Convergence of Genes and Cellular Pathways Dysregulated in Autism Spectrum Disorders

Dalila Pinto,[1,2,3,4,5,6,85] Elsa Delaby,[7,8,9] Daniele Merico,[10] Mafalda Barbosa,[1,2,3,4] Alison Merikangas,[11] Lambertus Klei,[12] Bhooma Thiruvahindrapuram,[10] Xiao Xu,[2,4,5] Robert Ziman,[10] Zhuozhi Wang,[10] Jacob A.S. Vorstman,[13] Ann Thompson,[14] Regina Regan,[15,16] Marion Pilorge,[7,8,9] Giovanna Pellecchia,[10] Alistair T. Pagnamenta,[17] Bárbara Oliveira,[18,19] Christian R. Marshall,[10,20] Tiago R. Magalhaes,[15,16] Jennifer K. Lowe,[21] Jennifer L. Howe,[10] Anthony J. Griswold,[22] John Gilbert,[22] Eftichia Duketis,[23] Beth A. Dombroski,[24] Maretha V. De Jonge,[13] Michael Cuccaro,[22] Emily L. Crawford,[25] Catarina T. Correia,[18,19] Judith Conroy,[16,26] Inês C. Conceição,[18,19] Andreas G. Chiocchetti,[23] Jillian P. Casey,[15,16] Guiqing Cai,[1,2,3] Christelle Cabrol,[7,8,9] Nadia Bolshakova,[11] Elena Bacchelli,[27] Richard Anney,[11] Steven Gallinger,[28] Michelle Cotterchio,[29] Graham Casey,[30] Lonnie Zwaigenbaum,[31] Kerstin Wittemeyer,[32] Kirsty Wing,[17] Simon Wallace,[33] Herman van Engeland,[13] Ana Tryfon,[1,2,80] Susanne Thomson,[25] Latha Soorya,[1,2,81] Bernadette Rogé,[34] Wendy Roberts,[35] Fritz Poustka,[23] Susana Mouga,[36,37] Nancy Minshew,[12] L. Alison McInnes,[1,2,82] Susan G. McGrew,[38] Catherine Lord,[39] Marion Leboyer,[40,41,42,43] Ann S. Le Couteur,[44] Alexander Kolevzon,[1,2,6] Patricia Jiménez González,[45] Suma Jacob,[46,47] Richard Holt,[17] Stephen Guter,[46] Jonathan Green,[48,49] Andrew Green,[16,50] Christopher Gillberg,[51] Bridget A. Fernandez,[52] Frederico Duque,[36,37] Richard Delorme,[40,53,54,55] Geraldine Dawson,[56] Pauline Chaste,[12,40] Cátia Café,[36] Sean Brennan,[11] Thomas Bourgeron,[40,53,54,57] Patrick F. Bolton,[58,59] Sven Bölte,[60,83] Raphael Bernier,[61] Gillian Baird,[62] Anthony J. Bailey,[33,84] Evdokia Anagnostou,[63] Joana Almeida,[36] Ellen M. Wijsman,[64,65] Veronica J. Vieland,[66] Astrid M. Vicente,[18,19] Gerard D. Schellenberg,[24] Margaret Pericak-Vance,[22] Andrew D. Paterson,[10,67] Jeremy R. Parr,[68] Guiomar Oliveira,[36,37] John I. Nurnberger,[69,70] Anthony P. Monaco,[17,71] Elena Maestrini,[27] Sabine M. Klauck,[72] Hakon Hakonarson,[73,74] Jonathan L. Haines,[25] Daniel H. Geschwind,[21] Christine M. Freitag,[23] Susan E. Folstein,[75] Sean Ennis,[16,50] Hilary Coon,[76] Agatino Battaglia,[77] Peter Szatmari,[14] James S. Sutcliffe,[25] Joachim Hallmayer,[78] Michael Gill,[11] Edwin H. Cook,[46] Joseph D. Buxbaum,[1,2,3,4,6,79] Bernie Devlin,[12] Louise Gallagher,[11] Catalina Betancur,[7,8,9,85,*] and Stephen W. Scherer[10,20,85,*]

Rare copy-number variation (CNV) is an important source of risk for autism spectrum disorders (ASDs). We analyzed 2,446 ASD-affected families and confirmed an excess of genic deletions and duplications in affected versus control groups (1.41-fold, $p = 1.0 \times 10^{-5}$) and an increase in affected subjects carrying exonic pathogenic CNVs overlapping known loci associated with dominant or X-linked ASD and intellectual disability (odds ratio = 12.62, $p = 2.7 \times 10^{-15}$, ~3% of ASD subjects). Pathogenic CNVs, often showing variable expressivity, included rare de novo and inherited events at 36 loci, implicating ASD-associated genes (*CHD2*, *HDAC4*, and *GDI1*) previously linked to other neurodevelopmental disorders, as well as other genes such as *SETD5*, *MIR137*, and *HDAC9*. Consistent with hypothesized gender-specific modulators, females with ASD were more likely to have highly penetrant CNVs ($p = 0.017$) and were also overrepresented among subjects with fragile X syndrome protein targets ($p = 0.02$). Genes affected by de novo CNVs and/or loss-of-function single-nucleotide variants converged on networks related to neuronal signaling and development, synapse function, and chromatin regulation.

[1]Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [2]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [3]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [4]The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [5]The Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [6]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [7]Institut National de la Santé et de la Recherche Médicale U1130, 75005 Paris, France; [8]Centre National de la Recherche Scientifique UMR 8246, 75005 Paris, France; [9]Neuroscience Paris Seine, Université Pierre et Marie Curie (Paris 6), Sorbonne Universités, 75005 Paris, France; [10]Program in Genetics and Genome Biology, The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON M5G 1L7, Canada; [11]Discipline of Psychiatry, School of Medicine, Trinity College Dublin, Dublin 8, Ireland; [12]Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA; [13]Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, 3584CX Utrecht, the Netherlands; [14]Department of Psychiatry and Behavioural Neurosciences, Offord Centre for Child Studies, McMaster University, Hamilton, ON L8S 4K1, Canada; [15]National Children's Research Centre, Our Lady's Children's Hospital, Dublin 12, Ireland; [16]Academic Centre on Rare Diseases, School of Medicine and Medical Science, University College Dublin, Dublin 4, Ireland; [17]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK; [18]Instituto Nacional de Saúde Doutor Ricardo Jorge, 1649-016 Lisboa, Portugal; [19]Center for Biodiversity, Functional, & Integrative Genomics, Faculty of Sciences, University of Lisbon, 1749-016 Lisboa, Portugal; [20]McLaughlin Centre, University of Toronto, Toronto, ON M5S 1A1, Canada; [21]Department of Neurology and Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; [22]John P. Hussman Institute for Human Genomics and Dr. John T. Macdonald Foundation Department of Human Genetics, Miller School of Medicine, University of Miami, Miami, FL 33136, USA; [23]Department of Child and Adolescent Psychiatry, Psychosomatics, and Psychotherapy, Goethe University, 60528 Frankfurt am Main, Germany; [24]Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; [25]Vanderbilt Brain Institute, Center for Human Genetics Research, and Department of Molecular Physiology & Biophysics, Vanderbilt University, Nashville, TN 37232, USA; [26]Children's University Hospital Temple Street, Dublin 1, Ireland;

Introduction

Autism spectrum disorders (ASDs) affect ~1% of the population and are characterized by impairments in social interaction and communication, as well as by repetitive and restricted behaviors. ASDs include mild to severe levels of impairment—cognitive function ranges from above average to intellectual disability (ID)—and are often accompanied by seizures and other medical problems. There is a ~4:1 male-to-female gender ratio in ASD.

ASDs are highly heritable,[1] and genomic studies have revealed that a substantial proportion of ASD risk resides in high-impact rare variation, ranging from chromosome abnormalities and copy-number variation (CNV)[2–6] to single-nucleotide variation (SNV).[7–11] These studies have highlighted a striking degree of genetic heterogeneity, implicating both de novo germline mutation and rare inherited ASD variation distributed across numerous genes. De novo CNVs are observed in 5%–10% of screened ASD-affected individuals, and after further follow-up studies, some of them have proven to alter high-risk genes (e.g., *NRXN1*[12] [MIM 600565]). De novo or transmitted CNVs, such as 15q11.2–q13 duplications of the affected region in Prader-Willi syndrome (PWS [MIM 176270]) and Angelman syndrome (AS [MIM 105830]), 16p11.2 deletion (MIM 611913), 16p11.2 duplication (MIM 614671), and X-linked deletions including the *PTCHD1-PTCHD1AS* locus (MIM 300828), have also been found to contribute to risk.[6,13,14] Exome and whole-genome sequencing studies have estimated at least another ~6% contribution to ASD[7–10,15] and an additional 5% conferred by rare inherited recessive or X-linked loss-of-function (LoF) SNVs.[11,16] A genetic overlap between ASD and other neuropsychiatric conditions has also been increasingly recognized.

Interestingly, CNV testing and exome sequencing have so far yielded mostly nonoverlapping genes, which might reflect different mutational mechanisms, although they

[27]Department of Pharmacy and Biotechnology, University of Bologna, 40126 Bologna, Italy; [28]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON M5G 1X5, Canada; [29]Cancer Care Ontario, Toronto, ON M5G 2L7, Canada; [30]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA; [31]Department of Pediatrics, University of Alberta, Edmonton, AB T6B 2H3, Canada; [32]School of Education, University of Birmingham, Birmingham B15 2TT, UK; [33]Department of Psychiatry, University of Oxford and Warneford Hospital, Oxford OX3 7JX, UK; [34]Unité de Recherche Interdisciplinaire Octogone, Centre d'Etudes et de Recherches en Psychopathologie, Toulouse 2 University, 31058 Toulouse, France; [35]Autism Research Unit, The Hospital for Sick Children, Toronto, ON M5G 1X8, Canada; [36]Unidade de Neurodesenvolvimento e Autismo do Serviço do Centro de Desenvolvimento da Criança and Centro de Investigação e Formação Clinica, Pediatric Hospital, Centro Hospitalar e Universitário de Coimbra, 3000-602 Coimbra, Portugal; [37]University Clinic of Pediatrics and Institute for Biomedical Imaging and Life Science, Faculty of Medicine, University of Coimbra, 3000-354 Coimbra, Portugal; [38]Department of Pediatrics, Vanderbilt University, Nashville, TN 37232, USA; [39]NewYork-Presbyterian/Weill Cornell Medical Center, New York, NY 10065, USA; [40]FondaMental Foundation, 94010 Créteil, France; [41]Institut National de la Santé et de la Recherche U955, Psychiatrie Génétique, 94010 Créteil, France; [42]Faculté de Médecine, Université Paris Est, 94010 Créteil, France; [43]Department of Psychiatry, Henri Mondor-Albert Chenevier Hospital, Assistance Publique – Hôpitaux de Paris, 94010 Créteil, France; [44]Institute of Health and Society, Newcastle University, Newcastle upon Tyne NE1 4LP, UK; [45]Child Developmental and Behavioral Unit, Hospital Nacional de Niños Dr. Sáenz Herrera, Caja Costarricense de Seguro Social, San José, Costa Rica; [46]Institute for Juvenile Research, Department of Psychiatry, University of Illinois at Chicago, Chicago, IL 60608, USA; [47]Institute of Translational Neuroscience and Department of Psychiatry, University of Minnesota, Minneapolis, MN 55455, USA; [48]Institute of Brain, Behaviour, and Mental Health, University of Manchester, Manchester M13 9PL, UK; [49]Manchester Academic Health Sciences Centre, Manchester M13 9NT, UK; [50]National Centre for Medical Genetics, Our Lady's Children's Hospital, Dublin 12, Ireland; [51]Gillberg Neuropsychiatry Centre, University of Gothenburg, 41119 Gothenburg, Sweden; [52]Discipline of Genetics, Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL A1B 3V6, Canada; [53]Human Genetics and Cognitive Functions Unit, Institut Pasteur, 75015 Paris, France; [54]Centre National de la Recherche Scientifique URA 2182 (Genes, Synapses, and Cognition), Institut Pasteur, 75015 Paris, France; [55]Department of Child and Adolescent Psychiatry, Robert Debré Hospital, Assistance Publique – Hôpitaux de Paris, 75019 Paris, France; [56]Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, NC 27710, USA; [57]University Paris Diderot, Sorbonne Paris Cité, 75013 Paris, France; [58]Institute of Psychiatry, King's College London, London SE5 8AF, UK; [59]South London & Maudsley Biomedical Research Centre for Mental Health, London SE5 8AF, UK; [60]Department of Child and Adolescent Psychiatry, Psychosomatics, and Psychotherapy, University of Frankfurt, 60528 Frankfurt, Germany; [61]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA 98195, USA; [62]Paediatric Neurodisability, King's Health Partners, King's College London, London WC2R 2LS, UK; [63]Bloorview Research Institute, University of Toronto, Toronto, ON M4G 1R8, Canada; [64]Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA 98195, USA; [65]Department of Biostatistics, University of Washington, Seattle, WA 98195, USA; [66]Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital, Columbus, OH 43205, USA; [67]Dalla Lana School of Public Health, Toronto, ON M5T 3M7, Canada; [68]Institute of Neuroscience, Newcastle University, Newcastle upon Tyne NE2 4HH, UK; [69]Institute of Psychiatric Research, Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN 46202, USA; [70]Department of Medical and Molecular Genetics and Program in Medical Neuroscience, Indiana University School of Medicine, Indianapolis, IN 46202, USA; [71]Office of the President, Tufts University, Medford, MA 02155, USA; [72]Division of Molecular Genome Analysis, German Cancer Research Center (Deutsches Krebsforschungszentrum), 69120 Heidelberg, Germany; [73]Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; [74]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; [75]Division of Child and Adolescent Psychiatry, Department of Psychiatry, Miller School of Medicine, University of Miami, Miami, FL 33136, USA; [76]Utah Autism Research Program, Department of Psychiatry, University of Utah School of Medicine, Salt Lake City, UT 84108, USA; [77]Stella Maris Clinical Research Institute for Child and Adolescent Neuropsychiatry, 56128 Calambrone, Pisa, Italy; [78]Department of Psychiatry, Stanford University Medical School, Stanford, CA 94305, USA; [79]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[80]Present address: International Laboratory for Brain, Music, and Sound Research, University of Montreal and McGill University, Montreal, QC H3C 3J7, Canada

[81]Present address: Department of Psychiatry, Rush University Medical Center, Chicago, IL 60612, USA

[82]Present address: Department of Psychiatry, Kaiser Permanente, San Francisco, CA 94118, USA

[83]Present address: Department of Women's and Children's Health, Center of Neurodevelopmental Disorders, Karolinska Institutet, 11330 Stockholm, Sweden

[84]Present address: Department of Psychiatry, University of British Columbia, Vancouver, BC V6T 2A1, Canada

[85]These authors contributed equally to this work

*Correspondence: catalina.betancur@inserm.fr (C.B.), stephen.scherer@sickkids.ca (S.W.S.)

http://dx.doi.org/10.1016/j.ajhg.2014.03.018.

might still perturb connected biological pathways.[17] Although numerous ASD-associated loci have been recognized to date,[18] they only account for a small fraction of the overall estimated heritability, consistent with predictions that there might be ~1,000 loci underlying ASD[19] and that many associated genes and risk variants remain to be identified.

Here, we have assessed the impact of de novo and inherited rare CNV in 2,446 ASD individuals and their parents from the Autism Genome Project (AGP), along with 2,640 unrelated controls, by applying a series of approaches to characterize candidate ASD-associated genes disrupted by CNVs and to identify the biological relationships and common pathways they share. Using evidence from multiple sources, we were able to directly implicate numerous dosage-sensitive genes as risk factors and provide insights into different but related mechanisms underlying ASD.

Subjects and Methods

ASD Samples

The samples were collected as part of the AGP, an international consortium with over 50 sites in North America and Europe. The first phase of the AGP involved examining genetic linkage and chromosomal rearrangements in 1,168 families with at least two ASD-affected individuals.[5] In the second phase, we genotyped simplex and multiplex families by using high-resolution microarrays to examine the contribution of rare CNVs and common SNPs to ASD. The second phase was divided in two stages; the results of stage 1, involving the first half of the families, were published in 2010.[6,20] In stage 2, we genotyped the remaining families (n = 1,604) for a total of over 2,845 families and performed genome-wide CNV (this study) and association studies.[21] Informed consent was obtained from all participants, and all procedures followed were in accordance with the ethical standards on human experimentation of the participating sites. The AGP sample set is a collection of families comprising an affected proband and two parents, as previously described in Pinto et al.[6] and Anney et al.[20,21] Many of the subjects at the recruiting sites were tested for fragile X syndrome (FXS [MIM 300624]) and assessed for chromosomal rearrangements with karyotype, fluorescence in situ hybridization, or multiplex-ligation-dependent probe amplification (MLPA); subjects with known karyotypic abnormalities, FXS, or other genetic disorders were typically excluded. The main analyses presented here were restricted to subjects of European ancestry.[21] All diagnostic, clinical, and cognitive assessments were carried out at each contributing site. All data were gathered at a central coordination site for standardization of data formatting and data quality assurance.

Autism Classification

Affected AGP participants were classified according to the Autism Diagnostic Observation Schedule (ADOS)[22] and the Autism Diagnostic Interview, Revised (ADI-R).[23] The ADOS is a semistructured, clinically administered instrument for assessing and diagnosing ASD. The ADI-R is a structured clinical interview conducted with the parents or caregivers; spectrum classification on the ADI-R was based on Risi et al.[24] The AGP *strict* and *spectrum* classifications are based on both instruments (Table S1A, available online). To meet criteria for strict autism, affected individuals must have an autism classification on both measures, whereas for the spectrum classification, individuals must meet the autism spectrum criteria on both measures or meet criteria for autism on one measure if the other measure was not available or not administered. The mean age of ADI-R assessment was 8 years.

Simplex and Multiplex Classification

Family type was classified as simplex, multiplex, or unknown. Simplex families had one known affected individual among the first- to third-degree relatives (cousins only) and included affected monozygotic twins. Multiplex families had at least two first- to third-degree relatives (cousins only) with a validated, clinical ASD diagnosis. All other situations, including instances where a family history of autism was not assessed explicitly, were coded as unknown.

