cases</i>							
<i>CAMTA1</i>	1p36.31	984 383	67	7 063 417	33 667	Loss	•
			68	7 545 052	44 523	Loss	
<i>DNM3</i>	1q24.3	571 237	87	170 223 915	47 814	Loss	•
			168	170 450 307	4136	Loss	
<i>HIST3H3</i>	1q42.13	481	69	226 634 235	54 736	Gain	•
			70	226 673 588	109 976	Gain	•
<i>COMMD1</i>	2p15	230 403	66	62 049 330	36 475	Loss	•
			32	62 106 806	24 384	Loss	
<i>LIMS1</i>	2q13	152 892	71	108 619 802	61 009	Loss	•
			72	108 619 802	61 583	Loss	•
<i>DPP6</i>	7q36.2	679 607	2	153 637 309	9260	Loss	
			73	153 737 928	6935	Loss	
			74	154 060 964	31 082	Loss	•
<i>PPP3CC</i>	8p21.3	100 043	75	22 265 902	166 778	Gain	•
			76	22 265 902	166 778	Gain	•
<i>JAK2<sup>a,b</sup></i>	9p24.1	142 939	7	4 527 834	486 659	Loss	•
			77	5 014 345	82 361	Gain	•
<i>ERC1</i>	12p13.33	504 696	78	865 867	157 163	Gain	•
			50	1 255 099	369 587	Loss	•
<i>SOX5<sup>a</sup></i>	12p12.1	1 030 150	79	23 826 011	11 461	Loss	
			80	23 828 090	9382	Loss	
			81	24 063 813	25 105	Loss	
			82	24 561 569	35 937	Loss	
<i>DNMIL</i>	12p11.21	66 448	83	32 679 398	45 218	Gain	•
			84	32 679 398	47 929	Gain	•
<i>RBFOX1<sup>a</sup></i>	16p13.2	1 694 209	54	6 171 253	32 382	Loss	
			85	6 754 460	50 650	Loss	•
			86	6 766 601	9285	Loss	
			87	6 973 749	104 964	Loss	•
			62	6 992 360	143 753	Loss	•
<i>PRKCA<sup>a</sup></i>	17q24.2	507 937	61	61 941 341	277 949	Loss	•
			88	62 145 527	7081	Loss	
<i>DOK6</i>	18q22.2	448 040	89	65 614 643	10 824	Loss	
			90	65 639 534	10 209	Loss	
<i>PTPRT<sup>b</sup></i>	20q12	1 117 166	91	40 639 325	86 602	Loss	
			92	40 778 749	187 266	Gain	•
			93	40 887 741	7643	Loss	
<i>Very rare CNVs overlapping promising candidate genes for schizophrenia found in a single schizophrenia case</i>							
<i>PTGER3<sup>b</sup></i>	1p31.2	195 456	141	71 150 238	44 186	Loss	•
<i>USH2A<sup>b</sup></i>	1q41	800 503	104	213 914 685	64 806	Loss	•
<i>DISC1</i>	1q42.2	414 458	38	230 150 208	121 262	Loss	•
<i>NRXN1<sup>a,b</sup></i>	2p16.3	1 114 032	170	503 371 141	18 515	Loss	
<i>GRM7<sup>a,b</sup></i>	3p26.1	880 417	146	7 058 814	35 895	Loss	
<i>GRK4</i>	4p16.3	77 132	100	2 994 521	31 371	Loss	•
<i>DOK7<sup>a</sup></i>	4p16.2	31 177	169	3 439 241	13 595	Gain	•
<i>FGF2</i>	4q26	71 528	198	123 980 237	58 004	Loss	•
<i>GRM4</i>	6p21.31	111 816	95	34 175 770	43 039	Gain	•
<i>RUNX2</i>	6p21	222 766	231	45 129 193	298 662	Loss	•
<i>GABRR1</i>	6q15	40 274	182	89 894 102	52 079	Loss	•
<i>CALN1<sup>a</sup></i>	7q11.22	632 885	223	71 331 369	131 729	Loss	•
<i>CACNA2D1<sup>b</sup></i>	7q21.11	493 614	166	81 829 583	187 332	Gain	•
<i>SEMA3A</i>	7q21.11	236 559	195	83 380 795	87 642	Gain	•
<i>RELN</i>	7q22.1	517 733	131	103 178 239	90 728	Loss	•
<i>SLC1A1<sup>a,b</sup></i>	9p24	97 043	7	4 527 834	486 659	Loss	•
<i>ABLIM1</i>	10q25	253 546	200	116 351 384	17 427	Loss	•
<i>RHOG</i>	11p15.4	14 006	188	3 817 748	11 292	Loss	•
<i>MTNR1B</i>	11q21	13 160	240	92 336 224	10 613	Loss	•
<i>WNT5B</i>	12p13.3	30 157	50	1 255 099	369 587	Loss	•
<i>ATN1<sup>c</sup></i>	12p13.31	17 859	208	6 837 052	74 565	Loss	•
<i>ENO2<sup>c</sup></i>	12p13.31	9 246	208	6 837 052	74 565	Loss	•
<i>GRIN2A<sup>b</sup></i>	16p13.2	429 347	32	9 982 679	10 953	Loss	
<i>PTPRM<sup>a,b</sup></i>	18p11.2	839 546	119	7 899 131	55 764	Loss	•
<i>PIK3C3<sup>b</sup></i>	18q12.3	126 250	132	37 880 681	27 546	Loss	•
<i>AP3D1<sup>a</sup></i>	19p13.3	50 564	195	2 015 619	180 285	Loss	•
<i>JAK3</i>	19p13.11	23 251	271	17 754 374	68 190	Gain	•

Continued

Table 4. Continued

Candidate gene	Cytoband	Gene size (nt)	Case #	CNV start	CNV size (bp)	CN	Exonic
<i>GRIK1</i>	21q21.3	403 029	166	30 082 515	11 811	Loss	
<i>S100B</i>	21q22.3	6505	181	46 843 667	16 998	Loss	•

Candidate gene, selected based on the reported neuropsychiatric/neurodevelopmental phenotype identified from systematic searches of human (e.g. Online Mendelian Inheritance in Man; <http://www.omim.org/>) and model organism (e.g. Mouse Genome Informatics; <http://www.informatics.jax.org/>) databases; Cytoband, cytogenetic location of candidate gene; Gene size, in nucleotides; Case #, subjects from the discovery sample ( $n = 454$ ) with schizophrenia; CNV start, hg18 (NCBI Build 36.1, March 2006); CNV size, in base pairs; CN, type of copy number aberration; Exonic, CNV overlaps exon(s) of candidate gene (•).

