



University Medical Center Groningen

University of Groningen

Genome-Wide Analysis Shows Increased Frequency of Copy Number Variation Deletions in Dutch Schizophrenia Patients

Buizer-Voskamp, Jacobine E.; Muntjewerff, Jan-Willem; Strengman, Eric; Sabatti, Chiara; Stefansson, Hreinn; Vorstman, Jacob A. S.; Ophoff, Roel A.; Genetic Risk and Outcome of Psychosis (G.R.O.U.P.); GROUP investigators, [No Value]

Published in:
Biological Psychiatry

DOI:
[10.1016/j.biopsych.2011.02.015](https://doi.org/10.1016/j.biopsych.2011.02.015)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2011

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Buizer-Voskamp, J. E., Muntjewerff, J-W., Strengman, E., Sabatti, C., Stefansson, H., Vorstman, J. A. S., ... GROUP investigators, N. V. (2011). Genome-Wide Analysis Shows Increased Frequency of Copy Number Variation Deletions in Dutch Schizophrenia Patients. *Biological Psychiatry*, 70(7), 655-662. <https://doi.org/10.1016/j.biopsych.2011.02.015>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Genome-Wide Analysis Shows Increased Frequency of Copy Number Variation Deletions in Dutch Schizophrenia Patients

Jacobine E. Buizer-Voskamp, Jan-Willem Muntjewerff, Genetic Risk and Outcome in Psychosis (GROUP) Consortium, Eric Strengman, Chiara Sabatti, Hreinn Stefansson, Jacob A.S. Vorstman, and Roel A. Ophoff

Background: Since 2008, multiple studies have reported on copy number variations (CNVs) in schizophrenia. However, many regions are unique events with minimal overlap between studies. This makes it difficult to gain a comprehensive overview of all CNVs involved in the etiology of schizophrenia. We performed a systematic CNV study on the basis of a homogeneous genome-wide dataset aiming at all CNVs ≥ 50 kilobase pair. We complemented this analysis with a review of cytogenetic and chromosomal abnormalities for schizophrenia reported in the literature with the purpose of combining classical genetic findings and our current understanding of genomic variation.

Methods: We investigated 834 Dutch schizophrenia patients and 672 Dutch control subjects. The CNVs were included if they were detected by QuantiSNP (<http://www.well.ox.ac.uk/QuantiSNP/>) as well as PennCNV (<http://www.neurogenome.org/cnv/penncnv/>) and contain known protein coding genes. The integrated identification of CNV regions and cytogenetic loci indicates regions of interest (cytogenetic regions of interest [CROIs]).

Results: In total, 2437 CNVs were identified with an average number of 2.1 CNVs/subject for both cases and control subjects. We observed significantly more deletions but not duplications in schizophrenia cases versus control subjects. The CNVs identified coincide with loci previously reported in the literature, confirming well-established schizophrenia CROIs 1q42 and 22q11.2 as well as indicating a potentially novel CROI on chromosome 5q35.1.

Conclusions: Chromosomal deletions are more prevalent in schizophrenia patients than in healthy subjects and therefore confer a risk factor for pathogenicity. The combination of our CNV data with previously reported cytogenetic abnormalities in schizophrenia provides an overview of potentially interesting regions for positional candidate genes.

Key Words: Candidate gene, copy number variation, cytogenetic abnormality, deletion, duplication, schizophrenia

Schizophrenia is a debilitating psychiatric disease showing a heterogeneous clinical phenotype and a lifetime risk of .46%–1% (1,2). It is characterized by psychotic symptoms, including delusions and hallucinations, reduced interest and drive, altered emotional reactivity, and disorganized behavior (3). Its predisposition is influenced by a complex interaction of genetic and environmental factors (4,5). Family, twin, and adoption studies have provided evidence for a substantial genetic contribution to the phenotypic variance of schizophrenia (6). It has been estimated that the heritability of developing schizophrenia is up to 80% (7), but the

pattern of inheritance and related pathogenic pathways remain elusive.

Early genetic studies in schizophrenia were based on linkage studies in pedigrees and association testing of candidate genes (8). Linkage and association studies have been complemented by identification of chromosomal abnormalities in patients as well as the more recently genome-wide association studies (GWAS) with single nucleotide polymorphisms (SNPs). These genomic microarrays also allowed for the systematic genome-wide analysis of submicroscopic cytogenetic variation (i.e., genomic copy number variation [CNV]), which includes genomic deletions and duplications of more than 1 kilobase pair (kb) in size. Many recent studies support a polygenic basis of schizophrenia. Important in the understanding of the genetic architecture of schizophrenia is the “common disease—common variant” versus the “common disease—rare variant” hypothesis. The first refers to the possibility that common alleles with small-to-moderate disease risks might have an additive or multiplicative effect on schizophrenia. Recent large-scale association studies have supported this hypothesis for schizophrenia (9). However, common variants are unlikely to explain the total heritability of schizophrenia. The latter hypothesis suggests that multiple rare variants with relatively large effect play a role in the etiology of schizophrenia. It is likely that both common and rare alleles lead to the genetic heterogeneity of schizophrenia. Current CNV studies have mainly focused on extremely rare and de novo variants. However, advances in CNV discovery techniques as well as the increasing number of data releases from SNP association studies might allow for the detection of smaller, more frequent variants (8).

Since 2008, several large-scale whole-genome schizophrenia association studies have reported on the detection of CNVs (10–15).

From the Rudolf Magnus Institute of Neuroscience (JEB-V, JASV, RAO); Department of Psychiatry; and Department of Medical Genetics (JEB-V, ES, RAO), University Medical Center Utrecht, Utrecht, the Netherlands; Department of Psychiatry (J-WM), University Medical Centre St. Radboud, Nijmegen, the Netherlands; Department of Health Research and Policy (CS), Stanford University School of Medicine, Stanford; and Population genomics (HS), deCODE genetics, Reykjavik, Iceland; Center for Neurobehavioral Genetics (RAO), University of California, Los Angeles, California. The GROUP Consortium members are listed fully in the Acknowledgments section.

Authors JEB-V and J-WM contributed equally to this work.

Address correspondence to Roel A. Ophoff, Ph.D., Center for Neurobehavioral Genetics, University of California, 4357C Gonda, Box 951761, 695 Charles E. Young Drive South, Los Angeles, California 90095-1761; E-mail: ROphoff@mednet.ucla.edu.

Received Jul 9, 2010; revised Feb 9, 2011; accepted Feb 11, 2011.

