SUPPLEMENTARY INFORMATION

Functional impact of global rare copy number variation in autism spectrum disorders

Dalila Pinto[1], Alistair T. Pagnamenta[2], Lambertus Klei[3], Richard Anney[4], Daniele Merico[5], Regina Regan[6], Judith Conroy[6], Tiago R. Magalhaes[7,8], Catarina Correia[7,8], Brett S. Abrahams[9], Joana Almeida[10], Elena Bacchelli[11], Gary D. Bader[5], Anthony J. Bailey[12], Gillian Baird[13], Agatino Battaglia[14], Tom Berney[15,56], Nadia Bolshakova[4], Sven Bölte[16], Patrick F. Bolton[17], Thomas Bourgeron[18], Sean Brennan[4], Jessica Brian[19], Susan E. Bryson[20], Andrew R. Carson[1], Guillermo Casallo[1], Jillian Casey[6], Brian H.Y. Chung[1], Lynne Cochrane[4], Christina Corsello[21], Emily L. Crawford[22], Andrew Crossett[23], Cheryl Cytrynbaum[1], Geraldine Dawson[24,25], Maretha de Jonge[26], Richard Delorme[27], Irene Drmic[19], Eftichia Duketis[16], Frederico Duque[10], Annette Estes[28], Penny Farrar[2], Bridget A. Fernandez[29], Susan E. Folstein[30], Eric Fombonne[31], Christine M. Freitag[16], John Gilbert[30], Christopher Gillberg[32], Joseph T. Glessner[33], Jeremy Goldberg[34], Andrew Green[6], Jonathan Green[35], Stephen J. Guter[36], Hakon Hakonarson[33,37], Elizabeth A. Heron[4], Matthew Hill[4], Richard Holt[2], Jennifer L. Howe[1], Gillian Hughes[4], Vanessa Hus[21], Roberta Igliozzi[14], Cecilia Kim[33], Sabine M. Klauck[38], Alexander Kolevzon[39], Olena Korvatska[40], Vlad Kustanovich[41], Clara M. Lajonchere[41], Janine A. Lamb[42], Magdalena Laskawiec[12], Marion Leboyer[43], Ann Le Couteur[15,56], Bennett L. Leventhal[44,45], Anath C. Lionel[1], Xiao-Qing Liu[1], Catherine Lord[21], Linda Lotspeich[46], Sabata C. Lund[22], Elena Maestrini[11], William Mahoney[47], Carine Mantoulan[48], Christian R. Marshall[1], Helen McConachie[15,56], Christopher J. McDougle[49], Jane McGrath[4], William M. McMahon[50], Alison Merikangas[4], Ohsuke Migita[1], Nancy J. Minshew[51], Ghazala K. Mirza[2], Jeff Munson[52], Stanley F. Nelson[53], Carolyn Noakes[19], Abdul Noor[54], Gudrun Nygren[32], Guiomar Oliveira[10], Katerina Papanikolaou[55], Jeremy R. Parr[56], Barbara Parrini[14], Tara Paton[1], Andrew Pickles[57], Marion Pilorge[58], Joseph Piven[59], Chris P. Ponting[60], David J. Posey[49], Annemarie Poustka[38]‡, Fritz Poustka[16], Aparna Prasad[1], Jiannis Ragoussis[2], Katy Renshaw[12], Jessica Rickaby[1], Wendy Roberts[19], Kathryn Roeder[23], Bernadette Roge[48], Michael L. Rutter[61], Laura J. Bierut[62], John P. Rice[62], Jeff Salt[36], Katherine Sansom[1], Daisuke Sato[1], Ricardo Segurado[4], Ana F. Sequeira[7,8], Lili Senman[19], Naisha Shah[6], Val C. Sheffield[63], Latha Soorya[39], Inês Sousa[2], Olaf Stein[64], Nuala Sykes[2], Vera Stoppioni[65], Christina Strawbridge[34], Raffaella Tancredi[14], Katherine Tansey[4], Bhooma Thiruvahindrapduram[1], Ann P. Thompson[34], Susanne Thomson[22], Ana Tryfon[39], John Tsiantis[55], Herman Van Engeland[26], John B. Vincent[54], Fred Volkmar[66], Simon Wallace[12], Kai Wang[33], Zhouzhi Wang[1], Thomas H. Wassink[67], Caleb Webber[60], Rosanna Weksberg[1], Kirsty Wing[2], Kerstin Wittemeyer[48], Shawn Wood[3], Jing Wu[23], Brian L. Yaspan[22], Danielle Zurawiecki[39], Lonnie Zwaigenbaum[68], Joseph D. Buxbaum[39*], Rita M. Cantor[53*], Edwin H. Cook[36*], Hilary Coon[50*], Michael L. Cuccaro[30*], Bernie Devlin[3*], Sean Ennis[6*], Louise Gallagher[4*], Daniel H. Geschwind[9*], Michael Gill[4*], Jonathan L. Haines[69*], Joachim Hallmayer[46*], Judith Miller[50], Anthony P. Monaco[2*], John I. Nurnberger Jr[49*], Andrew D. Paterson[1*], Margaret A. Pericak-Vance[30*], Gerard D. Schellenberg[70*], Peter Szatmari[34*], Astrid M. Vicente[7,8*], Veronica J. Vieland[64*], Ellen M. Wijsman[71*], Stephen W. Scherer[1,72*], James S. Sutcliffe[22*] & Catalina Betancur[58*]

May 27, 2010

[1]The Centre for Applied Genomics and Program in Genetics and Genomic Biology, The Hospital for Sick Children, Toronto, Ontario M5G 1L7, Canada. [2]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK. [3]Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15213, USA. [4]Autism Genetics Group, Department of Psychiatry, School of Medicine, Trinity College, Dublin 8, Ireland. [5]Banting and Best Department of Medical Research, Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario, M5S 3E1, Canada. [6]School of Medicine and Medical Science University College, Dublin 4, Ireland. [7]Instituto Nacional de Saude Dr Ricardo Jorge, 1649-016 Lisbon, and Instituto Gulbenkian de Cîencia, 2780-156 Oeiras, Portugal. [8]BioFIG—Center for Biodiversity, Functional and Integrative Genomics, Campus da FCUL, C2.2.12, Campo Grande, 1749-016 Lisboa, Portugal. [9]Program in Neurogenetics, Department of Neurology and Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine at UCLA.[10]Hospital Pediátrico de Coimbra, 3000 – 076 Coimbra, Portugal. [11]Department of Biology, University of Bologna, 40126 Bologna, Italy. [12]Department of Psychiatry, University of Oxford, Warneford Hospital, Headington, Oxford OX3 7JX, UK. [13]Newcomen Centre, Guy's Hospital, London SE1 9RT, UK. [14]Stella Maris Institute for Child and Adolescent Neuropsychiatry, 56128 Calambrone (Pisa), Italy. [15]Child and Adolescent Mental Health, University of Newcastle, Sir James Spence Institute, Newcastle upon Tyne NE1 4LP, UK. [16]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, J.W. Goethe University Frankfurt, 60528 Frankfurt, Germany. [17]Department of Child and Adolescent Psychiatry, Institute of Psychiatry, London SE5 8AF, UK. [18]Human Genetics and Cognitive Functions, Institut Pasteur; University Paris Diderot-Paris 7, CNRS URA 2182, Fondation FondaMental, 75015 Paris, France. [19]Autism Research Unit, The Hospital for Sick Children and Bloorview Kids Rehab, University of Toronto, Toronto, Ontario M5G 1X8, Canada. [20]Department of Pediatrics and Psychology, Dalhousie University, Halifax, Nova Scotia B3K 6R8, Canada. [21]Autism and Communicative Disorders Centre, University of Michigan, Ann Arbor, Michigan 48109-2054, USA. [22]Department of Molecular Physiology and Biophysics, Vanderbilt Kennedy Center, and Centers for Human Genetics Research and Molecular Neuroscience, Vanderbilt University, Nashville, Tennessee 37232, USA. [23]Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA. [24]Autism Speaks, New York 10016, USA. [25]Department of Psychiatry, University of North Carolina, Chapel Hill, North Carolina 27599-3366, USA. [26]Department of Child Psychiatry, University Medical Center, Utrecht 3508 GA, The Netherlands. [27]INSERM U 955, Fondation FondaMental, APHP, Hôpital Robert Debré, Child and Adolescent Psychiatry, 75019 Paris, France. [28]Department of Speech and Hearing Sciences, University of Washington, Seattle, Washington 98195, USA. [29]Disciplines of Genetics and Medicine, Memorial University of Newfoundland, St John's Newfoundland A1B 3V6, Canada. [30]The John P. Hussman Institute for Human Genomics, University of Miami, Miami, Florida 33101, USA. [31]Division of Psychiatry, McGill University, Montreal, Quebec H3A 1A1, Canada. [32]Department of Child and Adolescent Psychiatry, Göteborg University, Göteborg S41345, Sweden. [33]The Center for Applied Genomics, Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. [34]Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Ontario L8N 3Z5, Canada. [35]Academic Department of Child Psychiatry, Booth Hall of Children's Hospital, Blackley, Manchester M9 7AA, UK. [36]Institute for Juvenile Research, Department of Psychiatry, University of Illinois at Chicago, Chicago, Illinois 60612,, USA. [37]Department of Pediatrics, Children's Hospital of Philadelphia, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA. [38]Division of Molecular Genome Analysis, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany. [39]The Seaver Autism Center for Research and Treatment, Department of Psychiatry, Mount Sinai School of Medicine, New York 10029, USA. [40]Department of Medicine, University of Washington, Seattle, Washington 98195, USA. [41]Autism Genetic Resource Exchange, Autism Speaks, Los Angeles, California 90036-4234, USA. [42]Centre for Integrated Genomic Medical Research, University of

Manchester, Manchester M13 9PT, UK. [43]INSERM U995, Department of Psychiatry, Groupe Hospitalier Henri Mondor-Albert Chenevier, AP-HP; University Paris 12, Fondation FondaMental, Créteil 94000, France. [44]Nathan Kline Institute for Psychiatric Research (NKI), 140 Old Orangeburg Road, Orangeburg, New York 10962, USA. [45]Department of Child and Adolescent Psychiatry, New York University and NYU Child Study Center, 550 First Avenue, New York, New York 10016, USA. [46]Department of Psychiatry, Division of Child and Adolescent Psychiatry and Child Development, Stanford University School of Medicine, Stanford, California 94304, USA. [47]Department of Pediatrics, McMaster University, Hamilton, Ontario L8N 3Z5, Canada. [48]Centre d'Eudes et de Recherches en Psychopathologie, University de Toulouse Le Mirail, Toulouse 31200, France. [49]Department of Psychiatry, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. [50]Psychiatry Department, University of Utah Medical School, Salt Lake City, Utah 84108, USA. [51]Departments of Psychiatry and Neurology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15213, USA. [52]Department of Psychiatry and Behavioural Sciences, University of Washington, Seattle, Washington 98195, USA. [53]Department of Human Genetics, University of California—Los Angeles School of Medicine, Los Angeles, California 90095, USA. [54]Centre for Addiction and Mental Health, Clarke Institute and Department of Psychiatry, University of Toronto, Toronto, Ontario M5G 1X8, Canada. [55]University Department of Child Psychiatry, Athens University, Medical School, Agia Sophia Children's Hospital, 115 27 Athens, Greece. [56]Insitutes of Neuroscience and Health and Society, Newcastle University, Newcastle Upon Tyne NE1 7RU, UK. [57]Department of Medicine, School of Epidemiology and Health Science, University of Manchester, Manchester M13 9PT, UK. [58]INSERM U952 and CNRS UMR 7224 and UPMC Univ Paris 06, Paris 75005, France. [59]Carolina Institute for Developmental Disabilities, University of North Carolina at Chapel Hill, North Carolina 27599-3366, USA. [60]MRC Functional Genomics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, OX1 3QX, UK. [61]Social, Genetic and Developmental Psychiatry Centre, Institute Of Psychiatry, London SE5 8AF, UK. [62]Department of Psychiatry, Washington University in St Louis, School of Medicine, St Louis, Missouri 63130, USA. [63]Department of Pediatrics and Howard Hughes Medical Institute Carver College of Medicine, University of Iowa, Iowa City, Iowa 52242, USA. [64]Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital and The Ohio State University, Columbus, Ohio 43205, USA. [65]Neuropsichiatria Infantile, Ospedale Santa Croce, 61032 Fano, Italy. [66]Child Study Centre, Yale University, New Haven, Connecticut 06520, USA. [67]Department of Psychiatry, Carver College of Medicine, Iowa City, Iowa 52242, USA. [68]Department of Pediatrics, University of Alberta, Edmonton, Alberta T6G 2J3, Canada. [69]Center for Human Genetics Research, Vanderbilt University Medical Centre, Nashville, Tennessee 37232, USA. [70]Pathology and Laboratory Medicine, University of Pennsylvania, Pennsylvania 19104, USA. [71]Departments of Biostatistics and Medicine, University of Washington, Seattle, Washington 98195, USA. [72]Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A1, Canada.

CONTENTS

A. AUTISM SPECTRUM DISORDER (ASD) SAMPLE AND CONTROL COLLECTIONS 4

ASD samples 4

Control cohorts 4

B. GENOTYPING AND DATA CLEANING 6

SNP quality control 6

Intensity quality control for CNV detection 6

C. CNV DETECTION AND QUALITY CONTROL EVALUATION 7

Pilot experiment to evaluate the quality of detected stringent CNVs 8

D. CNV VERIFICATION 9

Confirmation using computational methods 9

Experimental CNV validation 9

E. RARE CNV BURDEN ANALYSIS 11

CNV global burden analysis 12

CNV Region (CNVR) burden analysis 12

CNV-based gene association test 13

Burden analysis for genes known to be implicated in ASD and/ or ID 13

Population attributable risk 13

F. GENE-SET ENRICHMENT AND FUNCTIONAL MAP 14

Analytical synopsis 14

Deriving gene sets 14

Gene-set enrichment test 15

Cryptic bias 15

Network visualization of gene-sets enriched for deletions: functional enrichment map 15

Expanded functional map between deletion-enriched gene sets & known ASD/ID genes 16

G. SUPPLEMENTARY FIGURES AND TABLES 18

H. ACKNOWLEDGEMENTS 72

I. SUPPLEMENTARY REFERENCES 73

LIST OF FIGURES

Figure 1. Quality control and analysis flow chart 19
Figure 2. Results from ancestry analysis using genome-wide genotype data 20
Figure 3. Global CNV measures for blood versus cell-line DNA-derived samples 21
Figure 4. Examples of CNVs overlapping ASD candidate genes or loci 22
Figure 5. Number of enriched gene-sets at different FDR q-value thresholds 26
Figure 6. Overview of all enriched gene-sets in deletions 27
Figure 7. Detailed annotation of the main enriched clusters 28
Figure 8. Control for bias in length and number of deletions 29
Figure 9. Bias control for genome proximity 31
Figure 10. Distribution of q-values per gene-set 32
Figure 11. Gene coverage between ASD and/or ID gene-sets and sets enriched for deletions 33
Figure 12: Expanded functional map with high-level of annotation detail 34

LIST OF TABLES

Table 1. Quality control (QC) steps for CNV analysis of ASD cases and controls 35
Table 2. Chromosome abnormalities larger than 7.5 Mb detected during QC 36
Table 3: Summary of characteristics of stringent CNVs in cases and controls 38
Table 4. Characteristics of rare CNVs in European ASD probands and controls 39
Table 5. Global rare CNV burden analyzes with respect to CNV size and CNV rate 42
Table 6. Examples of ASD candidate genes or loci 43
Table 7. Rare CNVs confirmed experimentally 44
Table 8. Rare CNVs in 996 ASD cases 60
Table 9. List of known ASD genes, ID genes, and ASD-candidates 61
Table 10. Clinically-relevant findings 66
Table 11. Population attributable risk (PAR) 69
Table 12. Number of analyzed gene-sets in the functional map enriched for deletions 70
Table 13. List of gene-sets enriched for deletions 71

A. AUTISM SPECTRUM DISORDER (ASD) SAMPLE AND CONTROL COLLECTIONS

ASD samples

In an ongoing effort, the international Autism Genome Project (AGP) Consortium is collecting ASD families for ongoing genetic studies. The first phase of this initiative involved examining genetic linkage and chromosomal rearrangements in 1,168 families having at least two ASD individuals[1]. In this second phase of the project, we collected more families and genotyped them to examine for CNVs and SNPs affecting risk for ASD. Here, we present the analysis of rare CNVs, which is outlined in Fig. 1 in the main text. As discussed, 1,275 ASD cases (or 1,256 cases with both parents) were available for genotyping for our study (Fig.1). DNA was obtained from blood (63%), buccal-swabs (10%) or cell-lines (22%) (in 5% the DNA source was not available). The Autism Diagnostic Interview-Revised (ADI-R)[2] and Autism Diagnostic Observation Schedule (ADOS)[3] were used for research diagnostic classification. Subjects with previously known karyotypic abnormalities or other genetic disorders associated with ASD were excluded.

Our ascertainment of ASD cases was based on the following criteria. Affected subjects were grouped in three classes (strict, broad and spectrum ASD) based on proband diagnostic measures. To qualify for the strict class, affected individuals met criteria for autism on both primary instruments, the ADI-R and the ADOS. The broad class included individuals who met ADI-R criteria for autism and ADOS criteria for ASD, but not autism, or vice versa. ADI-R-based diagnostic classification of subjects as ASD followed criteria published by Risi *et al*.[4]. Specifically, individuals who almost met ADI criteria for autism were classified as ASD if (1) they met criteria on social and either communication or repetitive behavior domains; or (2) met criteria on social and within 2 points of criteria for communication, or met criteria on communication and within 2 points of social criteria, or within 1 point on both social and communication domains[4]. Finally, the spectrum class included all individuals who were classified as ASD on both the ADI-R and ADOS or who were not evaluated on one of the instruments but were diagnosed with autism on the other instrument. Subjects from all classifications (strict, broad, and spectrum) were included in the CNV analysis.

Family-history reports were taken to inform on the family type. Multiplex (MPX) families had at least two individuals receiving validated ASDs diagnoses who were first to third degree relatives (for third degree, only considered cousins). This included families with affected dizygotic twins. Simplex (SPX) families had only one known individual with ASD in first to third (cousin) degree relatives. Families with only affected monozygotic twins were considered SPX. Unknown (UKN) families were any families that did not fall into the MPX or SPX criteria above. Given the international and multi-site nature of the project and range of chronological and mental age of the probands, a range of cognitive tests were administered, and standard scores were combined across tests to provide consolidated IQ estimates.

Control cohorts

Our primary considerations for selecting control groups for the genome-wide CNV comparison studies included using individuals with no obvious psychiatric history that were ancestrally matched samples to minimize potential confounds, and that were genotyped on the same Illumina 1M platform used for genotyping cases. It was not possible under these circumstances to control for age. Wherever possible we also considered and controlled for influences of sample source (eg. blood or cell line DNA), DNA preparation methods and experimental batch effects. We note that at the time of initiation, this project was one of the first large-scale projects to use the high-resolution Illumina Infinium 1M-single SNP microarray and this needed to be considered in our planning.

Two control groups, both genotyped with Illumina Human 1M-single BeadChip arrays, were assembled and used in the primary analysis: subjects from the Study on Addiction: Genetics and Environment (SAGE) and from HapMap CEPH Utah (HapMap CEU).

1. *SAGE cohort*: 1,880 control subjects from the larger SAGE case-control study[5] were made available. The consented sample included 31% males and 69% of females, with mean age of 39.2 (SD 9.1) and 73% of subjects self-identified as European-American, 26% as African-American and 1% as other (http://zork.wustl.edu/gei/study_description.htm). Both raw intensities and genotypes were obtained from NHGRI-dbGaP (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000092.v1.p1). The SAGE control subjects may have had exposure to alcohol (and possibly to other drugs), but did not meet criteria for alcohol or other drug dependence. The sub-set of control dataset used in the specific CNV analyses in this paper is composed of 1,287 unrelated European control samples that passed all quality control filters (75% had DNA extracted from whole blood and 25% had DNA extracted from cell lines).

2. *HapMap CEU*: data from 120 samples were obtained directly from Illumina. Data from 26 HapMap CEU samples that passed all quality control filters were used in the CNV analysis. Experiments followed the same quality control procedures as applied to the ASD family samples and SAGE controls.

To further evaluate specific CNVs identified in the primary analysis as potential risk factors in ASD (Supplementary Table 6), we tested them against an additional 3,677 European controls. These controls were selected because we had previous experience in using them in CNV comparison studies. As discussed below, while this data comes from other microarray platforms, we were able to use established methods to compare presence or absence of a particular CNV at a given locus:

1. *Coronary artery disease case-control GWAS study*: This “Ottawa Heart Institute (OHI)” study[6] was undertaken in the Ottawa valley population, province of Ontario, Canada. DNA samples were all extracted from whole blood. Both raw intensities and genotypes were obtained directly from the authors of the study. We used CNV data from 1,234 unrelated controls, genotyped with the Affymetrix Genome-Wide Human SNP 6.0 array, which were analyzed using two algorithms Birdsuite[7] and iPattern[8]. CNVs detected by both algorithms were used to define a stringent set of CNVs, in which CNVs detected in a subject by both algorithms were merged using the outside probe boundaries of the union of those CNVs. The control dataset used in the CNV analysis was composed of 53.7% of female and 46.3% male samples that passed quality control.

2. *PopGen study*[9]: 1,123 European control subjects were assembled from the PopGen study and genotyped with the Affymetrix Genome-Wide Human SNP 6.0 array. The data were analyzed for CNVs in the same way for as the OHI dataset (see above). All DNA samples were all extracted from whole blood.

3. *CHOP study***:** 1,320 European control subjects routinely seen at primary care and well-child clinic practices within the Children's Hospital of Philadelphia (CHOP) Health Care were genotyped with Illumina 550K BeadChip and analyzed using the circular binary segmentation method[10]. We used the CNV data as listed in the Database of Genomic Variants (ie. CNVs found in two or more samples were merged to obtain a stringent set of CNVs).

B. GENOTYPING AND DATA CLEANING

Samples were genotyped using the Illumina 1M-single array, and we performed stringent, uniform quality control (QC) procedures on the resulting data. The Illumina 1M-single Infinium BeadChip contains a total of 1,072,820 markers (50-mer probes) for SNP and CNV analyses. Samples were processed using the manufacturer's recommended protocol with no modifications for Infinium II arrays, and BeadChips were scanned on the Illumina BeadArray Reader using default settings. Analysis and intra-chip normalization were performed using Illumina's BeadStudio software v.3.3.7, with a GenCall cutoff of 0.1. Built-in controls, both sample independent (including staining controls, extension controls, target removal controls, and hybridization controls) and sample-dependent (including stringency controls, non-specific binding controls, and non-polymorphic controls), were inspected to assess the quality of the experiment. For genotype calling we followed the manufacturer's protocols[11] and used technical controls. Trios consisting of an affected offspring and both parents were genotyped, and in total genotyping was completed for 1,275 individuals (1,256 complete trios). For the control sample, 1,880 samples were genotyped on the 1M Illumina platform, as described elsewhere (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000092.v1.p1)[5].

SNP quality control

QC for individual SNPs was performed at family and individual levels as part of a whole genome association study of ASD[12] (Supplementary Fig. 1). We first assessed gender miscalls based on X chromosome genotypes and allele calls for Y, adjusting gender when appropriate (e.g., miscoding) and dropping samples (e.g., Klinefelter syndrome) or genotypes (e.g., loss of X in cell line) from the X chromosome. We searched the database for duplicate samples using a subset of 5,254 SNPs that were independent and had a >99.9% completion rate for genotypes at this QC stage. Duplicates, and highly-related samples based on identity by state (IBS), from 2 families were removed. Data were subsequently checked for Mendelian errors, and 13 families with large numbers of errors were removed from the analysis. Following this QC step, 1,213 genotyped cases were retained for 1,184 complete trios and 93 parent-child duos (Supplementary Table 1).

Ancestry was then determined for the proband by using 5,239 widely-spaced, independent SNPs that had a genotype completion rate of ≥99.9%. Software used was SpectralGEM[13], which estimated 5 significant dimensions of ancestry (Supplementary Fig. 2 and unpublished data). Subsequent clustering on dimensions of ancestry resulted in 6 clusters: 3 clusters of European ancestry (EA), and 3 clusters reflecting other major ancestral groups (e.g., African and Asian). We have used only the three clusters of European ancestry for the CNV analyses.

Intensity quality control for CNV detection

Prior to CNV analyses, QC criteria were applied as outlined in Supplementary Fig. 1. These QC criteria were selected to ensure that ascertainment of CNVs was consistent between cases and parents, to reduce heterogeneity due to varying sample sources and DNA treatment, and to avoid batch effects. To minimize potential intensity batch effects for CNV analysis, we generated a cluster file made of 1,550 samples internal to the project (i.e., mostly parents) that was used to normalize and re-cluster all samples of the project for better CNV calling. This internal cluster file was inspected manually for all markers with GenTrain score <0.6 and markers with low AB R Mean values (i.e., low normalized intensity of the heterozygote cluster) and any edits were validated by two independent raters. The edited cluster was then used to normalize and re-cluster all samples of the project, and to generate genotype calls using a GenCall (Illumina) cutoff score of 0.10. Additional thresholds were set for a variety of QC measures (Supplementary Fig. 1, Supplementary Table 1), and we excluded samples if:

i) based on the SNP QC (see above): if they were closely related, duplicates, had gender mismatches, Mendelian error rate >1%;
ii) array call rate <97%;
iii) cross-sample normalized ratio standard deviation >0.27 (i.e., sample-batch level QC);
iv) standard deviation for log R ratio values in the autosomes >0.27;
v) standard deviation of the B Allele Frequency values (i.e., allelic ratios within the 0.25 to 0.75 ranges) >0.13.

C. CNV DETECTION AND QUALITY CONTROL EVALUATION

For 1,159 autism samples (1,045 trios and 106 duos) and 1,907 controls that passed the above SNP and intensity QC filters, we ran three CNV calling algorithms, namely,QuantiSNP[14], iPattern[8] and PennCNV[15] to obtain high-confidence call. The required data for CNV analysis, i.e. within-sample normalized fluorescence (ie. X and Y normalized values), between-sample normalized fluorescence (ie. Log R ratios (LRR) and B allele frequency (BAF) values) and genotypes for each sample, were exported directly from Illumina's Beadstudio software. Mitochondrial markers were not used in the CNV analysis.

The objective of using multiple algorithms was to minimize the number of potential false discoveries. CNVs were identified by using QuantiSNP and iPattern, while PennCNV, and specifically its trio option, was used to confirm inheritance status of the resulting CNV calls. Because each of these algorithms employs unique strategies for CNV calling, their strengths can be leveraged to ensure maximum specificity. We note that the lead individuals who developed the QuantiSNP (Jiannnis Ragoussis and team), iPattern (Dalila Pinto/ Steve Scherer and team) and PennCNV (Kai Wang and team) algorithms from Oxford, Toronto, and Philadelphia are all co-authors of this study. In our collective experience through this and other projects, these three algorithms currently remain the most reliable for calling CNVs using Illumina 1M data.

iPattern implements a non-parametric density-based clustering model that integrates intensity data across samples to assign individual samples to distinct copy number states. iPattern data pre-processing produces a single one-dimensional summary of the relative intensity of test to reference samples. Specifically, data pre-processing evaluates the background signal to noise ratio for each batch of tested samples, and outliers from the standard deviation of the sample batch are removed. Normalization of chromosome X probes was performed separately for males and females before CNV calling of this chromosome. A two-stage analytical framework is then used to identify CNV regions, with a moving window-based approach followed by secondary boundary refinement. The largest cluster of unrelated samples is dynamically chosen as reference, and samples with higher or lower intensities are assigned as relative CNV gains or losses. CNV lengths were calculated based on the distance between the first and last array probes internal to the variant. QuantiSNP uses an Objective Bayes (OB) Hidden-Markov Model (HMM) approach for CNV calling in which OB measures are used to set hyperparameters (false positive rates) and the copy number state is inferred with an HMM (for details see Colella *et al.*[14]). The PennCNV algorithm uses an HMM to detect CNV from multiple sources of information, including the total signal intensity and allelic intensity ratio at each SNP marker (for details see Wang *et al.*[15]).

We excluded CNVs when they failed stringent quality control criteria: <5 probes and low confidence score log Bayes factor <15; if they resided in regions of extreme GC content (>70%); or if they were within centromere proximal cytobands. We removed samples (Supplementary Table 1) that were outliers with respect to: (1) Excessive number of CNVs detected by either of

the two algorithms – we defined an outlier as the mean plus 3 standard deviations; (2) excessive aggregate length of CNVs (samples with CNVs larger than 7.5 Mb were removed after visual inspected by plotting their intensities and genotype ratios —Supplementary Table 2— these likely correspond to large karyotypic chromosome abnormalities, or cell line artifacts[16], and were confirmed for most of the cases); and, (3) excessive number of *de novo* CNVs (>5 *de novo* CNVs). All CNVs by any algorithm with size larger than 1 Mb were inspected manually, and all samples that passed all above QC filters were inspected for chromosomes X and Y. A total of 1,132 cases (of which 991 are complete trios, 125 are child-parent duos and 16 are case only) and 1,858 controls passed the above filters.

CNVs detected in one individual with a minimum 5 consecutive probes covering 5 kb of sequence were merged using outside probe boundaries (i.e., union of the CNVs). As a final step, we joined any CNVs that appeared to be artificially split by either of the calling algorithms and also removed any CNVs that spanned known large assembly gaps in hg18 (greater than 200 kb). The final CNV list comprised 18,075 events in 1,132 cases, 41,946 in 2,712 parents, and 34,394 events in 1,858 controls (Supplementary Table 3A). To avoid confounding by ancestry[13,17], all downstream CNV analyses used EA-only cases and controls. This final sample list comprised 15,583 events in 996 EA cases, 33,704 in 2,196 parents, and 22,573 events in 1,287 European controls (Supplementary Table 3B).

Pilot experiment to evaluate the quality of detected stringent CNVs

We experimentally tested the quality of the stringent CNV calls (i.e. calls by both QuantiSNP and iPattern, each with 5 probes, 5 kb cutoff; see further details regarding stringent CNVs in the main paper) by randomly selecting CNV regions in distinct size ranges, followed by experimental validation using quantitative PCR (qPCR). Four size ranges/bins were considered (<20 kb, 20-50 kb, 51-100 kb, >101 kb) and an equal number of CNV regions from each size bin were initially selected for validation. One initial assay was designed per CNV region and tested in triplicate. If this assay was not validated, two other assays were designed and tested in triplicate. In cases where it was not possible to design a workable qPCR assay for a CNV region (either due to the presence of segmental duplications, repeat elements or SNPs), then that CNV region was excluded and other CNV regions were chosen and tested in a similar manner. In total, 45 regions (90 assays) were tested and 87% (39/45) were validated. The six CNVs that were not validated are due to a combination of false positives (FPs) within the CNV calling step and false negatives (FNs) within the qPCR validation step (see below).

Of the 45 CNVs detected in children, 21 were predicted to be inherited (19 were validated; 90%) while 24 were predicted to be *de novo* (20 were validated; 83%) using child-parent trios. To determine the accuracy of these predictions, additional qPCR was conducted using parental DNA. All 19 validated CNVs that were predicted to be inherited were confirmed to be inherited by qPCR and 16/20 validated CNVs that were predicted to be *de novo* were confirmed to be *de novo*. The other 4 CNVs were found to be inherited, representing FNs in the parents during the CNV calling step.

Upon closer inspection of the FNs and FPs, we found that most of these CNV regions overlap segmental duplications (i.e., blocks >10 kb in size, >95% identity) or were small in size (<30 kb). CNVs overlapping segmental duplications are usually highly polymorphic in the general population[18] and show poor probe coverage in commercial SNP-based arrays. Smaller CNVs are also more likely to have poor probe coverage. As a result, the signal to noise ratio at these CNVs does not allow reliable identification of distinct copy-number classes[19]. Removal of these CNVs then resulted in a validation rate >90% using laboratory techniques. Therefore, we found that

stringent application of quality score thresholds to CNV data substantially reduced false discoveries (FPs and FNs) arising from CNV calling error. To further evaluate if this generalization was valid, we tested an additional 40 predicted CNVs (from the above four size bins) that did not overlap segmental duplications. We observed a validation rate of 95% (38/40), similar to the stringent validation rate above. The lowest level of validation was observed in the 5-20 kb size bin.

Thus, all subsequent analyses described in the main paper were performed using a stringent set of CNVs with less than 50% of their length overlapping segmental duplications, ≥30 kb in size, and detected by at least 5 probes (median: 25 probes, 25th quartile: 6-17 probes). To improve the accuracy of predictions for inherited vs. *de novo* CNVs, additional computational methods were incorporated into CNV verification (see below), and the final compiled inherited status used in downstream analyses (described in the main paper). Finally, we further evaluated for differences in global CNV measures for blood versus cell-line samples and detected no significant differences and no significant differences in global CNV measures were detected for blood versus cell-line samples (Supplementary Fig. 3A).

D. CNV VERIFICATION

Confirmation using computational methods

Confirmatory evidence for CNVs was obtained from independent laboratory experiments (including qPCR, standard and long-range PCR, independent microarray platforms), and computational methods. The computational analyses included:

1. A third CNV calling method: PennCNV- trio option[15] was used for further confirmation of the inherited CNV status.

2. Proportion of genotype calls to verify deletions (% of homozygous calls *vs.* % of heterozygous calls *vs.* % of "no-calls"): deletions were considered to be verified when ≥ 90% SNPs were homozygous or "no calls.

3. Verification of *de novo* deletions seen in children: we used combined evidence from genotype proportions and fold intensity difference between the average intensity in the child-CNV and either parent-CNV. Specifically, deletions were considered to be *de novo* when ≥ 90% SNPs were homozygous or "no-calls" and the child-CNV average intensity was <4 fold lower than either of the parents.

4. Verification of *de novo* duplications seen in children: duplications were considered to be *de novo* when the child-CNV average intensity was >4 fold higher than either of the parents.

5. Visual inspection of chr. X and chr. Y intensities and genotype ratios B allele frequency for all samples that passed previous QC filters.

6. Visual inspection of all CNVs larger than 1 Mb by plotting their Log R ratios (LRR) and B allele frequencies (BAF). Samples (including ASD trios and controls) that passed all QC filters and showed CNVs larger than 7.5 Mb were further inspected manually by plotting their LRR as well as BAF. A molecular size cutoff of >7.5 Mb was selected to be consistent with large most cytogenetically visible chromosome abnormalities, and they were excluded; they are listed in Supplementary Table 2.

Experimental CNV validation

Putative *de novo* as well as inherited rare CNV were further confirmed by a qPCR, multiplex ligation-dependent probe amplification (MLPA), long-range PCR and/or independent arrays:

Quantitative-PCR: CNV regions were validated by at least two independent assays that showed concordant results in a family trio, using either Sybr-Green (Stratagene), or the Universal Probe

Library (UPL, Roche). For Sybr-Green qPCR, each assay was conducted in triplicate for both target region probe-set and control region probe-sets. Relative levels of region dosage (i.e., ratio for assay-test/assay-control) were determined using both comparative CT method and the standard curve methods as described in the company manuals and a fold-change less than 0.7 (deletion) or greater than 1.25 (duplication) was considered to constitute a true event.

