Supplementary Online Content


eMaterial 1. Samples, genotyping methods and quality control (QC) for the replication cohorts

eMaterial 2. Distal 16p11.2 probe coverage

eMaterial 3. Estimating the selection pressure against the distal 16p11.2 deletion

eMaterial 4. Members of the Molecular Genetics of Schizophrenia Consortium

eMaterial 5. Members of the Wellcome Trust Case Control Consortium 2

eFigure 1. z-Score histogram for distal 16p11.2 in the Bulgarian trio data

eFigure 2. Principal component analysis PCA plot for the Ashkenazi discovery cohort

eFigure 3. 16p11.2 region

eTable 1. Previous observations of distal 16p11.2 deletions

eTable 2. Phenotypic data on carriers of 16p11.2 distal deletions

eTable 3. Publications used to estimate the selection pressure (s) operating against deletions at distal 16p11.2

eTable 4. CNVs in the Ashkenazi Jewish (AJ) and Bulgarian (BG) cohorts at loci previously implicated in schizophrenia

eReferences

This supplementary material has been provided by the authors to give readers additional information about their work.
eMaterial 1. Samples, genotyping methods and quality control (QC)

for the replication cohorts

Japanese sample

Sample description: The data from the Japanese sample are based on those reported by Ikeda et al, 2010: 1139 age- and sex-matched unrelated subjects of Japanese ethnicity (575 schizophrenic [SZ] patients and 564 control subjects). Control subjects were members of the general public who had no personal history of mental disorders. This was ascertained during face-to-face interviews where subjects were asked if they had suffered an episode of depression, mania, or psychotic experiences or if they had received treatment for any psychiatric disorder. Patients were entered into the study if they 1) met DSM-IV criteria for schizophrenia; 2) were physically healthy and had normal routine laboratory tests; and 3) had no mood disorders, substance abuse, neurodevelopmental disorders, epilepsy, or known mental retardation. Consensus diagnoses were made by at least 2 experienced psychiatrists according to DSM-IV criteria on the basis of unstructured interviews with patients and families and review of medical records. Major medical conditions were obtained and body mass index was recorded based on available hospital records. After description of the study, written informed consent was obtained from each subject. This study was approved by the ethics committees of each participating university.

Genotyping and QC: All cases and controls were genotyped with Affymetrix 5.0 arrays. For the purpose of this study, we reanalyzed the raw data of that study. We used similar QC criteria as for the Bulgarian (BG) sample (except the z-score method). The total number of cases and controls in the Japanese sample was 490 and 516, respectively, after applying these filtering criteria.
Molecular Genetics of Schizophrenia (MGS)

**Sample description:** The MGS consists of cases and controls from European and African ancestry, described by Levinson et al., 2011. In total (after QC), there are **2671 cases and 2648 controls of European American ancestry and 1274 cases and 963 controls of African American ancestry.** Details of the diagnostic methods have been described by Levinson et al., 2011; patients meeting *DSM-IV* criteria for SZ or schizoaffective disorder were included. Concomitant medical diagnoses were not available for this cohort.

**Genotyping and QC:** The specimens were assayed at the Broad Institute, Cambridge, Massachusetts, by using Affymetrix 6.0 genotyping arrays. Copy number variations (CNVs) were detected, or “called”, with the Birdseye module of the Birdsuite software package, version 2 (internal version 1.3), which uses a hidden Markov model algorithm. The data were normalized within plates of up to 92 DNA samples. HG18 human genome build locations are reported.

The CNV calls were merged if nearby pairs (or sequential pairs) of deleted or duplicated segments flanked a “normal” segment containing less than 20% of the probes in the merged CNV (primarily in segmental duplication regions). CNV calls were excluded if they overlapped (50%) with telomeres (100 000 base pairs [bp]) and centromeres, where CNV calls may be unreliable, or immunoglobulin gene regions where Epstein-Barr virus transformation causes structural changes. CNVs were also excluded if seen predominantly on 1 or 2 plates, suggesting artifact.

