Microduplications of 16p11.2 are associated with schizophrenia

Shane E McCarthy[*,1], Vladimir Makarov[1], George Kirov[2], Anjene M Addington[3], Jon McClellan[4], Seungtai Yoon[1], Diana O Perkins[5], Diane E Dickel[6], Mary Kusenda[1,7], Olga Krastoshevsky[8], Verena Krause[8], Ravinesh A Kumar[9], Detelina Grozeva[2], Dheeraj Malhotra[1], Tom Walsh[6], Elaine H Zackai[10], Paige Kaplan[11], Jaya Ganesh[11], Ian D Krantz[10], Nancy B Spinner[10], Patricia Roccanova[1], Abhishek Bhandari[1], Kevin Pavon[1], B Lakshmi[1,12], Anthony Leotta[1], Jude Kendall[1], Yoon-ha Lee[1], Vladimir Vacic[1], Sydney Gary[1], Lilia M Iakoucheva[13], Timothy J Crow[14], Susan L Christian[9], Jeffrey A Lieberman[15,16], T Scott Stroup[15], Terho Lehtimäki[17], Kaija Puura[18], Chad Haldeman-Englert[10], Justin Pearl[19], Meredith Goodell[20], Virginia L Willour[20], Pamela DeRosse[21], Jo Steele[19], Layla Kassem[19], Jessica Wolff[19], Nisha Chitkara[21], Francis J McMahon[19], Anil K Malhotra[21], James B Potash[20], Thomas G Schulze[19,22], Markus M Nöthen[23,24], Sven Cichon[23,24], Marcella Rietschel[22,25], Ellen Leibenluft[26], Vlad Kustanovich[27], Clara M Lajonchere[27], James S Sutcliffe[28], David Skuse[29], Michael Gill[30], Louise Gallagher[30], Nancy R Mendell[31], Wellcome Trust Case Control Consortium[32] ,Nick Craddock[2], Michael J Owen[2], Michael C O'Donovan[2], Tamim H Shaikh[10], Ezra Susser[15], Lynn E DeLisi[33,34], Patrick F Sullivan[35], Curtis K Deutsch[33,36], Judith Rapoport[3], Deborah L Levy[8,33], Mary-Claire King[6] & Jonathan Sebat[1]

Recurrent microdeletions and microduplications of a 600-kb genomic region of chromosome 16p11.2 have been implicated in childhood-onset developmental disorders[1–3]. We report the association of 16p11.2 microduplications with schizophrenia in two large cohorts. The microduplication was detected in 12/1,906 (0.63%) cases and 1/3,971 (0.03%) controls ($P = 1.2 \times 10^{-5}$, OR = 25.8) from the initial cohort, and in 9/2,645 (0.34%) cases and 1/2,420 (0.04%) controls ($P = 0.022$, OR = 8.3) of the replication cohort. The 16p11.2 microduplication was associated with a 14.5-fold increased risk of schizophrenia (95% CI (3.3, 62)) in the combined sample. A meta-analysis of datasets for multiple psychiatric disorders showed a significant association of the microduplication with schizophrenia ($P = 4.8 \times 10^{-7}$), bipolar disorder ($P = 0.017$) and autism ($P = 1.9 \times 10^{-7}$). In contrast, the reciprocal microdeletion was associated only with autism and developmental disorders ($P = 2.3 \times 10^{-13}$). Head circumference was larger in patients with the microdeletion than in patients with the microduplication ($P = 0.0007$).

Rare structural mutations play an important role in schizophrenia. Recent studies have shown that the genome-wide burden of rare copy number variants (CNVs) is significantly greater in individuals with schizophrenia than in healthy controls[4–6]. In addition, multiple structural variants have been implicated in schizophrenia. Seminal examples include the recurrent microdeletion of 22q11.2 (ref. 7) and a balanced translocation disrupting the gene *DISC1* (ref. 8). More recently, recurrent microdeletions at 1q21.1, 15q13.3 (refs. 5,9) and 15q11.2 (refs. 6,9) and copy number mutations at other genomic loci[10–12] have been associated with schizophrenia in large cohorts.

We previously reported two individuals with childhood-onset schizophrenia who carry a 600-kb microduplication of 16p11.2 (ref. 4). This region is a well-documented hot spot for recurrent rearrangements that are associated with autism-spectrum disorders and mental retardation[1–3,13,14]. Genomic hot spots such as this are important candidate loci in genetic studies of schizophrenia.

We tested the hypothesis that microduplications of 16p11.2 are associated with schizophrenia by analysis of microarray intensity data in an initial cohort that included 1,906 affected individuals (cases) and 3,971 controls from several different sources. Sample collection is described in the **Supplementary Note** and **Supplementary Table 1**. Samples were analyzed with one of four microarray platforms (NimbleGen HD2, Affymetrix 6.0, Affymetrix 500K and ROMA 85K). Only the 16p11.2 region was examined. Thirteen microduplications and four microdeletions were detected in our primary sample using standard segmentation algorithms (**Fig. 1a**, **Supplementary Fig. 1a**). Microduplications at 16p11.2 were detected in 12/1,906 cases (0.63%) and 1/3,971 controls (0.03%), demonstrating a statistically significant association (**Table 1**, $P = 1.2 \times 10^{-5}$, OR = 25.8 (3.3, 199)).

