Neuron
Article

Rare De Novo and Transmitted Copy-Number Variation in Autistic Spectrum Disorders

Dan Levy,[1] Michael Ronemus,[1] Boris Yamrom,[1] Yoon-ha Lee,[1] Anthony Leotta,[1] Jude Kendall,[1] Steven Marks,[1] B. Lakshmi,[1,4] Deepa Pai,[1,5] Kenny Ye,[2] Andreas Buja,[3] Abba Krieger,[3] Seungtai Yoon,[1,6] Jennifer Troge,[1] Linda Rodgers,[1] Ivan Iossifov,[1] and Michael Wigler[1,*]
[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA
[2]Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY 10461, USA
[3]Department of Statistics, the Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA
[4]Present address: Ontario Institute for Cancer Research, Toronto, Ontario, Canada M5G 0A3
[5]Present address: Mumbai, India
[6]Present address: Department of Psychiatry, Mount Sinai School of Medicine, New York, NY 10026, USA
*Correspondence: wigler@cshl.edu
DOI 10.1016/j.neuron.2011.05.015

SUMMARY

To explore the genetic contribution to autistic spectrum disorders (ASDs), we have studied genomic copy-number variation in a large cohort of families with a single affected child and at least one unaffected sibling. We confirm a major contribution from de novo deletions and duplications but also find evidence of a role for inherited "ultrarare" duplications. Our results show that, relative to males, females have greater resistance to autism from genetic causes, which raises the question of the fate of female carriers. By analysis of the proportion and number of recurrent loci, we set a lower bound for distinct target loci at several hundred. We find many new candidate regions, adding substantially to the list of potential gene targets, and confirm several loci previously observed. The functions of the genes in the regions of de novo variation point to a great diversity of genetic causes but also suggest functional convergence.

INTRODUCTION

Autism spectrum disorders (ASDs) are among the most genetically determined of developmental and cognitive abnormalities, with concordance between identical twins reported at nearly 90% in some studies (Muhle et al., 2004; Rosenberg et al., 2009). There is a strong gender bias, with much higher incidence in males than in females, especially for higher-functioning children (Newschaffer et al., 2007). Previous studies found a higher incidence of new copy-number mutation in autistic children from simplex (only one affected child) ASD families than in typical children or in children from multiplex (multiple affected children) ASD families (Marshall et al., 2008; Sebat et al., 2007; see also Itsara et al., 2010). Based on these earlier findings, we proposed a role for new (or de novo) germline variation in simplex families, distinct from transmitted variation that might predominate in multiplex families. Similar findings have been reported for sporadic and inherited schizophrenia (Xu et al., 2008). Further analysis of the incidence of male probands in multiplex families led us to derive a risk function for the population and to propose that much of ASD arises from de novo variants of strong penetrance and that some de novo variants of high penetrance are transmitted by relatively asymptomatic carriers in a dominant fashion (Zhao et al., 2007).

In a continuing effort to explore ASDs and to reveal the targets of mutation, we have participated in a large study of simplex families: the Simons Simplex Collection (SSC), consisting of approximately 1000 families at the time of this analysis (Fischbach and Lord, 2010). Families with only a single child on the spectrum were recruited. In nearly all cases there was at least one unaffected sibling, and multiplex families were specifically excluded. We analyze copy-number variation (CNV) in SSC families by comparative genomic hybridization (Iafrate et al., 2004; Sebat et al., 2004), using the NimbleGen HD2 2.1 million probe microarray platform (http://www.nimblegen.com/products/cgh/wgt/human/2.1m/index.html) with oligonucleotides optimized for both hybridization performance and uniform genome coverage. An accompanying paper (Sanders et al., 2011) also reports findings on the same set of SSC families using a similar approach but different CGH platforms, the Illumina 1M and 1M Duo microarrays.

Due to the size of the present study, we are better able than before to assess the contribution of de novo CNVs to autism. Because this study utilizes a CGH platform with greater than twice than the number of unique probes than earlier published studies of similar family number (Pinto et al., 2010), we can in theory detect smaller regions of variation. Both de novo deletions and duplications contribute substantially to ASDs, and overall we find a greater number of regions at a finer scale than was previously possible. We also establish and estimate the contribution of the transmission of "ultrarare" variants to ASDs, in particular inherited duplications. These findings add substantially to the list of ASD candidate genes. Our results also reveal the gender bias of autism in greater depth and raise the puzzle of the fate of female carriers. By considering the number and proportion of variant loci that are recurrent, we are able to give a lower bound on the total number of target loci where copy-number

mutation can lead to the disorder. The functions of some of the genes in the de novo rare and ultrarare variation are considered briefly here and assessed in greater depth in an accompanying paper (Gilman et al., 2011).

The focus of this work is on rare events, in fact, "rare" almost to the point of uniqueness within the cohort. There are good reasons for this, both theoretical and practical (Xu et al., 2008; McClellan and King, 2010). The hypothesis that autism results from an unfortunate combination of common low-risk variants (Wang et al., 2009; Weiss et al., 2009) can be safely rejected. More generally, it flouts reason to believe that mutations of high penetrance would ever be common for a disorder that so drastically reduces fecundity. On the other hand, all genomes are under mutational pressure, and so constantly give rise to many variants that will be under strong negative selection. Some of this negative selection will not be readily apparent, occurring in utero. The rest will manifest as infant mortality and disorders of childhood (such as ASDs) and early adolescence. Each individual variant will be rare—extremely so—as it expands in the population only if it offers some compensatory advantage.

Table 1. Composition of Sample Subpopulations

	Total	Valid	HQ	HQ Quads
family	915	887	787	510
children	1784	1721	1475	1020
autM	772	747	641	443
autF	117	111	93	67
sibM	428	414	343	238
sibF	467	449	398	272
aut	889	858	734	510
sib	895	863	741	510

CGH was performed on 915 families, with a total of 1784 children (probands and unaffected siblings). Families with pedigree and/or gender mismatches were removed, leaving 887 families ("valid") comprising 1721 trios (defined as both parents and a single proband or unaffected child). To facilitate some computations, a restricted set of 1475 trios composed of only high-quality ("HQ") hybridizations was used. For another subset of computations, only HQ families with exactly one affected child and one unaffected sibling ("HQ quads") were considered. The subset of families used in each analysis is indicated where appropriate. Children were tabulated by affected status and gender (e.g., "autM," "autF," "sibM," and "sibF").

RESULTS

Study Design and Data Processing

The Simons Simplex Collection is being assembled at 13 clinical centers, accompanied by detailed and standardized phenotypic analysis. An ongoing study of the correlations between our genetic findings and the phenotypic data is not included in the present study. Families with single high-functioning probands, usually with unaffected siblings, are preferentially recruited, and families with two probands are specifically excluded (Fischbach and Lord, 2010). Bloods, drawn from parents and children (affected and unaffected), are sent to the Rutgers Cell Repository (RUCDR) for DNA preparation. Blood DNAs (and a few rare cases, EBV-immortalized DNAs) from nearly 1000 families (of the 3000 planned) were sent to our group for processing and analysis. Approximately one-tenth of the families we analyzed are not yet officially in the SSC databases. DNA samples were shipped to NimbleGen's Icelandic facility, where two-color hybridizations using a single reference male genome were performed. SSC samples were labeled with Cy3, and the reference was labeled with Cy5. Ninety-seven percent of families passed gender and pedigree checks for all members and are called "valid" herein. Those are the only families considered in this report. We define a trio as consisting of a mother, a father, and a child, either affected or unaffected. If each member of a trio has a hybridization that passes minimum quality thresholds (see Experimental Procedures), that trio and its associated hybridizations are called "high quality" (or "HQ"). Out of 1721 valid trios from 887 families, 1475 (86%) are HQ. For convenience, throughout this report we refer to the children with diagnosed ASDs as "probands" and to the children who do not have ASDs as "sibs." For purposes of statistical evaluation, we establish the "HQ quads," a subset of 510 HQ families with exactly one proband and one sib each.

The composition of the children and families for the various subpopulations under study is summarized in Table 1. There are roughly equal proportions of probands and sibs. The male-to-female ratio among the probands is 7:1, typical of high-functioning ASDs (Newschaffer et al., 2007). We mention here the observation (to be discussed later) that there are fewer male sibs than female sibs.