Developmental Impairment

Cognitive functioning and adaptive function were measured with an appropriate standardized cognitive-testing instrument and the Vineland Adaptive Behavior Scale (VABS),[25] respectively. To maximize the available data, we created a developmental-impairment variable by using a hierarchical combination of scores on full-scale, performance, and verbal IQ measures and the VABS composite score. A cutoff of 70 was applied on all measures; subjects who could not complete an IQ assessment because of low functioning or behavior were assigned to the "low" category. In the hierarchy, full-scale IQ (followed by performance IQ, verbal IQ, and finally the VABS composite score) was the preferred measure. For example, a subject with a full-scale IQ < 70 but a performance IQ ≥ 70 was considered positive for developmental impairment. Additionally, subjects missing all IQ information with a "low" VABS composite score were also assigned to the developmental-impairment category.

Control Subjects

Unrelated control subjects were assembled from three studies in which individuals had no obvious psychiatric history: the Study of Addiction Genetics and Environment (SAGE),[26] the Ontario Colorectal Cancer Case-Control Study,[27,28] and Health, Aging, and Body Composition (HABC) (Table S1B).[29] Samples were genotyped on the same array platforms (Illumina 1M single or duo arrays) as those of ASD subjects and parents and were analyzed with the same quality-control (QC) procedures and CNV analysis pipeline. The control data set used in the primary CNV analysis was composed of 2,640 control individuals of European ancestry (1,241 males and 1,399 females) who passed QC (Table S1B). Secondary analyses included 1,843 subjects from other ancestries (SAGE and HABC non-European control individuals), giving a total of 4,768 control subjects of all ancestries.

Data Analysis

We performed genotyping and data cleaning, including SNP and intensity QC for CNV detection, as described previously[6] to ensure that CNV ascertainment was consistent among affected subjects, parents, and control subjects (see Table S1B for detailed QC steps). Samples not meeting our quality thresholds were excluded.

CNV Analysis

CNVs were detected with our analytical pipeline of Illumina 1M arrays (v.1 and v.3)[6,30] and analyzed for case-control differences

in burden with PLINK v.1.0730, R stats, and custom scripts. The p values associated with odds ratios (ORs) were calculated with Fisher's exact test. Rare de novo CNVs, clinically relevant CNVs, and other selected rare CNVs were validated by at least one method (quantitative PCR, MLPA, and/or long-range PCR). Table S4 shows all validated de novo CNVs. A list of CNV calls passing QC in affected subjects, including all experimentally validated CNVs, is available in Tables S17A, S17B, and S17C.

Secondary analyses included comparisons of CNV number, length, and intersected gene number between our 102 de novo CNVs identified in affected subjects and the 76 de novo CNVs in control subjects of two published data sets: (1) 17 de novo CNVs identified in 15 unaffected siblings from 872 families with a single ASD-affected offspring and an unaffected sibling from the Simons Simplex Collection[4] and (2) 59 de novo CNVs detected in 57 out of 2,623 Icelandic control trios.[31]

The clinical relevance of CNVs was interpreted according to the American College of Medical Genetics guidelines[32] irrespective of the subjects' affected status, and CNVs were classified as pathogenic, uncertain, or benign. Pathogenic CNVs are documented as clinically significant in multiple peer-reviewed publications and databases (e.g., OMIM and GeneReviews), even if the penetrance and the expressivity might be variable.

Gene Lists

In order to perform burden analyses, we compiled a series of lists:

(1) Genes and loci implicated causally in ASD (updated from Betancur),[18] all of which have also been implicated in ID, as well as genes and loci implicated in ID, but not yet in ASD (Tables S6A–S6D). Note that the list of genes and loci involved in ASD was updated independently of the data from AGP stage 1;[6] thus, genes and loci were included only if there was independent evidence from other studies.
(2) Highly-brain-expressed genes defined by a log(RPKM [reads per kb per million reads]) > 4.5 by the BrainSpan resource (n = 5,610 genes).
(3) Functionally characterized control genes not expressed in the brain (log(RPKM) < 1; n = 5,410 genes).
(4) Postsynaptic density (PSD) genes.[33]
(5) Genes found to interact with fragile X mental retardation protein (FMRP).[34]
(6) Genes associated with neurological phenotypes compiled from the Human Phenotype Ontology (HPO) and Mammalian Phenotype Ontology (MPO).
(7) Genes grouped by their probability of haploinsufficiency (pHI)[35] into three subgroups: pHI > 0.15 (n = 8,862 genes), pHI > 0.35 (n = 4,136 genes), and pHI > 0.55 (n = 2,214 genes).

One-Gene- and Multiple-Gene-Hit Burden Analysis

One-gene-hit burden analyses were performed with Fisher's exact test. When considering the possibility that multiple genes within a CNV event or across events in the same subject act in concert to increase risk (i.e., multiple-gene-hit burden), we fit a series of logit models to the data. For the logit model, which is a special case of generalized linear model, log odds of case status (logit) was fit to predictor variables, namely the number of brain-expressed genes (BrainSpan) covered by the CNV and the level of gene expression. To analyze the expression data, we transformed the normalized RPKM value of each gene in the neocortex to log(1+RPKM). All analyses were performed in the statistical package R with the function "glm" and the logit link.

Functional Enrichment and Network Analyses

Functional-enrichment association tests and pathway and network analyses were performed with custom scripts,[6] Bioconductor, NETBAG,[36] and DAPPLE.[37]

Results

Excess Genome-wide Burden of Rare and De Novo Genic CNVs

To explore the contribution of CNV to ASD, we expanded our previous study (stage 1)[6] with an additional 1,604 families (stage 2), bringing the total to 9,050 individuals from 2,845 ASD-affected families. We used an analytical pipeline of Illumina 1M arrays[6,30] to detect rare CNV in families and applied a series of QC filters, including validation of all de novo events by at least one method (Tables S1A–S1C). In total, 1,359 stage 2 families passed QC, and 2,446 families were used in the combined analyses of both stages (Tables S2A and S2B). Of these, 2,147 families were European, and 299 were of other ancestries.[21] We used the same pipeline to analyze 2,640 control individuals of European ancestry[26,27,29] who were genotyped with the same array platforms. Ancestry was inferred by analysis of SNP genotype data (Table S1B). The rate, size, and number of genes affected by rare (<1% frequency) CNVs were assessed. Consistent with our previous data, we observed that compared to control subjects, affected subjects had an increased burden in the number of genes affected by rare CNVs (1.41-fold increase, empirical $p = 1 \times 10^{-5}$; Table 1). This enrichment was apparent for both deletions and duplications and remained after we controlled for potential case-control differences (Table 1). Similar findings were obtained when each stage was considered separately (Tables S3A–S3C).

Array- and exome-based studies have revealed a substantial contribution of de novo variation to ASD risk,[19] prompting us to assess this further. After screening 2,096 trios (of all ancestries), we found 102 rare de novo CNVs in 99 affected subjects (three of whom had two events; Table S4). Overall, 4.7% of trios had at least one de novo CNV, whereas control subjects had a frequency of 1%–2%.[4,31,38] The average length of de novo events in our affected subjects (1.17 Mb) was larger than that of de novo CNVs in unaffected siblings from the Simons Simplex Collection (0.67 Mb, p = 0.01)[4] and in control trios (0.55 Mb, p = 0.01).[31] The average size of de novo CNVs was also larger than the size of all rare CNVs in our affected (188 kb) and control (159 kb) subjects. De novo CNVs affected 3.8-fold more genes in affected subjects than in control subjects[4,31] (2.6-fold for deletions and 6.1-fold for duplications). Even after controlling for the difference in CNV size by proportionally scaling the number of intersected genes in each group, we observed a 1.77-fold

Table 1. Genome-wide Burden of Genes Intersected by Rare CNVs in a Combined Sample of 2,147 European ASD Affected Subjects and 2,640 European Control Subjects

Type	Group Size	No. of Rare Genic CNVs	No. of Genes Intersected by Rare CNVs	Baseline Gene Rate (Control)[a]	Case-Control Gene Ratio	p_{corr}[b]
All	all	6,859	6,745	3.55	1.41	0.00001*
Deletions	all	2,946	2,804	1.23	1.40	0.00049*
Duplications	all	3,913	5,217	2.32	1.41	0.00001*
All	30–500 kb	6,307	5,163	2.89	1.07	0.03628*
	>500 kb	552	2,491	0.66	2.88	0.00001*
	>1 Mb	187	1,337	0.26	4.48	0.00001*
Deletions	30–500 kb	2,795	2,014	1.07	1.07	0.20110
	>500 kb	151	947	0.16	3.60	0.00051*
	>1 Mb	63	647	0.08	4.58	0.02289*
Duplications	30–500 kb	3,512	3,934	1.83	1.08	0.03750*
	>500 kb	401	1,896	0.50	2.64	0.00026*
	>1 Mb	124	890	0.18	4.43	0.00036*

Rare CNVs in samples of European ancestry were defined as ≥ 30 kb in size and present in the total sample set at a frequency < 1%. Gene coordinates were defined by the RefSeq boundaries plus a 10 kb region on either side. All genomic analyses used UCSC Genome Browser hg18. *Significant differences ($p \leq 0.05$) are indicated.
[a]The baseline gene rate (control) is defined as the average number of genes intersected by CNVs per control subject.
[b]Genome-wide p values were estimated in 100,000 permutations (one sided) and additionally corrected (p_{corr}) for global case-control differences in CNV rate and size. Analyses were further stratified according to CNV type (deletions or duplications) and size.

difference (1.2-fold for deletions and 2.8-fold for duplications, p = 0.02). Furthermore, de novo CNVs in simplex families intersected 4.0-fold more genes than did CNVs in controls[4,31] (1.8-fold after size correction, p = 0.01). There were no significant differences between subjects from simplex families and those from multiplex families in the frequency (5% and 4.2%, respectively) or gene content (n = 18.7 and 18.8, respectively) of de novo CNVs. Similarly, no significant difference was found between males and females in the size (1.17 and 1.2 Mb, respectively) or gene content (n = 18 and 17.3, respectively) of de novo CNVs. For 85 of 102 de novo events, it was possible to determine the parent of origin, and roughly equal numbers of events originated on the paternal allele (n = 45) and the maternal allele (n = 40) (Tables S5A–S5H). Taken together, our data indicate that there is an increased burden of de novo events in ASD-affected subjects. The clinical relevance of de novo CNVs in ASD is confirmed by the fact that among 102 such events identified, half (n = 46) are considered etiologically relevant, including 40 loci known to be involved in ASD and ID (see below).

We replicated previous observations, such as a de novo deletion intersecting *PTCHD1AS* in a male (adding to the evidence that both *PTCHD1* and *PTCHD1AS* contribute to ASD risk[14]) and de novo events involving the miRNA *miR137* (MIM 614304) in 1p21.2–p21.3 in two subjects. Microdeletions of *miR137* have been reported in ASD,[39] ID,[40] and schizophrenia.[41] Examples of ASD candidate genes identified by small de novo CNVs include *SETD5*, *DTNA* (MIM 601239), and *LSAMP* (MIM 603241) (Supplemental Data section "Highlighted Genes," Figures S9, S10, and S14).

CNV Burden in Autosomal-Dominant or X-Linked Genes and Loci Implicated in ASD and/or ID

At least 124 genes and 55 genomic loci have been implicated in ASD to date (Tables S6A and S6B; updated from Betancur[18]), all of which have also been implicated in ID. In addition, we compiled a list of genes and loci that have been implicated in ID, but not yet in ASD (Tables S6C–S6D). When we analyzed samples of inferred European ancestry, we found that 4% (87/2,147) of ASD-affected subjects had CNVs overlapping autosomal-dominant or X-linked genes and loci implicated in ASD and/or ID; this percentage was significantly higher than that in controls (OR = 4.09, 95% confidence interval [CI] = 2.64–6.32, $p = 5.7 \times 10^{-12}$; Figure 1A). We further classified these events into pathogenic, uncertain, or benign according to the American College of Medical Genetics guidelines.[32] Pathogenic (or clinically significant) CNVs were identified in 2.8% (60/2,147) of affected subjects (OR = 12.62, 95% CI = 5.44–29.27, $p = 2.74 \times 10^{-15}$), and pathogenic deletions showed a striking estimated OR of 23.13 (95% CI = 5.57–96.08, $p = 2.6 \times 10^{-11}$; Figure 1B). Furthermore, the enrichment of pathogenic CNVs overlapping genes involved in ASD and/or ID was independently observed when the data were broken down by stages: 2.6% (25/979) of affected subjects in stage 1 (OR = 7.61, $p = 1.22 \times 10^{-5}$) carried pathogenic CNVs, whereas 3.0% (35/1,168) in stage 2 (OR = 6.47, $p = 2.89 \times 10^{-7}$) carried pathogenic CNVs. Some of these CNVs (e.g., *NRXN1* deletion, 1q21 duplications [MIM 612475], and 16p11.2 duplications) were seen in a small fraction of control subjects, consistent with their variable

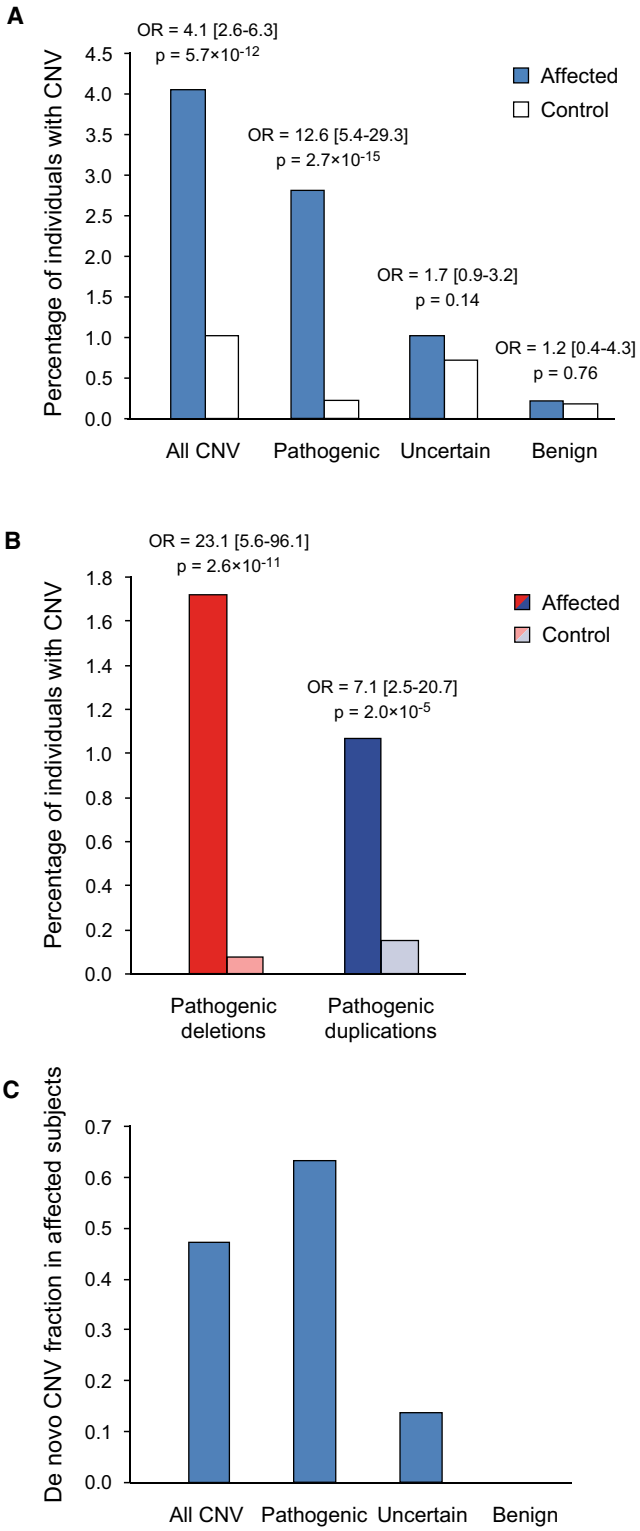


Figure 1. CNV Burden in Genes and Loci Implicated in ASD and/or ID

CNV data from 2,147 European affected subjects and 2,640 European control subjects were analyzed for overlap with genes and loci implicated in ASD and/or ID (results including non-European affected and control subjects are shown in Figure S1). Only CNVs affecting autosomal-dominant and X-linked dominant genes or loci in both genders (132 genes, 56 loci), as well as X-linked recessive genes or loci in males (52 genes, 2 loci), were considered ("all CNV"). Exonic ≥30 kb CNVs affecting an ASD- and/or ID-associated gene or overlapping at least 50% of the target loci were selected for further analysis. Rare CNVs were divided into three categories—pathogenic, uncertain clinical significance, or benign—without regard to affected status.

(A) Percentage of individuals with CNVs overlapping genes and loci implicated in ASD and/or ID ("all CNV"), pathogenic CNVs, uncertain CNVs, or benign CNVs; and OR in affected and control subjects.

(B) Percentage of individuals with pathogenic deletions or duplications and OR in affected and control subjects.

(C) Fraction of de novo CNVs in each category of affected subjects.

expressivity and/or incomplete penetrance. Among the affected subjects with pathogenic CNVs, 63% (38/60) carried de novo events (Figure 1C), including two subjects with two pathogenic events each.

When we further considered affected subjects of all ancestries (n = 2,446) and included chromosome abnormalities (>7.5 Mb), select large rare de novo events, and select experimentally validated smaller CNVs (<30 kb), we identified pathogenic CNVs in ~3.3% of individuals with unexplained ASD (84 pathogenic events in 82/2,446 subjects; Figures 2A and S1A–S1C; Tables S7A and S7B). This most likely represents an underestimate of the true etiologic yield, given that many of the subjects were assessed with clinical diagnostic methods and excluded if positive; similarly, those individuals with known congenital malformations or dysmorphic features were not enrolled. Interestingly, 83% (64/77 [5 without information]) of carriers of pathogenic CNVs were nonsyndromic (i.e., ASD without reported accompanying physical or neurological abnormalities), and 57% (44/77 [5 without information]) had no ID (Figure 2B). The fraction of subjects with ID among carriers of pathogenic CNVs (42%) was not significantly different from the fraction of ID among all affected subjects (46%).

Inheritance data showed that 64% (54/84) of pathogenic CNVs were de novo events (59% were deletions, and 41% were duplications) and that the remaining (36%) were inherited, including seven X-linked CNVs maternally transmitted to males and 23 (13 maternal and 10 paternal [27%]) on autosomes (Figure 2C). Pathogenic deletions tended to be smaller than duplications (Figure 2D). As expected, pathogenic de novo events were on average significantly larger than inherited ones (3.14 Mb—excluding three affected subjects with whole-chromosome aneuploidy—versus 1.44 Mb, respectively). We also observed that the proportion of females was significantly increased among carriers of highly penetrant pathogenic CNVs (male-to-female ratio of 2:1 versus 6:1 among all affected subjects; two-tailed Fisher's exact test p = 0.017; Figure 2E). In contrast, the male-to-female ratio among individuals with CNVs associated with variable expressivity was 6:1.