In a subset of individuals evaluated at Cold Spring Harbor Laboratory (CSHL, Cold Spring Harbor, NY) consisting of 1,352 cases and 1,179 controls, CNV calls were verified by median *z*-score outlier

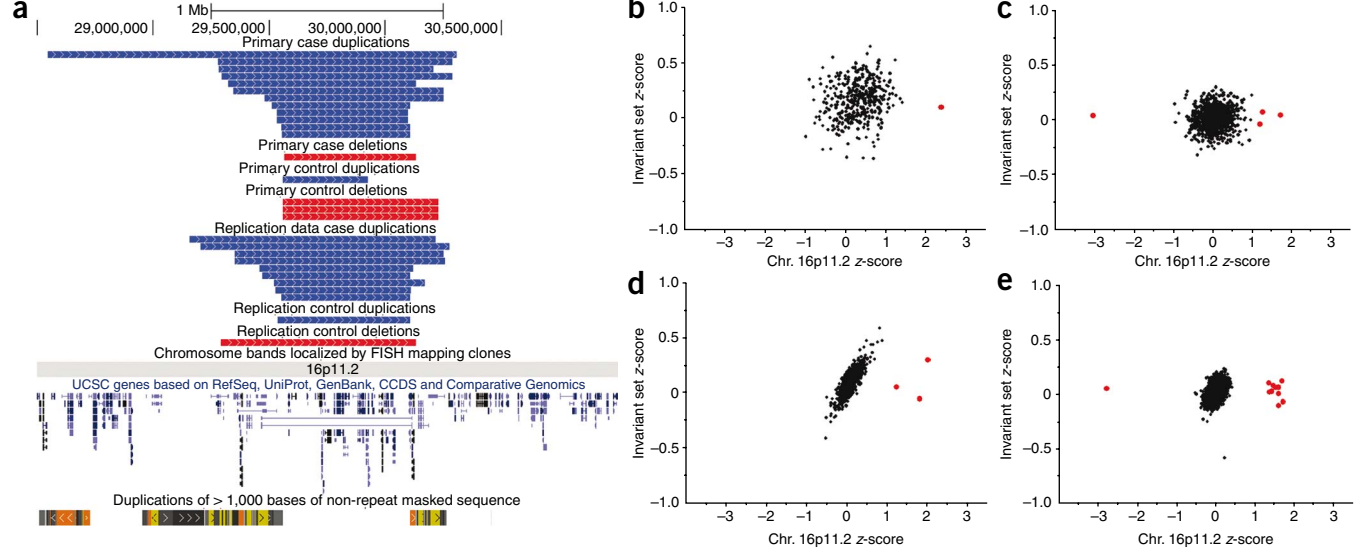

Figure 1 Microduplications and microdeletions at 16p11.2 in persons with schizophrenia and controls. (**a**–**e**) 16p11.2 rearrangements were detected in a primary sample of 1,906 cases and 3,971 controls (**a**–**d**) and a replication sample of 2,645 cases and 2,420 controls (**a**,**e**). The single microduplication and three microdeletions detected in the primary control set are presented based on the Affymetrix 500K coordinates (hg18). All other CNVs were validated in the NimbleGen HD2 platform and are illustrated based on the validation coordinates. In **b**–**e**, the median *z*-score for the 535-kb 16p11.2 target region is plotted on the *x* axis and the median *z*-score of flanking invariant probes is plotted on the *y* axis. Data are presented separately for the ROMA (**b**), Affymetrix500K (**c**), NimbleGen HD2 (**d**) and Affymetrix 6.0 (**e**) platforms. CNVs were called using thresholds of >2 s.d. for ROMA and >1 s.d. for all other platforms (red dots). MeZOD and the HMM algorithms detected the same deletions and duplications at 16p11.2.

detection (MeZOD), an independent CNV genotyping algorithm that identifies outliers in the sample based on the median probe *z*-score of the target region. These results are illustrated as cluster plots (**Fig. 1**). All microduplications and microdeletions detected in the combined sample were experimentally validated using an independent microarray platform (**Supplementary Table 2**).

To replicate this association, we evaluated the 16p11.2 region using microarray data (Affymetrix 6.0 platform) from an independent sample of 2,645 individuals with schizophrenia and 2,420 controls. These data were collected as part of a case-control study of schizophrenia supported by the Genetic Association Information Network (GAIN, phs000021.v2.p1). We detected ten duplications and one deletion using standard hidden Markov model (HMM) calling algorithms (**Fig. 1a**). The same events were also detected using MeZOD (**Fig. 1e**). All 16p11.2 rearrangements were validated by an independent microarray platform (**Supplementary Table 2**). The microduplication was detected in 9/2,645 cases and 1/2,420 controls, demonstrating a significant association ($P = 0.022$, OR = 8.3 (1.3, 50.5)).

The odds ratios in our primary and replication datasets were not significantly different (Breslow-Day-Tarone test $P = 0.46$). Thus, our initial result was replicated in an independent sample. For the combined sample, the association of schizophrenia with microduplication at 16p11.2 was highly significant ($P = 4.3 \times 10^{-7}$, OR = 14.5 (3.3,62)). Sex of the subject did not have a significant effect on the association (**Supplementary Note**).

Our present findings and those from previous studies[1–3,13,14] suggest that mutations at 16p11.2 confer high risk for schizophrenia and for other neuropsychiatric disorders. Clinical variability associated with the 16p11.2 microduplication is evident from the heterogeneity of psychiatric diagnoses among microduplication carriers in five families in our series (**Supplementary Fig. 2**). In these families, ten relatives carried the microduplication found in the proband. The diagnoses of these relatives were schizophrenia ($n = 3$), bipolar disorder ($n = 1$), depression ($n = 2$), psychosis signs not otherwise specified ($n = 1$) and no mental illness ($n = 3$). We were able to determine the parent of origin for the microduplication in four families, and in all cases the microduplications were inherited from a non-schizophrenic parent. The observations in these few families suggest that penetrance of the duplication is incomplete, though substantial (perhaps 30–50%), and that expression is highly variable.

In order to more precisely define the spectrum of psychiatric phenotypes associated with rearrangements of 16p11.2, we performed a meta-analysis of data on

Table 1 Duplications and deletions at 16p11.2 among persons with schizophrenia and controls

		Subjects	Deletions		Duplications			
Series	Diagnosis	*n*	*n*	%	*n*	%	OR (95% CI)	*P* value
Primary	Schizophrenia	1,906	1	0.05	12	0.63	25.8 (3.3,199)	1.2×10^{-5}
	Controls	3,971	3	0.08	1	0.03		
Replication	Schizophrenia	2,645	0	0.00	9	0.34	8.3 (1.3, 50.5)	0.022
	Controls	2,420	1	0.04	1	0.04		
Combined	Schizophrenia	4,551	1	0.02	21	0.46	14.5 (3.3, 62.0)	4.3×10^{-5}
	Controls	6,391	4	0.06	2	0.03		

In the primary sample, which consisted of patients and controls genotyped using one of three microarray platforms, association was calculated using the Cochran-Mantel-Haenszel exact test using array type as a stratifying variable. Combined odds ratio estimates and 95% confidence intervals were calculated using a logistic regression with disease group and array-type as factors. In the replication sample, which consisted of affected individuals and controls assessed on a single microarray platform, association was calculated using a Fisher's exact test. Deletions did not show a significant association with schizophrenia or in controls.

