

Recurrent 16p11.2 microdeletions in autism

Ravinesh A. Kumar¹, Samer KaraMohamed¹, Jyotsna Sudi¹, Donald F. Conrad¹,
Camille Brune⁵, Judith A. Badner⁴, T. Conrad Gilliam¹, Norma J. Nowak⁶, Edwin H. Cook Jr⁵,
William B. Dobyns^{1,2,3} and Susan L. Christian^{1,*}

¹Department of Human Genetics, ²Department of Neurology, ³Department of Pediatrics and ⁴Department of Psychiatry, University of Chicago, Chicago, IL 60637, USA, ⁵Department of Psychiatry, University of Illinois at Chicago, Chicago, IL 60612, USA and ⁶Department of Cancer Genetics, Roswell Park Cancer Institute, Buffalo, NY 14236, USA

Received November 21, 2007; Revised and Accepted December 19, 2007

Autism is a childhood neurodevelopmental disorder with a strong genetic component, yet the identification of autism susceptibility loci remains elusive. We investigated 180 autism probands and 372 control subjects by array comparative genomic hybridization (aCGH) using a 19K whole-genome tiling path bacterial artificial chromosome microarray to identify submicroscopic chromosomal rearrangements specific to autism. We discovered a recurrent 16p11.2 microdeletion in two probands with autism and none in controls. The deletion spans ~500-kb and is flanked by ~147-kb segmental duplications (SDs) that are >99% identical, a common characteristic of genomic disorders. We assessed the frequency of this new autism genomic disorder by screening an additional 532 probands and 465 controls by quantitative PCR and identified two more patients but no controls with the microdeletion, indicating a combined frequency of 0.6% (4/712 autism versus 0/837 controls; Fisher exact test $P = 0.044$). We confirmed all 16p11.2 deletions using fluorescence *in situ* hybridization, microsatellite analyses and aCGH, and mapped the approximate deletion breakpoints to the edges of the flanking SDs using a custom-designed high-density oligonucleotide microarray. Bioinformatic analysis localized 12 of the 25 genes within the microdeletion to nodes in one interaction network. We performed phenotype analyses and found no striking features that distinguish patients with the 16p11.2 microdeletion as a distinct autism subtype. Our work reports the first frequency, breakpoint, bioinformatic and phenotypic analyses of a *de novo* 16p11.2 microdeletion that represents one of the most common recurrent genomic disorders associated with autism to date.

INTRODUCTION

Autism (OMIM 209850) is a childhood neurodevelopmental disorder apparent by 3 years of age and characterized by qualitative impairments in reciprocal social interaction, deficits in verbal communication, restricted interests and repetitive behaviors. Autism comprises the severe end of the autism spectrum disorders (ASD), which also include Asperger syndrome, pervasive developmental disorder not otherwise specified and Rett syndrome. Prevalence rates for autism and ASD are 0.2 and 0.6%, respectively. Firm evidence of a genetic basis for autism has been demonstrated by twin studies that show

60–91% concordance rates in monozygotic twins (1,2), making autism the most heritable of all complex neuropsychiatric disorders. The search for autism susceptibility loci has involved several approaches including genome-wide linkage analysis, and association and mutation studies of candidate genes (3,4). However, the molecular basis of autism remains largely unknown.

A complementary approach to identify genetic susceptibility variants for autism involves searching for chromosomal abnormalities in patients. Microscopically visible chromosomal rearrangements such as deletions and duplications have been identified in 1–3% of autism subjects (5). The most

*To whom correspondence should be addressed at: Department of Human Genetics, University of Chicago, 920 East 58th Street, CLSC 319, Chicago, IL 60637, USA. Tel: +1 7738342971; Fax: +1 7738348470; Email: schrist@bsd.uchicago.edu

© 2007 The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

frequent are maternal 15q11-13 duplications; other known recurrent rearrangements include duplications of 17p12, and deletions of 7q11.23, 17p11.2 and 22q11.2 (3). Recent high resolution microarray-based studies in autism have identified a number of novel submicroscopic copy number variants (CNVs) including both deletions and duplications (6,7), although all are rare and few are recurrent. We hypothesize that more remain to be discovered.

We undertook a genome-wide investigation of autism-specific CNVs in 180 autism probands and 372 control subjects using a 19K whole-genome tiling path bacterial artificial chromosome (BAC) microarray and identified a 16p11.2 microdeletion in two patients but no controls. We screened an additional 532 autism probands and 465 controls by quantitative PCR (qPCR), and identified two additional probands deleted for 16p11.2. Here we provide phenotype data for patients with microdeletions at this locus, present detailed molecular characterization of the 16p11.2 deletion breakpoints, and perform computational network studies of the ~25 genes within the microdeletion. Our results demonstrate that microdeletions of 16p11.2 are among the most frequent recurrent genomic disorders associated with autism to date.

RESULTS

We investigated 180 unrelated autism subjects and 372 control subjects by array comparative genomic hybridization (aCGH) using a whole-genome tiling path microarray comprised of ~19 000 BACs. In this initial study, we identified a recurrent ~500-kb microdeletion in two unrelated patients with autism (HI0646 and HI0624) that was not detected in 372 control subjects, indicating a frequency of 1.1% (2/180 autism versus 0/372 controls; Fisher exact test $P = 0.11$) (Fig. 1A and B). The microdeletion has not been reported as a CNV in healthy controls [Database of Genomic Variants (Build 36); <http://projects.tcag.ca/variation/>]. We confirmed both deletions by fluorescence *in situ* hybridization (FISH). Microsatellite analyses and quantitative real-time PCR (qPCR) confirmed paternity and showed that the deletions were *de novo* and maternally derived (Fig. 1C–F). Chromosome analysis was reported as normal for HI0624 and not listed for HI0646 on the Autism Genetics Resource Exchange (AGRE) website (<http://www.agre.org/>). A summary of our molecular studies is presented in Table 1. We did not identify the 16p11.2 microdeletion in an affected sibling of HI0646, or in one affected and one unaffected sibling of HI0624.

We assessed the frequency of the 16p11.2 microdeletion in autism by screening an additional 532 probands and 465 controls using qPCR. We identified two additional probands (HI2467 and HI2997) and no controls deleted for 16p11.2, giving a frequency of 0.4% in this cohort (2/532 autism versus 0/465 controls; Fisher exact test $P = 0.50$). The frequency of the 16p11.2 microdeletion combining both cohorts is 0.6% (4/712 autism versus 0/837 controls; Fisher exact test $P = 0.044$). We confirmed the new microdeletions by FISH, microsatellite analysis and aCGH (Table 1), also confirming paternity (data not shown). Chromosome analysis was reported as normal for HI2977 and not listed for HI2467. We demonstrated that the deletion in HI2977 was

de novo and paternally derived; his affected sibling did not carry the microdeletion. The last family proved to be more complex; the deletion in HI2467 was paternally inherited and apparently *de novo* as the deletion was not found in either parent (or in her unaffected sister). However, her affected brother also had the microdeletion, demonstrating post-zygotic mosaicism for the 16p11.2 deletion and complete association of the microdeletion with autism in this family. The pedigree structures of all four families are shown in Figure 2. We also identified the reciprocal duplication product of the 16p11.2 microdeletion in one subject with autism (HI0128), her unaffected mother, and two controls (Table 1).

