

Global increases in both common and rare copy number load associated with autism

Santhosh Girirajan^{1,2,3,*}, Rebecca L. Johnson^{4,5}, Flora Tassone^{6,8}, Jorune Balciuniene^{4,5,10}, Neerja Katiyar², Keolu Fox¹, Carl Baker¹, Abhinaya Srikanth², Kian Hui Yeoh², Su Jen Khoo², Therese B. Nauth^{4,5}, Robin Hansen^{6,7}, Marylyn Ritchie², Irva Hertz-Picciotto⁶, Evan E. Eichler¹, Isaac N. Pessah^{6,9} and Scott B. Selleck^{2,4,5,*}

¹Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195, USA,

²Department of Biochemistry and Molecular Biology and ³Department of Anthropology, The Pennsylvania State University, University Park, PA 16802, USA, ⁴The Developmental Biology Center, Department of Genetics,

Cell Biology and Development and ⁵Department of Pediatrics, University of Minnesota, Minneapolis, MN 55455, USA,

⁶The MIND (Medical Investigation of Neurodevelopmental Disorders) Institute and ⁷Department of Pediatrics, Davis School of Medicine, University of California, Sacramento, CA 95817, USA, ⁸Department of Biochemistry and

Molecular Medicine and ⁹Department of Molecular Biosciences, School of Veterinary Medicine, University of California Davis, Davis, CA 95616, USA and ¹⁰Department of Biology, Temple University, Philadelphia, PA 19122, USA

Received February 4, 2013; Revised and Accepted March 20, 2013

Children with autism have an elevated frequency of large, rare copy number variants (CNVs). However, the global load of deletions or duplications, *per se*, and their size, location and relationship to clinical manifestations of autism have not been documented. We examined CNV data from 516 individuals with autism or typical development from the population-based Childhood Autism Risks from Genetics and Environment (CHARGE) study. We interrogated 120 regions flanked by segmental duplications (genomic hotspots) for events >50 kbp and the entire genomic backbone for variants >300 kbp using a custom targeted DNA microarray. This analysis was complemented by a separate study of five highly dynamic hotspots associated with autism or developmental delay syndromes, using a finely tiled array platform (>1 kbp) in 142 children matched for gender and ethnicity. In both studies, a significant increase in the number of base pairs of duplication, but not deletion, was associated with autism. Significantly elevated levels of CNV load remained after the removal of rare and likely pathogenic events. Further, the entire CNV load detected with the finely tiled array was contributed by common variants. The impact of this variation was assessed by examining the correlation of clinical outcomes with CNV load. The level of personal and social skills, measured by Vineland Adaptive Behavior Scales, negatively correlated (Spearman's $r = -0.13$, $P = 0.034$) with the duplication CNV load for the affected children; the strongest association was found for communication ($P = 0.048$) and socialization ($P = 0.022$) scores. We propose that CNV load, predominantly increased genomic base pairs of duplication, predisposes to autism.

INTRODUCTION

With the rapidly evolving tools of modern genomics, investigators have sought to understand the genetic contributors

to neurobehavioral and psychiatric disorders. Autism has received considerable attention on account of the high levels of heritability measured from twin studies as well as the documented increasing prevalence. Recent studies on copy number

*To whom correspondence should be addressed at: 206 Life Sciences Building, Pennsylvania State University, University Park, PA 16802, USA. Email: sbs24@psu.edu (S.B.S.); 205A Life Sciences Building, Pennsylvania State University, University Park, PA 16802, USA. Email: sxg47@psu.edu (S.G.)

variants (CNVs) have suggested that rare, large events contribute significantly to autism risk and are found at very low or negligible frequencies in control groups, allowing a statistically significant association for individual variants with disease (1). This strategy has successfully identified an ever-growing list of distinct CNVs that confer a significant risk for autism and has revealed a large number of genes or genomic loci that can affect autism susceptibility. Exome sequencing studies have also identified several non-recurrent *de novo* single-nucleotide variants in simplex autism families (2–4). Several important themes have emerged from these recent genetic studies on autism. First, the number of genes or genomic loci contributing to autism susceptibility is large, now estimated in the hundreds. A recent genome-wide copy number variation analysis projected between 156 and 280 genomic intervals contributing to autism and exome sequencing of over 900 individuals provided an estimate of nearly 1000 contributing genes (2–5). It is evident that the ‘genetic target’ contributing toward autism is large and the gene discovery process is still ongoing. Second, the contribution of common variants to autism, at present measured principally as single-nucleotide polymorphisms (SNPs), is modest, making it difficult to use these genetic variants for predicting clinical outcomes (6,7).

While the degree of heritability for autism is still debated, the estimates from the largest twin studies place it at 38–60% of the variance (8,9). The entire catalog of rare deletions and duplications to date accounts for perhaps 5–10% of clinical cases (10–12), a significant level, but still far short of the level of heritability estimated from twin studies (13). Although some of the missing heritability may reside in gene-by-environment and gene-by-gene interactions (14), some may still be explained by undiscovered genetic variation, in particular variants of intermediate frequency and penetrance. A recent review of the genetic architecture of psychiatric disorders pointed out a gap between rare, highly penetrant variants represented typically by large (>1 Mbp) CNVs and common but weakly penetrant common variants represented as SNPs in our current data sets (15). The existing genetic paradigms, which will allow the discovery of a single region or locus at a time, have not sufficiently explained the missing heritability and increases in the prevalence of autism (16). Although recent studies have shown an elevated frequency of large, rare CNVs in autism, the global load of deletions or duplications in base pairs and its correlation with the clinical manifestations have not been documented. One approach to this problem is a comprehensive CNV analysis that measures all events, without removing common events, and assessing the genetic contribution of these events collectively, as copy number load. In this study, we examined global CNV load and assessed for correlation with clinical outcomes for two independent groups of children with autism, using two separate, but complementary, approaches.

RESULTS

We analyzed 274 children with autism and 242 controls ascertained from the CHARGE study (Supplementary Material, Table S1) (17). We restricted our cases to include only those

children with an Autism Diagnostic Observation Schedule (ADOS) and Autism Diagnostic Interview, Revised (ADI-R) confirmed diagnosis of autism. We excluded children with developmental delay but lacking symptoms of autism as well as those with autism spectrum disorder as defined by the CHARGE protocol (17). Controls were identified by sampling the general population, using birth records, matching on age, sex and broad geographic region, and then restricting to those who were clinically confirmed to be without autism symptoms and in the normal range of development for age confirmed by clinical assessment.

For CNV analysis, we utilized a custom microarray targeted to 120 genomic regions flanked by segmental duplications at a high probe density (~2.6 kbp) and with a median probe density of ~36 kbp across the genomic backbone for comparative genomic hybridization experiments (10). Owing to the high sequence identity of segmental duplications, the unique sequence regions between segmental duplications, termed ‘genomic hotspots’, have a high propensity to undergo rearrangements by unequal crossover leading to deletions and duplications (18,19). In fact, these regions have a 25-fold increase in frequency for undergoing rearrangements compared with the other unique, non-hotspot regions in the genome (20) and therefore are useful sites for assessing the rate of genomic instability. Previous empirical estimates indicated a high accuracy (>99%) and sensitivity for events >50 kbp within the hotspots and >300 kbp in the genomic backbone using this microarray platform (10). Quality control filtering and validation of the discovered rare pathogenic events were performed as described previously (10).

Increased duplication CNV load in children with autism compared with controls

Previous studies have suggested an overrepresentation of large, rare (>1 Mbp) CNVs in individuals with autism compared with controls or individuals with subtle neurological disorders such as dyslexia (10). We hypothesized that although some genetic variants individually may not have a sufficiently large effect to show a statistically significant association with autism, in modest size samplings, these genetic changes might have an impact on disease susceptibility globally. Such global changes in copy number can then be expected to occur in regions prone for recurrent rearrangements, i.e. within genomic hotspots. Moreover, if these events affect autism susceptibility, collectively, there should be a difference in total copy number load between children with autism and unaffected, control individuals.