Over 41% of all CNVs identified overlap with known genes. This suggests that CNVs might play a substantial role in modulating gene expression (16). Two large-scale studies identified three recurrent but rare schizophrenia-associated microdeletions at chromosome 1q211, 15q13.3, and 15q11.2, each containing multiple genes (11,12). Many recurrent CNV loci are flanked by segmental duplications or low copy repeats (LCRs), which suggests that these deletions and duplications could be the result of nonallelic homologous recombination mediated by these LCRs. Large-scale CNV studies suggest that rare variants disrupting genes in neurodevelopmental pathways might significantly contribute to the risk for developing schizophrenia (odds ratios ranging from 2.7 to 14.8) (8,12,14). However, many regions reported in those studies are unique events, with minimal overlap between studies, with the exception of a few recurrent loci (8). Both phenotypic diversity associated with CNVs and heterogeneity in diagnoses and ethnicities could explain the difficulties in replicating disease variants (17).

Few studies provided genome-wide evaluation of CNVs (11,12,18). However, most CNV studies thus far present a targeted and limited number of strong CNV candidates involved in disease. Relatively few investigators have yet reported their complete CNV datasets or released raw data to the scientific community, minimizing the opportunity for comparing data or for meta-analyses (19). However, an overview of large-scale genomic variation will be a crucial resource in correlating genomic variations with experimental findings and clinical outcomes. A good example of a genome-wide overview of all CNVs found in schizophrenia is the ISC (International Schizophrenia Consortium) study (12) that reports all CNVs passing QC as an online datafile (<http://pngu.mgh.harvard.edu/isc/>). A systematic review of cytogenetic and CNV findings would provide researchers in the field of schizophrenia with an overview of potentially interesting regions for positional candidate genes. Good examples from other research fields are, for example, the Database of Genomic Variants (<http://projects.tcag.ca/variation/>) and systematic information on chromosome rearrangements for autism spectrum disorders (20,21).

We performed a two-step study. First, we carried out a systematic genome-wide CNV analysis in a homogeneous group of Dutch schizophrenia cases versus unaffected control subjects and provide data of all copy number variants of ≥ 50 kb, including common and rare variants. We further performed a systematic review of all cytogenetic and chromosomal abnormalities reported in the literature for schizophrenia, with the purpose of presenting a comprehensive overview and combining classical cytogenetic findings and our current understanding of genomic variation (22,23). We hypothesize that (new) risk loci for schizophrenia might be revealed through overlap between genomic regions affected by previously reported cytogenetic abnormalities and those affected by CNVs. In addition, we also investigated CNVs in known candidate gene loci from literature. We hypothesize that those candidate gene loci will be affected by CNVs more often in cases compared with control subjects. Both steps are a first effort to understand the role of genomic chromosomal variation and schizophrenia susceptibility. We believe that these data, which we have made publicly available, are a necessary step toward that goal.

Methods and Materials

Systematic Genome-Wide CNV Analysis

We studied CNVs within a cohort of 834 patients with schizophrenia and 672 unaffected control individuals. Inpatients and outpatients were recruited from a variety of psychiatric hospitals and institutions in the Netherlands, partly coordinated via academic

hospitals in Amsterdam, Groningen, Maastricht, and Utrecht (The Genetic Risk and Outcome of Psychosis [GROUP] project). All patients had been diagnosed for subtypes of schizophrenia according to the DSM-IV-TR. Detailed medical and psychiatric histories were collected, including the Comprehensive Assessment of Symptoms and History, an instrument for assessing diagnosis and psychopathology. The control subjects were volunteers and were all screened for any psychiatric history, the majority via the Comprehensive Assessment of Symptoms and History. Both cases and control subjects were of Dutch descent (with at least three of four grandparents of Dutch ancestry), and they all gave informed consent. The study was approved by the Ethics Committee of the University Medical Center Utrecht and by the appropriate local institutional review boards at all other participating hospitals.

Genomic DNA of all patients and control subjects was hybridized to HumanHap550v3 BeadArray (Illumina, San Diego, California) according to standard protocols. QuantiSNP (<http://www.well.ox.ac.uk/QuantiSNP/>) (24) and PennCNV (<http://www.neurogenome.org/cnv/penncnv/>) (25) were used to identify copy number deletions and duplications. Both QuantiSNP and PennCNV are based on a Hidden Markov Model for kilobase-resolution detection of CNVs from Illumina high-density SNP genotyping data. The PennCNV program is probably the most frequently used program for CNV studies in recent publications. This might in part be due to the user-friendly design of the program and free access to users. Its low false positive rate is a promising aspect. By contrast, QuantiSNP outperformed six other methods in a recent evaluation study of CNV calling algorithms (26).

For QuantiSNP, a quality control step for GC-content was performed. For PennCNV, a gc-model for GC-correction was used, and the following quality control steps were performed: 1) CNVs containing < 10 consecutive SNP markers were excluded, 2) CNVs with a value below .5 for the confidence score (value for quality of the CNV call) divided by the number of SNP markers were excluded, and 3) CNVs with an SNP density below 10 kb were excluded.

The CNVs were included only if they were detected by both PennCNV and QuantiSNP and meeting the aforementioned quality control criteria. Several CNV detection methods are available, but differences in characteristics exist, and every method has its own weaknesses (26). By including only overlapping CNVs, we made an effort to limit the false positive rate of CNV detection, as suggested by Winchester *et al.* (27). The CNVs detected by both algorithms were defined by overlapping start and stop positions with at least one position (start or stop SNP marker) similar or when a smaller CNV completely overlapped a larger CNV. Moreover, we only included CNVs containing known protein coding genes, with fuzzy border criteria of up to 50 kb surrounding the CNV boundaries. Compared with CNV size, gene content might be a more reliable indicator for clinical significance, such that small gene-rich CNVs are more likely to be pathogenic than larger gene-poor CNVs (28). The gene content of each CNV was defined with the UCSC genome browser (<http://genome.ucsc.edu/>). We investigated the reliability of our CNV dataset by: 1) calculating the concordance rate between calls from PennCNV and QuantiSNP, 2) calculating the percentage of overlap between calls from the two datasets when increasing filtering stringency, and 3) validating individual CNVs by multiplex ligation-dependent probe amplification and quantitative polymerase chain reaction (qPCR). Taken together, by focusing our analysis on those CNVs that were consistently called by the two algorithms and contain known protein coding (RefSeq) genes, we generated a reliable dataset for further testing.

We provide all data of CNVs ≥ 50 kb found in both cases and control subjects. Raw data are available for future research and meta-analysis. In addition to the total overview of CNVs, we com-

pared the number of deletions and duplications for different size categories in cases and control subjects.