Hybridization data underwent extensive processing before determining segments of altered copy number (Experimental Procedures, Supplemental Experimental Procedures, and Figure 1). We extracted signal and noise parameters from each hybridization and used these for quality control and to model integer copy-number states (Figure 2). For partitioning the genome into intervals of constant copy number, we used KS segmentation (Grubor et al., 2009). We also employed a trio-based Hidden Markov Model (HMM) to build databases of high-confidence events and transmissions. High-confidence events in 1500 parents were used to compile a frequency table of copy-number variation for all probes. We searched for de novo events in the 1475 HQ trios, initially restricting evidence to autosomal probes that did not have known extra mappings to the human genome (hg18 build) outside the event region, and probes that were rarely polymorphic in the high-confidence parental database (i.e., present in no more than 5/1500 parents). We compiled those events with high statistical significance of being de novo (p value $< 10^{-9}$), creating a "stringent" automated list of 70 de novo events (Table S1, "stringent"). Figure 3 illustrates the family probe ratio data for two typical de novo events, a duplication and a deletion.

We then relaxed these probe restrictions to consider lower-quality trios, probes on the X chromosome, probes with higher frequencies of polymorphism (but never more than 20/1500 parents), and de novo events of lower significance (p value $< 10^{-7}$). We curated the resulting list, accepting 24 additional de novo events, creating a "relaxed" manual list (Table S1, "relaxed"). All events on the stringent list passed manual inspection and are included in the "relaxed" list. We sent samples for

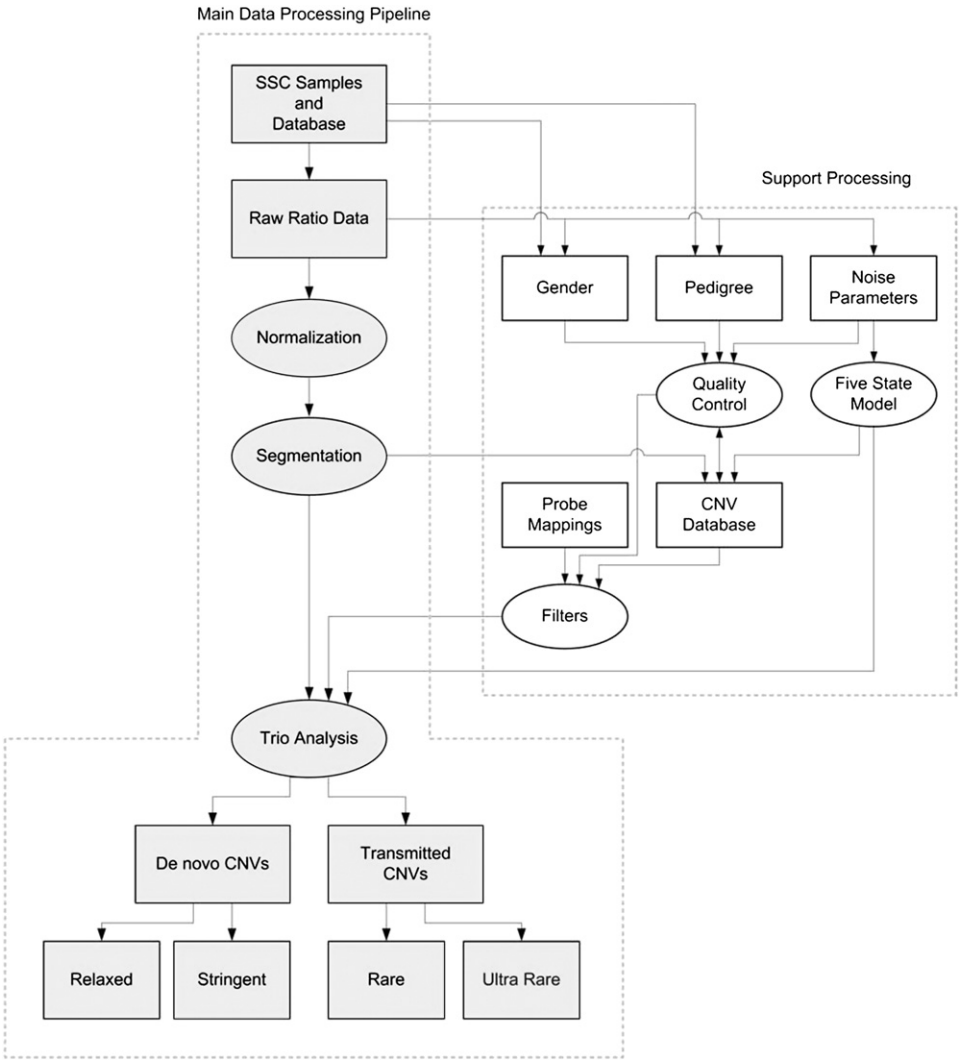

Figure 1. Data Processing Pipeline

The main data processing pipeline for identifying de novo and transmitted rare copy-number variants in the SSC is shown in gray at left, and the support pipeline, which includes quality control, is shown in white at right. Data sets are represented by boxes and computational processes by ovals. Raw ratio data were normalized and segmented prior to trio analysis. Supporting processes were used to determine hybridization quality, to exclude gender and pedigree mismatches, to establish probe filters, and to build models for various copy-number states.

validation by high-resolution CGH on Agilent 244K tiling arrays (Supplemental Experimental Procedures, Tables S1 and S2, and Figure S1), and 54/54 of the successfully completed hybridizations of trios confirmed calls of de novo events, giving us high confidence that these calls are true positives. We have even higher confidence on transmitted events, because of additional evidence, namely the presence of the event in both a parent and a child with nearly identical boundaries.

De Novo Variation

Our observations regarding de novo events are summarized (Table 2), and the events themselves are detailed individually (Table S1). In total, we observed 75 de novo events in 68 probands (7.9% of all probands) and 19 events in 17 sibs (2.0% of all sibs). These observations are consistent with the findings of previous studies that probands have a higher burden of de novo copy-number mutations (Marshall et al., 2008; Sebat et al., 2007). We also observe that females with ASDs have a higher frequency of de novo events than males (11.7% versus 7.4%, p value = 0.16) and that de novo deletions are more frequent than duplications in male probands (39 to 22, p value = 0.04).

We also looked at these data from the standpoint of gene "hits" (Table 3). We used RefSeq for gene and exon information, omitting snRNAs. A CNV is considered to "hit" a gene when at least one exon of the gene overlaps the CNV. Of the 75 de novo events in probands, 61 hit genes, as did nine of the 19 events in sibs (p value = 0.006). There were a total of 953 genes hit in de novo events in probands but only 59 in sibs. The difference was overwhelming when we looked only at genes involved

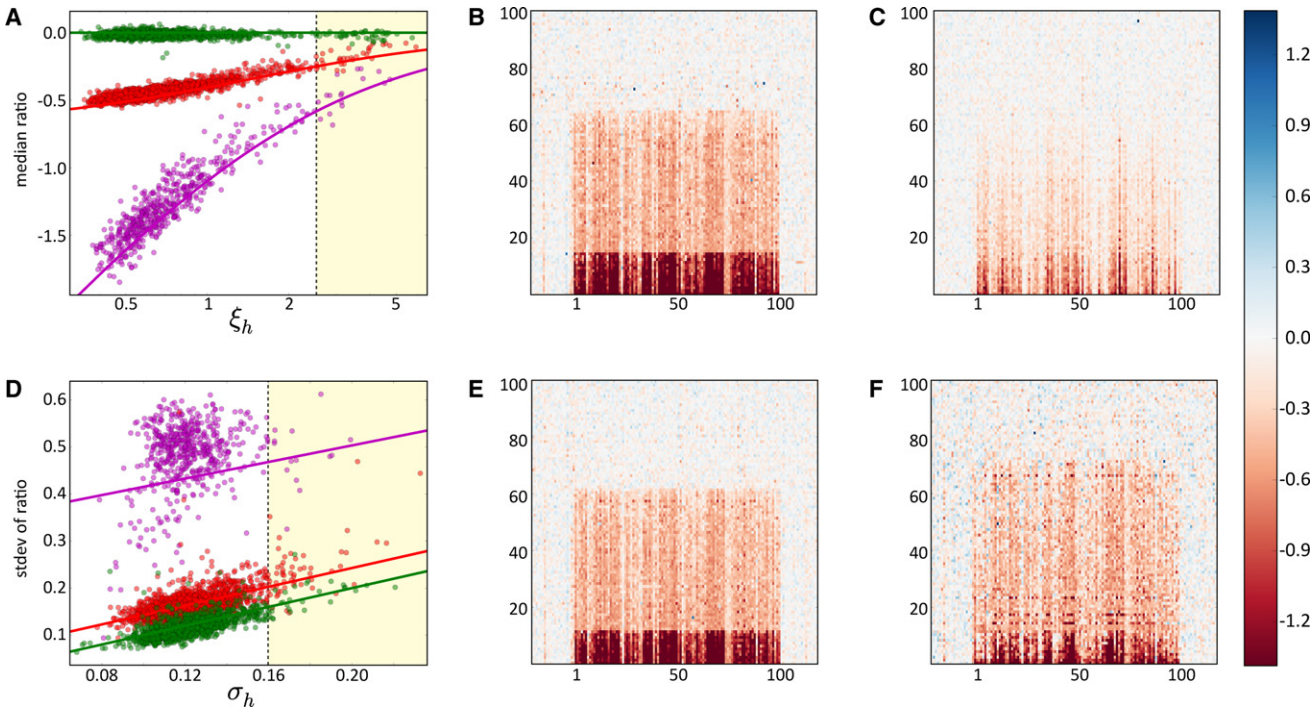

Figure 2. Noise Parameters and State Calls

Regions of common copy-number polymorphism in 3653 hybridizations were examined to validate the performance of signal and noise parameters. A commonly deleted region of 150 kb (representing a total of 99 probes) is displayed, along with the adjacent 20 probes from each flank.