We measured global changes in copy number as total length, i.e. base pairs of deletion or duplication in each individual, or collectively, as total base pairs of altered copy number (i.e. CNV load). We found that children with autism showed a significant increase in total base pairs of copy number change (deletions and duplications) (unpaired *t*-test, $P = 0.0003$) (Fig. 1A). A complete list of CNVs detected in autism and controls is shown in Supplementary Material, Tables S2 and S3. When variants were analyzed separately as deletions and duplications, the level of duplication load was significantly (unpaired *t*-test, $P = 0.0052$) elevated in individuals with autism but the deletion load was not (unpaired

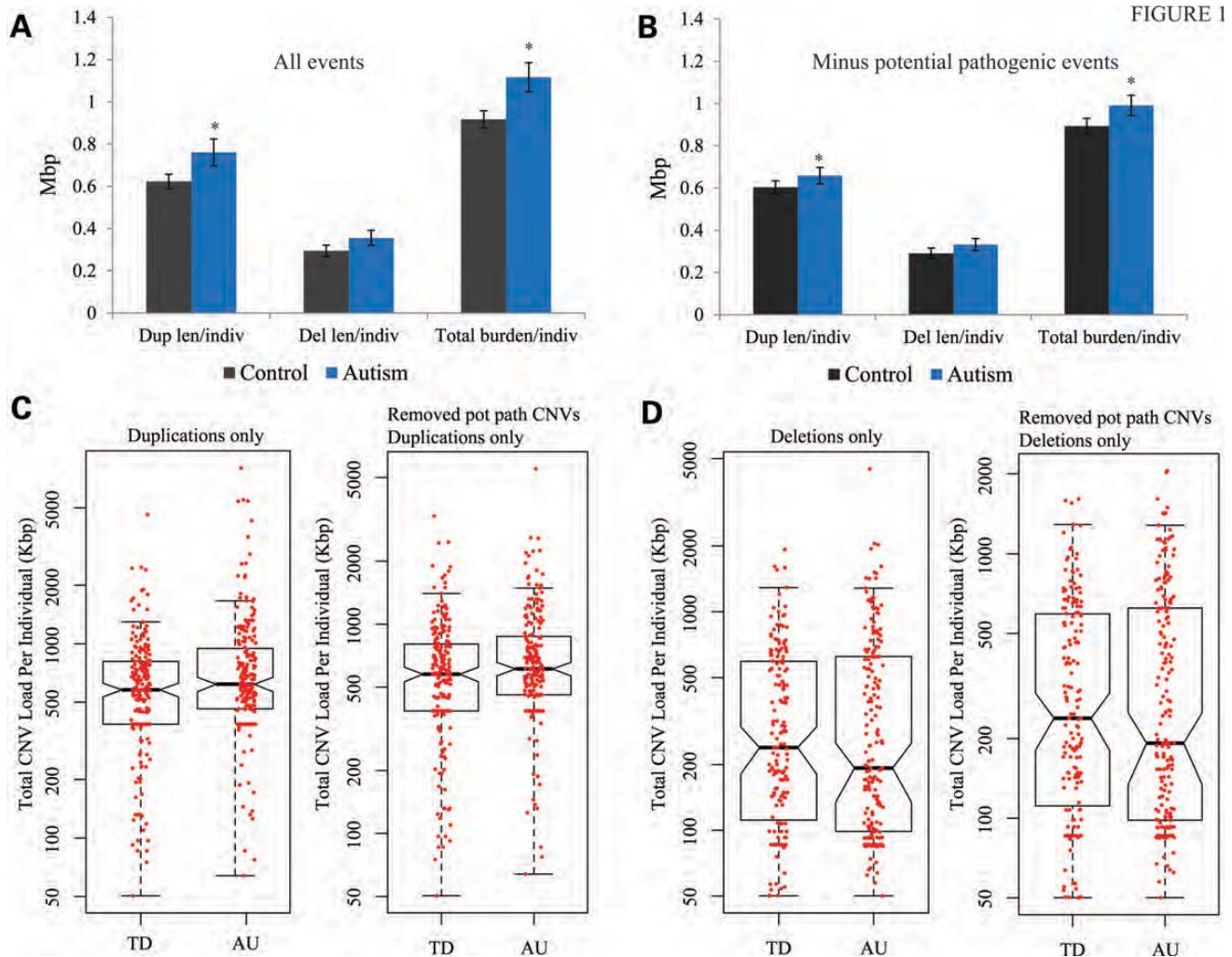


Figure 1. Copy number load and CNV size distributions for children in the CHARGE cohort. (A) Copy number load (Mbp) including all CNV events detected per individual is shown. Plots show duplication, deletion and total load length separately for controls and children with autism. (B) Copy number load per individual (deletion, duplication and total load) is shown after removing known and likely pathogenic events. Frequency distribution of total duplications (C) and deletions (D) for each individual for all events and after removing pathogenic events is also shown as boxplots. Horizontal line shows the median, and box limits represent 25th and 75th percentiles; 'whiskers' represent boundaries of all values within 1.5 SD of the mean. *P*-values were calculated using a *t*-test comparing control individuals and children with autism. Asterisks represent $P < 0.05$.

t-test, $P = 0.3$) (Fig. 1A). These findings show that autism is characterized by an increase in duplicated base pairs. We next sought to determine whether this elevated CNV load is independent of potentially pathogenic large CNVs and re-analyzed the data after removing these variants. We defined potential pathogenic CNVs as those that are rare [$<0.1\%$ ($<8/8629$) in the controls population] and/or those already known to be associated with neurodevelopmental phenotypes (21,22) (see Supplementary Material, Table S4 for potentially pathogenic CNVs). An increase in CNV load in autism cases compared with controls remained (total load, unpaired *t*-test, $P = 0.0054$), represented principally as duplications (unpaired *t*-test, $P = 0.032$) (Fig. 1B). These findings suggest that autism is associated with an increase in

duplications that are common in frequency in addition to an elevated level of rare, large genetic variants (22).

We also assessed the distribution of copy number load in the individuals to determine whether the increased duplication load is a consequence of a few outliers with a high degree of copy number changes or is an overall characteristic of the group. Boxplots showing all individuals with the median (central line), 25th and 75th percentile (box) distributions clearly show that the overall distribution of base pairs of duplication is shifted upward for the individuals with autism compared with the general population controls. This shift was not observed for deletions among these samples. The upward shift is also evident after the removal of the potential pathogenic events, showing that the statistically significant level of