The genomic structure and gene content of the 16p11.2 microdeletion were interrogated using the UCSC Browser (<http://genome.ucsc.edu>; hg 18). Segmental duplications (SDs) of ~147 kb and >99% identity were found to flank each end of the microdeletion (Fig. 3A), a common characteristic of genomic disorders. We tested the hypothesis that the 16p11.2 microdeletion is mediated by recombination between the SDs by using a custom-designed chromosome 16p11.2 oligonucleotide array. Although the array data produced low signal-to-noise ratio due to the highly repetitive genomic structure of our target region, we could approximate the breakpoints in all five patients (Fig. 3B). We estimated the telomeric and centromeric breakpoint boundaries at ~29.6 and ~30.2 Mb, respectively, which are in close proximity to the flanking SDs, thereby supporting our hypothesis of SD-mediated recombination.

To better map the deletion breakpoints, we performed replicate hybridization experiments in subject HI2467 and combined these results using segmentation analysis (Fig. 3C). Although complicated by the very high homology between the SDs, these results demonstrate a sharp breakpoint at the centromeric or internal end of the telomeric SD (arrow 1). However, the centromeric SD has two apparent breakpoints located near the internal (arrow 2) and external (arrow 3) ends of the repeat with an intermediate \log_2 ratio between them. We interpret the intermediate \log_2 ratio to reflect reduction of sequence within the SD from the normal four copies to three copies, as many of the probes from the SD regions have very high homology and cannot be reliably distinguished. This analysis unambiguously maps the breakpoints of the 16p11.2 microdeletion to the edges of flanking SDs, consistent with our previously estimated telomeric and centromeric breakpoint boundaries.

The microdeletion contains ~25 genes (Fig. 3A), with another four located within the SDs. To study the potential contribution of individual or sets of genes in the microdeletion to autism, we investigated whether these genes interact biologically using the Ingenuity Pathway Analysis (IPA) tool, which uses known functional relationships and interactions between gene products reported in the literature. Twelve of the 25 genes were mapped to a single genetic network (Fig. 4) that included genes involved with cell-to-cell signaling and interaction. The pathways for 3 of the 25 genes (*DOC2A*, *MAPK3* and *ALDOA*) included post-synaptic density genes that have been hypothesized to underlie autism (8).

We analyzed the behavioral profile on all patients with 16p11.2 microdeletions based on the phenotype data currently

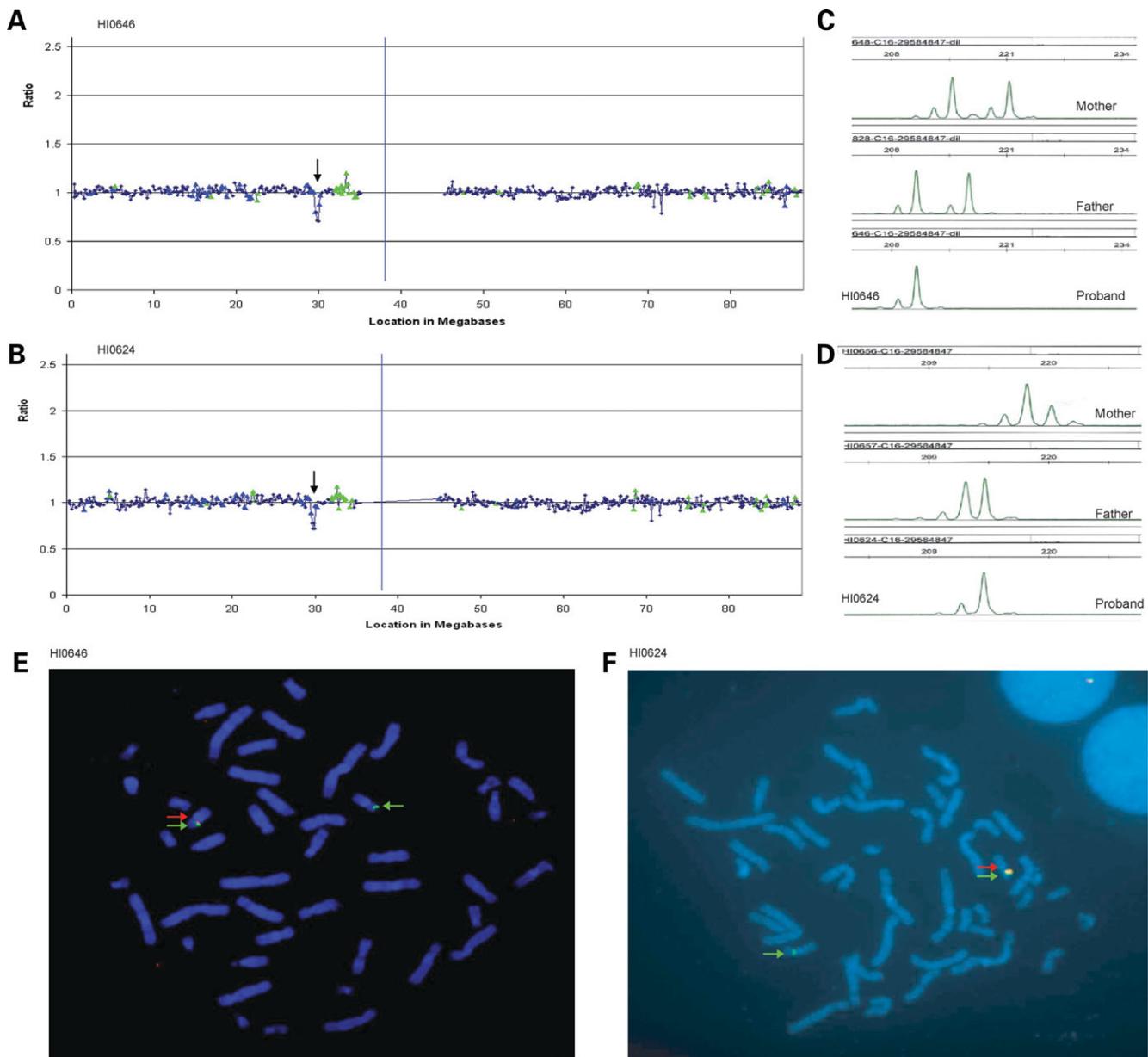


Figure 1. Identification of a new recurrent *de novo* 16p11.2 microdeletion in autism. (A and B) BAC aCGH demonstrates the presence of a deletion at 16p11.2 in two patients (black arrows highlight the vertical line of dots that indicate the presence of a deletion). The aCGH results show the \log_2 ratio of the reference versus patient DNA on the vertical axis. Each individual BAC is represented as a single blue dot. Known CNVs are represented as green dots. The horizontal axis shows the position of each BAC along the chromosome. (C and D) Microsatellite analysis demonstrates that the deletions are *de novo* and maternally inherited. (E and F) FISH results confirm the microdeletion at 16p11.2. BAC RP11-1107E4 (green arrow) and BAC RP11-114A15 (red arrow) show two copies of the green probe and a single copy of the red probe indicating a deletion of RP11-114A15. A similar result was obtained using RP11-74E23 (red) and RP11-455F5 (green) (data not shown).

available in the AGRE database (Table 2). We found no striking features that distinguish subjects with the 16p11.2 microdeletion as a distinct subtype of autism, although a majority had a trend toward behavioral difficulties involving aggression and overactivity. While all patients in the AGRE database have had physical examinations, data regarding facial appearance were entered for only one sibling pair. The boy with the 16p11.2 microdeletion (HI0646) had minor facial dysmorphism consisting of downslanting palpebral fissures, prominent ears and broad nasal root. However,

his brother with autism who did not have the microdeletion also had minor facial dysmorphism, which suggests that the minor facial dysmorphism may be familial rather than related to the microdeletion.