Table 2 Meta-analysis of 16p11.2 rearrangements in schizophrenia, autism and developmental delay, and bipolar disorder

	Subjects	Deletions				Duplications			
Diagnosis	*n*	*n*	%	OR (95% CI)	*P* value	*n*	%	OR (95% CI)	*P* value
Schizophrenia	8,590	3	0.03	NC[a]		26	0.30	8.4 (2.8, 25.4)	4.8×10^{-7}
Controls	28,406	9	0.03			8	0.03		
Autism or developmental delay	2,172	17	0.78	38.7 (13.4,111.8)	2.3×10^{-13}	10	0.46	20.7 (6.9,61.7)	1.9×10^{-7}
Controls	24,891	5	0.02			6	0.02		
Bipolar disorder	4,822	4	0.08	NC[a]		6	0.12	4.3 (1.3, 14.5)	0.017
Controls	25,225	6	0.02			7	0.03		

Data from four studies reporting microduplications and microdeletions of 16p11.2 in individuals with schizophrenia, autism and/or bipolar disorder were combined with data from the primary sample to assess the relative strength of the association of each variant with each disorder. Associations were calculated using the Cochran-Mantel-Haenszel exact test, using source as a stratifying variable. Combined odds ratio estimates and confidence intervals were calculated using logistic regression with disease group and source (study) as factors.

[a]Not calculated (NC) because significant heterogeneity among studies was detected by the Breslow-Day-Tarone test. The partial odds ratios (95%CI) for the deletion in schizophrenia were 0.69 (0.1, 4.9), 0.3 (0.05, 2.2), 14.6 (1.9, 111.2) and 0.3 (0.03, 3.7), and those for the deletion in bipolar disorder were 0.3(0.03,3.3), 0.55(0.05,6.7) and 25(5.4,117) in this study, the GAIN study and the Weiss *et al*. studies, respectively.

schizophrenia, bipolar disorder and childhood developmental disorders (combining autism and global developmental delays). We integrated data from this study with four publicly available datasets[1,3,5,15] to generate a combined sample of 8,590 individuals with schizophrenia, 2,172 with developmental delay or autism, 4,822 with bipolar disorder and 30,492 controls (**Supplementary Note**, **Supplementary Table 3**). In this combined sample, the microduplication of 16p11.2 was strongly associated with schizophrenia (**Table 2**, OR = 8.4 (2.8, 25.4), $P = 4.8 \times 10^{-7}$) and autism (OR = 20.7 (6.9, 61.7), $P = 1.9 \times 10^{-7}$). The association with bipolar disorder was also significant (OR = 4.3 (1.3, 14.5), $P = 0.017$). The reciprocal microdeletion of 16p11.2 was strongly associated with developmental delay or autism (OR = 38.7 (13.4, 111.8), $P = 2.3 \times 10^{-13}$), as reported previously[1–3]. However, the deletion was not associated with schizophrenia or bipolar disorder (**Supplementary Note**). These results suggest that the microduplication is associated with multiple psychiatric phenotypes, whereas the reciprocal microdeletion is more specifically associated with developmental delay and autism.

We explored the association of 16p11.2 microduplications and microdeletions with two clinical measures: head circumference and height. Available data were compiled from 32 patients with 16p11.2 mutations who had a diagnosis of schizophrenia, autism-spectrum disorder or developmental delay (**Supplementary Note**, **Supplementary Table 4** and refs. 13,16). *Z* scores for head circumference and height were calculated using standard growth charts from the Centers for Disease Control. Head circumference was greater among 23 individuals with microdeletions than among 9 individuals with microduplications (**Supplementary Table 5**). The mean orbital frontal circumference (OFC) values of patients with microdeletions and microduplications were 1.25 and −0.28, respectively (two-tailed Wilcoxon rank sum test $P = 0.0007$). In addition, mean head circumference of the microdeletion group was significantly greater than the population mean ($P = 0.0001$), whereas the mean head circumference of the microduplication group was not significantly different from the population mean ($P = 0.29$). The association between the 16p11.2 microdeletion and larger head circumference was observed in multiple diagnostic categories and was not specifically attributable to patients with autism (**Supplementary Table 5**). The microduplication and microdeletion groups did not significantly differ from each other in height.

We report here that microduplication of 16p11.2 is associated with an 8–24-fold increased risk of schizophrenia. This region joins a growing list of genomic hot spots that confer high risk for the disorder. The odds ratios in our series for the 16p11.2 microduplication and schizophrenia are comparable to odds ratios for deletions at other schizophrenia-associated genes and regions. Deletions of 1q21.1, 15q13.3 and *NRXN1* have reported odds ratios ranging from 7 to 18 (refs. 5,9,10).

Previous genome-wide studies of copy number variation did not find a significant association with the microduplication of 16p11.2 and schizophrenia. This microduplication event is rare, and its detection in a cohort may be influenced by several factors, including resolution of the platform, methods of analysis and chance. In the International Schizophrenia Consortium (ISC) study[5], microduplications spanning >50% of the 16p11.2 region were detected in 5/3,391 cases and 1/3,181 controls. These results are consistent with our findings, but the association did not meet the criteria for genome-wide significance in that study. In the Schizophrenia Gene (SGENE) consortium study of schizophrenia[9], the 16p11.2 microduplication was not selected as a candidate because the event was not observed in the initial phase of that study as a *de novo* mutation, which was the key criterion for inclusion in the association analyses.

Microduplication at 16p11.2 is associated with multiple neuropsychiatric phenotypes. Phenotypic heterogeneity has been observed for virtually all structural variants associated with schizophrenia. For example, in a large Scottish pedigree harboring a translocation disrupting *DISC1*, translocation carriers had diagnoses of schizophrenia, bipolar disorder, major depressive disorder or no mental illness[8]. Similarly, microdeletions of 1q21.1 (refs. 17,18), 15q13.3 (ref. 19), 22q11.2 (ref. 20) and *NRXN1* (refs. 10,12,21,22) are associated with adult psychiatric disorders and with autism and other pediatric neurodevelopmental disorders.