UPL probes were selected using the ProbeFinder v2.45 software (Roche, http://www.universalprobelibrary.com). All reactions were performed in triplicate; each 384-well plate included three control samples and three reference genes, as well as a no-template control for each gene. The plate was analyzed with a LightCycler 480 Real-Time PCR system (Roche). Raw data were obtained with the LightCycler 480 software and exported for analysis into the qBase software[20].

Multiplex ligation-dependent probe amplification (MLPA): MLPA probe-sets for CNV validation were purchased from MRC-Holland and used according to the manufacturer's protocols. Electrophoresis of PCR products was performed using an ABI 3730 sequencer (Applied Biosystems). Resulting data were analyzed using GeneMarker 1.70 software (SoftGenetics). After population normalization, the peak height from each sample was compared to a synthetic control, which represents the median of all normal samples in each experiment. Peak heights below 0.75 were considered as deletions and values above 1.3 as duplications.

Long-range PCR: Primers were designed using Primer3 (http://frodo.wi.mit.edu/primer3/input.htm) adjacent to the maximal deleted region, such that PCR products would only be expected in the presence of the deletion. Long-range PCR was carried out using the BIO-X-ACT long DNA polymerase kit (BIOLINE) or SequalPrep™ Long PCR Kit (Invitrogen), using the manufacturers' suggested protocol. PCR products were resolved using agarose gels and visualized with SYBR Safe DNA gel stain (Invitrogen) and UV illumination.

Independent microarray experiments: A CNV was considered confirmed if the CNV detected by the independent array overlapped ≥ 50% of the length of the CNV detected in the discovery phase. The arrays used were as follows:
1. Array experiments from samples used in previously published CNV studies: Affymetrix 500K[21], Illumina 550K[22,23], and Affymetrix SNP 5.0[24].
2. Illumina 1M-single arrays: additional Illumina 1M experiments were used in some instances to further examine the transmission of rare variants in families, or to re-genotype an independently obtained DNA sample from ASD cases or parents.
3. Agilent 1M comparative genomic hybridization (CGH) arrays: We further evaluated the validation rate for CNVs from a total of 174 of the Canadian samples that were consecutively selected from individuals that carried rare CNVs identified in the discovery phase. Two micrograms of DNA from each sample was arrayed onto genome-wide high-density Agilent 1x 1M CGH arrays (aCGH) following the manufacturer's recommended protocol. Experiments were sex-matched and a pool of 50 Caucasian control samples was used as reference. CNVs were detected using the built-in Aberration Detection Method-2 (ADM-2) algorithm DNA Analytics v.4.0.85 (Agilent Technologies) with threshold set to 5 and nested filter set to 2. At least 5 consecutive probes were utilized to call a CNV (same threshold was used on Illumina arrays). Experiments with poor derivative log ratio spread were repeated (DLRS > 0.2). A total of 2,375 CNVs detected by Illumina arrays were found to have sufficient coverage on the Agilent arrays (at least 5 probes coverage) and were used to evaluate the CNV validation rate.

In our initial analysis, 2,028 (85%) of the 2,375 Illumina CNVs were confirmed by the Agilent arrays. Confirmed CNVs met a minimum reciprocal overlap cutoff of ≥50% of the length of each

CNV from both platforms (Illumina and Agilent). The 347 CNVs (15%) that were not confirmed can be explained by a combination of factors. These include ~10% where a clear difference in probe distribution was seen between the platforms, leading to considerable size disparity between CNVs identified on the Illumina and Agilent arrays. After closer visual inspection, size overestimation or underestimation of CNVs was due to uneven probe distributions. Either case may have resulted in longer Illumina CNVs not meeting minimum overlap cutoffs with shorter Agilent CNVs, or vice-versa. In addition, the other ~5% of CNVs not confirmed by Agilent arrays could be due to false negatives from the Agilent arrays or false positives within the Illumina CNVs. The majority of these non-confirmed CNVs had ~10 or fewer probes spanning the corresponding region on the Agilent array, indicating that these could be enriched for false negatives.

In summary, besides computational verification of CNVs using a combination of methods (PennCNV, genotype proportions, child-parent intensity fold-changes and visual inspection), we further experimentally validated at least 40% of all case-CNVs that includes i) using qPCR for 51 de novo CNVs and 271 inherited CNVs (ie. all cases and their two parents were tested with at least 2 assays per CNV region), ii) ~200 Agilent one million feature aCGH microarrays (for 174 cases and 30 parents), and iii) ~80 additional Illumina 1M arrays to further examine family segregation in sibs, or other methods (such as FISH experiments and breakpoint sequencing). Information on the confirmed CNVs is listed in Supplementary Tables 7-8. Supplementary Table 7 lists rare CNVs confirmed experimentally (de novo and inherited) and their segregation in sibs, as well as phenotypes, where available. Supplementary Table 8 lists all rare CNVs detected in 996 ASD cases, as well as all types of verification and validation evidence accumulated for these CNVs.

E. RARE CNV BURDEN ANALYSIS

Stringent CNVs that passed all QC filters (≥5 probes, ≥30 kb size) were considered rare if they were found at a frequency ≤1% of the total sample set (2,283 subjects) and did not overlap, with >50% of its length, a CNV found at a frequency >1%. This resulted in a total of 5,478 rare stringent CNVs.
Supplementary Table 4A gives an overview of the characteristics of rare CNVs in ASD cases and controls. The average CNV size in ASD cases was 182.7 kb, the median number of CNVs per individual was 2, and 51.6% were deletions. We found 1.8% of ASD CNVs to be >1 Mb, and 43.7% were between 100 and 999 kb. We observed that the distributions of specific rare CNVs are similar in cases and controls for measures such as CNV size, number and proportion of duplications vs. controls.

We further examined parent-child transmission. Inheritance status was estimated for CNVs detected in 876 probands from complete trios where array data after QC was available from both parents. Using computational data only, the rate of *de novo* CNV events was initially estimated to be 6.9% (165/2,382) with 13.6% (119/876) of trio families having at least one putative *de novo* CNV. When additional computational and laboratory validation data (see section D. CNV verification) was added, we confirmed that at least 5.6% (49/876) of trio families carried at least one *de novo* CNV (average of 1.1 verified *de novo* CNVs/sample) (Supplementary Table 4A). By taking the family type into account, Supplementary Tables 4B and 4C list the summary characteristics of *de novo* rare CNVs and inherited CNVs, respectively, in multiplex (MPX) and simplex (SPX) families. No significant differences were found.

CNV global burden analysis
Global burden analyses for rare CNVs were performed using PLINK v1.07[25] and scripts developed in-house. A total of 5,478 rare stringent CNVs in a total sample set of 2,283 was used in the analyses. We tested for global increased burden in 996 cases compared to 1,287 controls for three measures: CNV rate, CNV size (Supplementary Table 5) and the average number of genes affected by CNVs (gene-count) (Table 1, main text).

CNV rate was evaluated in two ways, by the number of CNVs per sample and the proportion of samples with one or more CNVs. The CNV size was assessed as both the total genomic segment covered by CNVs, as well as the average CNV size (Supplementary Table 5).
Gene-count evaluated the average number of genes intersected by CNVs per case compared to that for control sample (Table 1, main text). Case-control contrasts were assessed using an adaptive permutation procedure for statistical significance of one-sided tests (i.e., hypothesizing that cases will show greater burden of rare CNVs than controls). For each of 100,000 permutations samples were randomly reassigned either case or control status. Genic regions were identified based on RefSeq annotations (UCSC, v. April 2009, NCBI v36, hg18) and defined by the outermost boundaries of the full set of transcript isoforms. Gene boundaries were extended with a 10 kb flanking region on either side, or to half the distance to the next neighboring gene when the 10 kb flanking regions of neighboring genes overlapped. To account for a potential bias in the global CNV rate and size between cases and controls, we also repeated the same test while controlling for these parameters using logistic regression (PLINK[25] and R stats). In a similar manner we also analyzed subsets of rare CNVs, deletions-only and duplications-only, separately.

We further explored subsets of rare CNVs in two additional frequency ranges: 2-6 occurrences and single-occurrences. Single occurrences were defined as CNVs that did not overlap any other CNVs in the dataset for more than 50% of their length. We also assessed duplications and deletions independently in the same frequency ranges. For these specific sets, deletions were considered as single occurrences even if they overlapped a duplication for more than 50%, and vice versa. Therefore, the number of single-occurrence deletions and duplications do not sum to the number of single-occurrence CNVs in the frequency filtered sets that consider both CNV types (see Table 1, main text and Supplementary Table 5). The set with 2-6 occurrences was obtained by selecting all rare CNVs that overlapped 6 or fewer other CNVs by at least 50% of their length, and subsequently removing CNVs that met the definition of single-occurrence. In Table 1 of the main paper, in total we found 1,419 single-occurrence genic CNVs (880 deletions and 896 duplications) and 1,831 genic CNVs (1,094 deletions and 977 duplications) in the 2-6 frequency range.

CNV Region (CNVR) burden analysis
We also tested for increased global burden of rare CNV regions (CNVRs) in cases compared to controls. Here, 2,181 non-redundant CNVRs were constructed by merging overlapping rare CNVs present in the total sample of cases and controls (n=2,283 samples) and taking the outermost boundaries of the union of those CNVs. Statistical significance for each gene or CNVR was assessed by Fisher's exact test. From these analyses, CNVR at *DDX53/PTCHD1* emerged as a significant ASD risk factor ($P = 3.1 \times 10^{-3}$ for the initial 1,287 EA controls). Specifically, we observed 7 ASD male cases with overlapping deletions at *DDX53/PTCHD1* (Xp22.1) and no CNVs were observed at this locus for the initial 1,287 controls (Supplementary Figure 4). We further inspected an additional set of 3,677 European controls amassed from three independent cohorts and none of them showed CNVs at this locus ($P = 3.57 \times 10^{-6}$ for the 4,964 combined controls (Supplementary Table 6). This result would be significant after Bonferroni correction for the total number of 2,181 CNVRs.

CNV-based gene association test
In Supplementary Table 6, we list examples of ASD candidate genes or loci identified by de novo and rare-inherited CNVs. Examples of novel ASD loci include *SHANK2*, *SYNGAP1*, and *DLGAP2* based on the observation that *de novo* CNV affects these genes in cases but not controls. Also, a combination of rare *de novo* and inherited CNVs affecting *NRXN1*, *IL1RAPL1, DMD,* and the DiGeorge 22q11.2 region in ASD cases replicate previous findings for these genes implicated in ASD[1], X-linked nonsyndromic ID and ASD[26], Duchenne and Becker muscular dystrophy associated with ASD/ID[27], and an established genomic disorder, respectively. We tested for specific hypothesis to identify regions associated with ASD. Significance was assessed using Fisher's exact test. Genic regions including 10 kb flanking regions were defined as previously described.

Burden analysis for genes known to be implicated in ASD and/ or ID
Through extensive review of the literature[28] and further scrutinizing of all available databases (up to December 2009), we compiled lists of genes previously described as being implicated in ASD and ID ('expert-curated' lists), and a third list of candidates for ASD (Supplementary Table 9):
- ASD-implicated: 36 genes and 10 loci strongly implicated in ASD and identified in subjects with ASD or ASD and ID;
- Intellectual disability (ID): 110 genes and 17 loci know to be implicated in ID but not yet in ASD;
- ASD candidates: 103 genes drawn from previous studies of common and rare variants for ASD. They include case reports of cytogenetic abnormalities, allelic association and CNV studies.

To select for CNVs with maximal impact, CNVs needed to intersect genes, or overlap loci (such as genomic disorder loci 1q21.1, 22q11.2) by an arbitrary cutoff of 50% or more of their total length. Linear logistic regression was used to specifically test for an increased CNV rate in cases vs. controls for the three different gene lists or combinations thereof. In this analysis, phenotype was regressed on the number of genes intersected by one or more CNVs, and significance was assessed by permutation with 100,000 iterations, where the phenotype assignment (cases/control status) was randomized. Beta values were exponentiated and odds ratios and corresponding 95% confidence intervals were obtained using the R statistics package 2.10.0 (http://www.r-project.org/).

We also repeated the analysis after manual inspection (i.e., not using an arbitrary cutoff) and eliminated CNVs less likely to be pathogenic, such as events that fall outside the critical region of overlap with genomic disorders loci, intronic CNVs in *NRXN1*, heterozygous CNVs disrupting autosomal recessive loci, X-linked genes in females inherited from non-ASD fathers, and duplications inherited from non-ASD parents. A list of clinically relevant CNVs is provided as Supplementary Table 10.

Population attributable risk
To determine the fraction (%) of ASD cases that can be attributed to rare CNVs affecting the combined list of ASD and ID genes/loci we calculated the population attributable risk (PAR) as follows:

PAR = P(A)*(R-1) / [1 + P(A)*(R-1)], in which:
P(A) = exposure to selected rare CNVs, which was estimated from the frequency in controls, and
R = the relative risk for ASD, which was estimated from the odds ratio.
Because ASD is relatively rare, the odds ratio is a good substitute for the relative risk and the observed frequency in controls is a reasonable estimate for P(A). Results are listed in Supplementary Table 11.

F. GENE-SET ENRICHMENT AND FUNCTIONAL MAP

Analytical synopsis

Gene-set enrichment analysis has been proposed as a means to identify biologically meaningful groups of genes having an impact on risk for disease and other phenotypes. Our analyses build on the work of Subramanian *et al.*[29], who develop a heuristic Benjamini and Hochberg's (1995)[30] false discovery rate (FDR) procedure for enrichment analysis of gene expression data. For a given threshold t, the FDR is defined as the expected fraction of false positive P-values $\leq t$ divided by the fraction of the total P-values $\leq t$. Gene sets are not necessarily independent because genes fall in more than one gene set. This dependency makes it challenging to estimate the numerator of the FDR. Subramanian *et al.*[29] assume that this quantity can be estimated by permutation of case-control status and re-analysis of the data under the null hypothesis. If this procedure is performed sufficient times (e.g., 1000 in their case) the resulting sets of P-values is presumed to be a reasonable approximation to the null distribution of the P-values, which can then be used to estimate the numerator for a given t. The denominator is the fraction of P-values $\leq t$ in the real data. Another related quantity is used in our analyses, the FDR q-value, which for an individual hypothesis test is the minimum FDR at which the test would be called significant[31,32]. For brevity we shall refer to the FDR q-value as the q-value.

Because we analyze data on rare events, specifically rare deletion CNVs overlapping genes, we use the Fisher's exact test (FET) to assess which gene-sets are more frequently affected by CNV events in cases compared to controls. Contrasting case and control CNV distributions focuses attention on gene sets most likely to be relevant to risk for autism. Following Subramanian *et al.*[29] to generate a null distribution of P-values, we permute case-control status sufficiently so that CNV events and case-control status are independent and then analyze the distribution of CNVs in cases and controls, by gene set, using FET. We performed this procedure 2000 times to build a distribution of P-values under the null hypothesis Permuting case-control labels has the benefit of maintaining the correlation of genes within sets as well as the correlation of CNV prevalence on genes. When the case/control status permutation is used to construct the null hypothesis distribution, the enrichment test belongs to the category of *self-contained* tests. In contrast, the usual application of FET to enrichment analysis, where an experimental gene-set is tested for overlap with functional gene-sets, as well as other enrichment methods relying on gene label permutations, belong to the *competitive* category of tests. More details on this difference can be found in Dinu *et al.*[33].

Deriving gene sets

Gene-sets were derived from the following resources (Supplementary Table 12): Gene Ontology (Biological Process, Molecular Function, and Cellular Component), NCI Pathways, KEGG Pathways, Reactome, PFAM protein domain families. Gene Ontology, KEGG and PFAM were downloaded from Bioconductor/R packages (org.Hs.eg.db, KEGG.db, GO.db); NCI pathways were directly downloaded from the NCI Pathway Interaction Database web site (http://pid.nci.nih.gov/); gene IDs were converted from Uniprot to EntrezGene using Bioconductor/R annotation packages (org.Hs.eg.db). Top-level Reactome pathways were downloaded from the Reactome web site (http://www.reactome.org/). All downloads were performed on 27th November 2009. The version of all Bioconductor packages is 2.2.11.

Terms annotating more than 700 or less than 5 genes were discarded (Supplementary Table 12). Large gene-sets often represent broad categories without much useful biological meaning (e.g. *regulation of physiological process*, *zinc ion binding*). Small groups of genes, on the other hand, are not likely to produce statistically meaningful results.

Gene-set enrichment test

The overall analytic plan is described above. We require a score to identify which genes and thus which gene sets are hit by a CNV in a particular individual. Let *i* index samples and *j* index gene sets. Define indicator I(*i*,*j*) = 1 if individual *i* carries a CNV overlapping at least one gene in set *j*, and zero otherwise. Note a sample can therefore contribute to multiple gene-sets but cannot contribute more than once to the same gene-set. The total score for gene set *j* is obtained by summing over *i*. For testing, this sum is partitioned between cases and controls.

After obtaining the null distribution of *P*-values as described in §1, we calculate the q-value for each gene set. For subsequent analysis and illustration, we use an arbitrary threshold requiring the q-value ≤ 12.5% (Supplementary Fig. 5), recognizing that about 1 in 10 gene sets will be false positives. According to this threshold, 76 gene-sets affected by deletions (2.18% of the total tested sets) were found to be enriched in cases compared to controls and were used to construct a functional enrichment map (Fig. 3a, main paper and Supplementary Fig. 6-7).

Cryptic bias

It is possible that some unknown bias differentiates cases and controls and confounds our gene enrichment analysis. If there were such a bias, however, we would expect to see some difference in the size or number of deletion CNVs in cases versus controls. Because we see no noteworthy or significant difference in summary statistics for these attributes (see main text, Supplementary Fig. 8, Supplementary Table 4A and 5), we believe the possibility of bias is remote. We also evaluated whether there was a relationship between enrichment significance and genome proximity. To do so, we broke the q-value scale into arbitrary bins and grouped gene sets into these bins. For every gene-set, we calculated two genomic proximity indexes: (a) the fraction of gene pairs located on the same chromosome, (b) the fraction of gene pairs having mutual distance less than 1 Mb. When the distribution of these indexes across gene-sets was plotted by bin (Supplementary Fig. 9), their median values were essentially the same across bins. In fact bins with low q-values (i.e. significant enrichment) tended to have fewer gene-sets with high proximity scores.

Network visualization of gene-sets enriched for deletions: functional enrichment map

Enriched gene-sets are graphically organized into a network, where each set is a node and edges represent gene overlap between sets; gene sets map to specific biological processes/pathways involved in autism susceptibility. The Cytoscape network software v.2.6.3[34] and the plugin "Enrichment Map"[35,36] were used to build the network. Plugin and source code are available at http://baderlab.org/Software/EnrichmentMap. Node color encodes the enrichment q-value (Fig. 3a, main text and Supplementary Fig. 6-7, 10 and 12). Node size is proportional to the total number of genes belonging to the corresponding gene-set. Edge thickness is proportional to the overlap score, as defined below.

Functional gene-set clusters, identified by shaded ovals, represent groups of gene-sets with highly overlapping support genes (Supplementary Table 13). Gene-set clusters were manually identified and annotated. The size of the network justifies the sole use of the layout to identify gene-set clusters[37]. Satisfactory algorithmic solutions are currently not available to generate cluster summaries. Modified enrichment methods have been proposed to prioritize specific gene-sets among many redundant ones[38,39]; however, these tend to either favor smaller or bigger gene-sets, whereas neither option is ideal[36]. For these reasons we preferred to annotate clusters through an expert-based curation process.

Gene-set overlap score: Define "support genes" as those genes more frequently overlapped by deletions in cases than in controls and let "gene-set overlap" be proportional to the number of "support genes" that two gene-sets share between them, normalized by their size. In previous work we investigated two measures of gene overlap, Jaccard Coefficient JC = $\frac{|A \cap B|}{|A \cup B|}$ and the Overlap Coefficient *OC* = $\frac{|A \cap B|}{\min(|A|,|B|)}$ in which *A* and *B* are two gene-sets connected by an edge. Both have weakness when there are imbalances of counts of events in the sets. We find that the arithmetic average of these two behaves better, and that is what we use here and refer to as the weighted overlap.

Expanded functional map between deletion-enriched gene sets and known ASD/ID genes
An expanded enrichment map was constructed to evaluate how genes affected by deletions discovered in our study functionally relate to known ASD/ID genes (ie. listed in Supplementary Table 13) (Fig. 3b, main text).

Visualization strategy
The map is composed of three node types:
a) *functional* gene-sets enriched in deletions, according to the self-contained enrichment test;
b) disease gene-sets, comprising genes known to be implicated in ASD and/or ID;
c) *functional* gene-sets enriched in disease genes, according to the competitive enrichment test with Fisher's exact test, but not in deletions.

Type (a) and (c) belong to the same collection of functional gene-sets, derived from Gene Ontology, pathway and protein domain annotations. Type (b) are "expert-curated" ASD/ID genes based on published findings, as described elsewhere in the paper. Type (a) nodes have round shape and are colored according to a red-to-white gradient correlated to enrichment FDR (red corresponds to FDR ~ 0%). Type (b) and (c) nodes have different shapes depending on disease (ASD: parallelogram, intellectual disability: triangle, both: octagon); type (b) are colored in green and type (c) in yellow. All node sizes are proportional to the number of member genes.

Different node types are connected by different edge types:
1) *type (a) gene-sets are inter-connected* using the weighted overlap coefficient between their support genes. The specific choice of support genes, rather than all member genes, reflects more accurately CNV incidence on genes (**green color**);
2) *type (c) are inter-connected, and also connected to type (a)* using the weighted overlap coefficient between all their member genes. Type (a) are not limited to their support genes, to take full advantage of the generalization from genes to functions operated by the enrichment test (**orange color**);
3) *type (b) gene-sets are connected to enriched gene-sets of type (a) and (c).* The presence of a connection corresponds to the presence of significant enrichment (**blue color**, Fisher's exact test *P*-value < 0.001).
Note the different overlap index used for a-a, a-c, c-c (above in 1-2), on one hand, and b-a, b-c (above in 3) on the other hand. The rationale for this solution is that functional sets (a, c) belong to highly inter-related resources, where overlaps tend to be large; using an index evaluating overlap significance, as done for (3), would result in excessive connectivity, and would also be systematically larger for pairs of gene-sets with more member genes. On the other hand, the overlaps from disease genes to functional sets are expected to be comparatively sparse also because there are no ontological relations pre-determining overlaps, hence the choice of the FET test to evaluate overlap significance appears more natural.

Edge thickness is proportional to the value of the overlap coefficient for type (1) and (2) or to –log(*P*-value) for type (3).

Gene coverage analysis

Deletion gene-set enrichment identifies functions that are significantly affected in autism. The enrichment map (Fig. 3b, main text) displays a detailed account of the relationships existing between gene-sets enriched in known disease genes and in deletions. Here we evaluate if the major functional groups enriched for deletions (Fig. 3a, main text) significantly overlap with known disease genes (Fig. 3b, main text). Since only a minority of genes within enriched sets is affected by deletions, the enrichment analysis not only provides a functional summary, but also extends the number of potential candidate genes for contribution to pathogenesis.

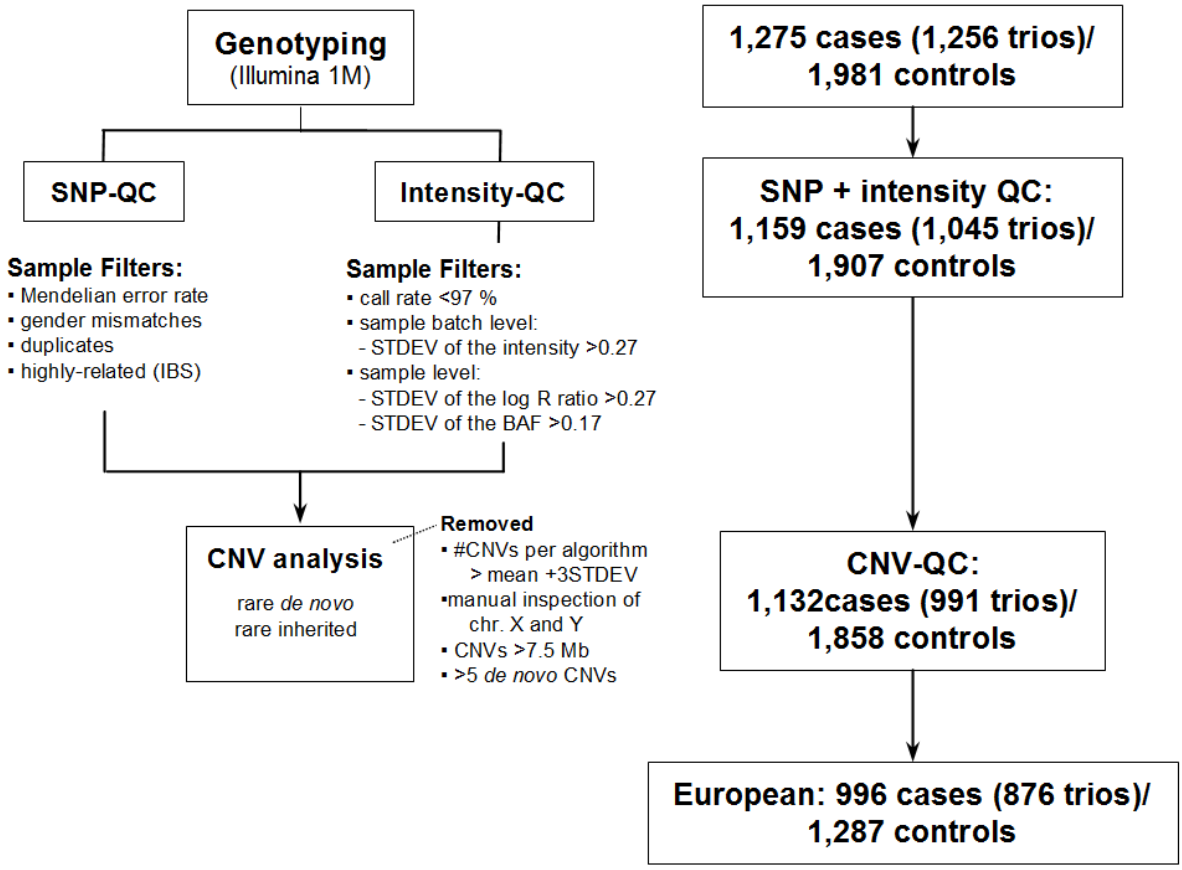
We evaluated how many genes from the ASD-implicated and Intellectual disability sets are also present in sets enriched for deletions. Considering the 17,660 genes in 6,129 gene-sets that passed the size filter, only a small fraction of genes is not present in the analyzed gene-sets (8.6% and 2.7% respectively), arguing for good background coverage. Both disease gene-sets have higher than expected overlap with gene-sets enriched in deletions, considering that the latter contain 30.3% (5,353/17,660) of the genes covered by all filtered gene-sets (Supplementary Fig. 11A).

In addition, we calculated the ratio between the observed and the expected number of overlapping genes between disease gene-sets and the major functional clusters of gene-sets identified as enriched in deletions: (1) all gene-sets enriched in deletions, (2) cell projection, motility and cell proliferation, (3) GTPase and Ras signaling (e.g., ref [40]) and 4) kinase activity/regulation. For comparison, we also included 5) a collection of less significant and less connected gene-sets (Supplementary Fig. 11B). Group (2) and (3) are particularly interesting because they have several gene-sets with top enrichment scores (i.e. small q-value) and they are highly connected to disease gene-sets, either directly, or through other gene-sets. Both disease gene-sets display enrichment in genes from deletion enriched gene-sets, with higher enrichment for groups (2) and (3), and little or no enrichment for group (5). ASD-implicated genes have higher overlap rates than Intellectual Disability genes.

G. SUPPLEMENTARY FIGURES AND TABLES

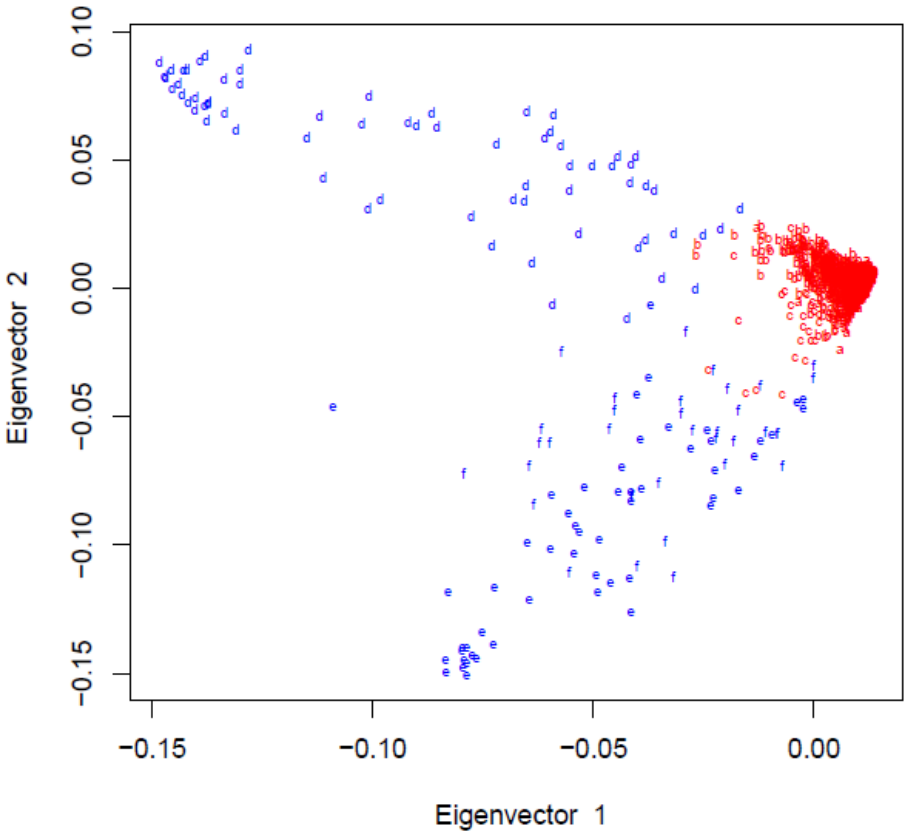
Supplementary Figure 1. Quality control and analysis flow chart

Flow chart (left) shows the quality control (QC) filters performed following genotyping (SNP and intensity QC), as well as the criteria used to remove CNVs following detection. Numbers of cases (trios) and controls (right) are shown corresponding to the number remaining after each QC or removal step. The final numbers correspond to the number of cases (trios) and controls of European ancestry that were included in the rare CNV discovery. *CNV analysis:* First, we used two CNV prediction algorithms (QuantiSNP[14] and iPattern[8]) to examine the ASD family and control samples passing QC-filtering. We established a stringent dataset of non-redundant CNVs (having ≥5 consecutive probes spanning 5 kb) called by both algorithms in an individual (Fig. 1). Stringent CNVs were excluded if they were pericentromeric, had >70% GC-content or overlapped segmental duplications over ≥50% of their length. We also excluded 19 samples (six cases and 13 parents) with CNVs >7.5 Mb, which likely represent karyotypic abnormalities or cell-line artifacts[16] (Supplementary Table 2). A series of validation experiments using quantitative-PCR (qPCR) and independent microarrays on subsets of samples showed validation rates of >90% for stringent CNV calls and >95% for stringent CNVs ≥30 kb in size (see sections *Pilot experiment to evaluate the quality of detected stringent CNVs* and *CNV verification)*.

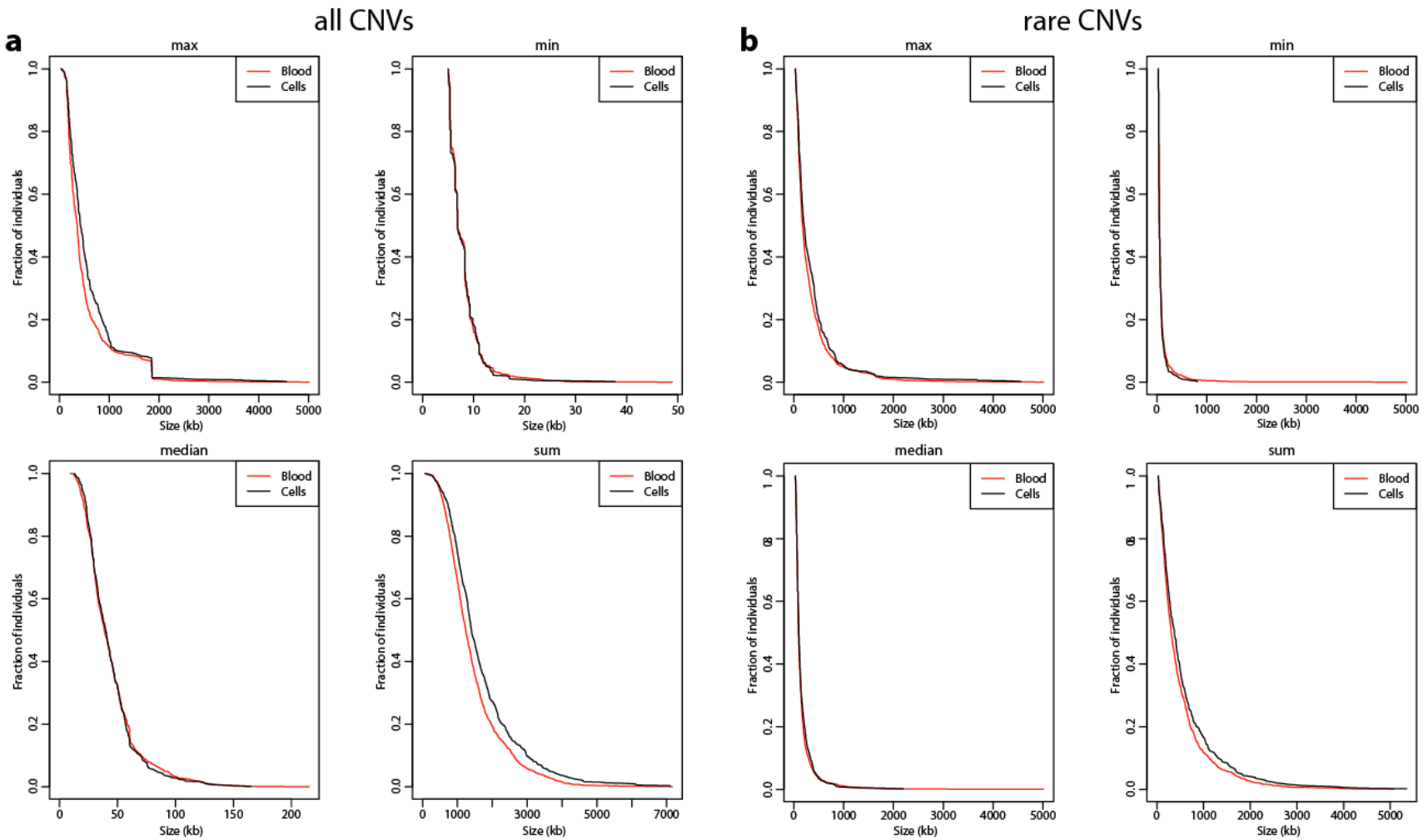


Supplementary Figure 2. Results from ancestry analysis using genome-wide genotype data

Presented are the first and second eigenvectors from the spectral decomposition of the genotype data. Red maps to Europeans versus non-European (blue) ancestry. Letters correspond (roughly) to Northwestern Europe (a), Southern Europe (b), other European ancestry (c), African/African-American (d), Asian (e) and Latino (f).