Pathogenic CNVs included well-characterized highly penetrant disorders associated with de novo CNVs, such as Phelan-McDermid syndrome (MIM 606232, 22q13.3 deletion including *SHANK3* [MIM 606230]), Smith-Magenis syndrome (MIM 182290, 17p11.2 deletion including

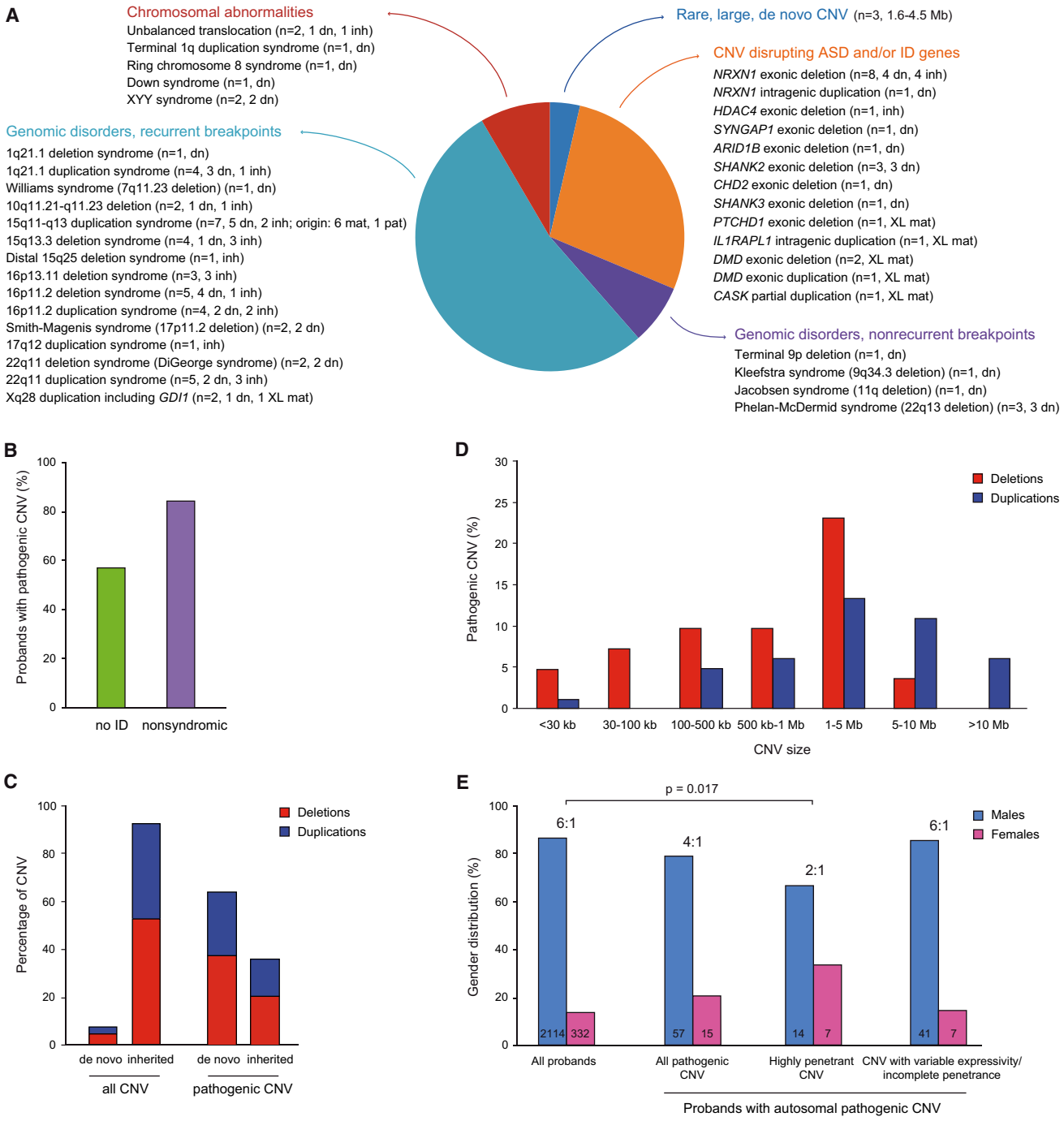


Figure 2. All Pathogenic CNVs Identified in Affected Subjects

CNVs overlapping genes and loci implicated in ASD and/or ID in 2,446 affected subjects irrespective of ancestry, plus chromosomal abnormalities, other large rare de novo events, and further experimentally validated CNVs < 30 kb. Pathogenic CNVs identified in affected subjects (84 CNVs in 82 probands) were divided into different categories: CNVs disrupting genes implicated in ASD and/or ID, genomic disorders with recurrent breakpoints, genomic disorders with nonrecurrent breakpoints, chromosomal abnormalities, and other rare, large de novo CNVs.

(A) Pie chart displaying the proportion for each of these categories. The number of events and inheritance are in parentheses.

(B) Percentage of probands with no ID or with nonsyndromic ASD among carriers of pathogenic CNVs.

(C) Distribution of de novo and inherited deletions and duplications in all CNVs versus in pathogenic CNVs in affected subjects.

(D) Size distribution of pathogenic CNVs.

(E) Gender distribution in all probands (n = 2,446) versus in probands with autosomal pathogenic CNVs (n = 72). Autosomal pathogenic CNVs were partitioned into two categories: highly penetrant CNVs (n = 21) and pathogenic CNVs with variable expressivity and/or incomplete penetrance (n = 48). The male-to-female ratio is shown above each group. The number of affected subjects is shown at the bottom of each bar. The proportion of females was increased among carriers of pathogenic CNVs associated with high penetrance.

RAI1 [MIM 607642]), Kleefstra syndrome (MIM 610253, 9q34.3 deletion including *EHMT1* [MIM 607001]), Williams syndrome (MIM 194050, 7q11.23 deletion), and large chromosomal abnormalities (Figure 2A; Table S7B). Recurrent deleterious CNVs mediated by segmental duplications affecting 12 distinct regions were identified in 44 individuals. For example, two unrelated males were found to harbor Xq28 duplications (MIM 300815), one de novo and one maternal, corresponding to a ~0.3 Mb segmental-duplication-mediated gain (153.2–153.5 Mb), which was previously reported in X-linked ID.[42] *GDI1* (MIM 300104), mutations of which are linked to ID, is the most likely gene involved (Figure S8). Thus, our findings implicate abnormal *GDI1* dosage in ASD. Interestingly, one AGP proband with the duplication had autism and a normal IQ, whereas the second had a borderline IQ (72) (see Table S8 for phenotype information of all affected subjects with pathogenic CNVs). Some other findings include a 1.7 Mb de novo deletion encompassing *ARID1B* (MIM 614556), recently implicated in ID and Coffin-Siris syndrome (MIM 135900), and a small maternally inherited intragenic deletion of *HDAC4* (MIM 605314), involved in brachydactyly-mental-retardation syndrome (MIM 600430; Figure S7). Although many 2q37 deletions have been described in ASD, the deletion found in our proband directly implicates *HDAC4* haploinsufficiency in autism.

In Table S9, we analyzed data across three ASD cohorts, including a total of 5,106 nonoverlapping affected subjects and 3,512 control subjects from the AGP, the Simons Simplex Collection, and the Autism Genetic Resource Exchange (AGRE), for 17 loci and genes commonly reported as implicated in ASD. The most frequent deletions involved 16p11.2 and *NRXN1*, accounting for 0.31% and 0.32% of affected subjects, respectively. Typical 15q11–q13 duplications of the imprinted PWS-AS critical region were found in 0.25% (13/5,106) of affected subjects, reaffirming this region's importance in ASD. The majority of these duplications were of maternal origin, but two were paternally derived (one without information; Table S9). Although paternally derived duplications appear to have incomplete penetrance in comparison to maternal ones, there have been several cases reported in subjects with ASD.[43]

FMRP Targets, PSD Genes, and Other Neuronal Genes Are Implicated in ASD

We expanded our analysis to lists of genes important for neurological function, such as highly-brain-expressed genes, PSD genes,[33] genes implicated in neurological diseases,[44,45] genes with a high pHI,[35] and FMRP targets,[34] the latter of which have been reported to be enriched in de novo LoF SNVs.[7] Our analysis focused on exonic events, and deletions and duplications were analyzed separately (Figures 3A and 3B).

FMRP targets (n = 842) and PSD genes (n = 1,453) carried a significant excess of both deletions and duplications in affected subjects (Figures 3A and 3B). Five percent (73/1,486) of affected subjects with exonic CNVs, including 52 subjects with genes not previously implicated in ASD and/or ID, carried deletions overlapping one or more FMRP targets, yielding 43 ASD candidate genes (Figure 3A; Table S10). Given that the lists of FMRP targets and PSD genes shared 279 genes, we performed conditional analyses showing that the excess of affected subjects carrying deletions overlapping PSD genes was independent of the signal in FMRP targets (OR = 2.62, 95% CI = 1.62–4.32, $p = 2.24 \times 10^{-5}$) and represented 4% of subjects with exonic events (59/1,486) or 3% after exclusion of pathogenic events ($p = 0.007$). Notably, females were overrepresented among affected subjects carrying exonic deletions overlapping FMRP targets (17 females in 73 affected subjects, 1.98-fold more than males, $p = 0.022$, 95% CI = 1.06–3.52).

Brain-expressed genes showed significant excess in affected versus control subjects for deletions only (OR = 1.89, 95% CI = 1.51–2.37, Fisher's exact test $p = 2.6 \times 10^{-8}$; Figures 3A and 3B). Similarly, deletions (and not duplications) overlapping genes implicated in dominant neurological diseases and orthologous genes associated with abnormal phenotypes in heterozygous knockout mice conferred significant increase in ASD risk (OR = 2.94, 95% CI = 1.76–4.93, $p = 2.5 \times 10^{-5}$). Many of the genes implicated in dominant diseases have been related to loss of function or haploinsufficiency, previously suggested to be more frequent and penetrant when deletions rather than duplications are involved.[46] Accordingly, we detected an excess of affected subjects carrying deletions overlapping genes with a high pHI (>0.35) (OR = 1.41, 95% CI = 1.13–1.76, $p = 0.002$).

Increased Multigene Burden in ASD-Affected Subjects

We tested whether multiple genes within a CNV or across unlinked genetic lesions in the same individual might act in concert to increase risk of ASD. In logit modeling, the number of genes overlapped by CNVs, the average brain-expression value for those genes, and the deletion or duplication status were used as predictors with the case-control status as the outcome (Figures 3C and 3D).

We found that ASD risk increased (as measured by the predicted OR) as the numbers of deleted brain-expressed genes increased (generalized linear model chi-square goodness of fit, $p = 3.2 \times 10^{-5}$ and $p = 4.7 \times 10^{-7}$, respectively; Figures 3C and 3D). These results were consistent across the various models tested (Tables S11A–S11E). There was a decrease in signal after removal of affected subjects with at least one de novo event, suggesting that most of the risk can be attributed to de novo CNVs. Notably, the signal further decreased 2-fold when the remaining pathogenic CNVs were removed, confirming that pathogenic inherited CNVs alone also carry risk. Moreover, we found that gene density contributed significantly to increased risk only when de novo CNVs and inherited pathogenic CNVs were considered, whereas a higher-than-average level of gene expression in deletions (but not duplications) was a contributor irrespective of CNV status (i.e., even after

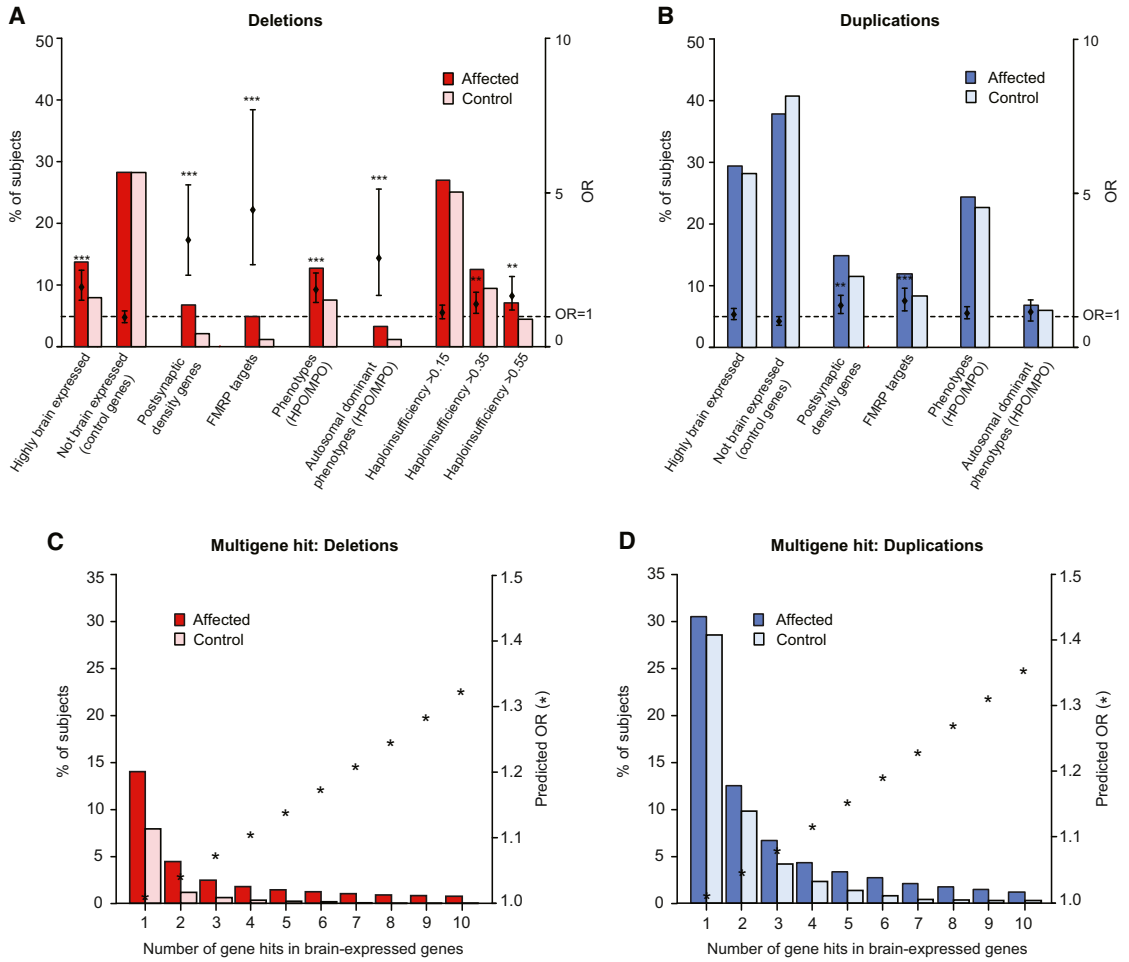


Figure 3. Enrichment of Functional Gene Sets Affected by Rare Exonic CNVs in Affected versus Control Subjects
Overrepresentation of deletions (A) and duplications (B) in various functional gene sets. ORs, with 95% CIs, and the percentages of affected subjects (n = 1,486) and control subjects (n = 1,820) with exonic CNVs overlapping genes are given for the following gene sets: (1) highly-brain-expressed genes (log(RPKM) > 4.5, BrainSpan; n = 5,610); (2) functionally characterized control genes not expressed in the brain (log(RPKM) < 1, BrainSpan; n = 5,410); (3) PSD genes (n = 1,453);[33] (4) FMRP interactors (n = 842);[34] (5) genes associated with neurological phenotypes compiled from the HPO and MPO (n = 3,112); (6) genes as described in (5) but filtered for autosomal-dominant genes (n = 739); and (7) genes grouped by their pHI[35] into three subgroups: pHI > 0.15 (n = 8,862), pHI > 0.35 (n = 4,136,) and pHI > 0.55 (n = 2,214). Genes with a pHI > 0.35 were considered haploinsufficient. The p values for affected and control subjects were estimated with two-tailed Fisher's exact tests (*$p < 0.01$, **$p < 0.001$, ***$p < 0.0001$).
(C–D) Pattern of increased burden as the number of brain-expressed genes affected by deletions (C) or duplications (D) increased. The percentages of affected and control subjects with CNVs overlapping genes are shown for deletions and duplications separately. For estimating the expected ORs (stars), a logit model of case status (affected or control) was fit to covariates, namely CNV status, the number of genes covered by each CNV, and their average brain expression levels (neocortex, BrainSpan). See Tables S11A–S11E for the results of alternative models, all of which showed that ASD risk increased as a function of the number of brain-expressed genes affected by a CNV, even after within-subject dependency of CNVs was accounted for.

removal of both de novo and inherited pathogenic CNVs) (Tables S11A–S11E). Thus, it is likely that de novo and pathogenic CNVs contribute to risk by altering the expression of more than one gene, suggesting that genetic interactions between these genes can underlie ASD risk.