<sup>a</sup>Gene implicated in a previous CNV study of schizophrenia: *JAK2* (32), *SOX5* (36), *RBFOX1* (11), *PRKCA* (32), *NRXN1* (9,10,36,67,68), *GRM7* (11,36), *DOK7* (32), *CALN1* (68), *SLC1A1* (11,68), *PTPRM* (36) and *AP3D1* (11).

<sup>b</sup>Gene implicated in a next-generation sequencing study of schizophrenia: *JAK2* (34), *PTPRT* (34), *PTGER3* (34), *USH2A* (34), *NRXN1* (34), *GRM7* (34), *CACNA2D1* (34), *SLC1A1* (34), *GRIN2A* (34), *PTPRM* (35) and *PIK3C3* (34).

<sup>c</sup>Genes overlapped by the same CNV.

'syndromic' features (2,17,18). To minimize the potential impact of false positives (16,20), we applied identical molecular and conservative analytic methods to unrelated cases with schizophrenia and to a similar-sized Canadian sample of epidemiologic controls, an unprecedented approach in neuropsychiatric disease recently validated in the study of congenital cardiac disease (18). The strategy emphasized CNVs likely to have an enhanced effect size because we used a more restricted level to define rarity ( $<0.1\%$ ) than that usually employed ( $<1\%$ ) (9–11). This could mean missing more common variants with a lower effect size. Using an independent population-based control set ensured equal effects for both cases and OPGP controls, however (see Materials and Methods), and thus all comparative analyses (including the burden and gene enrichment mapping) are robust to the potential effects of systematic Type 1 or 2 errors. Published guidelines for CNV analysis support our multi-algorithm approach (54), and we report a very high positive validation rate by independent methods (100% of 58 CNVs tested, including 17 smaller than 50 kb in size) that is in keeping with our previous experiences (5,6,18,19). All studies will be affected by the cases and controls studied, and the controls used to adjudicate rarity of CNVs. We acknowledge a focus on individuals of European ancestry. No genome-wide CNV results of our cases have been previously published, so there will be no sample overlap with other datasets.

Adjudicating findings for individually rare variants is challenging—indeed, this was a major rationale for our study. Blinded assessment of large rare CNVs by clinical laboratory directors provided a unique advantage. We also employed conservative qualitative and quantitative methods in order to identify putative candidate genes for schizophrenia overlapped by rare CNVs whose expression may be affected by the associated gene dosage changes. We report the details of a gene network derived from rare CNVs overlapping exons that differentiated schizophrenia cases from controls (Supplementary Material, Tables S5, S6 and S7, Figs S5 and S6; Fig. 2) and provide data on all rare CNVs identified in cases and controls (Supplementary Material, Table S4) to facilitate interpretation of findings of future studies (7). The network was further validated by demonstrating significant over-representation with known genes for ASD (Supplementary Material, Fig. S7). Nonetheless, all clinical and research interpretations are constrained by knowledge available at the time. We expect that the understanding of coding regions, and non-genic and regulatory sequences, will continue to evolve. Although the number of cases

in our study appears to be small in comparison to the combined numbers reported by large consortia, we were sufficiently powered to (i) arrive at a confidence interval (CI) for the prevalence of clinically significant CNVs in schizophrenia with a lower bound still  $>5\%$ , and (ii) demonstrate within-sample recurrence implicating individual genes (Table 4) and specific loci (Tables 1 and 2), including the novel discovery of enrichment of CNVs at the 2q13 locus in schizophrenia.

Our results provide clinical adjudication of previously reported CNVs for schizophrenia (7–11,32,36,46). Family studies, rare in schizophrenia, will be essential to delineate the inheritance status, penetrance and variable neuropsychiatric expression, segregation pattern and extended phenotype associated with those CNVs believed to underlie emerging genetic subtypes of schizophrenia (1,7). The exception is 22q11.2 deletion syndrome, where these are well established (1). *De novo* status is neither a necessary nor a sufficient criterion for pathogenicity, and many of the CNVs highlighted in this study may have been inherited. Many well-established schizophrenia risk variants of major effect are found in the majority of cases to be inherited (13), usually from parents with mild or no overt neuropsychiatric phenotype (4,47). We elected to use previously validated criteria for assessing syndromic features, shown to (i) have a good discriminant ability to identify at least one genomic disorder (22q11.2 deletion syndrome) (17) and (ii) be feasible for administration in a brief clinical encounter (2,17)—a key prerequisite for translation into the psychiatric clinic. There may exist more complex clinical screening criteria with sufficient discriminant ability to identify patients with large, rare, clinically significant CNVs. The collectively high prevalence of such variants, and the non-significant findings for clinical variables in this study, however, support the potential universal use of clinical microarray testing in schizophrenia. Definitively proving causality of specific genetic variants for schizophrenia is beyond the scope of our and other genome-wide CNV studies (5,6,8–11,18,19). Ideally, this would comprise direct functional evidence that can be related to gene dosage effects in the developing and adult human brain. For many candidate genes implicated by these CNV results, the functional significance has already been validated in model organism and/or human cellular models, as outlined in Tables 1 and 2. Such models are useful in a disease where access to the target organ at precise developmental time points is an impossibility.