Samples were excluded based on 1) samples with total numbers of narrowly defined deletions or duplications exceeding the group mean by 3 SDs, 2) those with more than 2 chromosomes with outlier call numbers, 3) data for outlier chromosomes for subjects with 1 or 2 such chromosomes, and 4) samples (mostly lymphoblastic cell lines) with probe intensity variances exceeding the group mean by 4 SDs (predicting fewer CNV calls).
International Schizophrenia Consortium (ISC)

Sample description: The data from the ISC study are from 6 European populations and have been published before. The sample included **3045 cases and 3185 controls**, after excluding the BG cases that are probands in the trios from the current study.

Genotyping and QC: The samples were genotyped on Affymetrix 5.0 or 6.0 arrays at the Broad Institute. For the purpose of this study we examined the raw data and applied to it the same QC criteria as for the BG sample above, (except the \( z \)-score method).

Vacic et al, 2011

Sample description: The initial discovery data set was composed of 1761 unrelated subjects analyzed on the NimbleGen HD2 Array-CGH platform. The unfiltered sample consisted of 913 patients and 848 controls ascertained at 10 sites. Samples from the Trinity College Dublin cohort (n = 45) were removed for the purposes of this analysis, to avoid redundancy with the Wellcome Trust Case Control Consortium 2 (WTCCC2)/Irish sample (described below). The final discovery data set (after QC) consisted of **757 unrelated patients and 742 controls**.

Genotyping and QC: Microarray hybridizations using the NimbleGen HD2 platform were performed at the service laboratory of Roche NimbleGen according to the manufacturer’s specifications. All processing of microarray data was performed at Cold Spring Harbor Laboratory. Two segmentation algorithms were used to discover CNVs in the GC-corrected NimbleGen HD2 data: HMMSeg\(^ \text{8} \) and Genome Alteration Detection Analysis\(^ \text{9} \) (GADA; http://biron.usc.edu/~piquereg/GADA). CNVs detected by both algorithms were used for downstream data processing and analysis. CNVs detected by only 1 algorithm were excluded. In addition, CNVs of the same type (ie, deletion or duplication) that were separated by \( \leq 3 \)
probes were merged into 1 contiguous segment. The proximal and distal boundaries of overlapping and adjacent CNVs were defined by the minimal chromosomal start position and the maximal chromosomal end position of the CNVs. CNVs with frequency >1% were removed. Rare CNV calls that passed the 1% frequency filter were further filtered based on the confidence score (CS). As the CS score of a CNV call we used the \( P \) value derived from our outlier detection genotyping method, median z-score outlier detection (MeZOD), as previously published.\(^{10}\)

**WTCCC2/Irish sample**

**Sample description:** Controls in this sample were from 3 sources: 2663 persons from the 1958 British Birth Cohort and 2533 persons from the National Blood Service in Britain who were genotyped as part of the WTCCC2 study (http://www.wtccc.org.uk/ccc2/). The third control sample consisted of 979 Irish people obtained with written informed consent from the Irish GeneBank and represented blood donors from the Irish Blood Transfusion Service (IBTS) recruited in the Republic of Ireland. Blood donors were included in the study if they met ethnicity criteria (having 4 ethnically Irish grandparents). In Ireland and in the United Kingdom, blood donors are not screened formally for mental disorders, but people who report taking regular prescribed medication are excluded from donation. Donors are not financially renumerated. The 1958 Birth Cohort (also known as the National Child Development Study) includes all births in England, Wales, and Scotland, during 1 week in 1958. From an original sample of over 17 000 births, 9377 cohort members were visited at home for a biomedical examination at the age of 44-45 years (http://www.b58cgene.sgul.ac.uk/followup.php), and provided 7692 blood samples. The samples from these control cohorts are anonymized and we cannot obtain any further information about the medical or psychiatric histories of the carries with deletions.
The case samples were recruited through community mental services and inpatient units in the Republic of Ireland and Northern Ireland following similar research protocols and with local ethics approval. Diagnosis of a major psychotic disorder was made by the consensus lifetime best estimate method using DSM-IV criteria with all available information (interview, family or staff report, and chart review). The final number of subjects after QC filtering and relevant exclusions was 1269 patients (all from Ireland) and 6175 controls.

**Genotyping and Quality Control (QC):** All samples were genotyped using Affymetrix 6.0 arrays, and CNV calls were generated using Birdseye from Birdsuite, version 1.5.5. CNV calls were restricted to autosomal regions and standard QC included removal of CNVs of <100 kilobase (kb) and logarithm of odds score of <10. Common CNVs were excluded based on >50% overlap with a region that was copy number variable in at least 1% of samples. Individuals with more than 30 CNV calls or a total CNV sequence length of >10M bp were excluded, as well as any sample failing standard QC metrics based on single-nucleotide polymorphism (SNP) data.