Network Analysis Links Exonic Deletions to Neurodevelopmental Processes

We performed a gene-set enrichment analysis[6] on our expanded sample set after refining our criteria to consider only exonic events (see Supplemental Data for details) and found only deletions to be significantly enriched in gene sets in affected versus control subjects (Figure 4A). We found 86 significantly enriched gene sets, including MAPK signaling components and neuronal synaptic functions and processes, in 42.5% (335/789) of affected subjects with exonic deletions (Figure 4A; Tables S12A–S12D). Enrichment of synaptic functioning has also been reported among inherited events in the AGRE families[47] and among de novo events in the Simons Simplex Collection.[36] Enriched sets delineate candidate genes disrupted by deletions not found in control subjects; these genes notably include those in the KEGG glutamatergic pathway (e.g., *GRIK2* [MIM 138244], *GRM5* [MIM 604102], *SHANK1*

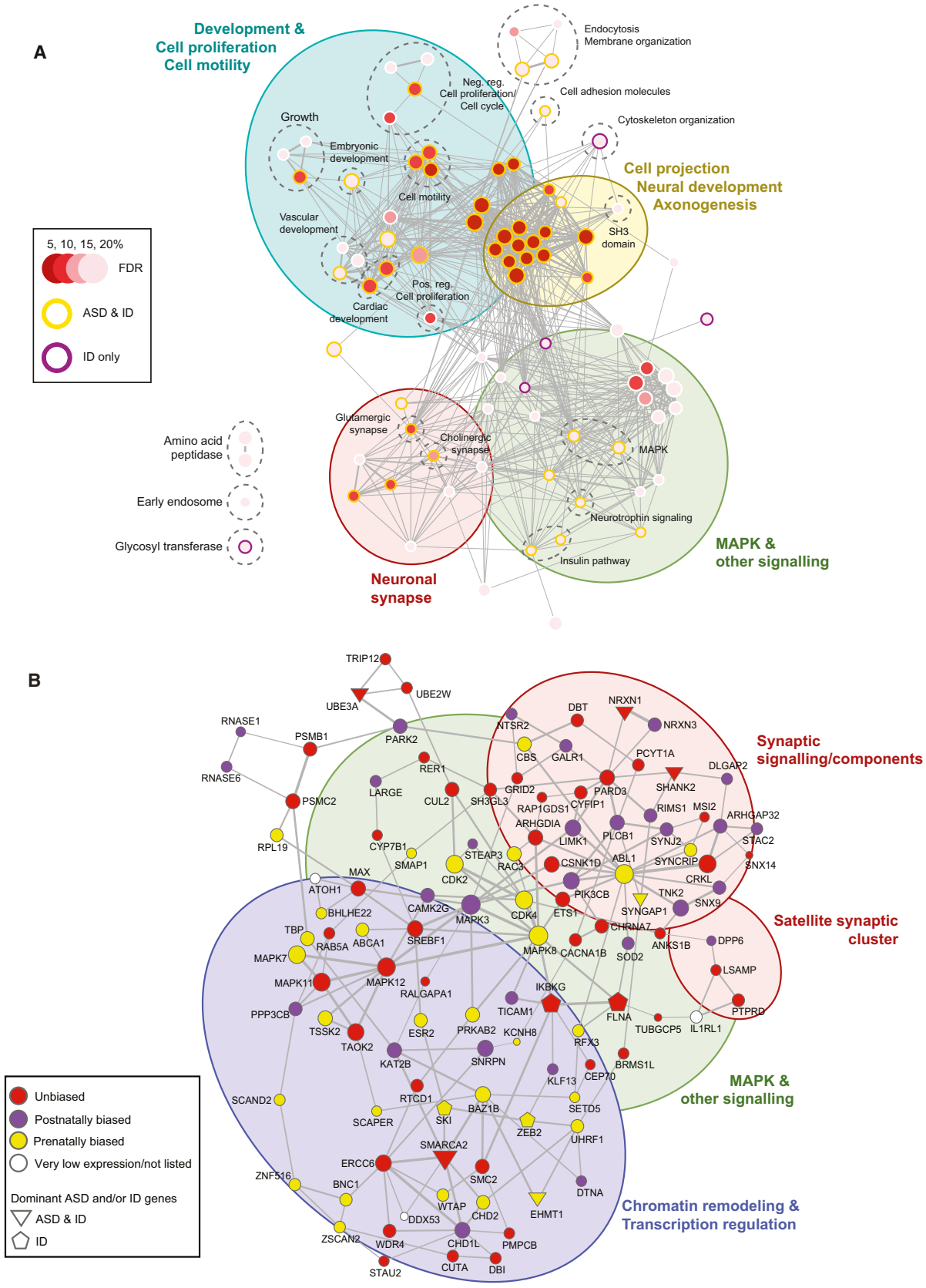


Figure 4. Functional ASD Maps

(A) Gene-set enrichment for rare exonic deletions (de novo and inherited) in affected versus control subjects. Enrichment results were mapped as a network of gene sets (nodes) related by mutual overlap (edges). Node size is proportional to the gene-set size, and edge thickness scales with the number of genes overlapping between sets. Only gene sets enriched in affected subjects with a FDR $\leq$ 20% are

(legend continued on next page)

[MIM 604999], *SHANK2* [MIM 603290], and *SHANK3* [MIM 606230]; Figure S2A and Tables S13A–S13D), the KEGG cholinergic pathway (e.g., *KCNJ12* [MIM 602323], *CHAT* [MIM 118490], and *SLC18A3* [MIM 600336], the latter two of which are within a recurrent 10q11.21–q11.23 deletion, recently reported in individuals with ID and ASD;[48] Figure S2B and Tables S13A–S13D), or in both pathways (e.g., *GNG13* [MIM 607298], *PRKACB* [MIM 176892], *PLCB1* [MIM 607120], *CAMK2G* [MIM 602123], and *PPP3CB* [MIM 114106]). When analyzing human homologs of mouse genes, we also found enrichment of phenotypes mostly related to the brain, including abnormal telencephalon morphology, neuron morphology, behavior, and nervous system physiology (Table S12D).

Genes within De Novo CNVs Cluster in a Gene Network

De novo deletions (52 in European affected subjects) were found to be significantly enriched within each gene set or pathway cluster (for 85 of the 86 gene sets or pathways), as well as across clusters (chi-square test $p = 9.8 \times 10^{-9}$) (Table S12B), prompting us to search for enriched biological functions within de novo events separately. Taking into account our observations of a significant multigene burden in ASD subjects (Figures 3C and 3D), we analyzed de novo events by using NETBAG[36] to identify up to two ASD candidate genes per CNV among 102 de novo events in 99 subjects. NETBAG identifies networks of genes under the premise that if genomic regions are perturbed by genetic variants associated with the same phenotype, they will contain genes forming connected clusters. The NETBAG analysis resulted in a network of 113 genes (global cluster p value = 0.02; Figure 4B). Ten genes have been previously implicated in autosomal-dominant or X-linked forms of ASD and ID (*UBE3A* [MIM 601623], *NRXN1*, *SHANK2*, *EHMT1*, *SYNGAP1* [MIM 603384], and *SMARCA2* [MIM 600014]) or ID only (*ZEB2* [MIM 605802], *FLNA* [MIM 300017], *SKI* [MIM 164780], and *IKBKG* [MIM 300248]). On the basis of cumulative evidence from various sources, an additional 68% (67/98) are likely to affect ASD risk (Tables S14A–S14E); 27/67 of these are either FMRP targets or PSD genes. Compared to all other genes within de novo CNVs (or deletion CNVs only), genes in the network exhibited a significantly higher pHI (Wilcoxon rank sum $p = 7.07 \times 10^{-8}$), and 55% (59/107 [6 without information]) had a pHI > 0.35. A similar NETBAG analysis of de novo CNVs in control subjects did not yield significant results.[36]

We further characterized the biological processes related to the NETBAG cluster (Tables S14B–S14E; Figures 4B and S3) and found a significant enrichment (false-discovery rate [FDR] < 10%) of genes involved in chromatin and transcription regulation, MAPK signaling, and synaptic signaling and components (Figure 4B). We recapitulated many of the results of our gene-set analysis (Figure 4A), notably for synapse functions and processes, and also identified genes involved in chromatin and transcription regulation. The latter category included a high-risk gene associated with ASD, the chromatin gene *CHD2* (MIM 602119), which is affected by a de novo 83 kb deletion in a male with ASD, mild ID, and dysmorphic features including micrognathia and protruding ears. His ASD-affected brother has mild ID, similar dysmorphic features, and epilepsy with onset at age 9 years and carries the same deletion, which removes the first six exons of *CHD2*. Neither parent carries the deletion, suggesting germline mosaicism (the deletion arose on the paternal chromosome). De novo SNVs in *CHD2* have been reported in an ASD subject and in several individuals with a broad spectrum of neurodevelopmental disorders, including ID and epileptic encephalopathy[8,49,50] (Figure S6). Two other genes in the chromodomain family have been linked to neurodevelopmental disorders: *CHD7* (MIM 608892) in CHARGE syndrome (MIM 214800) and *CHD8* (MIM 610528) in ASD. Another example is *TRIP12* (MIM 604506), encoding an E3 ubiquitin ligase that can regulate chromatin function to maintain genome integrity (Figure S12). The chromatin and transcription module showed a predominance of genes with a prenatally biased expression profile (Figures 4B and S4).

De Novo CNVs and LoF SNVs Converge on Functional Gene Networks

We expanded our analysis to genes altered by both our de novo CNVs (Figure 5A) and de novo LoF SNVs compiled from four exome sequencing studies in ASD.[7–10] Eleven genes affected by de novo CNVs (*NRXN1*, *SHANK2*, *ARID1B*, *RIMS1* [MIM 606629], *TRIP12*, *SMARCC2* [MIM 601734], *DLL1* [MIM 606582], *TM4SF19*, *MLL3* [MIM 606833], *PHF2* [MIM 604351], and *CSTF2T* [MIM 611968]) were found to be altered by de novo LoF SNVs among autism cohorts, and three of them (*NRXN1*, *SHANK2*, and *RIMS1*) were selected by NETBAG (Figure 4B).

shown; gene sets are colored by different red intensity scales on the basis of their FDR. The node stroke color (orange or purple) indicates whether the gene set is also enriched with genes known to cause ASD and/or ID. Groups of functionally related gene sets are circled and labeled (groups are filled green or blue circles; subgroups are dashed lines), and the functions of prominent clusters are shown.
(B) Network of genes affected by rare de novo CNVs in affected subjects. Shown are NETBAG results from the analysis of 102 rare de novo CNVs (11 large de novo chromosome abnormalities were not considered; Table S1C), representing 75 nonredundant genic CNV regions. Nodes in the network correspond to genes, and edges correspond to interactions. Node sizes are proportional to the gene's contribution to the overall cluster score. Edge widths are proportional to the prior likelihood that the two corresponding genes contribute to a shared genetic phenotype. Nodes are colored on the basis of whether genes show prenatal- or postnatal-biased brain expression, or have no biased expression, in an analysis of 12 developmental stages of the BrainSpan data set (Figure S4). Shaded ovals represent enriched biological functions (Tables S14A–S14E), and their colors represent functional themes shared among Figures 4A, 4B, and 5B.

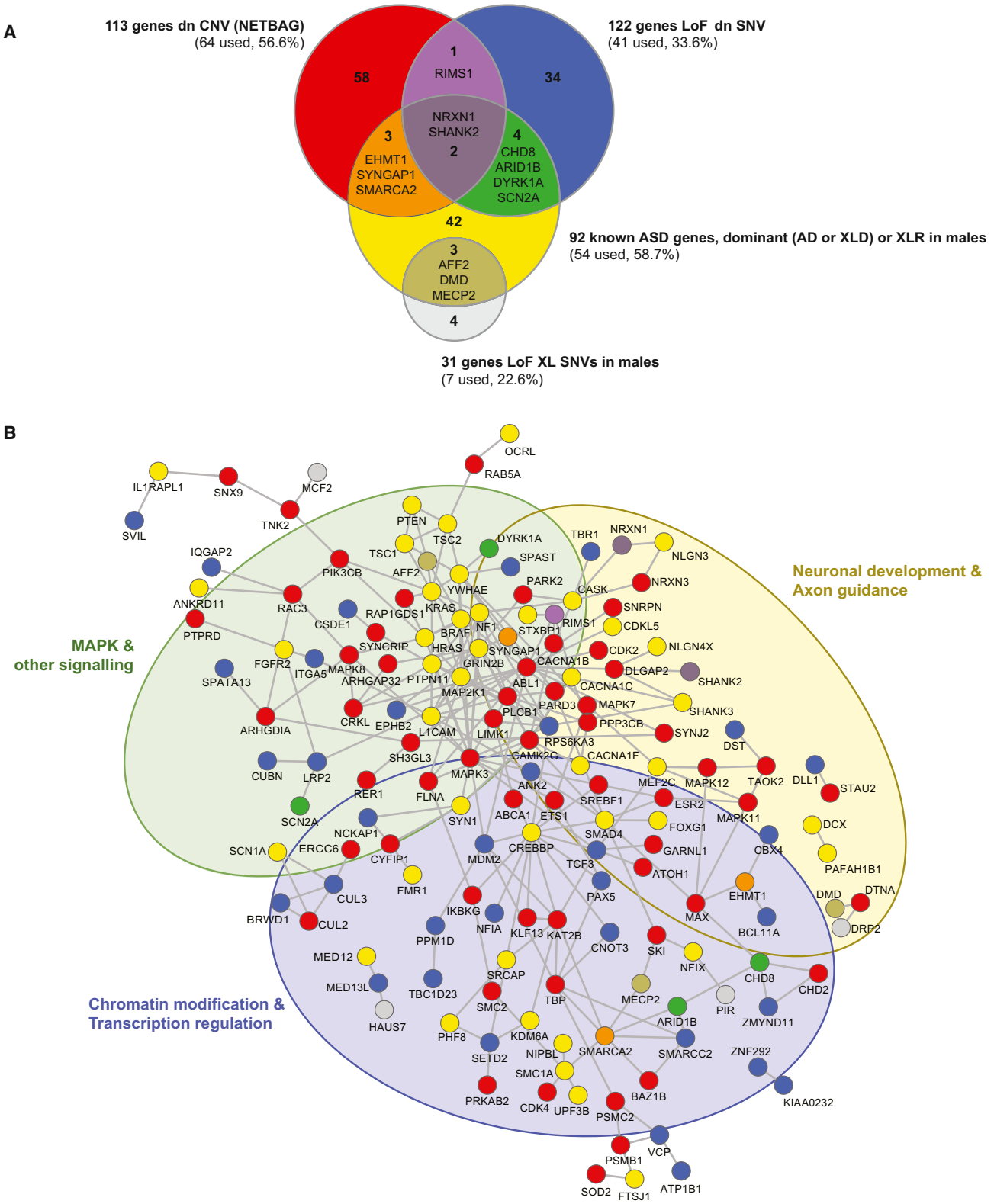


Figure 5. Genes Affected by CNVs and SNVs Converge on Functional Gene Networks

(A) Venn diagram showing the overview of 151 genes resulting from a DAPPLE analysis of 336 unique genes. A similar diagram of DAPPLE input genes is shown in Figure S5. For the DAPPLE analysis, we compiled the following lists of genes: (1) 113 genes identified from our de novo CNVs by NETBAG; (2) 122 genes with de novo LoF SNVs from four published exome sequencing studies;[7–10] (3) 31 genes with hemizygous LoF SNVs on the X chromosome of male ASD subjects and not observed in male control subjects;[16] and (4) 92 ASD-implicated genes previously described as autosomal dominant, X-linked dominant, or X-linked recessive in males.[18]

(B) A DAPPLE network of 151 genes (Table S15) from the genes in (A) shows direct interactions between associated proteins according to the InWeb database. Nodes represent genes and are colored according to gene-set membership depicted in (A): genes identified from our

(legend continued on next page)

Of the 11 genes, only *TM4SF19* (3q29) belongs to a known locus associated with a recurrent genomic disorder. Despite the limited overlap observed among genes altered by de novo CNVs and LoF SNVs, there is reinforcing evidence of the role of *NRXN1* and *SHANK2* in ASD (Figure 5A). In addition, *RIMS1*, altered by both a de novo duplication and a LoF SNV (Figure 5A) and encoding a brain-specific synaptic Rab3a-binding protein, emerges as an ASD candidate gene. *RIMS1* has a regulatory role in synaptic-vesicle exocytosis modulating synaptic transmission and plasticity.[51] Three other genes affected by de novo CNVs (also picked by NETBAG)—*CHD2*, *SYNGAP1*, and *SYNCRIP*—are also affected by LoF de novo SNVs in ID,[50,52] and two (*SYNGAP1* and *DPYD* [MIM 612779]) are altered in schizophrenia.[53] *CHD2* (discussed above) and *SYNGAP1* (encoding a Ras/Rap GTP-activating protein) are both involved in autosomal-dominant ID, ASD, and epilepsy.[54]

Under the assumption that different genes harboring suspected causative mutations for the same disorder often physically interact, we sought to evaluate protein-protein interactions (PPIs) encoded by genes known to be implicated in ASD and genes affected by rare CNVs or SNVs (data drawn from our de novo CNVs and published de novo ASD LoF SNVs; Figures 5A, 5B, and S5). The union set of 336 unique genes (Table S15) analyzed by DAPPLE[37] resulted in a network of direct PPIs encoded by 151 genes (Figure 5A) from each of the three main lists: 54/92 (58.7%) genes involved in ASD, 64/113 (56.6%) de novo CNV genes, and 41/122 (33.6%) de novo LoF SNV genes. The number of direct PPIs was 1.5-fold higher than expected ($p < 0.001$, Figure 5B), suggesting that many of the de novo CNV or SNV genes and ASD-implicated genes function cooperatively. Overrepresentation analysis identified convergent functional themes related to neuronal development and axon guidance, signaling pathways, and chromatin and transcription regulation. Overall, these findings are consistent among the three different types of analyses shown in Figures 4A, 4B, and 5.

Although 54 genes were previously implicated in ASD, the DAPPLE analysis singled out an additional 97 CNV or SNV high-confidence candidate genes (Figure 5B; Table S16). We found that compared to the 54 ASD-implicated genes, the newly selected 97 CNV or SNV genes had a comparably high pHI (median pHI = 0.58, Figure 6A). This is consistent with the observation that ASD subjects have more deletions with haploinsufficient genes than do controls (Figure 6B). Furthermore, similar to genes with known disease-causing mutations (Figure 6C), those genes have high functional-indispensability scores and a comparable high degree of centrality (i.e., high number of direct neighbors) and number of networks in which they are involved (Figures 6D and 6E). Compared to the genome average, they are also among the top 75% of more-conserved genes (on the basis of Genomic Evolutionary Rate Profiling scores and PhyloP) and are highly expressed in the brain. Interestingly, 39 of the 97 genes are either FMRP-related or PSD genes (of the initial 151 genes identified by DAPPLE, 51 are FMRP interactors and 24 are PSD genes). Thus, despite little overlap in genes, the strong interconnectedness between the resulting networks identifies pathways through which the effects of distinct mutations might converge.

Discussion

We used multiple approaches to prioritize key candidate ASD-associated genes disrupted by CNVs and further identified biological relationships and common pathways shared among those genes. Our data (1) confirm excess burden of genome-wide rare genic CNVs in an independent set of ASD subjects versus control subjects; (2) further reveal an extreme degree of etiological heterogeneity (36 different genetic loci were found among 82 individuals with pathogenic CNVs); (3) confirm the contribution of de novo CNVs to the etiology of autism and highlight the contribution of inherited pathogenic imbalances (36%); (4) show an increased proportion of females among carriers of highly penetrant pathogenic CNVs, as well as among carriers of deletions affecting FMRP targets; (5) show no significant difference in the frequency of de novo CNVs between simplex and multiplex families; (6) show that both deletions and duplications involving FMRP targets and PSD genes increase ASD risk; (7) show evidence of multigene contributions to ASD; (8) show that ASD-associated deletions impair synapse function and neurodevelopmental processes; (9) implicate chromatin and transcription regulation genes in ASD in a network analysis of de novo CNVs; and (10) show that genes affected by de novo CNVs and de novo LoF SNVs converge on functional gene networks.