We also reviewed available phenotype data for subjects with duplications of 16p11.2. One girl (HI0128) with autism and duplication 16p11.2 had severe compulsions and rituals consistent with obsessive-compulsive disorder. However, she also had a 1.0-Mb deletion of 7q31.2 and 890-kb duplication of 16q22.1, making interpretation of her phenotype difficult;

Table 1. Summary of molecular studies in patients with 16p11.2 rearrangements

Individual	Phenotype	Rearrangement	Inheritance	BAC aCGH	FISH	Microsatellite	qPCR	Oligo aCGH
H10646	Autism	Deletion	Maternal (<i>de novo</i>)	+	+	+	+	+
H10624	Autism	Deletion	Maternal (<i>de novo</i>)	+	+	+	+	+
H12467 ^a	Autism	Deletion	Paternal (mosaic)	+	+	+	+	+
H12466 ^a	Autism	Deletion	Paternal (mosaic)	+	+	+	+	+
H12997	Autism	Deletion	Paternal (<i>de novo</i>)	+	+	+	+	+
H10128	Autism	Duplication	Maternal (inherited)	+	+	+	+	-
04C34861	Control	Duplication	na	+	-	na	+	-
04C36902	Control	Duplication	na	+	-	na	+	-

na, parental DNA not available.

+, completed.

-, not done.

^asiblings.

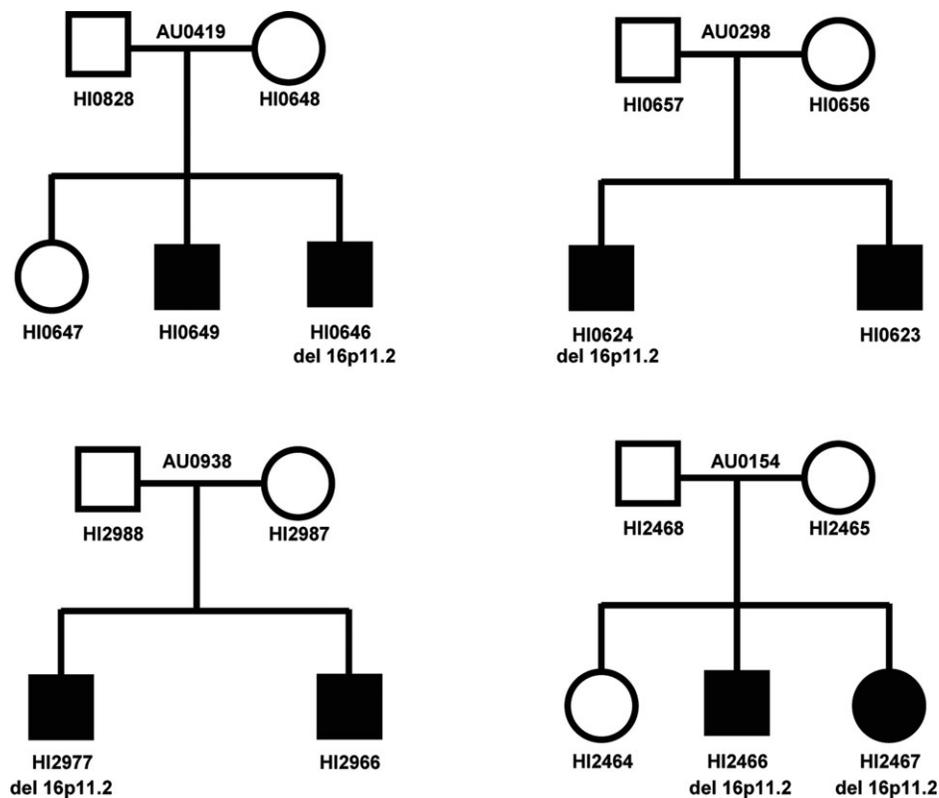


Figure 2. Pedigree structure of families with 16p11.2 microdeletions. Squares indicate males and circles indicate females. Individuals affected with autism are presented as black symbols while white symbols are unaffected. The family ID is presented at the top of each pedigree. Individual IDs are presented beneath each symbol. Individuals with the 16p11.2 microdeletion are indicated.

only the 16q22.1 duplication was *de novo*. Her mother (HI0126), who carries both the 7q31.2 deletion and 16p11.2 duplication, has depressive and anxiety symptoms (onset 13–19 years for both) without formal diagnosis of depression or anxiety disorder, as well as undiagnosed learning disabilities and other unspecified behavioral problems. The two control subjects (04C34861 and 04C36902) with duplication 16p11.2 noted several minor behavioral abnormalities such as compulsions, generalized anxiety, specific phobias and panic attacks, but no mental health diagnoses. While both had compulsions—defined as an inability to resist certain

activities in the presence of obsessions—the frequency was nominally or not significant in comparison to other controls.

DISCUSSION

The present study reports the recurrent nature and first frequency assessment of a novel and typically *de novo* 16p11.2 microdeletion that is significantly associated with autism: 4/712 autism versus 0/837 controls; Fisher exact test $P = 0.044$. We are aware of no other studies of autism that have