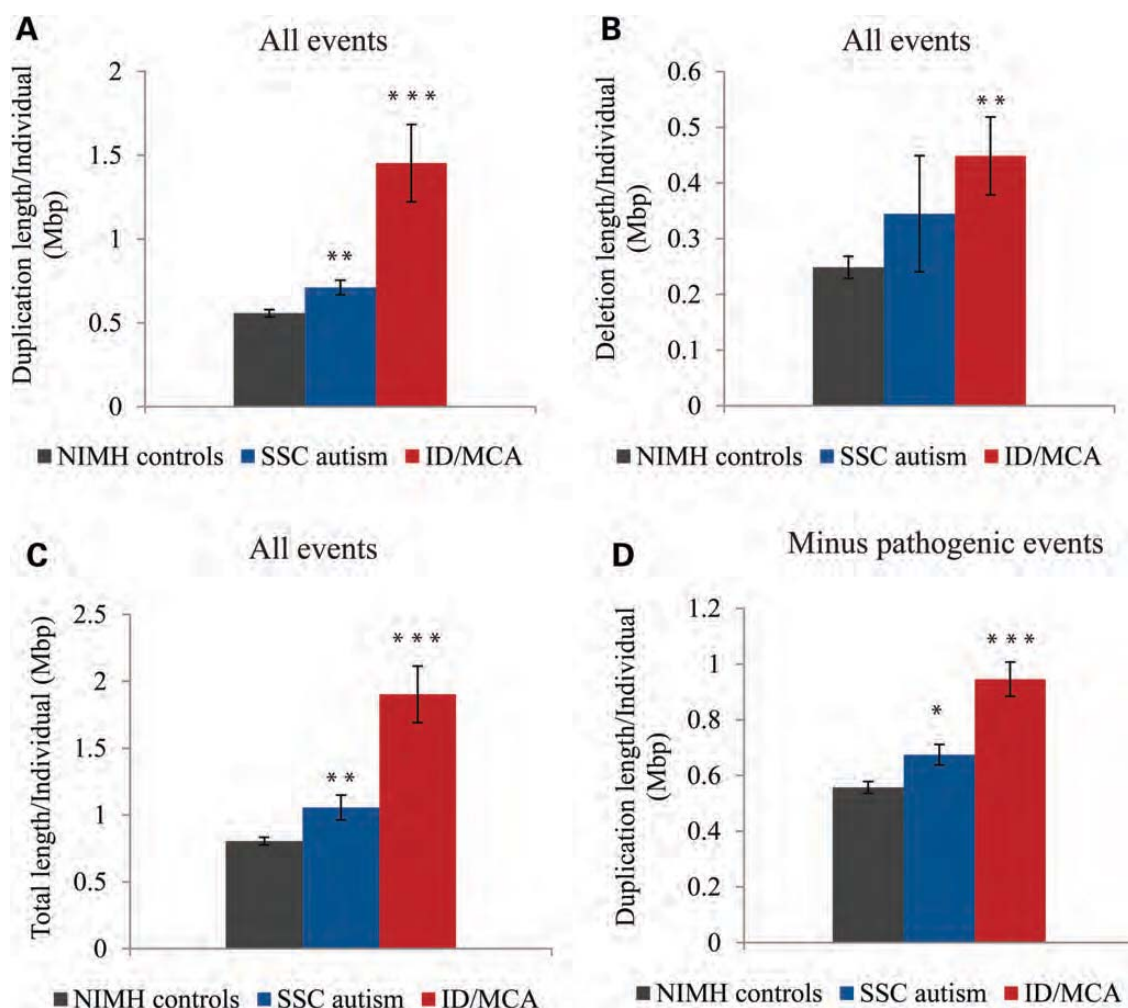


Figure 2. Copy number load comparison for neurodevelopmental disorders, including the Simons Simplex Cohort of children with autism and individuals with intellectual disability. Copy number load per individual (in Mbp) for controls, children with autism (Simons Simplex Cohort; AU SSC) and children with intellectual disability (ID) or multiple congenital anomalies (MCA) are shown as base pairs of duplications (A), deletions (B) and total load (C). Note that all events, rare pathogenic and common CNVs, were included in this analysis. (D) Duplication load for the three groups after the removal of identified pathogenic events is shown. *P*-values were determined using *t*-test comparisons with control individuals. **P* < 0.05, ***P* < 0.01 and ****P* < 0.001.

increased duplication load is not due to a few outlier samples with large events (Fig. 1C and D).

Replication of increased duplication load in an independent sample cohort

To replicate these findings in an independent cohort, we re-examined CNV data from 350 individuals with sporadic autism ascertained from the Simons Simplex Collection compared with a set of 337 control subjects ascertained by the NIMH Genetics Initiative (23,24). These samples were analyzed using the same hotspot array platform as the CHARGE samples (10). Since there was no bias in the array platforms used, this cohort served as an appropriate replication set. We measured copy number load as base pairs of deletion or duplication per individual both before and after the removal of rare, large events of potential pathogenic significance. Individuals with autism have significantly elevated total copy number

load compared with controls (unpaired *t*-test, *P* = 0.012), and duplications (unpaired *t*-test, *P* = 0.013), not deletions (unpaired *t*-test, *P* = 0.28), show statistically increased levels (Fig. 2A–C). Similar to the analysis on the CHARGE cohort, after the removal of known or likely pathogenic events in the Simons Simplex Collection, only the duplication load showed statistical significance (unpaired *t*-test, *P* = 0.045) when the autism group was compared with controls (Fig. 2D, Supplementary Material, Fig. S1). In addition, these data show that children with intellectual disability and multiple congenital anomalies, unlike those with autism, have both increased duplication and deletion load compared with controls after the removal of known pathogenic variants (Fig. 2A–D).

In order to replicate this finding in an orthogonal CNV platform, we re-analyzed CNV data from 1124 autism families reported by Sanders *et al.* (12) on Simons Simplex collection. These samples were processed using a high-density Illumina

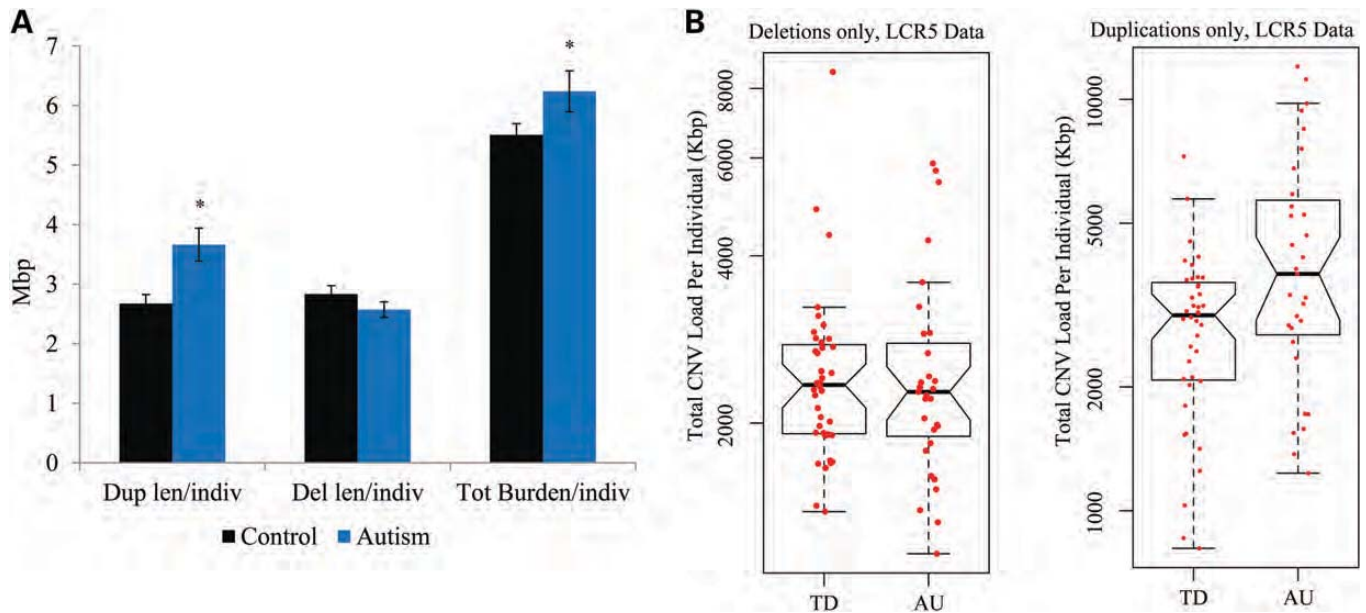


Figure 3. Copy number load measures using a high-resolution, finely tiled LCR5 array to examine five SD-rich intervals. (A) Copy number load for all events, separated into duplication and deletion load per individual (in Mbp), is shown. (B) Total copy number load per individual for all events (in Mbp). * $P < 0.05$. (C and D) Frequency distribution for duplications (C) and deletions (D) for each individual for all events and after removing pathogenic variants is shown as boxplots. The graphs show a comparison between controls and individuals with autism. P -values were calculated after a t -test comparison of cases and controls.