**Swedish sample**

**Sample description:** All subjects were born in Sweden. Schizophrenia cases were identified via the Swedish Hospital Discharge Register containing all individuals hospitalized in Sweden since 1973, described in Bergen et al, 2012. Diagnoses were established by the attending physician and confirmed in a subset of subjects by medical record review. Cases, aged 18-65 years, must have had at least 2 hospitalizations with a SZ diagnosis, both parents born in Sweden, and signed written informed consent. Subjects were interviewed twice. A checklist for ancestry, somatic diseases (self-reported), and alcohol/drug use was completed. DNA was extracted from peripheral blood. Prevalence and recurrence risks for this definition...
of schizophrenia are almost identical to that accepted by the field. A medical record review in the project showed 97% concordance with DSM-IV criteria. Control subjects, also selected through registers, were group matched by age, sex, and county of residence. Controls had never received a discharge diagnosis of schizophrenia or schizoaffective disorder. All subjects were at least 18 years old and gave written informed consent to participate (however, we note that the control with a 16p11.2 deletion was recruited at the age of 45).

**Genotyping and QC:** Genotyping was performed using Genome-Wide Human SNP Array 6.0. Subjects and SNPs with genotype call rates <95% were excluded. Population outliers assessed by multidimensional scaling were removed. One member of any pair of subjects demonstrating high relatedness (PL_HAT >.20) was arbitrarily selected for removal. SNPs with marked departures from Hardy-Weinberg equilibrium (P<1×10^-6) or very low minor allele frequencies (<1%) were excluded. SNPs with nonrandom genotyping failure, inferred by the flanking haplotypic background using the PLINK^13 mishap test, P < 1×10^-10, were removed from analysis. Plate-based associations of P <1×10^-6 were taken as evidence of gross nonrandom plate failure based on a comparison of allele frequency of each plate to all others and were removed on a plate-by-plate basis. Intensity data from both SNP and CNV probes were used to identify CNVs based on a hidden Markov model. Only subjects who passed SNP quality control filters were considered for CNV analyses. CNVs with >1% frequency, as well as common CNV regions indexed by HapMap, common rearrangements observed in blood, and any CNVs spanning large gaps (such as a centromere), were excluded. Analyses were restricted to autosomal CNVs greater than 100 kb. Seven subjects were removed for having either >10 Mb or >20 total CNVs. This resulted in 4429 CNVs (duplication:deletions ratio = 1.68) in **1506 schizophrenia cases** and **2089 controls** that remained for analysis (834 bipolar cases are not used in the current analysis).
German sample

A total of 1581 patients were included in the CNV analyses. These were recruited from consecutive admissions to psychiatric inpatient units and were all of German descent. Part of the sample (n = 487) was included in a previously published genome-wide association study (GWAS) and are described in detail as patient sample “GWAS Bonn-Mannheim” in Rietschel et al, 2011. A lifetime “best estimate” diagnosis of SZ or schizoaffective disorder according to DSM-IV criteria was assigned on the basis of the Structured Clinical Interview or the OPCRIT, medical records, and family history. Medical conditions relevant to psychiatric diagnosis were also recorded on the basis of these sources of information. After quality control, the sample consisted of 1386 patients with a male/female ratio of 1.14.

We used data from 1643 controls. Of these, 286 were recruited within the “Imaging Genetics” subproject of the German Integrated Genome Research Network “Systematic Investigation of the Molecular Causes of Major Mood Disorders and Schizophrenia.” None of these controls had a lifetime or family history of schizophrenia or any affective disorder. The remaining controls were drawn from 3 population-based epidemiological studies and are described in detail as control sample “GWAS Bonn-Mannheim” in Rietschel et al. After QC, the sample consisted of 1627 controls with a male/female ratio of 0.5.