The association between the 16p11.2 microdeletion and increased head circumference is notable given that the microdeletion appears to be specific to autism and developmental delay. Several studies have found increased head circumference in patients with autism[23–30], leading to the suggestion that early brain overgrowth may be a key neurobiological mechanism in the disorder[31]. A recent study has shown that microdeletions and microduplications of 1q21.1 are associated with microcephaly and macrocephaly, respectively[18]. Taken together, these studies suggest that some mutations underlying neurodevelopmental disorders may also lead to changes in brain volume.

The 16p11.2 microduplication spans a region of approximately 600 kb containing 28 genes (**Supplementary Fig. 1b**), including numerous genes with potential roles in neurodevelopment. At least 17 of the 28 genes in this region are expressed in the mammalian brain (**Supplementary Table 6**). Behavioral features have been reported in mouse *Mapk3*$^{-/-}$ *Doc2a*$^{-/-}$ and *Sez6l2*$^{-/-}$ knockout models[32–34].

Further studies are needed in order to identify the specific gene of genes in this region for which dosage effects contribute to increased risk for psychiatric and neurodevelopmental disorders.

Our findings further strengthen the evidence demonstrating a role for rare mutations in schizophrenia[4–6,9]. Collectively, these studies demonstrate that schizophrenia is characterized by marked genetic heterogeneity. Although the 16p11.2 locus by itself may account for only a small proportion of cases, the duplication of this region confers substantial risk to the individuals who carry it. In addition, although this single mutation is rare, the collective effect of rare mutations at many different loci may account for a substantial proportion of schizophrenia-affected individuals[4,5] and will likely influence overlapping neurobiological pathways. Characterizing these pathways will contribute substantially to our understanding of the origins of schizophrenia and suggest targets for treatment development.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Accession code. dbGaP (genotype data): GAIN study of schizophrenia (phs000021.v2.p1) and GAIN study of bipolar disorder (phs000017.v2.p1).

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

Schizophrenia: Funding support for the genome-wide association of schizophrenia study was provided by the National Institute of Mental Health (NIMH) (R01 MH67257, R01 MH59588, R01 MH59571, R01 MH59565, R01 MH59587, R01 MH60870, R01 MH59566, R01 MH59586, R01 MH61675, R01 MH60879, R01 MH81800, U01 MH46276, U01 MH46289 U01 MH46318, U01 MH79469 and U01 MH79470), and the genotyping of samples was provided through the genetic association information network (GAIN). The datasets used for the analyses described in this manuscript were obtained from the database of genotype and phenotype (dbGaP) found at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession number phs000021.v2.p1. Samples and associated phenotype data for the genome-wide association of schizophrenia study were provided by the Molecular Genetics of Schizophrenia Collaboration (PI: P.V. Gejman, Evanston Northwestern Healthcare (ENH) and Northwestern University, Evanston, Illinois, USA).

Bipolar Disorder: Funding support for the whole-genome association study of bipolar disorder was provided by the NIMH and the genotyping of samples was provided through GAIN. The datasets used for the analyses described in this manuscript were obtained from dbGaP found at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession number phs000017.v2.p1. Samples and associated phenotype data for the collaborative genomic study of bipolar disorder were provided by the NIMH genetics initiative for bipolar disorder. Data and biomaterials were collected in four projects that participated in the NIMH bipolar disorder genetics initiative. From 1991–1998, the principal investigators and co-investigators were: Indiana University, Indianapolis, Indiana, USA, U01 MH46282, J. Nurnberger, M. Miller and E. Bowman; Washington University, St. Louis, Missouri, USA, U01 MH46280, T. Reich, A. Goate and J. Rice; Johns Hopkins University, Baltimore, Maryland, USA, U01 MH46274, J.R. DePaulo Jr., S. Simpson and C. Stine; NIMH Intramural Research Program, Clinical Neurogenetics Branch, Bethesda, Maryland, USA, E. Gershon, D. Kazuba and E. Maxwell. Data and biomaterials were collected as part of ten projects that participated in the NIMH bipolar disorder genetics initiative. From 1999–2003, the principal investigators and co-investigators were: Indiana University, Indianapolis, Indiana, USA, R01 MH59545, J. Nurnberger, M.J. Miller, E.S. Bowman, N.L. Rau, P.R. Moe, N. Samavedy, R. El-Mallakh (at University of Louisville, Louisville, Kentucky, USA), H. Manji (at Wayne State University, Detroit, Michigan, USA), D.A. Glitz (at Wayne State University, Detroit, Michigan, USA), E.T. Meyer, C. Smiley, T. Foroud, L. Flury, D.M. Dick and H. Edenberg; Washington University, St. Louis, Missouri, USA, R01 MH059534, J. Rice, T. Reich, A. Goate and L. Bierut; Johns Hopkins University, Baltimore, Maryland, USA, R01 MH59533, M. McInnis, J.R. DePaulo Jr., D.F. MacKinnon, F.M. Mondimore, J.B. Potash, P.P. Zandi, D. Avramopoulos and J. Payne; University of Pennsylvania, Philadelphia, Pennsylvania, USA, R01 MH59553, W. Berrettini; University of California at Irvine, Irvine, California, USA, R01 MH60068, W. Byerley and M. Vawter; University of Iowa, Iowa City, Iowa, USA, R01 MH059548, W. Coryell and R. Crowe; University of Chicago, Chicago, Illinois, USA, R01 MH59535, E. Gershon, J. Badner, F. McMahon, C. Liu, A. Sanders, M. Caserta, S. Dinwiddie, T. Nguyen and D. Harakal; University of California, San Diego, La Jolla, California, USA, R01 MH59567, J. Kelsoe and R. McKinney; Rush University, Chicago, Illinois, USA, R01 MH059556, W. Scheftner, H.M. Kravitz, D. Marta, A. Vaughn-Brown and L. Bederow; NIMH Intramural Research Program, Bethesda, Maryland, USA, 1Z01MH002810-01, F.J. McMahon, L. Kassem, S. Detera-Wadleigh, L. Austin and D.L. Murphy.