(A) Median log ratio on the interval as a function of the signal parameter ξ_h. Magenta, red, and green curves correspond to the mean as predicted by the noise model for the three copy-number states 0, 1, and 2, respectively. The highlighted region to the right of the dashed line shows the values obtained in the 100 worst hybridizations.

(B and C) Probe log ratio values on the region for a subset of hybridizations. Experiments are sorted by the median probe ratio value of the polymorphic region. (B) shows 100 hybs selected at random from the left side of (A). (C) shows the 100 worst hybs from the highlighted (rightmost) region of (A).

(D)–(F) are as in (A)–(C) for the noise parameter σ_h.

in deletions: 534 in probands and two in sibs (Table 3, Figure 4). De novo events in probands typically involved many more genes than de novo events in sibs. Another disparity was evident by gender; more genes were present in events from female probands than in those from male probands. The median number of genes in a de novo event in a female proband was 15.5, but only 2.0 in males, with a high significance (p value = 0.05) as determined by a rank-sum permutation test. All genes hit by de novo events, whether in a proband or a sib, are listed in detail in Table S3.

Most de novo events were unique. There were, however, 16 events in probands that overlapped at four distinct loci (Table S4). We define events as "overlapping" if either the events intersect or they both hit the same gene, and we refer to the locus at which they overlap as a "recurrent" locus. Ten of these overlapping events occurred at 16p11.2, both as deletions (six) and as duplications (four). The incidence of 16p11.2 copy-number events we observe is consistent with previous studies (Mefford et al., 2009; Pinto et al., 2010; Weiss et al., 2008). Out of the ten 16p11.2 CNVs, all but one occur in a male, and—assuming the incidence of the mutation is equal in males and females—this observation implies gender bias in penetrance of the ASD phenotype at this locus. We observe a single instance of transmission of the 16p11.2 deletion from a mother (family 12010, Table S5). Inheritance of duplications at this locus had been previously reported in a number of cases of ASDs, but there are relatively few reports of transmitted deletions linked to ASDs (Bijlsma et al., 2009; Fernandez et al., 2010; Shinawi et al., 2010).

A notable recurrence occurs at 7q11.23, as a duplication at the Williams-Beuren Syndrome (WBS) locus. Deletion at this locus is associated with mental retardation, and—in contrast to ASDs—the deletion is characterized by precocious verbal ability, avid eye contact, and a highly sociable disposition (Merla et al., 2010). A third recurrent locus at 16p13.2 contains *USP7* (encoding a deubiquitinase), an intriguing finding given existing links between *USP7* and the spinocerebellar ataxia type 1 (Hong et al., 2002). The fourth occurs at the *NIPA* locus on 15q11.2, a region in which CNVs have been previously associated with ASDs as well as epilepsy and schizophrenia (de Kovel et al., 2010; Mefford et al., 2010; Stefansson et al., 2008; van der Zwaag et al., 2010). For reasons discussed later, we discount the significance of this recurrence. A fifth recurrence, de novo duplication on 16p13.11, occurs in both a proband and a sib from different families. This is a known locus of instability, wherein deletions but not duplications are thought to be associated with cognitive problems (Hannes et al., 2009). We observe

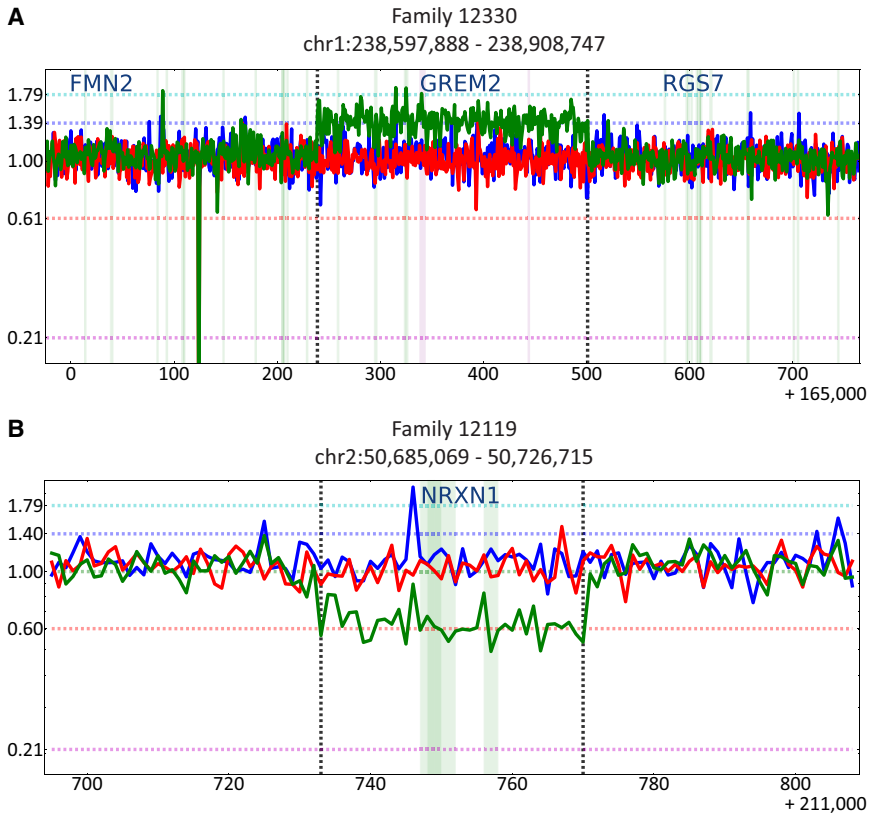

Figure 3. Ratio Data for De Novo Events

(A and B) Two representative de novo events, one duplication (A) and one deletion (B). In each panel, the mother's probe log ratio values appear in red, the father's in blue, and the proband's in green. Exons are shaded in green and purple, with colors alternating by gene (with gene names indicated in blue). Dashed horizontal lines show the mean ratios for the copy-number states 0, 1, 2, 3, and 4, as predicted by the noise parameters of the proband's hybridization. The boundaries of the deletion are indicated by the dashed vertical lines. The x axis gives the probe coordinates as an offset to the number in the lower right of the respective panel.

a single rare transmission of a deletion at this locus to a child with ASD (family 11450, Table S5).

To study recurrence further, we looked at a recently published study, which reported 56 de novo events in a mixed set of simplex and multiplex autism trios (Pinto et al., 2010). We omit ten of these because (1) they occur within regions that are commonly polymorphic in our cohort; and (2) when these regions are polymorphic, the polymorphisms are transmitted without bias to probands and sibs (Table S6). Of the remaining 46 events from that report, 12 events overlapped our set of de novo events in probands, at six distinct loci. Counting both data sets in total, recurrent de novo events were observed at 12 distinct loci (Table S4, Figure 5).

Transmitted Variation

Collections of families with multiple affected children will be significantly enriched for families at high risk of transmission. The SSC was designed to exclude such multiplex families, maximizing chances to discover de novo mutations. However, as most families are not large, it must be expected that the SSC is far from entirely depleted for families at high risk. In fact, we estimate that fewer than 15% of such high-risk families with two children, at least one with autism, would be excluded by the study design, and fewer than 30% of those with three children. Accordingly, we studied the SSC families for evidence of transmitted risk factors.

We saw no statistically significant difference between probands and sibs when we looked at total numbers of transmitted copy-number events or the numbers of genes hit by transmission (Table S7). We explored evidence for transmission distortion of many individual common copy-number polymorphisms, but found no statistically significant signal when adjusted for multiple hypotheses (data not shown).

