Structural Variation of Chromosomes in Autism Spectrum Disorder

Christian R. Marshall,[1] Abdul Noor,[2] John B. Vincent,[2] Anath C. Lionel,[1] Lars Feuk,[1] Jennifer Skaug,[1] Mary Shago,[3] Rainald Moessner,[1] Dalila Pinto,[1] Yan Ren,[1] Bhooma Thiruvahindrapduram,[1] Andreas Fiebig,[6] Stefan Schreiber,[6] Jan Friedman,[7] Cees E.J. Ketelaars,[8] Yvonne J. Vos,[8] Can Ficicioglu,[9] Susan Kirkpatrick,[10] Rob Nicolson,[11] Leon Sloman,[2] Anne Summers,[12] Clare A. Gibbons,[12] Ahmad Teebi,[4] David Chitayat,[4] Rosanna Weksberg,[4] Ann Thompson,[13] Cathy Vardy,[14] Vicki Crosbie,[14] Sandra Luscombe,[14] Rebecca Baatjes,[1] Lonnie Zwaigenbaum,[15] Wendy Roberts,[5,16] Bridget Fernandez,[14] Peter Szatmari,[13] and Stephen W. Scherer[1,*]

Structural variation (copy number variation [CNV] including deletion and duplication, translocation, inversion) of chromosomes has been identified in some individuals with autism spectrum disorder (ASD), but the full etiologic role is unknown. We performed genome-wide assessment for structural abnormalities in 427 unrelated ASD cases via single-nucleotide polymorphism microarrays and karyotyping. With microarrays, we discovered 277 unbalanced CNVs in 44% of ASD families not present in 500 controls (and re-examined in another 1152 controls). Karyotyping detected additional balanced changes. Although most variants were inherited, we found a total of 27 cases with de novo alterations, and in three (11%) of these individuals, two or more new variants were observed. De novo CNVs were found in ~7% and ~2% of idiopathic families having one child, or two or more ASD siblings, respectively. We also detected 13 loci with recurrent/overlapping CNV in unrelated cases, and at these sites, deletions and duplications affecting the same gene(s) in different individuals and sometimes in asymptomatic carriers were also found. Notwithstanding complexities, our results further implicate the *SHANK3-NLGN4-NRXN1* postsynaptic density genes and also identify novel loci at *DPP6-DPP10-PCDH9* (synapse complex), *ANKRD11*, *DPYD*, *PTCHD1*, 15q24, among others, for a role in ASD susceptibility. Our most compelling result discovered CNV at 16p11.2 ($p = 0.002$) (with characteristics of a genomic disorder) at ~1% frequency. Some of the ASD regions were also common to mental retardation loci. Structural variants were found in sufficiently high frequency influencing ASD to suggest that cytogenetic and microarray analyses be considered in routine clinical workup.

Introduction

Autism (MIM 209850) is a neurodevelopmental disorder that manifests in the first three years of life. The group of pervasive developmental disorders (PDDs), also termed autism spectrum disorders (ASDs), includes autism as well as PDD-not otherwise specified (PDD-NOS) and Asperger's disorder. The three core characteristics of the ASDs are impairments of reciprocal social interactions, problems in communication, and a restricted range of behaviors and interests. Asperger's disorder differs from autism in that individuals have an absence of clinically significant cognitive and language delay. ASDs are observed in all populations with an incidence of 6 cases per 1000, with about four times more males diagnosed than females.[1] The estimated prevalence of autism in siblings is 5%–10%.[2]

ASDs are etiologically heterogeneous. They are associated with a recognized cause in about 10% of cases, most commonly with fragile X (MIM 300624) and Rett syndrome (MIM 312750), tuberous sclerosis (MIM 191100), and other medical genetic conditions. Heritability estimates for ASDs, as determined from twin and family studies, are ~90%,[3] and linkage scans have mapped candidate risk loci.[4]

Based on a recent systematic review, cytogenetically detectable chromosome abnormalities are found in 7.4% (129/1749) of ASD cases with a range from 0% to 54%.[5,6] The highest occurrence of events is observed in syndromic forms of ASD.[6] Balanced translocations and inversions accounted for 17% (22/129) of rearrangements. Whereas the most frequent anomaly observed is maternally derived duplication of chromosome 15q11–q13 in 1%–3% of cases,[7] little is known about the proportion of inherited compared to spontaneous karyotypic changes at other sites. With chromosome abnormalities as the initial step to identify ASD candidate loci, mutations have most convincingly been reported in *SHANK3* on chromosome

[1]The Centre for Applied Genomics and Program in Genetics and Genomic Biology, The Hospital for Sick Children, and Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario M5G 1L7, Canada; [2]Centre for Addiction and Mental Health and Department of Psychiatry, University of Toronto, Toronto, Ontario M5T 1R8, Canada; [3]Paediatric Laboratory Medicine, [4]Division of Clinical and Metabolic Genetics, [5]Autism Research Unit, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada; [6]Institute for Clinical Molecular Biology, Christian-Albrechts-University, 24105 Keil, Germany; [7]Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada; [8]Department of Child Psychiatry, University Medical Center Groningen, 9700 RB Groningen, The Netherlands; [9]Division of Metabolism, Children's Hospital of Philadelphia, Philadelphia, PA 19104-4318, USA; [10]Department of Medical Genetics, University of Wisconsin, Madison, WI 53706-1580, USA; [11]Department of Psychiatry, University of Western Ontario, London, Ontario N6A 4G5, Canada; [12]Department of Genetics, North York General Hospital, Toronto, Ontario M2K 1E1, Canada; [13]Offord Centre for Child Studies, Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Ontario L8P 3B6, Canada; [14]Disciplines of Genetics and Medicine, Memorial University of Newfoundland, St. John's, Newfoundland A1B 3V6, Canada; [15]Department of Pediatrics, University of Alberta, Edmonton, Alberta T6G 2J3, Canada; [16]Bloorview Kids Rehab, University of Toronto, Toronto, Ontario M4G 1R8, Canada
*Correspondence: steve@genet.sickkids.on.ca
DOI 10.1016/j.ajhg.2007.12.009.

Table 1. Summary of CNV in ASD and Controls

	Popgen Controls		Autism Probands			
	All CNVs		All CNVs		Autism Specific[a]	
	Full	Stringent[b]	Full	Stringent[b]	Full	Stringent[b]
# Samples	500	500	427	427	427	427
# CNVs	3695	1558	3396	1315	889	277
CNV/Genome[c]	7.4	3.1	8.0	3.1	2.1	0.65
Mean/Median Size (kb)	315/151	470/224	390/162	603/219	518/121	1082/194
% Gain/Loss	59/41%	70/30%	58/42%	62/38%	61/39%	57/43%
Overlapping CNV/Loci (%)[d]	3005/333 (81%)	1226/142 (78%)	2734/277 (80%)	983/94 (74%)	398/122 (44%)	31/13 (11%)
>1 Mb CNV(%)	343 (9%)	250 (16%)	339 (10%)	212 (16%)	63 (7%)	32 (12%)

[a] Not seen in the initial 500 controls. There were 50 probands run in a previous study with the Affymetrix 10K array.[4] The 10K array detected five different CNVs in five individuals. Of these putative CNVs, three were confirmed by the current 500K experiments, one was a false positive, and one had no probe coverage on the 500K platform so its status is unknown.

[b] Stringent data set as called by >1 algorithms or arrays. Analysis with dChip was performed in batches of ~100 probands. For CNAG version 2.0, we set the reference pool to include all samples and performed an automatic batch pair-wise analysis with sex-matched controls. For GEMCA, we used two designated DNA samples (NA10851 and NA15510) as references for pair-wise comparison to all proband experiments. We further filtered these results by including only those CNVs that were common to both pair-wise experiments. In all instances, CNVs were merged if they were detected in the same individual by more than one algorithm with the outside probe boundaries.

[c] CNV/genome breakdown by algorithm: dChip Merged (3.0/genome), CNAG Merged (5.6/genome), GEMCA (5.5/genome). Validation experiments with q-PCR and FISH are described in the text. Another form of validation comes from examining the trios where we can demonstrate inheritance in 48 (maternal is 25, paternal is 23) of the autism-specific stringent data set. Also from the trios, 148 confirmed regions (inheritance assignment) in the stringent data set that overlap with controls (maternal is 65, paternal is 83).