Reviewing Literature

Studies were identified by searching the electronic National Library of Medicine MEDLINE and PubMed databases with different combinations of the search terms "schizophren*", "genetic*", "cytogenetic*", "copy number", "CNV," and "chromosom*" for reports published until January 2010. All studies were examined with a special emphasis on the quality of the schizophrenia diagnosis and the definition of the chromosomal aberration. The quality of the phenotype was rated for each case by two investigators (JWM and JEB) (Cohen's κ .68; $p < .001$). Ratings were based on the clinical description, the presence of a psychiatric diagnosis, or a classification system (Table S4 in Supplement 1). This permitted a ranking of regions on the basis of the most valid clinical diagnosis of schizophrenia.

Because we focused on gene-containing CNVs most probable to have pathogenic effects in our CNV analysis, we compared the whole-genome CNV data with cytogenetic abnormalities found in case reports. Cytogenetic visible deletions and duplications might constitute a less common but potentially stronger influence on risk for schizophrenia, and these are likely to have direct effects on gene expression. Because of their relatively large size, these chromosomal aberrations might encompass multiple genes, which suggests a role in modulating gene expression (29). For example, it is suggested that the combined haploinsufficient expression of multiple genes in the 22q11 microdeletion causes the high rates of schizophrenia observed in patients with this 3-Mb deletion syndrome (encompassing 30 genes) (22). We were able to select, from both the CNV and cytogenetic abnormality loci, cytogenetic regions of interest (CROIs): loci where at least one case report and two Dutch schizophrenia CNV cases overlapped. Together with a systematic review of CNV findings from the literature, we highlight (new) chromosomal variation loci for schizophrenia susceptibility.

We were interested, in addition to selecting CROIs, to see whether we were able to combine our CNV findings and literature loci in a different way. For this sub-analysis we included all articles describing CNVs in schizophrenia patients (Table S3 in Supplement 1) with special emphasis on the description of candidate genes residing in these loci. We counted all deletions and duplications from our CNV analysis showing overlap with the candidate gene loci from the literature. Cases and control subjects were compared on the number of deletions and duplications within these regions.

See Tables S2, S3, and S4 in Supplement 1 for further details on the methods.

Results

Systematic Genome-Wide CNV Analysis

In total, 7211 CNVs passed the quality control for QuantiSNP, and 21,182 CNVs passed for PennCNV. More than 50% of all CNVs

were gene-containing: 3767 for QuantiSNP, and 13,849 for PennCNV. In total, 2437 gene-containing CNVs were called by both algorithms and were included in the study. These CNVs were found in 659 unique cases and 508 unique control subjects.

The poor overlap between QuantiSNP and PennCNV is striking, although not unusual (26,27,30). A recent study by Dellinger *et al.* (26) compared, among others, the PennCNV and QuantiSNP algorithms and found that the average number of detected CNVs differed, depending on the algorithm. The call numbers were correlated to sensitivity, specificity, and κ . Tsuang *et al.* (31) also demonstrated that the number of CNVs identified depends on the algorithm(s) used. They showed, in accordance with our own findings, that PennCNV called far more CNVs compared with QuantiSNP.

It is known that modifying parameters affects CNV detection. The number of calls as well as the size of predicted CNVs are affected (26,27). When relaxing some of the parameter settings, we saw differences in the number of CNV calls also influencing the number of overlapping CNVs of both algorithms (data not shown). When increasing the filtering stringency from 10 to 30 consecutive SNP markers, we observed more overlapping CNVs between PennCNV and QuantiSNP, which affect especially CNVs of larger size. The total overlap percentage between the two datasets increased from 27.4% to 46.8%. The use of overlapping calls from both algorithms should increase the confidence of our dataset and gives clearer indications of the CNV boundaries (27,31).

To investigate the reliability of our CNV dataset, we visually inspected a set of common and rare CNVs. Rare CNVs were detected in one patient only. Inspection of their intensity plots revealed that they were likely to be true CNV loci. We selected and validated, on the basis of their gene content, four CNVs by either multiplex ligation-dependent probe amplification (duplication 2p25.3, deletion 2p16.3, and deletion 9q33.1) or genomic qPCR (duplication 5p15.2), as described previously (13). All four CNVs were confirmed as true positive finding. We also investigated nine duplicate samples to calculate a concordance rate between calls from PennCNV and QuantiSNP. The calls were highly correlated with a correlation coefficient of .81.

The average number of gene containing CNVs/subject was 2.1 for both cases and control subjects. One subject can have multiple deletions or duplications of different lengths. Therefore, groups of deletions or duplications categorized by size across subjects are not mutually exclusive. When we limited our analysis to a subject having just one or more CNV/type and size, we indicated a total of 796 deletions (Table 1) and 827 duplications (Table 2). Cases seemed to have significantly more deletions compared with control subjects for all different size categories (Table 1). However, we did not find any significant difference between cases and control subjects for duplications (Table 2). This is in accordance with the literature, stating that deletions might be more pathogenic than duplications (28).

We identified a total of 1896 CNVs of ≥ 50 kb observed in 433 schizophrenia patients and 355 control subjects. For this effort we

Table 1. Deletions in Cases Versus Control Subjects

Size	Cases	Control Subjects	Total	No. CNVs/Subject Cases/Control Subjects	Case/Control Ratio	p (1-Sided)
All	470	326	796	.71/.64	1.11	.00559
≥ 50 kb	375	226	601	.57/.44	1.28	1.37e-5
≥ 500 kb	18	4	22	.027/.008	3.47	.00817
≥ 1 Mb	11	2	13	.017/.004	4.24	.02422

CNV, copy number variation.

Table 2. Duplications in Cases Versus Control Subjects

Size	Cases	Control Subjects	Total	No. CNVs/Subject Cases/Control Subjects	Case/Control Ratio	<i>p</i> (1-Sided)
All	455	372	827	.69/.73	.94	.0675
≥50 kb	417	344	761	.63/.68	.93	.0607
≥500 kb	39	27	66	.06/.05	1.11	.35135
≥1 Mb	14	9	23	.021/.018	1.2	.4162

CNV, copy number variation.

only included overlapping results of the two CNV detection methods with the purpose of increasing our confidence of the CNV calls. QuantiSNP yielded 3195 CNVs, whereas PennCNV analysis resulted in detection of 10,641 CNVs; overlap between the two methods was 27.4%. These findings (including exact base pair [bp] boundaries) are presented in Tables S1 and S2 in Supplement 1. Some of our CNVs overlap regions of LCRs. Previous analyses showed significant associations between LCRs and CNV regions (32–34). However, the HumanHap550v3 BeadArray we used for SNP genotyping primarily targets SNP loci at nonrepetitive genomic regions. Furthermore, the rate of LCR-flanking CNVs is not different between cases and control subjects (data not shown). The LCR-mediated CNVs should therefore have limited effect in our analyses. We found no evidence for increased double hit rate (i.e., multiple CNVs/subject) in schizophrenia patients compared with control subjects. In Figure 1, all CNVs ≥500 kb (*n* = 60) are represented to scale on each chromosome. Genes reported to be associated with schizophrenia are also indicated.

