LETTERS

Autism genome-wide copy number variation reveals ubiquitin and neuronal genes

Joseph T. Glessner[1], Kai Wang[1], Guiqing Cai[2], Olena Korvatska[3], Cecilia E. Kim[1], Shawn Wood[4], Haitao Zhang[1], Annette Estes[3], Camille W. Brune[5], Jonathan P. Bradfield[1], Marcin Imielinski[1], Edward C. Frackelton[1], Jennifer Reichert[2], Emily L. Crawford[6], Jeffrey Munson[3], Patrick M. A. Sleiman[1], Rosetta Chiavacci[1], Kiran Annaiah[1], Kelly Thomas[1], Cuiping Hou[1], Wendy Glaberson[1], James Flory[1], Frederick Otieno[1], Maria Garris[1], Latha Soorya[2], Lambertus Klei[4], Joseph Piven[7], Kacie J. Meyer[8], Evdokia Anagnostou[2], Takeshi Sakurai[2], Rachel M. Game[6], Danielle S. Rudd[8], Danielle Zurawiecki[2], Christopher J. McDougle[10], Lea K. Davis[8], Judith Miller[9], David J. Posey[10], Shana Michaels[4], Alexander Kolevzon[2], Jeremy M. Silverman[2], Raphael Bernier[3], Susan E. Levy[11], Robert T. Schultz[11], Geraldine Dawson[3], Thomas Owley[5], William M. McMahon[9], Thomas H. Wassink[8], John A. Sweeney[5], John I. Nurnberger Jr[10], Hilary Coon[9], James S. Sutcliffe[6], Nancy J. Minshew[12], Struan F. A. Grant[1,11], Maja Bucan[13], Edwin H. Cook Jr[5], Joseph D. Buxbaum[2,14], Bernie Devlin[4], Gerard D. Schellenberg[15] & Hakon Hakonarson[1,11]

Autism spectrum disorders (ASDs) are childhood neurodevelopmental disorders with complex genetic origins[1–4]. Previous studies focusing on candidate genes or genomic regions have identified several copy number variations (CNVs) that are associated with an increased risk of ASDs[5–9]. Here we present the results from a whole-genome CNV study on a cohort of 859 ASD cases and 1,409 healthy children of European ancestry who were genotyped with ~550,000 single nucleotide polymorphism markers, in an attempt to comprehensively identify CNVs conferring susceptibility to ASDs. Positive findings were evaluated in an independent cohort of 1,336 ASD cases and 1,110 controls of European ancestry. Besides previously reported ASD candidate genes, such as *NRXN1* (ref. 10) and *CNTN4* (refs 11, 12), several new susceptibility genes encoding neuronal cell-adhesion molecules, including *NLGN1* and *ASTN2*, were enriched with CNVs in ASD cases compared to controls ($P = 9.5 \times 10^{-3}$). Furthermore, CNVs within or surrounding genes involved in the ubiquitin pathways, including *UBE3A*, *PARK2*, *RFWD2* and *FBXO40*, were affected by CNVs not observed in controls ($P = 3.3 \times 10^{-3}$). We also identified duplications 55 kilobases upstream of complementary DNA *AK123120* ($P = 3.6 \times 10^{-6}$). Although these variants may be individually rare, they target genes involved in neuronal cell-adhesion or ubiquitin degradation, indicating that these two important gene networks expressed within the central nervous system may contribute to the genetic susceptibility of ASD.

ASDs, including autism, are neurodevelopmental disorders characterized by impairments in social and communication skills, as well as stereotyped and repetitive behaviours and/or a restricted range of interests. Current prevalence estimates in the United States are 0.1–0.2% for autism and 0.6% for ASDs[1,2].

Linkage and candidate gene association studies have implicated several chromosomal regions in autism[3,4]. However, positive findings in one study often fail to replicate in other studies, and a consistent picture of susceptibility loci in autism is still lacking. Some telling clues about ASD genetics arose from recent studies on CNVs[5], including the association of *de novo* CNVs with ASDs[6]. Although *de novo* CNVs that disrupt specific genes may contribute to the pathogenesis of ASDs, heritable CNVs are much more common but have been less studied as risk factors of ASDs. A family-based genome-wide linkage and CNV analysis by the Autism Genome Project Consortium using Affymetrix 10K single nucleotide polymorphism (SNP) arrays implicated chromosome 11p12-13 and neurexin 1 (*NRXN1*) as candidate loci[7]. A study using the Affymetrix 500K SNP array in a Canadian population reported 277 rare CNVs that were only observed in ASD patients but not in 1,652 healthy controls or in the Database of Genomic Variants[8]. Furthermore, 16p11.2 deletions and duplications have been reported in independent cohorts of autism patients[9]. These studies concordantly implicate a role for CNVs in the genetic susceptibility to ASDs.