Funding for this study was provided by grants from T. and V. Stanley, the Simons Foundation J.M and C.D. Stone, grants from NARSAD to F.J.M., T.G.S., D.L.L., M.-C.K. and T.W., and grants from the Essel Foundation and the Sidney R. Baer, Jr. Foundation to D.L.L. and from the Margaret Price Investigatorship to J.B.P. and V.L.W. This work was supported by grants from the National Institutes of Health (NIH) Intramural Research Program, National Institutes of Health, including National Institute of Mental Health(NIMH) grant MH076431 to J.S., which reflects co-funding from Autism Speaks, and the Southwestern Autism Research and Resource Center, as well as NIH grants to J.S. (HF004222), M.-C.K., T.W. and J.M.C. (MH083989), D.L.L. and N.R.M. (MH071523; MH31340), J.M.C. (RR000037), P.F.S. (MH074027 and MH077139), J.S.S. (MH061009), L.E.D. (MH44245), T.H.S. (GM081519) and C.K.D. (MH081810; DE016442; HD04147). We gratefully acknowledge the resources provided by the Autism Genetic Resource Exchange (AGRE) Consortium and the participating AGRE families. The Autism Genetic Resource Exchange is a program of Autism Speaks and is supported, in part, by grant 1U24MH081810 from the National Institute of Mental Health to C.M. Lajonchere (PI). Funding for G.K., N. Craddock., M.J.O. and M.C.O. was provided by the Medical Research Council, UK, and the Wellcome Trust. The Clinical Antipsychotic Trials of Intervention Effectiveness project was funded by NIMH contract N01 MH90001. Funding was provided by the German Ministry of Education and Research BMBF (National Genome Research Network, NGFNplus, MooDS-Net grant no: 01GS08144 to SC and MN; grant no: 01GS08147 to MR). Genotyping of the Molecular Genetics of Schizophrenia study (PI Pablo Gejman) was funded by GAIN of the Foundation for the US NIH. This study makes use of data generated by the Wellcome Trust Case Control Consortium (full list of contributors is presented in the **Supplementary Note**). Funding for that project was provided by the Wellcome Trust under award 076113. Microarray data and clinical information were provided by GAIN. Thanks to the New York Cancer Project, P. Gregersen and A. Lee for providing population control samples. Also, we wish to thank P. Gejman and D. Levinson for helpful discussions. Special thanks to J. Watson for helpful discussions and support.

AUTHOR CONTRIBUTIONS

J.S. organized and designed the study. S.E.M., J.S., S.Y., N.R.M., M.-C.K., G.K., D.G. and A.M.A. contributed to the analysis of genetic data. S.E.M., J.S., C.K.D. and D.L.L. contributed to the analysis of clinical data. S.E.M. and J.S prepared the manuscript. All authors contributed their critical reviews of the manuscript in its preparation. The following persons contributed to the collection of samples and data: (**Schizophrenia**) G.K., D.G., N. Craddock, M.J.O., M.C.O., WTCCC, A.M.A., J.R., D.O.P., J.A.L., J.S.S., P.F.S., J.M., D.E.D., T.W., M.-C.K., E.S., O.K., V. Kraus, D.L.L., T.J.C. and L.E.D.; (**Bipolar Disorder**) J.P., M.Goodell, V.L.W., P.D., S.G., J.S., L.K., J.W., N. Chitkara, F.J.M., A.K.M., J.B.P., T.G.S., M.M.N., S.C., M.R., E.L., G.K., D.G., N. Craddock, M.J.O., M.C.O. and WTCCC; (**Autism**) V. Kustanovich, C.M.L., E.H.Z., P.K., J.G., I.D.K., N.B.S., C.H-E., T.H.S., M.Gill, L.G., T.L., K. Puura, R.A.K., S.L.C., J.S.S. and D.S. Array-comparative genomic hybridization data collection, processing and management at CSHL were carried out by: S.E.M., D.M., V.M., S.Y., M.K., P.R., A.B., K. Pavon, B.L., A.L., J.K., Y.-H.L., L.M.I., V.V. and J.S.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Weiss, L.A. *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* **358**, 667–675 (2008).
2. Kumar, R.A. *et al.* Recurrent 16p11.2 microdeletions in autism. *Hum. Mol. Genet.* **17**, 628–638 (2008).
3. Marshall, C.R. *et al.* Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.* **82**, 477–488 (2008).
4. Walsh, T. *et al.* Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320**, 539–543 (2008).
5. Stone, J.L. *et al.* Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237–241 (2008).
6. Kirov, G. *et al.* Support for the involvement of large CNVs in the pathogenesis of schizophrenia. *Hum. Mol. Genet.* **14**, 796–803 (2009).
7. Karayiorgou, M. *et al.* Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proc. Natl. Acad. Sci. USA* **92**, 7612–7616 (1995).