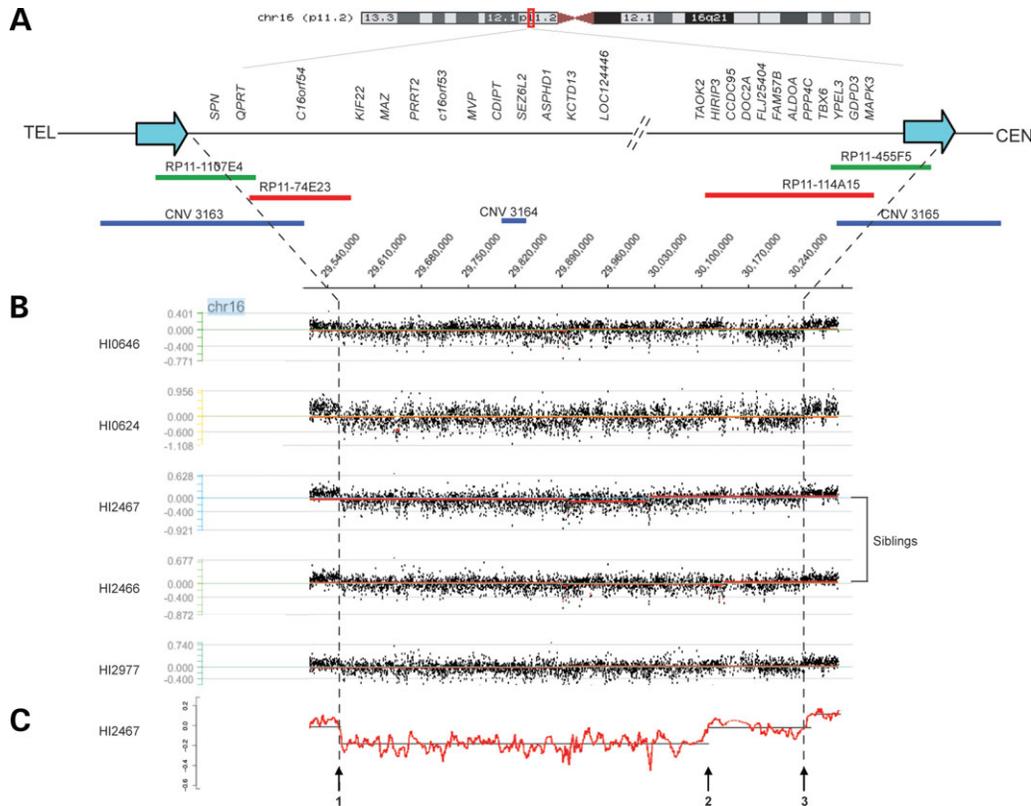


Figure 3. Breakpoint characterization of a recurrent 16p11.2 microdeletion in autism. (A) Schematic of the 16p11.2 microdeletion. The location of the microdeletion is indicated on a diagram of Chromosome 16 followed by a horizontal black line representing an enlargement of the region. Large blue arrows represent the ~147-kb SDs. The genes located within this region, not drawn to scale, are listed above the black line. BACs labeled for FISH are represented by green (overlap SDs) and red (within microdeletion) lines. Known CNVs are represented by blue lines. (B) The approximate breakpoints (dashed lines) are reported for all five patients using a high density custom chromosome 16 oligonucleotide microarray. The short normal flanking regions outside the dashed lines contain probes that are located mostly above the single horizontal red line, whereas the region between the dashed lines contain a large number of probes below the horizontal line (albeit noisy data), indicating copy number loss that is consistent with the FISH analyses. The array results show the \log_2 ratio of the reference versus patient DNA on the vertical axis. Each individual probe is represented as a single black dot. The horizontal axis shows the position of each probe along the chromosome. Vertical dashed lines represent the estimated telomeric and centromeric breakpoint boundaries. (C) The \log_2 ratios from two replicate experiments on HI2467 were averaged to create a lower noise data set. The red line shows a 1000-probe moving average of these data, while the 4 underlying black lines depict the mean probe intensity of (from left to right): the telomeric segmental duplication (SD), the unique deleted region, the centromeric SD, and copy-normal sequence. A sharp breakpoint at the centromeric or internal end of the telomeric SD (arrow 1) is indicated, as well as two apparent centromeric breakpoints located near the internal (arrow 2) and external (arrow 3) ends of the repeat.

reported a single microdeletion or duplication variant with $P < 0.10$. The frequency of the 16p11.2 microdeletion reported in this study is likely an underestimate of the actual occurrence in autism as we tested primarily multiplex families, which would be expected to have a higher frequency of inherited and lower frequency of *de novo* abnormalities compared with simplex families (6). Our study suggests that the 16p11.2 microdeletion is one of the most common recurrent genomic disorders associated with autism.

While our study is the first to report the recurrent nature of 16p11.2 microdeletions in autism, it is not the first report of this deletion. The 16p11.2 microdeletion was recently identified in a female with Asperger syndrome from a simplex family, without further details regarding the phenotype (6). A similar deletion was reported in monozygotic twins presenting with seizure disorder, mild mental retardation and aortic valve anomalies (9). Both twins and a third brother without the microdeletion had additional congenital anomalies, pointing to another genetic disorder segregating in the family that

may account for the aortic valve anomalies. Neither twin had formal testing for autism, although one had more severe expressive and receptive language deficits than expected from his overall level of functioning, as is characteristic of individuals with ASD (Supplementary Materials, Data S1 for details).

Finally, one patient with a large ~8.7-Mb deletion of 16p11.2–p12.2 that overlaps the region deleted in our patients was recently reported (10). This deletion encompasses ~104 genes, so that a more severe phenotype would be expected. Accordingly, she was reported to have developmental abnormalities most consistent with moderate mental retardation, facial dysmorphism including Robin sequence, multiple hand anomalies and unexplained syncope; however, no mention of autism was included. The additional abnormalities described in this girl are not likely due to deletion of genes in our 16p11.2 microdeletion, as only one proband in our study showed minor and non-specific craniofacial dysmorphism. However, follow-up studies to obtain additional phenotype

Table 2. Phenotype data for 16p11.2-deleted probands and their affected siblings

	H10646	Sibling ^a	H10624	Sibling ^a	H12467	H12466	H12997	Sibling ^a
Autism Diagnostic Interview-Review-Revised (ADI-R)								
Sex	M	M	M	M	F	M	M	M
Age (years)	5	7	4	2	12	14	9	7
Social domain	15	28	16	25	30	28	18	16
Communication—Verbal domain	15	—	14	18	—	—	12	13
Communication—Non-verbal domain	—	14	—	—	14	12	—	—
Restricted, repetitive behavior domain	6	6	3	5	4	4	6	4
Classification	Aut	Aut	Aut	Aut	Aut	Aut	Aut	Aut
Age of 1st words (mos)	48	85	27	11	—	84	30	36
Age of 1st phrases (mos)	60	NA	32	normal	NA	156	42	42
Any language regression	NR	NR	NR	NR	WI	NR	NR	—
Any regression of other skills	NR	NR	NR	S	CS/AM	NR	NR	—
Anxiety	NR	++	NR	NR	NR	NR	NR	—
Aggression	++	+++	+	+	++	NR	++	+
Overactivity	+	++	NR	+	+++	++	+++	+++
History of seizures	NR	NR	NR	NR	NR	yes	NR	NR
Coordination	F	G F	G F	G	F	G F	GF	GF
Savant skills	V	NR	—	NR	NR	NR	NR	—
Autism Diagnostic Observation Schedule (ADOS)	Aut	Aut	—	—	Aut	Aut	Aut	Aut
Physical measures and abnormalities ^b								
Age (years)	7	9	—	—	—	—	9	7
Birth weight (g)	3997 (82nd)	3685 (62nd)	—	—	—	—	—	—
Height (cm)	124 (40th)	—	—	—	—	—	—	—
Weight (kg)	24.5 (51st)	36.3 (82nd)	—	—	—	—	—	—
Head circumference (cm)	54 (82nd)	54 (75nd)	—	—	—	—	—	—
Ear length (cm)	5.8	6.8	—	—	—	—	—	—
Functioning/Psychiatric Symptoms/Medical History								
Age (years), IQ testing	7	9	—	—	—	—	9	7
Verbal IQ	74	50	—	—	—	—	74	72
Non-verbal IQ	108	78	—	—	—	70	80	70
Depressive symptoms	No	No	—	—	—	—	—	—
OCD symptoms	Yes	DX	—	—	—	—	—	—
Other diagnosis	—	DX	—	—	—	—	—	—
Stimulants (current)	Methylph, Dexedr	Methylph, Dexedr	—	—	—	—	—	—
Other medications (current)	Risper	Risper, Fluox, Gaba	—	—	—	—	—	—