To search systematically for CNVs that confer a risk to ASDs, we used a genome-wide approach with the Illumina HumanHap550 BeadChip. We assembled an Autism Case-Control (ACC) cohort by collecting 859 ASD cases (from a total of 1,246 ACC cases, parents, and siblings) of European ancestry, and 1,409 healthy controls. Among these case subjects, all met the diagnostic criteria for autism on the basis of Autism Diagnostic Interview (ADI), and 124 met the criteria for other ASDs on the basis of Autism Diagnostic Observation Schedule (ADOS)[13]. Fifty-four per cent were from simplex families, the rest were from multiplex families. In addition, we also analysed 1,336 ASD cases (from a total of 3,398 cases, parents, and siblings) in the Autism Genetic Resource Exchange (AGRE)[14] collection, as well as 1,110 control subjects as a replication cohort. Among the AGRE cases, 5% were from simplex families and 95%

[1]Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. [2]Seaver Autism Center for Research and Treatment, Department of Psychiatry, Mount Sinai School of Medicine, New York, New York 10029, USA. [3]University of Washington, Seattle, Washington 98105, USA. [4]Departments of Psychiatry and Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA. [5]Institute for Juvenile Research and Department of Psychiatry, University of Illinois at Chicago, Chicago, Illinois 60608, USA. [6]Center for Molecular Neuroscience and Vanderbilt Kennedy Center, Vanderbilt University, Nashville, Tennessee 37235, USA. [7]Neurodevelopmental Disorders Research Center and Department of Psychiatry, University of North Carolina, Chapel Hill, North Carolina 27412, USA. [8]University of Iowa, Iowa City, Iowa 52242, USA. [9]University of Utah, Salt Lake City, Utah 84112, USA. [10]Indiana University, Indianapolis, Indiana 46202, USA. [11]Department of Pediatrics, Children's Hospital of Philadelphia, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA. [12]Departments of Psychiatry and Neurology, University of Pittsburg School of Medicine, Pittsburg, Pennsylvania 15260, USA. [13]Dept of Genetics, Biology and Biostatistics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA. [14]Departments of Neuroscience, and Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, New York 10029, USA. [15]Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA.

©2009 Macmillan Publishers Limited. All rights reserved

Table 1 | CNVs in gene regions previously implicated in ASDs

Gene/region	Region of significance	Type	Validation	ACC case	ACC control	AGRE case	AGRE control	Inh	Combined *P* value	Permuted *P* value
UBE3A 15q11-13	chr15: 21200234–26208861	Dup	MLPA, qPCR	8	0	7 (5)	0	6:0:2 (75%)	1×10^{-5}	<0.001
NRXN1	chr2: 51120644–51147600	Del	qPCR	4	0	6 (4)	0	4:3:0 (100%)	4.7×10^{-4}	0.002
CNTN4	chr3: 1915190–1915922	Del	qPCR	3	0	7 (4)	0	5:3:0 (100%)	4.7×10^{-4}	0.004
22q11.21	chr22: 19351264–19358946	Dup	MLPA	5	0	4	0	2:4:0 (100%)	0.0010	0.008
CNTN4	chr3: 2548148–2548531	Dup	qPCR	1	1	8 (6)	0	2:3:1 (83%)	0.0078	0.013
16p11.2	chr16: 29554844–30085308	Dup	Affymetrix	2	3	7 (3)	1	3:4:1 (88%)	0.162	0.408
16p11.2	chr16: 29554844–30085308	Del	Affymetrix	3	2	5 (4)	2	0:0:5 (0%)	0.246	0.300
AUTS2	chr7: 68576554–68967283	Dup	Affymetrix	0	0	1	0	1:0:0 (100%)	0.466	0.425
NLGN3	chrX: 70288980–70460080	Dup	Affymetrix	0	0	1	1	0:0:0 (NA)	0.601	0.601
SHANK3	chr22: 49456858–49524956	Del	Affymetrix	2	1	0	1	4:3:0 (100%)	1	1