with schizophrenia. *Hum. Mol. Genet.* **9**, 1415–1423 (2000).
9. Stefansson, H. *et al.* Large recurrent microdeletions associated with schizophrenia. *Nature* **455**, 232–236 (2008).
10. Rujescu, D. *et al.* Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum. Mol. Genet.* **18**, 988–996 (2009).
11. Friedman, J.I. *et al.* CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. *Mol. Psychiatry* **13**, 261–266 (2008).
12. Kirov, G. *et al.* Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Hum. Mol. Genet.* **17**, 458–465 (2008).
13. Ghebranious, N., Giampietro, P.F., Wesbrook, F.P. & Rezkalla, S.H. A novel microdeletion at 16p11.2 harbors candidate genes for aortic valve development, seizure disorder, and mild mental retardation. *Am. J. Med. Genet. A.* **143**, 1462–1471 (2007).
14. Sebat, J. *et al.* Strong association of *de novo* copy number mutations with autism. *Science* **316**, 445–449 (2007).
15. Manolio, T.A. *et al.* New models of collaboration in genome-wide association studies: the Genetic Association Information Network. *Nat. Genet.* **39**, 1045–1051 (2007).
16. Bijlsma, E.K. *et al.* Extending the phenotype of recurrent rearrangements of 16p11.2: deletions in mentally retarded patients without autism and in normal individuals. *Eur. J. Med. Genet.* **52**, 77–87 (2009).
17. Mefford, H.C. *et al.* Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N. Engl. J. Med.* **359**, 1685–1699 (2008).
18. Brunetti-Pierri, N. *et al.* Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat. Genet.* **40**, 1466–1471 (2008).
19. Sharp, A.J. *et al.* Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat. Genet.* **38**, 1038–1042 (2006).
20. Ousley, O., Rockers, K., Dell, M.L., Coleman, K. & Cubells, J.F. A review of neurocognitive and behavioral profiles associated with 22q11 deletion syndrome: implications for clinical evaluation and treatment. *Curr. Psychiatry Rep.* **9**, 148–158 (2007).
Am. J. Hum. Genet. **82**, 199–207 (2008).
22. Szatmari, P. *et al.* Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat. Genet.* **39**, 319–328 (2007).
23. Woodhouse, W. *et al.* Head circumference in autism and other pervasive developmental disorders. *J. Child Psychol. Psychiatry* **37**, 665–671 (1996).
24. Fidler, D.J., Bailey, J.N. & Smalley, S.L. Macrocephaly in autism and other pervasive developmental disorders. *Dev. Med. Child Neurol.* **42**, 737–740 (2000).
25. Courchesne, E., Carper, R. & Akshoomoff, N. Evidence of brain overgrowth in the first year of life in autism. *J. Am. Med. Assoc.* **290**, 337–344 (2003).
26. Butler, M.G. *et al.* Subset of individuals with autism spectrum disorders and extreme macrocephaly associated with germline PTEN tumour suppressor gene mutations. *J. Med. Genet.* **42**, 318–321 (2005).
27. Dementieva, Y.A. *et al.* Accelerated head growth in early development of individuals with autism. *Pediatr. Neurol.* **32**, 102–108 (2005).
28. Redcay, E. & Courchesne, E. When is the brain enlarged in autism? A meta-analysis of all brain size reports. *Biol. Psychiatry* **58**, 1–9 (2005).
29. Lainhart, J.E. *et al.* Head circumference and height in autism: a study by the Collaborative Program of Excellence in Autism. *Am. J. Med. Genet. A.* **140**, 2257–2274 (2006).
30. Fukumoto, A. *et al.* Growth of head circumference in autistic infants during the first year of life. *J. Autism Dev. Disord.* **38**, 411–418 (2008).
31. Courchesne, E. *et al.* Mapping early brain development in autism. *Neuron* **56**, 399–413 (2007).
32. Mazzucchelli, C. *et al.* Knockout of ERK1 MAP kinase enhances synaptic plasticity in the striatum and facilitates striatal-mediated learning and memory. *Neuron* **34**, 807–820 (2002).
33. Sakaguchi, G. *et al.* Doc2alpha is an activity-dependent modulator of excitatory synaptic transmission. *Eur. J. Neurosci.* **11**, 4262–4268 (1999).
34. Miyazaki, T. *et al.* Disturbance of cerebellar synaptic maturation in mutant mice lacking BSRPs, a novel brain-specific receptor-like protein family. *FEBS Lett.* **580**, 4057–4064 (2006).

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA. [2]Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Department of Psychological Medicine and Neurology, School of Medicine, Cardiff University, Cardiff, UK. [3]Child Psychiatry Branch, National Institute for Mental Health, National Institutes of Health, Bethesda, Maryland, USA. [4]Department of Psychiatry, University of Washington, Seattle, Washington, USA. [5]Department of Psychiatry, University of North Carolina, Chapel Hill, North Carolina, USA. [6]Department of Genome Sciences, University of Washington, Seattle, Washington, USA. [7]Graduate Program in Genetics State University of New York, Stony Brook, New York, USA. [8]Psychology Research Laboratory, McLean Hospital, Belmont, Massachusetts, USA. [9]Department of Human Genetics, University of Chicago, Chicago, Illinois, USA. [10]Division of Human Genetics and [11]Section of Biochemical Genetics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. [12]Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [13]Laboratory of Statistical Genetics, The Rockefeller University, New York, USA. [14]The Prince of Wales International Center for SANE Research, Warneford Hospital, Oxford, UK. [15]College of Physicians and Surgeons of Columbia University, Columbia University, New York, USA. [16]New York State Psychiatric Hospital, New York, New York, USA. [17]Department of Clinical Chemistry. [18]Department of Child Psychiatry, Tampere University and University Hospital, Tampere, Finland. [19]Genetic Basis of Mood and Anxiety Disorders Unit, National Institute for Mental Health, National Institutes of Health, Bethesda, Maryland, USA. [20]Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, Maryland, USA. [21]Department of Psychiatry Research, The Zucker Hillside Hospital, Glen Oaks, New York, USA. [22]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, University of Heidelberg, Mannheim, Germany. [23]Department of Genomics, Life and Brain Center, [24]Institute of Human Genetics and [25]Department of Psychiatry and Psychotherapy, University of Bonn, Germany. [26]Mood and Anxiety Disorders Program, National Institute for Mental Health, National Institutes of Health, Bethesda, Maryland, USA. [27]Autism Genetic Resource Exchange, Autism Speaks, Los Angeles, California, USA. [28]Center for Molecular Neuroscience, Vanderbilt University, Nashville, Tennessee, USA. [29]Behavioral Sciences Unit, Institute of Child Health University College London, London, UK. [30]Department of Psychiatry, School of Medicine, Trinity College Dublin, Dublin, Ireland. [31]Department of Applied Mathematics and Statistics, State University of New York, Stony Brook, New York. USA. [32]Wellcome Trust Case Control Consortium Affiliations are provided in the **Supplementary Note**. [33]Department of Psychiatry, Harvard Medical School, Boston, Massachusetts, USA. [34]VA Boston Healthcare System, Brockton, Massachusetts, USA. [35]Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA. [36]Eunice Kennedy Shriver Center, University of Massachusetts Medical School, Waltham, Massachusetts, USA. Correspondence should be addressed to J.S. (sebat@cshl.edu).

ONLINE METHODS

Cohort description. For this study, data were collected on 8,800 individuals with schizophrenia, bipolar disorder and autism (cases), and 6,391 controls analyzed on one of four microarray platforms (ROMA, NimbleGen HD2, Affymetrix500K or Affymetrix 6.0). Ascertainment of samples in the primary (1,906 schizophrenia cases and 3,971 controls), replication (2,645 cases and 2,420 controls), autism (934 cases) and bipolar disorder (3,315 cases) datasets is provided in the **Supplementary Note**. A breakdown of samples by array is provided in the **Supplementary Table 1**.