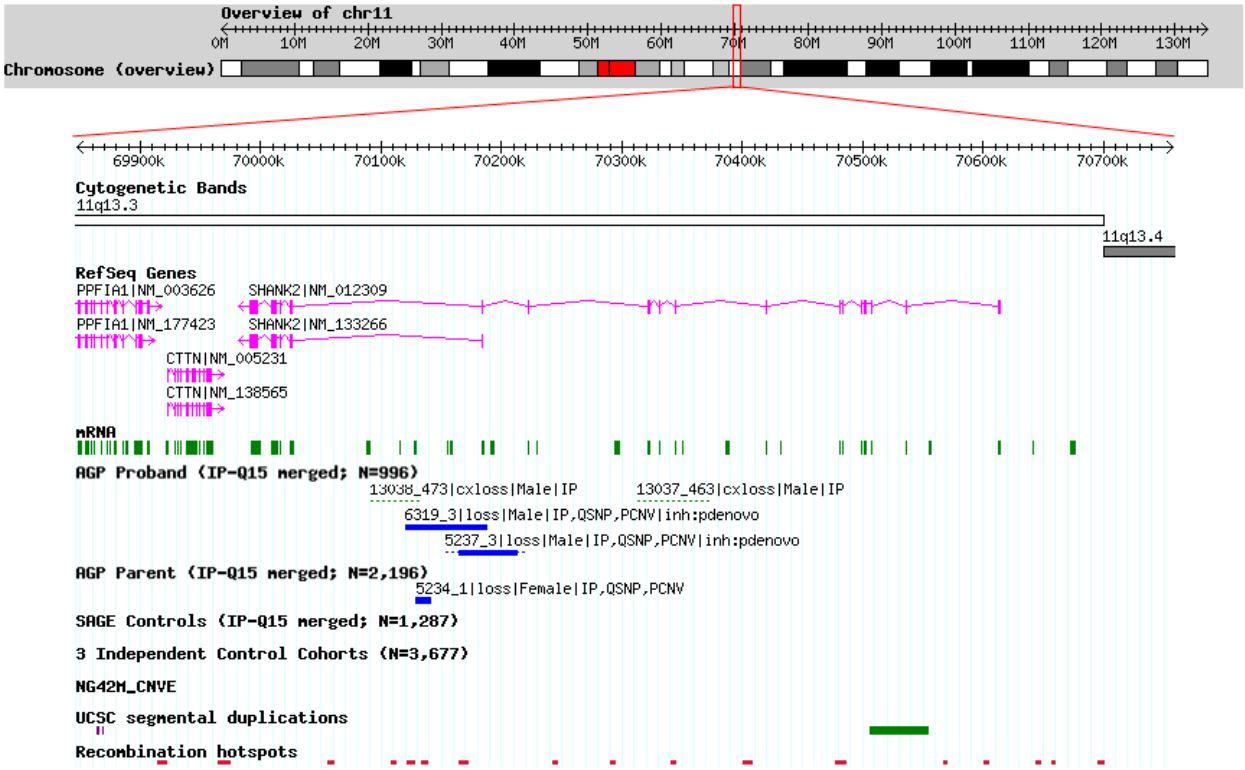
Supplementary Figure 3. Global CNV measures for blood versus cell-line DNA-derived samples



a. All CNVs, European-only samples; **b.** Rare CNVs, European-only samples. Cumulative distribution of the following global measures for blood (red color) and cell-lines (back color): largest (max) or smallest CNV (min.) per sample, median length and sum of all CNVs per sample. These plots indicate what proportion of samples have a median/sum/min/max CNV size greater or equal the size shown on the X-axis.

Supplementary Figure 4. Examples of CNVs overlapping ASD candidate genes or loci
Top level: Genome browser image of the cytogenetic band, genomic coordinates, genes and mRNA. **Middle level**: CNVs within the IP-Q15 merged dataset (ie. identified by two algorithms) depicted as blue (deletion) or red (duplication) bars or dotted lines (when identified by only a single algorithm; either QuantiSNP: QSNP or iPattern: IP) in AGP probands and parents. Common CNVs by Conrad et al (ie. detected using a Nimblegen 42M array set)[41], are also shown within the NG42M_CNVE track. As displayed in the context of information in the Database of Genomic Variants[42] these regions also contain a scarcity of segmental duplications, specifically with only a maximum of one element found at, or adjacent to, the CNV boundaries. Recombination hotspots are also highlighted in the last track.

A. CNVs identified in the *SHANK2* region of chromosome 11q13.3
Genome browser image of the region containing *SHANK2*. Two *de novo* deletion were identified in male probands (66 kb loss in ASD case 5237_3 and 68 kb loss in ASD case 6319_3). No CNVs were identified in four control datasets (SAGE controls and three independent control cohorts) containing 4,964 samples.

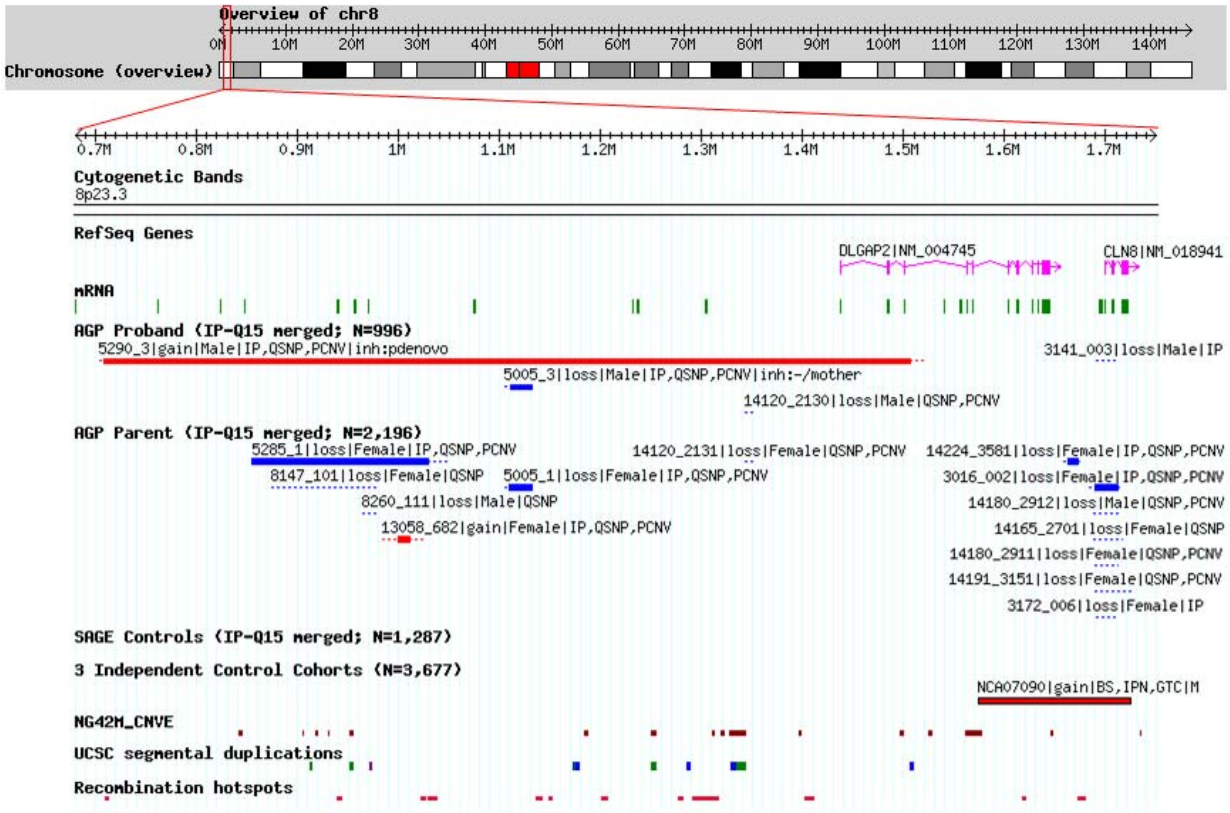


B. CNVs identified in the *SYNGAP1* region of chromosome 6p21.32

Genome browser image of the region containing *SYNGAP1*. A single 112 kb *de novo* deletion was identified in female ASD case 5353_3. No CNVs were identified in four control datasets (SAGE controls and three independent control cohorts) containing 4,964 samples.

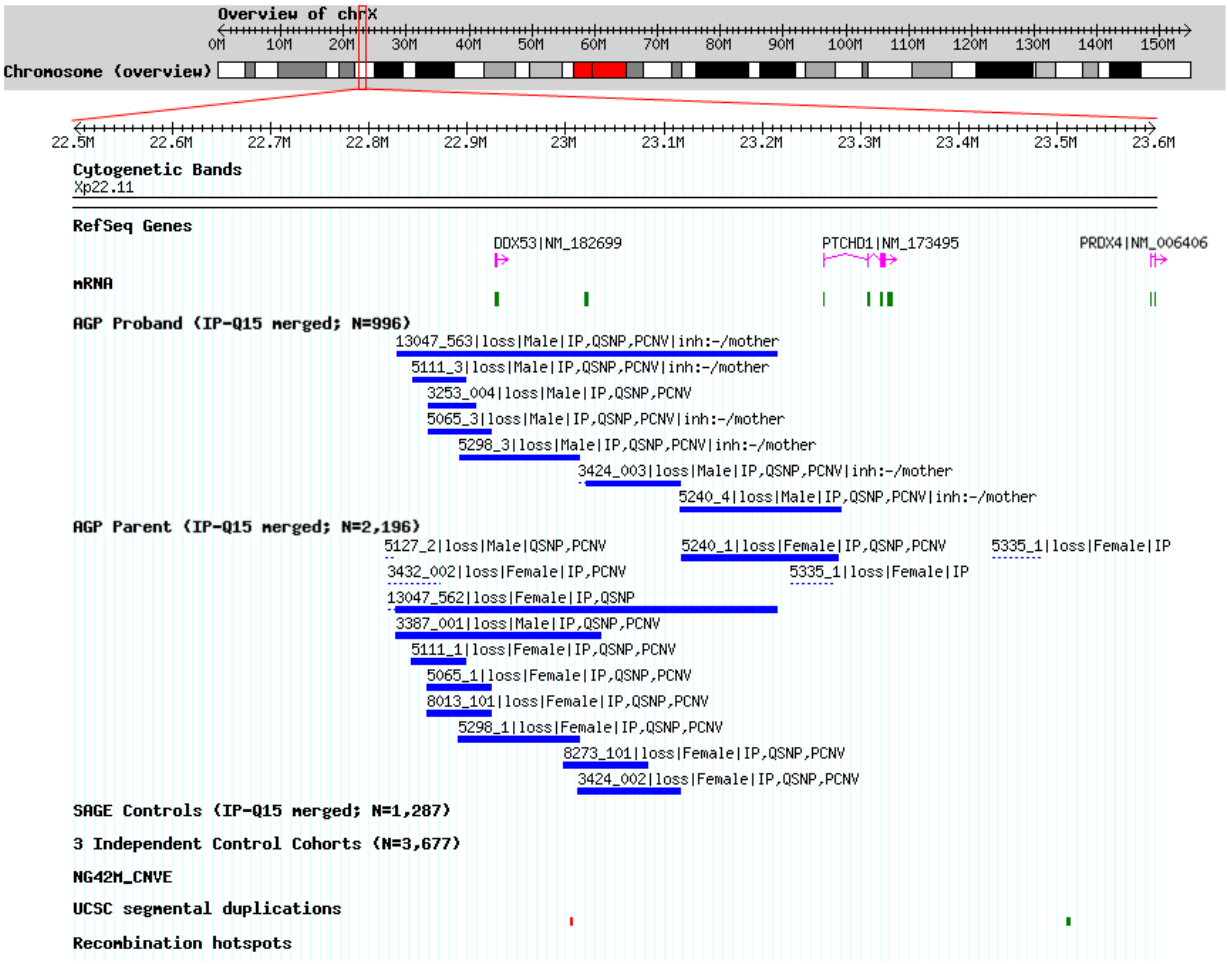
C. CNVs identified in the *DLGAP2* region of chromosome 8p23.3

Genome browser image of the region containing *DLGAP2*. A large 817 kb *de novo* duplication was identified in a male ASD case 5290_3 and was found to overlap the 5' end of *DLGAP2*. A single 151 kb gain overlapping *DLGAP2* was seen in a male from the Ontario Health Institute (OHI) controls[6] (total of 1 CNV identified in 4,964 controls from four control datasets).

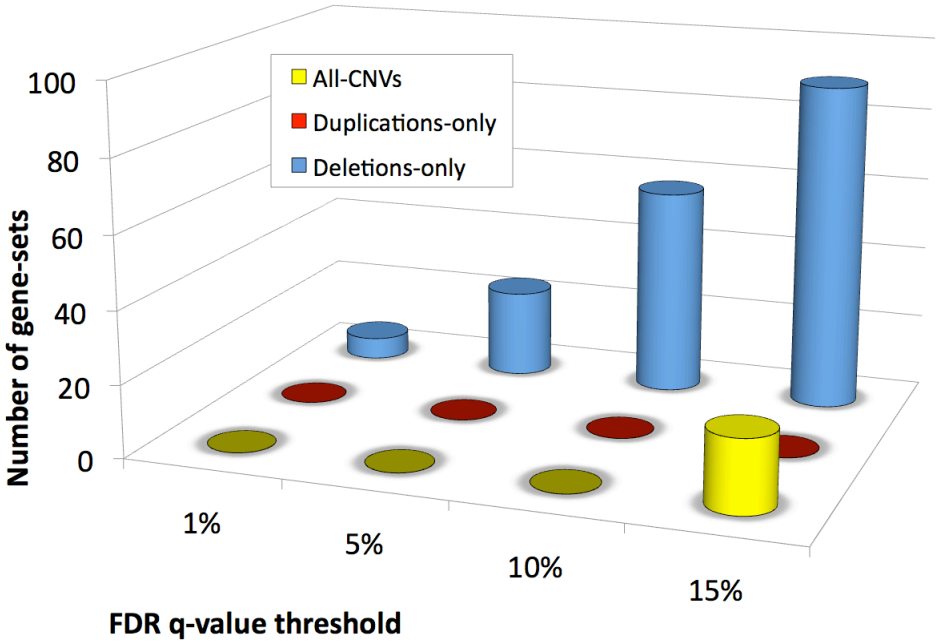


D. CNVs identified in the *DDX53/PTCHD1* region of chromosome Xp22.11

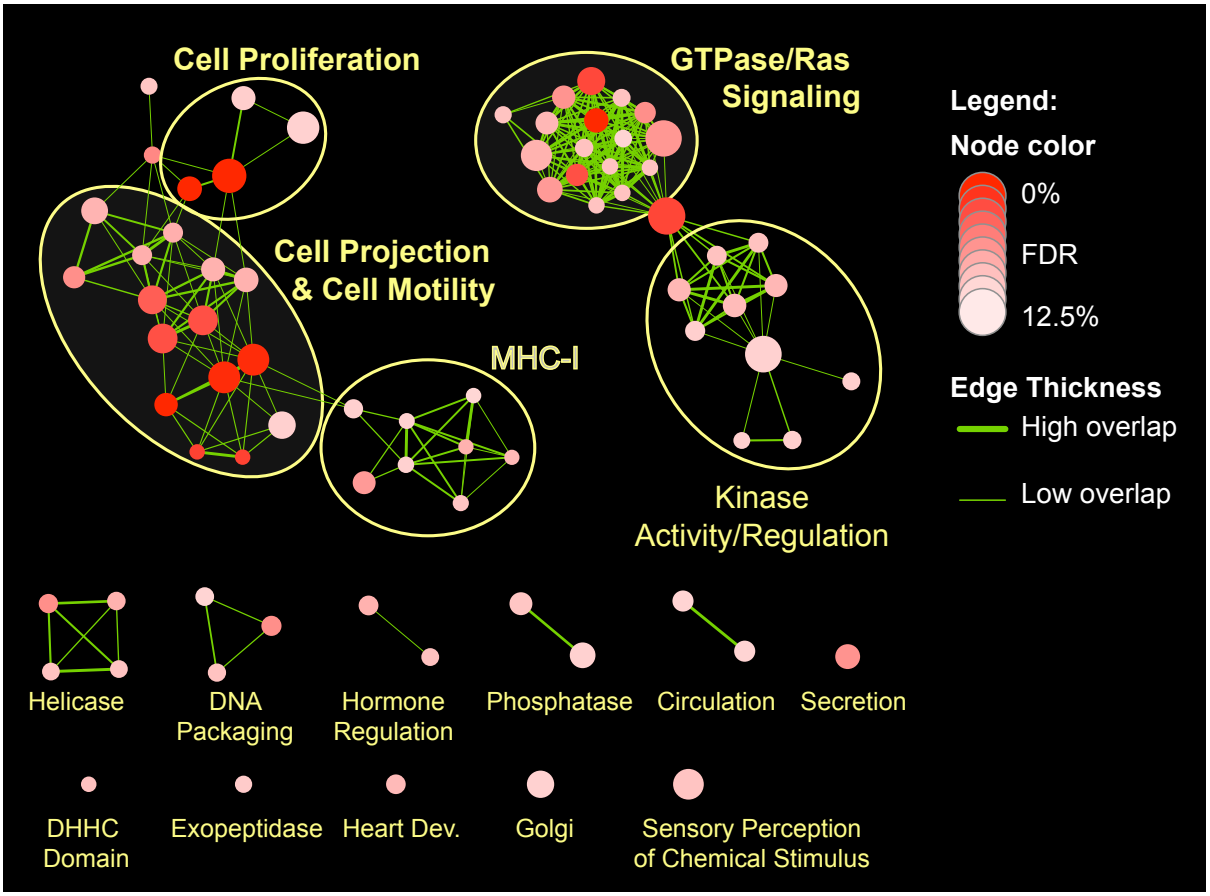
Genome browser image of the region containing *DDX53* and *PTCHD1*. ASD case 5240_4, which deletes exon 1 of *PTCHD1* was described previously in Marshall *et al.* 2008[21]. All ASD probands are males inheriting the CNV from mothers. No CNVs were identified in four control datasets (SAGE controls and three independent control cohorts) containing 4,964 samples.



Supplementary Figure 5. Number of enriched gene-sets at different FDR q-value thresholds

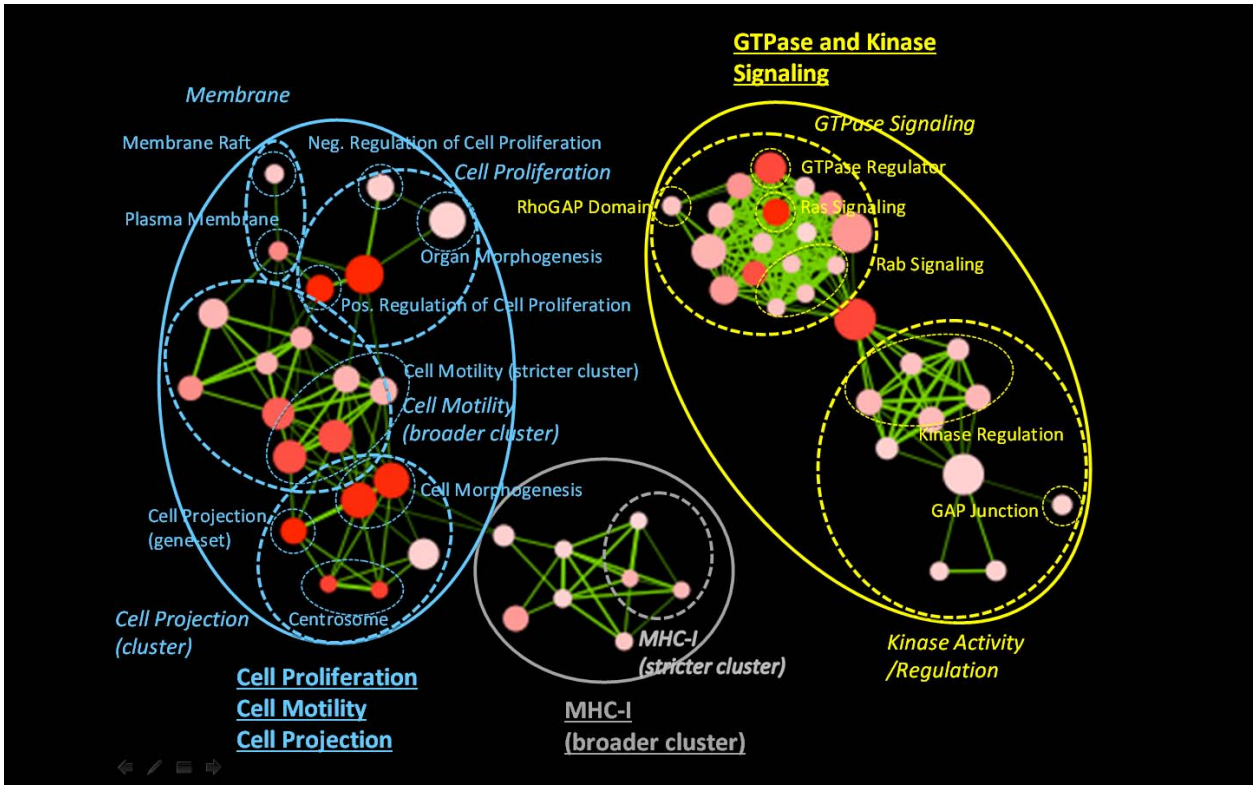


Supplementary Figure 6. Overview of all enriched gene-sets in deletions



Functional enrichment map, complete view. Major gene-set clusters (57 out 76 enriched gene-sets for deletions) are circled; they are also displayed in Fig. 3a of the main paper. The remainder 19 gene-sets form smaller clusters (i.e. composed of 2-4 gene-sets) or are not connected to any other enriched gene-set. Because they are poorly connected or 'disconnected', and their q-values are in general higher than 5% (except for protein-DNA complex assembly, helicase activity, and secretion), we chose to not display them in Fig. 3a, main text. We note though that the absence of connections to other gene-sets could be due to the paucity of gene-sets for certain biological functions, resulting in only less specific yet available gene-sets being enriched. Enrichment was also found in the MHC-I related gene-sets, however these were omitted from Fig. 3 of the main text as the enrichment was driven almost solely by the HLA-B locus. Most of such gene-sets were strictly related to MHC-I and antigen processing, with only one including a broader collection of adhesion proteins.

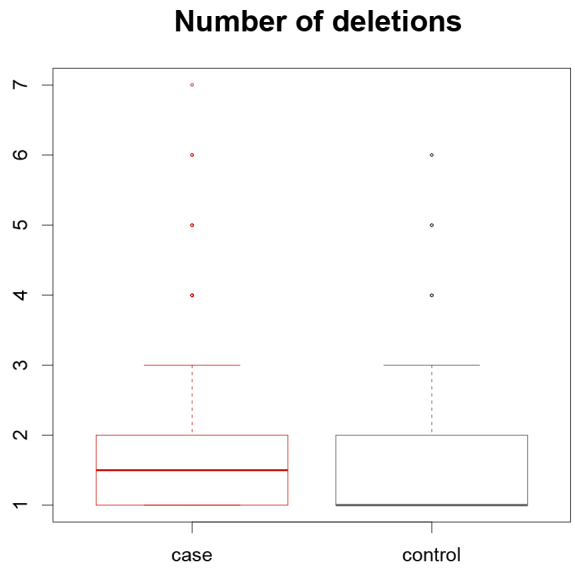
Supplementary Figure 7. Detailed annotation of the main enriched clusters



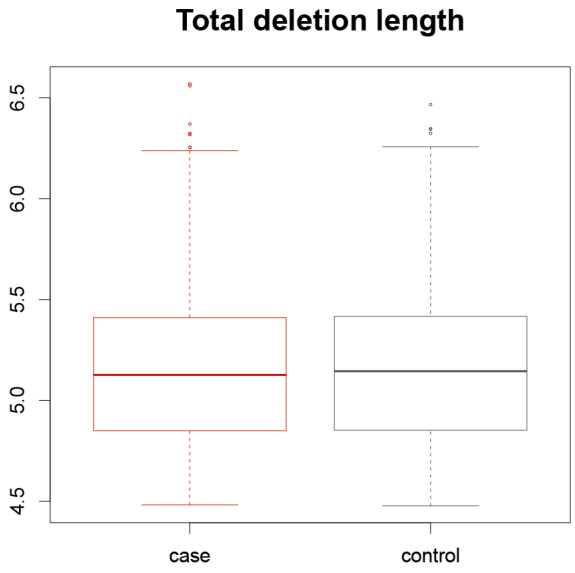
Disconnected and minor gene-set clusters are not shown. See Supplementary Fig. 6 legend for more details.

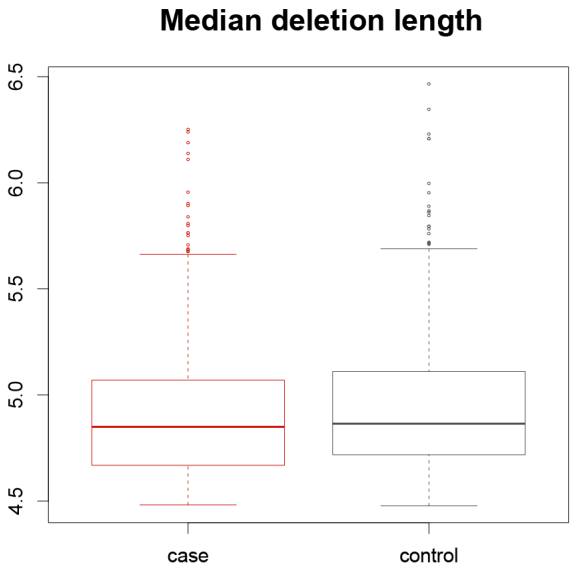
Supplementary Figure 8. Control for bias in length and number of deletions

A. Global measures



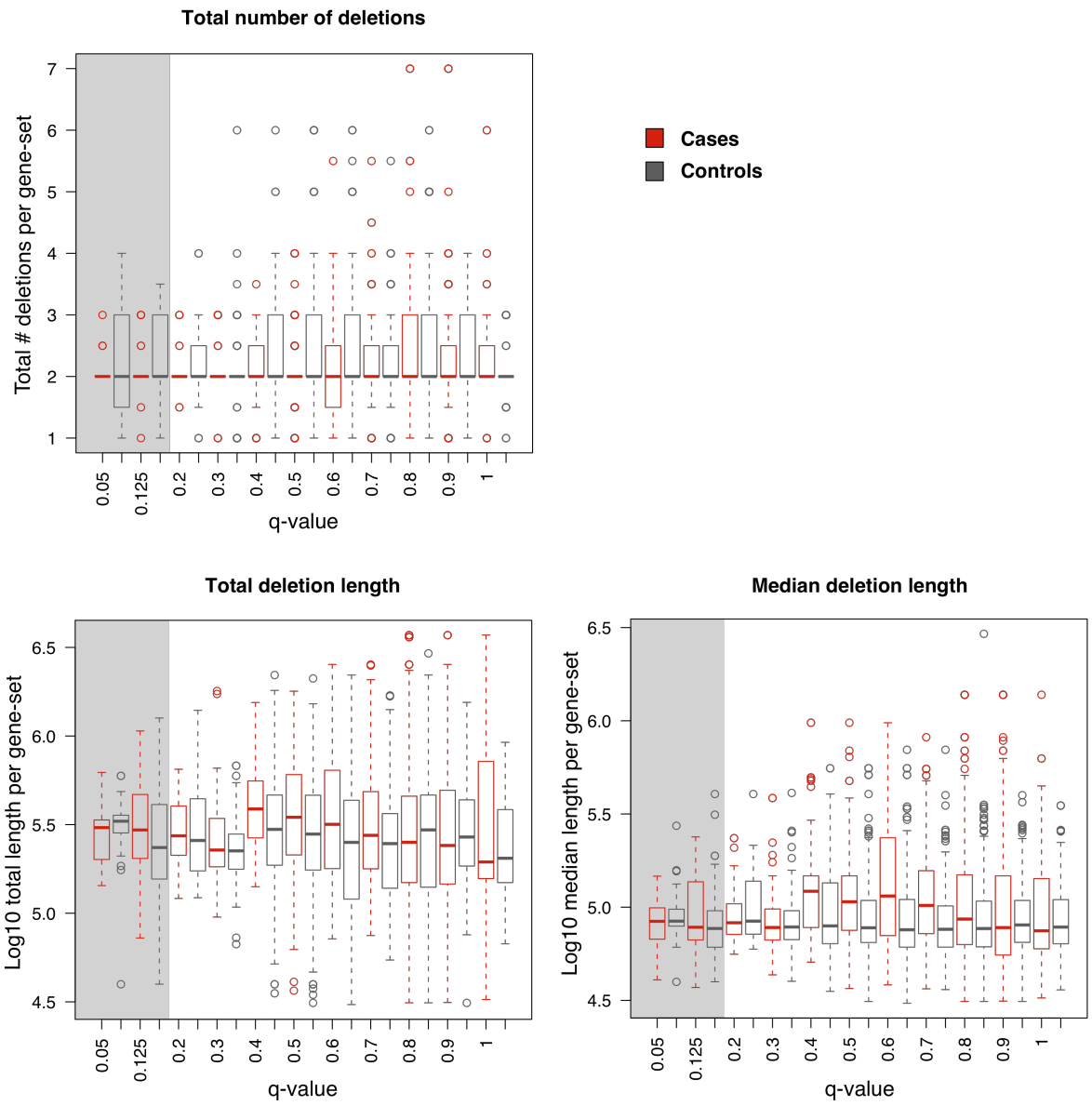
Wilcoxon Test p-value	greater	less
Total CNV length	0.7934	0.2067
Median CNV length	0.9734	0.0266
Number of CNV	0.2185	0.7816



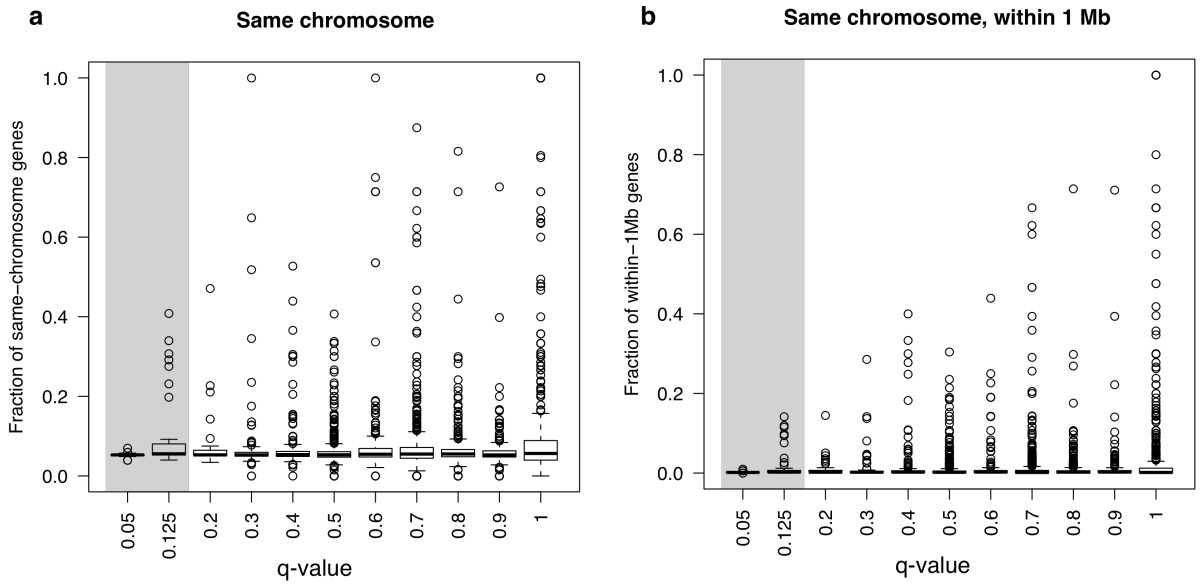


B. Measures per gene-set

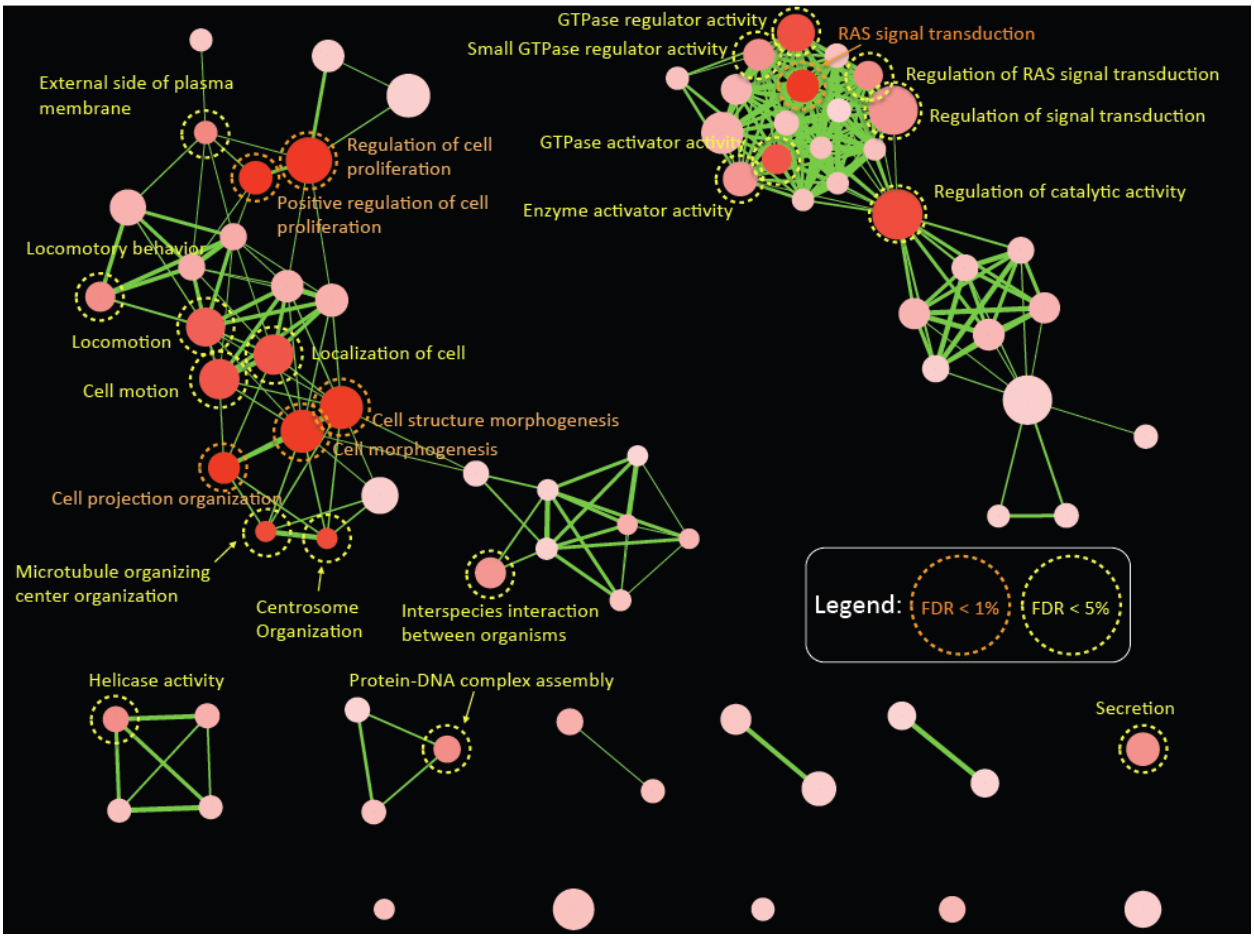
Gene-sets were grouped according to pre-fixed intervals of their q-values (0 < q-value ≤ 0.05, 0.05 < q-value ≤ 0.125, etc), as indicated on the x axis. The q-value on the tick marks identifies the right boundary of the interval. There were no gene-sets with q-value equal to zero. The area in grey corresponds to the sets used to draw the functional network. For every q-value interval, the total number of deletions, total deletion length and median length per gene-set were displayed as a boxplot.