The number in parenthesis refers to count of unrelated siblings or distinct unrelated families. The sample included 859 ASD cases from the ACC cohort, 1,336 ASD cases from the AGRE cohort, and 2,519 unaffected controls (1,409 ACC discovery controls and 1,110 AGRE replication controls). All CNVs, except 16p11.2, *AUTS2*, *NLGN3* and *SHANK3*, were experimentally validated in the ACC cohort. Regions listed represent the optimal overlap of cases and significance with respect to controls, as described in the Methods and Supplementary Fig. 5, upper panel. 'Inh' column lists the inheritance pattern of each CNV from parents to cases in the format <inherited from mother>:<inherited from father>:<denovo>. Pedigrees provided in Supplementary Fig. 9. The percentage of inheritance is listed in parenthesis next to these three values. Note that parents were not available for all cases.

were from multiplex families: 1,202 met the criteria for autism and 134 for other ASDs[13] (Supplementary Tables 1 and 2).

We generated 78,490 CNV calls (22,581 in the ACC series and 55,909 in the AGRE series) from all the ASD subjects and their family members that met strictly established data quality thresholds (Methods). On average, 15.5 CNV calls were made for each individual using the PennCNV software[15], with similar frequency observed in cases and controls (Supplementary Fig. 2).

We first examined eight genomic regions that have been previously implicated in ASDs. Among those, CNVs involving the 15q11-13, 22q11.21, and *NRXN1* regions have well-established associations with autism[10]. CNVs affecting *CNTN4* in ASD cases have also been reported in independent studies[11,12]. We statistically adjusted for relatedness of cases with permutation and our results demonstrate that duplications of 15q11-q13 and the 22q11.21 region, deletions of *NRXN1*, as well as deletions and duplications of *CNTN4* replicate in our cohorts (Table 1). Conversely, we did not obtain statistical support for several other genomic regions previously shown to associate with ASD, including *AUTS2* (ref. 16), *NLGN3* (ref. 17), *SHANK3* (ref. 18) and 16p11.2 (ref. 9) (Table 1). We observed a similar frequency of deletions and duplications of the 16p11.2 locus in the ASD cases (~0.3%) as previously reported[9]; however, the CNV frequency in the control subjects at this locus was also comparable to that of the cases (Supplementary Fig. 3). It is noteworthy that CNVs at the 16p11.2 locus do not segregate to all cases in three of the affected families and they are also transmitted to unaffected siblings (Supplementary Fig. 4). These results indicate that CNVs at the 16p11.2 locus may not be sufficient to be causal variants in ASD.

To identify other new genomic loci contributing to ASDs, we applied a segment-based scoring approach that scans the genome for consecutive SNPs with more frequent copy number changes in cases compared to controls. This approach defines copy number variation regions, or CNVRs (Supplementary Fig. 5, upper panel). In the ACC cohort, we identified four CNVRs that were observed in cases but not in controls, as well as five CNVRs that had significantly higher frequency in cases versus controls (Table 2).

To replicate the CNVRs exclusively observed in ACC cases, we examined the AGRE case-control data set; of the four case-specific CNVRs, two were also exclusive to AGRE cases (*PARK2* and *RFWD2*), whereas the other two (*AK057321* and *FBXO40*) were not observed in either the cases or controls (combined *P* values ranging from 3.57×10^{-6} to 0.1, unadjusted for multiple testing) (Table 2). Interestingly, four genes (*UBE3A*, *PARK2*, *RFWD2* and *FBXO40*) that were significantly enriched for CNVs and observed in the ASD cases only, belong to the ubiquitin gene family (UniProt category 'Ubl conjugation pathway', $P = 3.3 \times 10^{-3}$). The other five CNVRs, as well as being enriched in the ACC cases compared with controls, were over-represented in the AGRE cases compared with the independent controls (Table 2). Figure 1 shows the most statistically significant locus, a duplication 55 kb upstream of *AK123120*, using UCSC Genome Browser[19] with Build 36 (March 2006) of the human genome. To ensure reliability of our CNV detection method, we experimentally validated all the significant CNVRs using other methods, including quantitative PCR (qPCR) and multiplex ligation-dependent probe amplification (MLPA) (Fig. 2). Affymetrix 5.0 array data were also available for a subset of the AGRE subjects for validation.