[d] Represents the total number of overlapping and/or recurrent CNVs, the number of overlapping/CNV loci, and the percentage of overlapping CNVs, out of the total data set.

22q13,[8,9] two neuroligin (*NLGN3* and *NLGN4*) genes on the X chromosome,[10] and the neurexin 1 gene on chromosome 2p16.[4,11,12] Recent data have also revealed that submicroscopic copy number variants (CNVs) can have a role in ASD,[4,13] and de novo CNVs seem to be a more common risk factor in sporadic compared with familial forms of ASD.[13,14]

Our objective was to determine the potential contribution of all forms of genomic structural variation in ASD. We therefore used both high-resolution microarray technologies and karyotyping, allowing detection of both unbalanced and balanced as well as submicroscopic and cytogenetically visible structural variants. We also built a new Autism Chromosome Rearrangement Database (ACRD) allowing integration of ours and all other molecular information with the wealth of karyotypic data gathered over the years.[15] We have identified several new candidate ASD-susceptibility loci, some of which encode proteins that are known to function within the synapse. Although the association of these variants with ASD is often compelling, there are many genotypic and phenotypic complexities. Our data, however, do indicate that high-resolution microarray analysis can be a powerful tool for clinical categorization and diagnosis in ASD.

Material and Methods

DNA Samples and Population Structure

The study included 427 ASD families (Table S1 available online). All cases met Autism Diagnostic Interview-Revised (ADI-R) and Autism Diagnostic Observation Schedule (ADOS) criteria on a clinical best estimate.[16] Of these, 32 carried a cytogenetic chromosome rearrangement, 18 were detected through this study by karyotyping 313 of 413 idiopathic samples, and 14 were already known to carry karyotypic anomalies (Tables S1 and S2). Affected and unaffected siblings were also assessed, and 56% (237/427) had one child (simplex) and 44% (189/427) had more than one child (multiplex) with ASD. This ratio of multiplex to simplex is artificially elevated, reflecting our earlier interest in analyzing families in linkage studies.[4] Most cases were screened for fragile X mutations (75%), and if detected, they were not included. Experiments were performed on blood DNA (80%) or otherwise lymphoblasts. Population ancestry was estimated via STRUCTURE[17] (Table S1).

Affymetrix GeneChip Human Mapping 500K Array and Karyotyping Experiments

For each sample, approximately 500,000 SNPs were genotyped with the combined two-chip Affymetrix NspI and StyI GeneChip Human Mapping Commercial or Early Access Arrays according to standard protocols.[9,18] Data have been submitted to the Gene Expression Omnibus database (accession GSE9222). Karyotypes were generated by standard clinical diagnostic protocols.

Characterization of Copy Number Variation

NspI/StyI array scans were analyzed for CNV content with a combination of DNA Chip Analyzer (dChip),[4,19] Copy Number Analysis for GeneChip (CNAG),[20] and Genotyping Microarray based CNV Analysis (GEMCA)[21] (Table 1).

Controls and the ACRD

Initial control samples consisted of (1) CNVs observed in 500 Europeans from the German PopGen project[22] and (2) entries in the Database of Genomic Variants (containing 8006 CNVs at 3933 loci).[23] A CNV was considered ASD specific if it was >10 kb,

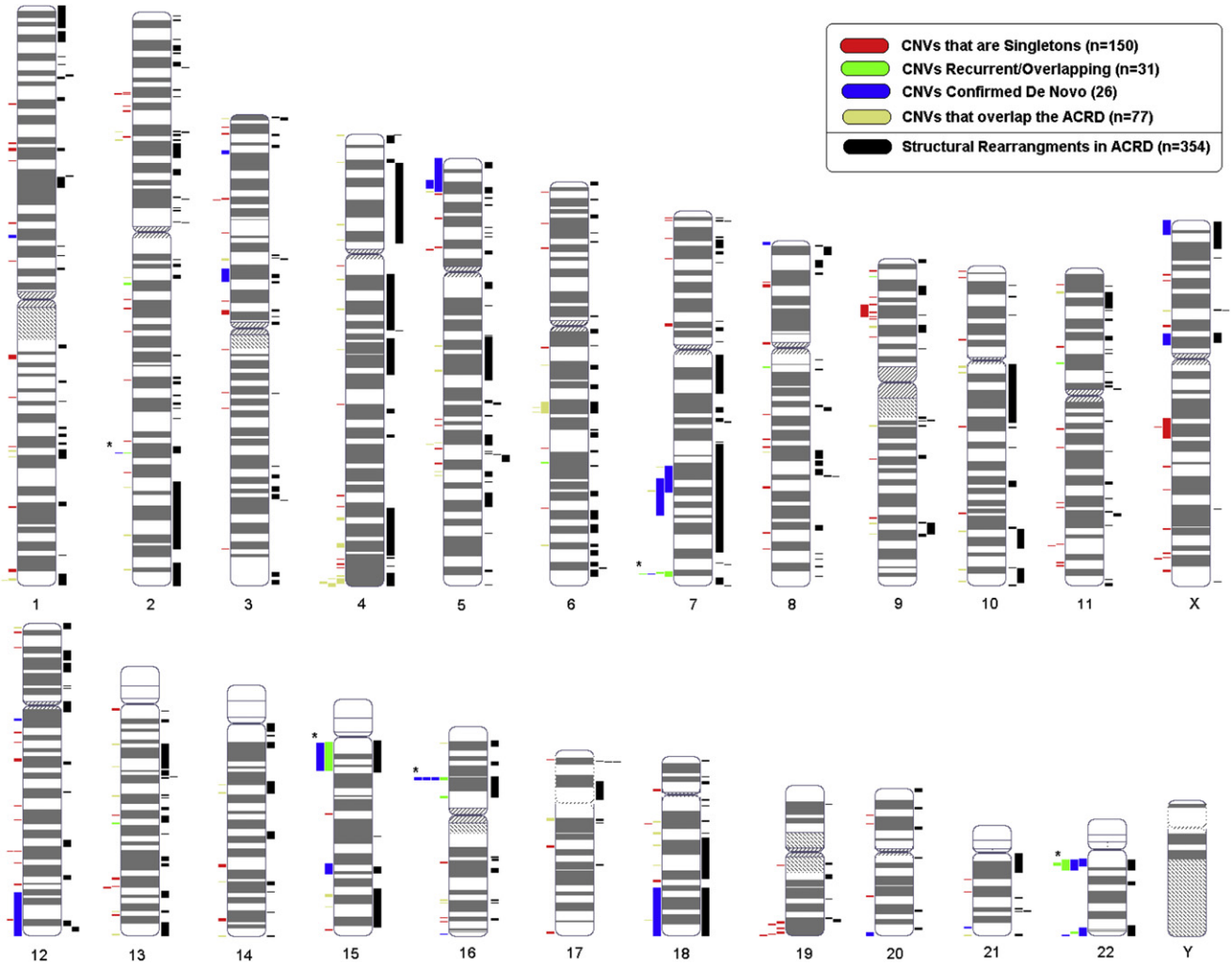

Figure 1. Genome-wide Distribution of CNVs

CNVs from the Autism Chromosome Rearrangement Database (ACRD) are plotted to the right of each chromosome (black). CNV data from the autism-specific stringent data set from the current study are shown to the left of the chromosome and is categorized as de novo (blue), overlapping/recurrent (green), CNVs overlapping with structural variation from the ACRD (yellow), and singleton CNVs (red). Note that five CNVs belong to both de novo and the recurrent categories and these are denoted by an asterisk (see Tables 2 and 3). All CNV data sets are described in Table 1 and the characteristics of the 277 CNVs depicted here are described in Table S3. The raw data from all are present in the Gene Expression Omnibus database.

contained at least three probes, and at least 20% of its total length was unique when compared to the controls. To further test whether CNVs found in probands were specific to ASD, we also compared to CNV found in a cohort of 1152 nondisease controls of European origin from the Ontario population.[24] In addition, we established the ACRD that had 834 putative CNVs or breakpoints mapped to the genome.