Genotyping and QC: Venous blood samples were obtained from all participants. These were genotyped separately using the following Illumina BeadArrays: HumanHap550v3, Human610-Quadv1, and Human660W-Quad. Only those markers common to all 3 arrays were analyzed. SNPs with a call rate of < 97% were omitted and individuals were excluded from the dataset for the following reasons: 1) SNP call rate < 97%; 2) differences between X-
chromosomally inferred and phenotypic sex; 3) DNA sample doublets identified by identity-by-state estimates (defined as IBS = 2); 4) cryptic relatedness (IBS ≥ 1.6); and 5) population outliers according to multidimensional scaling with HapMap phase 2 data.

**CNV detection:** To identify potential CNVs, the SNP-chip data of each participant was analyzed using QuantiSNP (version 2.1, [http://www.well.ox.ac.uk/QuantiSNP](http://www.well.ox.ac.uk/QuantiSNP)) and PennCNV (version 2010May01, [http://www.openbioinformatics.org/penncnv/](http://www.openbioinformatics.org/penncnv)). Participants were removed from the dataset if their SD from the log R ratio calculated over all SNPs exceeded 0.30.
**eMaterial 2. 16p11.2 Probe Coverage**

We examined the probe coverage of all arrays used in the current study, to ensure that the region of the 16p11.2 distal CNV is densely covered on each array type used in the studies, and therefore it is unlikely that any CNVs in this region could have been completely missed by some of the studies. The positions are shown in eFigure 3 and demonstrate a dense coverage of the region on all arrays.
eMaterial 3. Estimating the selection pressure against the distal 16p11.2 deletion

We were able to estimate the selection pressure (s) operating against deletions at distal 16p11.2 using the methods described in our previous publication.24 Briefly, assuming a mutation-selection balance model, s should equal the deletion de novo rate [de novo/(inherited + de novo)] for this locus. We estimate the de novo rate, and therefore s, for deletions at distal 16p11.2 to be 0.25 (95% CI = 0.06-0.44), based on papers that have reported in a systematic way the rate of de novo and inherited CNVs in this region (eTable 3). This estimation is based on the observation of 5 de novos out of 20 systematically ascertained carriers from 5 available datasets, where inheritance had been examined (current study and refs 19-22). An s of 0.25 falls within the selection pressure range (0.12-0.88) of the other known schizophrenia-associated CNVs.24
eMaterial 4. Members of the Molecular Genetics of Schizophrenia Consortium

Douglas F. Levinson, MD1, Jubao Duan, PhD2, Sang Oh1, Kai Wang, PhD3, Alan R. Sanders, MD2, Jianxin Shi, PhD4, Nancy Zhang, PhD1, Bryan J. Mowry, MD, FRANZCP5, Ann Olincy, MD6, Farooq Amin, MD7, C. Robert Cloninger, MD8, Jeremy M. Silverman, MD9, Nancy G. Buccola, APRN, BC10, William F. Byerley, MD11, Donald W. Black, MD12, Kenneth S. Kendler, MD13,14, Robert Freedman, MD6, Frank Dudbridge, PhD15, Itsik Pe’er, PhD16, Hakon Hakonarson, MD, PhD17, Sarah E. Bergen, PhD18, Ayman H. Fanous, MD19, Peter A. Holmans, PhD20, and Pablo V. Gejman, MD2

1Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California;

2NorthShore University HealthSystem Research Institute, Evanston, Illinois;

3Departments of Psychiatry and Preventive Medicine, University of Southern California, Los Angeles;

4National Cancer Institute, Bethesda, Maryland;

5Queensland Centre for Mental Health Research and Queensland Institute for Medical Research, Brisbane, Queensland, Australia;

6Department of Psychiatry, University of Colorado Denver, Aurora;

7Department of Psychiatry and Behavioral Sciences, Atlanta Veterans Affairs Medical Center, and Emory University, Atlanta, Georgia;

8Department of Psychiatry, Washington University, St. Louis, Missouri;

9Department of Psychiatry, Mount Sinai School of Medicine, New York, New York;

10School of Nursing, Louisiana State University Health Sciences Center, New Orleans;

11Departments of Psychiatry, and Neurology, University of California at San Francisco;

© 2013 American Medical Association. All rights reserved.
12Mental Health Clinical Research Center and Department of Psychiatry, University of Iowa Carver College of Medicine, Iowa City;

13Departments of Psychiatry and

14Human Genetics, Virginia Commonwealth University, Richmond;

15Department of Statistical Genetics and Epidemiology, London School of Hygiene & Tropical Medicine, London, England;