Table 2. Distribution of De Novo CNVs by Polarity and Gender

	Counts of De Novo Events			Children with De Novo Events			Frequency in Children		
	Combined	Del	Dup	Combined	Del	Dup	Combined	Del	Dup
autM	61	39	22	55	37	21	7.4%	5.0%	2.8%
autF	14	7	7	13	7	6	11.7%	6.3%	5.4%
sibM	10	3	7	9	3	6	2.2%	0.7%	1.4%
sibF	9	6	3	8	5	3	1.8%	1.1%	0.7%
aut	75	46	29	68	44	27	7.9%	5.1%	3.1%
sib	19	9	10	17	8	9	2.0%	0.9%	1.0%

In the set of valid trios (see Table 1), 75 discrete de novo CNVs were observed in 68 probands (7.9%), whereas 19 de novo events were detected in 17 unaffected siblings (2.0%). The frequency of de novo CNVs in affected females is 11.7%, as compared to 7.4% in the affected males. De novo events in affected females are evenly split between deletions and duplications (7:7), but in affected males, de novo deletions are nearly twice as frequent as duplications (39:22).

Table 3. Breakdown of De Novo Gene Disruptions by Polarity, Gender, and Gene Count

	Gene Disruption Count			Median Genes per Event			Median Genes per Child		
	Total	Del	Dup	All	Del	Dup	All	Del	Dup
autM	650	409	241	2	1	5.5	3	1	6
autF	303	125	178	15.5	9	26	19	9	27
sibM	33	0	33	1	0	4	2	0	4
sibF	26	2	24	0	0	7	0.5	0	7
aut	953	534	419	4	2	7	4	2	8
sib	59	2	57	0	0	4	1	0	4

Genes were considered "disrupted" when at least one exon of that gene overlapped a de novo deletion or duplication. From the set of valid families, a total of 953 genes were disrupted in probands. Only 59 genes were disrupted in the unaffected siblings, all but two of which were within duplication events. Median gene counts were computed by event and by child. There was a significant disparity between the median genes per de novo CNV in affected males (2) compared to affected females (15.5). Also notable in both affected males and females was the smaller median number of genes disrupted in de novo deletions than in duplications.

However, a role for transmission can be seen if carefully restricted to extremely rare events (Xu et al., 2008). We limited ourselves to the 510 HQ quads: families with high-quality data and exactly one affected and one unaffected child. To minimize false signal, we considered only events of at least 20 probes. Operationally, we define the "family hit count" for each RefSeq gene as the number of families in which we observe a transmitted event that overlaps an exon of that gene. In Table S8 we list all genes with a positive family hit count and provide counts for each time a given gene had an exon within an event transmitted to a sib or a proband.

We define an "ultrarare gene" as a gene with a family hit count of one and then define an ultrarare event as an event that overlaps at least one exon of an ultrarare gene. In other words, an ultrarare event is one that hits at least one gene that is not hit by any other transmitted event over the population of HQ quads. The 458 ultrarare events are summarized in Table S9. These events are further characterized by the gender of the recipient children, their affected status, and by the pattern of transmission ("singly" transmitted, either to a proband or a sib, or "doubly" transmitted, to both). Additional features are listed, such as parent of origin, the ultrarare genes overlapping the event, and whether the event is a duplication or deletion. One strong asymmetry is between the counts of ultrarare deletions (178) and duplications (432), in excess of the overall bias in all transmitted events (3119 deletions versus 3875 duplications, p value = 3×10^{-15}). Note that this bias for duplications is the opposite bias seen for de novo events in male probands, for which deletions exceed duplications.

For singly transmitted ultrarare events, we find a slight excess of events going to the proband rather than the sib (Table 4). The signal is even stronger when we consider the number of ultrarare gene hits (p value = 0.23 for events, p value = 0.13 for genes). We see no bias in families with female siblings, in keeping with the hypothesis that females are less likely to display the symptoms of ASDs. The entirety of the bias for singly transmitted events is in those quad families for which the unaffected sibling is male (p value = 0.05 for events, p value = 0.01 for genes). Moreover, doubly transmitted events occur more often when the sib is female than male (p value = 0.09 for events, p value = 0.02 for genes). Recalling that the SSC cohort excluded families with two affected children, these transmission biases are joint and independent evidence that there are fewer transmissions of ultrarare events to a male sib than to the autistic child in our cohort and support the hypothesis that a portion of ultrarare transmission events are causal in males.

There appears to be no gender bias in the parent of origin of ultrarare events. Overall, the sources of transmissions of ultrarare events were 233 from the fathers and 223 from the mothers. For events that were transmitted but not to the unaffected male siblings, the sources were 125 from the fathers and 125 from the mothers. The possible implications for this observation will be discussed later.

Combining Data for De Novo and Transmitted CNVs

We combined evidence from all CNVs, exploring transmitted events that overlap de novo events (Table S3). We also compiled lists of transmitted events with boundaries similar to those found in de novo events (Tables S5 and S10). Because ultrarare transmitted events and de novo events are sparse data sets, we cannot expect to draw strong conclusions for specific loci by combining these data. Rather, in these tables one can find anecdotal information that informally raises or lowers the suspicion that various loci are contributory. For example, transmission

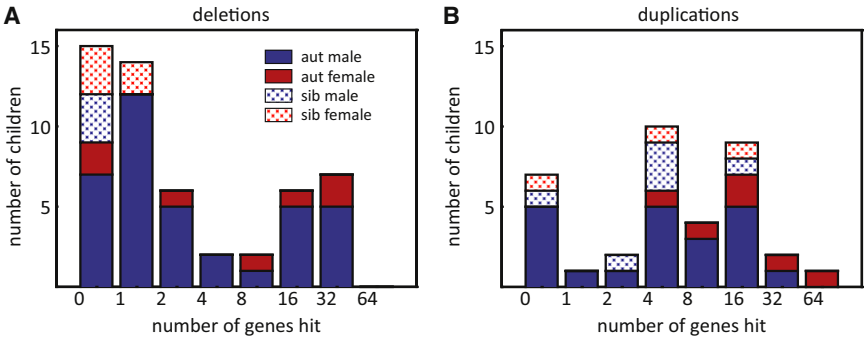

Figure 4. Histogram of De Novo Gene Hits

Eighty-five children in the study had detectable de novo lesions: 33 had duplications, 49 had deletions, and three had both a deletion and a duplication. We determined the number of gene hits per child and aggregated by affected status and gender. The boundaries of the histogram bins are log scaled.

(A) Distribution of de novo deletions. Gender and affected status are shown in the box at upper right.

(B) Distribution of de novo duplications. Gender and affected status are as in (A).

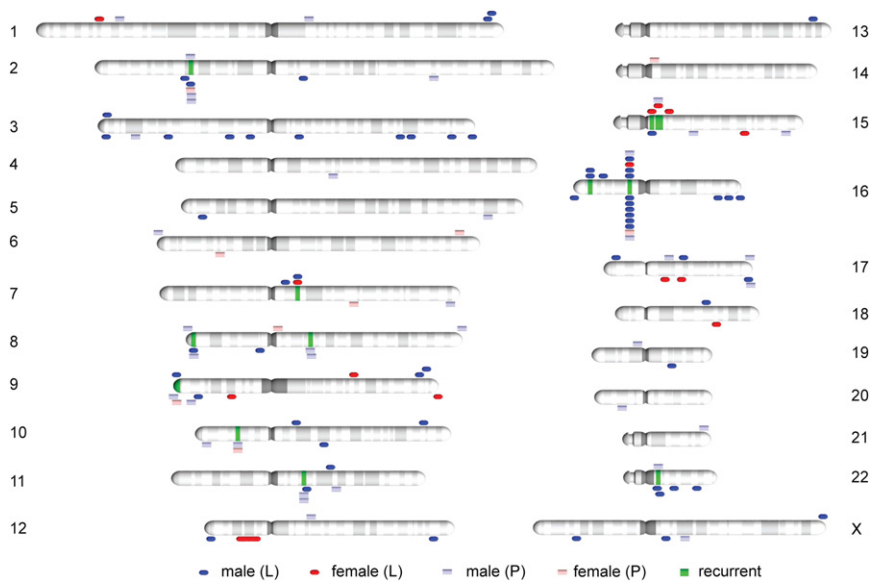

Figure 5. Distribution of De Novo Events in ASD Study Populations

A total of 75 de novo copy-number variants in autistic children were found in this study (designated by "L" in the legend). These observations were combined with the set of de novo CNVs identified in an earlier study (indicated by "P") of similar scale (Pinto et al., 2010). Ten of the CNVs from the earlier study appeared to be common variants against the background of the SSC, and consequently were omitted (Table S6). Chromosomes are pictured with alternating dark and light regions for each cytoband and aligned by centromeres. Duplications are pictured above and deletions below the respective chromosomal locus. Events in males and females from the SSC study ("L") appear in blue and red, respectively. Events in males and females from the Pinto et al. (2010) study ("P") appear in light blue and light red, respectively. Green bars spanning chromosomes indicate the positions of the 12 loci with recurrent de novo events.

data raise the suspicion for *USP7*, *CTNNA3*, and genes encoding several related voltage-gated calcium channels (see next section) but diminish suspicion for the loci at *NIPA* (15q11.2) and *NPHP1* (2q13). The latter loci appear to be mainly unstable, and parental variants transmit equally to sibs and probands.