Sex (M, male, F, female); ADI-R classification (Aut, Autism); Phrases (NA, skill not acquired); Language regression (loss of W, words, I, intent); Skill regression (loss of skills; C, communication, S, social, A, academic, M, motor); Anxiety includes social anxiety or fears; Aggression norms (V, visuospatial), ADOS classification (Aut, Autism; ASO, Autism Spectrum Disorder). Percentiles for birth weight, height, and weight based on median age (e.g. 6.5 years for a 6-year-old) CDC growth charts unless percentile provided in AGRE database. Verbal IQ is from the Peabody Picture Vocabulary Test (PPVT) standard score; non-verbal IQ is from the Raven's Progressive Matrices estimated non-verbal IQ. For psychiatric symptoms, yes, presence of symptoms. no, absence of symptoms, DX, diagnosis of disorder. Medications (methylphenidate).

^aRefers to affected sibling without a deletion of 16p 11.2.

^bAbnormal physical features (including congenital defects) and dysmorphic features were absent in H10646 and sibling.

Considering all of these observations together, we hypothesize that 16p11.2 microdeletions are a risk factor for ASD generally, and may be causal for mild ASD in some families. Further, our study raises interesting questions that have not been discussed in previous reports of CNV in autism, such as the role of genetic heterogeneity within families. Had the current study been performed primarily in simplex families, the discordant genotypes between affected siblings would not have been detected. The present work emphasizes the importance of studying multiplex as well as simplex families for the identification of both inherited and spontaneous mutations.

For a few deletion syndromes such as 22q11.2, both the deletion and reciprocal duplication have been associated with autism (16–18) (Christian *et al.*, submitted). It is not yet clear whether this will be true for the 16p11.2 locus as well. The reciprocal duplication product of the 16p11.2 microdeletion has been reported in a mother and daughter with mild

mental retardation (19) and in two more unrelated patients with autistic behavior (20). However, both studies used lower resolution techniques such as karyotyping and FISH so that the duplicated region may be larger than the ~500-kb segment highlighted in the current study. Our work and that of others suggest that analysis of 16p11.2 duplications should be explored further in autism as well as obsessive-compulsive disorder.

Several of the genes that reside within the 16p11.2 microdeletion represent promising candidates for autism based on known expression and functional data. *MAPK3* is expressed in human fetal and adult brains; *Mapk3*^{-/-} mice display abnormal avoidance behavior, hyperactivity, reduced long-term potentiation and immune system abnormalities (21). *MAZ* is expressed in several tissues with the highest expression in brain found in motor and midfrontal cortex. *MAZ* directly regulates genes involved in GABA signaling (22), neuronal differentiation (23) and the serotonin pathway

(24). *DOC2A* is expressed predominantly in the brain, localizes to synaptic vesicles, and is hypothesized to regulate synaptic activity through Ca-dependent mechanisms (25), consistent with the proposed role of Ca²⁺ signaling in autism (26). *Doc2a*^{-/-} mice display alterations in synaptic transmission and long-term potentiation as well as learning and behavioral deficits that include abnormal passive avoidance behavior (27). *SEZ6L2* is expressed at high levels in human brain; *Sez6l2*^{-/-} mice display behavior and neurological abnormalities that include impaired motor coordination and abnormal excitatory postsynaptic currents (28). *HIRIP3* is a widely expressed gene that interacts directly with HIRA, a major candidate for the DiGeorge syndrome and related developmental disorders (29). *IL6*, which is not located within the deletion, was found as a node in the identified network. *IL6* is known to play a role in brain development, learning and memory, response to disease injury in the nervous system and regulation of the balance between neurogenesis and gliogenesis (30,31), and has recently been identified as a key mediator of the effects of maternal immune activation on fetal brain development (32).

In summary, our work provides the first evidence of a recurrent 16p11.2 microdeletion in autism and ASD, and strengthens evidence for the involvement of 16p11.2 in other cognitive, language and social impairments. We hypothesize that perturbations in the network of genes identified in this study, especially those within the microdeletion region, underlie the autism phenotype. Our discovery of affected siblings without the microdeletion supports a role for genetic heterogeneity in these families, and leads us to hypothesize that the same phenomenon may be found for other autism loci and potentially for other common complex disorders.

MATERIALS AND METHODS

Subjects

The Institutional Review Board at the University of Chicago approved this study. The original 180 AGRE samples were simplex in 19 and multiplex in 161 families. For the additional 230 AGRE samples, 11 were simplex and 219 multiplex families, while the 302 National Institutes of Mental Health (NIMH) Genetics Initiative autism samples included 57 simplex and 245 multiplex families. Therefore, the total cohort of 712 autism probands consisted of 87 simplex and 625 multiplex families. DNA from 837 control subjects was selected from the NIMH Genetics Initiative Control sample set, and were used for both the aCGH ($n = 372$) and qPCR ($n = 465$) experiments. All control subjects were characterized for Axis I disorders. Although individuals were characterized for Axis I disorders, individuals meeting criteria for diagnosis of an Axis I disorder were not excluded from the sample. Autistic traits were not assessed in this sample.

Phenotypic evaluation

Detailed data of all subjects with autism, their parents and some unaffected siblings are available from the AGRE website, and were reviewed, compiled and analyzed by the authors (C.B. and E.H.C.). These data included detailed

medical histories and physical exams, and results from the Autism Diagnostic Interview—Revised (33,34), the Autism Diagnostic Observation Schedule (35) and other cognitive testing. Data represented as missing in Table 2 was unavailable in the AGRE database.