Why does the prevalence of large, rare, recurrent CNVs in schizophrenia appear to be higher in this study than estimates

reported in previous studies (10)? First, this is a prospective study using a consistent ascertainment, recruitment and assessment strategy for patients with chronic schizophrenia at community clinics. Although demographic features and age at onset were similar to those in other studies, the chronicity of the illness may have meant that a relatively few individuals with mild forms of schizophrenia and/or evolving psychiatric diagnoses were included (55). Second, although all studies may be expected to have volunteer bias, we took an inclusive approach to recruitment. We did not exclude individuals, e.g., with learning difficulties or residual symptoms, and obtained a surrogate consent for individuals where necessary. Also, recruitment has persisted over several years, and particularly in our catchment-based sample we have made every effort to include all patients (e.g. waiting until their clinical state improved, making home visits). We did not recruit from university health clinics, use groups of patients already enrolled in and/or volunteering for other studies or preferentially enroll patients from intact families with both parents available for study or those with familial schizophrenia. Such factors can affect the observed prevalence of 22q11.2 deletions (7) and likely that of other rare CNVs with high penetrance for severe neuropsychiatric phenotypes. Nevertheless, given evidence for premature death in some CNV-related conditions (56), we would have missed individuals who died before enrollment and those too ill to participate. Replication of our findings in an independent cohort should be a key goal of future research in this area, and our results may encourage others to adopt our clinically driven approach, including more epidemiologically-based strategies for patient ascertainment.

## CONCLUSIONS

For the first time, we have a reliable estimate of the collective prevalence of clinically significant CNVs in a community-based sample of schizophrenia. Most individuals with large rare CNVs, including many of those with clinically significant CNVs, will not be readily distinguishable from the rest of the general schizophrenia population based on developmental (syndromic) features. The prevalence of clinically significant CNVs discovered in community-based schizophrenia brings us a step closer to recommending clinical microarray testing for patients with schizophrenia (10,38), and may initiate a debate among clinicians, scientists, families and policy makers. As for children with developmental disorders found to have clinically significant CNVs, clinical genetic counseling could be provided to patients with schizophrenia and their families (1,40), regardless of the acknowledged need for more information on key genetic parameters, such as penetrance and variable expression across the lifespan. The complex network of genes being discovered through studies of rare variants mirrors the complexity of the neural networks involved. Further research efforts promise to bring order to these for schizophrenia and related neuropsychiatric disorders (6), as they have for cancer.

## MATERIALS AND METHODS

### Schizophrenia sample ascertainment and assessment

We prospectively recruited 459 unrelated Canadian patients meeting DSM-IV criteria for chronic schizophrenia or

schizoaffective disorder from four community mental health clinics. The majority ( $n = 248$ ) comprised a 'catchment sample' that originated from the only community mental health clinic in a catchment area of  $\sim 150\,000$  people (2). For the current study, we excluded the five subjects with 22q11.2 deletions with genome-wide CNV data published elsewhere (57), two of whom were from the catchment area sample (2). None of the remaining 454 subjects [317 (69.8%) male; mean age 44.8 (SD 12.1) years] had previously published genome-wide CNV data, though four had previously published abnormal karyotype results (2,58) and two with typical 1q21.1 gains were recently reported by us in a case series (46). After description of the study to the subjects, a written informed consent was obtained. The study was approved by local hospital and university institutional review boards.

All subjects underwent direct clinical screening assessments for potential syndromic features using a standardized protocol that included a review of available lifetime medical records and assessment of physical features (2,17,18). Using the criteria previously validated for identifying adults with 22q11.2 deletion syndrome [two or more of: global dysmorphic facial features, history of learning difficulties, abnormal (hypernasal) voice (17)], syndromic [ $n = 78$  (17.2%)] and non-syndromic subgroups were designated (18). No individual had a pediatric diagnosis of moderate or severe intellectual disability, multiple major congenital anomalies or autism (2). All phenotyping was done blind to genotype. The age at onset was considered the age at first treatment for psychosis [mean 22.9 (SD 6.8) years], and a positive family history defined as having one or more first degree relatives with a psychotic disorder [ $n = 70$  (16.8%) of 416].

### Control sample for formal analyses

To optimize our analyses, we used an independent Canadian control sample from the OPGP genetic epidemiologic project comprising 416 unrelated adults [208 (50.0%) male; mean age 45.0 (SD 12.1) years] of European ancestry (18). The OPGP controls are independent of the 2357 population-based CNV adjudication controls described below. The OPGP was conceived as an opportunity to establish a collection of fully consented, Ontario-based population control DNA samples, as a resource for researchers in Ontario and elsewhere. The OPGP was constructed in two phases. The first phase involved collaborations with the Ontario Familial Breast Cancer Registry (OFBCR) (59) and the Ontario Familial Colorectal Cancer Registry (OFCCR) (60). Population controls from the OFBCR and OFCCR that had completed the study questionnaires and provided a blood sample were eligible to participate as controls in the OPGP. The second phase involved collaboration with the Institute of Social Research at York University (Toronto, ON, Canada). A database of information on individuals from the population of Ontario ( $n = 14\,000$ ) was obtained, based on a random sample of households across Ontario identified initially on the basis of telephone directory listings. These households were contacted by means of a mailed letter and follow-up telephone calls were made by a trained interviewer who, after establishing eligibility within the household, randomly chose one person as the eligible control. Individuals were eligible if they were 20–79 years of age and resided in Ontario. In both the

phases, subjects were re-consented for OPGP, and surveyed for comprehensive demographic and health information. Blood samples were collected and sent via courier to The Centre for Applied Genomics (TCAG) Biobanking Facility for transformation, DNA preparation and distribution. The total OPGP collection consists of samples from 1134 males (42%) and 1556 females (58%), ranging in age from 20 to 82 years. This study uses data from 416 of these OPGP samples of European ancestry that were genotyped on the Affymetrix 6.0 platform. To minimize laboratory-related artefacts (54,61), identical experimental and analytical protocols for array genotyping and CNV analyses were followed at the same facility for the OPGP control and schizophrenia case samples (see below).

### Genotyping and CNV determination

High-quality genomic DNA was genotyped using the Affymetrix® Genome-Wide Human SNP Array 6.0 at TCAG in Toronto. Arrays meeting Affymetrix's recommended quality control guideline of contrast QC >0.4 were used for further analysis as outlined below and in Supplementary Material, Fig. S1.