1Mv1 or 1Mv3 Duo Bead arrays, and CNV predictions were made using PennCNV as described previously (12). We calculated the ratio of total base pairs of duplications and deletions among the rare, high-confidence data set of CNVs for each proband compared with their corresponding parents (Supplementary Material, Fig. S2). We found a >7-fold increase in duplicated base pairs in the probands, in aggregate, compared with father (mean proband/father ratio = 7.5) or mother (mean proband/father ratio = 7.7) and a more modest increase in deleted base pairs (mean proband/father ratio = 2.3; mean proband/mother ratio = 1.7). These results show a preferential accumulation of duplication load over deletions among children with autism and replicate our findings for the rare, high-confidence set of CNVs detected with a different CNV detection platform.

High-resolution microarray detects increased duplication load contributed by common variants

To determine whether a more sensitive assay would uncover significantly greater levels of copy number change, we analyzed 71 autism cases and 71 controls matched for ethnicity and sex using a finely tiled array. Of note, 34 of the autism individuals included in this analysis were from the CHARGE study sample, also evaluated using the genomic hotspot array, whereas the remainder was obtained from the Autism Genetic Resource Exchange (AGRE) study. We selected five highly dynamic, segmental duplication-rich genomic hotspots including 7q11.2 associated with Williams syndrome (25), 10q22q23 associated with autism and other developmental deficits (26), 15q11.2q13 associated with Prader-Willi/Angelman syndrome (27) and autism (28), 17q12 associated with renal cysts and diabetes (29) and autism (30) and 22q11.2 region associated with DiGeorge/velocardiofacial

syndrome (31). These regions were selected based on the length and size distribution of flanking segmental duplications and severity of the disorders associated with the CNV mapping within these regions. We designed a custom microarray with average probe spacing of one probe every 180 bp providing a CNV size resolution as high as 1–2 kbp (designated as LCR5 array). The sensitivity and specificity of the LCR5 arrays were evaluated by comparing CNV data from two HapMap control samples with CNV data from orthogonal whole-genome platforms reported previously (32) (see Supplementary Material, Fig. S3). Utilizing two complementary CNV detection algorithms and quantitative polymerase chain reaction validations, we estimate a true positive rate between 71.25 and 80% and false positive rate between 5.1 and 6.6% (see Materials and Methods). We performed LCR5 array hybridizations on 71 autism and 71 control subjects matched for age, sex and ethnicity. For the entire 75 Mbp of DNA interrogated by the LCR5 array, the autism samples showed a significantly higher total copy number load (unpaired t -test, $P = 0.032$) represented mainly as an increase in total duplication length in base pairs (unpaired t -test, $P = 0.011$, autism/control ratio = 1.37) (Fig. 3A and B). The load of deletions was not significantly different between autism cases and controls (unpaired t -test, $P = 0.089$). Boxplot representation of the copy number load (Fig. 3B), shown separately as duplications or deletions, reveals that the statistically significant increase in duplication load is not achieved through a few outliers, but rather is reflected in an elevated median and 25th and 75th percentile measures for the group. Twenty of the 71 autism individuals show a greater total length of duplication in the size range of ≥ 4 Mbp per person, emphasizing that the duplication load detected was not a consequence of a few outliers (Supplementary Material, Fig. S4). These differences in distributions were not detected for the measures of

Table 1. Copy number load and VABS correlations

Score	Load	Spearman's correlation (<i>r</i>)	<i>P</i> -value (uncorrected)	Bonferroni-corrected <i>P</i> -value* (five tests)
Vineland total	Total load	-0.1689	0.0058	0.029
	Duplications	-0.126	0.0338	0.169
	Deletions	-0.032	0.325	1
Socialization	Total load	-0.1575	0.0093	0.0465
	Duplications	-0.1376	0.0232	0.116
	Deletions	-0.0074	0.4582	1
Communication	Total load	-0.1975	0.0016	0.008
	Duplications	-0.12	0.0476	0.238
	Deletions	-0.065	0.1765	0.8825
Daily living skills	Total load	-0.1265	0.03	0.15
	Duplications	-0.0612	0.188	0.94
	Deletions	-0.0556	0.216	1
Motor skills	Total load	-0.099	0.0692	0.346
	Duplications	-0.092	0.09	0.45
	Deletions	0.0017	0.49	1

*Cutoff for corrected (for five tests) *P*-value = 0.01.

base pairs of deletion. We note that the amount of copy number variation detected using LCR5 array was greater than the whole-genome hotspot array reflecting the increased sensitivity gained with a finely tiled design. None of the variants detected with this platform included rare or previously identified pathogenic variants, suggesting that the increase in copy number load is contributed by common variants, at least for these five highly variant regions of the genome.

Measures of clinical severity show correlation with levels of copy number load

Our data indicate that autism and other neurodevelopmental disorders show an elevated level of global copy number load beyond the increased frequency of rare and large CNVs that have been characterized to date. If this copy number change has an effect on the susceptibility or level of deficit, then we would expect a relationship between the clinical outcome and the level of copy number change. Several measures of neurodevelopmental outcomes were obtained for CHARGE participants, including Mullen Scales of Early Learning (MSEL) (33) for cognitive function and the Vineland Adaptive Behavior Scales (VABS) (34) for assessments of social and adaptive behaviors on all study children, as well as the direct assessment of autism using the ADOS, given to those recruited with a diagnosis of autism (35). We examined the correlation between these clinical assessments and the levels of copy number load, measured as base pairs of deletion, duplication or total load. Negative correlations at significant levels, using the Spearman rank test, were observed for total CNV load as well as duplication load when assessed with VABS scores including communication, socialization and total VABS scores. Table 1 shows both corrected and uncorrected *P*-values for the five Spearman correlation tests performed using the VABS scores. We did not observe any significant level of correlation, with and without multiple testing corrections, between copy number load and either MSEL or ADOS scores (36) (data not shown). It is of interest that VABS are based on interviews of caregivers and seek to provide a detailed context for the child's current functional

daily living skills and activities, whereas MSEL and ADOS measures are taken in a 30–40 min assessment that can be affected by the age and attendance-to-task of the subject. Recently, similar observations were also made on autism samples from the Simons Simplex Collection, with the size of the duplication directly correlating with autism severity and with no impact on verbal IQ whereas the size of the deletion inversely correlating with IQ (37). Further replication of these results is warranted to fully understand the role of duplications toward autism-specific features.

Novel potentially pathogenic CNVs found in the CHARGE cohort

A number of CHARGE study children with a diagnosis of autism were found to have CNVs not previously reported and not found in an additional set of 8329 control individuals (22). We report these rare variants (Table 2) in order that other studies can attempt to replicate these events to determine their contribution to autism susceptibility.

4q32.3 deletion

A 3.6 Mbp, paternally inherited deletion was detected in one autism proband. No clinical phenotypes were reported for the father. This is a novel CNV not observed in 15 767 developmental delay (DD) children or 8329 controls (22). A large, 19 Mbp *de novo* deletion associated with autism has been reported for this region, indicating that this genomic interval may contribute to autism and other behavioral disorders (38).

4q13.1 duplication

We detected a *de novo* 1.5 Mbp duplication in a child with autism near the *EPHA5* gene. Previous genome-wide association study analysis identified a marker near *EPHA5* that had a significant association with autism (39).

10q11.23 duplication

A novel 290 kbp maternally inherited duplication spanning the *DRG11* gene was detected in an individual with autism. This CNV was not detected in any controls or DD individuals (22).