Supplementary Figure 9. Bias control for genome proximity
Genomic proximity scores for gene-set as a whole. Proximity criterion: **a,** Genes located on the same chromosome. **b,** Genes located on the same chromosome, within 1 Mb. Gene-sets were grouped according to pre-fixed intervals of their q-values, as described in legend of Supplementary Fig. 8B.



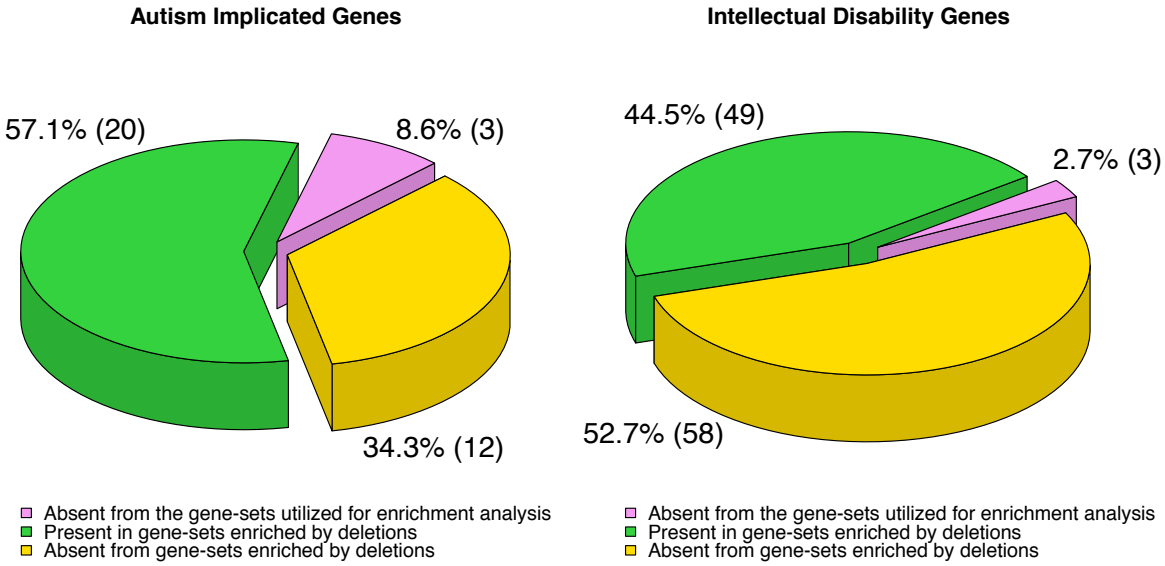
Supplementary Figure 10. Distribution of q-values per gene-set



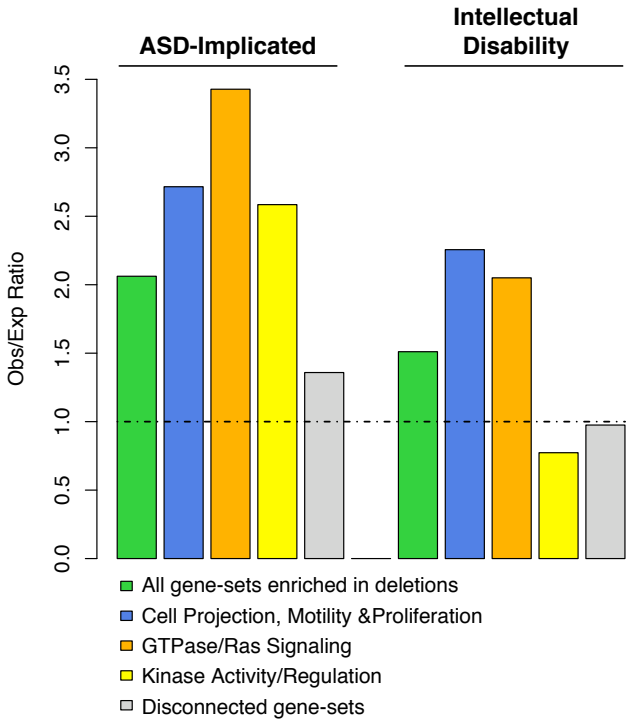
Sets colored by q-value threshold. Gene-sets passing more stringent q-value thresholds were identified in orange (<1%) and yellow (<5%) respectively. Notice that the two major clusters, Cell proliferation, motility and cell projection (1) and GTPase/RAS signaling (2) have at least one or a few gene-set passing the 1% threshold, and several passing the 5% threshold. This demonstrates the statistical reliability of larger clusters.

Supplementary Figure 11. Gene coverage between ASD and/or ID gene-sets and sets enriched for deletions

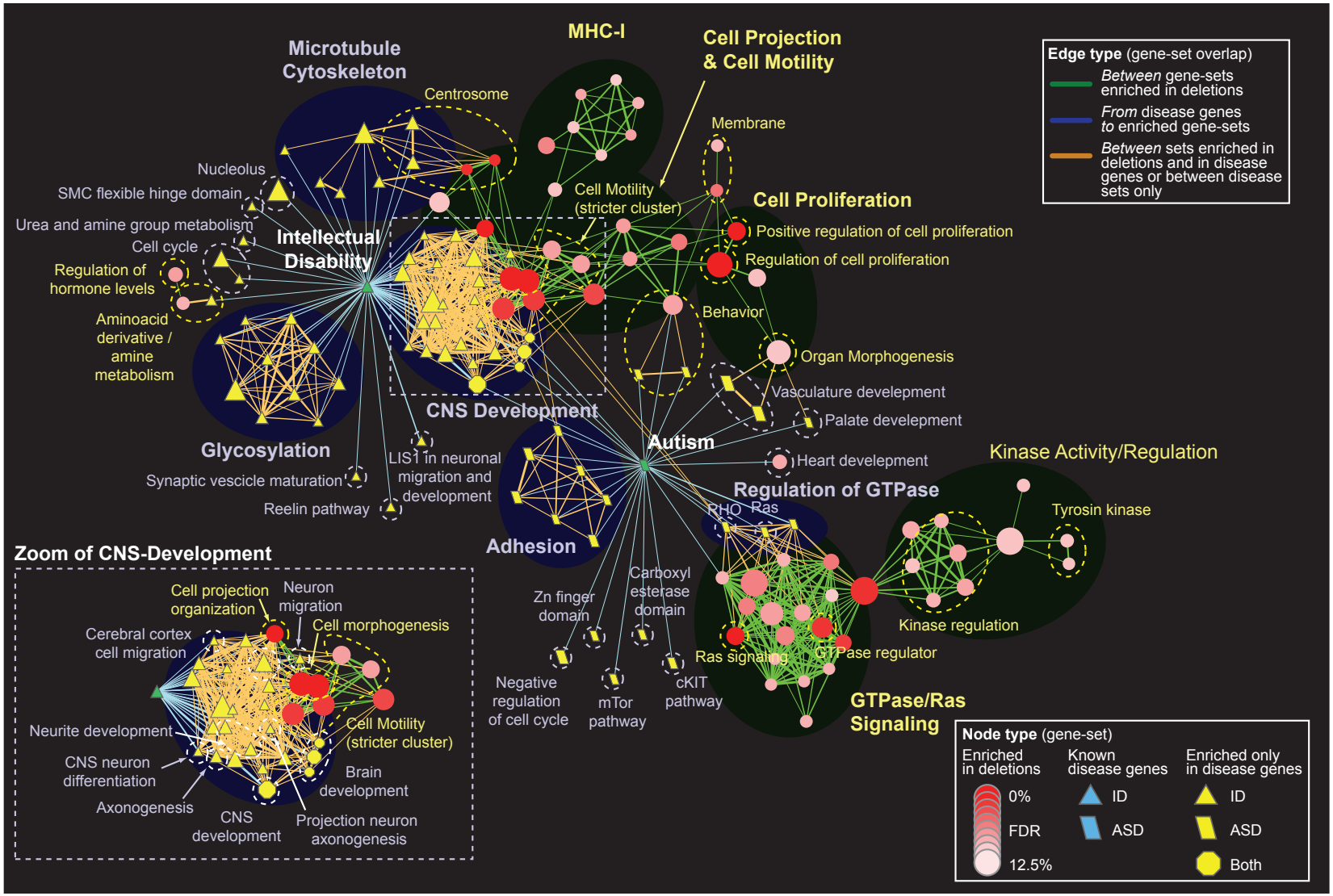
A. Fraction of ASD/ID genes that are also present in gene-sets enrichment in deletions



B. Ratio of observed over expected number of overlapping genes between ASD/ID gene-sets and individual deletion-enriched gene clusters



Supplementary Figure 12: Expanded functional map with high-level of annotation detail
Most of the gene-sets were annotated, especially gene-sets highly significant or with evident connections to brain function. Zoom for the CNS Development functional group is also shown.



Supplementary Table 1. Quality control (QC) steps for CNV analysis of ASD cases and controls

QC Evaluation (sample filters)	#AGP samples removed	#cases after QC	# trios (#duos) [4]	#AGP samples after QC[5]	#Controls removed (SAGE + HapMap)	#Controls after QC (SAGE + HapMap)	#Controls after QC (SAGE + HapMap) [6]
Basic QC							
# Samples	--	1,275	1,256 (16)	4,331	--	1,880 + 101	1,981
Related samples and duplicates	31	1,273	1,254 (16)	4,300	18 + 23	1,862 + 78	1,940
Technical controls	29	1,273	1,254 (16)	4,271	--	--	1,940
Probands with incomplete phenotype	39	1,234	1,216 (26)	4,232	--	--	1,940
Chip call rate (<97%)	34	1,226	1,197 (26)	4,198	1 + 0	1,861 + 78	1,939
Mendelian error Rate (>2%)	41	1,213	1,184 (93)	4,157	--	--	1,939
Sample batch level QC							
STDEV of the intensity [1] (>0.27)	191	1,166	1,065 (93)	3,966	31 + 0	1,830 + 78	1,908
Sample level QC							
STDEV of the Log R ratio [1] (>0.27)	19	1,160	1,059 (102)	3,947	1 + 0	1,829 + 78	1,907
STDEV of the BAF [1] (>0.17)	22	1,159	1,045 (106)	3,925	--	--	1,907
#CNV calls by iPattern (>mean + 3SD)	28	1,150	1,028 (118)	3,897	15 + 0	1,814 + 78	1,892
#CNV calls by quantiSNP (>mean + 3SD)	29	1,141	1,007 (125)	3,868	9 + 0	1,805 + 78	1,883
Chromosomal abnormalities (CNVs >7.5 Mb in size) [2]	14	1,136	995 (125)	3,854	23 + 2	1,782 + 76	1,858
Excessive de *novo* CNVs (>5 *de novo* CNVs)	6	1,134	993 (125)	3,848	--	--	--
From same family	2	1,132	991 (125)	3,846	--	--	--
European ancestry (EA) [3]	652	996	876 (104)	3,194	521 + 50	1,261 + 26	1,287

[1]STDEV= standard deviation; BAF = B allele frequency; [2]In total 19 samples (six cases and 13 parents) with CNVs >7.5 Mb were excluded, ie. five of them had been excluded during previous QC steps; The two cases with duplication of the chr.Y were kept in the analyses (since CNV analyses only used autosomes and chrX; they are listed in Supplementary Table 2; [3]EA: European ancestry, as estimated by spectral GEM[13].

[4]1,275 cases (1,256 complete trios and 16 duos) were genotyped and those passing the QC filtering criteria were used in the rare CNV analysis (see also Supplementary Fig. 1). Incomplete families, where proband-father/mother duos passed QC filters were also analyzed for CNVs; [5]Total count includes partial families with only parents; [6]Two control groups genotyped with the Illumina 1M beadchip array (totaling 1,287 unrelated EA passing QC filters); i) 1,261 subjects (unrelated EA after QC) from the Study on Addiction - Genetics and Environment (SAGE) case-control GWAS study obtained through dbGAP (Study Accession: phs000092.v1.p1), and ii) 26 HapMap CEU samples (unrelated EA after QC) obtained from Illumina.

Supplementary Table 2. Chromosome abnormalities larger than 7.5 Mb detected during QC

Sample group	Sample ID	Status	Gender	Tissue	Chr.	Anomaly type	Karyotype	Family Type	Comments
AGP	1108_2	Father	M	L	2p	p.arm.b.allele.freq.split	n/a	MPX	Mosaicism
AGP	5008_1	Mother	F	L	5q14.3-telom	q.arm.high.inten, q.arm.b.allele.freq.split	n/a	MPX	Partial duplication (5q14.3-telom); NOT found in another DNA batch by Affy500K-EA
AGP	14217_3491	Mother	F	B	9q-telom	q.arm.b.allele.freq.split	n/a	SPX	Mosaicism
AGP	14123_2171	Mother	F	B	11p12-telom	p.arm.b.allele.freq.split	n/a	SPX	Mosaicism (telom-p12)
AGP	3047_001	Father	M	BBC	12	high.inten, b.allele.freq.split	n/a	MPX	Whole chr12 duplication
AGP	1163_2	Father	M	L	14	high.inten, b.allele.freq.split	n/a	MPX	Whole chr14 duplication
AGP	1763_211	Father	M	L	14q12-q32.33	q.arm.high.inten, q.arm.b.allele.freq.split	n/a	MPX	Partial duplication (14q12-q32.33)
AGP	9518_102	Father	M	n/a	16	high.inten, b.allele.freq.split	n/a	MPX	Whole chr16 duplication
AGP	5355_2	Father	M	B	22q11.21-telom	q.arm.b.allele.freq.split	n/a	SPX	Mosaicism (q11.21-telom)
AGP	1341_1	Mother	F	CL	X	female.low.X.inten, b.allele.freq.split	XX.X0	MPX	Mosaicism
AGP	3100_002	Mother	F	BBC	X	female.high.X.inten. b.allele.freq.split	XX.XXX	MPX	Mosaicism
AGP	5134_1	Mother	F	L	X	b.allele.freq.split	XX.X0	MPX	Mosaicism; NOT found in another DNA batch by Affy500K
AGP	5294_1	Mother	F	B	X	female.low.X.inten, b.allele.freq.split	XX.X0	SPX	Mosaicism; NOT found in another DNA batch by Affy500K
AGP	5010_3	Proband	M	L	1q44, 9q-telom	1q44.low.inten & b.allele.freq.split; 9q.arm.high.inten, 9q.arm.b.allele.freq.split	n/a	MPX	Partial duplication (9q-telom); found in the same DNA batch by Affy500K-EA
AGP	14270_3930	Proband	F	B	6q25.3-q27	high.inten, b.allele.freq.split	46,XX.ish der(22)t(6;22) (6q25.3;p11.2) pat(6qtel+)	SPX	~10 Mb duplication (6q25.3-q27); confirmed by FISH, resulting from a balanced translocation in the father
AGP	5467_3	Proband	M	B	1q42.3-q44	high.inten, b.allele.freq.split	n/a	SPX	13.5 Mb duplication (1q42-q44); confirmed by qPCR (probes placed in *PD5*, *SMYD3*)
AGP	6379_4	Proband	M	L	14	high.inten, b.allele.freq.split	46,XY	SPX	Mosaicism, whole chr14 duplication by one algorithm only; karyotyping excluded a chr14 trisomy; cell line artifact
AGP	5420_3	Proband	M	L	21	male.high.X.inten, b.allele.freq.split	47,XY+21	UKN	confirmed by karyotyping

Sample group	Sample ID	Status	Gender	Tissue	Chr.	Anomaly type	Karyotype	Family Type	Comments
AGP	13137_1543	Proband	F	B	8p12-8q12.1	8p-q duplication	47,XX,+r[10]/46,XX[70]	UKN	26 Mb duplication of 8p12-8q12.1; karyotype: mosaic for a supernumerary ring chromosome
AGP	5257_3	Proband	M	B	Y	male.high.Y.inten	47,XYY	SPX	confirmed by karyotyping
AGP	5515_3	Proband	M	B	Y	male.high.Y.inten	47,XYY	UKN	confirmed by karyotyping
SAGE controls	B453672_1007871661	Control	F	L	7p21.2-telom	high.inten, b.allele.freq.split	n/a	–	Partial duplication (14.56 Mb; telom-p21.2)
SAGE controls	B192507_1007872574	Control	M	L	10q21.1-q22.1	high.inten, b.allele.freq.split	n/a	–	Partial duplication (17.9 Mb; 10q21.1-q22.1)
SAGE controls	B538284_1007874460	Control	F	L	12q24.31-q24.33	low.inten, b.allele.freq.split	n/a	–	Partial deletion (11 Mb; 12q24.31-q24.33)
SAGE controls	B875847_1007874425	Control	F	L	12q14-telom	b.allele.freq.split	n/a	–	Mosaicism
SAGE controls	B712092_1007853733	Control	M	L	19q13.2-telom	b.allele.freq.split	n/a	–	Mosaicism
SAGE controls	B182886_0057061564	Control	F	L	X	b.allele.freq.split	XX.X0	–	Mosaicism
SAGE controls	B227532_1007875264	Control	F	L	X	female.low.X.inten, b.allele.freq.split	XX.X0	–	Mosaicism
SAGE controls	B272418_1007873376	Control	F	L	X	female.low.X.inten	X0	–	–
SAGE controls	B331706_1007843500	Control	F	L	X	female.low.X.inten	X0	–	–
SAGE controls	B363038_1007872238	Control	F	L	X	female.low.X.inten, b.allele.freq.split	XX.X0	–	Mosaicism
SAGE controls	B400207_1007874431	Control	F	L	X	b.allele.freq.split	XX.X0	–	Mosaicism
SAGE controls	B442361_1007872258	Control	F	L	X	b.allele.freq.split	XX.X0	–	Mosaicism
SAGE controls	B443117_1007872671	Control	F	L	X	female.low.X.inten, b.allele.freq.split	XX.X0	–	Mosaicism
SAGE controls	B449686_1007840765	Control	F	L	X	b.allele.freq.split	XX.X0	–	Mosaicism
SAGE controls	B617470_1007872267	Control	F	L	X	b.allele.freq.split	XX.X0	–	Mosaicism
SAGE controls	B650091_1007872185	Control	F	L	X	b.allele.freq.split	XX.X0	–	Mosaicism
SAGE controls	B709845_1007875224	Control	F	L	X	b.allele.freq.split	XX.X0	–	Mosaicism
SAGE controls	B717554_1007875274	Control	F	L	X	female.low.X.inten, b.allele.freq.split	XX.X0	–	Mosaicism
SAGE controls	B729328_1007872228	Control	F	L	X	b.allele.freq.split	XX.X0	–	Mosaicism
SAGE controls	B747337_1007873392	Control	M	L	XY	male.med.lo.Y.inten	XY.X0	–	Mosaicism
SAGE controls	B821840_1007852894	Control	F	L	X	female.low.X.inten, b.allele.freq.split	XX.X0	–	Mosaicism
SAGE controls	B888881_1007873634	Control	M	B	XY	Multiple	XY/XXY/XYY	–	Mosaicism
SAGE controls	B915478_1007845341	Control	M	B	X	male.high.X.inten	XXY	–	–

Samples including ASD trios and controls that passed all QC filters and showed CNVs by any algorithm larger than 7.5 Mb on any of the autosomes or chrX were further inspected manually by plotting their log2 ratio intensities as well as allelic genotype ratios. A molecular size cutoff of >7.5 Mb was selected to be consistent with large cytogenetically visible chromosome abnormalities**.** Samples containing such alterations were excluded from the analyses, except for two AGP probands with an XYY karyotype that were retained since chrY markers were not used in the CNV analysis.

Abbreviations: B: blood, BBC: buccal swab, EA: early access (Affy500K-EA vs. Affy500K), L: cell line, n/a: not available, UKN: unknown family type (extended family not evaluated for ASD).

Supplementary Table 3: Summary of characteristics of stringent CNVs in cases and controls

A. All-CNVs, All ancestries

	Autism Probands	Controls	Parents
	All CNVs	All CNVs	All CNVs
#Samples	1,132	1,858	2,712
#CNVs [1]	18,075	34,394	41,946
#CNVs/genome [2] (mean / median)	16 / 16	18.5 / 18	15.5 / 15
Mean / Median Size (kb)	111.6 / 47.5	92.9 / 36.6	106.2 / 47.5
% Gain/Loss	26.2 / 73.1	24.5 / 75.1	25.2 / 74.3
#Recurrent/overlapping CNVs (%)/ and –Loci [3]	16,650 (92.1%)/ 827	32,741 (95.2%)/ 1,239	39,936 (95.2%)/ 1,380
#CNVs >1 Mb (%)	230 (1.3%)	268 (0.8%)	461 (1.1%)
#CNVs 100 kb-999 kb (%)	4,881 (27%)	8,017 (23.3%)	11,298 (26.9%)

[1] CNVs detected in the same individual by at least two algorithms were merged with the outside probes used as boundaries to identify a stringent dataset of CNVs, each CNV containing at least five consecutive probes and at least 5 kb in size.
[2] The number of stringent (iPattern and QuantiSNP) CNVs per genome is shown. Using iPattern and QuantiSNP alone 27 CNVs/sample and 35 CNVs/sample are detected, respectively. PennCNV called ~45 CNVs/sample, and was used to confirm inheritance status (not shown).
[3] The number and percentage of recurrent and/or overlapping CNVs in the dataset (%), and corresponding number of CNV loci.

B. All-CNVs, European-only

	Autism Probands	Controls	Parents
	All CNVs	All CNVs	All CNVs
#Samples	996	1,287	2,196
#CNVs [1]	15,583	22,573	33,704
#CNVs/genome [2] (mean / median)	15.6 / 15	17.5 / 17	15.3 / 15
Mean / Median Size (kb)	110.1 / 47.5	97.9 / 39.7	104 / 46.2
% Gain/Loss	26 / 73.4	24.9 / 74.7	25.2 / 74.3
#Recurrent/overlapping CNVs (%)/ and –Loci [3]	14,349 (92.1%) / 722	21,279 (94.3%)/ 855	31,962 (94.8%)/ 1,152
#CNVs >1 Mb (%)	189 (1.2%)	201 (0.9%)	356 (1.1%)
#CNVs 100 kb-999 kb (%)	4,205 (27%)	5,624 (24.9%)	9,045 (26.8%)

The table shows a summary of the characteristics of all CNVs detected in ASD cases and controls of European ancestry as estimated by Spectral GEM[13]. See Supplementary Table 3A for further legend details.

Supplementary Table 4. Characteristics of rare CNVs in European ASD probands and controls

A. Summary of rare CNVs in European probands and controls

	ASD probands, European (n = 996 probands, 876 trios)				Controls, European (n = 1,287)	
	All CNVs	Rare CNVs[1]	Rare putative *de novo* CNVs[2]	Rare inherited CNVs[2 ‡]	All CNVs	Rare CNVs[1]
# Samples	996	889	119	858	1,287	1,146
# CNVs[3]	15,583	2,382	165	2,137	22,573	3,096
Mean/median # CNVs/genome[4]	15.6/15	2.4/2	0.2/0	2.1/2	17.5/17	2.4/2
Mean/median CNV size (kb)	110.1/47.5	182.7/89.9	310.6/106.5	170/88.3	97.9/39.7	176/90.9
% Gain/Loss	26/73.4	48.4/51.6	54.5/45.5	47.4/52.6	24.9/74.7	50.7/49.3
# Recurrent/overlapping CNVs (%) / # Loci[5]	14,349 (92.1%) /722	1,477 (62%) /366	64 (39%) /26	1,266 (59.2%) /314	21,279 (94.3%) /855	2,049 (66.2%) /475
# CNVs >1 Mb (%)	189 (1.2%)	44 (1.8%)	8 (4.3%)	32 (1.5%)	201 (0.9%)	59 (1.9%)
# CNVs >100 kb - 999 kb (%)	4,205 (27%)	1,040 (43.7%)	78 (47.6%)	919 (43%)	5,624 (24.9%)	1,354 (43.7%)

[1] Rare stringent CNVs ≥30 kb present in the total sample at a frequency <1%**.** CNVs detected in the same individual by at least two algorithms were merged with the outside probes used as boundaries.

[2] Inheritance state was estimated for CNVs detected in 876 probands from complete trios where array data after QC was available from both parents. Using computational data only, the rate of *de novo* CNV events was initially estimated to be 6.9% (165/2,382) with 13.6% (119/876) of trio families having at least one putative *de novo* CNV. When additional computational and laboratory validation data was added (see section D. CNV Verification), we confirmed that at least 5.6% (49/876) of trio families carried at least one *de novo* CNV (average of 1.1 verified *de novo* CNVs/sample). ‡ Inheritance status could not be assigned to 80 CNVs, either because at least one of the parents did not pass QC or because the results were ambiguous. Therefore, inherited and *de novo* CNVs do not sum to 100%.

[3] Samples containing CNVs larger than 7.5 Mb were excluded (Supplementary Table 2).

[4] The number of stringent (iPattern and QuantiSNP) CNVs per genome is shown. Using iPattern and QuantiSNP alone 27 CNVs/sample and 35 CNVs/sample are detected, respectively. PennCNV called ~45 CNVs/sample.

[5] The number and percentage of recurrent and/or overlapping CNVs in the dataset (%), and corresponding number of CNV loci.

B. Rare *de novo* and inherited CNVs in ASD probands

	ASD Probands, European						
***De novo* CNVs**	**Rare (all) (N=996)**	**Rare *de novo* (N=876)**	**Rare *de novo* (MPX) (N=348)**	**Rare *de novo* (SPX) (N=393)**	**Rare *de novo* - after exp. validation (N=876)**	**Rare *de novo* - after exp. validation (MPX) (N=348)**	**Rare *de novo* - after exp. validation (SPX) (N=393)**
#Samples	889	119	46	50	50	19	22
#CNVs [3]	2,382	165	69	67	55	20	26
Mean / median #CNVs/genome	2.4 / 2	0.2 / 0	0.2 / 0	0.2 / 0	0.06 / 0	0.06 / 0	0.07 / 0
Mean / median CNV size (kb)	182.7 / 89.9	310.6 / 106.5	306.6 / 117.9	265.5 / 112.2	524.5 / 167.9	531.3 / 129.1	341.1 / 194.7
% Gain/Loss	48.4 / 51.6	54.5 / 45.5	55 / 44.9	47.8 / 52.2	33.4 / 65.5	25 / 75	38.5 / 61.5
#Recurrent/overlapping CNVs (%)/ and Loci [1]	1,477 (62%) / 366	64 (39%) / 26	16 (23.2%) / 7	14 (20.9%) / 5	19 (35.2%) / 9	2 (10%) / 1	6 (23.1%) / 3
#CNVs >1 Mb (%)	44 (1.8%)	8 (4.3%)	3 (4.3%)	3 (4.5%)	5 (9.1%)	2 (10%)	1 (3.8%)
#CNVs 100 kb-999 kb (%)	1,040 (43.7%)	78 (47.6%)	36 (52.2%)	33 (49.3%)	29 (53.7%)	10 (50%)	16 (61.5%)

This table shows a summary of the characteristics of *de novo* rare CNVs in multiplex (MPX) and simplex (SPX) families, before and after adding evidence from experimental laboratory validation. Inheritance state was estimated computationally for CNVs detected in 876 probands from complete trios where array data after quality control was available from both parents (i.e., up to a maximum of 876 complete trios). The *de novo* CNV rate was initially estimated to be 6.9% (165/2,382) with 13.6% of trio families (119/876) having at least one *de novo* CNV.
SPX= simplex ASD family (no first-third degree relatives with ASD); MPX= multiplex ASD family (two or more first-third degree relatives affected with ASD); U= unknown family type (extended family not evaluated);
[1] The number and percentage of recurrent and/or overlapping CNVs in the dataset, and corresponding number of CNV loci.

C. Rare inherited CNVs in ASD probands

	ASD Probands, European								
Inherited CNVs	**Rare inherited (all) (N=876)**	**Rare inherited (MPX) (N=389)**	**Rare inherited (SPX) (N=434)**	**Rare inherited (maternal) (N=996)**	**Rare inherited (maternal-MPX) (N=389)**	**Rare inherited (maternal-SPX) (N=434)**	**Rare inherited (paternal) (N=996)**	**Rare inherited (paternal-MPX) (N=389)**	**Rare inherited (paternal-SPX) (N=434)**
#Samples	858	340	382	666	261	297	606	240	281
#CNVs[3]	2,137	850	977	1,107	436	491	983	387	469
Mean / median #CNVs/genome	2.1 / 2	2.2 / 2	2.3 / 2	1.1 / 1	1.1 / 1	1.1 / 1	1 / 1	1 / 1	1.1 / 1
Mean / median CNV size (kb)	170 / 88.3	182.8 / 88.3	160 / 88.3	163.8 / 88.6	176.8 / 92.9	148.9 / 85.2	178.1 / 88.3	190.2 / 82.4	173.6 / 94.5
% Gain/Loss	47.4 / 52.6	51.1 / 48.9	45.8 / 54.2	47.2 / 52.8	48.2 / 1.8	46.8 / 53.2	47.7 / 52.3	53.7 / 46.3	45 / 55
#Recurrent/overlapping CNVs (%)/ and –Loci[1]	1,266 (59.2%) / 314	389 (45.8%) / 136	455 (46.6%) / 155	551 (49.8%) / 181	149 (34.2%) / 58	164 (33.4%) / 68	459 (46.7%) / 142	126 (32.6%) / 53	155 (33%) / 62
#CNVs >1 Mb (%)	32 (1.5%)	17 (2%)	11 (1.1%)	16 (1.4%)	8 (1.8%)	5 (1%)	16 (1.6%)	9 (2.3%)	6 (1.3%)
#CNVs 100 kb-999 kb (%)	919 (43%)	362 (42.6%)	425 (43.5%)	475 (42.9%)	193 (44.3%)	204 (41.5%)	418 (42.5%)	149 (38.5%)	216 (46.1%)

This table shows a summary of the characteristics of rare inherited CNVs in multiplex (MPX) and simplex (SPX) families, as well as further before and after experimental validation. See Supplementary Table 4B for further legend details. ‡ Inheritance status could not be assigned to 80 CNVs, either because at least one of the parents did not pass all QC filters or because the results were ambiguous. Therefore the columns “Rare *de novo* (N=876)” in Supplementary Table 4B and “Rare inherited (all) (N=876)” in this table do not sum up to 100%.

[1] The number and percentage of recurrent and/or overlapping CNVs in the dataset, and corresponding number of CNV loci.

Supplementary Table 5. Global rare CNV burden analyzes with respect to CNV size and CNV rate

			CNV rate[1]			CNV sample proportion[2]			Total CNV size (kb)			Average CNV size (kb)		
Type	**Classification**	**Total CNVs (n)**	***P***	**Case/ctrl ratio**	**Baseline rate (ctrl)**	***P***	**Case/ctrl Ratio**	**Baseline rate (ctrl)**	***P***	**Case/ctrl ratio**	**Baseline rate (ctrl)**	***P***	**Case/ctrl ratio**	**Baseline rate (ctrl)**
All	All	5,478	0.578	0.99	2.41	0.465	1.00	0.89	0.293	1.03	475.7	0.276	1.03	176.2
Deletions	All	2,757	0.158	1.04	1.19	0.299	1.02	0.68	0.534	0.99	236.7	0.652	0.97	138.2
Duplications	All	2,721	0.884	0.95	1.22	0.892	0.97	0.69	0.124	1.08	381.4	0.165	1.07	217.1
CNV frequency[3]														
All	2-6x	1,831	0.437	1.01	0.80	0.805	0.97	0.53	0.030	1.17	222.9	0.233	1.06	148.7
	1x	1,419	0.469	1.01	0.62	0.520	1.00	0.46	0.817	0.93	183.1	0.874	0.91	137.9
Deletions only	2-6x	1,094	0.111	1.08	0.46	0.486	1.00	0.37	0.079	1.20	171.4	0.188	1.13	138.6
	1x	880	0.231	1.05	0.38	0.165	1.06	0.32	0.550	0.99	125.6	0.613	0.97	107.9
Duplications only	2-6x	977	0.809	0.94	0.44	0.946	0.91	0.34	0.035	1.23	287.6	0.167	1.12	232.2
	1x	896	0.825	0.94	0.40	0.853	0.94	0.33	0.758	0.93	227.6	0.772	0.93	183.3
CNV size														
All	30 – 500 kb	5,086	0.642	0.99	2.24	0.429	1.00	0.87	0.665	0.98	305.9	0.457	1.00	116.9
	≥ 500 kb	392	0.293	1.06	0.17	0.571	0.99	0.15	0.061	1.12	1,016.0	0.193	1.07	941.6
Deletions only	30 – 500 kb	2,645	0.132	1.05	1.14	0.294	1.02	0.67	0.629	0.99	170.8	0.841	0.96	101.1
	≥ 500 kb	112	0.740	0.90	0.05	0.686	0.93	0.05	0.198	1.10	957.7	0.146	1.13	934.5
Duplications only	30 – 500 kb	2,441	0.944	0.93	1.10	0.863	0.97	0.65	0.416	1.01	234.8	0.112	1.04	136.9
	≥ 500 kb	280	0.167	1.14	0.12	0.432	1.03	0.11	0.130	1.11	992.5	0.333	1.04	939.0

Samples and CNVs that failed stringent quality criteria had previously been excluded, maintaining CNVs with ≥30 kb in size. We tested for global CNV burden in European cases (n=996) compared to European controls (n=1,287) considering rare CNVs, i.e., CNVs that are present in less than 1% of our total sample. Analyses were further stratified according to CNV type (deletions-only and duplications-only) and frequency (CNVs observed 2-6 times and in isolated cases). Genome-wide *P*-values were estimated by permutation (one-sided, 100,000 permutations), and report on four tests for CNV burden: number of CNVs (CNV rate), CNV sample proportion (proportion of samples with one or more CNVs), total kb size spanned, and average CNV size. The baseline rate in controls and the fold increase in cases (case/control ratio) are listed for each analysis.