Importantly, when considering highly penetrant pathogenic CNVs, we found a 2:1 male-to-female ratio (deviating from the overall ratio of 6:1 in the study sample). In contrast, the ratio was unchanged among carriers of CNVs characterized by variable expressivity and/or incomplete penetrance. Moreover, among affected subjects, females were twice as likely as males to have exonic deletions involving FMRP targets. Given the sex bias of ASD toward males, it has been suggested that females require

de novo CNVs by NETBAG (red nodes), genes affected by de novo LoF SNVs from published exome sequencing studies (blue nodes), genes affected by hemizygous LoF SNVs on the X chromosome of males (white nodes), and genes known to be implicated in ASD (yellow nodes). Other node colors (orange, purple, green, dark yellow, or dark purple) correspond to genes present in two or more lists. Edges represent significant direct protein-protein interactions (as defined by a common interactor binding degree of 2) in the InWeb database. Shaded ovals represent enriched biological functions common among 10% or more genes in the network, and their colors represent functional themes shared among Figures 4A, 4B, and 5B.

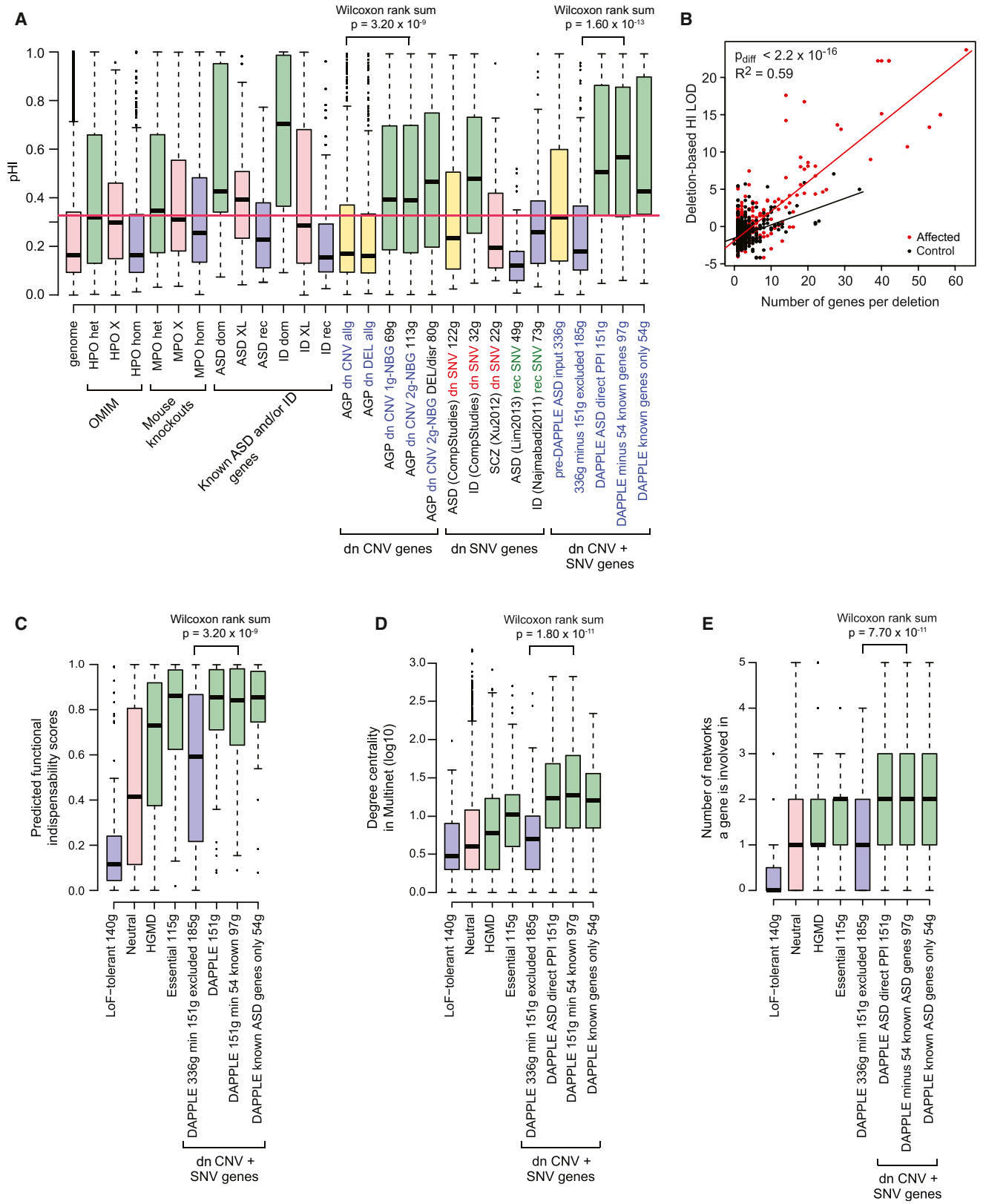


Figure 6. Functional Metrics for Various Gene Sets Derived from CNV and SNV Studies, as well as HI Scores for Genic Deletions
(A) Box plots of pHIs for various genes sets. Boxes correspond to the spread between the upper and lower quartiles; medians are indicated by a solid horizontal line, and whiskers extend up to 1.5× the interquartile range. "Genome" indicates all 16,781 genes with an available pHI from an imputed data set, excluding seed genes. Only genes implicated in dominant, recessive, or X-linked disorders with neurological phenotypes in the HPO database ("HPO het," "HPO hom," "HPO X," respectively) and mouse genes whose homozygous, heterozygous, or X-linked knockout ("MPO het," "MPO hom," "MPO X," respectively) causes various abnormal phenotypes were considered.

(legend continued on next page)

a higher genetic load to express ASD,[56] and our foundational data support this general hypothesis. This same phenomenon has recently been shown for *SHANK1*[57] and the 16p13.11 CNV.[58]

We observed significant enrichment of both deletion and duplication events overlapping FMRP and PSD targets, indicating that altered dosage of these genes can underlie ASD susceptibility. This is consistent with evidence that FMRP targets belong to multiple signaling and interconnected pathways such as PI3K-Akt-TSC-PTEN-mTOR and PI3K-RAS-MAPK,[59,60] which have been linked to ASD through both underexpression and overexpression of genes in these pathways. Although both deletions and duplications can contribute to risk, we found that deletions can have a stronger impact when highly-brain-expressed genes, genes conferring dominant phenotypes in humans and mice, or genes with a high pHI are considered. We also found that ASD risk increases as a function of the number of brain-expressed genes affected by rare de novo and pathogenic CNVs, consistent with an additive model of risk underlying ASD etiology.

We developed an expanded and extensively interconnected network of high-confidence ASD candidate genes by integrating protein products from CNV and SNV genes and ASD-implicated genes. Overall, these results demonstrate that genes involved in ASD participate in a wide array of processes, from neuronal development and axon guidance to MAPK and other kinase signaling cascades (including the PI3K-Akt-mTOR and PI3K-RAS-MAPK pathways) to chromatin modification and transcription regulation. An increasing number of genes involved in chromatin structure and epigenetic regulation have been implicated in a variety of developmental disorders.[61] Other chromatin regulator genes, such as *MBD5* (MIM 611472) and *KMT2D* (MIM 602113), have been implicated in ID and ASD, highlighting the need to further study this category of genes as ASD risk factors.

In addition to underlining important pathways, our results highlight specific genes in ASD risk. Whereas the majority of the 97 genes in the CNV or SNV network (not including the genes already known to be involved in ASD) most likely act via haploinsufficiency, a few are affected by duplications. One example is the duplication of *PIK3CB* (MIM 602925), which is likely to increase its expression and thus lead to excessive phosphatidylinositol 3-kinase (PI3K) activity. PI3K, which is regulated by FMRP,[59] is elevated in FXS mouse knockouts,[62,63] and downregulation of this pathway has been shown to have therapeutic effect in ASD and FXS mouse models. *RAC3* (MIM 602050), another example of a gene affected by duplication, encodes a Rho family GTPase that enhances neuritogenesis and neurite branching when overexpressed.[64]

Our findings implicate many ASD candidate genes altered by de novo, inherited, or X-linked CNVs (e.g., *SETD5*, *miR137*, and *HDAC9* [MIM 606543]; Supplemental Data section "Highlighted Genes") or altered by both de novo CNVs and LoF SNVs (e.g., *RIMS1*, *TRIP12*, and *DLL1*; Figures 4B and 5). Taken together, our results suggest that rare variants affecting ASD risk in the population collectively encompass hundreds of genes. Despite this heterogeneity, many genes converge in interconnected functional modules, providing diagnostic and therapeutic targets.

Supplemental Data

Supplemental Data include 18 figures, 19 tables, and Supplemental Acknowledgments and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2014.03.018.

The median pHI for HPO het was selected as the threshold to differentiate between dominant and recessive genes (red horizontal line). Genes implicated in ASD and ID were further annotated into dominant (dom), recessive (rec), or X-linked (XL) genes. Other abbreviations are as follows: dn, de novo; allg; all genes; DEL, deletion; "1g-NBG 69g," 69 genes selected by NETBAG analysis of 102 de novo CNVs with up to one gene per each CNV region; "2g-NBG 113g," 113 genes selected by NETBAG analysis with up to two genes per CNV (as depicted in Figure 4B); "2g-NBG DEL/disr 80g," subset of NETBAG genes completely overlapped or disrupted by deletions (no duplications were considered); "ASD (CompStudies) dn SNV 122g," 122 genes affected by de novo LoF SNVs from ASD exome sequencing studies; "ID (CompStudies) dn SNV 32g," 32 genes affected by de novo LoF SNVs from ID exome sequencing studies; "SCZ (Xu2012) dn SNV 22g," 22 genes affected by de novo LoF SNVs from schizophrenia exome sequencing studies; "ASD (Lim2013) rec SNV 49g," 49 genes affected by hemizygous LoF SNVs on the X chromosome of ASD males; "ID (Najmabadi2011) rec SNV 73g," 73 genes hit by recessive SNVs in consanguineous ID-affected families; "pre-DAPPLE ASD input 336g," 336 DAPPLE input genes; "336g minus 151g excluded 185g," 185 genes not used by DAPPLE; "DAPPLE ASD direct-PPI 151g," 151 genes depicted in the network of Figure 5B; "DAPPLE minus 54 known genes 97g," 97 genes depicted in the network of Figure 5B (and listed in Table S16), not including the 54 genes previously implicated in ASD (yellow nodes); and "DAPPLE known genes only 54g," 54 genes known to be involved in ASD.

(B) LOD scores of the probability that at least one gene within a rare deletion will cause haploinsufficiency were calculated for affected and control subjects. Deletion-based LOD scores are plotted as a function of the number of genes in each event for rare genic deletions in affected and control subjects. The p value for the difference in the slope of the two regression lines is indicated.

(C) Box plots with the distribution of predicted functional indispensability scores for gene categories from Khurana et al.[55] (LoF-tolerant genes, neutral genes, genes with known mutations as listed in the Human Genome Mutation Database, and essential genes [i.e., genes in which LoF mutations result in infertility or death before puberty]) and CNV or SNV genes from our DAPPLE analysis (185 genes excluded by DAPPLE, 151 genes selected by DAPPLE, 54 known ASD-implicated genes selected by DAPPLE, and 97 genes selected by DAPPLE after exclusion of the 54 ASD-implicated genes).

(D) Box plots with the distributions of degree centrality in Multinet[55] for the same gene categories as in (C).

(E) Box plots with the distributions of the number of networks in which a gene is involved in Multinet for the same gene categories as in (C) and (D).

Acknowledgments

The authors thank the main funders of the Autism Genome Project: Autism Speaks (USA), the Health Research Board (Ireland; AUT/2006/1, AUT/2006/2, PD/2006/48), the Medical Research Council (UK), the Hilibrand Foundation (USA), Genome Canada, the Ontario Genomics Institute, and the Canadian Institutes of Health Research (CIHR). Additional support for individual groups is shown in the Supplemental Acknowledgments. D.P. is the Abraham & Mildred Goldstein Seaver Center Faculty Fellow, J.D.B. holds the G. Harold and Leila Y. Mathers Professorship, C.B. is the recipient of a NARSAD Independent Investigator Grant from the Brain & Behavior Research Foundation, and S.W.S. holds the GlaxoSmithKline-CIHR Pathfinder Chair in Genome Sciences at the University of Toronto and The Hospital for Sick Children. E.H.C. is an advisor of Seaside Therapeutics, G.D. is a member of the Scientific Advisory Board at Integragen Inc., and S.W.S. is an advisor to Population Diagnostics and advisor and founder of YouNique Genomics. D.P., P.S., J.S.S., J.H., M.G., E.H.C., J.D.B., B.D., L.G., C.B., and S.W.S were leading contributors to the design and analysis of this study, and D.P., E. Delaby, D.M., M.B., J.D.B., B.D., L.G., C.B., and S.W.S wrote the manuscript.

Received: November 15, 2013
Accepted: March 25, 2014
Published: April 24, 2014

Web Resources

The URLs for data presented herein are as follows:

BrainSpan, http://brainspan.org/
The ConSurf Server, http://consurf.tau.ac.il
Database of Genomic Variants (DGV), http://dgv.tcag.ca/dgv/app/home
dbGaP, http://www.ncbi.nlm.nih.gov/gap
DECIPHER, http://decipher.sanger.ac.uk
European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations (ECARUCA), http://www.ecaruca.net
Genomic Evolutionary Rate Profiling (GERP), http://mendel.stanford.edu/SidowLab/downloads/gerp/index.html
International Standards for Cytogenomic Arrays (ISCA) Consortium, https://www.iscaconsortium.org
MutationTaster, http://www.mutationtaster.org
NHLBI Exome Sequencing Project (ESP) Exome Variant Server, http://evs.gs.washington.edu/EVS/
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org
PANTHER, http://www.pantherdb.org
PolyPhen-2, http://genetics.bwh.harvard.edu/pph2
SIFT, http://sift.jcvi.org/
SNAP, https://rostlab.org/services/snap
UCSC Genome Browser, http://genome.ucsc.edu

Accession Numbers

The dbGaP accession number for the raw data from the ASD-affected families is phs0000267.v4.

References

1. Constantino, J.N., Todorov, A., Hilton, C., Law, P., Zhang, Y., Molloy, E., Fitzgerald, R., and Geschwind, D. (2013). Autism recurrence in half siblings: strong support for genetic mechanisms of transmission in ASD. Mol. Psychiatry *18*, 137–138.
2. Levy, D., Ronemus, M., Yamrom, B., Lee, Y.H., Leotta, A., Kendall, J., Marks, S., Lakshmi, B., Pai, D., Ye, K., et al. (2011). Rare de novo and transmitted copy-number variation in autistic spectrum disorders. Neuron *70*, 886–897.
3. Marshall, C.R., Noor, A., Vincent, J.B., Lionel, A.C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y., et al. (2008). Structural variation of chromosomes in autism spectrum disorder. Am. J. Hum. Genet. *82*, 477–488.
4. Sanders, S.J., Ercan-Sencicek, A.G., Hus, V., Luo, R., Murtha, M.T., Moreno-De-Luca, D., Chu, S.H., Moreau, M.P., Gupta, A.R., Thomson, S.A., et al. (2011). Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron *70*, 863–885.
5. Szatmari, P., Paterson, A.D., Zwaigenbaum, L., Roberts, W., Brian, J., Liu, X.Q., Vincent, J.B., Skaug, J.L., Thompson, A.P., Senman, L., et al.; Autism Genome Project Consortium (2007). Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat. Genet. *39*, 319–328.
6. Pinto, D., Pagnamenta, A.T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S., et al. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. Nature *466*, 368–372.
7. Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y.H., Narzisi, G., Leotta, A., et al. (2012). De novo gene disruptions in children on the autistic spectrum. Neuron *74*, 285–299.
8. Neale, B.M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K.E., Sabo, A., Lin, C.F., Stevens, C., Wang, L.S., Makarov, V., et al. (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature *485*, 242–245.
9. Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature *485*, 237–241.
10. O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D., et al. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature *485*, 246–250.
11. Yu, T.W., Chahrour, M.H., Coulter, M.E., Jiralerspong, S., Okamura-Ikeda, K., Ataman, B., Schmitz-Abe, K., Harmin, D.A., Adli, M., Malik, A.N., et al. (2013). Using whole-exome sequencing to identify inherited causes of autism. Neuron *77*, 259–273.
12. Béna, F., Bruno, D.L., Eriksson, M., van Ravenswaaij-Arts, C., Stark, Z., Dijkhuizen, T., Gerkes, E., Gimelli, S., Ganesamoorthy, D., Thuresson, A.C., et al. (2013). Molecular and clinical characterization of 25 individuals with exonic deletions of NRXN1 and comprehensive review of the literature. Am. J. Med. Genet. B. Neuropsychiatr. Genet. *162B*, 388–403.
13. Moreno-De-Luca, D., Sanders, S.J., Willsey, A.J., Mulle, J.G., Lowe, J.K., Geschwind, D.H., State, M.W., Martin, C.L., and Ledbetter, D.H. (2013). Using large clinical data sets to infer pathogenicity for rare copy number variants in autism cohorts. Mol. Psychiatry *18*, 1090–1095.
14. Noor, A., Whibley, A., Marshall, C.R., Gianakopoulos, P.J., Piton, A., Carson, A.R., Orlic-Milacic, M., Lionel, A.C.,