BAC CGH array

The minimal tiling RPCI BAC array contains ~19 000 BAC clones that were chosen by virtue of their STS content, paired BAC end-sequence and association with heritable disorders and cancer. The backbone of the array consists of 4603 BAC clones that were directly mapped to specific, single chromosomal positions by FISH. Each clone was printed in duplicate on amino-silanated glass slides type A+ (Schott Nexterion typeA+) using a MicroGrid II TAS arrayer (Apogent Discoveries). The BAC DNA printing solutions were prepared from sequence connected RPCI-11 BACs by ligation-mediated PCR as described previously (36). The BAC DNA spots are ~80 μm in diameter with a 150 μm center-to-center spacing creating an array of ~38 000 elements. The printed slides were dried overnight and UV-crosslinked (350 mJ) in a Stratalinker 2400 (Stratagene) immediately before hybridization. A complete list of the BAC clones printed on the 19K array can be found on the RPCI microarray facility website. Reference and test sample genomic DNA (1 μg each) was fluorescently labeled using the BioArray CGH Labeling System (Enzo Life Sciences). For this study, each control DNA was used as an independent test sample against a reference DNA pool of 20 sex-mismatched individuals, and each AGRE patient was used as the test sample against the same sex-mismatched reference DNA pools. The DNA was denatured in the presence of random primers at 99°C for 10 min in a thermocycler, quickly cooled to 4°C and transferred onto ice. The samples were labeled with a dNTP-cyanine 3 mix (or dNTP-cyanine 5) and Klenow followed by incubation overnight at 37°C. Unincorporated nucleotides were removed using a QIAquick PCR purification column (Qiagen) and the labeled probe was eluted with 2 \times 25 μl washes. Prior to hybridization, the test and reference probes were resuspended in 110 μl SlideHyb Buffer #3 (Ambion) containing 5 μl of 20 $\mu\text{g}/\mu\text{l}$ Cot-1 DNA and 5 μl of 100 $\mu\text{g}/\mu\text{l}$ Yeast tRNA (Invitrogen), heated to 95°C for 5 min and placed onto ice. Hybridization to the 19K BAC microarrays was performed for 16 h at 55°C using a GeneTAC hybridization station (Genomic Solutions, Inc.) as described previously (37). After hybridization, the slides were automatically washed in the GeneTAC station with reducing concentrations of SSC and SDS. The slides were scanned using a GenePix 4200AL Scanner (Molecular Devices) to generate high-resolution (5 μm) images for both Cy3 (test) and Cy5 (control) channels. Image analysis was performed using the ImaGene (version 6.1.0) software from BioDiscovery, Inc. The log₂ test/control ratios were normalized using a sub-grid Loess correction. Mapping data for each BAC was acquired from the March 2006 version of the human genome and added to the resulting log₂ test/control values. BACs overlapping SDs or published CNVs were flagged (38–41). To select abnormal loci for further analysis, the aCGH data were analyzed using two methods. (i) All clones outside of

five SD were identified using software developed by Shuang Liu. The autosomes and each sex chromosome were analyzed separately. All pseudoautosomal BACs were included with the autosome analyses. (ii) Manual curation was also performed to identify loci with multiple abnormal clones that were detected as a single clone using a threshold of five SD. All loci with two contiguous abnormal BACs were then screened against the control data. Only CNVs not present in the controls were studied further.

Oligonucleotide CGH arrays and breakpoint analyses

We characterized the breakpoints using a NimbleGen Systems (Madison, WI) custom oligonucleotide Chr. 16 array (HG18; NCBI build 36) covering 29 318 000–30 304 400 Mb designed under the following specifications: isothermal probes (50–74 bp) with target T_m of 76; synthesis cycle limit of 148 cycles; filtered at average 15-mer frequency of 50; interval spacing at 2 bp (mean) and 3 bp (median). We also employed a fine tiling commercial design for chr16 from human genome (hg18; NCBI Build 36) designed as above except that interval spacing was 158 bp (mean) and 165 bp (median). All hybridizations were performed as previously described (42). Our estimates of the telomeric breakpoint boundaries were consistent with the data generated from a lower density Chromosome 16 microarray (NimbleGen); however, the centromeric boundary could not be accurately recovered due to a paucity of probes in this region on the lower resolution array (data not shown). A Gaussian mixture model approach was used to identify change points between two adjacent segments of different copy number. Briefly, a large window was manually designed that spanned an equal number of probes in each segment. The model was then constructed as a mixture of two probe sets, with the mixture proportions determined by the (unknown) location of the breakpoint. Data from the probes spanning the interval from the left edge of the design window to the breakpoint parameterize one of the mixture distributions; the other set of probes, the other distribution. The change point was then identified by evaluating the likelihood of the data across a grid of points (one point per interval between probes), re-estimating the other parameters in the model at each point. The maximum likelihood estimates for the three change points apparent in Figure 1H are: 29 557 173–29 559 112 (change point 1); 30 107 181 (change 2); 30 253 847–30 253 852 (change 3). All coordinates are relative to NCBI 36. Strikingly, these estimated change points correspond closely to the edges of known SDs: a 147-kb duplication on the centromeric flank of the deletion covers the sequence from 30 107 356–30 254 364. The other half of this pair is located at 29 368 017–29 514 353; however, there is a third, 22-kb duplication from 29 534 625–29 557 497, whose centromeric edge corresponds exactly with change point 1 identified here. It is possible that a CNV within this 22-kb SD is creating additional changes between HI2467 and the reference, thus complicating the analysis, or that the 22-kb SD is directly involved in NAHR with the centromeric 147-kb SD. Moving averages of the NimbleGen data were constructed with the 'caTools' package

written in the 'R' programming language (<http://cran.r-project.org/src/contrib/Descriptions/caTools.html>).

Quantitative real-time PCR

qPCR was performed using two TaqMan[®] primer/probe amplicon sets designed against the 16p11.2 microdeletion region (Assays-by-Design; Applied Biosystems) as well as a reference primer/probe set (TaqMan[®] RNaseP gene, Applied Biosystems). Sequences for the 16p11.2-region primers and FAM-labeled probes are as follows: SEZ6L2-F, 5'-CCTCTCTCTCCCCACAAAGG-3'; SEZ6L2-R, 5'-TG GACAGCCTGGTTCTCTCT-3'; SEZ6L2-probe, 5'- CCTC TGCTCAACCTCA-3'; C16orf54-F, 5'-ACTGCCCCA GCGAAGATG-3'; C16orf54-R, 5'-GCCACCACCCAG ACC-3'; C16orf54-probe, 5'-CTTCCCAGGCTCC-3'. Reactions were run in triplicate in 6 μ l total volume containing 40 ng genomic DNA, 2X TaqMan Universal Master Mix (ABI part number 4304437) and 20X primer/probe mix. Each assay included a no-template control and three controls subjects each validated by FISH for 16p11.2 copy number (i.e. one, two and three copies, respectively). We repeated qPCR for every subject showing copy number alterations at 16p11.2 by re-running each sample an additional six times for the target and reference probes. Each experiment was performed using a 384-well optical PCR plate and the 7900 HT Real-Time PCR System (Applied Biosystems) under the following cycling conditions: initial step at 50°C for 2 min, denaturation at 95°C for 10 min, followed by 40 cycles at 95°C for 15 s. Data analysis was performed using the SDS v2.1 software's Relative Quantitation feature. Copy number of the 16p11.2 target region was defined as $2^{-\Delta\Delta C_T}$, where ΔC_T is the difference threshold cycle number for the test and reference loci. For each subject showing copy number alterations at the target loci, we confirmed our finding by re-running each sample six times for the target and reference probes.