Table 2 | New common CNVRs over-represented in ASD patients

									Combined statistics		
CNVR	Gene	Description	Type	ACC case	ACC control	AGRE case	AGRE control	Inh	OR	*P* value	Permuted *P* value
chr6: 162584576–162587001	*PARK2*	E3 ubiquitin-protein ligase	Del	3	0	4 (3)	0	0:5:0 (100%)	NA	***0.0047***	0.005
chr1: 174500555–174543675	*RFWD2, PAPPA2*	E3 ubiquitin-protein ligase	Dup	3	0	3	0	1:1:1 (67%)	NA	**0.0102**	0.011
chr7: 32667087–32770713	*AK057321*	Testis expression	Dup	4	0	0	0	0:0:1 (0%)	NA	0.0469	0.034
chr3: 122826190–122870474	*FBXO40, GOLGB1*	Protein-ubiquitin ligase	Dup	3	0	0	0	0:0:0 (NA)	NA	0.1009	0.094
chr2: 13119667–13165898	*AK123120*	Cerebellum expression	Dup	9	3	24 (20)	4	16:8:1 (96%)	5.547	$\boldsymbol{3.57 \times 10^{-6}}$	<0.001
chr3: 4199731–4236304	*UNQ3037*	Sulphatase	Del	3	1	12 (8*)	2	6:7:0 (100%)	5.804	***0.0017***	0.003
chr10: 87941666–87949029	*GRID1*	Glutamate receptor	Del	4	1	10 (8)	2	4:7:0 (100%)	5.412	***0.0031***	0.011
chr3: 174754378–174771975	*NLGN1*	Neuroligin	Dup	40	52	74 (57†)	40	43:29:2 (97%)	1.471	**0.0101**	0.039
chr4: 144847402–144854579	*GYPELOC441046*	Glycophorin E precursor	Dup	3	0	7 (5)	2	2:2:2 (67%)	5.782	**0.0166**	0.037

The number in parenthesis refers to count of unrelated siblings or distinct unrelated families. The sample included 859 ASD cases from the ACC cohort, 1,336 ASD cases from the AGRE cohort and 2,519 unaffected controls (1,409 ACC discovery controls and 1,110 AGRE replication controls). All loci were validated with qPCR. 'Inh' column lists the inheritance pattern of each CNV from parents to cases in the format <inherited from mother>:<inherited from father>:<denovo>. The percentage of inheritance is listed in parenthesis. Italicized *P* values denote CNVs that survive multiple testing with Bonferroni adjustment in the discovery phase (that is, $P < 0.05$ after correction for five deletion and nine duplication CNVRs), and bold *P* values denote CNVRs that survived both the replication and experimental validation. OR, odds ratio.

* AGRE family 574 has three affected siblings.

† Three families had three affected siblings (AGRE families 656, 955 and 1559).