The graphical overview of this systematic genome-wide CNV analysis shows sites of increased deletions in cases versus control subjects for: 4q35.2, 15q13.2-q13.3, and 22q11.21. Interesting sites of increased duplications in cases versus control subjects are 1q42.3-43, 2p25.3, and 17q25.1. When examining loci of candidate genes for schizophrenia indicated by previous CNV studies, we observe significantly more CNV deletions of ≥50 kb within these gene regions for cases compared with control subjects (*p* value = .0004). However, there is no significant difference for CNV duplications of ≥50 kb within these candidate gene regions (*p* value = 1) (Table 3). The regions on chromosome 1q, 2p, 15q, and 22q have previously been highlighted in cytogenetic and CNV studies (10–15,18,35–38). The 4q35.2 region (188,329,837–191,164,126 bp) is a possible novel CNV region-of-interest, indicated by three schizophrenia cases with the deletion versus only one control subject. Our duplications at chromosome 17q (69,345,596–71,746,800 bp) in four cases and no control subjects were also not previously described to be associated with schizophrenia.

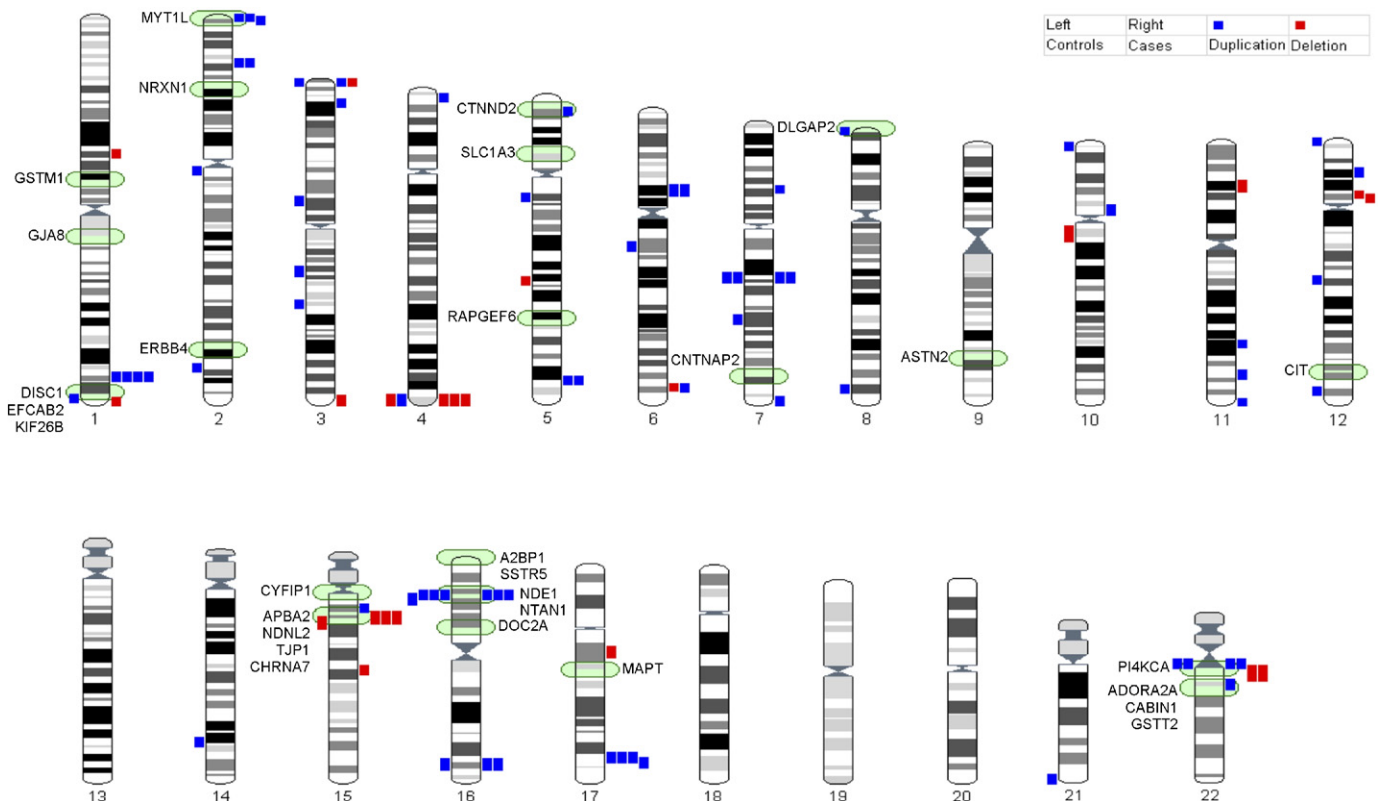


Figure 1. CNVs in 834 Dutch schizophrenia patients and 672 Dutch control subjects. Representation of a systematic genome-wide survey for copy number variations (CNVs) in 834 Dutch schizophrenia cases and 672 Dutch control subjects. On the left side of each chromosome, CNVs of control subjects are indicated. On the right side of each chromosome, CNVs of schizophrenia cases are indicated. Blue bars represent duplications, red bars represent deletions. In green, schizophrenia-associated genes from CNV literature are indicated.