Table 2. Rare CNVs detected in children with autism from CHARGE cohort

Chr. band	Start	End	Size	CNV	Sample ID	Ethnicity	Mullen	Vineland	Inheritance	Comments	Reference
4q32.3	166 240 300	169 876 274	3 635 974	Del	45804101669	Hispanic	49	55	Paternal	Novel event not found in 8329 controls or 15 767 DD. Includes <i>SPOCK3</i> and <i>KLHL2</i>	(22)
15q11.2q12	18 259 360	26 226 639	7 967 279	Dup	19305103869	Mixed	49	50	-	Represents classic duplication found in 1–3% of all autism cases; 0 out of 8329 controls and 27 out of 15 767 DD cases	(28,42)
15q11.2q12	20 812 153	26 225 840	5 413 687	Dup	17209111199	Caucasian	49	57	<i>De novo</i>		
15q11.2q12	21 226 843	26 242 596	5 015 753	Dup	60905104617	Mixed	49	50	-		
15q13.3	28 913 092	30 246 254	1 333 162	Dup	9106105625	Caucasian	49	55	-	Duplication involving <i>CHRNA7</i>	(43,60)
17q12	32 539 340	32 627 777	88 437	Dup	8203100466	Mixed	49	51	-	Observed 0 out of 8329 controls, 1 out of 15 767 DD	(22)
16p11.2	28 624 768	29 080 269	455 501	Dup	8203100466	Mixed	49	51	Paternal	Implicated in AU, schizophrenia, ID. Observed: 2 out of 8329 controls, 28 out of 15 767 DD	(22,44,45,61)
1q21.1	143 665 017	144 529 218	864 201	Del	607107598	Caucasian	111	77	<i>De novo</i>	Observed: 2 out of 8329 controls, 19 out of 15 767 DD	(22,62)
4q13.1	62 808 305	64 394 393	1 586 088	Dup	38108110540	Asian	49	48	<i>De novo</i>	Observed: 0 out of 8329 controls, 1 out of 15 767 DD. Duplication 3' to <i>EPHA5</i> associated with autism	(22,45)
10q11.23	50 094 947	50 388 501	293 554	Dup	8705102962	Caucasian	61	68	Maternal	Novel duplication encompassing <i>DRG11</i> . Observed: 0 out of 8329 controls, 0 out of 15 767 DD	(22)
7q11.22	68 528 129	69 368 926	840 797	Del	52104101908	Caucasian	53	71	<i>De novo</i>	Includes <i>AUTS2</i> . Observed: 0 out of 8329 controls	(22,63)
6q22.31	120 625 209	122 799 034	2 173 825	Dup	29806106231	Caucasian	114	87	Maternal	Duplication spanning <i>GJAI</i> . Observed: 0 out of 8329 controls	(22)

DD, developmental delay.

6q22.31 duplication

We detected a 2.1 Mbp duplication spanning the *GJAI* gene. Loss of *GJAI* function is responsible for oculodentodigital dysplasia, and mutations at this locus have been associated with Sudden Infant Death Syndrome (40,41). The detection of a duplication affecting this channel gene, connexin43, suggests that gain-of-function mutations could have different neurodevelopmental and behavioral consequences.

Known potentially pathogenic CNVs found in the CHARGE cohort

The detection and mapping of CNVs in 516 autism and typically developing individuals provided the opportunity to evaluate the representation of CNVs previously reported with autism in this ethnically diverse cohort from California.

15q11–13 duplications

We identified three individuals with autism carrying 15q11.2q13 duplications. These duplications are perhaps one of the common genomic causes for autism occurring in 1–3% of cases (28,42); the frequency (3/256, 1.2%) of this CNV in our study is therefore consistent with previous reports. One of these events was *de novo*, and for the other two children, parental DNA was not available. We also detected 15q13.3 duplications in two autism probands encompassing *CHRNA7*. A recent study of a large set of individuals with duplications in this interval makes the direct association of these CNVs with autism uncertain but suggests the possibility that it is a common disease-contributing variation (43).

16p11.2 duplication

We detected a paternally inherited 450 kbp 16p11.2 duplication in an autism proband. This variant has been associated with autism, schizophrenia and intellectual disability (44,45). Earlier work has found this variant in 28 out of 15 767 cases of developmental delay cases and 2 out of 8329 controls ($P = 0.0004$; $OR = 7.41$).

7q11.22 deletion

We also detected in one individual an ~840 kbp deletion that includes the *AUTS2* gene. Translocations that disrupt *AUTS2* have been reported in autistic twins (46) and children with cognitive deficits (47). Our finding provides additional data supporting a role for functional changes in this gene contributing to autism susceptibility.

DISCUSSION

Recent progress in understanding the genetic underpinnings of autism has provided some unexpected results and outlined the challenges that remain toward understanding autism. First, it is clear from both copy number and single-nucleotide variant analysis that, at the molecular level, autism is a very heterogeneous disorder. The genetic target for autism is now estimated in the hundreds to thousands of genes or genomic loci. Second, it would appear that genetic variants associated with autism are rarely unique to the disorder. A number of genomic regions that have a significant role in autism are associated

with substantial clinical heterogeneity (48), including intellectual disability, seizure disorder, schizophrenia and developmental delay. These findings indicate that there will likely be few autism-specific genes, but rather a collection of genes affecting the phenotypes associated with autism, principally language and socialization deficits. Third, the autism-associated genetic variants discovered thus far only begin to account for the estimated heritability, emphasizing there is a great deal of genetic heterogeneity yet to be uncovered. Finally, the incidence of autism reported is increasing (49) and there is no biological model to account for these epidemiological findings. As an alternative to the traditional approach of searching for the enrichment of individual genetic variants associated with autism, we performed a genome-wide CNV study of two independent cohorts based on a hypothesis that children with autism will exhibit global increases in copy number variation, particularly in those regions of the genome prone to rearrangement. This search for relatively large variants (>50 kbp) both across the genome and within unstable segment of the genome was bolstered by a high-resolution analysis (CNVs >2 kbp) using a finely tiled array of five segmental duplication-rich intervals where rare variants are known to contribute to behavioral disorders.

We analyzed total base pairs of copy number change in a total of 624 children with autism compared with 579 controls. Our analysis draws from two independent groups of children with autism: the population-based case-control CHARGE study, and the Simons Simplex Cohort. These two groups of children were assessed with a microarray that provides coverage of the entire genome at 300 kbp resolution, and for 120 SD-rich intervals prone to rearrangement, a resolution of 50 kbp. In addition to the whole-genome assessment, we examined five SD-rich intervals at high resolution (>1 kbp) in 142 autism cases and controls matched for ethnicity and sex. In all three of these studies, children with autism exhibited a significantly elevated copy number load, represented principally as an increase in duplicated base pairs found in large CNVs (>200 kbp) (Supplementary Material, Fig. S5). The level of deletion load was not different from controls for these cohorts. These findings were further replicated by analysis of published data that examined rare, high-confidence CNVs in 1024 Simons Simplex trios using a 1 million feature Illumina array (12). This analysis, comparing copy number load as the ratio of base pairs of deletion or duplication in proband to parent, showed an ~7-fold increase in duplication base pairs and roughly a 2-fold increase in deletion base pairs. Consistent with previous reports on an increase in *de novo* CNV load (50), we find that the increase in global CNV load from data from Sanders *et al.* is mainly due to *de novo* events in the affected child. This preponderance of duplication load further emphasizes the large contribution of duplication events to autism.

In the autism cases, the majority of copy number load is represented as large duplications, in contrast with ID/MCA, where both large deletions and duplications are increased compared with control groups. These findings suggest that deletions are typically associated with more severe phenotypes (ID/MCA), and that autism, on average, is associated with genomic variants with more modest functional impact, namely duplications. This conclusion is supported by the

observation that there is no elevated load of large, rare events in children with dyslexia, a neurological disorder with a relatively narrow set of deficits (10). Our results, therefore, validate the previously observed graded distribution of phenotypic severity correlating with levels and type of copy number variation.

Significant levels of duplication load remained even after the removal of rare, potentially pathogenic variants for autism and ID/MCA, suggesting some contribution from variants occurring at higher frequencies than the large, highly penetrant CNVs classified as pathogenic. Current studies have overlooked variants of intermediate penetrance and frequency because of the practice of removing common variants from CNV studies and focusing only on those individually rare variants with high penetrance. In a recent review of the genetic architecture of psychiatric disorders, it was pointed out that our current understanding includes rare variants (CNVs) of high penetrance and common variants of low penetrance (SNPs) but little about the variants of intermediate penetrance (15). The variants we have detected, and assessed collectively as copy number load, are likely to encompass all types, including those of intermediate penetrance and frequency (Supplementary Material, Fig. S5).