CNV Validation Experiments and Balanced Rearrangement Breakpoint Mapping

PCR validation of CNV calls was performed with Quantitative Multiplex PCR of short fluorescent fragments (QMPSF) or SYBR-Green I-based real-time quantitative PCR (qPCR) with controls at the *ACCN1*, *CFTR*, or *FOXP2* loci. At least two independent qPCR assays were required for confirmation of a CNV. Balanced rearrangements were mapped primarily with FISH.[25] The microdel program[26] was used to score CNV losses.

Statistical Analysis

To assess the significance of the frequency of recurrent CNVs in ASD probands and controls, a 2-sided Fischer's exact test was used.

Results

Structural Variation Characteristics in ASD Cases

We initially tested 427 ASD probands for CNV content including 413 idiopathic cases (18 of which were found by karyotyping to have rearrangements) and 14 others that were enrolled based on prior knowledge of a cytogenetic abnormality (Table S1). We used the Affymetrix 500k SNP array because it provided the highest resolution screen available for both SNP genotype and CNV data. By using the SNPs, we could categorize the ancestry of the samples

Table 2. Twenty-Seven ASD Cases with De Novo Rearrangements

	FamID	Sex	Type[a]	Chromosome[b]	Size (bp)[c]	CNV	Genes[d]	Phenotype Comments
1	SK0181-004	M	CHR (SPX)	3p14.1-p13 (a)	5,346,900	loss	13 genes	IQ/LOF[e] 78, language mod delay,[f] speech 2+,[g]
				t(6;14)(q13;q21) (k)	N/A	none	11 genes	RB2+,[h] dysmorph 3+,[i] congenital unilateral sensorineural hearing loss
2	SK0152-003	M	CHR (MPX)	3p25.1-p24.3 (a)	1,409,600	loss	12 genes	LOF 31, RL/EL[j] mod/sev delay, speech 3+,
				5p15.31-p15.2 (a)	3,429,389	loss	8 genes	RB2+, dysmorph 0, hypotonia affecting FM &
				12q12 (a)	422,842	loss	4 genes	GM development
				t(5;7)(p15p13) (k)	N/A	none	*CDH18*	
3	SK0215-006	M	CHR (SPX)	1p21.3 (a)	1,092,500	loss	*DPYD* whole	LOF 31, language sev delay, RB3+, dysmorph N/A
4	SK0205-004	F	CHR (SPX)	5p15.33-p15.2 (k)	13,800,984	loss	46 genes	IQ/LOF & language mild delay, RB1+, dysmorph 3+, Cri du Chat syndrome
5	SK0083-003	M	CHR (SPX)	7q31.1-q31.31 (k)	11,023,507	loss	25 genes	IQ/LOF 56, language sev, speech 3+, RB1+, dysmorph 3+, febrile seizures
6	SK0131-003	F	CHR (SPX)	7q31.1-q32.2 (k)	15,486,722	loss	>50 genes	IQ/LOF 74, RL/EL mod/sev delay, speech 2+, RB2+, dysmorph 3+, microcephaly
7	SK0243-003	M	CHR (SPX)	15q23-q24.2 (k)	4,289,500	loss	>50 genes	LOF severely impaired, language sev delay, RB2+, dysmorph 3+, severe scoliosis, hiatal hernia
8	SK0073-003	F	CHR (SPX)	15q11.2-q13.3 (k)	11,922,600	gain	>50 genes	IQ/LOF 49, RL/EL mod delay, RB3+, dysmorph N/A, premature (34 wks), hypotonia
9	SK0245-005	M	CHR (SPX)	15q11.2-q13.3 (k)	11,871,747	gain	>50 genes	IQ/LOF 47, RL/EL sev delay, RB3+, dysmorph N/A
10	SK0218-003	F	CHR (MPX)	18q21.32-q23 (k)	20,358,999	loss	>50 genes	LOF severely impaired, RL/EL nonverbal, RB3+, dysmorph 3+, seizures, microcephaly, strabismus, hypotonia, cleft palate, club feet
11	NA0039-000	F	CHR (SPX)	22q13.31-q13.33 (k)	3,231,700	loss	41 genes	LOF < 40, language sev delay, RB0, dysmorph 2+, hypotonia, mega cisterna magna, submucous cleft palate, single umbilical artery; father has balanced reciprocal translocation (14;22), proband inherited der(22), sister with ADHD also has unbalanced karyotype with der(14)
12	NA0097-000	F	CHR (SPX)	Xp22.33- p22.31 (a)	5,825,311	loss	21 genes+*NLGN4*	IQ/LOF 117, language average, speech 2+, RB1+, dysmorph 0, carries maternal balanced reciprocal 11;12 translocation
13	SK0283-003	F	CHR (SPX)	47,XX, ring chr1 (k)	N/A	gain	>50 genes	IQ/LOF 30, RL/EL sev delay, RB2+, dysmorph 3+, microcephaly, blood dyscrasia, failure to thrive
14	SK0195-003	M	CHR (SPX)	t(5;8;17)(q31.1; q24.1;q21.3) (k)	N/A	none	5 genes	IQ/LOF unknown, language nonverbal, RB2+, dysmorph 3+, bifid uvula
15	SK0306-004	F	SPX	2q32.1 (a)	97,130	loss	None	IQ/LOF 68, RL/EL mild delay, speech 2+,[i] RB2+, dysmorph 3+, severe hypotonia
16	NA0002-000	M	SPX	7q36.2 (a)	66,462	loss	*DPP6* exonic	IQ/LOF unknown, RL/EL sev delay, RB3+, dysmorph 0
17	SK0262-003	M	SPX	8p23.3 (a)	791,089	gain	*DLGAP2* exonic	IQ/LOF 68, RL/EL mod delay, RB4+, dysmorph N/A
18	MM0278-003	M	SPX	12q24.21-q24.33 (a)	18,218,000	gain	>50 genes	IQ/LOF 24, language sev delay, speech 1+, RB3+, dysmorph 1+, seizures; scoliosis, lower limb anomaly
19	NA0067-000	M	SPX	16q24.3 (a)	265,667	loss	*ANKRD11* exonic	IQ/LOF unknown, language mod delay, RB3+, dysmorph 2+, motor delay, alloimmune ITP
20	MM0088-003	F	MPX	16p11.2 (a)	675,829	loss	28 genes	IQ/LOF 82, RL/EL mod delay, RB3+, dysmorph 0
21	SK0102-004	M	SPX	16p11.2 (a)	432,600	gain	24 genes	IQ/LOF 39, RL/EL sev delay, speech 2+, RB3+, dysmorph 2+, epilepsy, scoliosis, diaphragmatic hernia
22	SK0019-004	M	SPX	16p11.2 (a)	675,829	loss	28 genes	IQ/LOF 93, RL/EL average, speech 2+, RB3+, dysmorph 1, hyperphagia and severe obesity
23	SK0244-003	M	SPX	21q22.3 (a)	353,936	gain	4 genes	IQ/LOF 80, RL/EL mild delay, RB1+, dysmorph N/A
24	MM0109-003	F	SPX	20q13.33 (a)	1,427,661	gain	44 genes	IQ/LOF 27, language nonverbal, RB3+,
				22q13.33 (a)	276,702	loss	13 genes+*SHANK3*	dysmorph 1+

Table 2. ***Continued***

	FamID	Sex	Type[a]	Chromosome[b]	Size (bp)[c]	CNV	Genes[d]	Phenotype Comments
25	SK0119-003[k]	M	MPX	22q11.21 (a)	2,771,300	loss	>50 genes	IQ/LOF 77, RL/EL mod/sev delay, dysmorph 3+, velocardiofacial syndrome
26	SK0297-003	M	SPX-MZ	22q11.21 (a)	4,281,262	gain	>50 genes	IQ/LOF 75, RL/EL average, speech 1+, RB3+, dysmorph 2+, seizures, MZ twin (discordant for ASD)
27	SK0306-004	F	SPX	Xp11.23-p11.22 (a)	4,643,367	gain	>50 genes	IQ/LOF 68, RL/EL mild delay, speech 2+, RB2+, dysmorph 3+, severe hypotonia

[a] Probands with abnormal karyotypes (CHR) (1–14) are separated from probands belonging to simplex (SPX) and multiplex (MPX) families with normal karyotypes (15–27). Families are grouped and sorted based on simplex (SPX), multiplex (MPX), and chromosomal abnormalities (CHR). Simplex families with affected monozygotic twins are denoted as SPX-MZ. Some cases here are also recurrent and appear in Table 3 and some of the family pedigrees are shown in Figure 2 and Figure S2. For multiplex families, the de novo events were not detected in affected siblings.