16Department of Computer Science, Columbia University, New York, New York;

17Children’s Hospital of Philadelphia and University of Pennsylvania, Philadelphia;

18Psychiatric & Neurodevelopmental Genetics Unit, Center for Human Genetic Research, Massachusetts General Hospital Boston, and Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts;

19Washington VA Medical Center and Department of Psychiatry, Georgetown University School of Medicine, Washington, DC; and

20Biostatistics and Bioinformatics Unit, Cardiff University, Cardiff, Wales.
eMaterial 5. Members of the Wellcome Trust Case Control Consortium 2

Management Committee

Peter Donnelly (Chair)\textsuperscript{1,2}, Ines Barroso (Deputy Chair)\textsuperscript{3}, Jenefer M. Blackwell\textsuperscript{4,5}, Elvira Bramon\textsuperscript{6}, Matthew A. Brown\textsuperscript{7}, Juan P. Casas\textsuperscript{8}, Aiden Corvin\textsuperscript{9}, Panos Deloukas\textsuperscript{3}, Audrey Duncanson\textsuperscript{10}, Janusz Jankowski\textsuperscript{11}, Hugh S. Markus\textsuperscript{12}, Christopher G. Mathew\textsuperscript{13}, Colin N. A. Palmer\textsuperscript{14}, Robert Plomin\textsuperscript{15}, Anna Rautanen\textsuperscript{1}, Stephen J Sawcer\textsuperscript{16}, Richard C Trembath\textsuperscript{13}, Ananth C. Viswanathan\textsuperscript{17}, Nicholas W. Wood\textsuperscript{18}

Data and Analysis Group

Chris C. A. Spencer\textsuperscript{1}, Gavin Band\textsuperscript{1}, Céline Bellenguez\textsuperscript{1}, Colin Freeman\textsuperscript{1}, Garrett Hellenthal\textsuperscript{1}, Eleni Giannoulatou\textsuperscript{1}, Matti Pirinen\textsuperscript{1}, Richard Pearson\textsuperscript{1}, Amy Strange\textsuperscript{1}, Zhan Su\textsuperscript{1}, Damjan Vukcevic\textsuperscript{1}, Peter Donnelly\textsuperscript{1,2}

DNA, Genotyping, Data QC and Informatics Group

Cordelia Langford\textsuperscript{3}, Sarah E. Hunt\textsuperscript{3}, Sarah Edkins\textsuperscript{3}, Rhian Gwilliam\textsuperscript{3}, Hannah Blackburn\textsuperscript{3}, Suzannah J. Bumpstead\textsuperscript{3}, Serge Dronov\textsuperscript{3}, Matthew Gillman\textsuperscript{3}, Emma Gray\textsuperscript{3}, Naomi Hammond\textsuperscript{3}, Alagurevathi Jayakumar\textsuperscript{3}, Owen T. McCann\textsuperscript{3}, Jennifer Liddle\textsuperscript{3}, Simon C. Potter\textsuperscript{3}, Radhi Ravindrarajah\textsuperscript{3}, Michelle Ricketts\textsuperscript{3}, Matthew Waller\textsuperscript{3}, Paul Weston\textsuperscript{3}, Sara Widaa\textsuperscript{3}, Pamela Whittaker\textsuperscript{3}, Ines Barroso\textsuperscript{3}, Panos Deloukas\textsuperscript{3}.

Publications Committee

Christopher G. Mathew (Chair)\textsuperscript{13}, Jenefer M. Blackwell\textsuperscript{4,5}, Matthew A. Brown\textsuperscript{7}, Aiden Corvin\textsuperscript{9}, Chris C. A. Spencer\textsuperscript{1}

\textsuperscript{1} Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, United Kingdom;

\textsuperscript{2} Department of Statistics, University of Oxford, Oxford OX1 3TG, United Kingdom;
3 Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, United Kingdom;

4 Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia, 100 Roberts Road, Subiaco, Western Australia 6008;

5 Cambridge Institute for Medical Research, University of Cambridge School of Clinical Medicine, Cambridge CB2 0XY, United Kingdom;

6 Department of Psychosis Studies, National Institute for Health Research Biomedical Research Centre for Mental Health at the Institute of Psychiatry, King’s College London and The South London and Maudsley National Health Service Foundation Trust, Denmark Hill, London SE5 8AF, United Kingdom;