[1] #CNVs per sample

[2] Proportion of samples with one or more CNVs

[3] CNV frequency: CNVs observed 2-6 times in the total sample (2-6x) and one time (1x).

Supplementary Table 6. Examples of ASD candidate genes or loci identified by *de novo* and rare-inherited CNVs[1]

Gene/ locus[1]	Cyto-band	# *de novo* CNVs in ASD, sample id (gender:parental origin:family:tissue)	# rare inherited CNVs in ASD, sample id (gender:inheritance:family:tissue)	#rare CNVs in 1,287 controls	*P* (n=1,287 controls)[2]	Description
SHANK2[#]	11q13.3-q13.4	66 kb del, 5237_3 (M:Mat:SPX:B) 68 kb del, 6319_3 (M:Pat:SPX:B)	0	0	(n.s.) ‡	Exonic deletions in *SHANK2*. Case 6319_3 also carries a rare recurrent paternally inherited 468 kb deletion at 15q11.2 between BP1-BP2 of the Prader-Willi/Angelman syndrome region. Both have PDD-NOS and mild ID.
SYNGAP1[#]	6p21.32	112 kb del, 5353_3 (F:na:SPX:B)	0	0	(n.s.) ‡	Deletion of 5 genes including the *SYNGAP1* ID gene. Nonsyndromic ASD and ID.
DLGAP2[#]	8p23.3	817 kb dup, 5290_3 (M:Pat:SPX:B)	0	0	(n.s.)	Duplication intersects *DLGAP2 (*interacts with *DLG4* and *SHANK2)*; also carries an 828 kb rare paternal duplication intersecting *PSD3* (found as a rare inherited deletion or duplication in two other cases). Nonsyndromic ASD.
CSNK1D/ SLC16A3[#]	17q25.3	2 consecutive *de novo*, 5444_3 829 kb dup and 64 kb del, (M:Pat:SPX:B)	42 kb dup, 6164_3 (F:Mat:SPX:na)	0	(n.s.) ‡	Duplication in 5444_3 intersects *SLC16A3* and deletion removes exons of *SLC16A3* and *CSNK1D. CSNK1D* is involved in circadian rhythms. Translocation in parents has been ruled out. The maternal duplication in 6164_3 spans *SLC16A3* and intersects *CCDC57* and *CSNK1D*. Both subjects have nonsyndromic ASD.
NRXN1	2p16.3	184 kb dup, 14068_1180 (M:Mat:SPX:B) 191 kb del, 13017_223 (F:Pat:U:B) 155 kb del, 13037_463 (M:Pat:U:B) 232 kb del, 13153_1703 (M:Pat?:U:B)	4 del (4M:3xPat, 1xMat)	4 del, 1 dup	(n.s.)**	*De novo* CNVs delete or duplicate exons. Rare-inherited CNVs in ASD cases and controls are intronic. ASD cases with *de novo* CNVs have nonsyndromic ASD with varying degrees of ID; the patient with the duplication and one with a deletion have normal IQ. **There is an excess of exonic CNVs overlapping *NRXN1* in cases compared to controls. *P*-value for 4,964 controls is 7.7×10^{-4}.
DiGeorge syndrome region - 22q11 (CNVR)	22q11.21	2.58 Mb del, 3183_007 (M:Mat:MPX:BBC)	2.54 Mb dup, 3127_004 (M:Pat:MPX:L) 2.54 Mb dup, 5261_4 (M:Pat:MPX:L) ---------- 742 kb dup, 6125_4 (M:Pat:SPX:na) 744 kb dup, 3067_005 (M:Pat:MPX:L) 730 kb dup, 14230_3640 (M:Mat:SPX:B)	0 ---------- 1 dup	0.030	A typical 22q11.2 deletion associated with DiGeorge syndrome is in 3183_007 with Asperger syndrome; his first-degree cousin, affected with autism, does not have the deletion. Reciprocal 22q11.2 microduplications were identified in 3127_004 and 5261_4. Smaller CNVs outside the DiGeorge critical region were identified in 6125_4, 3067_005, 14230_3640, and in one control sample.
DDX53/ PTCHD1[#] (CNVR)	Xp22.1	0	385 kb del, 13047_563 (M:Mat:MPX:B) 54 kb del, 5111_3 (M:Mat:MPX:B) 47 kb del, 3253_004 (M:Mat:MPX:BBC) 63 kb del, 5065_3 (M:Mat:MPX:B) 121 kb del, 5298_3 (M:Mat:SPX:B) 103 kb del, 3424_003 (M:Mat:SPX:B) 164 kb del, 5240_4 (M:Mat:SPX:B)	0	3.1×10^{-3} ‡**	CNVR: All seven CNVs are in males and are maternally inherited. *DDX53* and *PTCHD1* are separated by ~330 kb. The first deletion encompasses *DDX53* and is less than 50 kb upstream of *PTCHD1*. The next three deletions are between 5-30 kb upstream of *DDX53*. The last three fall within 350 kb upstream of *PTCHD1* (with 5240_4 deleting exon 1)(see Supplementary Fig. 4). No CNVs seen in an additional 3,677 controls. ***P*-value for 4,964 controls is 3.6×10^{-6}.
IL1RAPL1	Xp21.2	0	547 kb dup, 5126_4 (M:Mat:MPX:L) 112 kb del, 5036_4 (M:Mat:MPX:L)	1 del	(n.s.)	Maternally-inherited CNVs in ASD males. The 547 kb intragenic duplication involves exons. The 112 kb deletion is intronic. *IL1RAPL1* is a known ID locus. Both subjects have ASD without ID. The control carrying a 412 kb deletion is a female.
DMD	Xp21.1	0	381 kb dup, 5126_4 (M:Mat:MPX:L) 215 kb dup, 3019_003 (M:Mat:MPX:BBC)	1 del	(n.s.)	Maternally-inherited exonic duplications in ASD males. Case 5126_4 also carries a rare duplication of *IL1RAPL1*. The control carrying the 128 kb exonic deletion is female. The ASD brother of 3019_003 also has the duplication.
AGBL4[#]	1p33	0	13 del (~85 kb)	6 del	0.02	ATP/GTP-binding protein-like 4 gene; highly expressed in brain.

[1]Based on the analysis of 996 EA ASD cases and 1,287 SAGE controls. Additional phenotypic data is found in Supplementary Tables 7A and 7B. # = novel ASD loci detected in this study; del = deletion; dup = duplication; M = male; F = female; Pat = CNV inherited or arising *de novo* on father's chromosome; Mat = CNV inherited or arising *de novo* on mother's chromosome; SPX = simplex ASD family; MPX = multiplex ASD family; U = unknown family type; B = blood-derived DNA; L = cell-line derived DNA; BBC = buccal swab DNA; na = origin of DNA not available; ID, intellectual disability; CNVR (copy number variable region) = CNVs were merged by any overlap into one CNVR. The dashed line within a cell represents CNV regions overlapping and/or immediately adjacent to each other. CNV comparison data is found at the Autism Chromosome Rearrangement Database (http://projects.tcag.ca/autism/)[21].

[2] Nominal *P*-values obtained by permutation for gene or CNVR tests. n.s. = non-significant. All *P*-values <0.1 are explicitly listed. ‡ No CNVs seen in additional 3,677 European controls.

Supplementary Table 7a. Rare *de novo* CNVs confirmed experimentally

#	AGP ID	Gender	ASD family type	Tissue[a]	Cytoband	Start-End[b]	Size (bp)	CNV	RefSeq Genes	Inheritance	Assay location/ Gene	Method	Segregation in sibs	Phenotype
1	5437_3	M	Familial	CL	1p31.3	61653274-61906852	253 579	Gain	*NFIA* exonic	*De novo*	*NFIA*	qPCR, Agilent 1M	MZ twin with ASD, no CNV	Autism (based on ADI-R and ADOS), average IQ, language delay, diffuse pontine glioma
2	5437_3	M	Familial	CL	16p12.3	16538314-18268877	1 730 564	Loss	*XYLT1* whole	*De novo*	*XYLT1*	qPCR, Agilent 1M	MZ twin with ASD, no CNV	(see above)
3	5437_3	M	Familial	CL	6p21.31	36440905-36633025	192 121	Gain	*ETV7, PXT1, KCTD20, STK38*	*De novo*	*KCTD20, STK38*	qPCR, Agilent 1M	MZ twin with ASD, no CNV	(see above)
4	5089_5	M	Familial	CL	2q12.1	102292943-102345460	52 518	Loss	*IL1RL1* whole, *IL18R1* exonic	*De novo*	*IL1RL1*	qPCR	MZ twin with autism also carries CNV (1 sister with ASD, no DNA)	Autism (based on ADI-R and ADOS), severe MR (unable to complete IQ measure), below average language, no epilepsy, aplastic kidney removed, no dysmorphic features
5	5451_3	M	Sporadic	Blood	2q31.2	179151463-179256105	104 643	Loss	*TTN* exonic	*De novo*	*TTN*	qPCR	(1 non-ASD brother and 1 non-ASD sister, not tested)	Autism (based on ADI-R and ADOS), average IQ, language delay, speech/language impairment, chronic abdominal discomfort, diarrhea, no dysmorphic features
6	5220_3	F	Sporadic	Blood	3p24.3	19127998-19640299	512 302	Gain	*KCNH8* whole	*De novo*	*KNCH8*	qPCR, Agilent 1M	(1 non-ASD maternal half-sister, no DNA)	ASD (based on ADI-R and ADOS), average IQ, no language delay, no epilepsy; born with asymmetric head, fontanelle closed at age 4 y; dysmorphic features
7	5245_3	M	Familial	CL	3q13.31	117285007-117477191	192 185	Loss	*LSAMP* exonic	*De novo*	*LSAMP*	qPCR, Agilent 1M (seems mosaic)	1 sister with ASD, no CNV detected by qPCR	Autism (based on ADI-R and ADOS), below average non-verbal IQ (<1%ile), language delay, below average language (1%ile), apraxia, abnormal sleep EEG without seizures; born at 29 wks, intraventricular hemorrhage, mild cerebral palsy; alopecia areata, no dysmorphic features
8	5353_3	F	Sporadic	Blood	6p21.32	33399849-33512042	112 194	Loss	*CUTA, LYPLA2P1, PHF1, SYNGAP1* exonic	*De novo*	*SYNGAP1, PHF1*	qPCR	(2 non-ASD brothers and 1 non-ASD sister, not tested)	Autism (based on ADI-R and ADOS), below average non-verbal IQ (<1%ile), language delay, very little spontaneous communication; EEG at age 1-2 y, possible seizure activity
9	5386_3	M	Familial	Blood	6q25.3	156785155-158489874	1 704 720	Loss	*SERAC1, ARID1B, SNX9, SYNJ2, ZDHHC14*	*De novo*	*ARID1B, SNX9*	qPCR, Illumina 1M-duo for sib	1 brother with high functioning form of ASD (with none of the abnormalities present in his brother), no CNV by Illumina 1M	Autism (based on ADI-R and ADOS), deafness (no words, 15-20 signs), epilepsy, facial dysmorphism, agenesis of the corpus callosum, loss of white matter in the occipital lobes, undescended testes
10	5370_3	M	Sporadic	Blood	7q36.2	153775586-153844747	69 162	Loss	*DPP6* exonic	*De novo*	*DPP6*	qPCR, Affy 500K (Marshall et al. 2008), Agilent 1M	(no sibs)	Autism (based on ADI-R and ADOS), no epilepsy, no dysmorphic features, brain CT normal
11	5290_3	M	Sporadic	Blood	8p23.3	704383-1521910	817 528	Gain	*DLGAP2* exonic	*De novo*	*DLGAP2*	qPCR, Affy 500K (Marshall AJHG 2008), Agilent 1M	1 brother with ASD features, no CNV by qPCR	Autism (based on ADI-R and ADOS), unable to complete IQ measure, below average language (1%ile), no epilepsy, no dysmorphic features
12	5032_4	M	Familial	Blood	9p24.3	98998-334508	235 510	Loss	*DOCK8* exonic	*De novo*	*DOCK8*	qPCR	1 sister with ASD, no CNV	Autism (based on ADI-R and ADOS), above average non-verbal IQ, below average language, history of seizures, no dysmorphic features
13	5237_3	M	Sporadic	Blood	11q13.3-q13.4	70154458-70220632	66 175	Loss	*SHANK2* exonic	*De novo*	*SHANK2*	qPCR, Agilent 1M	1 non-ASD sister, no CNV	Autism (based on ADI-R and ADOS), below average non-verbal IQ (1%ile), below average language (<1%ile), no epilepsy, 5th finger clinodactyly, several curled toes
14	5272_3	M	Familial	Blood	12q23.1	98445422-98540678	95 257	Loss	*ANKS1B* intronic	*De novo*	*ANKS1B*	qPCR, Agilent 1M, Illumina 1M-duo for sib	one brother with autism, no CNV	ASD (based on ADI-R and ADOS), average non-verbal IQ (45%ile), delayed early language development but average language abilities, no epilepsy, no dysmorphic features
15	5068_3	F	Familial	Blood	16p11.2	29502984-30127026	624 042	Loss	36 genes	*De novo*	*MVP, GDPD3*	qPCR, Affy 500K (Marshall et al. 2008), Agilent 1M	1 brother with ASD, no CNV	Autism (based on ADI-R and ADOS), low average IQ, language delay, motor delay, neurologic assessment negative, adherent ear lobes; 16p11.2 microdeletion syndrome, 50% mosaicism (reported in Marshall et al. 2008; Fernandez et al. 2009; MM0088-003)

#	AGP ID	Gender	ASD family type	Tissue[a]	Cytoband	Start-End[b]	Size (bp)	CNV	RefSeq Genes	Inheritance	Assay location/ Gene	Method	Segregation in sibs	Phenotype
16	5359_4	M	Sporadic	Blood	16p11.2	29554843-30195224	640 381	Loss	39 genes	*De novo*	*MVP, GDPD3*	qPCR, Affy 500K (Marshall et al. 2008), Agilent 1M	1 non-ASD sister, no CNV	Autism (based on ADI-R and ADOS), anxiety disorder, sleep disturbance, average IQ, delayed early language development but average language abilities, obese, macrocephaly, dysmorphic features, 2-3 toe syndactyly, micropenis, hemivertebra (T10); brain CT normal; 16p11.2 microdeletion syndrome (reported in Marshall et al. 2008; Fernandez et al. 2009; SK0019-004)
17	5262_4	M	Sporadic	CL	16p11.2	29502984-30210849	707 865	Gain	40 genes	*De novo*	*MVP, GDPD3*	qPCR, Affy 500K (Marshall et al. 2008)	1 non-ASD brother and 1 non-ASD sister, no CNV in either	Autism (based on ADI-R and ADOS), anxiety symptoms, below average non-verbal IQ (4%ile), below average language (<1%ile), cleft palate, congenital diaphragmatic hernia, epilepsy, recurrent ear infections, short stature, alopecia in left parietal area, hypertelorism, epicanthic folds, small ears with curved pinnae, smooth philtrum, 5th finger clinodactyly, hyperextensibility, scoliosis; 16p11.2 microduplication syndrome (reported in Marshall et al. 2008; Fernandez et al. 2009; SK0102-004)
18	5056_4	M	Familial	Blood	17q12	34612208-34732327	120 120	Gain	*STAC2* whole, *FBXL20* exonic	*De novo*	*FBXL20*	qPCR	one brother with autism (5056_3), no CNV; 3 other non-ASD brothers: 5056_5, CNV (could be mosaic); 5056_6, CNV; 5056_7, no CNV (all tested with Illumina 1M)	Autism (based on ADI-R and ADOS), severe MR, below average language (<1%ile), no epilepsy, no dysmorphic features, normal neurological exam
19	5444_3	M	Sporadic	Blood	17q25.3	76953064-77782267	829 204	Gain	40 genes; distal breakpoint intersects *SLC16A3*; tandem duplication	*De novo*	*BAHCC1, FASN*	qPCR, Affy 500K (Marshall et al. 2008), Agilent 1M, FISH	1 non-ASD sister, no CNV	Autism (based on ADI-R and ADOS), MR (unable to complete IQ measure), below average language (1%ile), regression of language and motor skills at age 3 y, myoclonic epilepsy, self-injurious behavior, sleep disorder (altered circadian rhythm according to parents), no dysmorphic features
20	5444_3	M	Sporadic	Blood	17q25.3	77785939-77849717	63 779	Loss	*SLC16A3*exonic, *CSNK1D* whole	*De novo*	*CSNK1D, SLC16A3*	qPCR, Agilent 1M	(see above)	(see above)
21	5046_3	M	Familial	CL	20p12.3	8607242-8637441	30 200	Loss	*PLCB1* exonic	*De novo*	*PLCB1*	qPCR	1 brother with ASD, no CNV	Autism (based on ADI-R and ADOS), unable to complete IQ measure, below average language (1%ile), no epilepsy, no dysmorphic features
22	5335_3	M	Sporadic	CL	20p12.1	14545734-14948785	403 052	Loss	*MACROD2* exonic	*De novo*	*MACROD2*	qPCR, Agilent 1M	one healthy sib (not tested)	Autism (based on ADI-R and ADOS), MR (IQ Leiter = 36)
23	6164_3	F	Sporadic	U (blood or CL)	6q25.3	160023074-160081618	58 545	Gain	*WTAP* exonic, *SOD2* exonic	*De novo*	*SOD2*	qPCR	(1 healthy sister, no DNA)	Autism, mild MR, language delay, normal physical exam, normal brain MRI
24	6321_3	M	Sporadic	U (blood or CL)	8q12.3-8q13.1	65354366-66254869	900 504	Loss	*CYP7B1* whole, *BHLHB5* whole	*De novo*	*CYP7B1*	qPCR	(2 healthy sisters, no DNA)	Autism, mild MR, language delay, no dysmorphic features, flat feet, normal neurological exam
25	6246_4	M	Sporadic	U (blood or CL)	9p23	9399606-9631169	231 564	Loss	*PTPRD* exonic	*De novo*	*PTPRD*	qPCR	absent in 1 healthy sister	Autism, mild MR, no language delay, neonatal hypotonia, marfanoid habitus, arachnodactylia
26	6319_3	M	Sporadic	Blood	11q13.3	70119917-70187872	67 956	Loss	*SHANK2* exonic	*De novo*	*SHANK2*	qPCR	(1 healthy brother, no DNA)	PDD-NOS, mild MR, language delay, functional language, no epilepsy, hypermetropia, large and prominent ears, no other dysmorphic features, flat feet, normal brain MRI
27	6240_4	M	Sporadic	Blood	11q24.2-q25	126633939-132060374	5 426 436	Loss	20 genes	*De novo*	*OPCML, KCNJ1*	qPCR	absent in 1 healthy sister	Autism, mild MR, language delay, macrocephaly, automutilations, minor dysmorphic facial features (low ears flat nose and nasal bridge high forehead), normal neurological exam; brain MRI: white matter abnormalities; chromosome 11q deletion syndrome (Jacobsen syndrome)
28	6053_3	M	Familial	U (blood or CL)	12q13.3-q14.1	54218922-58779615	4 560 694	Gain	94 genes	*De novo*	*TSFM, LRIG3*	qPCR	absent in 1 sister with Asperger syndrome	Autism, moderate MR, language delay, normal physical exam
29	6101_4	M	Familial	U (blood or CL)	15q24.3	74735339-74929817	194 479	Loss	*SCAPER* exonic	*De novo*	*SCAPER*	qPCR	absent in 1 brother with autism and 1 unaffected sister	Autism, moderate MR, language delay, articulation defect, normal physical exam

#	AGP ID	Gender	ASD family type	Tissue[a]	Cytoband	Start-End[b]	Size (bp)	CNV	RefSeq Genes	Inheritance	Assay location/ Gene	Method	Segregation in sibs	Phenotype
30	6358_6	M	Sporadic	U (blood or CL)	19p13.3	4548413-5287389	738 977	Loss	*PTPRS, TNFAIP8L1, ARRDC5, JMJD2B, M6PRBP1, DPP9, C19orf10, UHRF1, FEM1A*	*De novo*	*DPP9, JMJD2B, FEM1A*	qPCR	(3 healthy sibs, no DNA)	Autism, low normal IQ, language delay, cleft palate, retrognathia, short philtrum, hypotelorism, strabismus, sandal gap, normal neurological exam, normal brain MRI
31	14068_1180	M	Sporadic	Blood	2p16.3	50493827-50677835	184 009	Gain	*NRXN1* exonic	*De novo*	*NRXN1*	qPCR	(not tested)	Autism (ADOS and ADI-R positive), normal IQ, neurodevelopmental delay with onset at 2 y (first words 9 m, first phrases 36 m, walked at 14 m), functional language; no dysmorphic features, no sleep problems, no epilepsy; brain MRI cortical atrophy
32	14070_1230	M	Familial	Blood	15q26.1	91200007-91283004	82 998	Loss	*CHD2* exonic	*De novo*	*CHD2*	qPCR	(1 brother with ASD not tested)	Autism, mild MR, no language delay, no epilepsy; micrognatia, protruding ears; brain MRI: altered angular gyrus (normal variant, unknown pathological significance)
33	3174_003	M	Familial	Blood	3p24.3	19921061-20096832	175 772	Loss	*RAB5A, KAT2B, EFHB*	*De novo*	array CGH	Agilent 1M (SNP array indicated mosaicism)	absent in 1 brother with autism	ADI and ADOS dx Autism, low average IQ, language delay, some gross and fine motor coordination difficulties, no dysmorphic features or associated medical or psychiatric problems, no epilepsy. This patient also carries a paternally-inherited 16p12.1 microdeletion
34	3183_7	M	Familial	BBC	22q11.21	17241748-19819918	2 578 171	Loss	50 genes	*De novo*	*HIRA, SNAP29*	qPCR	absent in 1 first degree cousin with autism	Asperger (meets criteria for autism on ADI-R and ADOS), complex language disorder with receptive and expressive difficulties; neonatal feeding difficulties, frequent diarrhea, failure to thrive, constant noisy respirations, recurrent chest and ear infections, walked at 18 m, no epilepsy, no dysmorphic features noted, head circumference P93, height P20, intermittent squint; normal brain MRI and sleep EEG; 22q11 deletion syndrome (DiGeorge/velocardiofacial syndrome)
35	13041_503	M	Sporadic	Blood	1q24.2	167493526-167507362	13 837	Loss	*NME7* intronic	*De novo*	*NME7*	qPCR	(not tested)	Autism, non-verbal, severe MR
36	13153_1703	M	Sporadic	Blood	2p16.3	50990306-51222043	231 738	Loss	*NRXN1* exonic	*De novo*	*NRXN1*	qPCR	(not tested)	Autism, verbal, mild MR, Wilms tumor
37	13017_223	F	Sporadic	Blood	2p16.3	50539877-50730546	190 670	Loss	*NRXN1* exonic	*De novo*	*NRXN1*	qPCR	(not tested)	Autism by ADI-R and ADOS, severe learning disability (IQ < 30), non-verbal, no dysmorphic features or associated medical conditions
38	13037_463	M	Sporadic	Blood	2p16.3	51002576-51157742	155 167	Loss	*NRXN1* exonic	*De novo*	*NRXN1*	*qPCR*	(not tested)	Autism, mild MR, no dysmorphic features
39	13082_963	M	Sporadic	Blood	2p25.1	11712589-11741036	28 448	Gain	*NTSR2* whole	*De novo*	*NTSR2*	qPCR	(not tested)	Autism, verbal, normal IQ
40	13046_553	M	Sporadic	Blood	3q26.1	164004033-164101579	97 546	Loss	—	*De novo*	not in gene	qPCR	(not tested)	Autism, verbal, normal IQ
41	13046_553	M	Sporadic	Blood	12p13.2	11407199-11436086	28 887	Loss	*PRB2* exonic	*De novo*	not in gene	qPCR	(not tested)	(see above)
42	13108_1253	M	Familial	Blood	3q26.1	164004033-164101579	97 547	Loss	—	*De novo*	not in gene	qPCR	(not tested)	Autism, verbal, borderline IQ
43	13022_293	M	Sporadic	Blood	4q13.1	63820936-63833261	12 325	Loss	—	*De novo*	not in gene	qPCR	(not tested)	Autism, non-verbal, MR
44	13007_83	M	Sporadic	Blood	6p25.3	757136-794087	36 952	Gain	—	*De novo*	not in gene	qPCR	(not tested)	Autism, verbal, mild MR, bilateral strabismus, atypical gait
45	13094_1113	M	Familial	Blood	6q21	108415352-108444960	29 608	Gain	—	*De novo*	not in gene	qPCR	(not tested)	Autism, non-verbal, normal IQ

#	AGP ID	Gender	ASD family type	Tissue[a]	Cytoband	Start-End[b]	Size (bp)	CNV	RefSeq Genes	Inheritance	Assay location/ Gene	Method	Segregation in sibs	Phenotype
46	13123_1 403	F	Sporadic	Blood	9p24.3-9p24.2	98998-3682923	3 583 926	Loss	*DMRT1, DMRT2, DMRT3, CBWD1, FLJ35024, FOXD4, KIAA0020, DOCK8, RFX3, C9orf66, KANK1, SMARCA2, KCNV2, VLDLR*	*De novo*	*KCNV2*	qPCR	(not tested)	Autism by ADI and ADOS, moderate/mild MR, verbal, normal birth but developed bilateral congenital diaphragmatic hernia; epilepsy, on anti-convulsants; asthma, no dysmorphic features
47	13094_1 113	M	Familial	Blood	13q21.31	63231043-63277373	46 331	Gain	*OR7E156P* whole	*De novo*	*OR7E156P*	qPCR	(not tested)	(see above)
48	13050_5 93	M	Sporadic	Blood	15q11.2, 15q13.1, 15q12	21190624-26203954	5 013 331	Gain	100 genes	*De novo*	*UBE3A, GABRB3*	qPCR	(not tested)	Autism by ADI and ADOS, mild MR, verbal; anoxia at birth due to cord around neck; EEG yielded inconclusive results (abnormalities in one hemisphere); asthma, no dysmorphic features. Family history of autism (nephew Asperger syndrome), depression and Down's syndrome; maternal 15q11-13 duplication
49	1960_30 1	F	Familial	CL	7q22.1	102699832-102798745	98 914	Loss	*DNAJC2, DPY19L2P2, PMPCB, PSMC2, SLC26A5*	*De novo*	SNP array	Illumina 550K (Glessner et al. 2009)	(not tested)	Autism by ADI-R and ADOS, verbal, MR, floppy infant, gastrointestinal problems (AGRE ID AU1558301)
50	1142_4	F	Familial	CL	8q11.21	48631388-48802529	171 142	Gain	*KIAA0146* exonic	*De novo*	SNP array	Affy 5.0 (Bucan et al. 2009)	(not tested)	Autism by ADI-R and ADOS, verbal, normal IQ, seizures, gastrointestinal problems (AGRE ID AU035004)
51	1050_3	F	Familial	CL	14q11.2	20279711-20345174	65 464	Gain	*FAM12A, FAM12B, RNASE1, RNASE6*	*De novo*	SNP array	Illumina 550K (Glessner et al. 2009), Affy 5.0 (Bucan et al. 2009)	(not tested)	Autism by ADI-R and ADOS, verbal, no MR (AGRE ID AU013303)

[a] DNA source: CL, cell line (peripheral blood lymphoblastoid cell line); BBC, buccal swab; U, unknown (most were either blood or buccal swab).

[b] Human reference genome NCBI v36, hg18.

Abbreviations: ADI-R, Autism Diagnostic Interview-Revised; ADOS, Autism Diagnostic Observation Schedule; ASD, autism spectrum disorder; dx, diagnosis; FISH, fluorescent *in situ* hybridization; IQ, intellectual quotient; MR, mental retardation; MZ, monozygotic; PDD-NOS, pervasive developmental disorder-not otherwise specified

Supplementary Table 7b. Rare inherited CNVs confirmed experimentally

#	AGP ID	Gender	ASD family type	Tissue[a]	Cytoband	Start-End[b]	Size (bp)	CNV	RefSeq Genes	Inheritance	Assay location/ Gene	Method	Segregation in sibs	Phenotype	Parental phenotype
1	5072_3	M	Familial	CL	2p16.3	50912249-50955087	42 839	Loss	*NRXN1* intronic	Maternal	*NRXN1*	qPCR, Illumina 1M-duo for sib	present in 1 brother with autism	ASD (based on ADI-R and ADOS), average IQ, no epilepsy, no dysmorphic features	Mother unaffected
2	5110_3	M	Familial	Blood	2q35	217291508-217317701	26 194	Gain	—	Paternal	not in gene	qPCR	(1 brother with autism, not tested)	Autism (based on ADI-R and ADOS), average IQ	Father unaffected
3	5355_3	M	Sporadic	Blood	3p26.3	1978504-2151165	172 662	Gain	*CNTN4* exonic	Maternal	*CNTN4*	qPCR, Illumina 1M-duo for sib	present in 1 non-ASD sister	Autism (based on ADI-R and ADOS), average IQ, no language delay, no epilepsy, integrative sensory dysfunction	Mother unaffected, social difficulties
4	5269_3	M	Sporadic	Blood	3p26.3	2644708-2876647	231 940	Gain	*CNTN4* exonic	Paternal	*CNTN4*	qPCR, Illumina 1M-duo for sib	present in 1 non-ASD brother	Autism (based on ADI-R and ADOS), above average IQ, delayed language development but average language (53%ile), no epilepsy, no dysmorphic features	Father ADHD
5	5269_3	M	Sporadic	Blood	12p13.33	2115897-2127756	11 860	Loss	*CACNA1C* intronic	Paternal	*CACN1AC*	qPCR, Illumina 1M-duo for sib	1 non-ASD brother, no CNV	(see above)	(see above)
6	5267_3	M	Sporadic	Blood	3p26.3	3098326-3184518	86 193	Loss	*CRBN, IL5RA, TRNT1*	Paternal	*CRBN*	qPCR	(1 non-ASD sister, no DNA)	Autism (based on ADI-R and ADOS), average IQ, no language delay, no epilepsy, no dysmorphic features	Father unaffected
7	5210_4	M	Familial	Blood	3q26.31	174722147-174771975	49 829	Gain	*NLGN1* intronic	Paternal	not in gene	qPCR	(1 brother with ASD, not tested; 1 baby sister, no phenotype data, no DNA)	Autism (based on ADI-R and ADOS), average nonverbal IQ, language delay; born at 35 wks by emergency C-section, respiratory distress; cerebral palsy, right hemiplegia, anomalies on MRI compatible with motor deficits; at age 2 y: head circumference P20, weight P90; no epilepsy	Father unaffected
8	5003_3	F	Familial	CL	3p21.31	50,098,421-50,197,697	99 277	Gain	*RBM5* whole, *SEMA3F* exonic	Paternal	*SEMA3F*	qPCR, Affy500K (proband)	2 brothers with autism: 1 with CNV (5003_5), 1 without (5003_4)	ASD (based on ADI-R and ADOS), average IQ, average language, no epilepsy, curved 5th finger, toe syndactyly	Father unaffected
9	5070_4	M	Familial	Blood	4p15.31	20944461-20960842	16 382	Loss	*KCNIP4* intronic	Paternal	*KCNIP4*	qPCR	(1 brother with ASD, not tested)	Autism (based on ADI-R and ADOS), average nonverbal IQ (34%ile), below average language (<1%ile), no epilepsy, no dysmorphic features	Father unaffected
10	5332_3	M	Sporadic	CL	5p12	43037358-43096689	59 332	Loss	*C5orf39* whole, *LOC153684* whole	Maternal	*C5orf39*	qPCR, Illumina 1M-duo for sibs	1 non-ASD twin brother (5332_5, zygosity unknown) and 1 non-ASD brother (5332_4), no CNV in either	Autism (based on ADI-R and ADOS), ADHD, below average nonverbal IQ (<1%ile), language delay, no epilepsy; strabismus, mildly slanted eyes, epicanthic folds, sandal gap	Mother unaffected
11	5521_3	M	Sporadic	Blood	6p22.1	26240643-26359580	118 938	Gain	17 genes	Maternal	*HIST1H2AE*	qPCR	2 non-ASD sisters, no CNV in either	Autism (based on ADI-R and ADOS), below average IQ (<1%ile), nonverbal, seizure disorder, coarse facial features	Mother unaffected
12	5521_3	M	Sporadic	Blood	Xp22.12	19471138-19861338	390 201	Gain	*CXorf23* exonic, *SH3KBP1* exonic	Maternal	*SH3KBP1*	qPCR	2 non-ASD sisters, no CNV in either	(see above)	(see above)
13	5122_3	M	Familial	CL	7p21.1	17868378-18192733	324 356	Gain	*PRPS1L1* whole, *SNX13* exonic	Maternal	not in gene	qPCR, Illumina 1M-duo for sib	1 brother with autism, no CNV (1 non-ASD sister, not tested)	Autism (based on ADI-R and ADOS), severe MR, below average language (<1%ile), seizures, small built, wasted thin extremities, protruding abdomen	Mother unaffected
14	5200_3	M	Sporadic	Blood	7q21.13	89626742-89825983	199 242	Gain	*C7orf63, GTPBP10, STEAP1, STEAP2*	Paternal	*C7ORF63*	qPCR, Illumina 1M-duo for sib	1 non-ASD sister, no CNV	Autism (based on ADI-R and ADOS), below average IQ (5%ile), below average language (3%ile), no epilepsy, no dysmorphic features	Father unaffected
15	5014_4	M	Familial	CL	7q36.2	153118877-153487764	368 888	Gain	*DPP6* exonic	Paternal	*DPP6*	qPCR	(1 brother with ASD, 1 non-ASD brother and 1 non-ASD sister, not tested)	ASD (based on ADI and ADOS), normal IQ, below average language (1%ile), no epilepsy, no dysmorphic features	Father unaffected
16	5130_3	M	Familial	Blood	7q36.2	153742175-153788452	46 278	Loss	*DPP6* exonic	Paternal	*DPP6*	qPCR, Illumina 1M-duo for sib	present in 1 brother with ASD	Autism (based on ADI-R and ADOS), average IQ, oxygen deprivation at birth, tonic clonic seizures from birth to age 6 m, treated with phenobarbital until age 18 m, seizure free since then; right ptosis, no dysmorphic features	Father unaffected