[b] De novo event detected by either karyotype (k) or microarray (a).

[c] All de novo CNV detected by the array were validated with qPCR, whereas translocations and larger deletions/duplications detected by karyotyping were confirmed by FISH. In all cases where an unbalanced change occurs, CNV size is based on array results. The breakpoints have not been accurately defined, and CNVs may be smaller or larger than posted.

[d] If the CNV intersects only a single gene (suggesting that it may disrupt the gene), the term "exonic" is used, and if the CNV encompasses the entire gene, the term "whole" is used. The term "intronic" is used for CNV that overlaps noncoding parts of a single gene.

[e] IQ/LOF, level of functioning denotes average of Vineland Social, Communication, and daily living scores and nonverbal IQ, when available.

[f] Language was rated as average, nonverbal, mild, mod (moderately delayed), or sev (severely delayed).

[g] Speech refers to the severity of impaired speech intelligibility, most likely resulting from oral motor apraxia (1+, mild; 2+, moderate; 3+, severe unintelligibility; or 0, intelligible speech).

[h] RB: repetitive behavior score was derived from ADI and ADOS ratings (1+, mild; 2+, moderate; 3+, severe repetitive behaviors; or 0, no repetitive behaviors).

[i] Dysmorphology scores were based on anthropometric measurement abnormalities and qualitative features documented either by a clinical geneticist or a developmental pediatrician. Anomalies were reviewed by a single clinical geneticist and assigned score: 0, not dysmorphic; 1+, mild; 2+, moderate; 3+, severe degree of dysmorphism. Children with known genetic syndromes received a score of 3+. N/A, not assessed for dysmorphic features.

[j] RL, receptive language; EL, expressive.

[k] SK0119-003 originally entered the study with an ASD diagnosis but upon re-examination after CNV detection was assessed to be below cutoffs for ASD.

(to guide selection of controls) and found 90.3%, 4.5%, 4.5%, and 0.7% to have European, European/mixed, Asian, or African backgrounds, respectively.

To maximize CNV discovery, we used three calling algorithms (Figure S1) and merged sample-specific results between them to identify a "full" data set of 3396 independent CNVs (~8 CNVs per genome, mean size 390 kb) (Table 1). To minimize potential false positives, we generated a second data set whereby a CNV needed to be detected by two or more algorithms and/or on both the NspI or StyI microarrays (see Figure S2 for representative data).[27] This "stringent" data set contained 1315 CNVs (~3 CNVs per genome, mean size 603 kb), of which we were able to validate 10/10 (5 deletions and 5 duplications) randomly selected CNVs via q-PCR. In total, we validated 97% (89/92) of CNVs tested in the stringent data set.

We then examined 500 European control samples for their CNV content and found similar numbers of CNVs (3695 in the full and 1558 in the stringent data set) to those in the ASD cases (Table 1). This suggested that germline chromosome instability was not a significant contributing mechanism. We also compared the ASD CNVs against the 500 European controls and the Database of Genomic Variants (a repository of structural variation in "nondisease" populations)[23] to establish ASD-specific CNV data sets. Most of our subsequent analysis then focused on the 277 CNVs in our stringent autism-specific category, which mapped across all the 22 autosomes and X (but not Y) chromosome (Figure 1; Table S3). Additional ASD-relevant CNV data are also found in the other categories (Table 1; Table S2; and in the ACRD).

Wide-ranging prevalence frequencies of cytogenetically detectable chromosomal abnormalities in ASD and the inability of microarray scans to find balanced abnormalities prompted us to also perform karyotyping. Karyotyping (and FISH) also provided the ability to characterize the chromosomal context (e.g., ring chromosomes) of some of the CNV regions, something not possible with microarrays alone. We therefore examined consecutive idiopathic cases where blood was available and found that 5.8% (18/313) cases had balanced (11) or unbalanced (7) karyotypes. We note that all unbalanced karyotypic changes described in this study were also found by microarray analysis and are included in the CNV statistics, and further that microarray analysis confirmed all balanced karyotypic changes (i.e., there were no additional gains or losses at the breakpoints below cytogenetic resolution). The genomic characteristics of all CNVs are shown in the ACRD (see Figure S3 as one example). We note that from our data, a CNV loss and gain will typically equate to a standard deletion or duplication, but they could also be more complex.[18] In some cases, a duplication of only part of a gene could lead to its disruption, and there are also positional effects on gene expression to consider.[28]

Table 3. Recurrent and Overlapping Loci in ASD

	Chromosome	FamID	Sex	Type[a]	Size (bp)[b]	CNV	Origin	Genes[c]	Phenotype Comments
1	2q14.1	SK0147-003	F	SPX	478,370	loss	paternal	*DPP10* exonic	IQ/LOF[d] unknown, RL/EL[e] sev delay, RB2+,[f] dysmorph 3+,[g] neurofibromatosis type 1, abnormal EEG
		SK0288-003	F	SPX-MZ	105,120	gain	paternal	*DPP10* intronic	IQ/LOF 73, RL/EL average/mild delay, RB 2+, dysmorph N/A
2	2q32.1	SK0306-004	F	SPX	97,130	loss	de novo	none	IQ/LOF 68, RL/EL mild delay, speech 2+,[h] RB2+, dysmorph 3+, severe hypotonia
		NA0030-000	M	SPX	112,323	loss	maternal	none	IQ/LOF 69, RL/EL mod delay, RB2+, dysmorph 1+, seizures, medicated for OCD, motor incoordination
3	6q22.31	MM0220-003	M	MPX	318,000	gain	paternal	*PLN* whole	IQ/LOF 82, language average, dysmorph 2+, high myopia, fine motor difficulties
		NA0025-000	M	SPX	293,989	gain	paternal	*PLN* whole	IQ/LOF unknown, language: regression/mild delay, RB2+, dysmorph 0
4	7q36.2	SK0190-003	M	SPX	1,780,000	gain	maternal	*DPP6* whole	IQ/LOF 55, RL/EL sev delay, speech 2+, RB1+, dysmorph N/A, parents 1st cousins
		SK0115-003	M	SPX	274,000	gain	unknown	*DPP6* exonic	IQ/LOF 86, RL/EL average, RB2+, ysmorph N/A,
		SK0058-003	M	MPX	16,788	gain	maternal	*DPP6* intronic	IQ/LOF 111, RL/EL average, RB1+, ysmorph N/A
		NA0002-000	M	SPX	66,462	loss	de novo	*DPP6* exonic	IQ/LOF unknown, RL/EL severe delay, RB3+, dysmorph 0
5	8q11.23	SK0143-003	M	SPX	285,200	gain	unknown	*UNQ9433* whole, *RB1CC1* exonic	IQ/LOF 65, RL/EL sev delay, speech 2+, RB2+, dysmorph 2+, seizures, hypoplastic left heart syndrome, left hemidiaphragm paralysis
		MM0236-004	M	MPX	271,679	gain	unknown	*RB1CC1* exonic	IQ/LOF 84, language average, RB1+, dysmorph 0, central auditory processing difficulty
6	9p24.1	SK0270-003	M	SPX	38,900	loss	unknown	none	IQ/LOF 67, RL/EL mild/mod delay, speech 1+, RB3+, dysmorph N/A
		MM0103-003	M	MPX	34,950	loss	paternal	none	IQ/LOF 100, language mild delay, speech 0, RB2+, dysmorph 0, twin preg (other twin lost in 1st trimester), premature (34 wks with respiratory distress syndrome), club feet
7	11p12	MM0272-003	M	MPX	262,938	loss	maternal	none	IQ/LOF 74, language mild delay, speech 3+, RB2+, dysmorph N/A, seizures, unilateral congenital ptosis
		SK0167-003	F	MPX	192,846	loss	unknown	none	IQ/LOF 62, RL/EL average/mild delay, speech apraxia 2+, dysmorph N/A
8	13q21.32	SK0023-003	M	SPX	189,438	gain	unknown	*PCDH9* intronic	IQ/LOF 82, RL/EL mod/sev delay, speech 2+, RB2+, dysmorph N/A; seizures
		MM0299-003	F	MPX	172,401	gain	paternal	*PCDH9* intronic	IQ/LOF 43, language nonverbal, RB2+, dysmorph 0, hypotonia, gags/chokes, fine motor delay
9	15q11.2-q13.3	SK0073-003	F	CHR	11,922,600	gain	de novo	>50 genes	IQ/LOF 49, RL/EL mod delay, RB3+, dysmorph N/A, premature (34 wks)
		SK0245-005	M	CHR	11,871,747	gain	de novo	>50 genes	IQ/LOF 47, RL/EL sev delay, RB3+, dysmorph N/A
10	16p12.2	MM0109-003	F	SPX	1,246,288	gain	maternal	8 genes	IQ/LOF 27, language nonverbal, RB3+, dysmorph 1+
		MM0289-003	F	MPX	802,555	loss	maternal	5 genes	IQ/LOF 45, language delay, RB2+, dysmorph 0
11	16p11.2	NA0133-000	F	SPX	525,319	gain	maternal	29 genes	IQ/LOF & language moderate delay, speech 2+, RB1+, dysmorph 0, early motor delay
		SK0102-004	M	SPX	432,600	gain[i]	de novo	24 genes	IQ/LOF 39, RL/EL sev delay, speech 2+, RB3+, dysmorph 2+, epilepsy, congenital diaphragmatic hernia
		MM0088-003	F	MPX	675,829	loss	de novo	32 genes	IQ/LOF 82, RL/EL mod delay, RB3+, dysmorph 1+
		SK0019-004	M	SPX	675,829	loss	de novo	32 genes	IQ/LOF 93, RL/EL average, speech 2+, RB3+, dysmorph 1+, hyperphagia and severe obesity