7 University of Queensland Diamantina Institute, Brisbane, Queensland, Australia;

8 Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London WC1E 7HT and Dept Epidemiology and Public Health, University College London WC1E 6BT, United Kingdom;

9 Neuropsychiatric Genetics Research Group, Institute of Molecular Medicine, Trinity College Dublin, Dublin 2, Eire;

10 Molecular and Physiological Sciences, The Wellcome Trust, London NW1 2BE;

11 Department of Oncology, Old Road Campus, University of Oxford, Oxford OX3 7DQ, United Kingdom, Digestive Diseases Centre, Leicester Royal Infirmary, Leicester LE7 7HH, United Kingdom, and Centre for Digestive Diseases, Queen Mary University of London, London E1 2AD, United Kingdom;

12 Clinical Neurosciences, St George's University of London, London SW17 0RE;

13 King’s College London Department of Medical and Molecular Genetics, King’s Health Partners, Guy’s Hospital, London SE1 9RT, United Kingdom;
14 Biomedical Research Centre, Ninewells Hospital and Medical School, Dundee DD1 9SY, United Kingdom;

15 King’s College London Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Denmark Hill, London SE5 8AF, United Kingdom;

16 University of Cambridge Department of Clinical Neurosciences, Addenbrooke’s Hospital, Cambridge CB2 0QQ, United Kingdom;

17 National Institute for Health Research Biomedical Research Centre for Ophthalmology, Moorfields Eye Hospital National Health Service Foundation Trust and University College London Institute of Ophthalmology, London EC1V 2PD, United Kingdom;

18 Department of Molecular Neurosciences, Institute of Neurology, Queen Square, London WC1N 3BG, United Kingdom.
**eFigure 1.** z-Score histogram for distal 16p11.2 (chromosome 16:28615687-28951365) in the Bulgarian trio data. Callout boxes indicate 2 transmitted deletion copy number variations, where the first 4 digits represent family ID and the last digit represents individual ID (1 = proband; 3 = mother).
**eFigure 2.** Principal component analysis plot of the Ashkenazi discovery cohort using 762,372 genome-wide autosomal single-nucleotide polymorphisms. Green dots represent case and blue dots represent control. As can be observed in the plot, there is minimal residual population structure in this cohort, and no evidence of batch effects.
**eFigure 3.** 16p11.2 region, including the known schizophrenia-associated locus (proximal) and the new locus implicated in the current study (distal). Shown are the positions of probes on the various arrays used in the current study.
### eTable 1. Previous observations of distal 16p11.2 deletions

Further phenotypes tested for the 16p11.2 distal deletion. DD = developmental delay, ADHD = attention-deficit/hyperactivity disorder, SZ = schizophrenia, MCA = multiple congenital anomalies, MR = mental retardation. Note that the controls used in these studies are not independent from ours: Thus the Bochukova et al, 2010\(^{19}\) study used controls identical to those in the current study (Wellcome Trust Case Control Consortium 2 [WTCCC2] and Molecular Genetics of Schizophrenia [MGS]\(^3\)), the study by Bachmann-Gagescu et al, 2010\(^{21}\) used 3181 partially overlapping controls to those used in the current study (the International Schizophrenia Consortium ISC sample\(^6\)), and the study by Cooper et al, 2011\(^ {23}\) used some of the samples from WTCCC2. It is possible that all deletions in controls in these 3 studies are the same as those reported in the current study (from the WTCCC2 sample, where we report 2 deletions in the current study).