©2009 Macmillan Publishers Limited. All rights reserved

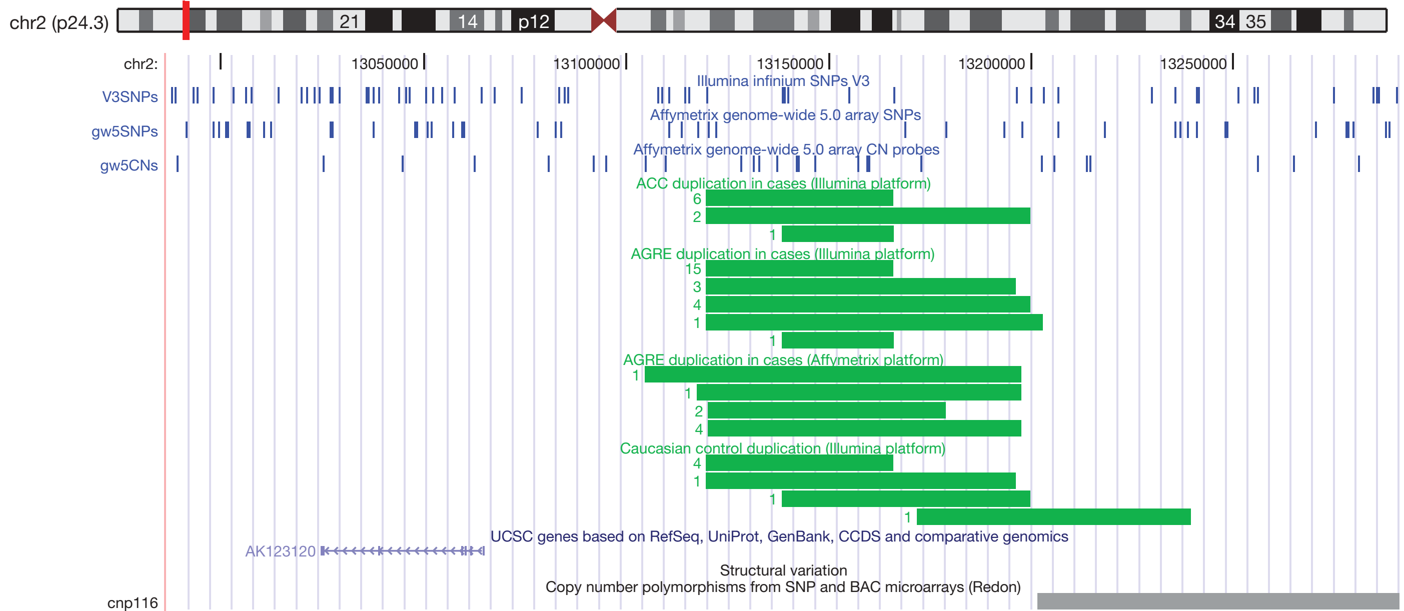

Figure 1 | *AK123120*: example of overrepresented CNVs. *AK123120* chromosome 2 (chr2): 12,986,750–13,291,000 divided into subsections with headers for ACC CNVs, AGRE CNVs, AGRE Affymetrix validation CNVs, and control CNVs. The AGRE Affymetrix Replication track is on the basis of genome-wide 5.0 SNP genotyping data from the Broad Institute (see Supplementary Methods and Acknowledgements), and were generated using the PennCNV-Affy algorithm (see Supplementary Methods), to serve as a further means to validate the Illumina-based CNV calls. SNP and copy number (CN) probe coverage are shown as blue lines across the top. Produced with custom tracks listing CNV calls uploaded to http://genome.ucsc.edu. Figures for all loci are included in Supplementary Information.

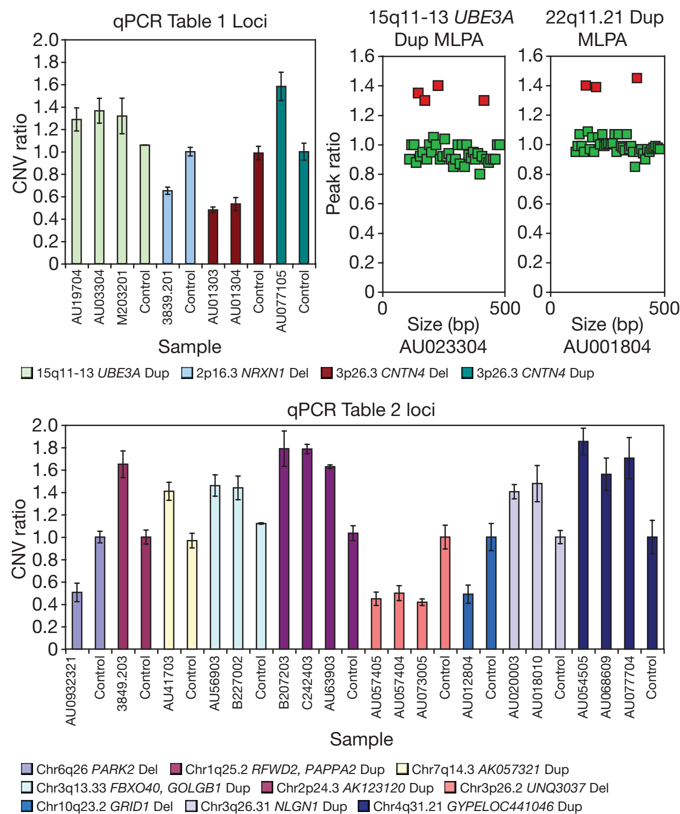

Figure 2 | Independent validation using qPCR and MLPA. Fluorescent probe-based qPCR assays using Roche Universal probe library and/or MLPA were designed to validate every candidate CNV with a completely independent test (representative series shown for each locus). Error bars denote the s.d. of quadruplicate runs. bp, base pairs; Del, deletion; Dup, duplication.

Besides a segment-based scoring approach for CNV association, an alternative method is the gene-based scoring approach that examines CNV calls affecting any region of the gene (Supplementary Fig. 5, lower panel). Using this approach, we further identified seven genes with an increased frequency of CNVs in ASD cases versus controls (Supplementary Table 3). For each gene, most CNVs target different parts of the gene and would have been missed by the segment-based approach. Of note, four of the genes identified by the segment- and gene-based approaches are involved in neuron development (*NRXN1*, *CNTN4*, *ASTN2* and *NLGN1*) (Gene Ontology term 'neuron development', $P = 9.5 \times 10^{-3}$). Therefore, by combining evidence from two complementary CNV association approaches, the large sample size has enabled us to implicate two specific gene networks or biological pathways in ASDs: the ubiquitination system and neuronal cell-adhesion molecules.