Table 3. ***Continued***

	Chromosome	FamID	Sex	Type[a]	Size (bp)[b]	CNV	Origin	Genes[c]	Phenotype Comments
12	22q11.2	SK0119-003[j]	M	MPX	2,771,300	loss	de novo	>50 genes	IQ/LOF 77, RL/EL mod/sev delay, dysmorph 3+, velocardiofacial syndrome
		SK0091-004	F	MPX	4,281,262	gain	paternal	>50 genes	IQ/LOF 92, RL/EL average/mod delay, RB3+, dysmorph 0, placental insufficiency
		SK0297-003	M	SPX-MZ	4,281,262	gain	de novo	>50 genes	IQ/LOF 74, language average; RB3+, dysmorph 2+, seizures, MZ twin (discordant for ASD)
		SK0323-003	M	MPX	743,100	gain	unknown	7 genes	IQ/LOQ 43, language sev delay, RB4+, dysmorph N/A
13	22q13.31	SK0123-004	M	MPX	601,528	gain	maternal	none	IQ 93, language mod delay, RB4+, dysmorph 0, nonidentical triplet, Hirschsprung disease
		MM0102-003	M	MPX	80,380	loss	maternal	none	IQ/LOF 60, language mild delay, dysmorph 0, weakness & hypotonia of arms

[a] Probands are ordered by chromosome location. Families are grouped based on simplex (SPX), multiplex (MPX), and chromosomal abnormalities (CHR). Simplex families with affected monozygotic twins are denoted as SPX-MZ. The de novo cases also appear in Table 2 and some of the family pedigrees are shown in Figure 2 and Figure S2.

[b] All recurrent and overlapping CNVs were detected by the array validated with qPCR. The breakpoints have not been accurately defined, and CNVs may be smaller or larger than noted.

[c] If the CNV intersects only a single gene (suggesting that it may disrupt the gene), the term "exonic" is used, and if the CNV encompasses the entire gene, the term "whole" is used. The term "intronic" is used for CNV that overlaps noncoding parts of a single gene.

[d] IQ/LOF (level of functioning) denotes average of Vineland Social, Communication and daily living scores and nonverbal IQ, when available.

[e] Language (RL, receptive language; EL, expressive language) was rated as average, nonverbal, or mild, mod (moderately), or sev (severely) delayed.

[f] RB (repetitive behavior) score was derived from ADI and ADOS ratings (1+, mild; 2+, moderate; 3+, severe repetitive behaviors; or 0, no repetitive behaviors).

[g] Dysmorphology scores were based on anthropometric measurement abnormalities and qualitative features documented either by a clinical geneticist or a developmental pediatrician. Anomalies were reviewed by a single clinical geneticist and assigned score 0, not dysmorphic; 1+, mild; 2+, moderate; 3+, severe degree of dysmorphism. Children with known genetic syndromes received a score of 3+. N/A, not assessed for dysmorphic features.

[h] Speech refers to the severity of impaired speech intelligibility, most likely resulting from oral motor apraxia (1+, mild; 2+, moderate; 3+, severe unintelligibility).

[i] CNV is called by only one algorithm.

[j] SK0119-003 originally entered the study with an ASD diagnosis but upon re-examination after CNV detection was assessed to be below cutoffs for ASD.

De Novo, Overlapping/Recurrent, and Inherited Structural Variants

We initially prioritized structural variants found in ASD cases to possibly be etiologic if they were not in our initial control set and (1) de novo in origin (27 cases) (Table 2), (2) overlapping (27 cases at 13 loci) in two or more unrelated samples (Table 3), (3) recurrent (same breakpoints) in two or more unrelated samples (four cases at two loci) (Table 3), or (4) inherited (the remainder). We found CNVs at previously reported ASD loci: *NLGN4* and 22q11.2, 15q11–q13, *SHANK3*, and *NRXN1* in categories 1, 2, 3, and 4, respectively. ASD structural variants found in controls (e.g., *NRXN1*) could also be involved.

By testing random families, a de novo CNV rate of 7.1% (4/56) and 2.0% (1/49) was observed in our idiopathic simplex and multiplex families, respectively, similar to rates in another project.[13] This is likely a minimum value because many smaller CNVs will be beyond detection with the current technology. The literature[13,29] and our previous work[18] have shown a <1% spontaneous CNV mutation rate in nondisease samples. Regarding balanced alterations, there was parental information for 13 of 18 cases discovered to carry cytogenetic anomalies (Table S1), and 7 (6 simplex, 1 multiplex) of these were de novo in origin. Because only 1/7 (from a simplex family) was balanced and directly interrupting a gene, we estimate that this class of rearrangements had much less of a contribution than CNVs to the total rate of de novo and inherited structural variation in our ASD cohort.

Our collective data identified 27 de novo cases (Table 2), and in three of these, two or more events were identified. Notably in family SK0152 (Figure 2A), there were four de novo events. In MM0109 (Figure 2B), there were two de novo CNVs, one leading to haploinsufficiency of *SHANK3*.

The 13 loci where overlapping ASD-specific CNVs were found are more likely to have a role in ASD susceptibility because they arise in two or more unrelated families. In six loci, gains and losses were observed at the same locus (Table 3), suggesting that general gene dysregulation may be involved. In those cases with shared CNVs not overlapping with a known medical genetic locus (see below), phenotypic re-examination yielded no obvious genotype-phenotype correlation beyond a trend for larger sporadic abnormalities to generally have a more complex clinical presentation. We also note that in the "full" data set, there are 397 overlapping CNVs at 122 loci, among which other bona fide candidate ASD loci may exist.