Sato, D., Pinto, D., et al.; Autism Genome Project Consortium (2010). Disruption at the PTCHD1 Locus on Xp22.11 in Autism spectrum disorder and intellectual disability. Sci. Transl. Med. *2*, 49ra68.
15. Jiang, Y.H., Yuen, R.K., Jin, X., Wang, M., Chen, N., Wu, X., Ju, J., Mei, J., Shi, Y., He, M., et al. (2013). Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. Am. J. Hum. Genet. *93*, 249–263.
16. Lim, E.T., Raychaudhuri, S., Sanders, S.J., Stevens, C., Sabo, A., MacArthur, D.G., Neale, B.M., Kirby, A., Ruderfer, D.M., Fromer, M., et al.; NHLBI Exome Sequencing Project (2013). Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. Neuron *77*, 235–242.
17. Devlin, B., and Scherer, S.W. (2012). Genetic architecture in autism spectrum disorder. Curr. Opin. Genet. Dev. *22*, 229–237.
18. Betancur, C. (2011). Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. Brain Res. *1380*, 42–77.
19. Buxbaum, J.D., Daly, M.J., Devlin, B., Lehner, T., Roeder, K., and State, M.W.; Autism Sequencing Consortium (2012). The autism sequencing consortium: large-scale, high-throughput sequencing in autism spectrum disorders. Neuron *76*, 1052–1056.
20. Anney, R., Klei, L., Pinto, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S., Sykes, N., Pagnamenta, A.T., et al. (2010). A genome-wide scan for common alleles affecting risk for autism. Hum. Mol. Genet. *19*, 4072–4082.
21. Anney, R., Klei, L., Pinto, D., Almeida, J., Bacchelli, E., Baird, G., Bolshakova, N., Bölte, S., Bolton, P.F., Bourgeron, T., et al. (2012). Individual common variants exert weak effects on the risk for autism spectrum disorders. Hum. Mol. Genet. *21*, 4781–4792.
22. Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., and Schopler, E. (1989). Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. J. Autism Dev. Disord. *19*, 185–212.
23. Lord, C., Rutter, M., and Le Couteur, A. (1994). 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*, 659–685.
24. Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., Cook, E.H., Jr., Leventhal, B.L., and Pickles, A. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorders. J. Am. Acad. Child Adolesc. Psychiatry *45*, 1094–1103.
25. Sparrow, S.S., Cicchetti, D.V., and Balla, D.A. (2005). Vineland Adaptive Behavior Scales, Second edition (Vineland-II) (San Antonio: Pearson).
26. Bierut, L.J., Agrawal, A., Bucholz, K.K., Doheny, K.F., Laurie, C., Pugh, E., Fisher, S., Fox, L., Howells, W., Bertelsen, S., et al.; Gene, Environment Association Studies Consortium (2010). A genome-wide association study of alcohol dependence. Proc. Natl. Acad. Sci. USA *107*, 5082–5087.
27. Figueiredo, J.C., Lewinger, J.P., Song, C., Campbell, P.T., Conti, D.V., Edlund, C.K., Duggan, D.J., Rangrej, J., Lemire, M., Hudson, T., et al. (2011). Genotype-environment interactions in microsatellite stable/microsatellite instability-low colorectal cancer: results from a genome-wide association study. Cancer Epidemiol. Biomarkers Prev. *20*, 758–766.
28. Newcomb, P.A., Baron, J., Cotterchio, M., Gallinger, S., Grove, J., Haile, R., Hall, D., Hopper, J.L., Jass, J., Le Marchand, L., et al.; Colon Cancer Family Registry (2007). Colon Cancer Family Registry: an international resource for studies of the genetic epidemiology of colon cancer. Cancer Epidemiol. Biomarkers Prev. *16*, 2331–2343.
29. Fox, C.S., Liu, Y., White, C.C., Feitosa, M., Smith, A.V., Heard-Costa, N., Lohman, K., Johnson, A.D., Foster, M.C., Greenawalt, D.M., et al.; GIANT Consortium; MAGIC Consortium; GLGC Consortium (2012). Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women. PLoS Genet. *8*, e1002695.
30. Pinto, D., Darvishi, K., Shi, X., Rajan, D., Rigler, D., Fitzgerald, T., Lionel, A.C., Thiruvahindrapuram, B., Macdonald, J.R., Mills, R., et al. (2011). Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. Nat. Biotechnol. *29*, 512–520.
31. Kirov, G., Pocklington, A.J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., Moran, J., Chambert, K., Toncheva, D., Georgieva, L., et al. (2012). De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol. Psychiatry *17*, 142–153.
32. Kearney, H.M., Thorland, E.C., Brown, K.K., Quintero-Rivera, F., and South, S.T.; Working Group of the American College of Medical Genetics Laboratory Quality Assurance Committee (2011). American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. Genet. Med. *13*, 680–685.
33. Bayés, A., van de Lagemaat, L.N., Collins, M.O., Croning, M.D., Whittle, I.R., Choudhary, J.S., and Grant, S.G. (2011). Characterization of the proteome, diseases and evolution of the human postsynaptic density. Nat. Neurosci. *14*, 19–21.
34. Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y., Mele, A., Fraser, C.E., Stone, E.F., Chen, C., Fak, J.J., Chi, S.W., et al. (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell *146*, 247–261.
35. Huang, N., Lee, I., Marcotte, E.M., and Hurles, M.E. (2010). Characterising and predicting haploinsufficiency in the human genome. PLoS Genet. *6*, e1001154.
36. Gilman, S.R., Iossifov, I., Levy, D., Ronemus, M., Wigler, M., and Vitkup, D. (2011). Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. Neuron *70*, 898–907.
37. Rossin, E.J., Lage, K., Raychaudhuri, S., Xavier, R.J., Tatar, D., Benita, Y., Cotsapas, C., and Daly, M.J.; International Inflammatory Bowel Disease Genetics Constortium (2011). Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. PLoS Genet. *7*, e1001273.
38. Malhotra, D., McCarthy, S., Michaelson, J.J., Vacic, V., Burdick, K.E., Yoon, S., Cichon, S., Corvin, A., Gary, S., Gershon, E.S., et al. (2011). High frequencies of de novo CNVs in bipolar disorder and schizophrenia. Neuron *72*, 951–963.
39. Carter, M.T., Nikkel, S.M., Fernandez, B.A., Marshall, C.R., Noor, A., Lionel, A.C., Prasad, A., Pinto, D., Joseph-George, A.M., Noakes, C., et al. (2011). Hemizygous deletions on chromosome 1p21.3 involving the DPYD gene in individuals with autism spectrum disorder. Clin. Genet. *80*, 435–443.

40. Willemsen, M.H., Vallès, A., Kirkels, L.A., Mastebroek, M., Olde Loohuis, N., Kos, A., Wissink-Lindhout, W.M., de Brouwer, A.P., Nillesen, W.M., Pfundt, R., et al. (2011). Chromosome 1p21.3 microdeletions comprising DPYD and MIR137 are associated with intellectual disability. J. Med. Genet. *48*, 810–818.
41. Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium (2011). Genome-wide association study identifies five new schizophrenia loci. Nat. Genet. *43*, 969–976.
42. Vandewalle, J., Van Esch, H., Govaerts, K., Verbeeck, J., Zweier, C., Madrigal, I., Mila, M., Pijkels, E., Fernandez, I., Kohlhase, J., et al. (2009). Dosage-dependent severity of the phenotype in patients with mental retardation due to a recurrent copy-number gain at Xq28 mediated by an unusual recombination. Am. J. Hum. Genet. *85*, 809–822.
43. Urraca, N., Cleary, J., Brewer, V., Pivnick, E.K., McVicar, K., Thibert, R.L., Schanen, N.C., Esmer, C., Lamport, D., and Reiter, L.T. (2013). The interstitial duplication 15q11.2-q13 syndrome includes autism, mild facial anomalies and a characteristic EEG signature. Autism Res. *6*, 268–279.
44. Robinson, P.N., Köhler, S., Bauer, S., Seelow, D., Horn, D., and Mundlos, S. (2008). The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. Am. J. Hum. Genet. *83*, 610–615.
45. Smith, C.L., Goldsmith, C.A., and Eppig, J.T. (2005). The Mammalian Phenotype Ontology as a tool for annotating, analyzing and comparing phenotypic information. Genome Biol. *6*, R7.
46. Turner, D.J., Miretti, M., Rajan, D., Fiegler, H., Carter, N.P., Blayney, M.L., Beck, S., and Hurles, M.E. (2008). Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. Nat. Genet. *40*, 90–95.
47. Gai, X., Xie, H.M., Perin, J.C., Takahashi, N., Murphy, K., Wenocur, A.S., D'arcy, M., O'Hara, R.J., Goldmuntz, E., Grice, D.E., et al. (2012). Rare structural variation of synapse and neurotransmission genes in autism. Mol. Psychiatry *17*, 402–411.
48. Stankiewicz, P., Kulkarni, S., Dharmadhikari, A.V., Sampath, S., Bhatt, S.S., Shaikh, T.H., Xia, Z., Pursley, A.N., Cooper, M.L., Shinawi, M., et al. (2012). Recurrent deletions and reciprocal duplications of 10q11.21q11.23 including CHAT and SLC18A3 are likely mediated by complex low-copy repeats. Hum. Mutat. *33*, 165–179.
49. Carvill, G.L., Heavin, S.B., Yendle, S.C., McMahon, J.M., O'Roak, B.J., Cook, J., Khan, A., Dorschner, M.O., Weaver, M., Calvert, S., et al. (2013). Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. Nat. Genet. *45*, 825–830.
50. Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endele, S., Schwarzmayr, T., Albrecht, B., Bartholdi, D., Beygo, J., Di Donato, N., et al. (2012). Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet *380*, 1674–1682.
51. Schoch, S., Castillo, P.E., Jo, T., Mukherjee, K., Geppert, M., Wang, Y., Schmitz, F., Malenka, R.C., and Südhof, T.C. (2002). RIM1alpha forms a protein scaffold for regulating neurotransmitter release at the active zone. Nature *415*, 321–326.
52. de Ligt, J., Willemsen, M.H., van Bon, B.W., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., de Vries, P., Gilissen, C., et al. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. N. Engl. J. Med. *367*, 1921–1929.
53. Xu, B., Ionita-Laza, I., Roos, J.L., Boone, B., Woodrick, S., Sun, Y., Levy, S., Gogos, J.A., and Karayiorgou, M. (2012). De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. Nat. Genet. *44*, 1365–1369.
54. Berryer, M.H., Hamdan, F.F., Klitten, L.L., Møller, R.S., Carmant, L., Schwartzentruber, J., Patry, L., Dobrzeniecka, S., Rochefort, D., Neugnot-Cerioli, M., et al. (2013). Mutations in SYNGAP1 cause intellectual disability, autism, and a specific form of epilepsy by inducing haploinsufficiency. Hum. Mutat. *34*, 385–394.
55. Khurana, E., Fu, Y., Chen, J., and Gerstein, M. (2013). Interpretation of genomic variants using a unified biological network approach. PLoS Comput. Biol. *9*, e1002886.
56. Robinson, E.B., Lichtenstein, P., Anckarsäter, H., Happé, F., and Ronald, A. (2013). Examining and interpreting the female protective effect against autistic behavior. Proc. Natl. Acad. Sci. USA *110*, 5258–5262.
57. Sato, D., Lionel, A.C., Leblond, C.S., Prasad, A., Pinto, D., Walker, S., O'Connor, I., Russell, C., Drmic, I.E., Hamdan, F.F., et al. (2012). SHANK1 deletions in males with autism spectrum disorder. Am. J. Hum. Genet. *90*, 879–887.
58. Tropeano, M., Ahn, J.W., Dobson, R.J., Breen, G., Rucker, J., Dixit, A., Pal, D.K., McGuffin, P., Farmer, A., White, P.S., et al. (2013). Male-biased autosomal effect of 16p13.11 copy number variation in neurodevelopmental disorders. PLoS ONE *8*, e61365.
59. Ascano, M., Jr., Mukherjee, N., Bandaru, P., Miller, J.B., Nusbaum, J.D., Corcoran, D.L., Langlois, C., Munschauer, M., Dewell, S., Hafner, M., et al. (2012). FMRP targets distinct mRNA sequence elements to regulate protein expression. Nature *492*, 382–386.
60. Darnell, J.C. (2011). Defects in translational regulation contributing to human cognitive and behavioral disease. Curr. Opin. Genet. Dev. *21*, 465–473.
61. Day, J.J., and Sweatt, J.D. (2011). Epigenetic mechanisms in cognition. Neuron *70*, 813–829.
62. Gross, C., Nakamoto, M., Yao, X., Chan, C.B., Yim, S.Y., Ye, K., Warren, S.T., and Bassell, G.J. (2010). Excess phosphoinositide 3-kinase subunit synthesis and activity as a novel therapeutic target in fragile X syndrome. J. Neurosci. *30*, 10624–10638.
63. Sharma, A., Hoeffer, C.A., Takayasu, Y., Miyawaki, T., McBride, S.M., Klann, E., and Zukin, R.S. (2010). Dysregulation of mTOR signaling in fragile X syndrome. J. Neurosci. *30*, 694–702.
64. Albertinazzi, C., Gilardelli, D., Paris, S., Longhi, R., and de Curtis, I. (1998). Overexpression of a neural-specific rho family GTPase, cRac1B, selectively induces enhanced neuritogenesis and neurite branching in primary neurons. J. Cell Biol. *142*, 815–825.

The American Journal of Human Genetics, Volume *94*
Supplemental Data

Convergence of Genes and Cellular Pathways Dysregulated in Autism Spectrum Disorders

Dalila Pinto, Elsa Delaby, Daniele Merico, Mafalda Barbosa, Alison Merikangas, Lambertus Klei, Bhooma Thiruvahindrapuram, Xiao Xu, Robert Ziman, Zhuozhi Wang, Jacob A.S. Vorstman, Ann Thompson, Regina Regan, Marion Pilorge, Giovanna Pellecchia, Alistair T. Pagnamenta, Bárbara Oliveira, Christian R. Marshall, Tiago R. Magalhaes, Jennifer K. Lowe, Jennifer L. Howe, Anthony J. Griswold, John Gilbert, Eftichia Duketis, Beth A. Dombroski, Maretha V. De Jonge, Michael Cuccaro, Emily L. Crawford, Catarina T. Correia, Judith Conroy, Inês C. Conceição, Andreas G. Chiocchetti, Jillian P. Casey, Guiqing Cai, Christelle Cabrol, Nadia Bolshakova, Elena Bacchelli, Richard Anney, Steven Gallinger, Michelle Cotterchio, Graham Casey, Lonnie Zwaigenbaum, Kerstin Wittemeyer, Kirsty Wing, Simon Wallace, Herman van Engeland, Ana Tryfon, Susanne Thomson, Latha Soorya, Bernadette Rogé, Wendy Roberts, Fritz Poustka, Susana Mouga, Nancy Minshew, L. Alison McInnes, Susan G. McGrew, Catherine Lord, Marion Leboyer, Ann S. Le Couteur, Alexander Kolevzon, Patricia Jiménez González, Suma Jacob, Richard Holt, Stephen Guter, Jonathan Green, Andrew Green, Christopher Gillberg, Bridget A. Fernandez, Frederico Duque, Richard Delorme, Geraldine Dawson, Pauline Chaste, Cátia Café, Sean Brennan, Thomas Bourgeron, Patrick F. Bolton, Sven Bölte, Raphael Bernier, Gillian Baird, Anthony J. Bailey, Evdokia Anagnostou, Joana Almeida, Ellen M. Wijsman, Veronica J. Vieland, Astrid M. Vicente, Gerard D. Schellenberg, Margaret Pericak-Vance, Andrew D. Paterson, Jeremy R. Parr, Guiomar Oliveira, John I. Nurnberger, Anthony P. Monaco, Elena Maestrini, Sabine M. Klauck, Hakon Hakonarson, Jonathan L. Haines, Daniel H. Geschwind, Christine M. Freitag, Susan E. Folstein, Sean Ennis, Hilary Coon, Agatino Battaglia, Peter Szatmari, James S. Sutcliffe, Joachim Hallmayer, Michael Gill, Edwin H. Cook, Joseph D. Buxbaum, Bernie Devlin, Louise Gallagher, Catalina Betancur, and Stephen W. Scherer

CONTENTS

SUPPLEMENTAL FIGURES

HIGHLIGHTED GENES

SUPPLEMENTAL TABLES

SUPPLEMENTAL DATAFILES

SUPPLEMENTAL FIGURES

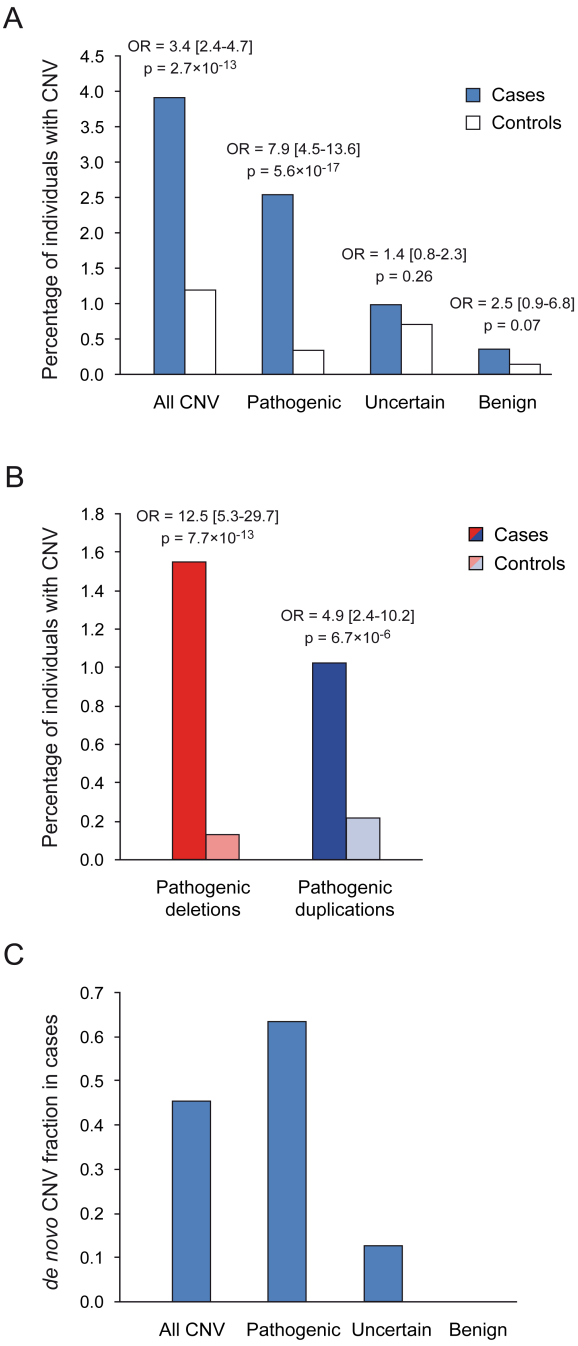
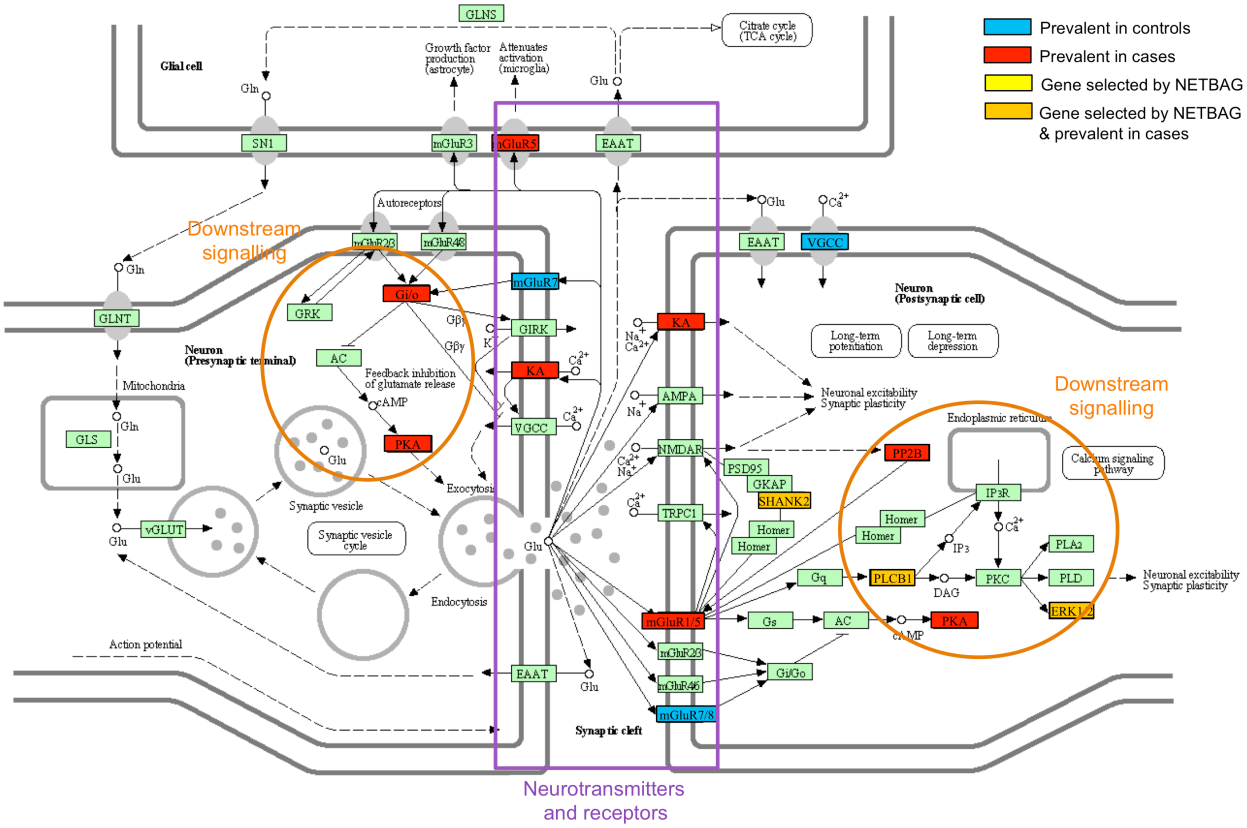