Table 3. Schizophrenia Candidate Genes Residing in CNVs in Dutch Case-Control Study

Gene	Position	No. CNVs Overlapping		No. Cases (%)		No. Control Subjects (%)	
		Del	Dup	Del	Dup	Del	Dup
<i>A2BP1</i> (15)	16p13.2	4	3	3 (.69)	—	1 (.28)	3 (.85)
<i>ADAMTSL3</i> (18)	15q25.2	0	0	—	—	—	—
<i>ADORA2A</i> (47)	22q11.23	0	0	—	—	—	—
<i>APBA2</i> (18,48,49)	15q13.1	0	0	—	—	—	—
<i>ASTN2</i> (13)	9q33.1	0	1	—	1 (.23)	—	—
<i>CABIN1</i> (47)	22q11.23	0	0	—	—	—	—
<i>CHRNA7</i> (11,12)	15q13.3	3	2	3 (.69)	—	—	2 (.56)
<i>CIT</i> (15)	12q24.23	0	0	—	—	—	—
<i>CNTNAP2</i> (50)	7q35–q36.1	3	0	3 (.69)	—	—	—
<i>COMT</i> (37,38)	22q11.21	1	0	1 (.23)	—	—	—
<i>CTNND2</i> (13)	5p15.2	0	1	—	1 (.23)	—	—
<i>CYFIP1</i> (11)	15q11.2	7	10	4 (.92)	6 (1.39)	3 (.85)	4 (1.13)
<i>DISC1</i> (51)	1q42.2	0	5	—	3 (.69)	—	2 (.56)
<i>DLGAP2</i> (48)	8p23.3	0	0	—	—	—	—
<i>DOC2A</i> (48)	16p11.2	0	0	—	—	—	—
<i>EFCAB2</i> (18)	1q44	0	1	—	1 (.23)	—	—
<i>ERBB4</i> (14)	2q34	1	0	1 (.23)	—	—	—
<i>GJA8</i> (11)	1q21.1	0	0	—	—	—	—
<i>GSTM1</i> (47)	1p13.3	0	0	—	—	—	—
<i>GSTT2</i> (47)	22q11.23	0	1	—	—	—	1 (.28)
<i>KIF26B</i> (18)	1q44	1	1	1 (.23)	1 (.23)	—	—
<i>MAPT</i> (48)	17q21.31	0	0	—	—	—	—
<i>MYT1L</i> (13)	2p25.3	0	2	—	2 (.46)	—	—
<i>NDE1</i> (52)	16p13.1	0	6	—	3 (.69)	—	3 (.85)
<i>NDNL2</i> (48)	15q13.1	0	0	—	—	—	—
<i>NOTCH4</i> (41,53)	6p21.32	0	0	—	—	—	—
<i>NRGN</i> (41)	11q24.2	0	0	—	—	—	—
<i>NRXN1</i> (13,14,18,48,49)	2p16.3	3	2	3 (.69)	1 (.23)	—	1 (.28)
<i>NTAN1</i> (52)	16p13.1	0	4	—	3 (.69)	—	1 (.28)
<i>PGBD1</i> (41)	6p22.1	0	0	—	—	—	—
<i>PI4KCA</i> (47)	22q11.21	2	1	2 (.46)	—	—	1 (.28)
<i>PRODH</i> (37,38)	22q11.21	11	23	9 (2.08)	13 (3.0)	2 (.56)	10 (2.82)
<i>RAPGEF6</i> (15)	5q31.1	0	0	—	—	—	—
<i>SLC1A3</i> (14)	5p13.2	0	0	—	—	—	—
<i>SSTR5</i> (47)	16p13.3	0	0	—	—	—	—
<i>TJP1</i> (48)	15q13.1	0	0	—	—	—	—
Total		36	63	30 (6.9)	35 (8.1)	6 (1.7)	28 (7.9)

Deletions (Del) $p = .0004$. Duplications (Dup) $p = 1$. CNV, copy number variation.

Comparison with Literature

An overview of all cytogenetic abnormalities described in literature can be found in Figure S2 and Table S4 in Supplement 1. It seems that schizophrenia susceptibility loci are not equally distributed across the human genome but rather cluster together in specific CROIs. When we compare cytogenetic abnormalities with CNVs found in Dutch schizophrenia cases and control subjects, we observe the following CROIs: 1q42, 5q35.1, 7q21.12–q21.13, and 22q11 (Figure 2). These regions are reported by at least one cytogenetic study and have CNVs in at least two cases. The regions on chromosome 1q and 22q have been described before (10–12,14,15,18,35–38) and also appeared as interesting sites in our CNV analysis. Our CNV on 1q42, although overlapping the large region from the 1;11 translocation, is not overlapping the exact boundaries of the *DISC1* gene (approximately 3 Mb between CNV and gene region). For the 22q11.2 region we find both deletions and duplications (Figure S3 in Supplement 1). This region is known from the 22q11.2 Deletion Syndrome (MIM ID #188400). Most patients with this syndrome share a 3-Mb loss, although a nested 1.5-Mb deletion is also ob-

served along with infrequent atypical deletions. The LCRs are flanking and mediating the deletion regions (39). We identified 12 deletions within the 22q11.2 interval. Of these, 1 (case) was consistent with the larger deletion, 0 were consistent with the shorter deletion, and 11 (9 cases and 2 control subjects; 729 kb: 19,063–19,792 kb; 156 kb: 17,258 kb –17,434 kb) were atypical. For chromosome 7, although found in two cases and one cytogenetic study, also two control subjects are identified to have a duplication at this site. Chromosome 5q (168,385,055–169,072,475 bp) is a possible novel region, indicated by clustering of cytogenetic abnormalities and CNV findings. However, the only investigated candidate genes residing in this locus are *SLIT3*, *GABRP*, and *FGF18*, with inconsistent results (<http://www.schizophreniaforum.org/res/sczgene>).

An overview of all CNV studies from literature can be found in Table S3 in Supplement 1. Lack of information about the real number of CNVs occurring in cases and control subjects and unclear diagnostic criteria limited our ability to give a complete and correct overview. However, comparison with our own CNV data and schizophrenia-associated genes from CNV literature indicates some clus-