By using q-PCR or by assessing SNP patterns, we confirmed 196 inherited CNVs (90 maternal and 106 paternal). No subgrouping of these demonstrated obvious parent-of-origin effects (the two chromosome 15q11–q13 duplications detected were both de novo in origin). We

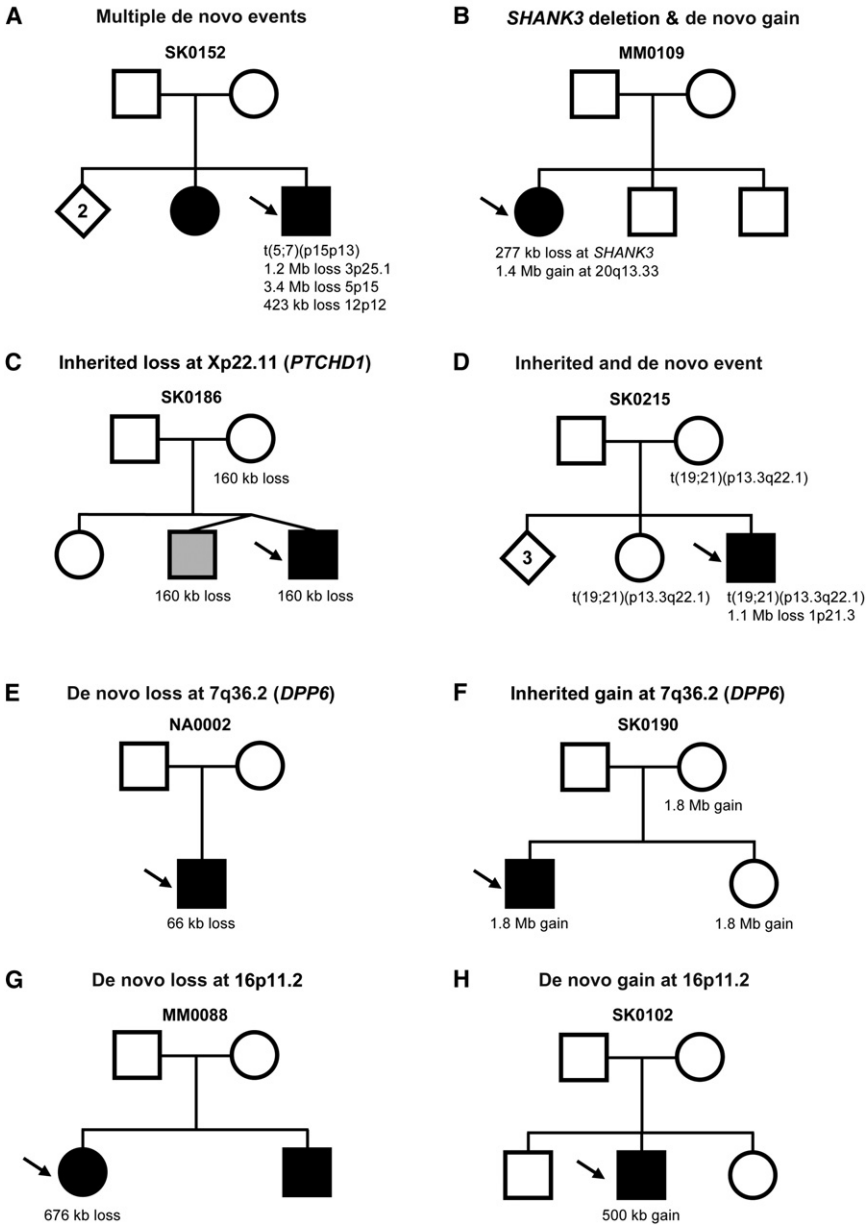

Figure 2. Examples of Complexities of Structural Variants Observed in ASD Families

Males are denoted by squares and females by circles. The size of each de novo or inherited event is shown below each family member. Arrows identify the proband, open shapes are unaffected, and filled have ASD diagnosis (gray denotes developmental delay but not a definitive ASD diagnosis). Diamonds indicate number of older unaffected siblings of unspecified gender. For ASD cases, probands may have multiple de novo events (A), including rearrangements overlapping genes known to be associated with ASD such as *SHANK3* (B). Probands may also inherit chromosome X deletions (at *PTCHD1*) from female carriers (C) or have an inherited translocation in addition to an unrelated de novo deletion (D). Overlapping events in unrelated probands may be either de novo (E) or inherited (F) at the *DPP6* locus. Recurrent de novo events in unrelated probands may also be either losses (G) or gains (H) at chromosome 16p11.2. Additional ASD families with CNVs at *DPP6* and 16p11.2 are shown in Figures S4 and S5, respectively.

did detect a 160 kb deletion in a male inherited from a carrier mother, leading to a null *PTCHD1* in the proband and his dizygotic twin brother (Figure 2C). We also found instances where apparently balanced inherited translocations were accompanied by de novo deletions elsewhere in the genome in the offspring (e.g., *DPYD*; Figure 2D).

Candidate ASD-Susceptibility Genes and Loci Identified

To further validate our findings, we compared our de novo and recurrent/overlapping CNVs from ASD cases to an additional 1152 matched controls. This sample consists of samples from Ontario, Canada, of European origin.[24] We analyzed these in an identical manner to our ASD data and were unable to find CNV matching identically to the 27 de novo cases nor to the 13 recurrent/overlapping ASD cases, further emphasizing the potential importance of these regions. There was, however, one new CNV overlapping the *DPP6* locus found in this control set (see below).

New ASD candidates identified were those with a structural change (either de novo, found in two or more unrelated ASD cases, or for the X chromosome an allele being transmitted maternally from an unaffected carrier) specific to that gene, including *ANKRD11*, *DLGAP2*, *DPP6*, *DPP10*, *DPYD*, *PCDH9*, and *PTCHD1* (Tables 2 and 3). As previously mentioned, *NLGN4*, *SHANK3*, and *NRXN1* were also identified. The *PCDH9* and *NRXN1* genes are also found as CNVs in controls in the Database of Genomic Variants. This suggests that there could also be other important CNVs that are not included in the 277 ASD-specific variants on which we based most of our analyses.

Additional positional candidate genes identified were those found interrupted by balanced cytogenetic breakpoints including *NEGR1*, *PIP5K1B*, *GABRG1*, *KLHL3*, *STK3*, *ST7*, and *SATB2* (Table S2). Moreover, 77 CNVs in our stringent data set overlapped with the ACRD, providing a second line of evidence for involvement (Figure 1). For example, a 4.6 Mb de novo duplication at Xp11.23–11.22 was detected in a female SK0306-004 (Table 2) and a male in

the database.[30] There is a wealth of other data that needs further analysis at the population level, such as the finding of an ~100 kb deletion at 2q37.3 (previously shown to be involved in ASD[6,7] that occurs in twice the frequency in ASD cases compared to controls) (Table S2).

DPP6 and *DPP10* emerge as being positional and functional candidates. *DPP6* (~1.5 Mb in size at 2q14.1) and *DPP10* (~1.3 Mb at 7q36.2) code for accessory transmembrane dipeptidyl peptidase-like subunits that affect the expression and gating of Kv4.2 channels (*KCND2*). Kv4.2 channels function in regulation of neurotransmitter release and neuronal excitability in the glutamatergic synapse[31] at the same sites where *SHANK3* and the *NLGN* gene products are found. Moreover, we have mapped autism balanced breakpoints near *KCND2* at 7q31 (see the ACRD).[25]

For *DPP10*, there are inherited CNV gains and losses (Table 3; Figure S4). De novo and inherited CNVs were found at the multitranscript *DPP6* gene (CNV in 4/427 ASD probands versus 1/1652 controls [initial 500 controls plus 1152 additional controls; Fischer exact test p value = 0.016]). A 66 kb de novo loss encompassing exons 2 and 3 is found in a male in family NA0002 (Figure 2E). In family SK0190, the male proband and an unaffected female sibling both carry a CNV gain inherited from an unaffected mother (Figure 2F) that encompassed the entire *DPP6*. A 270 kb gain is found in SK0115-003 that extends across the first exon (which may disrupt the functional gene), and SK0058-003 carries a maternally inherited 16 kb intronic CNV gain (Figure S4).[14]

Medical Genetics

We identify structural variants overlapping loci involved in medical genetic conditions including Waardenburg Type IIA (MIM 193510; 3p14.1), speech and language disorder (MIM 602081; 7q31), mental retardation (MR) (15q23–q24, 16p11.2), and velocardialfacial syndrome (VCFS) (MIM 192430; 22q11.2), among others (Tables 2 and 3; Table S2). Identification of the structural variant at these loci led to clinical reassessment and either identification or refinement of the diagnosis, for additional syndromic features. Other instances (e.g., SK0186-*PTCHD1* deletion) (Figure 2C) prompted retesting of the entire family and eventually a diagnosis of mild ASD in a previously undiagnosed sibling. This family would therefore move from a simplex to multiplex designation, underscoring the need for cautious interpretation of our statistics in any medical genetic application.