<table>
<thead>
<tr>
<th>Study</th>
<th>Phenotype</th>
<th>Platform</th>
<th>Case</th>
<th>Control</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bochukova et al, 2010(^ {19})</td>
<td>Obesity/DD</td>
<td>Affy 6.0</td>
<td>300 (3)</td>
<td>&gt;7000 (2), (WTCCC2 + MGS)</td>
<td>143 cases had DD</td>
</tr>
<tr>
<td>Barge-Schaapveld et al, 2011(^ {20})</td>
<td>DD/unusual facial morphology/obesity</td>
<td>aCGH Agilent.</td>
<td>Family 1 Two affected bothers carriers</td>
<td>Family 1 –Father died early of paraproteinema, very likely to have passed on deletion to both sons as mother does not carry the deletion.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachmann-Gagescu et al, 2010(^ {21})</td>
<td>Range of abnormal phenotypes, most commonly DD.</td>
<td>Multiple array platforms.</td>
<td>23 084 (31)</td>
<td>7700 (1). Controls come from 3 large published data sets.</td>
<td>Some patients with deletions have autism, ADHD, SZ, and Bipolar disorder</td>
</tr>
<tr>
<td>Bijlsma et al, 2009(^ {22})</td>
<td>MR/MCA</td>
<td>Various array platforms (distal deletion detected on Agilent 44k).</td>
<td>4284 (1)</td>
<td></td>
<td>Child with dismorphisms and hypotonia. Father transmitting the deletion had learning difficulties.</td>
</tr>
<tr>
<td>Cooper et al, 2011(^ {23})</td>
<td>DD</td>
<td>Multiple array</td>
<td>15 767 (15)</td>
<td>8329 (1)</td>
<td></td>
</tr>
</tbody>
</table>
**eTable 2. Phenotypic data on carriers of 16p11.2 distal deletions.** Phenotypic data on carriers of 16p11.2 distal deletions, identified in the current study. Sz Undiff = Undifferentiated type of schizophrenia; Sz Par = paranoid type of schizophrenia; SzAff = schizoaffective disorder, Bip = bipolar type, dep = depressive type, sib = sibling, AAO = age at onset; AJ = Ashkenazi Jewish, BG = Bulgarian, EA = European American, MGS = Molecular Genetics of Schizophrenia.

<table>
<thead>
<tr>
<th>ID</th>
<th>Cohort</th>
<th>Sex</th>
<th>Ethnicity</th>
<th>Family History</th>
<th>Diagnosis</th>
<th>AAO, y</th>
<th>Course/clinical features</th>
</tr>
</thead>
<tbody>
<tr>
<td>82840</td>
<td>MGS</td>
<td>F</td>
<td>EA</td>
<td>grandparent SZ (not known if carrier)</td>
<td>Sz Undiff</td>
<td>24</td>
<td>Continuous</td>
</tr>
<tr>
<td>83897</td>
<td>MGS</td>
<td>M</td>
<td>EA</td>
<td>1 sib SZ (not known if carrier)</td>
<td>Sz Undiff</td>
<td>30</td>
<td>Episodic</td>
</tr>
<tr>
<td>72800</td>
<td>MGS</td>
<td>M</td>
<td>EA</td>
<td>1 sib SZ (not known if carrier)</td>
<td>Sz Undiff</td>
<td>15</td>
<td>Episodic (mild deterioration)</td>
</tr>
<tr>
<td>84406</td>
<td>MGS</td>
<td>F</td>
<td>EA</td>
<td>-</td>
<td>SzAff, Dep</td>
<td>16</td>
<td>Continuous; depressed ~50% of the time</td>
</tr>
<tr>
<td>97911</td>
<td>MGS</td>
<td>F</td>
<td>EA</td>
<td>-</td>
<td>Sz Undiff</td>
<td>15</td>
<td>Continuous</td>
</tr>
<tr>
<td>SZPBYK0003</td>
<td>AJ</td>
<td>F</td>
<td>AJ</td>
<td>-</td>
<td>Sz Par</td>
<td>30</td>
<td>continuous no medical problems noted</td>
</tr>
<tr>
<td>SZAABR0015</td>
<td>AJ</td>
<td>F</td>
<td>AJ</td>
<td>1 sib SZ (not known if carrier)</td>
<td>SzAff, Bip</td>
<td>25</td>
<td>Course data not available/ T2 diabetes, hypertension</td>
</tr>
<tr>
<td>P1336</td>
<td>Japan</td>
<td>M</td>
<td>Japanese</td>
<td>1 sib SZ (carrier)</td>
<td>Sz Par</td>
<td>26</td>
<td>2 episodes BMI=22.4</td>
</tr>
<tr>
<td>2220-1</td>
<td>BG trios</td>
<td>F</td>
<td>BG</td>
<td>Mother depressed (carrier)</td>
<td>Sz Catatonic</td>
<td>28</td>
<td>Acute onset, 1st episode, voices, paranoid, stupor, ideas of guilt</td>
</tr>
<tr>
<td>4290-1</td>
<td>BG trios</td>
<td>M</td>
<td>BG</td>
<td>Father anxiety (not carrier)</td>
<td>SZ Par</td>
<td>20</td>
<td>Continuous, voices, paranoid, self-neglect, “Overweight”</td>
</tr>
<tr>
<td>sbn-128-01</td>
<td>Germany</td>
<td>F</td>
<td>German</td>
<td>Father major depressive disorder (not known if carrier)</td>
<td>SzAff, Bip</td>
<td>23</td>
<td>No medical problems noted</td>
</tr>
<tr>
<td>Case-1</td>
<td>Sweden</td>
<td>F</td>
<td>Swedish</td>
<td>-</td>
<td>Sz Undiff</td>
<td>28</td>
<td>T2 diabetes, recurrent depression</td>
</tr>
<tr>
<td>Case-2</td>
<td>Sweden</td>
<td>F</td>
<td>Swedish</td>
<td>-</td>
<td>SzAff</td>
<td>24</td>
<td>Hypothyroidism, obesity</td>
</tr>
</tbody>
</table>