Among the CHARGE study children with autism, we found significant negative correlations between the level of copy number load and socialization and communication modules of the VABS. In contrast, no such significant correlations were found for ADOS or Mullen scores. These findings suggest that measures of adaptive skills provide a very sensitive assessment of functional change. A comparison of VABS and ADOS scores for children assessed in two different clinical centers showed VABS to be significantly reduced in high functioning autistic children with normal IQ scores. In addition, there was only a weak correlation between VABS and ADOS scores (51). Indeed, Klin *et al.* (51) suggested the VABS may be particularly useful in genetic studies of autism since it was designed to assess graded changes in adaptive ability, including the normal range of function, whereas ADOS measures were designed to identify disability, and may therefore not be as sensitive to some functional deficits.

Increased duplication load associated with autism can be explained by more than one mechanism. It is known that deletions are generally deleterious and duplications are relatively well tolerated in the genome; our observations suggest a selection bias toward duplication in autism. In fact, SD-rich regions where deletions have been associated with developmental delay often have reciprocal duplications that produce autistic features. Examples include reciprocal duplication of the Williams syndrome region and 17p11.2 duplication or Potocki-Lupski syndrome, which is a reciprocal duplication of the Smith-Magenis syndrome interval. Notably, one of the first CNVs identified as predisposing for autism is the reciprocal duplication of the Prader-Willi/Angelman syndrome region on chromosome 15q11.2q13.1. Alternatively, the elevated level of duplication could be not only due to selection, but ascertainment; namely, the generation of these variants could be increased in the autism population. Segmental duplication-rich intervals possess an inherent instability and it is possible that these segments of the genome show an additional propensity to change in the autism population either from genetic or

environmental causes. There are several factors that could contribute to genomic instability, including imbalances in DNA replication or recombination mechanisms (52), maternal and paternal age (53,54) and DNA methylation (55). It is notable that perinatal vitamin/folate supplementation, which would potentially increase DNA methylation levels, is associated with a lower autism incidence (56). Given the large genetic target of neurodevelopmental disorders, estimated in the hundreds or even thousands of genomic loci, it stands to reason that anything that increases genomic instability could contribute to the genesis of these disorders.

MATERIALS AND METHODS

Ethics statement

Patients from each of study cohort were recruited after appropriate human subjects' approval and informed consent.

DNA samples

All samples were obtained after informed written consent and in accordance with IRB protocols at all facilities. Autism patients' DNA samples were acquired from the CHARGE study (17) conducted through the Medical Investigation of Neurodevelopmental Disorders (MIND) Institute at UC-Davis, and from the AGRE Repository (57). All patients for CHARGE and AGRE cohorts were selected based on meeting full criteria for Autistic Disorder (AU) (OMIM 209850) using ADOS, ADI-R and/or DSM-IV-TR criteria. The ADOS instrument consists of a series of structured and semistructured presses for interaction, accompanied by coding of specific target behaviors associated with particular tasks and by general ratings of the quality of behaviors (58). Based on documentation in the AGRE database, efforts were made to exclude patients with syndromic autism and known genetic causes of autism, including fragile X syndrome (OMIM 300624). Samples from the CHARGE study were screened for fragile X syndrome by PCR (59). For the custom targeted hotspot array study, samples only from CHARGE study participants were used. In summary, 274 autism and 242 control samples were hybridized, of which 243 and 223 passed QC criteria (10) and were included in the final analysis. Based on self-reported ethnicity, this cohort consists of 52% of individuals of European descent, 28% Hispanic, 2% African or African-American ancestry, 5% Asian ancestry and the remaining 12% of mixed ancestry (Supplementary Material, Table S1). The finely tiled analysis using the LCR5 array included a total of 71 autism patients, 37 samples from the MIND Institute and 34 samples from the AGRE repository, with a total of 34 Caucasian males, 21 Hispanic males, 8 Caucasian females and 8 Hispanic females. Typically developing control DNA samples were obtained from the MIND Institute, the University of Minnesota Psychology Department and the CEPH Utah Pedigree (Centre de'Etude du Polymorphisme Human, Coriell Institute for Medical Research, Camden, NJ, USA). DNA samples from the CHARGE study participants and the University of Minnesota were from whole blood, whereas the CEPH and AGRE samples were from transformed lymphoblastoid cells. A total

of 71 controls, matched to the autism patient sample set based on gender and ethnicity, were included in this study, with a total of 40 samples from the MIND Institute, 28 samples from the University of Minnesota collection and 3 samples were obtained from the CEPH collection.

Custom targeted hotspot analysis

Custom targeted hotspot arrays comprised 135 000 probes (Roche NimbleGen), with higher density probe coverage (median probe spacing 2.6 kbp) in the genomic hotspots (regions flanked by segmental duplications) and a lower probe density in the genomic backbone (median probe spacing 36 kbp). Hybridization, quality control and segmentation analysis were all conducted as previously described (10). All potentially pathogenic events detected with the hotspot array were confirmed with the higher resolution Agilent hotspot 2 × 400K array according to the manufacturer's instructions.

LCR5 finely tiled array design and data analysis

We designed a custom 385K oligonucleotide array from Roche NimbleGen Systems, Inc. (Madison, WI, USA) targeting five genomic regions with an average probe density of one probe every 120 bp in segmental duplication-containing regions and one probe every 200 bp in unique sequence regions. Probe sequences were based on Build 36 (Hg18) of the human reference genome. In SD-containing regions, probes were required to have a minimum of one unique nucleotide per probe; probes in unique sequences were required to have at least five unique nucleotides per probe. The five genomic regions were chr7: 61 058 424–82 000 033 (20.9 Mb), chr10: 77 000 071–91 999 959 (15.0 Mb), chr15: 18 260 026–34 999 973 (16.7 Mb), chr17: 12 000 112–22 187 066 (10.2 Mb) and chr22: 14 430 001–26 000 041 (11.6 Mb). SD-containing regions accounted for 24.5% of the sequence on the array (18.2 out of 74.4 Mb). Details for LCR5 experimental methods, including platform comparisons and CNV calling criteria, are provided in the Supplementary Material, Methods.

Enrichment analysis for unique and sequences embedded within segmental duplications

To identify the location of deletion, duplication or total copy number load, sequences within a CNV were parsed in to those that are entirely unique, those containing segmental duplications and unique sequences and segmental duplication-only sequences. Enrichment statistics were calculated for total base pairs of deletion or duplication for each of these categories.

Statistical analysis of CNV load

Statistical comparison of copy number load data described in this study was performed using the unpaired *t*-test reporting two-sided *P*-values.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We gratefully acknowledge the technical contributions and assistance of the following colleagues: Sarah Pendergrass, Emre Karakoc, Bradley Coe, Gregory Cooper, Caitlin M. Conboy, Angela N. Klossner, Ryan Davis, Jeff Gregg and Maria Krasilnikova. We also thank Majid Alsagabi, Abdullah Alqallaf and Ahmed Tewfik for the development of algorithms for segmentation assessment. We are also indebted to Tonya White and Monica Luciano (University of Minnesota) for providing control DNA samples from their University of Minnesota-based study.

Conflict of Interest statement. E.E.E. is on the scientific advisory boards for Pacific Biosciences, Inc., SynapDx Corp and DNAnexus, Inc.

FUNDING

This work received support from the Kempf Fund as well as from the University of Minnesota General Clinical Research Center Award. This work was supported by Autism Speaks/Cure Autism Now (to S.B.S.), the University of Minnesota Harrison Autism Initiative Fund (to S.B.S.), Pennsylvania State University (to S.B.S.), the Minnesota Medical Foundation (to R.L.J.) and Autism Speaks/Cure Autism Now Environmental Innovator Award (to I.N.P.) (matching funds for the CHARGE study), the MIND Institute [matching funds for the CHARGE Study (I.H.-P.)] and by R01-ES015359 and P01ES011269 from the NIEHS and Award Numbers R833292 and R829388 from the Environmental Protection Agency. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. Funding to pay the Open Access publication charges for this article was provided by The Pennsylvania State University to Scott B. Selleck.