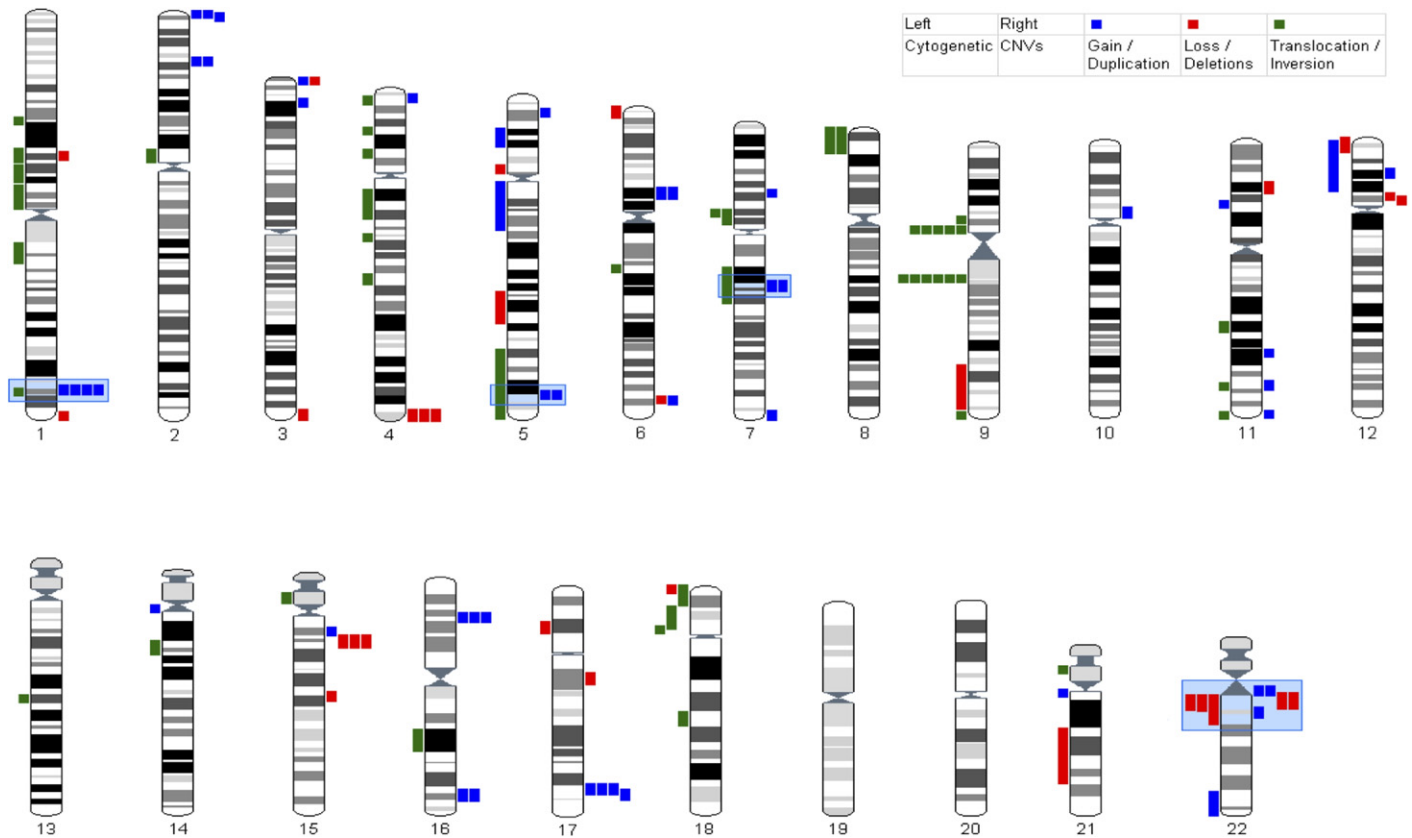


Figure 2. Copy number variations (CNVs) in 834 Dutch schizophrenia cases compared with cytogenetic aberrations from literature. On the left side of each chromosome, chromosomal aberrations from literature are indicated. On the right side of each chromosome, copy number variants in 834 Dutch schizophrenia cases are indicated. Blue bars represent duplications or gains, red bars represent deletions or losses, and green bars represent balanced translocations/inversions. Blue boxes indicate regions where at least one chromosomal aberration and two copy number cases overlapped. The regions 1q42.3-q43 and 22q11.2 were previously indicated by linkage (54,55) or association studies (56,57) and often supported by CNV findings from literature. The regions on chromosome 5q35 and on chromosome 7q21.12-q21.13 are potentially novel regions, not indicated in association or linkage studies (54–57). For the exact base pair boundaries of overlap regions we refer to Tables S1 and S4 in Supplement 1.

tered regions of interest (e.g., on chromosome 1q, 15q, and 22q11) (Figure S1 and Table S3 in Supplement 1).

Discussion

We performed a genome-wide analysis of CNVs in 834 schizophrenia cases and 672 unaffected control subjects from the Netherlands. The most apparent observation is that cases showed significantly more deletions for all size categories, an effect not seen for duplications. When focusing on previously reported schizophrenia-candidate CNV loci, a similar effect was observed, with cases showing significantly more deletions compared with control subjects. These findings suggest that deletion CNVs might be more prevalent in individuals with schizophrenia and therefore confer a higher risk factor for pathogenicity. This is in line with previous CNV studies that also suggested that duplications are genetic alterations that are better tolerated in the genome and that deletions have a higher likelihood of being pathogenic (19,28).

Additionally, we reviewed all cytogenetic and chromosomal abnormalities described for schizophrenia patients (Table S4 in Supplement 1). The integrated identification of CNV regions and cytogenetic and chromosomal loci highlight the following regions of interest (CROIs): 1q42, 2p25, 15q13, and 22q11 as loci previously described. Secondly, we identified 4q35 (increased number of deletions in cases compared with control subjects) and 17q25 (increased number of duplications in cases compared with control

subjects) as CROIs from our CNV analysis. An interesting region from the integration of CNVs and cytogenetic abnormality reports is 5q35. Furthermore, we see regions of genomic variation existing both in cases and in control subjects. These CNVs are very difficult to interpret in the absence of further correlative data.

It is probable that a combination of both common and rare risk variants is involved in schizophrenia etiology. Common CNVs (allele frequency >5%) are almost always inherited and comprise most CNV differences between individuals (19). In two recent large GWAS studies for schizophrenia, *TCF4* and *ZNF804A* have been found as best hits (40,41). These are transcription factors that might regulate expression of many genes. This raises the possibility that common variation might confer susceptibility to schizophrenia. However, until now, mostly rare (<.5%–1%) and large (>100 kb) CNVs have been implicated in schizophrenia (42). This study both confirms the involvement of the recurrent variants (e.g., for 1q42) and adds to the implication of rare CNVs with an elevated risk for schizophrenia (43).

Improvement of genomic array and sequencing technologies will provide higher-resolution and more accurate detection of CNVs (44). These new technologies will aid the discovery of new schizophrenia risk regions and candidate genes. In addition to technological advances, there is need for development of improved algorithms and statistical methods for reliable CNV calling in existing GWAS datasets. There is still a large-scale variability between calling

algorithms, so the use of multiple algorithms specific to the array platform used seems recommended (31). Moreover, it is important that studies with large sample sizes will make their CNV calls publicly available for further study and systematic meta-analyses (e.g., via the Database of Genomic Variants). New bioinformatic tools that combine different genomic data sources are necessary to gain better insight into this complex psychiatric disorder.