Intensity data processing. Processing of microarray data was performed at three different sites. Affymetrix 500 K data from the National Institute of Mental Health (NIMH) (83 cases) were processed at NIMH using published described methods[4]. Affymetrix 500 K data from Cardiff University and the Wellcome Trust Case Control Consortium (WTCCC) (471 schizophrenia cases, 1,697 bipolar disorder cases and 2,792 controls) were processed at Cardiff University using the same software and initial quality control parameters for CNV calls and arrays as published previously[6]; however, for these analyses, we did not remove CNVs that were detected by <10 probes on both the Affymetrix Nsp and Sty arrays.

All other data were processed at Cold Spring Harbor Laboratory (Cold Spring Harbor, NY) using different methods for dual-color intensity data (array–comparative genomic hybridization (array-CGH) platforms: ROMA 85K and NimbleGen HD2) and single-color intensity data (SNP genotyping arrays: Affymetrix 500K, Affymetrix 5.0 and Affymetrix 6.0). Processing of dual- and single-color intensity data is described in more detail in the **Supplementary Note**.

Array-CGH intensity data. Normalization of ROMA intensity data by locally weighted scatterplot smoothing (LOWESS), and geometric mean estimation of $\log_2$ ratios has been described previously[35].

NimbleGen HD2 dual-color intensity data were normalized in a two-step process: first, a 'spatial' normalization of probes was performed to adjust for regional differences in intensities across the surface of the array, and second, the Cy5 and Cy3 intensities were adjusted to a fitting curve by invariant set normalization, preserving the variability in the data. The $\log_2$ ratio for each probe was then estimated using the geometric mean of normalized and raw intensity data.

SNP genotyping data—Affymetrix 500K, Affymetrix 5.0 and Affymetrix 6.0. To analyze Affymetrix SNP array single-color intensity data, we developed a two-step process that (i) normalizes all arrays by invariant set normalization to a single reference array and (ii) calculates the ratio of intensities for each experiment in comparison to a sex-matched virtual reference genome.

GC correction of $\log_2$ ratios. The final step of data processing involved the correction of the effects of genomic waves in $\log_2$ ratios due to regional correlations with GC content based on the fitted linear regression model proposed by Diskin *et al.*[36]

Chr. 16p11.2 detection by HMM segmentation. To detect 16p11.2 rearrangements in our ROMA and GC-corrected Affymetrix $\log_2$ ratio data, we implemented the seven-state HMM algorithm described previously[35]. We used a modified version of this HMM algorithm to identify CNVs in our higher resolution Affymetrix 5.0, 6.0 and NimbleGen HD2 GC-corrected datasets[37]. The results of segmentation were examined for the presence of CNVs overlapping at least 50% of the 16p11.2 region (Chr. 16: 29,557,498–30,107,355 of the UCSC human genome version HG18 (NCBI Build 36.1)).

16p11.2 genotyping: rare CNV detection by outlier clustering. *Principles.* As an alternative method for genotyping rare CNVs, we developed an algorithm called median *z*-score outlier detection (MeZOD) to detect rare variants based on the probe intensity data across the population of experiments. The principles of this method are similar to other approaches that genotype common CNVs by probe intensity clustering[38,39]; however, in most cases very few individuals carry the rare genotype. Therefore, rather than using cluster analysis to identify variants in the population, our method detects rare outliers in the standardized probe intensity distribution.

Selection of target probes and flanking probes. The 16p11.2 rearrangements were genotyped using probes selected from within the target region (Chr. 16: 29,564,890–30,100,063). Two unique sequences, one proximal (Chr. 16: 27,388,307–28,952,358) and one distal (Chr. 16: 30,304,580–31,870,683) to the 16p11.2 target region, were combined into a single set of invariant probes. The results are displayed as a scatterplot. Median *z*-scores of target probes are shown on the *x* axis, and median *z*-scores of the invariant probes are shown on the *y* axis.

To avoid patterns of common copy number polymorphism, probes were excluded if the positive or negative Pearson correlations with neighboring probes exceeded conservative maximum or minimum thresholds, respectively. Probes not exceeding these thresholds were used for genotyping. **Supplementary Table 7** contains all platform-specific probes within the target and invariant regions. The selected genotyping probes in the target and invariant regions are represented in red and green, respectively, in the UCSC human genome browser.

Median z-score calculation and outlier detection. Calculation of the median *z*-scores was a three-stage process involving (i) experiment-wise $\log_2$ ratio standardization, (ii) probe-wise standardization of the genotyping probe *z*-scores and (iii), median *z*-score determination for the target and invariant region. For each probe m of experiment n, the standardized $\log_2$ ratio *z*-score z is simply calculated by:

$$z(m_n) = \frac{m_n - \mu_n}{\sigma_n}$$

where μ_n and σ_n are the mean and s.d. of probe ratios for experiment n, respectively. The *z*-score for each genotyping probe g in experiment n was then standardized probe-wise within the population of experiments for a given platform by:

$$Z(G_n) = \frac{G_n - \mu_g}{\sigma_g}$$

where μ_g is the mean and σ_g is s.d. of genotyping probe g. Finally, the median for experiment n was calculated for the target genotyping probes and the combined proximal and distal invariant genotyping probes.

To detect rearrangements of 16p11.2, outliers of the target median *z*-score distribution were analyzed. Thresholds were set for microduplications at target median *z*-scores >2 for the ROMA array and >1 for Affymetrix500K, Affymetrix 6.0 and NimbleGen HD2 arrays, whereas the outlier threshold for microduplications on all platforms was below a target median *z*-score of −2. As noted earlier, with the exception of Affymetrix 500K data analyzed locally by the NIMH, Cardiff University and the WTCCC, all intensity data was analyzed using MeZOD at CSHL. Further assessment of 16p11.2 HMM and MeZOD genotyping is provided in the **Supplementary Note** and **Supplementary Figure 3**.