The identification of a de novo deletion (2.7 Mb) at 22q11.2 in two brothers with ASD led to their re-examination and diagnosis for VCFS. The retesting also further defined the siblings to be at opposite ends of the ASD spectrum (Figure S5). Larger duplications (4.3 Mb) of this same region in two other ASD families (SK0289 and SK0091) do not cause VCFS (Table 3). However, in SK0091, the variant was inherited from a normal father and was not found in an affected male sibling. Other studies also reveal differential phenotypes, some having ASD associated with different duplications and deletions at this locus.[32]

We also discovered a recurrent ~500 kb duplication at 16p11.2 in two ASD families (SK0102 and NA0133) (Figure 2; Figure S5). The same region was recently shown to be involved in mental retardation, aortic valve development, and seizure disorder.[33] As with *DPP6*/*DPP10* and 22q11.2, there are carriers of these structural variants without ASD. At the same loci, we also detected a recurrent 676 kb de novo deletion in two ASD families (MM0088 and SK0019). Family SK0019 is a simplex family with the deletion in the proband (Figure S5), whereas in family MM0088, the deletion is detected in only one of two ASD siblings (Figure 2G). This brings a combined discovery rate of 4/427 16p11.2 CNVs in ASD and 0/1652 controls (2-sided Fischer's exact test p value = 0.002). We note that the 16p11.2 deletions and duplications may be reciprocal events, and we are testing this by fine-mapping the breakpoints.

Discussion

Our genome-wide analysis of structural variation yields many new results indicating that chromosomal abnormalities have a substantial role in ASD. In light of these and other findings,[4,13,34] we recommend that genomic analysis, in particular with high-resolution microarrays, be used in research studies of ASD and possibly in clinical assessment once we understand what the implications of these variants are. Such structural variation data could initially serve to focus clinical examination in a search for undetected syndromes. In cases where the loci detected show high penetrance such as haploinsufficiency of *SHANK3*[8,9] and 16p11.2, clinical diagnostic testing could be considered.

In our cohort, we observed that ~7% of randomly selected idiopathic cases harbor de novo genomic CNV rearrangements that are detectable by the microarray platform we used (including ASD cases with balanced translocation or inversion changes that affect genes increases the number slightly). Of these, we found that ~11% of cases actually carry two or more de novo events. Moreover, in other families (and in some cases in the general population), there were non-ASD carriers of the variants, as well as ASD siblings without the variant (Figure 2). These observations, coupled with the newly recognized abundance of structural variants in the genome[35] having a comparatively high new mutation rate,[29] cautions against assigning causation to all de novo rearrangements found. Certainly in some instances additional independent risk factors need to be considered, including those being genetic, epigenetic, environmental, or stochastic in origin.[36] Our finding of X-linked CNVs being maternally transmitted to males and that most structural variants detected on X are inherited from parents may contribute to explain gender prevalence differences and the high degree of heritability in ASD.

With the genome-wide scanning approach, we have identified numerous new putative ASD loci (Tables 2 and 3; Figure 1). Three themes that emerge include the identification of (1) loci that often contain genes functioning in the postsynaptic density (PSD), (2) and/or chromosomal regions previously shown to be involved in mental retardation, and (3) new loci where general dysregulation of gene expression may be involved.

First, we found CNVs that further implicate the *SHANK3*, *NLGN4*, and *NRXN1*-PSD genes and also identify others including *DPP6* and *DPP10* (and *PCDH9, RPS6KA2, RET* from the larger data set). Neuroligins have been shown to bind to the SHANK3 scaffolding protein and are required for the maturation of glutamatergic synapses. Moreover, the neurexins are ligands of the neuroligins, thus supporting a potential pathogenic construct of a network of interrelated molecules, from the *trans*-synaptic interaction between the neurexins and neuroligins to the PSD complex that includes SHANK3. Given that DPP6 and DPP10 are gating channels regulating neurotransmitter release, a functional complex emerges, in which perturbations at a number of potential PSD molecules acting alone or in combination may lead to ASD.[37,38]

Second, we identified at least two ASD loci (15q24, 16p11.2) that overlap with known MR sites,[33,39,40] suggesting that some of the genes causing MR also cause autism (we also found inherited CNVs at three other MR loci: *IDS*, *IL1RAPL1*, and *TSPAN7*). We did not detect structural variants at other chromosome loci associated with anomalies in syndromic forms of ASD such as 7q11.23 (MIM 609757) and 17p11.2 (MIM 610883), most likely because we limited our cases to more strictly defined ASD.

The 16p11.2 CNV region is particularly interesting because it is found at near to 1% frequency in our autism cohort and not controls. It also exhibits characteristics of a genomic disorder, including being flanked by a pair of segmental duplications with >99% identity, which likely mediate the deletion and duplication events through nonallelic homologous recombination (Figure S3). We found both duplication and deletions associated with ASD at this locus. Although no obvious genotype-phenotype correlation is so far obvious in the four families we have identified (Figure S5B), the deletion may be more penetrant because some unaffected individuals carry the duplication. Twenty-four genes map to the CNV interval, including five postsynaptic density (PSD) genes (*DOC2A*, *ALDOA*, *MAPK3*, *CORO1A*, and *CDIPT*) and two candidates for seizure (*SEZ6L2* and *QPRT*). The fact that both duplications and deletions exert phenotype suggests that some genes in the interval are sensitive to dosage.

Third, we discovered six unrelated ASD cases that had either CNV gains or losses in unrelated individuals at the same locus (including 16p11.2) (Table 3). This suggests that fine tuning of gene expression of critical genes in these regions can be crucial for the development of speech and language and/or social communication, as seems to be the case for *SHANK3*[8,9] and genes in the Williams-Beuren syndrome locus (MIM 194050).[41] In some cases, duplication of the Williams-Beuren syndrome region also exhibits an ASD-like phenotype (MIM 609757).

Our understanding of the full etiologic role of structural variation in ASD will require genomic and phenotypic analyses of more cases (and their families) and population controls.[42] To achieve the desired numbers, integration of multiple data sets will likely be necessary; hence, our establishing the ACRD resource as a first step. From our current data, it is already apparent that for a proportion of individuals, it will be possible to describe their ASD based on the underlying structural characteristics of their genome. The detection rate of structural variation and ability to discern the associated outcome will only increase as new higher-resolution genome-scanning methods become available and accessible.

Supplemental Data

Supplemental data include five figures and three tables and can be found with this article online at http://www.ajhg.org/.

Acknowledgments

We thank Lili Senman, Carolyn Noakes, Michelle Lee, Beata Stachowiak, Kathy Whitten, Barbara Noble, and The Centre for Applied Genomics for technical assistance. The work is supported by Genome Canada/Ontario Genomics Institute, the Canadian Institutes of Health Research (CIHR), the McLaughlin Centre for Molecular Medicine, the Canadian Institute of Advanced Research, Autism Speaks, the McMaster Children's Hospital Foundation, the Hospital for Sick Children (SickKids) Foundation, and the German National Genome Research Network (NGFN). C.R.M. is supported by the SickKids Foundation and the National Alliance for Research on Schizophrenia and Depression (NARSAD). L.F. is supported by Canadian Institute for Health Research (CIHR) Fellowship. D.P. is supported by the Netherlands Organization for Scientific Research and the Royal Netherlands Academy of Arts and Sciences. L.Z. is supported by a Health Scholar award from the Alberta Heritage Foundation for Medical Research and is a CIHR New Investigator. S.W.S. is an Investigator the GlaxoSmithKline-CIHR Pathfinder Chair in Genetics and Genomics at SickKids and the University of Toronto.

Received: December 13, 2007
Revised: December 18, 2007
Accepted: December 19, 2007
Published online: January 17, 2008

Web Resources

The URLs for data presented herein are as follows:

The Autism Chromosome Rearrangement Database, http://projects.tcag.ca/autism/
The Database of Genomic Variants, http://projects.tcag.ca/variation/
Gene Expression Omnibus, http://www.ncbi.nlm.nih.gov/geo/
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/sites/entrez?db=OMIM

Accession Numbers

Microarray data have been deposited in the Gene Expression Omnibus database under accession GSE9222.