The genes from the ubiquitin pathway (*UBE3A*, *PARK2*, *RFWD2* and *FBXO40*) represent a new CNV finding in ASD susceptibility. Ubiquitination is a post-translational modification which can rapidly alter protein function and target proteins for proteasome-mediated degradation. The ubiquitin–proteasome system operates pre- and post-synaptic compartments, regulating synaptic attributes, including neurotransmitter release, synaptic vesicle recycling in pre-synaptic terminals, and dynamic changes in dendritic spines and the post-synaptic density (PSD)[20]. Of the four ubiquitin-related genes highlighted in our study, *UBE3A*, a ubiquitin protein ligase, has been the most extensively studied in the context of autism. *PARK2* is a ubiquitin-protein ligase, mutations of which cause autosomal recessive juvenile Parkinson's disease[21], and *RFWD2* and *FBXO40* are also ubiquitin-protein ligases, but neither has been previously associated with disease-causing mutations. The role of ubiquitin in the turnover of synaptic components such as the neuronal cell-adhesion molecules in a process involving regulation of activity-dependent synaptic plasticity presents a mechanism that links these two major gene networks. In addition to the genes described above, several ubiquitin-related genes are involved in human neurological diseases. These include *NHLRC1*, *UBR*, *CUL4B*, *BRWD3* and *HUWE1*, genes that encode ubiquitin protein E3 ligases. Mutations in the latter three and in

©2009 Macmillan Publishers Limited. All rights reserved

UBE2A, an E2 ubiquitin-conjugating enzyme, cause syndromes that include intellectual disability[22].

Genes from the second group of genes implicated in our study, neuronal cell-adhesion molecules, are critical in the development of the nervous system, contributing to axonal guidance, synaptic formation and plasticity, and neuronal–glial interactions. Recent genetic evidence has suggested associations between autism susceptibility and neuronal cell-adhesion molecules, including *NRXN1* (ref. 10), *CNTNAP2* (ref. 23), *NLGN3* (ref. 17), *NLGN4X* (ref. 17), and specific cadherins. Our results provide support for some previously reported genes (*NRXN1* and *CNTN4*), and also implicate additional genes with cell-adhesion functions, including *NLGN1* and *ASTN2*. Mutations in neuroligin superfamily members have previously been found in individuals with autism, and have subsequently been shown to be functionally relevant[24]. *ASTN1*, a well-studied homologue of *ASTN2*, is a neuronal protein receptor integral in the process of glial-guided granule cell migration during development[25], and *ASTN2* deletions have recently been associated with schizophrenia[26].

Using a genome-wide approach for high-resolution CNV detection, we have identified candidate genomic loci with enrichment of CNVs in ASD cases as compared to controls, and have replicated many of them using an independent set of cases. Most of these genes fall within two pathways/networks involving neuronal cell-adhesion and ubiquitin degradation. The enrichment of genes within these molecular systems suggests new susceptibility mechanisms for ASDs. Our results call for functional and expression assays to be completed to assess the biological effects of CNVs in these candidate genes.

METHODS SUMMARY

All genome-wide SNP genotyping was performed using the InfiniumII HumanHap550 BeadChip at the Center for Applied Genomics at The Children's Hospital of Philadelphia (CHOP). We called CNVs with the PennCNV algorithm[15], which combines multiple values, including Log R Ratio, B Allele Frequency, SNP spacing and population frequency of the B allele into a hidden Markov model. The term 'CNV' represents individual CNV calls, whereas 'CNVR' refers to population-level variation. Quality control thresholds included a high success rate of attempted SNPs, low standard deviation of normalized intensity, genetically inferred European ancestry, low genomic wave artefacts, count of CNV calls per subject, and genotypic duplicate removal (Supplementary Table 4). CNV frequency between cases and controls was evaluated at each SNP using Fisher's exact test. We report statistical local minimums in reference to a region of nominal significance including SNPs residing within 1 Mb of each other (Supplementary Fig. 6). Resulting significant CNVRs were excluded if they were (1) residing on telomere or centromere proximal cytobands; (2) arising from a 'peninsula' of common CNV (Supplementary Fig. 7); (3) genomic regions with extremes in GC content[27]; or (4) samples contributing to multiple CNVRs. To adjust for siblings in the AGRE data, we calculated a permutation-based P value ($\times$1,000), in which disease labels for siblings were permutated together. DAVID (Database for Annotation, Visualization, and Integrated Discovery)[28] assessed the significance of functional annotation clustering. We considered loci significant between cases and controls ($P < 0.05$) where ACC discovery cases had overlapping variation, replicated in AGRE or were not observed in control subjects, and validated with another method (qPCR Roche Universal Probe Library using qBase[29], MRC-Holland MLPA, and Affymetrix 5.0 from Broad). Statistical correction of five deletion and nine duplication CNVRs, on the basis of discovery cohort (ACC) significance and signal review is appropriate for our study ('CNV Filtering Steps' in Supplementary Materials).