Validations of 16p11.2 rearrangements. All rearrangements of 16p11.2 detected in the primary and replication samples were validated using an additional microarray platform. Microduplications detected on the NimbleGen HD2 platform were confirmed using the Agilent 244K array. CNVs detected on other platforms, including the Cardiff schizophrenia cases, were validated on the NimbleGen HD2 array. Rearrangements detected in the WTCCC were detected independently on both Affymetrix Nsp and Sty arrays (**Supplementary Table 2**). Additional DNA was not available for WTCCC controls to perform additional fine-mapping of events detected in these samples. Of the 15 CNVs detected in additional cohorts of autism and bipolar disorder, genomic DNA was available for 12 (**Supplementary Table 2**), and all CNVs in 12 genomic DNA samples were validated.

Meta-analysis and strength of 16p11.2 associations in multiple psychiatric disorders. Data from this study were combined with data from three independent published studies[1,3,5] to obtain a combined sample of 8,590 schizophrenia, 4,822 bipolar disorder and 2,172 autism or developmental delay cases and a combined sample of 30,492 controls. Controlling for study, the control samples used for a particular disorder were derived only from those studies contributing

disorder and autism or developmental delay consisted of 28,406, 25,225 and 24,891 individuals, respectively. Additional information on each study included in the meta-analysis is provided in the in the **Supplementary Note**.

Statistical analysis. *Association of 16p11.2 microduplication with schizophrenia.* The primary sample consisted of data from multiple microarray platforms that varied in probe density. All had good sensitivity to detect CNVs that are 500 kb in size. However, subtle differences in sensitivity could have influenced the overall frequency of 16p11.2 microduplications when all platforms were combined into a single dataset. Therefore, we used array type as a stratifying variable when testing for association using the Cochran-Mantel-Haenszel (CMH) exact test. Logistic regression was used to estimate the combined ORs and the 95% confidence intervals and to measure the effect of array type based on the deviance in the *P* value. The Fisher's exact test was used to test the association of the 16p11.2 microduplications in the replication dataset (single array platform). The Breslow-Day-Tarone test was used to assess the homogeneity of the ORs between the primary and replication datasets. We also examined whether sex had an effect on the association by using sex as a covariate. Results of these analyses are discussed further in the **Supplementary Note**.

Meta-analysis of duplications and deletions of 16p11.2 in multiple psychiatric disorders. The association of the microduplication and microdeletion was examined independently in each disorder using the CMH exact test with source as a stratifying variable. The Breslow-Day-Tarone test was used to assess the homogeneity of the partial OR between the studies of each disorder used in the meta-analysis. A common *P* value and OR were reported from the CMH exact test and from the logistic regression, respectively, only if there was homogeneity in the ORs between the studies in the meta-analysis. Given the very small number of deletion observations in the GAIN schizophrenia study and in each of the BD studies, approximate ORs were calculated by replacing the number of deletions n with $n + 0.5$. Results of these analyses are discussed further in the **Supplementary Note**.

Analysis of quantitative clinical features with 16p11.2 rearrangements. Quantitative clinical data on height, weight and head circumference (occipital-frontal circumference, OFC) was collected from records on individuals carrying 16p11.2 in this study, in previously published studies (Weiss *et al.* and Ghebranious *et al.*[1,13]) and from unpublished data on individuals carrying 16p11.2 rearrangements ascertained by referral for global developmental delay (T. Shaikh, The Children's Hospital of Philadelphia, Philadelphia, PA, personal communication). We excluded from our analysis subjects with known Hispanic, Polynesian and African American ethnicity and any subjects with documented cytogenetic abnormalities.

OFC and height measurements were converted to percentile rankings, conditioned on age and gender, using clinical growth charts from the Center for Disease Control's National Center for Health Statistics (see URLs). OFC percentile rankings were further verified using the online tool developed by SimulConsult, which is based on the same reference database (see URLs). Height and OFC percentiles were converted to *z* scores using online resources (see URLs). *Z*-scores were contrasted among 16p11.2 microdeletions versus microduplications carriers using the Wilcoxon two-sample rank sum test. We repeated the analysis using the craniofacial normative database from Farkas *et al.*[40,41] (in individuals of European ancestry). The above analysis was also performed within subsets of samples defined by diagnoses of schizophrenia, developmental delay or autism spectrum disorders to further examine if the observed effect was present within each individual group. The results of these analyses are discussed further in the **Supplementary Note**.

Due to limited availability of data on the IQ of subject participants, we did not examine intellectual disability in microdeletion and microduplication cases. Because of the known influences of antipsychotic medication on body weight, differences in weight between individuals with microdeletions and duplications were not examined.

A description of the psychiatric symptoms in 16p11.2 carriers is provided in the **Supplementary Note** and **Supplementary Table 7**.

URLs. CDC Growth Charts, http://www.cdc.gov/growthcharts/; SimulConsult microcephaly calculator, http://www.simulconsult.com/resources/ftemp20.html; HyperStat Online normal distribution calculators, http://davidmlane.com/hyperstat/z_table.html; dbGaP, http://www.ncbi.nlm.nih.gov/dbgap; GAIN, http://www.genome.gov/19518664.

35. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
36. Diskin, S.J. *et al.* Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Res.* **36**, e126 (2008).
37. Grubor, V. *et al.* Novel genomic alterations and clonal evolution in chronic lymphocytic leukemia revealed by representational oligonucleotide microarray analysis (ROMA). *Blood* **113**, 1294–1303 (2009).
38. McCarroll, S.A. *et al.* Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat. Genet.* **40**, 1166–1174 (2008).
39. Cooper, G.M., Zerr, T., Kidd, J.M., Eichler, E.E. & Nickerson, D.A. Systematic assessment of copy number variant detection via genome-wide SNP genotyping. *Nat. Genet.* **40**, 1199–1203 (2008).
40. Deutsch, C.K. Head Circumference in autism. in *The Autism Encyclopedia* (eds. Neisworth, J. & Wolfe, P.) 96–97 (Brookes, Baltimore, MD, 2004).
41. Deutsch, C.K. & Farkas, L.G. Quantitative methods of dysmorphology diagnosis. in *Anthropometry of the Head and Face* (ed. Farkas, L.G.) 151–158 (Raven, New York, 1994).