© 2013 American Medical Association. All rights reserved.
eTable 3. Publications used to estimate the selection pressure(s) operating against deletions at distal 16p11.2, showing the numbers of reported \textit{de novo} and inherited observations.

<table>
<thead>
<tr>
<th>Study</th>
<th>De novo/ (inherited + de novo)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bochukova et al\textsuperscript{19}</td>
<td>0/2</td>
</tr>
<tr>
<td>Bachmann-Gagescu et al\textsuperscript{21}</td>
<td>5/13</td>
</tr>
<tr>
<td>Current study</td>
<td>0/2</td>
</tr>
<tr>
<td>Barge-Schaapveld et al\textsuperscript{20}</td>
<td>0/2</td>
</tr>
<tr>
<td>Bijlsma et al\textsuperscript{22}</td>
<td>0/1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>5/20</strong></td>
</tr>
</tbody>
</table>
eTable 4. CNVs in the Ashkenazi Jewish (AJ) and Bulgarian (BG) cohorts at loci previously implicated in schizophrenia. The list of loci is based on the reviews by Kirov, 2010\textsuperscript{25} and Malhotra and Sebat, 2012\textsuperscript{26} and represents only loci which we regard as having received strong statistical support, including replication evidence. CNV= copy number variant.

<table>
<thead>
<tr>
<th>CNV Locus</th>
<th>Position (Mb)</th>
<th>AJ cohort</th>
<th>BG cohort</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Case</td>
<td>Control</td>
<td>Transmitted (de novo)</td>
<td>Not transmitted</td>
</tr>
<tr>
<td>1q21.1 deletion</td>
<td>145.0-146.3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2p16.3 deletion (NRXN1)</td>
<td>50.0-51.1</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>3q29 deletion</td>
<td>197.2-198.8</td>
<td>1</td>
<td>0</td>
<td>1 de novo</td>
<td>0</td>
</tr>
<tr>
<td>7q11.23 duplication (WBS region)</td>
<td>72.4-73.8</td>
<td>0</td>
<td>0</td>
<td>1 de novo</td>
<td>0</td>
</tr>
<tr>
<td>7q36.3 duplication (VIPR2)</td>
<td>158.4-158.8</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>15q11.2 deletion</td>
<td>20.3-20.6</td>
<td>0</td>
<td>0</td>
<td>+ 2 de novo</td>
<td>2</td>
</tr>
<tr>
<td>15q11-13 duplication including PWS region</td>
<td>22.4-26.1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15q13.3 deletion</td>
<td>28.7-30.2</td>
<td>0</td>
<td>0</td>
<td>2 de novo</td>
<td>0</td>
</tr>
<tr>
<td>16p13.11 duplication</td>
<td>15.4-16.2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>16p11.2 duplication</td>
<td>29.5-30.1</td>
<td>1</td>
<td>0</td>
<td>+ 1 de novo</td>
<td>0</td>
</tr>
<tr>
<td>17p12/ deletion (HNPP)</td>
<td>14.0-15.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17q12 deletion</td>
<td>31.9-33.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>22q11.21 deletion</td>
<td>17.4-18.7</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>


13. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81: 559-575.