#	AGP ID	Gender	ASD family type	Tissue[a]	Cytoband	Start-End[b]	Size (bp)	CNV	RefSeq Genes	Inheritance	Assay location/ Gene	Method	Segregation in sibs	Phenotype	Parental phenotype
17	5354_3	F	Sporadic	Blood	8p23.1	10653990-10729568	75 579	Loss	*PINX1 exonic*	Paternal	*PINX1*	qPCR, Illumina 1M-duo for sib	1 non-ASD sister, no CNV	Autism (based on ADI-R and ADOS), below average IQ (1%ile), below average language (<1%ile), no epilepsy, no dysmorphic features	Father unaffected, with some social difficulties
18	5275_3	F	Sporadic (identical twins)	Blood	9p24.3	288719-676170	387 452	Gain	*DOCK8* exonic, *KANK1* exonic	Paternal	*DOCK8*	qPCR	(1 MZ twin with ASD, not tested)	ASD (based on ADI-R and ADOS), low average nonverbal IQ (13%ile), below average language (2%ile), no epilepsy, no dysmorphic features	Father unaffected
19	5136_4	M	Familial	Blood	9q34.3	138753347-138762844	9 498	Loss	*LCN10, LCN6, LCN8*	Maternal	*LCN10*	qPCR	(1 brother with ASD and 1 non-ASD half-sister, not tested)	Autism (based on ADI-R and ADOS), average IQ, language delay, no epilepsy, no dysmorphic features	Mother depression
20	5136_4	M	Familial	Blood	12p13.33	2115897-2127756	11 860	Loss	*CACNA1C* intronic	Paternal	*CACN1AC*	qPCR	(see above)	(see above)	Father unaffected, some obsessive behavior noted
21	5262_4	M	Sporadic	CL	10q11.21	42600836-43271395	670 560	Gain	*BMS1, CSGALNACT2, FXYD4, HNRNPF, RASGEF1A, RET, ZNF487*	Paternal	not in gene	qPCR, Illumina 1M-duo for sibs	1 non-ASD brother and 1 non-ASD sister, no CNV in either	Autism (based on ADI-R and ADOS), anxiety symptoms, below average non-verbal IQ (4%ile), below average language (<1%ile), cleft palate, congenital diaphragmatic hernia, epilepsy, recurrent ear infections, short stature, alopecia in left parietal area, hypertelorism, epicanthic folds, small ears with curved pinnae, smooth philtrum, fifth finger clinodactyly, hyperextensibility, scoliosis. This patient also has a *de novo* 16p11.2 microduplication syndrome	Father unaffected, cleft palate
22	5349_3	M	Sporadic	CL	10q21.1	58409439-58421439	12 001	Gain	—	Paternal	not in gene	qPCR	(1 non-ASD brother and 1 non-ASD sister, not tested)	Autism (based on ADI-R and ADOS), below average nonverbal IQ (<1%ile), below average language (<1%ile), no epilepsy	Father unaffected
23	5004_3	M	Familial	Blood	10q26.3	134591133-134689210	98 078	Gain	*C10orf93* whole	Paternal	not in gene	qPCR, Illumina 1M-duo for sib	present in 1 brother with autism	Autism (based on ADI-R and ADOS), below average nonverbal IQ (<1%ile), language delay, no epilepsy, no dysmorphic features	Father unaffected
24	5004_3	M	Familial	Blood	12p13.33	2115897-2127756	11 860	Loss	*CACNA1C* intronic	Maternal	*CACN1AC*	qPCR, Illumina 1M-duo for sib	1 brother with autism, no CNV	(see above)	Mother unaffected
25	5017_3	M	Familial	CL	10q26.3	135134088-135230489	59 332	Gain	*CYP2E1, LOC619207, SYCE1*	Maternal	*FLJ00268*	qPCR	(1 brother with autism, not tested)	Autism (based on ADI-R and ADOS), MR (unable to complete IQ measure), below average language (1 %ile), no epilepsy, no dysmorphic features	Mother unaffected
26	5263_3	M	Sporadic	Blood	11p15.4	5397196-5454375	57 180	Gain	*OR51I1, OR51I2, OR51Q1*	Maternal	not in gene	qPCR	(1 non-ASD brother with possible ADHD, not tested)	Autism (based on ADI and ADOS), below average IQ (<1%ile), below average language (1%ile), no epilepsy, no dysmorphic features	Mother unaffected
27	5263_3	M	Sporadic	Blood	15q11.2	20235613-20807351	571 739	Gain	*CYFIP1, GOLGA9P, LOC283767, NIPA1, NIPA2, TUBGCP5, WHDC1L1*	Maternal	*CYFIP1, NIPA1*	qPCR	(see above)	(see above)	(see above)
28	5119_3	F	Familial	CL	11q23.3	117452643-117537452	84 810	Gain	*SCN2B, SCN4B, TMPRSS4*	Paternal	*TMPRSS4*	qPCR	(1 sister with autism, not tested)	Autism (based on ADI-R and ADOS), average nonverbal IQ, language delay, no epilepsy; coarse facial features, high hairline, frontal bossing, hypertelorism, clinodactyly	Father unaffected
29	5106_3	M	Familial	Blood	12p13.33	2115897-2127756	11 860	Loss	*CACNA1C* intronic	Maternal	*CACN1AC*	qPCR	CNV present in 1 sister with autism (5106-4)	ASD (based on ADI-R and ADOS), below average non-verbal IQ (<1%ile), language delay, no epilepsy, no dysmorphic features, CT scan normal. Sister (5106_4): Autism (based on ADI-R and ADOS), below average non-verbal IQ (<1%ile), language delay, no dysmorphic features	Mother unaffected

#	AGP ID	Gender	ASD family type	Tissue[a]	Cytoband	Start-End[b]	Size (bp)	CNV	RefSeq Genes	Inheritance	Assay location/ Gene	Method	Segregation in sibs	Phenotype	Parental phenotype
30	5106_3	M	Familial	Blood	12p13.32	3244089-3279231	35 143	Gain	*TSPAN9* exonic	Paternal	*TSPAN9*	qPCR, Affy500K (Marshall AJHG 2008), Illumina 1M-duo for sib	CNV present in 1 sister with autism (5106-4)	(see above)	Father unaffected
31	5347_3	M	Sporadic	CL	12p13.31	9760538-9911534	150 997	Loss	*CD69, CLEC2B, CLECL1, KLRF1*	Paternal	*CD69*	qPCR	(1 non-ASD sister, not tested)	Autism (based on ADI-R and ADOS), average IQ, language delay, no epilepsy, no dysmorphic features	Father unaffected, mild academic difficulties
32	5065_3	M	Familial	Blood	12p13.33	2115897-2127756	11 860	Loss	*CACNA1C* intronic	Paternal	*CACN1AC*	qPCR	(1 brother with autism, not tested)	Autism (based on ADI-R and ADOS), below average nonverbal IQ (<1%ile), language delay, no epilepsy, spina bifida oculta, high arched palate, no other dysmorphic features noted	Father unaffected
33	5065_3	M	Familial	Blood	Xp22.11	22860224-22923580	63 356	Loss	upstream of *DDX53*	Maternal	upstream of *DDX53*	qPCR	1 brother with autism, no CNV	(see above)	Mother unaffected
34	5250_4	M	Familial	Blood	15q11.2	20303106-20800564	497 459	Gain	*CYFIP1, GOLGA9P, HERC2P2, LOC283767, NIPA1, NIPA2, TUBGCP5, WHDC1L1*	Paternal	*TUBGCP5, CYFIP1*	qPCR, Illumina 1M-duo for sib	present in 1 brother with ASD (5250_3)	Autism (based on ADI-R and ADOS), below average IQ (<1%ile), no epilepsy, adherent ear lobes	Father unaffected
35	5250_4	M	Familial	Blood	20p12.1	14729684-14773787	44 104	Loss	*MACROD2* intronic	Paternal	*MACROD2*	qPCR	(1 brother with ASD, no CNV data)	(see above)	(see above)
36	5453_4	M	Sporadic (but has a second cousin affected)	Blood	15q11.2	20303106-20836955	533 850	Loss	*HERC2P2, CYFIP1, NIPA2, NIPA1, TUBGCP5, WHDC1L1, GOLGA9P*	Paternal	*CYFIP1, NIPA1*	qPCR, Agilent 1M, Illumina 1M-duo for sibs	2 non-ASD brothers, no CNV	Autism (based on ADI-R and ADOS), low-functioning (unable to complete IQ measure), language delay, no epilepsy, no dysmorphic features	Father unaffected
37	5453_4	M	(see above)	Blood	22q12.1	25718110-25730447	12 338	Gain	—	Maternal	not in gene	qPCR	1 non-ASD brother, no CNV (another non-ASD brother, no CNV data)	(see above)	Mother unaffected
38	5225_3	M	Sporadic	Blood	15q13.1-q13.2	26887815-28157206	1 269 392	Gain	*TJP1, KIAA0574, NDNL2, APBA2*	Paternal	*APBA2*	qPCR, Agilent 1M, Illumina 1M-duo for sibs	present in 1 brother (5225_4, under phenotype assessement) and 1 sister (5225_5, mild cerebral palsy)	Autism (based on ADI-R and ADOS)	Mother has cerebral palsy
39	5114_3	M	Familial	CL	15q15.3	42350178-42386578	36 401	Loss	*CASC4* exonic	Maternal	*CASC4*	qPCR	1 brother with autism, no CNV	Autism (based on ADI-R and ADOS), low average IQ, no language delay, no epilepsy, no dysmorphic features	NA
40	5258_3	M	Sporadic	Blood	16p13.11	15387380-16270740	883 360	Gain	*MYH11, KIAA0430, C16orf63, MPV17L, NOMO3, NDE1, C16orf45, ABCC6, ABCC1*	Paternal	*NDE1*	qPCR, Illumina 1M-duo for sib	1 non-ASD paternal half-sister, no CNV	Autism (based on ADI-R and ADOS), average IQ, language delay, hx of torticollis and hydrocele, no epilepsy, brain CT normal; 16p13.11 microduplication	Father unaffected

#	AGP ID	Gender	ASD family type	Tissue[a]	Cytoband	Start-End[b]	Size (bp)	CNV	RefSeq Genes	Inheritance	Assay location/ Gene	Method	Segregation in sibs	Phenotype	Parental phenotype
41	5208_3	M	Familial	Blood	16p13.2	8636625-8714816	78 192	Gain	*ABAT* exonic, *C16orf68* exonic	Paternal	*ABAT*	qPCR, Illumina 1M-duo for sib	1 brother with autism, no CNV	ASD (based on ADI-R and ADOS), above average nonverbal IQ, low average language, no epilepsy; mother treated with valproic acid for epilepsy during pregnancy, fetal tachycardia during labor, forceps delivery at 33 1/2 wks, poor respiratory effort, on ventilation; CT scan: mild cortical atrophy, normal EEGs; neurological exam: tight hamstrings, toe walking; dysmorphic features: thin upper lip, small chin, flattened naso-labial fold, flattened occiput, small hands/fingers, lingual frenulum	Father unaffected, mild academic difficulties
42	5208_3	M	Familial	Blood	16q23.1	74997300-75007489	10 190	Loss	*CNTNAP4* intronic	Maternal	*CNTNAP4*	qPCR, Illumina 1M-duo for sib	1 brother with autism, no CNV	(see above)	Mother unaffected
43	5108_3	M	Familial	CL	19q13.32	52318905-52334620	15 716	Gain	*SAE1* exonic	Paternal	*SAE1*	qPCR	(1 brother with autism, not tested)	ASD (based on ADI-R and ADOS), average IQ, no language delay, no epilepsy, normal physical and neurological exams	Father unaffected
44	5228_3	F	Sporadic	CL	19q13.42	59844904-59870590	25 687	Gain	*LILRB4* exonic	Maternal	*LILRB4, LIR5*	qPCR	(1 non-ASD brother, SLI, hyperlexic, not tested)	ASD (based on ADI-R and ADOS), average non-verbal IQ (42%ile), below average language (3%ile), no epilepsy, no dysmorphic features	Mother unaffected
45	5021_5	M	Familial	CL	20p12.1	14999717-15091806	92 090	Loss	*MACROD2* intronic	Paternal	*MACROD2*	qPCR	(1 non-ASD brother and 1 sister with autism, not tested)	Autism (based on ADI-R and ADOS), average non-verbal IQ (55%ile), below average language (<1%ile), no epilepsy, no dysmorphic features	Father unaffected
46	5061_3	M	Familial	Blood	20p12.1	14820313-14879494	59 182	Loss	*MACROD2* intronic	Maternal	*MACROD2*	qPCR, Illumina 1M-duo for sibs	1 brother with ASD and 1 with autism, no CNV in either	ASD (based on ADI-R and ADOS), below average IQ (<1%ile), no epilepsy, premature (28 wks), brain dysfunction, soft neurological signs, amblyopia, hair whorls, 5th finger clinodactyly, long 3rd toe, toe syndactyly	Mother unaffected
47	5066_4	M	Familial	Blood	20p12.1	14818398-14879494	61 097	Loss	*MACROD2* intronic	Maternal	*MACROD2*	qPCR	(1 brother with autism and 1 non-ASD brother, not tested)	Autism (based on ADI-R and ADOS), average IQ, language delay, no epilepsy, malformed ears	Mother unaffected
48	5244_3	M	Sporadic	Blood	20p12.1	14778453-14888687	110 235	Loss	*MACROD2* intronic	Paternal	*MACROD2*	qPCR, Illumina 1M-duo for sibs	present in 1 non-ASD brother; 1 non-ASD sister, no CNV	Autism (based on ADI-R and ADOS), above average non-verbal IQ (95%ile), average language (55%ile), apraxia, possible seizures, microcephaly	Father unaffected; mother has epilepsy
49	5147_9	M	Familial	CL	21q21.1	18979790-19000295	20 505	Gain	no gene, BC028044 transcript	Maternal	BC028044	qPCR, Agilent 1M	(2 non-ASD sisters and 1 maternal first-degree male cousin with ASD, not tested)	Autism (based on ADI-R and ADOS), low average IQ, language delay, no epilepsy, no dysmorphic features	Mother unaffected
50	5261_4	F	Familial	Blood	22q11.21	17257787-19795780	2 537 993	Gain	34 genes	Paternal	*SNAP29, TBX1*	qPCR, Illumina 1M-duo for sibs	1 maternal half-brother with autism and 1 non-ASD sister; no CNV in either	Autism (based on ADI and ADOS), above average non-verbal IQ (96%ile), below average language (4%ile), no epilepsy, no dysmorphic features; 22q11 duplication syndrome	Father unaffected
51	5378_3	M	Familial	Blood	22q11.21	19063495-19358946	295 452	Gain	*MED15, SCARF2, ZNF74, KLHL22*	Paternal	intergenic region close to *MED15*	qPCR	present in 1 brother with ASD	Autism (based on ADI-R and ADOS), unable to complete IQ measure, below average language (1%ile), no epilepsy, sleep problems. Brother: ASD (based on ADI-R and ADOS), normal IQ, sleep problems	Father unaffected
52	5111_3	M	Familial	Blood	Xp22.11	22844170-22897714	53 544	Loss	upstream of *DDX53*	Maternal	upstream of *DDX53*	qPCR, Illumina 1M-duo for sibs	1 brother with autism, no CNV; present in 1 non-ASD sister	Autism (based on ADI-R and ADOS), below average nonverbal IQ (<1%ile), language delay, non-verbal at age 8, no epilepsy, physical examination normal except for some mild dysmorphic facial features and a large head circumference; brain MRI normal	Mother unaffected, some academic difficulties noted
53	5298_3	M	Sporadic (2nd degree cousin Asperger)	Blood	Xp22.11	22892380-23013494	121 114	Loss	*DDX53* exonic, upstream of *PTCHD1*	Maternal	*DDX53*	qPCR, Agilent 1M and Illumina 1M-duo for sibs	present in 1 non-ASD sister (5298_4)	Autism (based on ADI-R and ADOS), moderate MR; severe language impairment, speech and oral motor deficit; possible history of seizures; pes planus and genu varus bilaterally, no other dysmorphic features; normal brain CT scan	Mother unaffected

#	AGP ID	Gender	ASD family type	Tissue[a]	Cytoband	Start-End[b]	Size (bp)	CNV	RefSeq Genes	Inheritance	Assay location/ Gene	Method	Segregation in sibs	Phenotype	Parental phenotype
54	5240_4	M	Familial	Blood	Xp22.11	23116188-23280628	164 441		*PTCHD1* exonic	Maternal	*PTCHD1*	qPCR, Illumina 1M-duo for sibs	present in 1 DZ twin brother with autism (5240_5) and 1 non-ASD sister (5240_3)	Autism (based on ADI and ADOS), average non-verbal IQ (42%ile), low average language (18%ile), no epilepsy, sleep problems, no dysmorphic features. DZ twin brother: Autism (based on ADI and ADOS), average IQ	Mother unaffected, some academic difficulties noted
55	5036_4	M	Familial	CL	Xp21.2	29446046-29557942	111 897	Loss	*IL1RAPL1* intronic	Maternal	*IL1RAPL1*	qPCR, Agilent 1M	present in 1 sister with Asperger and 1 non-ASD DZ triplet brother; absent in 1 non-ASD DZ triplet brother and 1 non-ASD sister	Autism (based on ADI and ADOS), average non-verbal IQ (32%ile), below average language (1%ile), no epilepsy, no dysmorphic features; triplet with 2 non-ASD brothers	Mother unaffected
56	5126_4	M	Familial	CL	Xp21.3-Xp21.2	28931559-29478966	547 408	Gain	*IL1RAPL1* exonic	Maternal	*IL1RAPL1*	qPCR, Agilent 1M	(1 brother with ASD, average IQ, epilepsy, abnormal brain MRI, no DNA) 4 unaffected sisters: CNV present in 2, absent in 1, no DNA for the other 2	Autism (based on ADI-R and ADOS), anxiety disorder, low average IQ (5%ile), delayed language development, sat at 6 m and walked at 12 m, health history uneventful except for asthma as an infant. When last evaluated at age 10 y, no obvious muscular difficulties were noted	Mother unaffected
57	5126_4	M	Familial	CL	Xp21.1	32948977-33330592	381 615	Gain	*DMD* exonic	Maternal	*DMD*	Agilent 1M	(see above)	(see above)	(see above)
58	5241_3	M	Familial	Blood	Xp21.1	31793278-31822704	29 427	Loss	*DMD* exonic; exon 48	Maternal	*DMD*	qPCR	1 brother with ASD, no CNV	Autism (based on ADI-R and ADOS), low non-verbal IQ (9%ile), average language (47%ile), no language delay, no epilepsy, right 2-3 toe syndactyly, right ear malformation	Mother anxiety disorder, OCD, history of seizures, mild attention difficulties
59	5007_3	M	Familial	Blood	Xp11.3	46255974-46292959	36 986	Gain	*LOC401588* whole, *ZNF674* exonic	Maternal	*ZNF674*	qPCR, Illumina 1M-duo for sibs	1 brother with autism and 1 non-ASD brother, no CNV in either	Autism (based on ADI-R and ADOS), anxiety, below average IQ, language delay, no epilepsy, no dysmorphic features	Mother unaffected
60	5419_3	M	Sporadic	CL	Xp11.4	41441499-41478503	37 005	Gain	*CASK* intronic, *GPR82* whole	Maternal	*CASK*	qPCR	(no sibs)	Autism (based on ADI-R and ADOS), no epilepsy; long, wide palpebral fissures, wide mouth, protruding ears, no other dysmorphic features	NA
61	5209_3	M	Familial	CL	Xq21.33	95936387-95949943	13 557	Gain	*DIAPH2* intronic	Maternal	*DIAPH2*	qPCR	present in 1 brother with autism (5209_4); absent in 1 non-ASD sister	Autism (based on ADI-R and ADOS), average nonverbal IQ (50%ile), below average language (<1%ile), no epilepsy, no dysmorphic features	Mother unaffected
62	5468_3	M	Sporadic	Blood	Xq22.3	107765808-107890152	124 345	Gain	*COL4A5* exonic, *IRS4* whole	Maternal	*IRS4, COL4A5*	qPCR, Illumina 1M-duo for sib	present in 1 non-ASD sister	Autism (based on ADI-R and ADOS), average nonverbal IQ (27%ile), below average language (<1%ile), no epilepsy, no dysmorphic features	Mother unaffected
63	5286_3	M	Sporadic	Blood	Xq28	148493841-148543935	50 095	Loss	*HSFX1* intronic, *TMEM185A* exonic	Maternal	*HSFX1, TMEM185A*	qPCR, Illumina 1M-duo for sibs	present in 1 non-ASD brother and absent in 1 non-ASD maternal half-sister	Autism (based on ADI-R and ADOS), below average non-verbal IQ (4%ile), below average language (1%ile), no epilepsy, no dysmorphic features	NA
64	6180_4	M	Sporadic	Blood	1p31.1	75508955-75788048	279 094	Loss	*SLC44A5* exonic	Maternal	*SLC44A5*	qPCR	(1 unaffected sister, no DNA)	Autism, severe MR, non verbal, normal physical exam, no epilepsy	Mother Hashimoto's thyroiditis
65	6340_3	M	Sporadic	U (blood or CL)	1q25.1	172142024-172310899	168 876	Gain	*RC3H1* whole, *SERPINC1*	Paternal	*RC3H1*	qPCR	(3 sibs, no DNA)	Autism, mild MR, no language delay, chronic otitis, normal physical exam, no epilepsy, brain MRI: Arnold Chiari type I	Father unaffected
66	6340_3	M	Sporadic	U (blood or CL)	22q11.21	20247190-20277644	30 455	Loss	*UBE2L3* exonic	Both parents	*UBE2L3*	qPCR	(see above)	(see above)	Father unaffected, mother bulimia
67	6125_4	M	Sporadic	U (blood or CL)	2p16.3	50822312-50886363	64 052	Loss	*NRXN1* intronic	Maternal	*NRXN1*	qPCR	(1 healthy brother, no DNA)	Autism, moderate MR, hyperactivity, language delay, limited language, inguinal hernia, retrognathia, narrow palate, anteverted nares, 3rd finger clinodactyly. Two tonic-clonic seizures during adolescence, treated with Depakine; normal brain CT	Mother unaffected

#	AGP ID	Gender	ASD family type	Tissue[a]	Cytoband	Start-End[b]	Size (bp)	CNV	RefSeq Genes	Inheritance	Assay location/ Gene	Method	Segregation in sibs	Phenotype	Parental phenotype
68	6125_4	M	Sporadic	U (blood or CL)	22q11.21	19051464-19793730	742 267	Gain	17 genes	Paternal	*ZNF74, KLHL22, MED15, SNAP29*	MLPA	(see above)	(see above)	Father unaffected
69	6197_4	F	UKN (Asperger suspected in brother)	Blood	2p16.3	50822312-50900862	78 551	Gain	*NRXN1* intronic	Maternal	*NRXN1*	qPCR	(1 healthy brother, no DNA)	Autism, mild MR, phrase speech delay, irregularly implanted teeth, thin upper lip, physical exam otherwise normal, no epilepsy; normal brain CT	Mother unaffected
70	6203_4	M	Sporadic	Blood	3q11.2	94994003-95230688	236 686	Gain	*STX19* exonic, *ARL13B* exonic, *PROS1* whole	Maternal	*PROS1*	qPCR	absent in 1 unaffected brother	Autism, severe MR, language delay, normal physical exam, no epilepsy	Mother unaffected
71	6356_5	M	Sporadic	Blood	4q31.1	140478322-140522239	43 918	Loss	*NARG1* exonic	Paternal	*NARG1*	qPCR	(2 healthy sibs, no DNA)	Autism, moderate MR, no language delay, normal physical exam, no epilepsy	Father unaffected
72	6356_5	M	Sporadic	Blood	16q24.1	82800294-82887663	87 370	Loss	*WFDC1* exonic, *KCNG4* whole	Maternal	*KCNG4*	qPCR	(see above)	(see above)	Mother unaffected
73	6033_3	M	Familial	Blood	5p14.3	19532212-20357961	825 750	Loss	*CDH18* exonic	Maternal	*CDH18*	qPCR	absent in 1 sister with autism	High-functioning autism, IQ above normal, early speech development, neonatal hypothyroidism, normal physical exam, no epilepsy	Mother unaffected
74	6127_3	F	Sporadic	Blood	7p15.1	29175877-29503232	327 356	Gain	*CHN2* exonic	Paternal	*CHN2*	qPCR	(1 healthy sister, no DNA)	Autism, severe MR, language delay, normal physical exam, no epilepsy	Father unaffected
75	6266_3	F	Sporadic	Blood	7q35	145853798-145885593	31 796	Gain	*CNTNAP2* intronic	Maternal	*CNTNAP2*	qPCR	present in 1 brother with ADHD, absent in 1 healthy sister	Autism, ADHD, sleep disorder, mild MR, language delay, normal physical exam, no epilepsy	Mother unaffected
76	6320_4	M	Sporadic	Blood	7q35	145853798-145885593	31 796	Gain	*CNTNAP2* intronic	Maternal	*CNTNAP2*	qPCR	(3 healthy sibs, no DNA)	Autism, low normal IQ, language delay, normal physical exam, no epilepsy, normal brain MRI	Mother unaffected
77	6240_4	M	Sporadic	Blood	8q22.3	104451635-104607137	155 503	Loss	*WDSOF1, CTHRC1, RIMS2, SLC25A32*	Paternal	*RIMS2*	qPCR	(1 healthy sister, not tested)	Autism, mild MR, language delay, macrocephaly, automutilations, minor dysmorphic features (low ears, flat nose and nasal bridge, high forehead), normal neurological exam, no epilepsy, brain MRI: white matter abnormalities. This patient was also found to carry a 5.4 Mb *de novo* 11q deletion (Jacobsen syndrome)	Father unaffected
78	6191_3	M	Sporadic	U (blood or CL)	11q22.1	98412813-98476682	63 870	Gain	*CNTN5* intronic	Maternal	*CNTN5*	qPCR	absent in 1 unaffected sister (no DNA available for 1 brother with MR)	Autism, mild MR, no language delay, macrocephaly, no dysmorphic features, no epilepsy	Mother unaffected
79	6362_3	M	Sporadic	Blood	14q31.1	78536692-78578318	41 627	Loss	*NRXN3* intronic	Paternal	*NRXN3*	qPCR	(1 healthy paternal half-sister, no DNA)	Autism, moderate MR, language delay, normal physical exam, no epilepsy, normal brain MRI	Father dysthymia
80	6319_3	M	Sporadic	Blood	15q11.2	20305097-20773130	468 034	Loss	*NIPA2, NIPA1, CYFIP1,*	Paternal	*CYFIP1, TUBGCP5*	MLPA	(1 healthy brother, no DNA)	PDD-NOS, mild MR, language delay, functional language, no epilepsy, hypermetropia, no dysmorphic features, flat feet,	Father unaffected
81	6261_3	M	Sporadic	Blood	16p12.1	21854731-22331199	476 469	Loss	*EEF2K, C16orf65, CDR2, POLR3E, C16orf52, UQCRC2, VWA3A*	Maternal	*POLR3E, VWA3A*	qPCR	(2 healthy sibs, no DNA)	Autism, mild MR, language delay; hearing deficit detected at 1 y, treated with hearing aid, audiogram at 4 yrs normal; normal physical exam, no epilepsy, normal brain CT; 16p12.1 microdeletion	Mother unaffected
82	6261_3	M	Sporadic	Blood	16p12.1	61650712-61824330	173 619	Loss	*PRKCA* exonic, *APOH* exonic	Paternal	*APOH*	qPCR	(see above)	(see above)	
83	6164_3	F	Sporadic	U (blood or CL)	17q25.3	77760278-77802222	41 945	Gain	*CCDC57 exonic, SLC16A3* whole, *CSNK1D* exonic	Maternal	*CSNK1D, SLC64A*	qPCR	(1 healthy sister, no DNA)	Autism, mild MR, language delay, functional language, sleep difficulties, normal physical exam, no epilepsy, normal brain MRI	Mother unaffected, breast cancer

#	AGP ID	Gender	ASD family type	Tissue[a]	Cytoband	Start-End[b]	Size (bp)	CNV	RefSeq Genes	Inheritance	Assay location/ Gene	Method	Segregation in sibs	Phenotype	Parental phenotype
84	6164_3	F	Sporadic	U (blood or CL)	Xq28	153248612-153264501	15 890	Gain	*FLNA* exonic, *EMD* whole	Maternal	*EMD*	qPCR	(see above)	(see above)	(see above)
85	6023_3	M	Familial	U (blood or CL)	20p12.1	14948785-15161293	212 509	Loss	*MACROD2* exonic	Maternal	*MACROD2*	qPCR	present in 1 brother with PDD-NOS	High-functioning autism, normal IQ, language delay, normal physical exam, no epilepsy	Mother unaffected
86	6057_3	M	Familial	Blood	22q11.21	19221842-19341132	119 291	Gain	*MED15*	Paternal	*MED15*	qPCR, MLPA	present in 1 brother with autism	Autism, MR, no phrases, normal physical exam, no epilepsy; brain MRI: white matter hyperintensities. Younger brother: autism, MR, no language, normal physical exam, no epilepsy; brain MRI: delayed myelinization	Father unaffected
87	6046_4	M	Familial	Blood	22q13.33	49452422-49567307	114 886	Gain	*SHANK3*, *ACR*, *RABL2B*, *MGC70863*	Both parents	*SHANK3, ACR, RABL2B*	qPCR, MLPA	present in 1 brother with autism	Autism, severe MR, no language, normal physical exam, no epilepsy, normal brain MRI. Younger brother: autism, severe MR, no language, normal physical exam	Both parents unaffected
88	6239_3	M	Sporadic	U (blood or CL)	Xp22.2	11324498-11418698	94 201	Gain	*ARHGAP6* intronic	Maternal	*ARHGAP6*	qPCR	(no sibs)	Autism, severe MR, no language, strabism, no dysmorphic features, no epilepsy	Mother unaffected
89	6239_3	M	Sporadic	U (blood or CL)	Xq27.1	138549326-138716662	167 337	Gain	*MCF2* exonic, *ATP11C* exonic	Maternal	*ATP11C*	qPCR	(no sibs)	(see above)	Mother unaffected
90	6323_3	M	Sporadic	U (blood or CL)	Xp22.2	14603137-14746120	143	Loss	*GLRA2* exonic	Maternal	*GLRA2*	qPCR	(no sibs)	Autism, normal IQ, language delay, bilateral myopia, normal physical exam, no epilepsy	Mother unaffected
91	6032_4	F	Familial	U (blood or CL)	Xp21.3	28732962-28751750	18 789	Loss	*IL1RAPL1* intronic	Paternal	*IL1RAPL1*	qPCR	absent in 1 maternal half-brother with autism, present in 1 healthy sister	High-functioning autism, normal IQ, language delay, normal physical exam, no epilepsy	Father unaffected
92	6379_4	M	Sporadic	U (blood or CL)	Xq21.1	76915420-77030430	115 011	Gain	*MAGT1* exonic	Maternal	*MAGT1*	qPCR	(no sibs)	Autism, mild MR, language delay, normal physical exam, no epilepsy	Mother OCD
93	13135_1523	F	UKN	Blood	1q21.1	144838594-146308287	1 469 694	Gain	14 genes	Maternal	*BCL9*	qPCR	(1 brother with Aspergers and dyspraxia, not tested)	Autism, low average IQ, normal physical exam, no epilepsy; 1q21.1 microduplication syndrome	Mother unaffected
94	13133_1503	M	Sporadic	Blood	1q42.2	229977634-230015444	37 810	Gain	*DISC1* exonic	Maternal	*DISC1*	qPCR	(1 healthy sister, no DNA)	Autism, moderate-severe MR, non-verbal, normal physical exam, no epilepsy	NA
95	13043_523	M	Sporadic	Blood	15q11.2	21914925-22008680	93 755	Gain	*PWRN2* whole	Both parents	*PWRN2*	qPCR	(1 healthy sister and 2 healthy DZ twin brothers, no DNA)	Autism, normal IQ, fluent speech, no epilepsy, narrow face, elongated ears, physical exam otherwise normal	Father may have Asperger syndrome (not evaluated)
96	13047_563	M	Familial	Blood	Xp22.11	22829183-23214712	385 529	Loss	*DDX53* exonic, upstream of *PTCHD1*	Maternal	*DDX53*	qPCR	(2 brothers with autism and 1 sister with pragmatic language disorder, no DNA)	ASD, normal IQ, no dysmorphic features, history of allergies and recurrent ear infections	Mother unaffected
97	3017_3	M	Familial	Blood	2p12	76782603-76809810	27 207	Loss	downstream of *LRRTM4*	Paternal	downstream of *LRRTM4*	LR-PCR	absent in 1 affected brother	Autism, moderate MR	NA
98	3160_3	M	Familial	Blood	2p12	76782603-76809810	27 207	Loss	downstream of *LRRTM4*	Paternal	downstream of *LRRTM4*	LR-PCR	present in 1 affected sister	Autism, normal IQ, no language delay	Father unaffected
99	3092_3	F	Familial	Blood	2p12	76782603-76809810	27 207	Loss	downstream of *LRRTM4*	Maternal	downstream of *LRRTM4*	LR-PCR	absent in 1 affected brother	Autism, normal IQ	Mother unaffected
100	3106_3	M	Familial	Blood	2p12	76782603-76809810	27 207	Loss	downstream of *LRRTM4*	Paternal	downstream of *LRRTM4*	LR-PCR	absent in 1 unaffected brother and 1 affected sister	Autism, normal IQ	NA
101	3424_3	M	Sporadic	Blood	2q23.1	148881443-149078468	197 025	Gain	*MBD5* whole	Maternal	*MBD5*	qPCR	absent in 1 unaffected brother	Autism, mild MR, relative macrocephaly, no dysmorphic features	Mother unaffected
102	3424_3	M	Sporadic	Blood	Xp22.11	23013250-23116188	102 939	Loss	upstream of *PTCHD1*	Maternal	upstream of *PTCHD1*	qPCR	absent in 1 unaffected brother	(see above)	(see above)