It is possible that our study is subject to various kinds of biases. First, overrepresentation of some regions in our literature study might represent availability of probes rather than significant association with schizophrenia. This is clearly the case for the 22q11 deletion, which results in reporter bias. When a specific region has been linked to schizophrenia, the locus becomes more interesting for research, resulting in more tested cases. In our effort to include previously reported cytogenetic abnormalities in schizophrenia, we noticed that approximately one-third of the cases showed either mental retardation or physical anomalies. For CNV studies in literature, by contrast, this type of clinical data was not always mentioned. However, both features are associated with a high incidence of cytogenetic abnormalities (45). Therefore, the loci indicated in this study might be linked to mental retardation and dysmorphic features rather than to the schizophrenia phenotype alone.

For complex traits like schizophrenia, it is important to consider all classes of variation: cytogenetic abnormalities as well as SNPs and CNVs (46). This study provides a comprehensive overview of all CNVs ≥ 50 kb found in a homogeneous Dutch sample of schizophrenia cases and unaffected control subjects. Deletions are more prevalent in schizophrenia patients and might confer a higher risk factor than duplications. We combined these data with a systematic overview of cytogenetic and CNV findings from the literature. Our results confirm the already-known results for, for example, the 15q13 and 22q11.2 deletions and highlight novel candidate regions. These regions indicate how classical cytogenetic findings and our current understanding of genomic variation can be combined and might contain loci with rare variants contributing to disease susceptibility. By reporting our complete CNV dataset and systematically reviewing the literature, we hope this study is a step toward understanding the role of genomic chromosomal variation in schizophrenia etiology. Future studies will include genome-wide sequencing, so the whole spectrum of genomic and genetic variation can be examined for involvement in schizophrenia susceptibility.

The Genetic Risk and Outcome in Psychosis (GROUP) project was supported by a grant from ZON-MW (<http://www.zonmw.nl/en/programmes/all-programmes/mental-health/>), within the Mental Health program (project number: 10.000.1001). In addition, genome-wide single nucleotide polymorphism genotyping of the Dutch schizophrenia sample was funded by the National Institute of Mental Health Grant RO1 MG078075 to RAO. We thank Thomas Zhangzy for helping with the copy number variations calling procedures and Flip Mulder for technical support. We acknowledge members of the department of Medical Genetics and the Department of Psychiatry for critically reading the manuscript.

The GROUP Consortium members: René S. Kahn (Rudolf Magnus Institute of Neuroscience, Department of Psychiatry, University Medical Centre Utrecht, Utrecht, the Netherlands), Don H. Linszen (Academic Medical Centre University of Amsterdam, Department of Psychiatry, Amsterdam, the Netherlands), Jim van Os (Maastricht University Medical Centre, the Netherlands), Durk Wiersma (University Medical Centre Groningen, Department of Psychiatry, University of Groningen, the Netherlands), Richard Bruggeman (University Medical Centre Gro-

ningen, the Netherlands), Wiepke Cahn (Rudolf Magnus Institute of Neuroscience, Department of Psychiatry, University Medical Centre Utrecht, Utrecht, the Netherlands), Lieuwe de Haan (Academic Medical Centre University of Amsterdam, the Netherlands), Lydia Krabbendam (Maastricht University Medical Centre, South Limburg Mental Health Research and Teaching Network, Maastricht, the Netherlands), and Inez Myin-Germeys (Maastricht University Medical Centre, the Netherlands).

The authors report no biomedical financial interests or potential conflicts of interest.

Supplementary material cited in this article is available online.

- Gottesman I (1994): Schizophrenia epigenesis: Past, present, and future. *Acta Psychiatr Scand* 90:26–33.
- Jablensky A (2000): Epidemiology of schizophrenia: The global burden of disease and disability. *Eur Arch Psychiatry Clin Neurosci* 250:274–285.
- Andreasen N (1995): Symptoms, signs, and diagnosis of schizophrenia. *Lancet* 346:477–481.
- Burmeister M, McInnis M, Zöllner S (2008): Psychiatric genetics: Progress amid controversy. *Nat Rev Genet* 9:527–540.
- Tsuang M (2000): Schizophrenia: Genes and environment. *Biol Psychiatry* 47:210–220.
- Sullivan P (2005): The genetics of schizophrenia. *PLoS Med* 2:e212.
- Sullivan P, Kendler K, Neale M (2003): Schizophrenia as a complex trait: Evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry* 60:1187–1192.
- Tam G, Redon R, Carter N, Grant S (2009): The role of DNA copy number variation in schizophrenia. *Biol Psychiatry* 66:1005–1012.
- International Schizophrenia Consortium, Purcell S, Wray N, Stone J, Visscher P, O'Donovan M, *et al.* (2009): Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460:748–752.
- Kirov G, Grozeva D, Norton N, Ivanov D, Mantripragada K, Holmans P, *et al.* (2009): Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Hum Mol Genet* 18:1497–1503.
- Stefansson H, Rujescu D, Cichon S, Pietiläinen O, Ingason A, Steinberg S, *et al.* (2008): Large recurrent microdeletions associated with schizophrenia. *Nature* 455:232–236.
- The International Schizophrenia Consortium (2008): Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 455:237–241.
- Vrijenhoek T, Buizer-Voskamp JE, van der Stelt I, Strengman E, GROUP Consortium, Sabatti C, *et al.* (2008): Recurrent CNVs disrupt three candidate genes in schizophrenia patients. *Am J Med Genet Part B* 83:504–510.
- Walsh T, McClellan J, McCarthy S, Addington A, Pierce S, Cooper G, *et al.* (2008): Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 320:539–543.
- Xu B, Roos J, Levy S, van Rensburg E, Gogos J, Karayiorgou M (2008): Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet* 40:880–885.
- Choy K, Setlur S, Lee C, Lau T (2010): The impact of human copy number variation on a new era of genetic testing. *BJOG* 117:391–398.
- Mefford H, Sharp A, Baker C, Itsara A, Jiang Z, Buysse K, *et al.* (2008): Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med* 359:1685–1699.
- Need A, Ge D, Weale M, Maia J, Feng S, Heinzen E, *et al.* (2009): A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLoS Genet* 5:e1000373.
- Bassett A, Scherer SW, Brzustowicz L (2010): Copy number variations in schizophrenia: Critical review and new perspectives on concepts of genetics and disease. *Am J Psychiatry* 167:899–914.
- Iafate A, Feuk L, Rivera M, Listewnik M, Donahoe P, Qi Y, *et al.* (2004): Detection of large-scale variation in the human genome. *Nat Genet* 36:949–951.
- Vorstman JAS, Staal WG, van Daalen E, van Engeland H, Hochstenbach PFR, Franke L (2006): Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism. *Mol Psychiatry* 11:18–28.
- Bray N (2008): Gene expression in the etiology of schizophrenia. *Schizophrenia Bull* 34:412–418.