Figure S1. CNV burden in genes and loci implicated in ASD and ID in affected and control subjects of all ancestries

CNV data from 2,446 cases and 4,768 controls irrespective of ancestry were analyzed for overlaps with a list of genes and loci causally implicated in ASD/ID (**Tables S6A-S6D**). Overlap results include 299 non-European cases and 1,843 non-Europeans controls. Only CNVs affecting autosomal dominant and X-linked dominant genes/loci in both genders (132 genes, 56 loci) as well as X-linked recessive genes/loci in males (52 genes, 2 loci) were considered ('all CNV'). Exonic CNVs ≥30 kb affecting an ASD/ID gene and CNVs overlapping at least 50% of the target loci were selected for further analysis. After curation, CNVs were divided into three categories: pathogenic, uncertain clinical significance or benign. **(A)** Percentage of individuals with CNVs overlapping genes and loci implicated in ASD/ID ('all CNV'), pathogenic, uncertain or benign CNVs and odds ratio (OR) in cases and controls. **(B)** Percentage of individuals with pathogenic deletions or duplications and OR in cases and controls. **(C)** *De novo* CNV fraction in each category in cases.

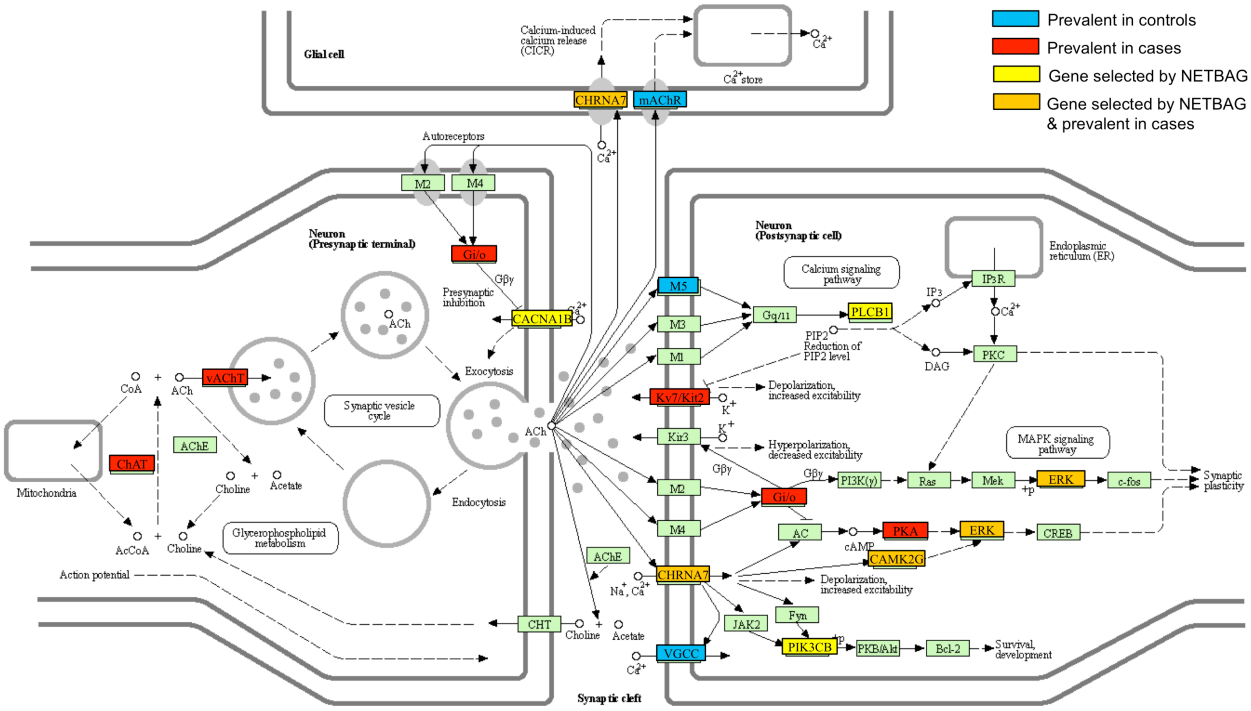
Figure S2. Projection of the glutamergic (A) and cholinergic (B) synapse subclusters enriched in cases over controls, onto the KEGG pathways

A. Glutamergic pathway



19 cases and 2 controls in the glutamatergic pathway: 11 deleted genes in 19 cases (average size= 635 kb, median size= 265 kb; average number of genes= 16, median number of genes= 3): *GNG13* (1), *MAPK3* (5), *GNG2* (2), *PRKACB* (1), *SHANK2* (3), *PPP3CB* (1), *SHANK3* (3), *GRIK2* (1), *GRM5* (1), *PLCB1* (1), *SHANK1* (1); 2 deleted genes in 2 controls (average/median size= 85 kb; average/median number of genes= 1): *CACNA1C*, *GRM7*. Genes selected by NETBAG (main **Figure 4B**) are highlighted with yellow or orange shaded boxes.

B. Cholinergic pathway



17 cases and 2 controls in the cholinergic pathway. Genes selected by NETBAG (main **Figure 4B**) are highlighted with yellow or orange shaded boxes

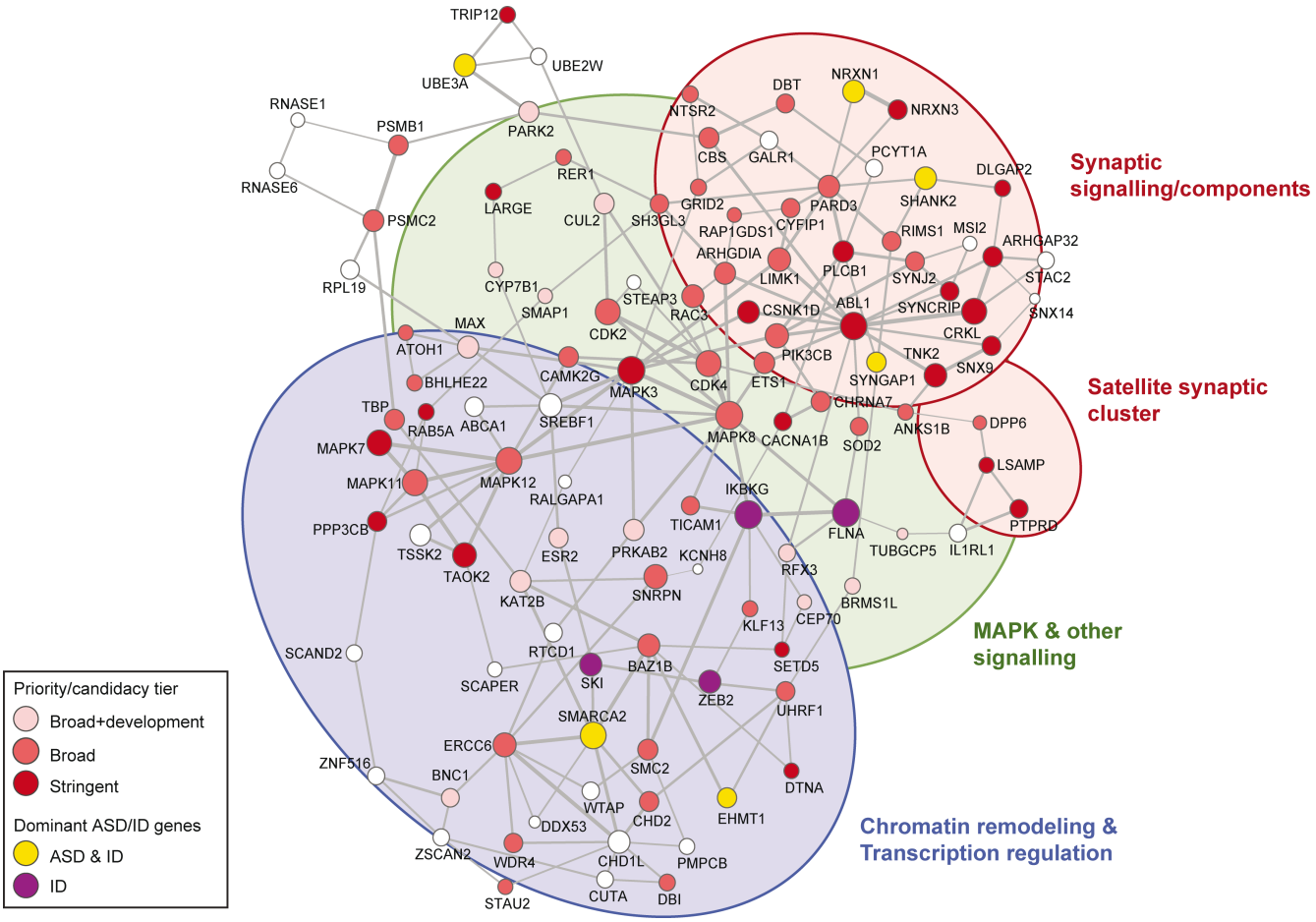


Figure S3. Candidacy tier for *de novo* CNV genes selected by NETBAG

Genes implicated in dominant forms of ASD and ID or in ID are highlighted in yellow and purple, respectively. Genes implicated in autosomal recessive ID (*ERCC6*, *CBS*, *DBT*, *PLCB1* and *LARGE*) were not given a specific color for ease of comprehension of the figure. Multiple lines of evidence from the literature and dedicated databases (DECIPHER, ISCA and ECARUCA) for the *CHD2* gene (recently involved in other neurodevelopmental disorders), and the candidate genes *SETD5*, *LSAMP*, *TRIP12*, *SYNCRIP*, *DTNA*, and *PIK3CB*, are given in the section '**Higlighted genes**'.

NETBAG analysis of *de novo* CNV genes: For this analysis, recurrent events were counted once (or combined in one region) and events that did not intersect any genes were removed; the resulting final set used in the analyses consisted of 75 CNV regions intersecting 874 unique RefSeq genes. Derived gene clusters were scored using the de-weighted method, where the contribution of each individual gene to the overall cluster score is given by the weighted sum of its connections (edges) to the other cluster genes,[1] and significance was obtained by generating random events with the same gene count as observed in the *de novo* CNV dataset with 1,000 randomizations (main **Figure 4B**; **Tables S14A-S14E**). The highest scoring cluster obtained using a searching procedure of up to two genes per CNV (global p-value 0.02) is shown. The list of up to 2 genes selected per CNV is given in **Table S14A**. The majority of the genes identified by NETBAG (92.0%, 104/113) are highly expressed in brain. When using a search procedure of up to one gene per CNV, a list of 69 genes was obtained (global p-value <0.01); this list is given in **Table S14A**. Analysis of established annotation resources, such as PubMed, OMIM, GeneReviews, EntrezGene and iHOP, suggested that a significant fraction of genes in the identified network either play a well-defined functional role in the brain or have been previously implicated in neurological and psychiatric disorders. An overview of functional information about each of the genes forming the cluster can be found in **Table S14B**.

Prenatal, postnatal or unbiased brain expression of genes selected by NETBAG (additional information relevant to main **Figure 4B**, based on NETBAG data). Nodes were colored according to whether a gene's brain expression is prenatally or postnatally biased, or has no biased expression in an analysis of 12 developmental stages of the BrainSpan RNA-seq dataset. Genes were considered biased if their mean expression in prenatal developmental stages was increased more than two-fold compared to their post-natal expression, or vice versa. Genes with related function according to Gene Ontology (GO)[2] and KEGG pathways[3] were indicated by shaded areas; a complete listing of over-represented GO terms and pathways is given in **Table S14D**. By further examining these genes (**Table S14C**), we found that 77% (87/113) of them have a strong link with neuronal and brain-specific genes. We explored whether affected genes were biased to certain developmental stages by interrogating the temporal expression of genes selected by NETBAG. We found no overall significant

bias to prenatally expressed genes, but we did observe prenatal bias for a cluster of genes participating in chromatin remodeling and transcription regulation (main **Figure 5B**, **Table S14E**). The prenatally-biased expression profile was characterized by high expression during development and sharp decrease after birth. We further compared expression levels (**Table S14B**) and found that most of the genes are highly expressed in the neocortex. Similar results were obtained when looking at LoF *de novo* single SNVs in 122 genes assembled from four ASD exome-sequencing studies.[4-7]

Candidacy tier of ASD candidate genes

To identify potential new ASD targets and gain mechanistic insights, we expanded our analysis to several gene lists described below. These gene lists were used to build a candidacy tier, by identifying which genes are more likely to be implicated in ASD based on membership in the gene lists. These gene lists were used for the analyses presented in main **Figure 3**, **Figure S3** and **Table S14C**. Our analyses focused on exonic events overlapping or disrupting exons, and deletions and duplications were analysed separately. P-values associated with odds ratios were calculated using Fisher exact tests.

1) ***Fragile X mental retardation protein (FMRP) targets*****:** the gene encoding FMRP, *FMR1*, responsible for fragile X syndrome, is mutated in ~2% of ASD cases.[8] Significant overlap between FMRP-targets and ASD candidate genes was reported recently,[9] and Iossifov et al.[4] had previously shown that rare *de novo* LoF SNVs identified in their exome sequencing study were enriched in this class of genes. We used a set of FMRP-RNA interactors identified experimentally (n=842), using crosslinking and immunoprecipitation (CLIP) experiments in mouse.[9] We note that a later paper[10] detected FMPR targets computationally, using human RNA sequence motif analysis (n=939 top genes as provided by the authors), but given the limited overlap between the genes in the two papers (~20%), we decided to use for our enrichment analyses only the list of targets that were experimentally identified by Darnell et al.[9].

2) ***Post-synaptic density genes*****:** The full list of experimentally identified human post-synaptic density genes was collected from Bayes et al. (n=1,453).[11]

3) ***Haploinsufficiency index*****:** We used the predicted haploinsufficiency index (HI) by Huang et al.[12] In particular, we used the imputed predictions as these have better gene coverage, and showed a similar performance compared to the original list of predicted genes pre-imputation when looking at dominant and recessive ASD and ID genes (data not shown). We defined four gene-sets: all genes with any value of HI (17,081 genes), the top 55% HI genes (pHI ≥0.15, 8,862 genes), top 26% (pHI ≥0.35, 4,136 genes), and top 14% (pHI ≥0.55, 2,214 genes). We focused our analysis on deletions because deletions, not duplications, are associated with HI, and show that HI is an important contributor to ASD etiology (duplication breakpoints can also disrupt genes or their regulatory elements and lead to haploinsufficiency).

4) ***Neurodevelopmental/neuropsychiatric phenotypes*****:** Genes associated with neurodevelopmental/neuropsychiatric phenotypes were mined from HPO (Human Phenotype Ontology)[13] for human and MGI/MPO for mouse (Mouse Genome Informatics/Mammalian Phenotype Ontology, The Jackson Laboratory, Bar Harbor, Maine, http://www.informatics.jax.org; data retrieved on November, 2012).[14]
   a) For human, we selected all genes annotated for "Behavioural/Psychiatric abnormality" (HP:0000708) and/or "Cognitive impairment" (HP:0100543), as well as children terms in the ontology. We only considered genes with HPO phenotype annotation derived from OMIM, as for these we could more reliably infer the mode of inheritance based on the first digit of the OMIM ID (1 = autosomal dominant, 2 = autosomal recessive, 3 = X-linked).
   b) For mouse, we selected all genes annotated for "nervous system phenotype" (MP:0003631) and "behavior/neurological phenotype" (MP:0005386), as well as children terms in the ontology; eQTL and complex phenotypes were removed. Autosomal dominant mode of inheritance was inferred by parsing MGI mutation allele tables and retaining only gene-phenotype annotations supported by a heterozygous mutated allele; genes on the human X chromosome were removed to obtain the autosomal dominant subset.

5) ***Neurodevelopmental function*****:** Two neurodevelopmental lists were generated to help evaluate the candidacy of the genes selected by NETBAG: "NeuroF" was purely based on the manual curation of Gene Ontology terms and pathways, and "NeuroF_EM" was based on gene-sets found associated by the logistic regression test in the enrichment map (EM) (results shown in main **Figure 4A**) and clustered as: (a) Cell projection/Neural development/Axonogenesis, or (b) Neuronal synapse. The Gene Ontology terms and pathways used to generate NeuroF are listed in **Table S13A**.