Received 13 November 2008; accepted 27 February 2009.
Published online 28 April 2009.

1. Autism and Developmental Disabilities Monitoring Network. ⟨http://www.cdc.gov/mmwr/pdf/ss/ss5601.pdf⟩ (2007).
2. Newschaffer, C. J. *et al.* The epidemiology of autism spectrum disorders. *Annu. Rev. Public Health* **28,** 235–258 (2007).
3. Gupta, A. R. & State, M. W. Recent advances in the genetics of autism. *Biol. Psychiatry* **61,** 429–437 (2007).
4. Klauck, S. M. Genetics of autism spectrum disorder. *Eur. J. Hum. Genet.* **14,** 714–720 (2006).
5. Vorstman, J. A. S. *et al.* Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism. *Mol. Psychiatry* **11,** 18–28 (2006).
6. Sebat, J. *et al.* Strong association of de novo copy number mutations with autism. *Science* **316,** 445–449 (2007).
7. Szatmari, P. *et al.* Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nature Genet.* **39,** 319–328 (2007).
8. Marshall, C. R. *et al.* Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.* **82,** 477–488 (2008).
9. Weiss, L. A. *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* **358,** 667–675 (2008).
10. Kim, H. G. *et al.* Disruption of neurexin 1 associated with autism spectrum disorder. *Am. J. Hum. Genet.* **82,** 199–207 (2008).
11. Roohi J. . *et al.* Disruption of contactin 4 in three subjects with autism spectrum disorder. *J. Med. Genet.* **46,** 176–182 (2008).
12. Fernandez, T. *et al.* Disruption of *Contactin 4* (*CNTN4*) results in developmental delay and other features of 3p deletion syndrome. *Am. J. Hum. Genet.* **82,** 1385 (2008).
13. Le Couteur, A. *et al.* Diagnosing autism spectrum disorders in pre-school children using two standardised assessment instruments: the ADI-R and the ADOS. *J. Autism Dev. Disord.* **38,** 362–372 (2008).
14. Geschwind, D. H. *et al.* The autism genetic resource exchange: a resource for the study of autism and related neuropsychiatric conditions. *Am. J. Hum. Genet.* **69,** 463–466 (2001).
15. Wang, K. *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **17,** 1665–1674 (2007).
16. Kalscheuer, V. M. *et al.* Mutations in *autism susceptibility candidate 2* (*AUTS2*) in patients with mental retardation. *Hum. Genet.* **121,** 501–509 (2007).
17. Jamain, S. *et al.* Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nature Genet.* **34,** 27–29 (2003).
18. Moessner, R. *et al.* Contribution of SHANK3 mutations to autism spectrum disorder. *Am. J. Hum. Genet.* **81,** 1289–1297 (2007).
19. Kent W. J. . *et al.* The human genome browser at UCSC. *Genome Res.* **12,** 996–1006 (2002).
20. Yi, J. J. & Ehlers, M. D. Ubiquitin and protein turnover in synapse function. *Neuron* **47,** 629–632 (2005).
21. Kitada, T. *et al.* Mutations in the *parkin* gene cause autosomal recessive juvenile parkinsonism. *Nature* **392,** 605–608 (1998).
22. Tai, H.-C. & Schuman, E. Ubiquitin, the proteasome and protein degradation in neuronal function and dysfunction. *Nature Rev. Neurosci.* **9,** 826–838 (2008).
23. Alarcón, M. *et al.* Linkage, association, and gene-expression analyses identify *CNTNAP2* as an autism-susceptibility gene. *Am. J. Hum. Genet.* **82,** 150–159 (2008).
24. Chubykin, A. A. *et al.* Dissection of synapse induction by neuroligins: effect of a neuroligin mutation associated with autism. *J. Biol. Chem.* **280,** 22365–22374 (2005).
25. Zheng, C., Heintz, N. & Hatten, M. E. CNS gene encoding astrotactin, which supports neuronal migration along glial fibers. *Science* **272,** 417–419 (1996).
26. Kahler, A. K. *et al.* Association analysis of schizophrenia on 18 genes involved in neuronal migration: MDGA1 as a new susceptibility gene. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **147B,** 1089–1100 (2008).
27. Diskin, S. *et al.* Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Res.* **36,** e126 (2008).
28. Dennis, G. Jr *et al.* DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biol.* **4,** R60 (2003).
29. Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. & Vandesompele, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* **8,** R19 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We gratefully thank all the ASD children and their families at the participating study sites who were enrolled in this study and all the control subjects who donated blood samples to Children's Hospital of Philadelphia for genetic research purposes. We thank the technical staff in the Center for Applied Genomics, Children's Hospital of Philadelphia for generating the genotypes used in this study. We thank S. Diskin for her contribution to the discussion on the effect of wave artifacts on CNV calling. We also thank S. Kristinsson, L. A. Hermannsson and A. Krisbjörnsson for their software design and contribution. This research was financially supported by The Children's Hospital of Philadelphia, Autism Speaks, and NICHD (HD35476). We also gratefully acknowledge the resources provided by the AGRE Consortium (D. H. Geschwind, M. Bucan, W. T. Brown, J. D. Buxbaum, R. M. Cantor, J. N. Constantino, T. C. Gilliam, C. M. Lajonchere, D. H. Ledbetter, C. Lese-Martin, J. Miller, S. F. Nelson, G. D. Schellenberg, C. A. Samango-Sprouse, S. Spence, M. State, R. E. Tanzi) and the participating families. AGRE is a program of Autism Speaks and is at present supported, in part, by grant 1U24MH081810 from the National Institute of Mental Health to C. M. Lajonchere (PI), and formerly by grant MH64547 to D. H. Geschwind (PI). The AGRE data set was genotyped by the Center for Applied Genomics at the Children's Hospital of Philadelphia, which funded all genotyping for this study, and the complete sets of genotyping data have been released to the public domain. AGRE-approved academic researchers can acquire the data sets from AGRE (http://www.agre.org). The study is supported in part by Research Award from the Margaret Q. Landenberger Foundation (H.H.), UL1-RR024134-03 (H.H.), a Research Development Award from the Cotswold Foundation (H.H. and S.F.A.G.), the Beatrice and Stanley A. Seaver Foundation (J.D.B.), the Department of Veterans Affairs (G.D.S.), and National Institute of Health grants HD055782-01 (J.Munson, A.E., O.K., G.D. and G.D.S.), MH0666730 (J.D.B.), MH061009 and NS049261 (J.S.S.), HD055751 (E.H.C.), U10MH66766-02S1