Rare Homozygous Deletions

Although our focus has been on rare variants that contribute to phenotypes in a dominant fashion, it has been documented that some autism can result from the combined action of recessive alleles (Morrow et al., 2008). Therefore, we scanned the genomes of probands and sibs looking for rare homozygous deletions that hit both alleles. Two were found, both occurring in probands (Figure S2). One disrupted *COMMD1* (2p15) in a female. Homozygous loss of this gene is implicated in copper toxicosis in dogs (van De Sluis et al., 2002) but has not been previously reported in humans. The deletion initially appeared as a de novo event; the father, but not the mother, carried a hemizygous deletion. However, the boundaries of the homozygous loss in the child matched those of the hemizygous loss in the father precisely, which raised our suspicion that the child had an instance of a rare but known occurrence of uniparental disomy of chromosome 2 (Kotzot and Utermann, 2005). Further analysis of chromosome 2 of the proband revealed multiple small regions of homozygosity and heightened our suspicion, which was eventually confirmed (M. State, personal communication). The event is not counted in our tables or statistics of de novo events because there may be other homozygous recessive mutations elsewhere on the father's chromosome 2 that are not copy-number variants.

The second rare homozygous deletion occurred in a male proband and disrupted *CACNA2D4* (12p13.33). Both parents were in the hemizygous state. This gene encodes a voltage-dependent calcium channel. Although there are no known autism-related phenotypes associated with homozygous mutations in *CACNA2D4* (Wycisk et al., 2006), defects in *CACNA1C* are known to be the basis of Timothy syndrome, a rare disorder with symptoms including autism. We observe a de novo two-gene deletion disrupting *CACNA1B*, another voltage-gated calcium channel, and a transmission of a rare variant of *CACNA1C* (a disruptive intragenic duplication) in one family.

Table 4. Statistics on Ultrarare Transmissions

	Single Transmissions			Double Transmissions		
	Aut	Sib	P Value	QuadF	QuadM	P Value
All HQ Quads						
event	157	142	0.23	94	65	0.09
child	128	113	0.15	77	59	0.21
gene	359	293	0.13	276	122	0.02
QuadF						
event	75	81	0.64			
child	63	63	0.54			
gene	150	176	0.73			
QuadM						
event	82	61	0.05			
child	65	50	0.06			
gene	209	117	0.01			

Ultrarare events (see text and Table S9) from HQ quads were tabulated and classified as either single transmissions (to one child only, whether affected or unaffected) or double transmissions (to both children). All HQ quad transmissions were further broken down into family type by the gender of the unaffected sibling ("QuadF" and "QuadM"). Counts are given for number of events, number of recipient children and number of ultrarare gene disruptions. The p values were computed from 10,000 permutations and showed the strongest disparity for singly transmitted events in families with male siblings. Doubly transmitted events occurred disproportionately in families with female siblings.

DISCUSSION

We find de novo events in 8% of children with ASDs and only in 2% of their unaffected siblings, in keeping with other reports (Marshall et al., 2008; Sebat et al., 2007). The observed frequency of de novo events in children with autism from simplex families that we observe in our present study is slightly lower than that observed in our previous study (10%), despite the fact that our discovery tools are much more powerful than before (Sebat et al., 2007). This may be related to ascertainment biases in the two studies. The simplex population from the first study may have been based on larger families with a single proband, and so may have had fewer cryptic multiplex families than are undoubtedly present in the current study. Also, the present study is biased to higher-functioning probands, and as a consequence, there is a lower ratio of female probands than in our earlier study. Observable de novo events are more frequent in females, so the first study—which recruited a higher proportion of females—contained a higher proportion of children with observable events. Finally, the first study was smaller, and the de novo events were not filtered with the same exacting care as in the present study.

It is reasonable to infer that most of de novo copy-number mutations are at least contributory to the disorder. Taken in isolation, the observation is also compatible with another explanation: that de novo mutation is evidence of genome instability, the actual underlying causal condition. However, the latter view is not consistent with a decreased association of de novo mutation in multiplex autism, nor with additional observations made in this report, namely duplication-deletion imbalances, frequency and size imbalances in the de novo events by gender, bias in transmission of ultrarare copy-number variation to probands, and bias in transmission by gender.

To help form a genetic theory of the basis of autism, we find it useful to provide a summary of observations in the form of lists of observed biases, or "asymmetries." These observations are derived from both this study and the previous literature (Marshall et al., 2008; Pinto et al., 2010; Sebat et al., 2007; Zhao et al., 2007). We organize these observations, each made with varying degrees of confidence, as follows. (1) There is a higher incidence of de novo copy-number mutation in children with ASDs from simplex families than in their siblings. (2) There is a higher incidence of de novo copy-number mutation in children with ASDs from simplex families than in children with ASDs from multiplex families. (3) For transmitted rare variants, duplications greatly outweigh deletions. (4) Deletions outweigh duplications in de novo events in children with ASDs. (5) There is evidence of transmission distortion for ultrarare events to children with ASDs, and (6) this bias arises from families in which the sibling is an unaffected male. (7) Females are less likely to be diagnosed with ASDs than are males. (8) A higher proportion of females with ASDs have detectable de novo copy-number events than do males with ASDs, and the events are larger.

The asymmetries are readily explained by a plausible genetic theory. De novo mutation of high penetrance contributes to autism, more so in families of low risk than in families at high risk. In the latter, transmission genetics plays a greater role. Deletions are generally more likely to be harmful than duplications. By selection, a mutation of recent vintage but carried by an unaffected parent is thus more likely to be a duplication. Females appear to be more resistant than males to developing ASDs, and large-copy-number events are observed more frequently in affected females because such events are more harmful, because there are fewer target genes that induce ASDs in females than males, or both; see also the accompanying paper by Gilman et al. (2011) for independent evidence. Sexual dimorphism in brain development may explain the relative lack of females with ASDs. Relative to males, females have an accelerated timescale for a number of cognitive milestones; for example, generally speaking their first words at an earlier age (Richler et al., 2010; Roze et al., 2010). A quicker pace of development might reflect a robustness that offers females protection.

There is one asymmetry that is conspicuous by its absence, a puzzle buried in the transmission data. If females are resistant to ASDs and children with ASDs have reduced fecundity, then simple genetic theory predicts that mothers would be more the likely sources of a risk allele than fathers. But we see no bias in the parent of origin among transmitted ultrarare events. However, we cannot reject such a hypothesis based on the observed data. There is insufficient power under reasonable assumptions of the rate of observable contributory transmitted CNVs (7%) and a strong bias toward transmission from mothers of contributory events (75%). Moreover, we lack a longitudinal study of high-functioning children with ASDs and cannot know that males will display reduced fecundity. Nevertheless, the fate of females with a risk genotype remains a puzzle worth contemplating. Females with higher-risk genotypes may encounter difficulties at later stages of their lives that manifest as a different diagnostic category, or that reduces fecundity. If true, the disorder would most likely be one with a gender bias opposite that of ASDs, such as anorexia nervosa (Fairburn and Harrison, 2003).

Our genetic theory of autism, as discussed above, largely depends on dominant acting genetic variants of variable penetrance. We think the theory is sufficient to explain most of the genetic basis of autism, both simplex and multiplex, but certainly not all. For example, the role of recessive mutations in individuals from consanguineous marriages has been demonstrated (Morrow et al., 2008). We have observed only a single case of inheritance of a rare homozygous null state.

A striking finding of all the studies of de novo mutation in children with ASDs is the apparent number of distinct target loci. Even discounting 25% of events as incidental (based on a 2% frequency in sibs and 8% in probands), there are a large number of target regions and few recurrences. Only CNVs at 16p11.2 are present in more than 1% of cases (ten out of 858 children). We can make an estimate of the minimum number of target regions by analysis of recurrence. Combining two large studies (ours and that of Pinto et al., 2010), we observe 39 overlaps at 12 recurrent loci in 121 events. Excluding the highly recurrent 16p11.2 locus (with 13 hits in the combined dataset) and discounting one-quarter of the remaining 108 events as incidental, we observe 11 recurrent loci in approximately 80 causal events. If we assume a uniform rate of copy-number mutation, we estimate the number of target loci at 250–300. However,

targets do not have a uniform rate of copy-number mutation, so this figure would be an underestimate of total targets. We derive a similar estimate for target size by a completely different method, based on many assumptions including the rate of new mutations that damage a gene in humans (about one gene per three births), the incidence of ASDs among males (approximately 1 in 100), a genetic model that predicts that about half of ASDs result from new mutations (Zhao et al., 2007), and high penetrance of a select set of single mutational hits. The latter assumption is based on the observation of dominant transmission in multiplex families (Zhao et al., 2007).