To accurately estimate ancestry, genotypes of the schizophrenia cases from 1120 genome-wide unlinked SNPs were clustered by the program STRUCTURE (62) together with genotypes from 270 HapMap samples, which were used as references of known ancestry during clustering. Ancestries were assigned with a threshold of coefficient of ancestry >0.9: 420 (92.5%) of European, 29 (6.4%) of Admixed, 4 (0.9%) of East Asian and 1 (0.2%) of African ancestry. Recent findings of unexpected relatedness in widely used databases (63) prompted a systematic analysis of genetic relatedness to ensure that all cases and controls were unrelated to each other. Pair-wise identity by descent (IBD) was calculated for every pair of cases from the genotypes of 934 968 SNP probes on the Affymetrix 6.0 platform, using the PLINK toolset (64). Those pairs of samples with PI\_HAT values [defined as  $P(\text{IBD} = 2) + 0.5 \times P(\text{IBD} = 1)$ ] >0.1, corresponding to first cousins or closer, were further investigated for relatedness and only one sample from each pair was used in the final dataset. The results using the initial samples tested showed that (i) four subjects in the schizophrenia group were related to other probands (subsequently confirmed by further pedigree data), and (ii) one male OPGP control was related to another control subject. Data for these five individuals were therefore excluded as a preliminary step in compiling the samples of 454 cases and 416 controls. Notably, the schizophrenia case and OPGP control samples of European ancestry showed a similar pair-wise within-sample background relatedness (median PI\_HAT values of 0.032 and 0.046, respectively, within pre-established ranges of 0.031–0.099).

Genome-wide CNVs were detected using a multiple-algorithm approach to maximize sensitivity and specificity of CNV calling, as described previously (19). Briefly, for each subject we defined 'stringent' CNV calls as those detected by at least two of three CNV calling algorithms: Birdsuite (65), iPattern (54), and Affymetrix Genotyping Console, and spanning at least 10 kb in length and five or more consecutive array probes. All subsequent analyses focused on these stringent CNVs, which in our past experience and in this study (see below) have very high positive validation rates by independent methods, e.g., quantitative PCR (qPCR) (5,6,18,19). Each CNV identified in

the schizophrenia case and OPGP control samples was then adjudicated for rarity by comparison to those identified, using an identical microarray platform and CNV analysis strategy, in two large population-based control cohorts comprising 2357 individuals of European ancestry from Ontario and Germany (18,19). We used a conservative definition of 'rare' CNVs: those present in <0.1% of the population controls (18). All rare CNVs identified in schizophrenia cases and OPGP controls are presented in Supplementary Material, Table S4.

### CNV prioritization, assessment of pathogenicity and prevalence calculations

We prioritized two main groups of rare (present in <0.1% of population-based controls) CNVs in the schizophrenia cases for qualitative analysis: (i) large CNVs (i.e., those defined as >500 kb in size) and (ii) CNVs <500 kb in size that were 'very rare' (i.e., not present in the population-based control cohorts, using a 50% reciprocal overlap criterion) (19), and that overlapped the same CNS-related gene(s) in two or more unrelated schizophrenia cases (18). We also identified other very rare CNVs that overlapped promising candidate genes for schizophrenia (as determined by known biological function and/or previous reports), but that were found only in a single schizophrenia case.

Each large (>500 kb) rare CNV in schizophrenia case and OPGP control individuals was assessed independently by two experienced clinical cytogenetic laboratory directors blind to case–control status. According to the ACMG guidelines for interpretation of postnatal constitutional CNVs (21), each CNV event was assigned to one of the five main categories: 'Pathogenic', 'Uncertain clinical significance; likely pathogenic', 'Uncertain clinical significance (no sub-classification)', 'Uncertain clinical significance; likely benign' and 'Benign'. Disagreements between the two laboratory directors in initial interpretations were subsequently resolved through discussion to arrive at a consensus determination (while still blind to case-control status). In the text, variants deemed to be 'Pathogenic' or of 'Uncertain clinical significance; likely pathogenic' were collectively termed 'clinically significant', and those of 'Uncertain clinical significance (no subclassification)' were termed 'variants of unknown significance' (VUS); the remainder were termed 'benign'.

We used our schizophrenia community catchment sample (2) to estimate the minimum prevalence of large, rare, clinically significant structural variants that would be detectable on a clinical microarray. We estimated that there were 370 unrelated adults with schizophrenia in this area, including the 248 consecutive subjects for whom genome-wide CNV data were available (67.0%), as well as individuals who refused or were unable to consent to participation in the study, and others known to but not attending the clinic or hospital system; all non-participants were assumed to be unrelated to probands in this study or to each other.

### Experimental validation of CNVs

Confirmatory studies of possible schizophrenia-associated CNVs used Stratagene SYBR® Green based qPCR. Each qPCR assay was performed simultaneously in triplicate for

both the test region and a control region of known copy number 2 on chromosome 7 (26). The ratio of the average value for the test region to that for the control region had to be  $>1.3$  or  $<0.7$  in order for the CNV to be confirmed as a duplication or deletion, respectively. In addition, the standard error of the ratio had to be  $<1.0$  on the same scale in order for the assay to be considered reliable. We validated 58 (100.0%) of 58 CNVs tested, i.e., all 58 assays met both of these criteria for CNV confirmation. The validated CNVs spanned a range of sizes, and included 17 that were  $<50$  kb. Post-study molecular genetic results from clinical laboratories confirmed a further 11 CNVs, and also the parental origin of the 15q11-q13 duplications. All clinically significant variants were confirmed in a CLIA-approved laboratory and returned to patients and their families.