©2009 Macmillan Publishers Limited. All rights reserved

(C.J.M.), MH69359, M01-RR00064 and the Utah Autism Foundation (H.C., J.Miller, and W.M.M.). All genotyping for this study was supported by an Institutional Development Award to the Center for Applied Genomics (H.H.) from the Children's Hospital of Philadelphia. We acknowledge the Autism Genome Project Consortium (J.P., C.W.B., T.H.W., W.M.M., H.C., J.I.N., J.S.S., E.H.C., J.Munson, A.E., O.K., J.D.B., B.D. and G.D.S.) funded by Autism Speaks, the Medical Research Council (UK) and the Health Research Board (Ireland). We are grateful for the public access to the Affymetrix 5.0 data available for a subset of AGRE families (http://www.agre.org), which served as important validation of CNV calls made.

Author Contributions H.H. and G.D.S. designed the study and supervised the data analysis and interpretation. J.T.G., K.W. and B.D. conducted the statistical analyses. C.E.K. and E.C.F. directed the genotyping of stage 1. J.D.B. coordinated the validation. G.C. and O.K. performed qPCR and MLPA validation of CNVs and edited the manuscript. J.T.G., K.W., H.H., G.D.S. and B.D. drafted the manuscript. G.D.S., N.J.M., E.H.C., W.M.M., H.C., T.H.W., J.D.B., T.O., J.I.N., E.A., L.S., J.R., T.S., C.B., C.J.M., D.J.P. and D.Z. collected samples and contributed phenotype data for the study and assisted with data collection and manuscript preparation. E.C., S.F.A.G., P.S., M.I., B.D., L.K., S.W. and K.W. reviewed the data, assisted with interpretation of the data, and edited the manuscript. Other authors contributed to sample acquisition and processing or to data analysis and interpretation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to H.H. (hakonarson@chop.edu) or G.D.S (gerardsc@mail.med.upenn.edu).

©2009 Macmillan Publishers Limited. All rights reserved