23. Moore J, Williams S (2009): Epistasis and its implications for personal genetics. *Am J Med Genet A* 85:309–320.
24. Colella S, Yau C, Taylor J, Mirza G, Butler H, Clouston P, *et al.* (2007): QuantiSNP: An objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res* 35:2013–2025.
25. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant S, *et al.* (2007): PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 17:1665–1674.
26. Dellinger A, Saw S, Goh L, Seielstad M, Young T, Li Y (2010): Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays. *Nucleic Acids Res* 38:e105.
27. Winchester L, Yau C, Ragoussis J (2009): Comparing CNV detection methods for SNP arrays. *Brief Funct Genomics Proteomics* 8:353–366.
28. Lee C, Iafrate A, Brothman A (2007): Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. *Nat Genet* 39(suppl): S48–S54.
29. Muir W, Pickard B, Blackwood D (2006): Chromosomal abnormalities and psychosis. *Br J Psychiatry* 188:501–503.
30. Marenne G, Chanoch S, Rothman N, Rodriguez-Santiago B, Rico D, Pita G, *et al.* (2009): CNV assessment using Illumina Infinium 1M platform: Agreement according to algorithm and source of DNA. *Ann Hum Genet* 73:658–669.
31. Tsuang D, Millard S, Ely B, Chi P, Wang K, Raskind W, *et al.* (2011): The effect of algorithms on copy number variant detection. *PLoS ONE* 5:e14456.
32. Bailey J, Kidd J, Eichler E (2008): Human copy number polymorphic genes. *Cytogenet Genome Res* 123:234–243.
33. Freeman J, Perry G, Feuk L, Redon R, McCarroll S, Altshuler D, *et al.* (2006): Copy number variation: New insights in genome diversity. *Genome Res* 16:949–961.
34. Sharp A, Locke D, McGrath S, Cheng Z, Bailey J, Vallente R, *et al.* (2005): Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* 77:78–88.
35. Blackwood D, Fordyce A, Walker M, St. Clair D, Porteous D, Muir W (2001): Schizophrenia and affective disorders—cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: Clinical and P300 findings in a family. *Am J Hum Genet* 69:428–433.
36. Hoogendoorn M, Vorstman J, Jalali G, Selten J, Sinke R, Emanuel B, Kahn RS (2008): Prevalence of 22q11.2 deletions in 311 Dutch patients with schizophrenia. *Schizophr Res* 98:84–88.
37. Karayiorgou M, Gogos J (2004): The molecular genetics of the 22q11-associated schizophrenia. *Mol Brain Res* 132:95–104.
38. Weksberg R, Stachon A, Squire J, Moldovan L, Bayani J, Meyn S, *et al.* (2007): Molecular characterization of deletion breakpoints in adults with 22q11 deletion syndrome. *Hum Genet* 120:837–845.
39. Shaikh T, Kurahashi H, Saitta S, Mizraha O'Hare A, Hu P, Roe B, *et al.* (2000): Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: Genomic organization and deletion endpoint analysis. *Hum Mol Genet* 9:489–501.
40. O'Donovan M, Craddock N, Norton N, Williams H, Peirce T, Moskvina V, *et al.* (2008): Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet* 40:1053–1055.
41. Stefansson H, Ophoff R, Steinberg S, Andreassen OA, Cichon S, Rujescu D, *et al.* (2009): Common variants conferring risk of schizophrenia. *Nature* 460:744–747.
42. Manolio T, Collins F, Cox N, Goldstein D, Hindorff L, Hunter D, *et al.* (2009): Finding the missing heritability of complex diseases. *Nature* 461:747–753.
43. Duan J, Sanders A, Gejman P (2010): Genome-wide approaches to schizophrenia. *Brain Res Bull* 83:93–102.
44. Alkan C, Kidd J, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F, *et al.* (2009): Personalized copy number and segmental duplication maps using next-generation sequencing. *Nat Genet* 41:1061–1067.
45. Riegel M, Baumer A, Jamar M, Delbecq K, Herens C, Verloes A, Schinzel A (2001): Submicroscopic terminal deletions and duplications in retarded patients with unclassified malformation syndromes. *Hum Genet* 109:286–294.
46. Conrad D, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, *et al.* (2010): Origins and functional impact of copy number variation in the human genome. *Nature* 464:704–712.
47. Rodriguez-Santiago B, Brunet A, Sobrino B, Serra-Juhé C, Flores R, Armengol L, *et al.* (2009): Association of common copy number variants at the glutathione S-transferase genes and rare novel genomic changes with schizophrenia. *Mol Psychiatry* 15:1023–1033.
48. Guilmatre A, Dubourg D, Mosca A, Legallic S, Goldenberg A, Drouin-Garraud V, *et al.* (2009): Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation. *Arch Gen Psychiatry* 66:947–956.
49. Kirov G, Gumus D, Chen W, Norton N, Georgieva L, Sari M, *et al.* (2008): Comparative genome hybridization suggests a role for *NRXN1* and *APBA2* in schizophrenia. *Hum Mol Genet* 17:458–465.
50. Friedman J, Vrijenhoek T, Markx S, Janssen I, van der Vliet W, Faas B, *et al.* (2008): *CNTNAP2* gene dosage variation is associated with schizophrenia and epilepsy. *Mol Psychiatry* 13:261–266.
51. St. Clair D, Blackwood D, Muir W, Carothers A, Walker M, Spowart G, *et al.* (1990): Association within a family of a balanced autosomal translocation with major mental illness. *Lancet* 336:13–16.
52. Ingason A, Rujescu D, Cichon S, Sigurdsson E, Sigmundsson T, Pietiläinen O, *et al.* (2011): Copy number variations of chromosome 16p13.1 region associated with schizophrenia. *Mol Psychiatry* 16:17–25.
53. The International Schizophrenia Consortium (2009): Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460:748–752.
54. Badner J, Gershon E (2002): Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. *Mol Psychiatry* 7:405–411.
55. Lewis C, Levinson D, Wise L, DeLisi L, Straub R, Hovatta I, *et al.* (2003): Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am J Hum Genet* 73:34–48.
56. Allen N, Bagade S, McQueen M, Ioannidis J, Kavvoura F, Khoury M, *et al.* (2008): Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: The SzGene database. *Nat Genet* 40:827–834.
57. Wall J, Pritchard J (2003): Assessing the performance of the haplotype block model of linkage disequilibrium. *Am J Hum Genet* 73:502–515.