### CNV burden analyses

For the CNV burden analyses, we compared the proportion of subjects of European ancestry with one or more rare CNVs in the 420 [ $n = 291$  (69.3%) male] unrelated schizophrenia cases with that in the 416 [ $n = 208$  (50.0%) male] unrelated OPGP controls (18). We used only autosomal CNVs because of the significant sex differences between these groups ( $\chi^2 = 32.31$ ,  $df = 1$ ,  $P < 0.001$ ). As an additional conservative step, variants that might reasonably be discovered without use of a genome-wide microarray, i.e., rare CNVs  $>6.5$  Mb in size potentially detectable by routine G-band analysis (cases of European ancestry:  $n = 6$ , controls:  $n = 0$ ; Table 1) and 22q11.2 deletions (cases of European ancestry:  $n = 4$ , controls:  $n = 0$ ), were not used in these formal burden calculations. Statistical analyses were performed using SAS® software. Comparisons between schizophrenia cases and OPGP controls, and between syndromic and non-syndromic schizophrenia cases, were performed using standard two-sided statistical tests. ORs and 95% CIs were calculated using standard methods. A  $P$ -value of  $<0.05$  was considered to be significant. As described above, identical experimental and analytical protocols for array genotyping and CNV analyses were followed at the same facility for the schizophrenia case and OPGP control samples, and each CNV was then adjudicated for rarity by comparison to those identified, using an identical microarray platform and CNV analysis strategy, in two large population-based control cohorts distinct from the OPGP controls. Thus, the rate of Type 1 and 2 errors would be expected to be the same in schizophrenia cases and OPGP controls, with no resulting impact on the relative inter-group CNV burden.

### Functional gene enrichment mapping

To identify biologically meaningful groups of genes (i.e., those sharing a common function) affected by CNVs using the same samples as for the main burden analyses, we used a gene enrichment mapping method similar to that previously published for autism (6,37) and congenital cardiac disease (18). Briefly, we first compiled comprehensive collections of gene sets (Supplementary Material, Table S5). Then, for each functional gene set, we tested whether the prevalence of subjects with very rare CNVs was higher in schizophrenia cases than in OPGP controls using a one-tailed Fisher's exact test (FET) (6,18). Nominal  $P$ -values were corrected for multiple testing using a case-control class permutation procedure to estimate an empirical

false discovery rate (FDR) (6,18). Case-control labels were randomly permuted 5000 times and, for each  $P$ -value, the empirical FDR was computed as the average number of gene sets with equal or smaller  $P$ -value over permutations. We selected 10% as the empirical FDR significance threshold for final results. Additional details are presented in Supplementary Material, Table S6. As above, the rate of Type 1 and 2 errors would be expected to be the same in schizophrenia cases and OPGP controls, with no resulting impact on the gene enrichment mapping. Finally, we highlighted the clusters of the gene sets that were significantly associated with schizophrenia where for at least half of the gene sets there was also significant over-representation of known ASD genes (66). Over-representation was tested using FET and Benjamini-Hochberg FDR for multiple test correction, with a significance threshold of 5% FDR.

### ACKNOWLEDGEMENTS

The authors thank the patients and their families for their participation, colleagues for referring patients, research assistants, staff at the Saint John Community Mental Health Services, the Saint John Regional Hospital, Hillsborough Hospital and Hamilton clinic, staff at The Centre for Applied Genomics (TCAG), fellows and students who assisted in the collection and analysis of data for the study, F. Fu and S. Bekeschus for helping with the preparation of the figures, and G. Wong, M. Torsan, and W. Warnica for helping with the preparation of the manuscript. We thank A. Fiebig, A. Franke, and S. Schreiber at POPGEN (University of Kiel, Germany), and A. Stewart, R. McPherson and R. Roberts of the University of Ottawa Heart Institute (University of Ottawa, Canada) for providing control data. In assessing pathogenicity, the clinical cytogenetic laboratory directors made use of the International Standards for Cytogenomic Arrays (ISCA) Consortium database ([www.iscaconsortium.org/](http://www.iscaconsortium.org/)), which generates information using NCBI's database of genomic structural variation (dbVar, [www.ncbi.nlm.nih.gov/dbvar/](http://www.ncbi.nlm.nih.gov/dbvar/)) with samples and associated phenotype data provided by ISCA Consortium member laboratories.

Extended control data for 2q13 variant comparisons were obtained, along with permission for use, from the database of Genotypes and Phenotypes (dbGaP) found at <http://www.ncbi.nlm.nih.gov/gap> through accession numbers phs000143.v1.p1 (Starr County Health Studies' Genetics of Diabetes Study), phs000091.v2.p1 (GENEVA NHS/HPFS Diabetes study) and phs000169.v1.p1 (Whole Genome Association Study of Visceral Adiposity in the HABC Study). The Starr County Health Studies Genetics of Diabetes Study was supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) and the NIDDK Central Repositories. Support for the GWAS of Gene and Environment Initiatives in Type 2 Diabetes was provided through the NIH Genes, Environment and Health Initiative [GEI] (U01HG004399). The human subjects participating in the GWAS derive from The Nurses' Health Study and Health Professionals' Follow-up Study and these studies are supported by National Institutes of Health grants CA87969, CA55075, and DK58845. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the Gene Environment Association Studies, GENEVA Coordinating Center (U01

HG004446) and the National Center for Biotechnology Information. Support for genotyping, which was performed at the Broad Institute of MIT and Harvard, was provided by the NIH GEI (U01HG004424). Support for the “CIDR Visceral Adiposity Study” was provided through the Division of Aging Biology and the Division of Geriatrics and Clinical Gerontology, National Institute on Aging. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by Health ABC Study (HABC) Investigators.

**Conflict of Interest statement:** S.W.S. is on the Scientific Advisory Board of Population Diagnostics, a US company that could use data from this study. The other authors declare that they have no competing interests.

## FUNDING

This work was supported by the Canadian Institutes of Health Research (CIHR) (MOP-89066 to A.S.B., MOP-111238 to A.S.B.). A.S.B. holds the Canada Research Chair in Schizophrenia Genetics and Genomic Disorders and the Dalglish Chair in 22q11.2 Deletion Syndrome. A.C.L. holds a NeuroDevNet doctoral fellowship. G.C. holds a CIHR Vanier Canada Graduate Scholarship. S.W.S. is supported by grants from the University of Toronto McLaughlin Centre, NeuroDevNet, Genome Canada and the Ontario Genomics Institute, the CIHR, the Canadian Institute for Advanced Research, the Canada Foundation for Innovation, the government of Ontario, Autism Speaks, and The Hospital for Sick Children Foundation. S.W.S. holds the GlaxoSmithKline-CIHR Chair in Genome Sciences at the University of Toronto and The Hospital for Sick Children.