An organism will be vulnerable to a single mutational hit at only a small subset of its genetic elements. We imagine that vulnerable targets may arise by two distinct cellular mechanisms: insufficient or uncorrectable dosage compensation resulting from (for example) altered stoichiometry of protein complexes; and monoallelic gene expression, which could result in subpopulations of functionally null neurons, perhaps confined to specific subtypes (Gimelbrant et al., 2007; Gregg et al., 2010). Many more gene products undoubtedly function on pathways related to the vulnerable target genes, but these genes will very rarely be found in genetic screens of the proband because of either functional redundancy or the robustness of dosage compensation. The conclusion that there are a huge number of potential targets for ASDs is all but unavoidable.

Despite the large number of target loci we identify and the small number of recurrent loci detected in this analysis, several of the events that we find supplement previous studies. For example, *NRXN1* (encoding neurexin 1) is a well-established candidate gene underlying ASDs as well as schizophrenia (Ching et al., 2010; Kim et al., 2008; Pinto et al., 2010); the 44 kb deletion in family 12119 extends the number of known ASD-causing variants in the 2p16.3 region. Similarly, homozygous mutations in *ADSL* lead to adenylosuccinate lyase deficiency (OMIM #103050) and autistic features (Marie et al., 1999; Stone et al., 1992); *ADSL* haploinsufficency (family 12224) may also lead to an ASD phenotype. More recently, maternally inherited deletions at the X-linked *DDX53* locus (encoding a DEAD-box RNA helicase of unknown function) have been linked to ASDs in males (Pinto et al., 2010). The deletion of *DDX53* in a male proband from family 12561 is the first known ASD-associated de novo mutation at this locus. The linkage of the X chromosomal *NLGN3* locus (encoding neuroligin 3) to ASDs has been somewhat unclear, as this conclusion was based on a single maternally inherited missense mutation that cosegregated with autistic diagnoses in two brothers from one family (Jamain et al., 2003). The 33 kb deletion in *NLGN3* (family 11689) discovered in this study provides the first independent confirmation for a role of *NLGN3* mutations in the pathogenesis of ASDs.

At the present time, target genes in most de novo events cannot be known with certainty. First, mutations in any given candidate loci, even the recurrent ones, might be coincidental and unrelated to ASDs. Second, most events are large, disrupting more than one gene (and often dozens). Third, multiple genes within an event might act in concert. Fourth, attempts by biologists to discern the true functional subsets of genes in candidate loci cannot easily be subjected to rigorous statistical evaluation. For this reason, we have attempted to perform automated functional network analysis in a companion paper (Gilman et al., 2011). That study concludes that among the diversity there is also evidence of functional convergence upon synaptogenesis, axon guidance, and neuron motility.

Although the studies of Gilman et al. and others (Bill and Geschwind, 2009; Pinto et al., 2010) argue for functional convergence, there is "evidence" to support almost any mechanism. Some potential targets encode proteins involved in neurotransmitter metabolism (*ABAT* in family 11551), synaptic proteins (*NRXN1* and *NLGN3*, as mentioned above), and growth cones (*BAIAP2* in family 11186). But other candidate genes encode proteins active in protein degradation pathways (*USP7* in families 11551 and 11168), in intracellular signaling such as catenins (*CTNNA3* in family 11705 and *CTNND2* in family 12289), and in the *WNT* pathways (*WNT3* and *WNT9B* in family 11982). Others encode proteins that participate in metabolism (*ADSL* in family 12224, as previously mentioned), inflammation (*CSMD1* in family 11225), and possibly environmental detoxification (*COMMD1* in family 11482). Although a significant fraction of perturbed genes converge on several well-defined processes, the causes of autism are likely to be very diverse, and some causes may be treatable. However, the diversity implies that a treatment for one form of autism may be specific for only a narrow subset of genotypes and have no value for the majority. Once the specific genes mutated in ASDs are known with confidence, we can begin to think with clarity about the problems specific to individuals within categories of causation rather than attempting to manage a conglomerate disorder.

To achieve this clarity, copy-number studies may not suffice. Even with 3000 families, searching for large-scale deletions and amplifications will be inadequate to define the majority of mutational targets with the certainty that is required to further deepen understanding of the disorder at the mechanistic level. We expect that single genes will be frequent targets. If so, then we calculate that identifying the recurrent targets of de novo mutation by sequencing the exome from 3000 families will provide the yield and certainty that is needed to identify conclusively the genetic causes of ASDs.

EXPERIMENTAL PROCEDURES

An outline of the overall study design is shown in Figure 1. The institutional review board of Cold Spring Harbor Laboratory approved this study, and written informed consent from all subjects was obtained by SFARI.

System Noise Correction and Segmentation

Complete details for system noise correction follow those in Lee et al. (2011). In brief, we used standard schema (local and Lowess normalization), and we also performed self-self hybridizations (using multiple reference genomes) throughout the course of the SSC analysis. Based on singular value decomposition of the self-self data, we were able to determine the principal components of system noise and to minimize the distortion of genetic signal. We then used KS segmentation (Grubor et al., 2009), which utilizes minimization of variance to segment the data and Kolmogorov-Smirnov statistics to judge the significance of the segments. The generation of noise parameters is detailed in the Supplemental Experimental Procedures.

Pedigree and Gender Analysis

We identified a set of 974 copy-number variant regions (CNVRs) in which cluster analysis allowed us to make genotyping calls for integer copy-number

states. We selected the 837 CNVRs for which less than 5% of trios (child, father, and mother) appear to have an inconsistency in inheritance. For each pair of hybridizations, we then calculated two relatedness measures: the Mean Square Sum of the difference in copy-number state between two individuals, and a measure of the degree of similarity between polymorphisms present in the DNA profiles of two individuals (Glaubitz et al., 2003). For these measures, the thresholds for relatedness were 0.425 and 0.35, respectively. A family was flagged when more than 10% of CNVRs showed inconsistency between at least one parent and a child. For reasons of pedigree, we excluded 24 families from further analysis.

We compared the gender of a person as determined by probes on the X and Y with the information supplied in the SSC databases. If any member was discordant, the entire family was excluded, for a total of 26 families. There were two cases of Kleinfelter syndrome in unaffected siblings; these families were considered valid.

Five-State Model

We used signal/noise parameters to determine probabilities of copy-number states for segments from normalized ratio data (Supplemental Experimental Procedures). In our analysis, we restricted the state space in two ways. First, we assumed that the reference is in copy-number state 2. For uniquely mapping autosomal probes, this was almost always the correct state. The handful of regions where our reference genome was not in copy-number state 2 was filtered later for polymorphism frequency.

Our second assumption limited the test genome to five integer copy-number states, 0 to 4. Assuming a reference copy state of 2, this provided a reasonable range of variability in the test genome, more than sufficient for handling all but a few highly polymorphic regions. With the signal/noise parameters and the state model, we determined a distribution for the normalized ratio values at each of the five states within each hyb. We refer to this as the five-state model.

CNV Database

For each hybridization, we applied the five-state model to determine the most likely copy-number state for each interval in the KS segmentation. For each segment, we determine the most likely copy state for each probe. If the majority of the probes are in the 0 or 1 state, the segment is a potential deletion; if the majority of the probes are in the 3 or 4 state, the segment is a potential duplication. For a potential N-probe deletion, we apply a binomial distribution to determine the likelihood of observing M or more probes in the 0 or 1 state if the segment is really in copy state 2. An analogous procedure was used for determining a p value for potential duplications. By applying a reasonable threshold for the p value (less than 10^{-7}), we established a database of CNVs. This database served two main purposes: (1) identifying failed hybs with too many segments; and (2) generating a probe-wise map of copy-number polymorphisms over a set of 1500 high-quality parental hybridizations.

Quality Control

We used three parameters to determine the quality of a hyb: the number of autosomal segments in the CNV database, the signal parameter ξ_h, and the noise parameter σ_h (Supplemental Experimental Procedures). The quality measures for a trio are the maximum of the measures taken over the hybs of the three members in the trio. We ranked the valid trios for each quality parameter and designated as HQ only those trios in the lower 95th percentile for all three parameters. The effects of poor noise and signal parameters on the ability to distinguish copy-number states are demonstrated in Figure 2.