Fluorescence *in situ* hybridization

Lymphoblastoid cell lines for each proband and selected family members were acquired from the Rutgers University Cell and DNA Repository and cultured using standard techniques. RPCI-11 BACs that defined the boundaries of the selected CNVs were acquired from several sources including the Wellcome Trust Sanger Institute and the Roswell Park Cancer Institute. BAC probes were labeled by nick translation using either Spectrum Green or Spectrum Orange fluorescent dyes (Abbott Labs). FISH was performed using standard techniques. Slides were analyzed with a Zeiss Axioplan 2 fluorescent microscope with a cooled CCD camera and Applied Imaging CytoVision v3.7 software.

Microsatellite analysis

Microsatellites were selected from the UCSC Genome Browser microsatellite or simple repeat tracks and primers were designed using the MIT Primer3 program. For a single reaction, a master mix of 1 μ l 10 \times PCR buffer with 15 mM MgCl₂ (2), 1 μ l 10 mM dNTP, 0.1 μ l Ampli Taq Gold enzyme, 0.8 μ l 10 μ M primers (forward and reverse) and

7.1 μ l sterile H₂O was prepared. One microliter DNA (10 ng/ μ l) was added to each reaction. The PCR reaction was run in ABI 9700 thermocyclers using the following conditions: hot start at 94°C for 10 min, 94°C for 30 s, 55°C for 30 s, 72°C for 30 s for 35 cycles followed by a final extension step at 72°C for 10 min. Samples were analyzed on an ABI 3730 XL DNA sequencing analyzer and processed using GeneMapper 3.7 software (Applied Biosystems).

Pathway analyses

Gene accession numbers for our 'focus' genes (i.e. the 25 genes residing within the 16p11.2 microdeletion) were imported into the IPA software (Systems, Mountain View, CA, USA; <https://analysis.ingenuity.com/pa>), a web-delivered application that enables discovery, visualization and exploration of molecular interaction networks. Genetic networks were ranked by a score that takes into account the number of focus genes and size of networks. The probability that a collection of genes equal to or greater than the number in a network could be achieved by chance alone was calculated. A score of three indicates that there is a 1/1,000 chance that the focus gene(s) are in a network due to random chance. Scores of three or higher have a 99.9% confidence level of not being generated by random chance alone and was used as the cut-off for identifying gene networks.

Statistical analysis

To assess the significance of the difference of the frequency of the 16p11.2 microdeletion between autistic subjects and controls, the Fisher Exact Test was used, as implemented at <http://www.matforsk.no/ola/fisher.htm>.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS

We gratefully acknowledge the resources provided by the Autism Genetic Resource Exchange (AGRE) Consortium and the participating AGRE patients and families. Additional autism families and the control samples were acquired as part of the NIMH Center for Collaborative Genetic Studies on Mental Disorders.

Conflict of Interest statement. None declared.

FUNDING

This study was supported by the National Institutes of Health, National Institute of Neurological Diseases and Stroke (1R01NS51812 to S.L.C.) and National Institute of Mental Health (1R01MH64547-01 to T.C.G.), the National Alliance for Autism Research to S.L.C., and an Autism Speaks Post-Doctoral Fellowship Award to E.H.C. Jr. and C.B. AGRE is supported, in part, by a grant to D.H.G. from the National Institute of Mental Health (NIMH) (MH64547).

REFERENCES

- Bailey, A., Le Couteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E. and Rutter, M. (1995) Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol. Med.*, **25**, 63–77.
- Steffenburg, S., Gillberg, C., Hellgren, L., Andersson, L., Gillberg, I.C., Jakobsson, G. and Bohman, M. (1989) A twin study of autism in Denmark, Finland, Iceland, Norway and Sweden. *J. Child Psychol. Psychiatry*, **30**, 405–416.
- Veenstra-VanderWeele, J. and Cook, E.H. Jr. (2004) Molecular genetics of autism spectrum disorder. *Mol. Psychiatry*, **9**, 819–832.
- Freitag, C.M. (2007) The genetics of autistic disorders and its clinical relevance: a review of the literature. *Mol. Psychiatry*, **12**, 2–22.
- Vorstman, J.A., Staal, W.G., van Daalen, E., van Engeland, H., Hochstenbach, P.F. and Franke, L. (2006) Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism. *Mol. Psychiatry*, **11**, 18–28.
- Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J. *et al.* (2007) Strong association of de novo copy number mutations with autism. *Science*, **316**, 445–449.
- Autism Genome Project Consortium Szatmari, P., Paterson, A.D., Zwaigenbaum, L., Roberts, W., Brian, J., Liu, X.Q., Vincent, J.B., Skaug, J.L., Thompson, A.P. *et al.* (2007) Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat. Genet.*, **39**, 319–328.
- Geschwind, D.H. and Levitt, P. (2007) Autism spectrum disorders: developmental disconnection syndromes. *Curr. Opin. Neurobiol.*, **17**, 103–111.
- Ghebranious, N., Giampietro, P.F., Wesbrook, F.P. and Rezkalla, S.H. (2007) 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.
- Ballif, B.C., Hornor, S.A., Jenkins, E., Madan-Khetarpal, S., Surti, U., Jackson, K.E., Asamoah, A., Brock, P.L., Gowans, G.C., Conway, R.L. *et al.* (2007) Discovery of a previously unrecognized microdeletion syndrome of 16p11.2-p12.2. *Nat. Genet.*, **39**, 1071–1073.
- de Vries, B.B., Pfundt, R., Leisink, M., Koolen, D.A., Vissers, L.E., Janssen, I.M., Reijmersdal, S., Nillesen, W.M., Huys, E.H., Leeuw, N. *et al.* (2005) Diagnostic genome profiling in mental retardation. *Am. J. Hum. Genet.*, **77**, 606–616.
- Koolen, D.A., Vissers, L.E., Pfundt, R., de Leeuw, N., Knight, S.J., Regan, R., Kooy, R.F., Reyniers, E., Romano, C., Fichera, M. *et al.* (2006) A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. *Nat. Genet.*, **38**, 999–1001.
- Fan, Y.S., Jayakar, P., Zhu, H., Barbouth, D., Sacharow, S., Morales, A., Carver, V., Benke, P., Mundy, P. and Elsas, L.J. (2007) Detection of pathogenic gene copy number variations in patients with mental retardation by genomewide oligonucleotide array comparative genomic hybridization. *Hum. Mutat.*, **28**, 1124–1132.
- Friedman, J.M., Baross, A., Delaney, A.D., Ally, A., Arbour, L., Armstrong, L., Asano, J., Bailey, D.K., Barber, S., Birch, P. *et al.* (2006) Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation. *Am. J. Hum. Genet.*, **79**, 500–513.
- Wagenstaller, J., Spranger, S., Lorenz-Depiereux, B., Kazmierczak, B., Nathrath, M., Wahl, D., Heye, B., Glaser, D., Liebscher, V., Meitinger, T. *et al.* (2007) Copy-number variations measured by single-nucleotide-polymorphism oligonucleotide arrays in patients with mental retardation. *Am. J. Hum. Genet.*, **81**, 768–779.
- Vorstman, J.A., Morcus, M.E., Duijff, S.N., Klaassen, P.W., Heineman-de Boer, J.A., Beemer, F.A., Swaab, H., Kahn, R.S. and van Engeland, H. (2006) The 22q11.2 deletion in children: high rate of autistic disorders and early onset of psychotic symptoms. *J. Am. Acad. Child Adolesc. Psychiatry*, **45**, 1104–1113.
- Simon, T.J., Bish, J.P., Bearden, C.E., Ding, L., Ferrante, S., Nguyen, V., Gee, J.C., McDonald-McGinn, D.M., Zackai, E.H. and Emanuel, B.S. (2005) A multilevel analysis of cognitive dysfunction and psychopathology associated with chromosome 22q11.2 deletion syndrome in children. *Dev. Psychopathol.*, **17**, 753–784.
- Mukaddes, N.M. and Herguner, S. (2007) Autistic disorder and 22q11.2 duplication. *World J. Biol. Psychiatry*, **8**, 127–130.
- Engelen, J.J., de Die-Smulders, C.E., Dirckx, R., Verhoeven, W.M., Tuinier, S., Curfs, L.M. and Hamers, A.J. (2002) Duplication of