## REFERENCES

- Costain, G. and Bassett, A.S. (2012) Clinical applications of schizophrenia genetics: genetic diagnosis, risk, and counseling in the molecular era. *Appl. Clin. Genet.*, **5**, 1–18.
- Bassett, A.S., Costain, G., Fung, W.L.A., Russell, K.J., Pierce, L., Kapadia, R., Carter, R.F., Chow, E.W. and Forsythe, P.J. (2010) Clinically detectable copy number variations in a Canadian catchment population of schizophrenia. *J. Psychiatr. Res.*, **44**, 1005–1009.
- Cook, E.H. Jr. and Scherer, S.W. (2008) Copy-number variations associated with neuropsychiatric conditions. *Nature*, **455**, 919–923.
- Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V. *et al.* (2011) A copy number variation morbidity map of developmental delay. *Nat. Genet.*, **43**, 838–846.
- Marshall, C.R., Noor, A., Vincent, J.B., Lionel, A.C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y. *et al.* (2008) Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.*, **82**, 477–488.
- Pinto, D., Pagnamenta, A.T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S. *et al.* (2010) Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*, **466**, 368–372.
- Bassett, A.S., Scherer, S.W. and Brzustowicz, L.M. (2010) Copy number variations in schizophrenia: critical review and new perspectives on concepts of genetics and disease. *Am. J. Psychiatry*, **167**, 899–914.
- Kirov, G., Pocklington, A.J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., Moran, J., Chambert, K., Toncheva, D., Georgieva, L. *et al.* (2012) De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol. Psychiatry*, **17**, 142–153.
- International Schizophrenia Consortium. (2008) Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*, **455**, 237–241.
- Levinson, D.F., Duan, J., Oh, S., Wang, K., Sanders, A.R., Shi, J., Zhang, N., Mowry, B.J., Olincy, A., Amin, F. *et al.* (2011) Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. *Am. J. Psychiatry*, **168**, 302–316.
- Melhem, N., Middleton, F., McFadden, K., Klei, L., Faraone, S.V., Vinogradov, S., Tiobech, J., Yano, V., Kuartei, S., Roeder, K. *et al.* (2011) Copy number variants for schizophrenia and related psychotic disorders in Oceanic Palau: risk and transmission in extended pedigrees. *Biol. Psychiatry*, **70**, 1115–1121.
- McClellan, J.M., Susser, E. and King, M.C. (2007) Schizophrenia: a common disease caused by multiple rare alleles. *Br. J. Psychiatry*, **190**, 194–199.
- Sahoo, T., Theisen, A., Rosenfeld, J.A., Lamb, A.N., Ravnan, J.B., Schultz, R.A., Torchia, B.S., Neill, N., Casci, I., Bejjani, B.A. *et al.* (2011) Copy number variants of schizophrenia susceptibility loci are associated with a spectrum of speech and developmental delays and behavior problems. *Genet. Med.*, **13**, 868–880.
- Lionel, A.C., Vaags, A.K., Sato, D., Gazzellone, M.J., Mitchell, E.B., Chen, H.Y., Costain, G., Walker, S., Egger, G., Thiruvahindrapuram, B. *et al.* (2013) Rare exonic deletions implicate the synaptic organizer Gephyrin (GPHN) in risk for autism, schizophrenia and seizures. *Hum. Mol. Genet.*, **22**, 2055–2066.
- Rujescu, D., Ingason, A., Cichon, S., Pietilainen, O.P., Barnes, M.R., Touloupoulou, T., Picchioni, M., Vassos, E., Ettinger, U., Bramon, E. *et al.* (2009) Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum. Mol. Genet.*, **18**, 988–996.
- Raychaudhuri, S., Plenge, R.M., Rossin, E.J., Ng, A.C., Purcell, S.M., Sklar, P., Scolnick, E.M., Xavier, R.J., Altshuler, D. and Daly, M.J. (2009) Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.*, **5**, e1000534.
- Fung, W.L., Chow, E.W., Webb, G.D., Gatzoulis, M.A. and Bassett, A.S. (2008) Extracardiac features predicting 22q11.2 deletion syndrome in adult congenital heart disease. *Int. J. Cardiol.*, **131**, 51–58.
- Silversides, C.K., Lionel, A.C., Costain, G., Merico, D., Migita, O., Liu, B., Yuen, Y., Rickaby, J., Thiruvahindrapuram, B., Marshall, C.R. *et al.* (2012) Rare copy number variations in adults with tetralogy of Fallot implicate novel risk gene pathways. *PLoS Genet.*, **8**, e1002843.
- Lionel, A.C., Crosbie, J., Barbosa, N., Goodale, T., Thiruvahindrapuram, B., Rickaby, J., Gazzellone, M., Carson, A.R., Howe, J.L., Wang, Z. *et al.* (2011) Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. *Sci. Transl. Med.*, **3**, 95 ra75.
- Raychaudhuri, S., Korn, J.M., McCarroll, S.A., International Schizophrenia, C., Altshuler, D., Sklar, P., Purcell, S. and Daly, M.J. (2010) Accurately assessing the risk of schizophrenia conferred by rare copy-number variation affecting genes with brain function. *PLoS Genet.*, **6**, e1001097.
- Kearney, H.M., Thorland, E.C., Brown, K.K., Quintero-Rivera, F. and South, S.T. and Working Group of the American College of Medical Genetics Laboratory Quality Assurance, C. (2011) American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet. Med.*, **13**, 680–685.
- Yu, H.E., Hawash, K., Picker, J., Stoler, J., Urion, D., Wu, B.L. and Shen, Y. (2012) A recurrent 1.71 Mb genomic imbalance at 2q13 increases the risk of developmental delay and dysmorphism. *Clin. Genet.*, **81**, 257–264.
- Stankiewicz, P., Kulkarni, S., Dharmadhikari, A.V., Sampath, S., Bhatt, S.S., Shaikh, T.H., Xia, Z., Pursley, A.N., Cooper, M.L., Shinawi, M. *et al.* (2012) Recurrent deletions and reciprocal duplications of 10q11.21q11.23 including CHAT and SLC18A3 are likely mediated by complex low-copy repeats. *Hum. Mutat.*, **33**, 165–179.
- Molin, A.M., Andrieux, J., Koolen, D.A., Malan, V., Carella, M., Colleaux, L., Cormier-Daire, V., David, A., de Leeuw, N., Delobel, B. *et al.* (2012) A novel microdeletion syndrome at 3q13.31 characterised by developmental delay, postnatal overgrowth, hypoplastic male genitals, and characteristic facial features. *J. Med. Genet.*, **49**, 104–109.
- Zhang, X., Snijders, A., Seagraves, R., Zhang, X., Niebuhr, A., Albertson, D., Yang, H., Gray, J., Niebuhr, E., Bolund, L. *et al.* (2005) High-resolution mapping of genotype–phenotype relationships in cri du chat syndrome using array comparative genomic hybridization. *Am. J. Hum. Genet.*, **76**, 312–326.