Probe Mapping and Filtering

For a given KS segment, states were only computed for probes that passed our filters. The first filter was the number of mappings of the probe sequence in the genome (hg18 build). We excluded probes with more than two mappings, resulting in the exclusion of ~3% of the probes. Further, we only considered probes with two mappings if the second mapping was to a site within the segment. This ensured that most probes behaved according to the five-state model. Our second filter was based on the frequency of polymorphism, or "population threshold." If a probe was in a segment deemed amplified or deleted in five or more parents, we excluded that probe from our analysis of the segment. This eliminated most regions where our reference genome was not in the standard copy-number state and guarded against cryptic de novo events, for which parents carry both a duplication and a deletion of the same locus.

Mendel and De Novo Violation

To analyze trios for de novo mutations, we used KS segmentation of the child and generated three five-state models, one for each member of the trio. For each interval in the child's segmentation, we determined the most likely copy-number state for each probe. If the majority of the child's probes were most likely in the 0 or 1 state, the segment was flagged as a potential deletion event. If they were most likely in the 3 or 4 state, the segment was flagged as a potential duplication.

If the segment was flagged, we decided whether each probe was a "Mendel violator." A probe is a deletion Mendel violator if the child probe is most likely in the 0 or 1 state and if both parents are most likely in the 2, 3, or 4 state. A probe is a duplication Mendel violator if the child probe is most likely in the 3 or 4 state and if both parents are most likely in the 0, 1, or 2 state. For each potential deletion (duplication) segment, we recorded the total number of probes and the number of deletion (duplication) Mendel-violating probes.

For each trio, we used the five-state model to simulate ratio data for all 125 trio states (0 to 4 for child, father, and mother.) Of the 125 states, 36 are "Mendel violator" states (child = 1, father = 2, mother = 2; child = 1, father = 2, mother = 3, etc.) and the remaining 89 trio states are "Mendel obedient" (child = 2, father = 2, mother = 2; child = 1, father = 1, mother = 2, etc.). For each trio state, we compute the probability that a probe drawn from that distribution is classified as a deletion (or duplication) Mendel violator. We apply that probability to parameterize a binomial distribution. This allows us to determine the likelihood that an N-probe segment in that trio state would generate M or more probes classified as Mendel violators. The p value for an N-probe segment with M Mendel violators is the maximum likelihood computed over all Mendel-obedient states.

Stringent De Novo Discovery

We set a strict threshold for the Mendel violation p value of 10^{-9} such that in 500 trios, we expected less than one false positive. As previously indicated, we also set a strict threshold for the population filter of no more than five parents showing a lesion involving a given probe. This method identified 70 de novo copy-number events in 67 trios.

Relaxed De Novo Discovery

We performed manual curation, in which we relaxed the p value threshold to 10^{-7} and the population threshold to 20. This yielded 241 de novo candidate (DNC) events in 216 children. For each DNC, we assessed a variety of information such as family ratio data, modeled state means, population polymorphism, quantile quality scores, and systematic noise. A total of 91 events passed curation, including all 70 stringent events. A full list of de novo events and their method of discovery can be found in Table S1.

Given the limited size of the X and Y chromosomes, we chose not to automate de novo discovery over these chromosomes. We altered the five-state model to use a reference copy-number state of 1 and modified the Mendel violation rules for a probe to reflect the gender of the child and the parents. We then manually inspected all segments with greater than 70% of the probes reporting as Mendel violators. Using this method, we identified three X chromosome de novo events (Table S1).

Transmitted CNVs

To identify transmitted copy-number events, we developed a 125-state HMM that operates simultaneously on the normalized ratio data of the child, father, and mother. To determine emission probabilities, we used the product of the five-state model for each member of the trio. We limited the effect of isolated failed probes by setting a minimum emission probability calibrated to the rate of single probe outliers. Transition probabilities were computed from the average CNV frequency based on KS segmentation. An additional penalty was applied for entering a "Mendel-violating" state. We then employed the Viterbi algorithm to find the most likely path through the state space. Restricting to events in which the child showed deletions or duplications, we then determined whether any parent shared the event. For each of the eight

possibilities (del/dup; from mother/father/both/neither), we constructed a measure of support similar to that of Mendel violators. Worst-case false-positive rates were determined and p values assigned to each transmitted (and de novo) event using a binomial distribution.

Permutation Tests

To determine the statistical significance of asymmetries, we performed random permutations of the data. Typically, we used 10,000 permutations for each test. See Supplemental Experimental Procedures for more details.

ACCESSION NUMBERS

The data used in this work have been deposited in NCBI's Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession number GSE23682.

SUPPLEMENTAL INFORMATION

Supplemental Information includes two figures, ten tables, and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.neuron.2011.05.015.

ACKNOWLEDGMENTS

This work was supported by a grant from the Simons Foundation (SFARI award number SF51 to MW).

We are grateful to all of the families at the participating SFARI Simplex Collection (SSC) sites, as well as the principal investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, D. Grice, A. Klin, R. Kochel, D. Ledbetter, C. Lord, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, B. Pelphrey, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, E. Wijsman).

The DNA samples used in this work include families from SSC versions 1 through 5. Approved researchers can obtain the SSC population dataset described in this study by applying at https://base.sfari.org

We thank Roche NimbleGen and Oxford Gene Technology for extensive technical assistance. We also thank Gerald Fischbach, Marian Carlson, Marilyn Simons, Catherine Lord, Matthew State, David Donoho and James Simons for helpful discussions. MW is an American Cancer Society Research Professor.

Accepted: May 10, 2011
Published: June 8, 2011

REFERENCES

Bijlsma, E.K., Gijsbers, A.C., Schuurs-Hoeijmakers, J.H., van Haeringen, A., Fransen van de Putte, D.E., Anderlid, B.M., Lundin, J., Lapunzina, P., Pérez Jurado, L.A., Delle Chiaie, B., et al. (2009). Extending the phenotype of recurrent rearrangements of 16p11.2: deletions in mentally retarded patients without autism and in normal individuals. Eur. J. Med. Genet. *52*, 77–87.

Bill, B.R., and Geschwind, D.H. (2009). Genetic advances in autism: heterogeneity and convergence on shared pathways. Curr. Opin. Genet. Dev. *19*, 271–278.

Ching, M.S., Shen, Y., Tan, W.H., Jeste, S.S., Morrow, E.M., Chen, X., Mukaddes, N.M., Yoo, S.Y., Hanson, E., Hundley, R., et al; Children's Hospital Boston Genotype Phenotype Study Group. (2010). Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders. Am. J. Med. Genet. B. Neuropsychiatr. Genet. *153B*, 937–947.

de Kovel, C.G., Trucks, H., Helbig, I., Mefford, H.C., Baker, C., Leu, C., Kluck, C., Muhle, H., von Spiczak, S., Ostertag, P., et al. (2010). Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies. Brain *133*, 23–32.

Fairburn, C.G., and Harrison, P.J. (2003). Eating disorders. Lancet *361*, 407–416.

Fernandez, B.A., Roberts, W., Chung, B., Weksberg, R., Meyn, S., Szatmari, P., Joseph-George, A.M., Mackay, S., Whitten, K., Noble, B., et al. (2010). Phenotypic spectrum associated with de novo and inherited deletions and duplications at 16p11.2 in individuals ascertained for diagnosis of autism spectrum disorder. J. Med. Genet. *47*, 195–203.

Fischbach, G.D., and Lord, C. (2010). The Simons Simplex Collection: a resource for identification of autism genetic risk factors. Neuron *68*, 192–195.

Gilman, S.R., Iossifov, I., Levy, D., Ronemus, M., Wigler, M., and Vitkup, D. (2011). Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. Neuron *70*, this issue, 898–907.

Gimelbrant, A., Hutchinson, J.N., Thompson, B.R., and Chess, A. (2007). Widespread monoallelic expression on human autosomes. Science *318*, 1136–1140.

Glaubitz, J.C., Rhodes, O.E., and Dewoody, J.A. (2003). Prospects for inferring pairwise relationships with single nucleotide polymorphisms. Mol. Ecol. *12*, 1039–1047.

Gregg, C., Zhang, J., Butler, J.E., Haig, D., and Dulac, C. (2010). Sex-specific parent-of-origin allelic expression in the mouse brain. Science *329*, 682–685.

Grubor, V., Krasnitz, A., Troge, J.E., Meth, J.L., Lakshmi, B., Kendall, J.T., Yamrom, B., Alex, G., Pai, D., Navin, N., et al. (2009). Novel genomic alterations and clonal evolution in chronic lymphocytic leukemia revealed by representational oligonucleotide microarray analysis (ROMA). Blood *113*, 1294–1303.