- chromosome region (16)(p11.2 -> p12.1) in a mother and daughter with mild mental retardation. *Am. J. Med. Genet.*, **109**, 149–153.
20. Finelli, P., Natacci, F., Bonati, M.T., Gottardi, G., Engelen, J.J., de Die-Smulders, C.E., Sala, M., Giardino, D. and Larizza, L. (2004) FISH characterisation of an identical (16)(p11.2p12.2) tandem duplication in two unrelated patients with autistic behaviour. *J. Med. Genet.*, **41**, e90.
 21. Mazzucchelli, C., Vantaggiato, C., Ciamei, A., Fasano, S., Pakhotin, P., Krezel, W., Welzl, H., Wolfer, D.P., Pages, G., Valverde, O. *et al.* (2002) Knockout of ERK1 MAP kinase enhances synaptic plasticity in the striatum and facilitates striatal-mediated learning and memory. *Neuron*, **34**, 807–820.
 22. Murakami, A., Matsuda, M., Nakasima, A. and Hirata, M. (2006) Characterization of the human PRIP-1 gene structure and transcriptional regulation. *Gene*, **382**, 129–139.
 23. Okamoto, S., Sherman, K., Bai, G. and Lipton, S.A. (2002) Effect of the ubiquitous transcription factors, SP1 and MAZ, on NMDA receptor subunit type 1 (NR1) expression during neuronal differentiation. *Brain Res. Mol. Brain Res.*, **107**, 89–96.
 24. Parks, C.L. and Shenk, T. (1996) The serotonin 1a receptor gene contains a TATA-less promoter that responds to MAZ and Sp1. *J. Biol. Chem.*, **271**, 4417–4430.
 25. Groffen, A.J., Friedrich, R., Brian, E.C., Ashery, U. and Verhage, M. (2006) DOC2A and DOC2B are sensors for neuronal activity with unique calcium-dependent and kinetic properties. *J. Neurochem.*, **97**, 818–833.
 26. Krey, J.F. and Dolmetsch, R.E. (2007) Molecular mechanisms of autism: a possible role for Ca²⁺ signaling. *Curr. Opin. Neurobiol.*, **17**, 112–119.
 27. Sakaguchi, G., Manabe, T., Kobayashi, K., Orita, S., Sasaki, T., Naito, A., Maeda, M., Igarashi, H., Katsuura, G., Nishioka, H. *et al.* (1999) Doc2alpha is an activity-dependent modulator of excitatory synaptic transmission. *Eur. J. Neurosci.*, **11**, 4262–4268.
 28. Miyazaki, T., Hashimoto, K., Uda, A., Sakagami, H., Nakamura, Y., Saito, S.Y., Nishi, M., Kume, H., Tohgo, A., Kaneko, I. *et al.* (2006) Disturbance of cerebellar synaptic maturation in mutant mice lacking BSRPs, a novel brain-specific receptor-like protein family. *FEBS Lett.*, **580**, 4057–4064.
 29. Llevadot, R., Scambler, P., Estivill, X. and Pritchard, M. (1996) Genomic organization of TUPLE1/HIRA: a gene implicated in DiGeorge syndrome. *Mamm. Genome*, **7**, 911–914.
 30. Bauer, S., Kerr, B.J. and Patterson, P.H. (2007) The neuropoietic cytokine family in development, plasticity, disease and injury. *Nat. Rev. Neurosci.*, **8**, 221–232.
 31. He, F., Ge, W., Martinowich, K., Becker-Catania, S., Coskun, V., Zhu, W., Wu, H., Castro, D., Guillemot, F., Fan, G. *et al.* (2005) A positive autoregulatory loop of Jak-STAT signaling controls the onset of astroglialogenesis. *Nat. Neurosci.*, **8**, 616–625.
 32. Smith, K.M., Bauer, L., Fischer, M., Barkley, R. and Navia, B.A. (2005) Identification and characterization of human NR4A2 polymorphisms in attention deficit hyperactivity disorder. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **133**, 57–63.
 33. Lord, C., Rutter, M. and Le Couteur, A. (1994) Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Dev. Disord.*, **24**, 659–685.
 34. Lord, C., Risi, S., Lambrecht, L., Cook, E.H. Jr., Leventhal, B.L., DiLavore, P.C., Pickles, A. and Rutter, M. (2000) The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disord.*, **30**, 205–223.
 35. Nowak, N.J., Gaile, D., Conroy, J.M., McQuaid, D., Jordan, H., Mawhood, L. and Schopler, E. (1989) Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. *J. Autism Dev. Disord.*, **19**, 185–212.
 36. Goggins, M.G., Hruban, R.H. and Maitra, A. (2005) Genome-wide aberrations in pancreatic adenocarcinoma. *Cancer Genet. Cytogenet.*, **161**, 36–50.
 37. Nowak, N.J., Wang, Y.D., Head, K., Conroy, J., McQuaid, D. and Nowak, N.J. (2004) Identification and characterisation of constitutional chromosome abnormalities using arrays of bacterial artificial chromosomes. *Br. J. Cancer*, **90**, 860–865.
 38. Sharp, A.J., Locke, D.P., McGrath, S.D., Cheng, Z., Bailey, J.A., Vallente, R.U., Pertz, L.M., Clark, R.A., Schwartz, S., Seagraves, R. *et al.* (2005) Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.*, **77**, 78–88.
 39. Tuzun, E., Sharp, A.J., Bailey, J.A., Kaul, R., Morrison, V.A., Pertz, L.M., Haugen, E., Hayden, H., Albertson, D., Pinkel, D. *et al.* (2005) Fine-scale structural variation of the human genome. *Nat. Genet.*, **37**, 727–732.
 40. Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W. and Lee, C. (2004) Detection of large-scale variation in the human genome. *Nat. Genet.*, **36**, 949–951.
 41. Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M. *et al.* (2004) Large-scale copy number polymorphism in the human genome. *Science*, **305**, 525–528.
 42. Selzer, R.R., Richmond, T.A., Pofahl, N.J., Green, R.D., Eis, P.S., Nair, P., Brothman, A.R. and Stallings, R.L. (2005) Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH. *Genes Chromosomes Cancer*, **44**, 305–319.