REFERENCES

- Malhotra, D. and Sebat, J. (2012) CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell*, **148**, 1223–1241.
- Neale, B.M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K.E., Sabo, A., Lin, C.F., Stevens, C., Wang, L.S., Makarov, V. *et al.* (2012) Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*, **485**, 242–245.
- O'Roak, B.J., Deriziotis, P., Lee, C., Vives, L., Schwartz, J.J., Girirajan, S., Karakoc, E., Mackenzie, A.P., Ng, S.B., Baker, C. *et al.* (2011) Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat. Genet.*, **43**, 585–589.
- Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L. *et al.* (2012) De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*, **485**, 237–241.
- Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y.H., Narzisi, G., Leotta, A. *et al.* (2012) De novo gene disruptions in children on the autistic spectrum. *Neuron*, **74**, 285–299.
- Anney, R., Klei, L., Pinto, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S., Sykes, N., Pagnamenta, A.T. *et al.* (2010) A genome-wide scan for common alleles affecting risk for autism. *Hum. Mol. Genet.*, **19**, 4072–4082.
- Wang, K., Zhang, H., Ma, D., Bucan, M., Glessner, J.T., Abrahams, B.S., Salyakina, D., Imitowski, 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.
- Hallmayer, J., Cleveland, S., Torres, A., Phillips, J., Cohen, B., Torigoe, T., Miller, J., Fedele, A., Collins, J., Smith, K. *et al.* (2011) Genetic heritability and shared environmental factors among twin pairs with autism. *Arch. Gen. Psychiatry*, **68**, 1095–1102.
- Pennington, B.F., McGrath, L.M., Rosenberg, J., Barnard, H., Smith, S.D., Willcutt, E.G., Friend, A., Defries, J.C. and Olson, R.K. (2009) Gene X environment interactions in reading disability and attention-deficit/hyperactivity disorder. *Dev. Psychol.*, **45**, 77–89.
- Girirajan, S., Brkanac, Z., Coe, B.P., Baker, C., Vives, L., Vu, T.H., Shafer, N., Bernier, R., Ferrero, G.B., Silengo, M. *et al.* (2011) Relative burden of large CNVs on a range of neurodevelopmental phenotypes. *PLoS Genet.*, **7**, e1002334.
- Girirajan, S., Campbell, C.D. and Eichler, E.E. (2011) Human copy number variation and complex genetic disease. *Annu. Rev. Genet.*, **45**, 203–226.
- 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**, 863–885.
- 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.
- Zuk, O., Hechter, E., Sunyaev, S.R. and Lander, E.S. (2012) The mystery of missing heritability: genetic interactions create phantom heritability. *Proc. Natl Acad. Sci. USA*, **109**, 1193–1198.
- Sullivan, P.F., Daly, M.J. and O'Donovan, M. (2012) Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat. Rev. Genet.*, **13**, 537–551.
- Eichler, E.E., Flint, J., Gibson, G., Kong, A., Leal, S.M., Moore, J.H. and Nadeau, J.H. (2010) Missing heritability and strategies for finding the underlying causes of complex disease. *Nat. Rev. Genet.*, **11**, 446–450.
- Hertz-Picciotto, I., Croen, L.A., Hansen, R., Jones, C.R., van de Water, J. and Pessah, I.N. (2006) The CHARGE study: an epidemiologic investigation of genetic and environmental factors contributing to autism. *Environ. Health Perspect.*, **114**, 1119–1125.
- Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W. and Eichler, E.E. (2002) Recent segmental duplications in the human genome. *Science*, **297**, 1003–1007.
- Sharp, A.J., Hansen, S., Selzer, R.R., Cheng, Z., Regan, R., Hurst, J.A., Stewart, H., Price, S.M., Blair, E., Hennekam, R.C. *et al.* (2006) Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat. Genet.*, **38**, 1038–1042.
- Itsara, A., Cooper, G.M., Baker, C., Girirajan, S., Li, J., Absher, D., Krauss, R.M., Myers, R.M., Ridker, P.M., Chasman, D.I. *et al.* (2009) Population analysis of large copy number variants and hotspots of human genetic disease. *Am. J. Hum. Genet.*, **84**, 148–161.
- Girirajan, S. and Eichler, E.E. (2010) Phenotypic variability and genetic susceptibility to genomic disorders. *Hum. Mol. Genet.*, **19**, R176–R187.
- Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V. *et al.* (2011) A copy number variation morbidity map of developmental delay. *Nat. Genet.*, **43**, 838–846.
- Moldin, S.O. (2003) NIMH Human Genetics Initiative: 2003 update. *Am. J. Psychiatry*, **160**, 621–622.
- Kessler, R.C. and Ustun, T.B. (2004) The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *Int. J. Methods Psychiatr. Res.*, **13**, 93–121.
- Pober, B.R. (2010) Williams-Beuren syndrome. *N. Engl. J. Med.*, **362**, 239–252.
- Balciuniene, J., Feng, N., Iyadurai, K., Hirsch, B., Charnas, L., Bill, B.R., Easterday, M.C., Staaf, J., Oseth, L., Czapansky-Beilman, D. *et al.* (2007) Recurrent 10q22-q23 deletions: a genomic disorder on 10q associated with cognitive and behavioral abnormalities. *Am. J. Hum. Genet.*, **80**, 938–947.