26. Feuk, L., Kalervo, A., Lipsanen-Nyman, M., Skaug, J., Nakabayashi, K., Finucane, B., Hartung, D., Innes, M., Kerem, B., Nowaczyk, M.J. *et al.* (2006) Absence of a paternally inherited FOXP2 gene in developmental verbal dyspraxia. *Am. J. Hum. Genet.*, **79**, 965–972.
27. Paridaen, J.T., Danesin, C., Elas, A.T., van de Water, S., Houart, C. and Zivkovic, D. (2009) Apc1 is required for maintenance of local brain organizers and dorsal midbrain survival. *Dev. Biol.*, **331**, 101–112.
28. Youle, R.J. and Strasser, A. (2008) The BCL-2 protein family: opposing activities that mediate cell death. *Nat. Rev. Mol. Cell Biol.*, **9**, 47–59.
29. Binder, M.D. and Kilpatrick, T.J. (2009) TAM receptor signalling and demyelination. *Neuro-Signals*, **17**, 277–287.
30. Bassett, A.S. (2011) Parental origin, DNA structure, and the schizophrenia spectrum. *Am. J. Psychiatry*, **168**, 350–353.
31. Vrijenhoek, T., Buizer-Voskamp, J.E., van der Stelt, I. and Strengman, E., Genetic Risk and Outcome in Psychosis (GROUP) Consortium, Sabatti, C., Geurts van Kessel, A., Brunner, H.G., Ophoff, R.A. and Veltman, J.A. (2008) Recurrent CNVs disrupt three candidate genes in schizophrenia patients. *Am. J. Hum. Genet.*, **83**, 504–510.
32. Xu, B., Roos, J.L., Levy, S., van Rensburg, E.J., Gogos, J.A. and Karayiorgou, M. (2008) Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat. Genet.*, **40**, 880–885.
33. Steinberg, S., Mors, O., Borglum, A.D., Gustafsson, O., Werge, T., Mortensen, P.B., Andreassen, O.A., Sigurdsson, E., Thorgeirsson, T.E., Botcher, Y. *et al.* (2011) Expanding the range of ZNF804A variants conferring risk of psychosis. *Mol. Psychiatry*, **16**, 59–66.
34. Need, A.C., McEvoy, J.P., Gennarelli, M., Heinzen, E.L., Ge, D., Maia, J.M., Shianna, K.V., He, M., Cirulli, E.T., Gumbs, C.E. *et al.* (2012) Exome sequencing followed by large-scale genotyping suggests a limited role for moderately rare risk factors of strong effect in schizophrenia. *Am. J. Hum. Genet.*, **91**, 303–312.
35. Xu, B., Ionita-Laza, I., Roos, J.L., Boone, B., Woodruff, S., Sun, Y., Levy, S., Gogos, J.A. and Karayiorgou, M. (2012) De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat. Genet.*, **44**, 1365–1369.
36. Walsh, T., McClellan, J.M., McCarthy, S.E., Addington, A.M., Pierce, S.B., Cooper, G.M., Nord, A.S., Kusenda, M., Malhotra, D., Bhandari, A. *et al.* (2008) Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*, **320**, 539–543.
37. Merico, D., Isserlin, R., Stueker, O., Emili, A. and Bader, G.D. (2010) Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One*, **5**, e13984.
38. Miller, D.T., Adam, M.P., Aradhya, S., Biasecker, L.G., Brothman, A.R., Carter, N.P., Church, D.M., Crolla, J.A., Eichler, E.E., Epstein, C.J. *et al.* (2010) Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am. J. Hum. Genet.*, **86**, 749–764.
39. Shen, Y., Dies, K.A., Holm, I.A., Bridgemohan, C., Sobehi, M.M., Caronna, E.B., Miller, K.J., Frazier, J.A., Silverstein, I., Picker, J. *et al.* (2010) Clinical genetic testing for patients with autism spectrum disorders. *Pediatrics*, **125**, e727–e735.
40. Costain, G., Chow, E.W., Ray, P.N. and Bassett, A.S. (2012) Caregiver and adult patient perspectives on the importance of a diagnosis of 22q11.2 deletion syndrome. *J. Intellect. Disabil. Res.*, **56**, 641–651.
41. Trottier, M., Roberts, W., Drmic, I., Scherer, S.W., Weksberg, R., Cyttrynbaum, C., Chitayat, D., Shuman, C. and Miller, F.A. (2013) Parents' perspectives on participating in genetic research in autism. *J. Autism Dev. Disord.*, **43**, 556–568.
42. Riggs, E., Wain, K., Riethmaier, D., Smith-Packard, B., Faucett, W., Hoppman, N., Thorland, E., Patel, V. and Miller, D. (2013) Chromosomal microarray impacts clinical management. *Clin. Genet.* doi:10.1111/cge.12107.
43. Scherer, S.W. and Dawson, G. (2011) Risk factors for autism: translating genomic discoveries into diagnostics. *Hum. Genet.*, **130**, 123–148.
44. Miller, F.A., Hayeems, R.Z. and Bytautas, J.P. (2010) What is a meaningful result? Disclosing the results of genomic research in autism to research participants. *Eur. J. Hum. Genet.*, **18**, 867–871.
45. Bassett, A.S., McDonald-McGinn, D.M., Devriendt, K., Digilio, M.C., Goldenberg, P., Habel, A., Marino, B., Oskarsdottir, S., Philip, N., Sullivan, K. *et al.* (2011) Practical guidelines for managing patients with 22q11.2 deletion syndrome. *J. Pediatr.*, **159**, 332–339.