#	AGP ID	Gender	ASD family type	Tissue[a]	Cytoband	Start-End[b]	Size (bp)	CNV	RefSeq Genes	Inheritance	Assay location/ Gene	Method	Segregation in sibs	Phenotype	Parental phenotype
103	3423_003	M	Sporadic	Blood	2q31.1	170311824-170375059	63 236	Loss	*KLHL23, SSB*	Maternal	*KLHL23, SSB*	qPCR	(no sibs)	Autism, moderate MR, no dysmorphic signs, weight >P97, height P90, head circumference P98	Mother unaffected
104	3211_3	M	Familial	Blood	3p26.2	4069293-4236304	167 011	Loss	*SUMF1* exonic	Paternal	*SUMF1*	LR-PCR	absent in 1 affected brother	Autism, normal IQ, no language delay, height P97, head circumference P50, epicanthus, steepled palate, physical and neurological exam otherwise normal	NA
105	3211_3	M	Familial	BBC	16p11.2	29502984-30127026	624 043	Gain	37 genes	Maternal	*IMAA*	qPCR	absent in 1 affected brother	(see above); 16p11.2 microduplication syndrome	NA
106	3319_3	F	Familial	BBC	3p26.2	4063576-4076356	12 780	Loss	— (*SUMF1* intron according to UCSC/Ensembl but not Refseq)	Maternal	*SUMF1* (UCSC/Ensembl but not Refseq gene)	LR-PCR	present in 1 of 2 unaffected sisters but not in affected sister	Autism, MR, no functional language, no epilepsy	NA
107	3072_008	M	Familial	Blood	4q35.2	189540090-190704726	1 164 636	Loss	None	Paternal	11 qPCR assays across region	qPCR	(one affected cousin in maternal side of family, not tested)	Autism, normal IQ, language delay, no epilepsy, height P97, head circumference P98, mild eczema, no other associated medical or psychiatric problems	NA
108	3072_008	M	Familial	Blood	4q35.2	190922297-190993476	71 179	Loss	None	Paternal	11 qPCR assays across	qPCR	(see above)	(see above)	NA
109	3081_5	F	Familial	BBC	7q31.1	110835815-110867477	31 663	Loss	*IMMPL2* intronic	Maternal	*IMMPL2*	LR-PCR	absent in 1 affected brother and 1 unaffected sister	Autism, moderate to severe MR, no dysmorphic features, head circumference +2 SD, self-injurious behavior, no epilepsy	NA
110	3320_003	M	Familial	Blood	7q31.1	110879714-110924988	45 274	Loss	*IMMP2L* exonic	Paternal	7 qPCR assays across region	qPCR	present in the unaffected sister, but absent in the brother with PDD-NOS	Autism, no language delay, normal IQ, 40% hearing in one ear, no other associated medical or psychiatric problems, no epilepsy	Father unaffected
111	3018_3	M	Familial	CL	7q35	146059148-146251008	191 861	Gain	*CNTNAP2* exonic	Maternal	*CNTNAP2*	qPCR	absent in 1 affected brother and 1 unaffected brother, present in 1 unaffected sister	High-functioning autism, delayed language; abnormal EEG without seizures, treated with carbamazepine; born at 38 wks, oxygen deprivation due to cord around neck, 48 h in ICU; eczema, hydronephrosis of right kidney, weight problem, gross motor difficulties, polydactyly on right hand and left foot, head circumference P98	Mother unaffected, depression, psoriasis
112	3307_3	M	Familial	BBC	8p23.1	6316818-6341678	24 861	Loss	*MCPH1* exonic	Paternal	LR-PCR across deletion	LR-PCR	present in 1 unaffected brother, and absent in 2 affected sisters	Autism, average IQ, language delay, head circumference >P98, no dysmorphic features, no seizures	Father unaffected
113	3209_004	M	Familial	Blood	10q21.3	66980652-66983475	2 823	Loss	None	Paternal	8 qPCR assays across region	qPCR	present in sister with autism	ADI and ADOS dx Autism, normal IQ, delay in single words but not in phrase speech, no other associated medical or psychiatric problems, no epilepsy. Sister: ADI and ADOS dx Autism, performance IQ 70, language delay, no other associated medical or psychiatric problems, no epilepsy	Father unaffected
114	3135_4	M	Familial	Blood	10q21.3	67748487-67785209	36 723	Loss	*CTNNA3* intronic	Maternal	*CTNNA3*	LR-PCR	absent in 1 affected brother and 1 socially awkward sister (no ASD diagnosis)	Autism, mild MR, language delay, head circumference P98, no dysmorphic features, no seizures	Mother had two miscarriages. Maternal grandfather (not tested for CNV): partial complex seizures with transient speech arrest with good response to carbamazepine
115	3156_3	M	Familial	Blood	10q21.3	67741619-67788456	46 837	Loss	*CTNNA3* intronic	Paternal	*CTNNA3*	LR-PCR	absent in 1 sister with autism	Autism, mild MR, epilepsy	NA
116	3209_4	M	Familial	Blood	10q21.3	67741619-67788456	46 837	Loss	*CTNNA3* intronic	Paternal	*CTNNA3*	LR-PCR	absent in 1 affected sibling	Autism, normal IQ, no language delay, no dysmorphic features, head circumference +2 SD, no epilepsy	NA

#	AGP ID	Gender	ASD family type	Tissue[a]	Cytoband	Start-End[b]	Size (bp)	CNV	RefSeq Genes	Inheritance	Assay location/ Gene	Method	Segregation in sibs	Phenotype	Parental phenotype
117	3321_3	M	Familial	Blood	10q21.3	67741619-67788456	46 837	Loss	*CTNNA3* intronic	Maternal	*CTNNA3*	LR-PCR	absent in 1 sister with autism	Autism, normal IQ, no epilepsy	NA
118	3311_3	M	Familial	Blood	10q21.3	67688367-67759307	70 941	Loss	*CTNNA3* exonic	Maternal	*CTNNA3*	LR-PCR	present in 1 brother with PDD-NOS, absent in 1 affected brother	Autism, moderate MR, language delay, congenital cataract, amblyopia, seizures in infancy. Brother: PDD-NOS, language delay, amblyopia	NA
119	3034_3	M	Familial	Blood	10q21.3	68099569-68167094	67 525	Loss	*CTNNA3* intronic	Paternal	*CTNNA3*	LR-PCR	absent in 1 affected sibling	Autism, low normal IQ, language delay, recurrent ear infections, no dysmorphic features, no epilepsy	Father type II diabetes
120	3134_4	M	Familial	Blood	10q21.3	67838960-67885161	46 202	Loss	*CTNNA3* intronic	Maternal	*CTNNA3*	LR-PCR	present in 1 affected brother	Autism, moderate MR, phrase speech delay, no dysmorphic features, no seizures, born at 35 wks. Brother: autism, profound MR, language delay (non-verbal at time of evaluation), head circumference P98, no dysmorphic features, no seizures, left extremities decreased muscle function, flat feet, full-term, C-section due to cord around neck	Mother unaffected
121	3093_4	M	Familial	Blood	10q21.3	67987089-68067310	80 222	Loss	*CTNNA3* exonic	Maternal	*CTNNA3*	LR-PCR	absent in 1 brother with autism	Autism, mild MR	Mother unaffected
122	3169_4	M	Familial	Blood	10q21.3	68029140-68183933	154 794	Loss	*CTNNA3* exonic	Maternal	*CTNNA3*	LR-PCR	present in 1 affected brother	Autism, low average IQ	Mother depression, anxiety
123	3174_003	M	Familial	Blood	16p12.1	21854731-22343312	488 582	Loss	*EEF2K, C16orf65, CDR2, POLR3E, C16orf52, UQCRC2, VWA3A*	Paternal	*POLR3E, VWA3A*	qPCR, Agilent 1M	present in 1 brother with autism	ADI and ADOS dx Autism, low average IQ, language delay, gross and fine motor coordination difficulties, no dysmorphic features, no epilepsy; 16p12.1 microdeletion; this patient also carries a de novo 3p24.3 deletion. Brother: ADI and ADOS dx Autism, low average IQ, language delay, some mild dysmorphic facial features, no epilepsy; 16p12.1 microdeletion	Evidence of broader autism phenotype (social and communication domains) in both parents, with father scoring higher
124	3099_8	M	Familial	CL	16q21	60027157-61668976	1 641 819	Loss	CDH8 whole	Maternal	*CDH8,* LR-PCR across deletion	LR-PCR, breakpoint sequencing, Affy 10K	present in 2 other affected brothers but not in 4 unaffected siblings	All 3 affected siblings have ASD and MR (for more details see Vieland et al., submitted) (sequencing showed true deletion coordinates are chr16: 60025584–61667839)	Mother unaffected
125	3435_003	F	Sporadic	Blood	18p11.21	11957792-12120394	162 603	Loss	*IMPA2*	Paternal	*IMPA2*	qPCR	present in an unaffected sister	Autism, mild MR, no dysmorphic signs, weight P90, height P50, head circumference P80	Father unaffected
126	3181_7	M	Familial	CL	19q13.32	52467295-52645983	178 689	Loss	*MEIS3, GPR77, C5AR1, LOC255783, SLC8A2, DHX34*	Maternal	LR-PCR across deletion	nested LR-PCR	no siblings; absent in affected cousin	ADI and ADOS dx Autism, no language delay, higher than average IQ, no dysmorphic features, height and weight 2-9th centiles, no epilepsy	Mother unaffected
127	3127_4	M	Familial	CL	22q11.21	17257787-19793730	2 535 944	Gain	57 genes	Paternal	*HIRA, SNAP29*	qPCR	present in 1 affected brother	ASD, normal IQ, violent outbursts of aggression, severe learning difficulties, no language delay, articulation difficulties, limited and repetitive speech, abnormal EEG but no seizures, fine and gross motor problems, visuo-spatial difficulties, protruding tongue, several raised skin lesions (amelanotic and café au lait) on trunk, forehead and back of his neck (history of café au lait spots in the maternal family)	Father trouble reading and writing, problems with coordination, normal IQ
128	3067_5	M	Familial	CL	22q11.21	19051464-19795780	744 317	Gain	17 genes	Paternal	*SNAP29*	qPCR	4 sibs: CNV present in 1 affected brother; absent in 1 affected male, 1 male with mild PDD, and 1 unaffected sister	Autism, normal IQ, no epilepsy	NA
129	3253_4	M	Familial	BBC	Xp22.11	22860224-22907454	47 231	Loss	upstream of *PTCHD1*	Maternal	upstream of *PTCHD1*	qPCR	absent in 1 brother with autism and 1 unaffected younger sister	Autism, mild MR	NA

#	AGP ID	Gender	ASD family type	Tissue[a]	Cytoband	Start-End[b]	Size (bp)	CNV	RefSeq Genes	Inheritance	Assay location/ Gene	Method	Segregation in sibs	Phenotype	Parental phenotype
130	3019_3	M	Familial	BBC	Xp21.1	32100618-32315937	215 319	Gain	*DMD* exonic	Maternal	*DMD*	qPCR	CNV present in affected brother, absent in healthy sister	Autism, normal IQ, articulation dyspraxia, poor coordination, odd head movements and frequent body contortions. Affected brother (3019_5) has a general developmental delay and as an adult has moderate learning difficulties. The medical notes do not mention any specific motor problems for either one of them.	NA
131	14244_3680	M	Sporadic	Blood	1p33	49688435-49770826	82 392	Loss	*AGBL4* intronic	Maternal	*AGBL4*	qPCR	(no sibs)	ASD, normal IQ, phrase speech delay, no epilepsy, asthma, no dysmorphic features	Mother depression
132	14246_3700	M	Sporadic	Blood	1p33	49688435-49770826	82 392	Loss	*AGBL4* intronic	Maternal	*AGBL4*	qPCR	(no sibs)	ASD, normal IQ, phrase speech delay, no epilepsy, no dysmorphic features	Mother unaffected
133	14080_1360	F	Sporadic	Blood	1p33	49688435-49770826	82 392	Loss	*AGBL4* intronic	Paternal	*AGBL4*	qPCR	(1 healthy sister, 1 healthy brother, no DNA)	Autism, moderate MR, no language, neurodevelopmental delay, no dysmorphic features, no epilepsy, sleep problems; great-uncle had autism	Father unaffected
134	14239_2920	M	Sporadic	Blood	1p33	49685647-49770826	85 180	Loss	*AGBL4* intronic	Maternal	*AGBL4*	qPCR	(no sibs)	Autism, normal IQ, phrase speech delay, no epilepsy, no dysmorphic features	Mother unaffected
135	14301_4220	M	Sporadic	Blood	2q34	214808080-214829997	21 918	Gain	*SPAG16* exonic	Paternal	*SPAG16*	qPCR	(1 healthy brother, no DNA)	Autism, normal IQ, phrase speech delay, no epilepsy, no dysmorphic features, vision problems (unspecific refraction defect)	Father unaffected
136	14309_4260	M	Sporadic	Blood	2q34	214808080-214829997	21 918	Gain	*SPAG16* exonic	Paternal	*SPAG16*	qPCR	(1 healthy brother, no DNA)	Autism, mild MR, phrase speech delay, no epilepsy, neurodevelopmental delay with onset at 2 y, no dysmorphic features	Father unaffected
137	14029_560	M	Sporadic	Blood	2q34	214808080-214829997	21 918	Gain	*SPAG16* exonic	Paternal	*SPAG16*	qPCR	(1 healthy brother, no DNA)	autism, normal IQ, language delay, neurodevelopmental delay, no epilepsy, no dysmorphic features; mother had 3 previous miscarriages; maternal great-aunt had autism	Father unaffected
138	14164_2680	M	Sporadic	Blood	2q34	214808080-214829997	21 918	Gain	*SPAG16* exonic	Maternal	*SPAG16*	qPCR	(1 brother language delay, no DNA)	Autism, normal IQ, phrase speech delay, no epilepsy, no dysmorphic features	Mother unaffected
139	14200_3240	M	Sporadic	Blood	2q34	214803990-214829997	26 008	Gain	*SPAG16* exonic	Paternal	*SPAG16*	qPCR	(1 healthy brother, no DNA)	Autism, normal IQ, phrase speech delay, no epilepsy, no dysmorphic features	Father unaffected
140	14061_1040	M	Sporadic	Blood	5q23.1	118544777-118617384	72 608	Gain	*DMXL1* exonic	Maternal	*DMXL1*	qPCR	(no sibs)	Autism, mild MR, phrase speech delay, no epilepsy, delivery complications, neurodevelopmental delay, no dysmorphic features	Mother unaffected
141	14072_1250	M	Sporadic	Blood	6q24.3	145841703-145852542	10 840	Loss	—	Paternal	not in gene	qPCR	(1 healthy brother, no DNA)	Autism, normal IQ, no language delay, no epilepsy, sleep problems, no dysmorphic features	Father unaffected
142	14230_3640	M	Sporadic	Blood	6q24.3	145841703-145857642	15 940	Loss	—	Maternal	not in gene	qPCR	(2 healthy half-sisters, no DNA)	ASD, mild MR, no epilepsy, no dysmorphic features	Mother unaffected
143	14171_2770	F	Sporadic	Blood	6q24.3	145841703-145852542	10 840	Loss	—	Paternal	not in gene	qPCR	(1 healthy sister, no DNA)	Autism, moderate MR, language delay, altered EEG, developmental delay, no dysmorphic features, hypotonia, myopia; brain MRI: cortico-subcortical atrophy	Father schizoaffective disorder
144	14111_2000	M	Sporadic	Blood	9q33.2	124209751-124398737	188 987	Gain	*OR1L8,OR1N2, OR1N1,OR1J4,OR1J2,OR1J1*	Paternal	*OR1J1*	qPCR	(1 healthy sister, no DNA)	Autism, normal IQ; mild asphyxia during delivery, did not require reanimation; no language delay, neurodevelopmental delay at 2 y, no epilepsy, no dysmorphic features	Father unaffected
145	14184_3020	M	Sporadic	Blood	9q33.2	124209751-124398737	188 987	Gain	*OR1L8,OR1N2, OR1N1,OR1J4,OR1J2,OR1J1*	Maternal	*OR1J1*	qPCR	(1 healthy brother, no DNA)	ASD, normal IQ, no language delay, no epilepsy, no dysmorphic features	Mother unaffected
146	14167_2720	M	Sporadic	Blood	15q13.2,15q13.3	28705540-30436163	1 730 623	loss	*KLF13, OTUD7A, MTMR15, MTMR10, CHRNA7, ARHGAP11B, TRPM1*	Paternal	not in gene	qPCR	(no sibs)	Autism, mild MR, phrase speech delay, no epilepsy, no dysmorphic features; 15q13.3 microdeletion syndrome	Father unaffected

#	AGP ID	Gender	ASD family type	Tissue[a]	Cytoband	Start-End[b]	Size (bp)	CNV	RefSeq Genes	Inheritance	Assay location/ Gene	Method	Segregation in sibs	Phenotype	Parental phenotype
147	14071_1240	M	Sporadic	Blood	16p12.1	21788416-22351124	562 709	Loss	*EEF2K, C16orf65, CDR2, POLR3E, C16orf52, UQCRC2, VWA3A*	Paternal	*POLR3E, VWA3A*	qPCR	(not tested)	Autism, mild MR, phrase speech delay, neurodevelopmental delay with onset at 2 y, no dysmorphic features, no epilepsy, sleep problems; 16p12.1 microdeletion	Father unaffected
148	14262_3850	M	Sporadic	Blood	18q12.2	35205049-35299318	94 270	Loss	*LOC647946* (non-coding RNA)	Paternal	not in gene	qPCR	(no sibs)	ASD, normal IQ, no language delay, no epilepsy, neurodevelopmental delay at 2 y, no dysmorphic features	Father unaffected
149	14049_850	M	Sporadic	Blood	18q12.2	35205049-35299318	94 270	Loss	*LOC647946* (non-coding RNA)	Maternal	not in gene	qPCR	(1 healthy sister, 1 healthy brother, no DNA)	Autism, mild MR, phrase speech delay, no epilepsy, gastrointestinal problems, growth delay, neurodevelopmental delay at 2 y, no dysmorphic features	Mother unaffected
150	14144_2420	M	Sporadic	Blood	Xq28	154569169-154582606	13 438	Loss	—	Maternal	not in gene	qPCR	(1 healthy sister, no DNA)	Autism, normal IQ, no language delay, no epilepsy, neurodevelopmental delay at 2 y, no dysmorphic features	Mother unaffected
151	14186_3050	M	Sporadic	Blood	Xq28	1545691691 54582606-	13 438	Loss	—	Maternal	not in gene	qPCR	(1 healthy brother, no DNA)	Autism, normal IQ, no language delay, no epilepsy, neurodevelopment delay, no dysmorphic features	Mother unaffected
152	14198_3220	M	Familial	Blood	Xq28	154569169-154582606	13 438	Loss	—	Maternal	not in gene	qPCR	(no sibs)	ASD, normal IQ, phrase speech delay, no epilepsy, no dysmorphic features, myopia	Mother depression
153	2232_1	M	Sporadic	Blood	1p36.33	1370430-1429557	59 128	Loss	*ATAD3B* whole, *ATAD3C* whole	Paternal	close to *ATAD3B*	qPCR	(not tested)	Autism, mild MR, language delay, overall nondysmorphic appearance, macrocephalic, wide nasal bridge, partial transverse palmar crease on left hand, leaky gut, numerous food allergies, wears glasses for mild hypermetropia, no epilepsy, normal brain MRI	NA
154	2232_1	M	Sporadic	Blood	3p11.1	89485137-89499861	14 724	Loss	*EPHA3* intronic	Both parents	*EPHA3*	qPCR, Affymetrix 6.0 SNP array	(not tested)	(see above)	NA
155	2230_1	M	Sporadic	Blood	1q24.2	167493526-167507362	13 837	Loss	*NME7* intronic	Maternal	*NME7*	qPCR	(not tested)	Autism, moderate MR, language delay, no dysmorphic features, fifth finger clinodactyly, obesity, urethral obstruction surgically repaired, bilateral inguinal hernias, chronic ear infections, allergies, no epilepsy, normal brain CT scan	NA
156	2230_1	M	Sporadic	Blood	5q11.2	54842833-54966863	124 031	Gain	*SLC38A9* exonic, *PPAP2A* exonic	Paternal	*SLC38A9*	qPCR, Affy 6.0	(not tested)	(see above)	NA
157	2230_1	M	Sporadic	Blood	14q31.1	78557401-78589854	32 454	Loss	*NRXN3* intronic	Paternal	*NRXN3*	qPCR, Affy 6.0	(not tested)	(see above)	NA
158	2230_1	M	Sporadic	Blood	22q11.1	15641026-15653178	12 153	Loss	*XKR3* exonic	Paternal	*XKR3*	qPCR, Affy 6.0	(not tested)	(see above)	NA
159	2241_1	M	Sporadic	Blood	5p14.2	24545450-24623946	78 497	Loss	*CDH10* exonic	Maternal	*CDH10*	qPCR	(not tested)	Autism, mild MR, no language delay, no dysmorphic features, growth delay, recurrent ear infections, hay fever, no epilepsy, normal brain CT scan	NA
160	2234_1	M	Sporadic	Blood	8p23.2	3773951-3780096	6 146	Loss	*CSMD1* intronic	Maternal	*CSMD1*	qPCR	(not tested)	Autism, language delay, no dysmorphic features, significant hypotonia but walked at 18 m, normal neurological exam, no epilepsy	NA
161	2234_1	M	Sporadic	Blood	9p24.3	175632-422918	247 286	Gain	*DOCK8* exonic, *C9orf66* whole	Maternal	*DOCK8*	qPCR, Affy 6.0	(not tested)	(see above)	NA
162	2234_1	M	Sporadic	Blood	12q12	38587878-38603284	15 407	Loss	*SLC2A13* intronic	Maternal	*SLC2A13*	qPCR, Affy 6.0	(not tested)	(see above)	NA

#	AGP ID	Gender	ASD family type	Tissue[a]	Cytoband	Start-End[b]	Size (bp)	CNV	RefSeq Genes	Inheritance	Assay location/ Gene	Method	Segregation in sibs	Phenotype	Parental phenotype
163	2236_1	M	Sporadic	Blood	8q13.3	72378104-72380244	2 141	Loss	*EYA1* intronic	Maternal	*EYA1*	qPCR	(not tested)	Autism, normal IQ, language delay, no dysmorphic features, mild ligamentous laxity, 4th toe clinodactyly, Mongolian spot around sacral area, slightly hypotonic, normal neurological exam, no epilepsy, recurrent ear infections, reactive airways disease when younger	NA
164	2236_1	M	Sporadic	Blood	15q14	32459510-32606592	147 083	Loss	*GOLGA8A* exonic,	Maternal	*GOLGA8*	qPCR, Affy 6.0	(not tested)	(see above)	NA
165	2238_1	M	Sporadic	Blood	9p24.3	600460-694075	93 615	Loss	*KANK1* exonic	Paternal	*KANK1*	qPCR, Affy 6.0	(not tested)	Autism, MR, language delay, limited language, no dysmorphic features, constipation, hay fever, no epilepsy	NA
166	1957_303	F	Familial	CL	9p24.3	243594-362341	118 748	Gain	*DOCK8* exonic	Maternal	SNP array	Illumina 550K (Glessner et al. 2009)	(not tested)	Autism by ADI-R and ADOS, normal IQ, verbal (AGRE ID: AU1453303)	NA
167	1987_301	F	Familial	CL	15q11.2	20090262-21038099	947 838	Gain	*CYFIP1, GOLGA8E, GOLGA9P, HERC2P2, LOC283767, NIPA1, NIPA2, TUBGCP5, WHDC1L1*	Maternal	SNP array	Illumina 550K (Glessner et al. 2009)	(not tested)	Autism by ADI-R and ADOS, normal IQ, verbal, no epilepsy; 15q11.2 BP1-BP2 microduplication (AGRE ID: AU1871301)	NA
168	1950_301	M	Familial	CL	15q13	26762141-30436163	3 674 020	Loss	14 genes	Maternal	SNP array	Illumina 550K (Glessner et al. 2009)	(not tested)	Autism by ADI-R and ASD by ADOS, no IQ available (untestable by Raven), non verbal, poor suck at birth, no epilepsy; 15q13.3 microdeletion syndrome (AGRE ID: AU1024301)	NA
169	1231_3	M	Familial	CL	15q13.2	28187888-28881771	693 884	Loss	*ARHGAP11B, CHRFAM7A, DKFZP434L187*	Paternal	SNP array	Illumina 550K (Glessner et al. 2009)	(not tested)	Autism by ADI-R and ADOS, normal IQ, verbal, delayed suture closure, complex partial seizures (CNV partially overlaps the 15q13.3 microdeletion syndrome region) (AGRE ID: AU062603)	NA
170	1956_302	M	Familial	CL	15q22.31	63073899-63183826	109 928	Gain	*MTFMT, OSTbeta, RASL12, hCG_1645727*	Paternal	SNP array	Illumina 550K (Glessner et al. 2009)	(not tested)	Autism by ADI-R and ASD by ADOS, MR (untestable by Ravens, low functioning), non verbal (AGRE ID: AU1377302)	NA
171	1165_3	M	Familial	CL	17q24.2	69345596-70202513	856 918	Gain	15 genes	Paternal	SNP array	Illumina 550K (Glessner et al. 2009), Affy 5.0 (Bucan et al. 2009)	(not tested)	Autism by ADI-R and ADOS, MR, non verbal, poor suck at birth, floppy infant, gastrointestinal problems (AGRE ID: AU043903)	NA

[a] DNA source: CL, cell line (peripheral blood lymphoblastoid cell line); BBC, buccal swab; U, unknown (most were either blood or buccal swab).

[b] Human reference genome NCBI v36, hg18.

Abbreviations: ADI-R, Autism Diagnostic Interview-Revised; ADOS, Autism Diagnostic Observation Schedule; ASD, autism spectrum disorder; CARS, Childhood Autism Rating Scale; dx, diagnosis; DZ, dizygotic; hx, history; IQ, intellectual quotient; LR-PCR, long range PCR; MLPA, multiplex ligation probe amplification; MR, mental retardation; MZ, monozygotic; NA, not available; OCD, obsessive-compulsive disorder; PDD-NOS, pervasive developmental disorder-not otherwise specified; SLI, specific language impairment

Supplementary Table 8. Rare CNVs in 996 ASD cases

Data provided separately, as an Excel workbook file.

Supplementary Table 9. List of known ASD genes, ID genes, and ASD-candidates

Gene symbol or Locus	Cytoband	Type	Associated condition
ASD Implicated			
NRXN1	2p16.3	gene	disrupted in ASD, MR, schizophrenia
SLC9A9	3q24	gene	rare homozygous deletions, sequence mutations
c3orf58	3q24	gene	*DIA1* (Deleted in Autism), rare homozygous deletions
NIPBL	5p13.2	gene	Cornelia de Lange syndrome
NSD1	5q35.3	gene	Sotos syndrome
AHI1	6q23.3	gene	Joubert syndrome (autosomal recessive)
CNTNAP2	7q36.1	gene	Cortical dysplasia-focal epilepsy syndrome
CHD7	8q12.2	gene	CHARGE syndrome
VPS13B	8q22.2	gene	Cohen syndrome (autosomal recessive)
TSC1	9q34	gene	Tuberous sclerosis
PTEN	10q23.3	gene	PTEN hamartoma tumor syndrome
DHCR7	11q13.2-q13.5	gene	Smith-Lemli-Optiz syndrome
CACNA1C	12p13.3	gene	Timothy syndrome
PTPN11	12q24	gene	Noonan syndrome
UBE3A	15q11-q13	gene	Angelman syndrome (gene within locus)
TSC2	16p13.3	gene	Tuberous sclerosis
CREBBP	16p13.3	gene	Rubinstein-Taybi syndrome
RAI1	17p11.2	gene	Smith-Magenis syndrome (deletion), Potocki-Lupski syndrome (duplication) (gene within locus)
NF1	17q11.2	gene	neurofibromatosis type 1, NF1 microdeletion/microduplicatiion syndrome
DMPK	19q13.3	gene	Myotonic dystrophy type 1
TBX1	22q11.21	gene	22q11 deletion syndrome (velocardiofacial/DiGeorge syndrome) (gene within locus)
ADSL	22q13.1	gene	Adenylosuccinate lyase deficiency (autosomal recessive)
SHANK3	22q13.3	gene	22q13 deletion syndrome
NLGN4X	Xp22.32-p22.31	gene	ASD, nonsyndromic XLMR
CDKL5	Xp22	gene	Rett like syndrome with infantile spasms
ARX	Xp21	gene	large spectrum of MR phenotypes: X-linked lissencephaly and abnormal genitalia, West syndrome, Partington syndrome, nonsyndromic MR
IL1RAPL1	Xp22.1-p21.3	gene	ASD, nonsyndromic XLMR
DMD	Xp21.2	gene	Muscular dystrophy, Duchenne and Becker types
FGD1	Xp11.21	gene	Aarskog-Scott syndrome, nonsyndromic XLMR
NLGN3	Xq13.1	gene	ASD, nonsyndromic XLMR
ATRX	Xq21.1	gene	large spectrum of phenotypes, including ATRX syndrome and nonsyndromic XLMR
FMR1	Xq27.3	gene	Fragile X mental retardation 1, ASD
AFF2	Xq28	gene	nonsyndromic XLMR
SLC6A8	Xq28	gene	Creatine deficiency syndrome, nonsyndromic XLMR
MECP2	Xq28	gene	Rett syndrome, syndromic and nonsyndromic XLMR, neonatal encephalopathy (in males)
RPL10	Xq28	gene	mutated in two ASD families, non syndromic XLMR
1p36_del	1p36	locus	1p36 microdeletion syndrome (chr1:1-5308621)
1q21_del_dup	1q21.1	locus	1q21.1 recurrent microdeletion/microduplication (chr1:144979000-146204000)
2q37_del	2q37	locus	2q37 monosomy (chr2:239619630-242951149)
4p16_WH_syndr	4p16	locus	Wolf-Hirschhorn syndrome (chr4:1-2043468)
7q11_del_dup	7q11.23	locus	Williams-Beuren syndrome (deletion), 7q11.23 duplication syndrome (chr7:71970679-74254837)
15q11-13_del_dup	15q11.2-q13.1	locus	Angelman syndrome (maternal deletion), Prader-Willi syndrome (paternal deletion), 15q11-q13 duplication (chr15:21309483-26230781)
15q13_del	15q13.3	locus	15q13.3 microdeletion syndrome (chr15:28557287-30488774)
15q24_del	15q24	locus	15q24 recurrent microdeletion syndrome (chr15:72164227-73949332)
16p11_del_dup	16p11.2	locus	16p11.2 autism susceptiblity locus (chr16:29550000-30200000)
22q11_del_dup	22q11.21	locus	22q11 deletion syndrome (velocardiofacial/DiGeorge syndrome), 22q11 duplication syndrome (chr22:16926349-20666469)
Intellectual disability (ID)			
POMGNT1	1p34.1	gene	Muscle-eye-brain disease (autosomal recessive)
ASPM	1q31	gene	Microcephaly and MR (autosomal recessive)
NPHP1	2q13	gene	Joubert syndrome (autosomal recessive)
MBD5	2q23.1	gene	autosomal dominant MR
CRBN	3p26.3	gene	autosomal recessive nonsyndromic MR
ARL13B	3q11.2	gene	Joubert syndrome (autosomal recessive)
CC2D2A	4p15.33	gene	Joubert syndrome, Meckel syndrome, COACH syndrome (autosomal recessive)
PRSS12	4q26	gene	autosomal recessive nonsyndromic MR
SYNGAP1	6p21.32	gene	nonsyndromic MR

Gene symbol or Locus	Cytoband	Type	Associated condition
GRIK2	6q16.3	gene	autosomal recessive nonsyndromic MR (gene also in ASD candidate list)
HOXA1	7p15.3	gene	Bosley-Salih-Alorainy/Athabaskan brainstem dysgenesis syndromes, autosomal recessive MR (gene also in ASD candidate list)
AP4M1	7q22.1	gene	autosomal recessive type of tetraplegic cerebral palsy with MR
RELN	7q22	gene	Lissencephaly
MCPH1	8p23.1	gene	Microcephaly and MR (autosomal recessive)
TUSC3	8p22	gene	nonsyndromic autosomal recessive MR
TMEM67	8q22.1	gene	Joubert syndrome (autosomal recessive)
VLDLR	9p24	gene	Cerebellar hypoplasia and MR (autosomal recessive)
FKTN	9q31.2	gene	Fukuyama congenital muscular dystrophy with type 2 lissencephaly (autosomal recessive)
CDK5RAP2	9q33.2	gene	Microcephaly vera (autosomal recessive)
STXBP1	9q34.11	gene	autosomal dominant MR and nonsyndromic epilepsy
POMT1	9q34.13	gene	Walker-Warburg syndrome (autosomal recessive)
EHMT1	9q34.3	gene	9q subtelomeric deletion syndrome (gene within locus)
SMC3	10q25.2	gene	Cornelia de Lange syndrome
ALG8	11q14.1	gene	Congenital disorder of glycosylation type Ih (autosomal recessive)
CEP290	12q21.32	gene	Joubert syndrome (autosomal recessive)
CENPJ	13q12.12	gene	Microcephaly vera (autosomal recessive)
FOXG1	14q12	gene	Congenital variant of Rett syndrome
POMT2	14q24.3	gene	Walker-Warburg syndrome (autosomal recessive)
RPGRIP1L	16q12.2	gene	Joubert syndrome (autosomal recessive)
PAFAH1B1	17p13.3	gene	Miller-Dieker lissencephaly syndrome (gene within locus)
YWHAE	17p13.3	gene	Miller-Dieker lissencephaly syndrome (gene within locus)
CC2D1A	19p13.12	gene	nonsyndromic autosomal recessive MR
DNMT3B	20q11.2	gene	Immunodeficiency, centromeric instability, and facial dysmorphism (ICF) syndrome (autosomal
EP300	22q13.2	gene	Rubinstein-Taybi syndrome
MID1	Xp22	gene	Opitz syndrome
HCCS	Xp22.2	gene	Microphthalmia with linear skin defects syndrome (MCOPS7)
OFD1	Xp22.2	gene	Oral facial digital syndrome type I, Simpson-Golabi-Behmel syndrome type 2
FANCB	Xp22.2	gene	Fanconi anemia complementation group B
AP1S2	Xp22.2	gene	nonsyndromic XLMR
NHS	Xp22.13	gene	Nance-Horan syndrome
PDHA1	Xp22.12	gene	Pyruvate decarboxylase deficiency
RPS6KA3	Xp22.12	gene	Coffin-Lowry syndrome, nonsyndromic MR
SMS	Xp22.1	gene	Snyder-Robinson syndrome
GK	Xp21.2	gene	Glycerol kinase deficiency, MR with hyperglycerolemia
OTC	Xp21.1	gene	Ornithine carbamoyltransferase deficiency
TSPAN7	Xp11.4	gene	nonsyndromic XLMR
BCOR	Xp11.4	gene	Lenz microphthalmia 2, oculofaciocardiodental syndrome
ATP6AP2	Xp11.4	gene	X-linked mental retardation with epilepsy
CASK	Xp11.4	gene	XLMR and microcephaly with pontine and cerebellar hypoplasia, XLMR with nystagmus
MAOA	Xp11.3	gene	Brunner syndrome
NDP	Xp11.4	gene	Norrie disease
ZNF674	Xp11.3	gene	nonsyndromic XLMR
ZNF41	Xp11.23	gene	nonsyndromic XLMR
SYN1	Xp11.23	gene	Epilepsy, X-linked, with variable learning disabilities and behavior disorders
ZNF81	Xp11.23	gene	nonsyndromic XLMR
FTSJ1	Xp11.23	gene	nonsyndromic XLMR
PORCN	Xp11.23	gene	Focal dermal hypoplasia
PQBP1	Xp11.23	gene	large spectrum of MR phenotypes, including nonsyndromic MR
SYP	Xp11.23	gene	nonsyndromic XLMR
SHROOM4	Xp11.22	gene	Stocco dos Santos X-linked mental retardation syndrome, nonsyndromic XLMR
JARID1C	Xp11.22	gene	Mental retardation, X-linked, JARID1C-related; syndromic and nonsyndromic MR
SMC1A	Xp11.22	gene	Cornelia de Lange syndrome
HSD17B10	Xp11.22	gene	2-methyl-3-hydroxybutyryl-CoA dehydrogenase (MHBD) deficiency
HUWE1	Xp11.22	gene	nonsyndromic XLMR
PHF8	Xp11.22	gene	Siderius-Hamel syndrome
KLF8	Xp11.21	gene	nonsyndromic XLMR
ARHGEF9	Xq11.1	gene	Hyperekplexia and epilepsy
OPHN1	Xq12	gene	MR with cerebellar and vermis hypoplasia
DLG3	Xq13.1	gene	nonsyndromic XLMR
MED12	Xq13.1	gene	Opitz-Kaveggia syndrome (FG syndrome-1)
ZMYM3	Xq13.1	gene	syndromic XLMR
SLC16A2	Xq13.2	gene	Allan-Herndon-Dudley syndrome
KIAA2022	Xq13.3	gene	syndromic XLMR
MAGT1	Xq21.1	gene	nonsyndromic XLMR
ATP7A	Xq21.1	gene	Menkes disease, occipital horn syndrome, X-linked cutis laxa
PGK1	Xq13	gene	Phosphoglycerate kinase 1 deficiency