Hannes, F.D., Sharp, A.J., Mefford, H.C., de Ravel, T., Ruivenkamp, C.A., Breuning, M.H., Fryns, J.P., Devriendt, K., Van Buggenhout, G., Vogels, A., et al. (2009). Recurrent reciprocal deletions and duplications of 16p13.11: the deletion is a risk factor for MR/MCA while the duplication may be a rare benign variant. J. Med. Genet. *46*, 223–232.

Hong, S., Kim, S.J., Ka, S., Choi, I., and Kang, S. (2002). USP7, a ubiquitin-specific protease, interacts with ataxin-1, the SCA1 gene product. Mol. Cell. Neurosci. *20*, 298–306.

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.

Itsara, A., Wu, H., Smith, J.D., Nickerson, D.A., Romieu, I., London, S.J., and Eichler, E.E. (2010). De novo rates and selection of large copy number variation. Genome Res. *20*, 1469–1481.

Jamain, S., Quach, H., Betancur, C., Råstam, M., Colineaux, C., Gillberg, I.C., Soderstrom, H., Giros, B., Leboyer, M., Gillberg, C., and Bourgeron, T.; Paris Autism Research International Sibpair Study. (2003). Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. Nat. Genet. *34*, 27–29.

Kim, H.G., Kishikawa, S., Higgins, A.W., Seong, I.S., Donovan, D.J., Shen, Y., Lally, E., Weiss, L.A., Najm, J., Kutsche, K., et al. (2008). Disruption of neurexin 1 associated with autism spectrum disorder. Am. J. Hum. Genet. *82*, 199–207.

Kotzot, D., and Utermann, G. (2005). Uniparental disomy (UPD) other than 15: phenotypes and bibliography updated. Am. J. Med. Genet. A. *136*, 287–305.

Lee, Y.-H., Ronemus, M., Kendall, J., Lakshmi, B., Leotta, A., Levy, D., Esposito, D., Grubor, V., Ye, K., Wigler, M., and Yamrom, B. (2011) *Removing system noise from comparative genomic hybridization data by self-self analysis. arXiv:1105.0900v1*.

Marie, S., Cuppens, H., Heuterspreute, M., Jaspers, M., Tola, E.Z., Gu, X.X., Legius, E., Vincent, M.F., Jaeken, J., Cassiman, J.J., and Van den Berghe, G. (1999). Mutation analysis in adenylosuccinate lyase deficiency: eight novel mutations in the re-evaluated full ADSL coding sequence. Hum. Mutat. *13*, 197–202.

Marshall, C.R., Noor, A., Vincent, J.B., Lionel, A.C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y., et al. (2008). Structural variation of chromosomes in autism spectrum disorder. Am. J. Hum. Genet. *82*, 477–488.

McClellan, J., and King, M.C. (2010). Genetic heterogeneity in human disease. Cell *141*, 210–217.

Mefford, H.C., Cooper, G.M., Zerr, T., Smith, J.D., Baker, C., Shafer, N., Thorland, E.C., Skinner, C., Schwartz, C.E., Nickerson, D.A., and Eichler, E.E.

(2009). A method for rapid, targeted CNV genotyping identifies rare variants associated with neurocognitive disease. Genome Res. *19*, 1579–1585.

Mefford, H.C., Muhle, H., Ostertag, P., von Spiczak, S., Buysse, K., Baker, C., Franke, A., Malafosse, A., Genton, P., Thomas, P., et al. (2010). Genome-wide copy number variation in epilepsy: novel susceptibility loci in idiopathic generalized and focal epilepsies. PLoS Genet. *6*, e1000962.

Merla, G., Brunetti-Pierri, N., Micale, L., and Fusco, C. (2010). Copy number variants at Williams-Beuren syndrome 7q11.23 region. Hum. Genet. *128*, 3–26.

Morrow, E.M., Yoo, S.Y., Flavell, S.W., Kim, T.K., Lin, Y., Hill, R.S., Mukaddes, N.M., Balkhy, S., Gascon, G., Hashmi, A., et al. (2008). Identifying autism loci and genes by tracing recent shared ancestry. Science *321*, 218–223.

Muhle, R., Trentacoste, S.V., and Rapin, I. (2004). The genetics of autism. Pediatrics *113*, e472–e486.

Newschaffer, C.J., Croen, L.A., Daniels, J., Giarelli, E., Grether, J.K., Levy, S.E., Mandell, D.S., Miller, L.A., Pinto-Martin, J., Reaven, J., et al. (2007). The epidemiology of autism spectrum disorders. Annu. Rev. Public Health *28*, 235–258.

Pinto, D., Pagnamenta, A.T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S., et al. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. Nature *466*, 368–372.

Richler, J., Huerta, M., Bishop, S.L., and Lord, C. (2010). Developmental trajectories of restricted and repetitive behaviors and interests in children with autism spectrum disorders. Dev. Psychopathol. *22*, 55–69.

Rosenberg, R.E., Law, J.K., Yenokyan, G., McGready, J., Kaufmann, W.E., and Law, P.A. (2009). Characteristics and concordance of autism spectrum disorders among 277 twin pairs. Arch. Pediatr. Adolesc. Med. *163*, 907–914.

Roze, E., Meijer, L., Van Braeckel, K.N., Ruiter, S.A., Bruggink, J.L., and Bos, A.F. (2010). Developmental trajectories from birth to school age in healthy term-born children. Pediatrics *126*, e1134–e1142.

Sanders, S.J., Ercan-Sencicek, A.G., Hus, V., Luo, R., Murtha, M.T., Moreno-De-Luca, D., Chu, S.H., Moreau, M.P., Gupta, A.R., Thomson, S.A., et al. (2011). Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron *70*, this issue, 863–885.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Månér, S., Massa, H., Walker, M., Chi, M., et al. (2004). Large-scale copy number polymorphism in the human genome. Science *305*, 525–528.

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.

Shinawi, M., Liu, P., Kang, S.H., Shen, J., Belmont, J.W., Scott, D.A., Probst, F.J., Craigen, W.J., Graham, B.H., Pursley, A., et al. (2010). Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. J. Med. Genet. *47*, 332–341.

Stefansson, H., Rujescu, D., Cichon, S., Pietiläinen, O.P., Ingason, A., Steinberg, S., Fossdal, R., Sigurdsson, E., Sigmundsson, T., Buizer-Voskamp, J.E., et al; GROUP. (2008). Large recurrent microdeletions associated with schizophrenia. Nature *455*, 232–236.

Stone, R.L., Aimi, J., Barshop, B.A., Jaeken, J., Van den Berghe, G., Zalkin, H., and Dixon, J.E. (1992). A mutation in adenylosuccinate lyase associated with mental retardation and autistic features. Nat. Genet. *1*, 59–63.

van De Sluis, B., Rothuizen, J., Pearson, P.L., van Oost, B.A., and Wijmenga, C. (2002). Identification of a new copper metabolism gene by positional cloning in a purebred dog population. Hum. Mol. Genet. *11*, 165–173.

van der Zwaag, B., Staal, W.G., Hochstenbach, R., Poot, M., Spierenburg, H.A., de Jonge, M.V., Verbeek, N.E., van 't Slot, R., van Es, M.A., Staal, F.J., et al. (2010). A co-segregating microduplication of chromosome 15q11.2 pinpoints two risk genes for autism spectrum disorder. Am. J. Med. Genet. B. Neuropsychiatr. Genet. *153B*, 960–966.

Wang, K., Zhang, H., Ma, D., Bucan, M., Glessner, J.T., Abrahams, B.S., Salyakina, D., Imielinski, M., Bradfield, J.P., Sleiman, P.M., et al. (2009). Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature *459*, 528–533.

Weiss, L.A., Shen, Y., Korn, J.M., Arking, D.E., Miller, D.T., Fossdal, R., Saemundsen, E., Stefansson, H., Ferreira, M.A., Green, T., et al; Autism Consortium. (2008). Association between microdeletion and microduplication at 16p11.2 and autism. N. Engl. J. Med. *358*, 667–675.

Weiss, L.A., Arking, D.E., Daly, M.J., and Chakravarti, A.; Gene Discovery Project of Johns Hopkins & the Autism Consortium. (2009). A genome-wide linkage and association scan reveals novel loci for autism. Nature *461*, 802–808.

Wycisk, K.A., Zeitz, C., Feil, S., Wittmer, M., Forster, U., Neidhardt, J., Wissinger, B., Zrenner, E., Wilke, R., Kohl, S., and Berger, W. (2006). Mutation in the auxiliary calcium-channel subunit CACNA2D4 causes autosomal recessive cone dystrophy. Am. J. Hum. Genet. *79*, 973–977.

Xu, B., Roos, J.L., Levy, S., van Rensburg, E.J., Gogos, J.A., and Karayiorgou, M. (2008). Strong association of de novo copy number mutations with sporadic schizophrenia. Nat. Genet. *40*, 880–885.

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.