References

1. Chakrabarti, S., and Fombonne, E. (2005). Pervasive developmental disorders in preschool children: confirmation of high prevalence. Am. J. Psychiatry *162*, 1133–1141.
2. Jones, M.B., and Szatmari, P. (1988). Stoppage rules and genetic studies of autism. J. Autism Dev. Disord. *18*, 31–40.
3. 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.
4. Autism Genome Project Consortium (AGPC), 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.
5. Xu, J., Zwaigenbaum, L., Szatmari, P., and Scherer, S.W. (2004). Molecular cytogenetics of autism. Curr. Genomics *5*, 347–364.
6. 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.
7. Veenstra-Vanderweele, J., Christian, S.L., and Cook, E.H. Jr. (2004). Autism as a paradigmatic complex genetic disorder. Annu. Rev. Genomics Hum. Genet. *5*, 379–405.
8. Durand, C.M., Betancur, C., Boeckers, T.M., Bockmann, J., Chaste, P., Fauchereau, F., Nygren, G., Rastam, M., Gillberg, I.C., Anckarsater, H., et al. (2007). Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. Nat. Genet. *39*, 25–27.
9. Moessner, R., Marshall, C.R., Sutcliffe, J.S., Skaug, J., Pinto, D., Vincent, J., Zwaigenbaum, L., Fernandez, B., Roberts, W., Szatmari, P., and Scherer, S.W. (2007). Contribution of SHANK3 mutations to autism spectrum disorder. Am. J. Hum. Genet. *81*, 1289–1297.
10. Jamain, S., Quach, H., Betancur, C., Rastam, M., Colineaux, C., Gillberg, I.C., Soderstrom, H., Giros, B., Leboyer, M., Gillberg, C., and Bourgeron, T. (2003). Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. Nat. Genet. *34*, 27–29.
11. Chubykin, A.A., Liu, X., Comoletti, D., Tsigelny, I., Taylor, P., and Sudhof, T.C. (2005). Dissection of synapse induction by neuroligins: effect of a neuroligin mutation associated with autism. J. Biol. Chem. *280*, 22365–22374.
12. Feng, J., Schroer, R., Yan, J., Song, W., Yang, C., Bockholt, A., Cook, E.H. Jr., Skinner, C., Schwartz, C.E., and Sommer, S.S. (2006). High frequency of neurexin 1beta signal peptide structural variants in patients with autism. Neurosci. Lett. *409*, 10–13.
13. 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.
14. Zhao, X., Leotta, A., Kustanovich, V., Lajonchere, C., Geschwind, D.H., Law, K., Law, P., Qiu, S., Lord, C., Sebat, J., et al. (2007). A unified genetic theory for sporadic and inherited autism. Proc. Natl. Acad. Sci. USA *104*, 12831–12836.
15. Gillberg, C. (1998). Chromosomal disorders and autism. J. Autism Dev. Disord. *28*, 415–425.
16. Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., Cook, E.H. Jr., Leventhal, B.L., and Pickles, A. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorders. J. Am. Acad. Child Adolesc. Psychiatry *45*, 1094–1103.
17. Falush, D., Stephens, M., and Pritchard, J.K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics *164*, 1567–1587.
18. Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., et al. (2006). Global variation in copy number in the human genome. Nature *444*, 444–454.
19. Zhao, X., Li, C., Paez, J.G., Chin, K., Janne, P.A., Chen, T.H., Girard, L., Minna, J., Christiani, D., Leo, C., et al. (2004). An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. Cancer Res. *64*, 3060–3071.
20. Nannya, Y., Sanada, M., Nakazaki, K., Hosoya, N., Wang, L., Hangaishi, A., Kurokawa, M., Chiba, S., Bailey, D.K., Kennedy, G.C., and Ogawa, S. (2005). A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. Cancer Res. *65*, 6071–6079.
21. Komura, D., Shen, F., Ishikawa, S., Fitch, K.R., Chen, W., Zhang, J., Liu, G., Ihara, S., Nakamura, H., Hurles, M.E., et al. (2006). Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays. Genome Res. *16*, 1575–1584.
22. Krawczak, M., Nikolaus, S., von Eberstein, H., Croucher, P.J., El Mokhtari, N.E., and Schreiber, S. (2006). PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. Community Genet. *9*, 55–61.
23. 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.
24. Zogopoulos, G., Ha, K.C., Naqib, F., Moore, S., Kim, H., Montpetit, A., Robidoux, F., Laflamme, P., Cotterchio, M., Greenwood, C., et al. (2007). Germ-line DNA copy number variation frequencies in a large North American population. Hum. Genet. *122*, 345–353.
25. Scherer, S.W., Cheung, J., MacDonald, J.R., Osborne, L.R., Nakabayashi, K., Herbrick, J.A., Carson, A.R., Parker-Katiraee, L., Skaug, J., Khaja, R., et al. (2003). Human chromosome 7: DNA sequence and biology. Science *300*, 767–772.
26. Kohler, J.R., and Cutler, D.J. (2007). Simultaneous discovery and testing of deletions for disease association in SNP genotyping studies. Am. J. Hum. Genet. *81*, 684–699.
27. Pinto, D., Marshall, C., Feuk, L., and Scherer, S.W. (2007). Copy-number variation in control population cohorts. Hum. Mol. Genet. *16*, 168–173.
28. Stranger, B.E., Forrest, M.S., Dunning, M., Ingle, C.E., Beazley, C., Thorne, N., Redon, R., Bird, C.P., de Grassi, A., Lee, C., et al. (2007). Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science *315*, 848–853.
29. Lupski, J.R. (2007). Genomic rearrangements and sporadic disease. Nat. Genet. *39*, S43–S47.

30. Bonnet, C., Gregoire, M.J., Brochet, K., Raffo, E., Leheup, B., and Jonveaux, P. (2006). Pure de-novo 5 Mb duplication at Xp11.22-p11.23 in a male: phenotypic and molecular characterization. J. Hum. Genet. *51*, 815–821.
31. Wong, W., Newell, E.W., Jugloff, D.G., Jones, O.T., and Schlichter, L.C. (2002). Cell surface targeting and clustering interactions between heterologously expressed PSD-95 and the Shal voltage-gated potassium channel, Kv4.2. J. Biol. Chem. *277*, 20423–20430.
32. Ensenauer, R.E., Adeyinka, A., Flynn, H.C., Michels, V.V., Lindor, N.M., Dawson, D.B., Thorland, E.C., Lorentz, C.P., Goldstein, J.L., McDonald, M.T., et al. (2003). Microduplication 22q11.2, an emerging syndrome: clinical, cytogenetic, and molecular analysis of thirteen patients. Am. J. Hum. Genet. *73*, 1027–1040.
33. 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.
34. Jacquemont, M.L., Sanlaville, D., Redon, R., Raoul, O., Cormier-Daire, V., Lyonnet, S., Amiel, J., Le Merrer, M., Heron, D., de Blois, M.C., et al. (2006). Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders. J. Med. Genet. *43*, 843–849.
35. Feuk, L., Carson, A.R., and Scherer, S.W. (2006). Structural variation in the human genome. Nat. Rev. Genet. *7*, 85–97.
36. Skuse, D.H. (2007). Rethinking the nature of genetic vulnerability to autistic spectrum disorders. Trends Genet. *23*, 387–395.
37. Zoghbi, H.Y. (2003). Postnatal neurodevelopmental disorders: meeting at the synapse? Science *302*, 826–830.
38. Mitchell, K.J. (2007). The genetics of brain wiring: from molecule to mind. PLoS Biol. *5*, e113. 10.1371/journal.pbio.0050113.
39. 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.
40. Sharp, A.J., Selzer, R.R., Veltman, J.A., Gimelli, S., Gimelli, G., Striano, P., Coppola, A., Regan, R., Price, S.M., Knoers, N.V., et al. (2007). Characterization of a recurrent 15q24 microdeletion syndrome. Hum. Mol. Genet. *16*, 567–572.
41. Somerville, M.J., Mervis, C.B., Young, E.J., Seo, E.J., del Campo, M., Bamforth, S., Peregrine, E., Loo, W., Lilley, M., Perez-Jurado, L.A., et al. (2005). Severe expressive-language delay related to duplication of the Williams-Beuren locus. N. Engl. J. Med. *353*, 1694–1701.
42. Yu, C.E., Dawson, G., Munson, J., D'Souza, I., Osterling, J., Estes, A., Leutenegger, A.L., Flodman, P., Smith, M., Raskind, W.H., et al. (2002). Presence of large deletions in kindreds with autism. Am. J. Hum. Genet. *71*, 100–115.