27. Cassidy, S.B., Dykens, E. and Williams, C.A. (2000) Prader-Willi and Angelman syndromes: sister imprinted disorders. *Am. J. Med. Genet.*, **97**, 136–146.
28. Cook, E.H. Jr, Lindgren, V., Leventhal, B.L., Courchesne, R., Lincoln, A., Shulman, C., Lord, C. and Courchesne, E. (1997) Autism or atypical autism in maternally but not paternally derived proximal 15q duplication. *Am. J. Hum. Genet.*, **60**, 928–934.
29. Mefford, H.C., Clauin, S., Sharp, A.J., Moller, R.S., Ullmann, R., Kapur, R., Pinkel, D., Cooper, G.M., Ventura, M., Ropers, H.H. *et al.* (2007) Recurrent reciprocal genomic rearrangements of 17q12 are associated with renal disease, diabetes, and epilepsy. *Am. J. Hum. Genet.*, **81**, 1057–1069.
30. Moreno-De-Luca, D., Mulle, J.G., Kaminsky, E.B., Sanders, S.J., Myers, S.M., Adam, M.P., Pakula, A.T., Eisenhauer, N.J., Uhas, K., Weik, L. *et al.* (2010) Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. *Am. J. Hum. Genet.*, **87**, 618–630.
31. Shprintzen, R.J. (1994) Velocardiofacial syndrome and DiGeorge sequence. *J. Med. Genet.*, **31**, 423–424.
32. Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shaper, M.H., Carson, A.R., Chen, W. *et al.* (2006) Global variation in copy number in the human genome. *Nature*, **444**, 444–454.
33. Akshoomoff, N. (2006) Use of the Mullen scales of early learning for the assessment of young children with autism spectrum disorders. *Child Neuropsychol.*, **12**, 269–277.
34. Sparrow, S.S. and Cicchetti, D.V. (1985) Diagnostic uses of the Vineland Adaptive Behavior Scales. *J. Pediatr. Psychol.*, **10**, 215–225.
35. Lord, C., Risi, S., Lambrecht, L., Cook, E.H. Jr, Leventhal, B.L., DiLavore, P.C., Pickles, A. and Rutter, M. (2000) The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disord.*, **30**, 205–223.
36. Lord, C., Leventhal, B.L. and Cook, E.H. Jr. (2001) Quantifying the phenotype in autism spectrum disorders. *Am. J. Hum. Genet.*, **105**, 36–38.
37. Girirajan, S., Dennis, M.Y., Baker, C., Malig, M., Coe, B.P., Campbell, C.D., Mark, K., Vu, T.H., Alkan, C., Cheng, Z. *et al.* (2013) Refinement and discovery of new hotspots of copy number variation associated with autism spectrum disorder. *Am. J. Hum. Genet.*, **92**, 1–17.
38. Ramanathan, S., Woodroffe, A., Flodman, P.L., Mays, L.Z., Hanouni, M., Modahl, C.B., Steinberg-Epstein, R., Bocian, M.E., Spence, M.A. and Smith, M. (2004) A case of autism with an interstitial deletion on 4q leading to hemizyosity for genes encoding for glutamine and glycine neurotransmitter receptor sub-units (AMPA 2, GLRA3, GLRB) and neuropeptide receptors NPY1R, NPY5R. *BMC Med. Genet.*, **5**, 10.
39. Weiss, L.A., Arking, D.E., Daly, M.J. and Chakravarti, A. (2009) A genome-wide linkage and association scan reveals novel loci for autism. *Nature*, **461**, 802–808.
40. Paznekas, W.A., Boyadjiev, S.A., Shapiro, R.E., Daniels, O., Wollnik, B., Keegan, C.E., Innis, J.W., Dinulos, M.B., Christian, C., Hannibal, M.C. *et al.* (2003) Connexin 43 (GJA1) mutations cause the pleiotropic phenotype of oculodentodigital dysplasia. *Am. J. Hum. Genet.*, **72**, 408–418.
41. Van Norstrand, D.W., Asimaki, A., Rubinos, C., Dolmatova, E., Srinivas, M., Tester, D.J., Saffitz, J.E., Duffy, H.S. and Ackerman, M.J. (2012) Connexin43 mutation causes heterogeneous gap junction loss and sudden infant death. *Circulation*, **125**, 474–481.
42. Thomas, J.A., Johnson, J., Peterson Kraai, T.L., Wilson, R., Tartaglia, N., LeRoux, J., Beischel, L., McGavran, L. and Hagerman, R.J. (2003) Genetic and clinical characterization of patients with an interstitial duplication 15q11-q13, emphasizing behavioral phenotype and response to treatment. *Am. J. Med. Genet. A*, **119A**, 111–120.
43. Szafranski, P., Schaaf, C.P., Person, R.E., Gibson, I.B., Xia, Z., Mahadevan, S., Wiszniewska, J., Bacino, C.A., Lalani, S., Potocki, L. *et al.* (2010) Structures and molecular mechanisms for common 15q13.3 microduplications involving CHRNA7: benign or pathological? *Hum. Mutat.*, **31**, 840–850.
44. McCarthy, S.E., Makarov, V., Kirov, G., Addington, A.M., McClellan, J., Yoon, S., Perkins, D.O., Dickel, D.E., Kusenda, M., Krastoshevsky, O. *et al.* (2009) Microduplications of 16p11.2 are associated with schizophrenia. *Nat. Genet.*, **41**, 1223–1227.
45. 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.* (2008) Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.*, **358**, 667–675.
46. Sultana, R., Yu, C.E., Yu, J., Munson, J., Chen, D., Hua, W., Estes, A., Cortes, F., de la Barra, F., Yu, D. *et al.* (2002) Identification of a novel gene on chromosome 7q11.2 interrupted by a translocation breakpoint in a pair of autistic twins. *Genomics*, **80**, 129–134.
47. Huang, X.L., Zou, Y.S., Maher, T.A., Newton, S. and Milunsky, J.M. (2010) A de novo balanced translocation breakpoint truncating the autism susceptibility candidate 2 (AUTS2) gene in a patient with autism. *Am. J. Med. Genet. A*, **152A**, 2112–2114.
48. Coe, B.P., Girirajan, S. and Eichler, E.E. (2012) The genetic variability and commonality of neurodevelopmental disease. *Am. J. Med. Genet. C Semin. Med. Genet.*, **160C**, 118–129.
49. Hertz-Picciotto, I. (2009) Commentary: diagnostic change and the increased prevalence of autism. *Int. J. Epidemiol.*, **38**, 1239–1241; author reply 1243–1234.
50. 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.
51. Klin, A., Saulnier, C.A., Sparrow, S.S., Cicchetti, D.V., Volkmar, F.R. and Lord, C. (2007) Social and communication abilities and disabilities in higher functioning individuals with autism spectrum disorders: the Vineland and the ADOS. *J. Autism Dev. Disord.*, **37**, 748–759.
52. Zhang, F., Carvalho, C.M. and Lupski, J.R. (2009) Complex human chromosomal and genomic rearrangements. *Trends Genet.*, **25**, 298–307.
53. Parner, E.T., Baron-Cohen, S., Lauritsen, M.B., Jorgensen, M., Schieve, L.A., Yeargin-Allsopp, M. and Obel, C. (2012) Parental age and autism spectrum disorders. *Ann. Epidemiol.*, **22**, 143–150.
54. Sandin, S., Hultman, C.M., Kolvezon, A., Gross, R., MacCabe, J.H. and Reichenberg, A. (2012) Advancing maternal age is associated with increasing risk for autism: a review and meta-analysis. *J. Am. Acad. Child Adolesc. Psychiatry*, **51**, 477–486.e1.
55. Li, J., Harris, R.A., Cheung, S.W., Coarfa, C., Jeong, M., Goodell, M.A., White, L.D., Patel, A., Kang, S.H., Shaw, C. *et al.* (2012) Genomic hypomethylation in the human germline associates with selective structural mutability in the human genome. *PLoS Genet.*, **8**, e1002692.
56. Schmidt, R.J., Hansen, R.L., Hartiala, J., Allayee, H., Schmidt, L.C., Tancredi, D.J., Tassone, F. and Hertz-Picciotto, I. (2011) Prenatal vitamins, one-carbon metabolism gene variants, and risk for autism. *Epidemiology*, **22**, 476–485.
57. Geschwind, D.H., Sowiński, J., Lord, C., Iversen, P., Shestack, J., Jones, P., Ducat, L. and Spence, S.J. (2001) The autism genetic resource exchange: a resource for the study of autism and related neuropsychiatric conditions. *Am. J. Hum. Genet.*, **69**, 463–466.
58. Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L. and Schopler, E. (1989) Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. *J. Autism Dev. Disord.*, **19**, 185–212.
59. Tassone, F., Pan, R., Amiri, K., Taylor, A.K. and Hagerman, P.J. (2008) A rapid polymerase chain reaction-based screening method for identification of all expanded alleles of the fragile X (FMR1) gene in newborn and high-risk populations. *J. Mol. Diagn.*, **10**, 43–49.
60. Ingason, A., Rujescu, D., Cichon, S., Sigurdsson, E., Sigmundsson, T., Pietilainen, O.P., Buizer-Voskamp, J.E., Strengman, E., Francks, C., Muglia, P. *et al.* (2011) Copy number variations of chromosome 16p13.1 region associated with schizophrenia. *Mol. Psychiatry*, **16**, 17–25.
61. Rosenfeld, J.A., Coppinger, J., Bejjani, B.A., Girirajan, S., Eichler, E.E., Shaffer, L.G. and Ballif, B.C. (2011) Speech delays and behavioral problems are the predominant features in individuals with developmental delays and 16p11.2 microdeletions and microduplications. *J. Neurodev. Disord.*, **2**, 26–38.
62. Klopocki, E., Schulze, H., Strauss, G., Ott, C.E., Hall, J., Trotier, F., Fleischhauer, S., Greenhalgh, L., Newbury-Ecob, R.A., Neumann, L.M. *et al.* (2007) Complex inheritance pattern resembling autosomal recessive inheritance involving a microdeletion in thrombocytopenia-absent radius syndrome. *Am. J. Hum. Genet.*, **80**, 232–240.
63. Elia, J., Gai, X., Xie, H.M., Perin, J.C., Geiger, E., Glessner, J.T., D'Arcy, M., deBerardinis, R., Frackelton, E., Kim, C. *et al.* (2010) Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. *Mol. Psychiatry*, **15**, 637–646.