46. Dolcetti, A., Silversides, C.K., Marshall, C.R., Lionel, A.C., Stavropoulos, D.J., Scherer, S.W. and Bassett, A.S. (2013) 1q21.1 Microduplication expression in adults. *Genet. Med.*, **15**, 282–289.
47. Costain, G., Chow, E.W.C., Silversides, C.K. and Bassett, A.S. (2011) Sex differences in reproductive fitness contribute to preferential maternal transmission of 22q11.2 deletions. *J. Med. Genet.*, **48**, 819–824.
48. Costain, G., Esplen, M.J., Toner, B., Hodgkinson, K.A. and Bassett, A.S. (2012) Evaluating Genetic Counseling for Family Members of Individuals With Schizophrenia in the Molecular Age. *Schizophr. Bull.* doi:10.1093/schbul/sbs124.
49. Costain, G., Esplen, M.J., Toner, B., Scherer, S.W., Meschino, W.S., Hodgkinson, K.A. and Bassett, A.S. (2012) Evaluating Genetic Counseling for Individuals With Schizophrenia in the Molecular Age. *Schizophr. Bull.* doi:10.1093/schbul/sbs138.
50. Costain, G. and Bassett, A.S. (2012) The ever-evolving concept of clinical significance and the potential for sins of omission in genetic research. *Am. J. Bioeth.*, **12**, 22–24.
51. Vorstman, J.A.S., Breetvelt, E.J., Thode, K.I., Chow, E.W.C. and Bassett, A.S. (2013) Expression of autism spectrum and schizophrenia in patients with a 22q11.2 deletion. *Schizophr. Res.*, **143**, 55–59.
52. Johnson, M.B., Kawasawa, Y.I., Mason, C.E., Krsnik, Z., Coppola, G., Bogdanovic, D., Geschwind, D.H., Mane, S.M., State, M.W. and Sestan, N. (2009) Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron*, **62**, 494–509.
53. Insel, T.R. (2010) Rethinking schizophrenia. *Nature*, **468**, 187–193.
54. Pinto, D., Darvishi, K., Shi, X., Rajan, D., Rigler, D., Fitzgerald, T., Lionel, A.C., Thiruvahindrapuram, B., Macdonald, J.R., Mills, R. *et al.* (2011) Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nat. Biotechnol.*, **29**, 512–520.
55. Bromet, E.J., Kotov, R., Fochtmann, L.J., Carlson, G.A., Tanenberg-Karant, M., Ruggero, C. and Chang, S.W. (2011) Diagnostic shifts during the decade following first admission for psychosis. *Am. J. Psychiatry*, **168**, 1186–1194.
56. Bassett, A.S., Chow, E.W., Husted, J., Hodgkinson, K.A., Oechslin, E., Harris, L. and Silversides, C. (2009) Premature death in adults with 22q11.2 deletion syndrome. *J. Med. Genet.*, **46**, 324–330.
57. Bassett, A.S., Marshall, C.R., Lionel, A.C., Chow, E.W. and Scherer, S.W. (2008) Copy number variations and risk for schizophrenia in 22q11.2 deletion syndrome. *Hum. Mol. Genet.*, **17**, 4045–4053.
58. Caluseriu, O., Mirza, G., Ragoussis, J., Chow, E.W., MacCrimmon, D. and Bassett, A.S. (2006) Schizophrenia in an adult with 6p25 deletion syndrome. *Am. J. Med. Genet. A*, **140**, 1208–1213.
59. Andrusis, I.L., Boyd, N.F. and Sutherland, H. (1997) New Ontario familial breast cancer registry to facilitate genetic and epidemiologic studies. *Can. Fam. Physician*, **43**, 949–950.
60. Cotterchio, M., McKeown-Eyssen, G., Sutherland, H., Buchan, G., Aronson, M., Easson, A.M., Macey, J., Holowaty, E. and Gallinger, S. (2000) Ontario familial colon cancer registry: methods and first-year response rates. *Chronic Dis. Can.*, **21**, 81–86.
61. Scherer, S.W., Lee, C., Birney, E., Altshuler, D.M., Eichler, E.E., Carter, N.P., Hurler, M.E. and Feuk, L. (2007) Challenges and standards in integrating surveys of structural variation. *Nat. Genet.*, **39**, S7–15.
62. Pritchard, J.K. (2001) Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet.*, **69**, 124–137.
63. Pemberton, T.J., Wang, C., Li, J.Z. and Rosenberg, N.A. (2010) Inference of unexpected genetic relatedness among individuals in HapMap Phase III. *Am. J. Hum. Genet.*, **87**, 457–464.
64. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
65. Korn, J.M., Kuruvilla, F.G., McCarroll, S.A., Wysoker, A., Nemesh, J., Cawley, S., Hubbell, E., Veitch, J., Collins, P.J., Darvishi, K. *et al.* (2008) Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat. Genet.*, **40**, 1253–1260.
66. Betancur, C. (2011) Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res.*, **1380**, 42–77.
67. Kirov, G., Grozeva, D., Norton, N., Ivanov, D., Mantripragada, K.K., Holmans, P., Consortium, I.S., Consortium, W.T.C.C., Craddock, N., Owen, M.J. *et al.* (2009) Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Hum. Mol. Genet.*, **18**, 1497–1503.
68. Stewart, L.R., Hall, A.L., Kang, S.H., Shaw, C.A. and Beaudet, A.L. (2011) High frequency of known copy number abnormalities and maternal duplication 15q11-q13 in patients with combined schizophrenia and epilepsy. *BMC Med. Genet.*, **12**, 154.