Gene symbol or Locus	Cytoband	Type	Associated condition
BRWD3	Xq21.1	gene	nonsyndromic XLMR
ZNF711	Xq21.1	gene	nonsyndromic XLMR
PCDH19	Xq22.1	gene	Female-restricted epilepsy and mental retardation syndrome
SRPX2	Xq22.1	gene	Rolandic epilepsy with speech dyspraxia
TIMM8A	Xq22.1	gene	Mohr-Tranebjaerg syndrome, Jensen syndrome
NXF5	Xq22.1	gene	syndromic XLMR
PLP1	Xq22.2	gene	Pelizaeus-Merzbacher disease
PRPS1	Xq22.3	gene	Phosphoribosylpyrophosphate synthetase I superactivity
ACSL4	Xq22.3	gene	nonsyndromic XLMR
PAK3	Xq22.3	gene	nonsyndromic XLMR
DCX	Xq22.3	gene	Type 1 lissencephaly
AGTR2	Xqq23	gene	nonsyndromic XLMR
UBE2A	Xq24	gene	syndromic XLMR *UBE2A*-related
UPF3B	Xq24	gene	nonsyndromic XLMR
NDUFA1	Xq24	gene	Mitochondrial complex I deficiency
LAMP2	Xq24	gene	Danon disease
CUL4B	Xq24	gene	nonsyndromic XLMR
GRIA3	Xq25	gene	nonsyndromic XLMR
OCRL	Xq25	gene	Lowe syndrome
ZDHHC9	Xq26.1	gene	nonsyndromic XLMR
GPC3	Xq26.1	gene	Simpson-Golabi-Behmel syndrome type 1
PHF6	Xq26.2	gene	Borjeson-Forssman-Lehmann syndrome
HPRT1	Xq26.2	gene	Lesch-Nyhan syndrome
SLC9A6	Xq26.3	gene	syndromic XLMR, Christianson syndrome
ARHGEF6	Xq26.3	gene	nonsyndromic XLMR
SOX3	Xq27.1	gene	XLMR with isolated growth hormone deficiency, X-linked panhypopituitarism
IDS	Xq28	gene	Mucopolysaccharidosis II
ABCD1	Xq28	gene	Adrenoleukodystrophy
L1CAM	Xq28	gene	X-linked hydrocephalus, MASA syndrome
AVPR2	Xq28	gene	X-linked nephrogenic diabetes insipidus
FLNA	Xq28	gene	Bilateral periventricular nodular heterotopia
GDI1	Xq28	gene	nonsyndromic XLMR
IKBKG	Xq28	gene	Incontinentia pigmenti
DKC1	Xq28	gene	Dyskeratosis congenita
2p15-p16_del	2p15-16.1	locus	2p15-16.1 microdeletion syndrome (chr2:57595300-61591838)
2q33_del	2q33.1	locus	2q33.1 deletion syndrome (chr2:196633334-204915185)
3q29_del_dup	3q29	locus	3q29 microdeletion syndrome, 3q29 microduplication syndrome (chr3:197156626-198982266)
5p_del_Cri-du-Chat-syndr	5p15	locus	Cri du Chat syndrome (5p deletion) (chr5:1-11776854)
8p23_del	8p23.1	locus	8p23.1 deletion syndrome (chr8:8156705-11803128)
9q_subtel_del	9q34.3	locus	9q subtelomeric deletion syndrome (chr9:139523184-140273252)
11p13_WAGR_del	11p13	locus	WAGR 11p13 deletion syndrome (chr11:31760085-32467564)
Potocki-Shaffer_syndr	11p11.2	locus	Potocki-Shaffer syndrome (chr11:43941853-46021136)
12q14_del	12q14	locus	12q14 microdeletion syndrome (chr12:63358186-66931792)
15q26_overgrowth_syndr	15q26.3	locus	15q26 overgrowth syndrome (*IGF1R* gene) (chr15:97175493-100338915)
ATR-16_syndr	16p13.3	locus	ATR-16 syndrome (alpha thalassemia/mental retardation syndrome) (chr16:1-774373)
16p13_del_dup	16p13.11	locus	16p13.11 recurrent microdeletion/microduplication (chr16:15411955-16191749)
16p11.2-p12.2_del	16p11.2-p12.2	locus	16p11.2-p12.2 microdeletion syndrome (chr16:21521457-28949693)
17p13_MD_lissen_syndr_del	17p13.3	locus	17p13.3 Miller-Dieker lissencephaly syndrome (chr17:1-2492179)
17q21_del	17q21.31	locus	17q21.31 microdeletion syndrome (chr17:40988249-41565982)
Cat-Eye_syndr	22q11.1	locus	Cat eye syndrome (type I) (chr22:1-16971860)
22q11_distal_del	22q11.2	locus	22q11.2 distal deletion syndrome (chr22:20445848-22026229)
ASD candidates			
B3GALT6	1p36.33	gene	*de novo* deletion
CA6	1p36.2	gene	rare CNV
MTF1	1p34.3	gene	association
RIMS3	1p34.2	gene	*de novo* deletion
NEGR1	1p31.1	gene	translocation breakpoint
DPYD	1p21.3	gene	*de novo* deletion
GPR89A	1q21.1	gene	rare CNV
RFWD2	1q25.1-q25.2	gene	rare CNV
PAPPA2	1q25.2	gene	rare CNV
ASTN1	1q25.2	gene	paralogue of *ASTN2*
DISC1	1q42.1	gene	association
DPP10	2q14.1	gene	rare CNV, paralogue of *DPP6*
GALNT13	2q24.1	gene	rare CNV
SLC4A10	2q24.2	gene	*de novo* deletion
SCN7A	2q24.3	gene	rare CNV

Gene symbol or Locus	Cytoband	Type	Associated condition
SLC25A12	2q31.1	gene	association, linkage
ITGA4	2q31.3	gene	association, translocation breakpoint
INPP1	2q32.2	gene	association
AGAP1	2q37.2	gene	association, linkage, rare CNV
CNTN4	3p26.3	gene	rare CNV
OXTR	3p25.3	gene	association
FHIT	3p14.2	gene	*de novo* deletion
SUCLG2	3p14.1	gene	rare CNV
CNTN3	3p12.3	gene	rare CNV
FBXO40	3q13.33	gene	rare CNV
NLGN1	3q26.31	gene	rare CNV
HTR3C	3q27.1	gene	association
GABRG1	4p12	gene	inversion breakpoint, chromosomal duplication
GABRA4	4p12	gene	association
GABRB1	4p12	gene	association
EIF4E	4q23	gene	translocation breakpoint
PCDH10	4q28.3	gene	rare CNV
TDO2	4q32.1	gene	association
TPPP	5p15.3	gene	rare CNV
SEMA5A	5p15.2	gene	association
CTNND2	5p15.2	gene	rare CNV
CDH18	5p14.3	gene	translocation breakpoint
CDH10	5p14.2	gene	association
CDH9	5p14.1	gene	association
KLHL3	5q31.2	gene	translocation breakpoint
RNF182	6p23	gene	*de novo* deletion
CD83	6p23	gene	*de novo* deletion
CAP2	6p22.3	gene	*de novo* gain
RNF8	6p21.3	gene	rare CNV
KIAA1586	6p12.1	gene	rare CNV
GRIK2	6q16.3	gene	association, linkage (gene also in ID list)
PLN	6q22.31	gene	rare CNV
PARK2	6q26	gene	rare CNV
TMEM195	7p21.1	gene	*de novo* deletion
HOXA1	7p15.3	gene	*de novo* deletion (gene also in ID list)
AUTS2	7q11.22	gene	rare CNV
SRPK2	7q22.2	gene	inversion breakpoint
PIK3CG	7q22.3	gene	association
LAMB1	7q31.1	gene	association
NRCAM	7q31.1	gene	association
MET	7q31.2	gene	association
ST7	7q31.2	gene	translocation breakpoint
WNT2	7q31.2	gene	association
CADPS2	7q31.3	gene	rare CNV, association
GRM8	7q31.33	gene	rare CNV
UBE2H	7q32.2	gene	association, paralogue of *UBE3A*
PRKAG2	7q36.1	gene	rare CNV
DPP6	7q36.2	gene	rare CNV
EN2	7q36.3	gene	association
DLGAP2	8p23.3	gene	*de novo* gain
RB1CC1	8q11.23	gene	rare CNV
STK3	8q22.2	gene	translocation breakpoint
PIP5K1B	9q21.11	gene	translocation breakpoint
ASTN2	9q33.1	gene	rare CNV
GATA3	10p14	gene	*de novo* deletion
JMJD1C	10q21.2-q21.3	gene	inversion breakpoint
REEP3	10q21.3	gene	inversion breakpoint
KCNMA1	10q22.3	gene	translocation breakpoint
GRID1	10q23.1	gene	rare CNV
SHANK2	11q13.3	gene	paralogue of *SHANK3*
HTR3A	11q23.1	gene	association
AVPR1A	12q14.2	gene	association
TPH2	12q21.1	gene	association
GALNT9	12q24.33	gene	*de novo* deletion
NBEA	13q13.3	gene	translocation breakpoint, *de novo* deletion
PCDH9	13q21.32	gene	rare CNV
MDGA2	14q21.3	gene	rare CNV
GABRB3	15q12	gene	association

Gene symbol or Locus	Cytoband	Type	Associated condition
GABRA5	15q12	gene	rare CNV
NDNL2	15q13.1	gene	rare CNV
DUOXA1	15q21.1	gene	rare CNV
A2BP1	16p13.3	gene	*de novo* deletion
PRKCB	16p11.2	gene	association
GRIN2A	16p13.2	gene	association
ANKRD11	16q24.3	gene	*de novo* deletion
SLC6A4	17q11.2	gene	association
ITGB3	17q21.32	gene	association
BZRAP1	17q22	gene	rare CNV
ANKRD12	18p11.22	gene	paralogue of *ANKRD11*
ZNF676	19p12	gene	rare CNV
SHANK1	19q13.3	gene	paralogue of *SHANK3*
PAK7	20p12	gene	rare CNV
PDE9A	21q22.3	gene	*de novo* duplication
WDR4	21q22.3	gene	*de novo* duplication
NDUFV3	21q22.3	gene	*de novo* duplication
PKNOX1	21q22.3	gene	*de novo* duplication
GRPR	Xp22.2	gene	translocation breakpoint
PTCHD1	Xp22.11	gene	*de novo* deletions

We defined three gene-sets based on evidence from previous studies of their involvement in autism spectrum disorders (ASDs): 1) 'ASD implicated' consisting of 36 disease genes and 10 loci (involved in microdeletion/microduplication syndromes) strongly implicated in ASD and identified in subjects with ASD only or ASD and ID; 2) 'ID' consisting of 110 disease genes and 17 loci known to be implicated in ID but not yet in ASD; and 3) 'ASD candidates' including 103 candidate genes drawn from previous ASD studies, including case reports of cytogenetic abnormalities, allelic association and CNV studies.
The genomic coordinates of the loci refer to the UCSC human genome assembly hg18 (NCBI build 36).
Abbreviations: del, deletion; dup, duplication; ID, intellectual disability; MR, mental retardation; XLMR, X-linked mental retardation

Supplementary Table 10. Clinically-relevant findings

Loci

AGP ID	Gender	Family type	Chr	Start-end	Size (bp)	CNV type	Locus	Inheritance	Comments
13135_1523	F	UKN	1	144838594-146308287	1 469 693	Gain	1q21.1 microduplication	N/A	1q21.1 microdeletions and microduplications have been reported in patients with MR and/or ASD. The phenotype is highly variable, specially for duplications; both deletions and duplications can be inherited from unaffected parents (incomplete penetrance)
13050_593	M	UKN	15	21190624-26203954	5 013 330	Gain	15q11-13 duplication	*De novo*	Maternally-derived 15q11-13 duplication; this is the most common chromosomal abnormality reported in autism. Many AGP sites screen subjects for chromosomal abnormalities and 15q11-q13 rearrangements (and exclude them if positive), explaining why only one case was identified
1950_301	M	MPX	15	26762141-30436163	3 674 022	Loss	15q13.3 microdeletion syndrome	Maternal	15q13.3 deletions have been described in patients with epilepsy, MR, ASD, bipolar disorder and schizophrenia; they can be inherited from unaffected parents and be present in healthy siblings, with highly variable intra- and inter-familial phenotype. An additional family with 15q13.3 deletion failed the stringent quality control used in this study, but is described elsewhere (PMID: 19050728).
14167_2720	M	SPX	15	28705540-30436163	1 730 623	Loss	15q13.3 microdeletion syndrome	Paternal	
14283_4060	M	UKN	16	14771033-16307313	1 536 280	Loss	16p13.11 microdeletion syndrome	Maternal	16p13.11 deletions have been reported in MR with congenital anomalies, idiopathic generalized epilepsy and schizophrenia; duplications have been reported in MR, ASD and schizophrenia. Both deletions/duplications can be inherited from unaffected parents, and their role as causal or risk factors is not clear at present; duplications appear to be frequent in controls (~0.3%), so they could be rare neutral variants. At present, the clinical significance of 16p13.11 duplications is uncertain.
14142_2400	M	SPX	16	15387380-16256106	868 726	Gain	16p13.11 microduplication	Paternal	
5258_3	M	SPX	16	15387380-16270740	883 360	Gain	16p13.11 microduplication	Paternal	
5068_3	F	MPX	16	29502984-30127026	624 042	Loss	16p11.2 microdeletion syndrome	*De novo*	16p11.2 deletions and duplications have been reported in patients with autism and MR; both types are associated with incomplete penetrance and variable expressivity, particularly in the case of duplications
5359_4	M	SPX	16	29554843-30195224	640 381	Loss	16p11.2 microdeletion syndrome	*De novo*	
5262_4	M	SPX	16	29502984-30210849	707 865	Gain	16p11.2 microduplication syndrome	*De novo*	
3211_003	M	SPX	16	29502984-30127026	624 042	Gain	16p11.2 microduplication syndrome	Maternal	
3183_007	M	MPX	22	17241748-19819918	2 578 170	Loss	22q11 deletion syndrome	*De novo*	Both 22q11.2 deletions (DiGeorge syndrome) and duplications have been reported in ASD. Incomplete penetrance and variable expressivity of DiGeorge deletions are well known; microduplications are associated with an even higher phenotypic variability and many are inherited from unaffected parents
3127_004	M	MPX	22	17257787-19793730	2 535 943	Gain	22q11 duplication syndrome	Paternal	
5261_4	F	MPX	22	17257787-19795780	2 537 993	Gain	22q11 duplication syndrome	Paternal	

Genes

AGP ID	Gender	Family type	Chr	Start-end	Size (bp)	CNV type	Gene name	Exonic/Intronic	Inheritance	Comments
13017_223	F	UKN	2	50539877-50730546	190 669	Loss	*NRXN1*	Exonic	*De novo*	All *de novo NRXN1* CNVs observed here are exonic, whereas all inherited CNVs in cases as well as CNVs in controls are intronic, suggesting that exonic deletions (and maybe duplications) of *NRXN1* could be clinically relevant
13153_1703	M	UKN	2	50990306-51222043	231 737	Loss	*NRXN1*	Exonic	*De novo*	
13037_463	M	UKN	2	51002576-51157742	155 166	Loss	*NRXN1*	Exonic	*De novo*	
14068_1180	M	SPX	2	50493827-50677835	184 008	Gain	*NRXN1*	Exonic	*De novo*	
5126_4	M	MPX	X	28931559-29478966	547 407	Gain	*IL1RAPL1*	Exonic	Maternal	*IL1RAPL1* is involved in non syndromic X-linked MR when mutated/deleted; the effects of duplications and intronic deletions are unknown. The duplication is intragenic and is likely to disrupt the gene; the intronic deletion could also be deleterious but further studies are required
5036_4	M	MPX	X	29446046-29557942	111 896	Loss	*IL1RAPL1*	Intronic	Maternal	
3019_003	M	MPX	X	32100618-32315937	215 319	Gain	*DMD*	Exonic	Maternal	At least 10% of males with Duchenne and Becker muscular dystrophies have duplications of *DMD*. ASD has been described in patients with DMD, and about a third have MR. We identified 2 males with exonic duplications inherited from their mothers. Another male proband was found to carry a maternally-inherited exonic deletion but was not included in the counts of CNVs affecting MR genes because the CNV was <30 kb; all 3 CNVs have been validated. When last evaluated, the patients did not exhibit any motor difficulties, but they will be re-evaluated. (Note that patient 5126_4 with the *DMD* duplication also has an *ILIRAPL1* duplication)
5126_4	M	MPX	X	32948977-33330592	381 615	Gain	*DMD*	Exonic	Maternal	
5241_3*	M	MPX	X	31793278-31822704	29 427	Loss	*DMD*	Exonic	Maternal	
14216_3470	M	SPX	X	153263157-153474401	211 244	Gain	*RPL10*	Exonic	*De novo*	14216_3470 carries a *de novo* Xq28 duplication affecting 15 genes, including 3 genes involved in ASD/MR, *RPL10, GDI1* and *IKBKG. RPL10* missense variants were described in 2 ASD subjects; *GDI1* is involved in nonsyndromic XLMR while *IKBKG* is involved in a syndromic form of XLMR, incontinentia pigmenti.
14216_3470			X	153263157-153474401	211 244	Gain	*GDI1*	Exonic	*De novo*	
14216_3470			X	153263157-153474401	211 244	Gain	*IKBKG*	Exonic	*De novo*	
5353_3	F	SPX	6	33399849-33512042	112 193	Loss	*SYNGAP1*	Exonic	*De novo*	*SYNGAP1* was recently shown to be involved in nonsyndromic MR; it had not been implicated in ASD yet.
5419_3	M	SPX	X	41441499-41478503	37 004	Gain	*CASK*	Intronic	Maternal	*CASK* was recently involved in syndromic XLMR with brain malformations or nystagmus. Only sequence mutations have been reported thus far; the consequences of duplications, and in particular of intronic duplications as in this case, are unknown. Re-evaluation of the patient to determine if he has a clinical presentation compatible with *CASK* mutation is underway.
5007_3	M	MPX	X	46255974-46292959	36 985	Gain	*ZNF674*	Exonic	Maternal	Sequence mutations of *ZNF674* cause non syndromic XLMR; the effect of duplications are unknown. A duplication overlapping this gene was also observed in an AGP father with learning difficulties; the possible pathogenic role of these CNVs needs to be further studied

Chromosomal abnormalities

AGP ID	Gender	Family type	Chr	Start-end	Size (bp)	CNV type	Rearrangement	Inheritance	Comments
5467_3	M	SPX	1	233476547-247165725	13 689 178	Gain	13.5 Mb duplication 1q42.3-q44	*De novo*	Confirmed by qPCR (no karyotype available); the boy has autism and moderate MR, with no obvious dysmorphic features
14270_3930	F	SPX	6	160773919-170761395	9 987 477	Gain	10 Mb duplication 6q25.3-q27 (>100 genes)	*De novo*	High resolution karyotype was normal; subtelomere FISH analysis revealed a 6q terminal duplication arising from a paternal balanced translocation, 46,XY, t(6;22)(q25.3;p11.2). The father is healthy. The patient's karyotype is: 46,XX.ish der(22)t(6;22)(6q25.3;p11.2)pat(6qtel+). The girl has nonsyndromic autism and mild MR
13137_1543	F	SPX	8	31,928,590 - 58,996,070	27 067 480	Gain	26 Mb duplication 8p12-8q12.1	*De novo*	Mosaic supernumerary ring chromosome (47, XX, +r[10]/46, XX[70]), of unknown origin according to the karyotype, shown to involve chromosome 8 by the SNP array
5420_3	M	UKN	21	1-247249719	247 249 719	Gain	47,XY+21	*De novo*	Down's syndrome
5257_3	M	SPX	Y	1-57772954	57 772 954	Gain	47,XYY	*De novo*	XYY syndrome; both confirmed by karyotyping. There have been several case reports of XYY syndrome in subjects with ASD, and conversely, individuals with XYY appear to be at increased risk for ASD. However, these observations are based on small samples and larger epidemiological studies are required to determine the contribution of an extra Y chromosome to ASD.
5515_3	M	UKN	Y	1-57772954	57 772 954	Gain	47,XYY	*De novo*	

The first 4 abnormalities are considered etiologic and were excluded from the analyses.

Other likely clinically relevant CNVs (non exhaustive list)

AGP ID	Gender	Family type	Chr	Start-end	Size (bp)	CNV type	Genes	Inheritance	Comments
5386_3	M	MPX	6	156785155-158489874	1 704 720	Loss	5 genes	*De novo*	1.7 Mb *de novo* 6q25.3 deletion
13123_1403	F	UKN	9	98998-3682923	3 583 926	Loss	14 genes	*De novo*	3.5 Mb *de novo* 9p24.3-9p24.2 deletion
6240_4	M	SPX	11	126633939-132060374	5 426 436	Loss	20 genes	*De novo*	5.4 Mb *de novo* 11q24.2-q25 deletion; Jacobsen syndrome (chromosome 11q deletion syndrome), previously reported in two individuals with ASD
6053_3	M	MPX	12	54218922-58779615	4 560 694	Gain	94 genes	*De novo*	4.5 Mb *de novo* 12q13.3-q14.1 duplication; younger sister with Asperger syndrome does not carry the duplication
5444_3	M	SPX	17	76953064-77782267	829 204	Gain	40 genes	*De novo*	This child has two consecutive *de novo* CNVs in chromosome 17q25.3, the 829 kb duplication listed here and a smaller deletion, both have their breakpoints within *SLC16A3*; the deletion also affects *CSNK1D* (see Supplementary Table 6)
6358_6	M	SPX	19	4548413-5287389	738 977	Loss	9 genes	*De novo*	739 kb *de novo* 19p13.3 deletion

Other CNVs affecting ASD/MR genes not likely to be pathogenic

AGP ID	Gender	Family type	Chr	Start-end	Size (bp)	CNV type	Gene name	Exonic/Intronic	Inheritance	Comments
Duplications without characteristic phenotype and/or inherited from healthy parents										
3424_003	M	SPX	2	148881443-149078468	197 025	Gain	*MBD5*	Exonic	Maternal	Haploinsufficiency of *MBD5* (methyl-CpG binding domain protein 5) is responsible for the 2q23.1 microdeletion syndrome; no duplications described thus far. The patient carries a duplication involving the whole gene and does not have any sign similar to those described in the 2q23.1 microdeletion syndrome. This, together with the fact that the duplication is inherited from a healthy mother, suggests it is not pathogenic
1265_8	F	MPX	X	9931816-10758861	827 045	Gain	*MID1*	Exonic	Paternal	*MID1* is mutated in Opitz syndrome, an XLMR multiple congenital anomalies syndrome. This duplication was detected in a female proband, who inherited it from her reportedly healthy father, suggesting that it is not pathogenic
1376_301	M	MPX	X	76933952-77030430	96 478	Gain	*MAGT1*	Exonic	Maternal	*MAGT1* (magnesium transporter 1) is an XLMR gene described in 2008; no duplications had been reported. Recurrent duplication were observed in several AGP parents, including a healthy father, suggesting that the duplication of *MAGT1* is unlikely to be pathogenic
5036_4	M	MPX	X	148075334-148617551	542 217	Gain	*IDS*	Exonic	Maternal	Mucopolysaccharidosis type II (Hunter syndrome); XLMR. Effect of duplication unknown. The patient does not have any of the clinical features commonly observed in Hunter syndrome (coarse facies, macrocephaly, short stature, hepatosplenomegaly, joint contractures) and urine muccopolysaccharides were normal, suggesting that whole gene duplications of *IDS* are not deleterious
Autosomal recessive genes										
5267_3	M	UKN	3	3098326-3184518	86 192	Loss	*CRBN*	Exonic	Paternal	mutations in *CRBN* cause autosomal recessive nonsyndromic MR
14061_1040	M	SPX	4	119424168-119702863	278 695	Gain	*PRSS12*	Exonic	Maternal	mutations in *PRSS12* cause autosomal recessive nonsyndromic MR
13072_853	M	UKN	8	6428786-6552017	123 231	Loss	*MCPH1*	Exonic	N/A	mutations in *MCPH1* cause autosomal recessive microcephaly and MR
5323_3	M	UKN	8	15446378-15476553	30 175	Loss	*TUSC3*	Intronic	Paternal	mutations in *TUSC3* cause autosomal recessive nonsyndromic MR
5378_3	M	MPX	8	100610705-100707665	96 960	Loss	*VPS13B*	Exonic	Paternal	mutations in *VPS13B* cause Cohen syndrome
13123_1403	F	UKN	9	98998-3682923	3 583 925	Loss	*VLDLR*	Exonic	*De novo*	mutations in *VLDLR* cause autosomal recessive cerebellar hypoplasia and MR
14219_3520	M	SPX	11	77245396-77562430	317 034	Gain	*ALG8*	Exonic	Paternal	mutations in *ALG8* cause autosomal recessive congenital disorder of glycosylation type Ih

AGP ID	Gender	Family type	Chr	Start-end	Size (bp)	CNV type	Gene name	Exonic/Intronic	Inheritance	Comments
Autosomal recessive genes *(continued)*										
3145_003	M	MPX	2	110198845-110583308	384 463	Loss	*NPHP1*	Exonic	Maternal	mutations in *NPHP1* cause Joubert syndrome type 4
5112_4	M	MPX	2	110198845-110340339	141 494	Loss	*NPHP1*	Exonic	Maternal	
3049_003	F	MPX	2	110206673-110615080	408 407	Loss	*NPHP1*	Exonic	Paternal	
3181_007	M	MPX	2	110206673-110615080	408 407	Loss	*NPHP1*	Exonic	N/A	
3266_003	M	MPX	2	110206673-110615080	408 407	Loss	*NPHP1*	Exonic	Maternal	
14064_1110	M	SPX	2	110210164-110615080	404 916	Loss	*NPHP1*	Exonic	N/A	
6279_3	M	UKN	2	110210164-110340339	130 175	Loss	*NPHP1*	Exonic	Paternal	

* The CNV in proband 5241_3 did not meet the criteria for inclusion in the analyses of CNVs overlapping ASD/MR genes (<30 kb)

Abbreviations: ASD, autism spectrum disorder; F, female; M, male; MR, mental retardation; MPX, multiplex ASD family (two or more first-third degree relatives affected with ASD); N/A, not available; SPX, simplex ASD family (no first-third degree relatives with ASD); UKN, unknown family type (extended family not evaluated); XLMR, X-linked mental retardation

Supplementary Table 11. Population attributable risk (PAR)

A. Counts

	Candidates		ASD-implicated		ID-implicated only		ASD- & ID-implicated	
	Exposed	Not Exposed	Exposed	Not Exposed	Exposed	Not Exposed	Exposed	Not Exposed
Cases	78	918	43	953	36	960	76	920
Controls	94	1,193	30	1,257	30	1,257	58	1,229

B. PAR Estimation

	P(A)	R	P(A)(R-1)	PAR
ASD Candidates	0.073	1.08	0.006	0.60%
ASD-implicated	0.023	1.89	0.02	2.00%
ID-implicated	0.023	1.57	0.013	1.30%
ASD- & ID implicated	0.045	1.75	0.034	3.30%

PAR = P(A)*(R-1) / [1 + P(A)*(R-1)] (i.e., the % of cases that can be attributed to rare CNV);
P(A) = exposure to selected rare CNVs, which was estimated from the frequency in controls;
R = the relative risk for ASD, which was estimated from the odds ratio.

Supplementary Table 12. Number of analyzed gene-sets in the functional map enriched for deletions

	Number of gene-sets			
	All	Filtered[1]	Tested (deletions)[2]	Enriched for deletions in ASD[3]
Gene Ontology	10,438	4,684	2,696	67
KEGG	205	194	152	5
NCI	162	151	115	-
Reactome	89	65	59	-
PFAM	3,539	1,035	471	4
Total	14,433	6,129[2]	3,493	76

[1] Gene-sets with less than 5 genes and larger than 700 genes were excluded from the analyses.
[2] Filtered gene-sets whose genes were never overlapped by a CNV were not tested for enrichment. Of the 6,129 filtered genes, 3,493 genes had deletion counts >0 for either cases or controls, and were therefore used in the enrichment test and q-value estimation.
[3] Gene-sets enriched in ASD compared to controls with q-value <12.5%. 76/3,493 (2.18%) gene-sets were found to be enriched for deletions.

Supplementary Table 13. List of gene-sets enriched for deletions

Data provided separately, as an Excel workbook file.

H. ACKNOWLEDGEMENTS

We wish to acknowledge SAGE as part of this study:
Funding support for the Study of Addiction: Genetics and Environment (SAGE) was provided through the NIH Genes, Environment and Health Initiative [GEI] (U01 HG004422). SAGE is one of the genome-wide association studies funded as part of the Gene Environment Association Studies (GENEVA) under GEI. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center (U01 HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Support for collection of datasets and samples was provided by the Collaborative Study on the Genetics of Alcoholism (COGA; U10 AA008401), the Collaborative Genetic Study of Nicotine Dependence (COGEND; P01 CA089392), and the Family Study of Cocaine Dependence (FSCD; R01 DA013423). Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIH GEI (U01 HG004438), the National Institute on Alcohol Abuse and Alcoholism, the National Institute on Drug Abuse, and the NIH contract "High throughput genotyping for studying the genetic contributions to human disease" (HHSN268200782096C).

I. SUPPLEMENTARY REFERENCES

1 Szatmari, P. *et al.*, Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* 39 (3), 319-328 (2007).

2 Lord, C., Rutter, M., & Le Couteur, A., Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* 24 (5), 659-685 (1994).

3 Lord, C. *et al.*, 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 (3), 205-223 (2000).

4 Risi, S. *et al.*, Combining information from multiple sources in the diagnosis of autism spectrum disorders. *J Am Acad Child Adolesc Psychiatry* 45 (9), 1094-1103 (2006).

5 Bierut, L.J. *et al.*, A genome-wide association study of alcohol dependence. *Proc Natl Acad Sci U S A* [Epub ahead of print] (2010).

6 Stewart, A.F. *et al.*, Kinesin family member 6 variant Trp719Arg does not associate with angiographically defined coronary artery disease in the Ottawa Heart Genomics Study. *J Am Coll Cardiol* 53 (16), 1471-1472 (2009).

7 Korn, J.M. *et al.*, Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet* 40 (10), 1253-1260 (2008).

8 Zhang, J. *et al.*, iPattern: a cross-sample copy number variation discovery method for multiple array platforms. *Submitted* (2010).

9 Krawczak, M. *et al.*, PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet* 9 (1), 55-61 (2006).

10 Shaikh, T.H. *et al.*, High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. *Genome Res* 19 (9), 1682-1690 (2009).

11 Peiffer, D.A. *et al.*, High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res* 16 (9), 1136-1148 (2006).

12 Anney, R. *et al.*, A genome-wide scan for common variants affecting risk for autism. *Submitted* (2010).

13 Lee, A.B., Luca, D., Klei, L., Devlin, B., & Roeder, K., Discovering genetic ancestry using spectral graph theory. *Genet Epidemiol* 34 (1), 51-59 (2010).

14 Colella, S. *et al.*, QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res* 35 (6), 2013-2025 (2007).

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 (11), 1665-1674 (2007).

16 Redon, R. *et al.*, Global variation in copy number in the human genome. *Nature* 444 (7118), 444-454 (2006).

17 Pinto, D., Marshall, C., Feuk, L., & Scherer, S.W., Copy-number variation in control population cohorts. *Hum Mol Genet* 16 Spec No. 2, R168-173 (2007).

18 Conrad, D.F. *et al.*, Origins and functional impact of copy number variation in the human genome. *Nature*, [Epub ahead of print] (2009).

19 The Wellcome Trust Case Control Consortium, Genome-wide association study of copy number variation in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* Accepted (2010).

20 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 (2), R19 (2007).

21 Marshall, C.R. *et al.*, Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* 82 (2), 477-488 (2008).
22 Glessner, J.T. *et al.*, Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 459 (7246), 569-573 (2009).
23 Bucan, M. *et al.*, Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes. *PLoS Genet* 5 (6), e1000536 (2009).
24 Weiss, L.A. *et al.*, Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* 358 (7), 667-675 (2008).
25 Purcell, S. *et al.*, PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81 (3), 559-575 (2007).
26 Piton, A. *et al.*, Mutations in the calcium-related gene IL1RAPL1 are associated with autism. *Hum Mol Genet* 17 (24), 3965-3974 (2008).
27 Hendriksen, J.G. & Vles, J.S., Neuropsychiatric disorders in males with duchenne muscular dystrophy: frequency rate of attention-deficit hyperactivity disorder (ADHD), autism spectrum disorder, and obsessive--compulsive disorder. *J Child Neurol* 23 (5), 477-481 (2008).
28 Betancur, C., Sakurai, T., & Buxbaum, J.D., The emerging role of synaptic cell-adhesion pathways in the pathogenesis of autism spectrum disorders. *Trends Neurosci* 32 (7), 402-412 (2009).
29 Subramanian, A. *et al.*, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102 (43), 15545-15550 (2005).
30 Benjamini, Y. & Hochberg, Y., Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 51 (1), 289-300 (1995).
31 Storey, J.D. & Tibshirani, R., Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 100 (16), 9440-9445 (2003).
32 Storey, J.D., The positive false discovery rate: a Bayesian interpretation and the q-value. *Annals of Statistics* 31 (6), 2013-2035 (2003).
33 Dinu, I. *et al.*, Gene-set analysis and reduction. *Brief Bioinform* 10 (1), 24-34 (2009).
34 Cline, M.S. *et al.*, Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* 2 (10), 2366-2382 (2007).
35 Isserlin, R. *et al.*, Pathway analysis of dilated cardiomyopathy using global proteomic profiling and enrichment maps. *Proteomics* [Epub ahead of print] (2010).
36 Merico, D., Isserlin, R., Stueker, O., Emili, A., & Bader, G.D., Enrichment Map: a network-based method for gene-set enrichment visualization and interpretation. *Submitted* (2010).
37 Merico, D., Gfeller, D., & Bader, G.D., How to visually interpret biological data using networks. *Nat Biotechnol* 27 (10), 921-924 (2009).
38 Grossmann, S., Bauer, S., Robinson, P.N., & Vingron, M., Improved detection of overrepresentation of Gene-Ontology annotations with parent child analysis. *Bioinformatics* 23 (22), 3024-3031 (2007).
39 Falcon, S. & Gentleman, R., Using GOstats to test gene lists for GO term association. *Bioinformatics* 23 (2), 257-258 (2007).
40 Newey, S.E., Velamoor, V., Govek, E.E., & Van Aelst, L., Rho GTPases, dendritic structure, and mental retardation. *J Neurobiol* 64 (1), 58-74 (2005).
41 Conrad, D.F. *et al.*, Origins and functional impact of copy number variation in the human genome. *Nature* 464 (7289), 704-712 (2010).
42 Iafrate, A.J. *et al.*, Detection of large-scale variation in the human genome. *Nat Genet* 36 (9), 949-951 (2004).