6) ***Developmental genes*****:** Genes were collected from all gene-sets found associated by the logistic regression test and clustered as Development/Cell Proliferation/Cell Motility in the enrichment map ("Dev_EM").

7) ***Brain-expressed genes*****:** Three gene lists were constructed based on the human BrainSpan RNA-seq dataset:
   a) *ExprBspan_BrainAny_log2rpkm3.0*: genes with RPKM ≥3.0 in at least five BrainSpan data-points (a data-point corresponds to any unique combination of donor subject x brain region x developmental time-point).
   b) *ExprBspan_BrainAny_log2rpkm4.5*: genes with RPKM ≥4.5 in at least five BrainSpan data-points.
   c) *ExprBspan_NotBrainExprF*: genes RPKM ≥1.0 not satisfying the above definition, yet annotated for at least one Gene Ontology term or pathway; these are similar in number to ExprBspan_BrainAny_log2rpkm4.5 and therefore were used as a negative control list, representing functionally characterized genes not expressed or expressed at very low levels in brain.

   Brain-expressed gene lists were also defined based on the Novartis expression atlas (HG-U133A Affymetrix arrays), and used only for the definition of candidate gene lists:

a) *Novartis Brain-expressed genes*: for a gene to be considered brain expressed it needed to have the robust multi-array average (RMA) gene expression values in whole-brain or fetal brain: (i) greater than the array median expression value (considering all tissues), and (ii) greater than the median expression in non-nervous tissues.
b) *Novartis Brain-specific genes*: for a gene to be considered brain-specific it needed to have the RMA gene expression values in whole brain or fetal brain: (i) greater than the array median expression value (considering all tissues), and (ii) greater than the double of the median expression in non-nervous tissues.

Three lists were compiled based on the results from the 1-gene hit odds ratio analysis of single gene-sets and their combinations (main **Figures 3A-B**):

1) Stringent (1,088 genes): clearly linked to synaptic components and/or autosomal dominant neurodevelopmental/neuropsychiatric phenotypes.
2) Broad (5,342 genes): linked to synaptic components, or neurodevelopmental/neuropsychiatric phenotypes, or brain expression.
3) Broad development (7,158 genes): extended by gene-sets corresponding to broader developmental functions found by the logistic regression gene-set test and enrichment map clusters.

In particular, the following definitions were used:

1) Stringent, union of:
   a) FMRP targets from Darnell et al.[9]
   b) Genes found in at least two of these lists:
      i) NeuroPheno Dom
      ii) Intersection of: (a) Novartis Brain-expressed and (b) HI ≥ 0.55
      iii) Postsynaptic density genes, full dataset from Bayes et al.[11]
      iv) FMRP targets from Ascano et al.[10]
2) Broad, union of:
   a) Stringent candidates (as defined above, in Stringent)
   b) Novartis Brain-specific
   c) Intersection of: (i) ExprBspan_BrainAny_log2rpkm4.5 and (ii) HI ≥ 0.35
   d) Intersection of: (i) NeuroF and (ii) HI ≥ 0.35
   e) Intersection of: (i) ExprBspan_BrainAny_log2rpkm3.0 and (ii) NeuroPheno All
3) Broad development, union of:
   a) Broad candidates (as defined above, in Broad)
   b) Intersection of: (i) NeuroF_EM and (ii) NeuroPheno All
   c) Development cluster in the enrichement map (**Figure 4A**, main text)

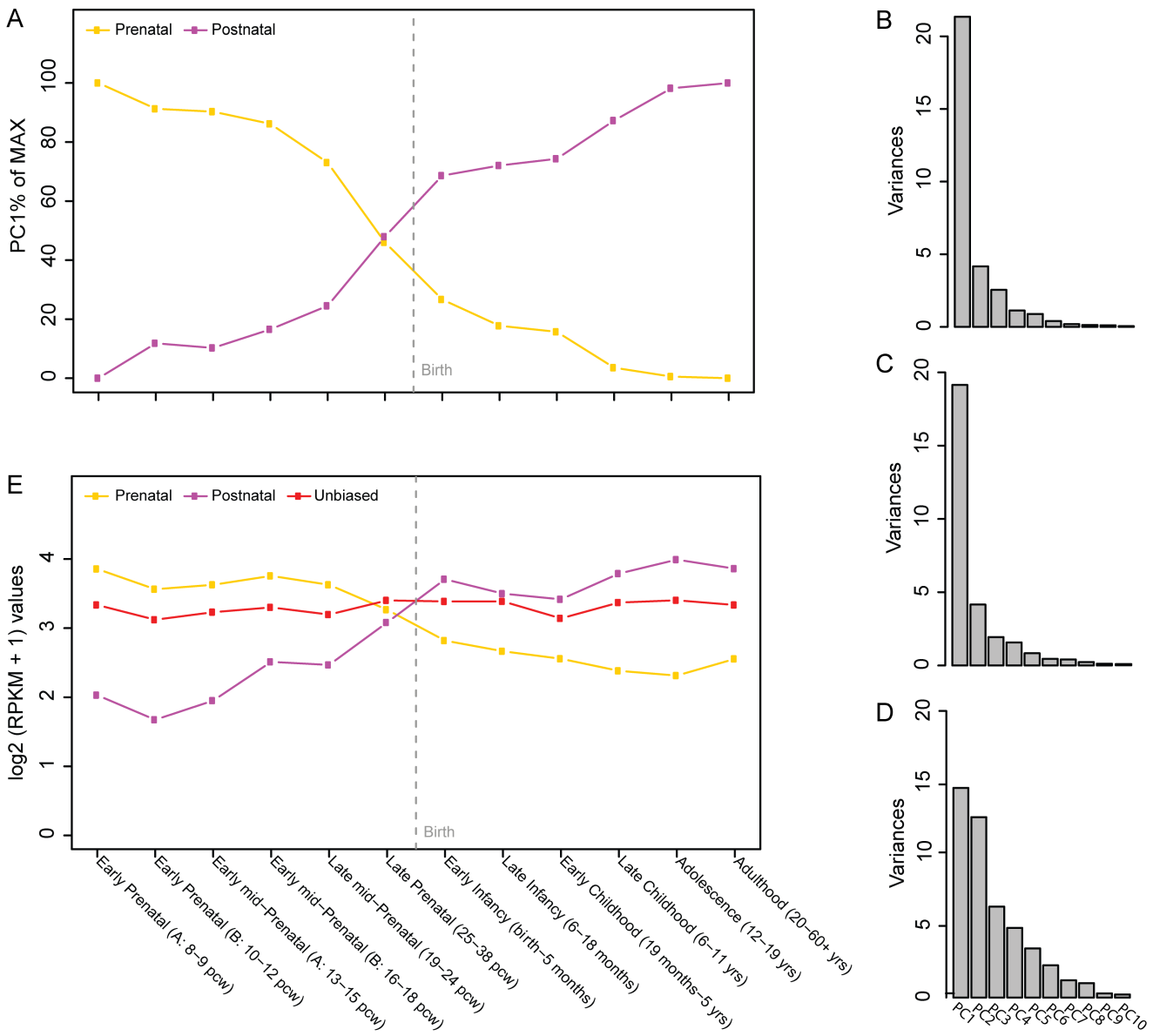


Figure S4. Expression profile for 113 *de novo* CNV genes selected by NETBAG in the neocortex across 12 developmental stages

The time course of BrainSpan RNA-seq gene expression data was quantile normalized and log2 transformed. **(A)** The first principle components (PC1, as percentage of the maximum, %/MAX) for the 113 CNV genes are shown in 2 expression groups (prenatally biased, postnatally biased). The amount of variance explained by each principle component is shown for **(B)** Prentatally-biased genes, **(C)** Postnatally-biased genes, and **(D)** Unbiased genes. **(E)** The group centroids (median) of the 3 expression groups (Prenatally biased, Postnatally biased, Unbiased) are plotted at each developmental stage with the Y axis showing log2 transformed expression profiles log2(RPKM + 1). PCW, post-conceptional weeks. Panels B-D compared the variance explained by PC1 for the 3 expression groups. PC1 for Unbiased genes is not representative, i.e., it does not explain more variance than other PCs and it is unable to profile the major direction of the expression group. Therefore we did not include the unbiased group in panel A.

RNA-Seq data preprocessing and normalization: The human brain developmental transcriptome data, consisting of a total of 12 developmental stages (from 8-9 post-conceptional weeks [PCW] to 20-60 years) in 524 samples, was downloaded from BrainSpan resource (http://brainspan.org/; Allen Institute for Brain Science, 2012; analyses presented here used the October 2013 version).[15,16] In this dataset, gene expression profiles measured by RNA-sequencing experiments were represented as reads per kilobase of transcript per million mapped reads (RPKM),[17] which were pre-summarized to gene levels based on a composite model defining a gene as the union of all exonic nucleotides across all transcripts. A sample level filtering procedure was implemented after comparison of the distribution of each sample to the entire distribution (all 524 samples); the top 2% outliers were removed based on the sum of chi-squares with visual assistance on boxplot. Subsequently, a gene level filter was applied so that only genes with a RPKM value > 0 in

more than 80% of the samples in any brain region of any developmental stage were kept. To ensure quality of the data, the samples were further filtered by RNA integrity number (RIN) > 8. As a result, 42,965 gene entries and 416 samples passed QC and entered the downstream analysis where quantile normalization and log2 transformation procedures were performed.

Gene expression level classification: We then sought to classify genes by their expression levels across multiple brain regions and developmental stages. The genes with lower 20% expression level in all 8 brain regions (neocortex, amygdala, hippocampus, striatum, diencephalon, upper rhombic lip, cerebellar cortex, and dorsolateral prefrontal cortex) were classified as "non-expressed" genes. The remaining 80% genes were classified into 3 expressional categories specifically for two brain regions, the neocortex (NCX) and dorsolateral prefrontal cortex (DFC): 1) lowly expressed (bottom 25% averaging across NCX or DFC samples); 2) highly expressed (top 25% averaging across NCX or DFC samples); and 3) the remainder of genes were considered moderately expressed for NCX or DFC regions.

Identification of differentially expressed genes: We analysed the 113 genes resulting from the NETBAG functional network analysis of 102 *de novo* CNVs, and determined their prenatal-biased or postnatal-biased status by a nonparametric algorithm, similar to the method used in Xu et al.[18] Specifically, for each of the 8 brain regions, we used the median of samples to represent the expression level of a certain given gene under a specific developmental stage. Subsequently, two groups were defined: prenatal (periods 2A, 2B, 3A, 3B, 4, 5) and postnatal (periods 6 through 11). Wilcoxon rank sum test with a p value cut-off of 0.05 and group median fold change cut-off of 1 were applied to determine differentially expressed genes between prenatal and postnatal groups. Genes with significantly higher expression in prenatal or postnatal group were considered prenatally or postnatally biased genes, while the remaining genes were classified as unbiased. Interestingly, we observed that some genes (e.g. *KCNH8*, *CRKL*) presented a peculiar profile: a "U shape", i.e., high expression in the first developmental stages followed by a decrease in expression and later, towards the last developmental stages, a new increase in expression. If there was a predominance of high expression in the left or in the right branch of the "U" curve, the genes were classified as prenatally-biased or postnatally-biased, respectively. The inverse could be observed for genes with a profile resembling an inverted U (e.g. *RAC3*).

Profile of gene expression across 12 developmental stages: Expression profile across 12 developmental stages of the 113 genes resulting from the NETBAG functional network categorized into 3 groups (prenatal, postnatal and unbiased) was plotted for 8 brain regions after Principle Component Analysis was performed with a custom R script. The first principle component of each expressional group was plotted in **Figure S4** for the neocortex region. Of note, the chromatin/transcription module in **Figures 4B** and **S4** showed a predominance of genes with a prenatally-biased expression profile, suggesting that besides the function of the altered gene, the timing of the effect of the genetic perturbation may also be of critical importance in determining disease risk.

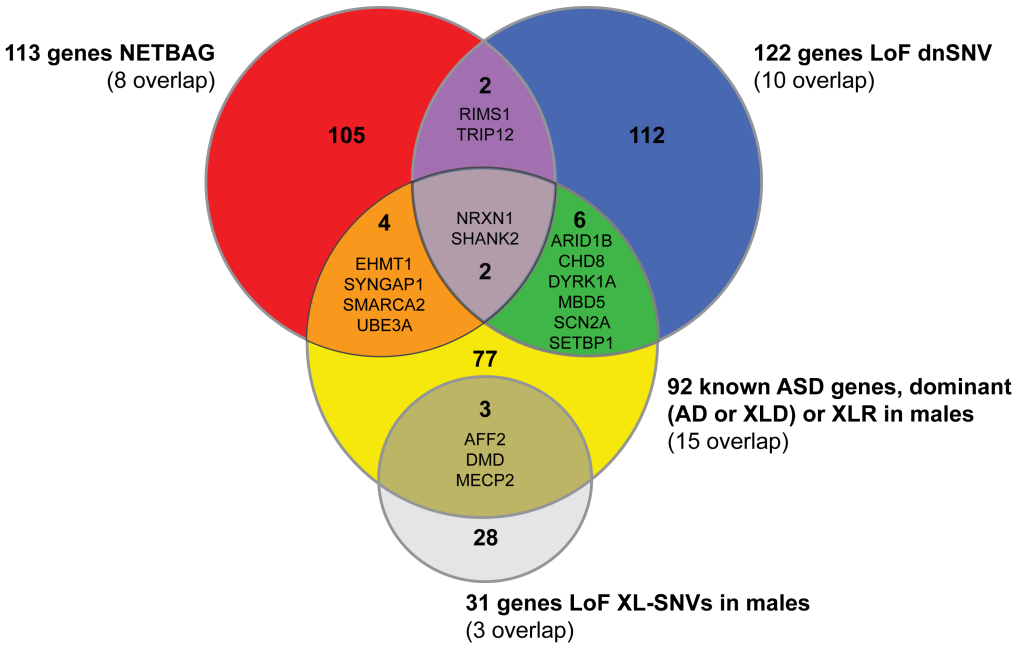


Figure S5. Overlap among gene lists pre-DAPPLE analysis

For the protein-protein interaction (PPI) network analyses presented in **Figures 5A and S5**, we used the Disease Association Protein-Protein Link Evaluator (DAPPLE)[19] under the assumption that causal genetic variants affect a limited number of mechanisms that are detectable by PPIs. DAPPLE takes a list of genes or loci as input and searches for significant physical connectivity among proteins encoded by those genes based on protein-protein interactions available in the InWeb database.[20] DAPPLE will then assess the statistical significance of several network connectivity parameters as well as of the connectivity of individual proteins to other seed proteins using a within-degree node-label permutation method. The Venn diagram shows the overlap between four lists used as input in DAPPLE (i.e. pre-DAPPLE):

1) 113 genes affected by *de novo* CNVs in AGP probands selected by NETBAG.
2) 122 genes with *de novo* LoF SNVs from four published whole exome sequencing studies comprising 965 ASD cases.[4-7] LoF variants included nonsense, frameshift, and splice site-disrupting mutations. One of the 122 genes is X-linked (*RPS6KA3*), and the remainder are autosomal. Three genes were not present in the InWeb database and were thus not included in the analyses (*DHRS4L1*, *PCDHA13*, and *RAD21L1*).
3) 92 genes known to be involved in ASD (list updated from Betancur, 2011)[21] (**Table S6A**); only autosomal dominant (AD) or X-linked (XL) genes were included (XL dominant, XLD, and XL recessive, XLR), since heterozygous CNVs affecting genes involved in autosomal recessive disorders are not deleterious *per se*.
4) 31 X-linked genes with hemizygous LoF SNVs in male ASD cases and not observed in male controls.[22]

Of the 336 genes analysed (339 unique genes pre-DAPPLE minus 3 not present in the Inweb database), 17 genes overlap between two or more lists. Specifically, the *de novo* CNV genes selected by NETBAG overlap with the list of genes with *de novo* LoF SNVs and with the list of genes known to be involved in ASD; two of the genes (*NRXN1*, *SHANK2*) were shared by the three groups. There is no overlap between the 113 genes selected by NETBAG and the X-linked genes with LoF SNVs. Likewise, the Venn diagram post-DAPPLE depicted in main **Figure 5A** shows the overlap among the 151 genes selected by DAPPLE. It resembles the pre-DAPPLE Venn diagram since the relationship between the four gene lists was maintained, including the sharing of two genes (*NRXN1*, *SHANK2*) by the three main lists of genes. Over-representation analysis identified convergent functional themes related to neuronal development and axon guidance, signalling pathways, and chromatin/transcription regulation. Moreover, 90.7% (137/151) of the gene products reside within the same interconnected cluster.

We then further analysed the 113 genes selected by NETBAG and the DAPPLE output network of 151 proteins in terms of brain expression during development and in terms of functional groups (**Table S14E** and **S15**, respectively). Regarding the 113 gene list, we observed that the functional group composed of chromatin remodelers and transcription factors is significantly enriched for prenatal-biased genes and that the functional group composed of signalling genes is enriched for postnatal-biased genes. Similarly, for the DAPPLE output network, the chromatin remodelers/transcription factors functional group is significantly enriched for prenatal-biased genes, with a smaller p-value than the one estimated for the same functional group on the 113 genes list. Interestingly, in this case the functional group composed of neuronal genes is significantly enriched for postnatal-biased genes.

Phylop and GERP scores: Basewise conservation scores (PhyloP) across 45 vertebrate genomes (including the human genome) were downloaded from the UCSC website (http://genome.ucsc.edu/)[23] and processed with reference to the UCSC gene track (known genes). Using only the exome, the average base score was calculated. When different variants were present in a gene, all isoforms were aggregated and the average score was used. Likewise, the Genomic Evolutionary Rate Profiling (GERP++) score was downloaded from

Sidow's Lab website (http://mendel.stanford.edu/SidowLab/downloads/gerp/index.html). GERP score was calculated by taking the average basewise score of a gene. Both Phylop and GERP++ scores of genes were ranked. The ranking percentage of the scores (e.g. X% meaning X% of genes in the entire genome are ranked behind the given gene in terms of GERP or Phylop scores) are calculated to characterize each of the genes. Genes in the DAPPLE network (main **Figure 5B**) are among the top 75% more conserved genes compared to the genome average (based on GERP scores and PhyloP), and are typically long (median 103 kb *vs*. 10 kb for the genes in whole genome). Functionally significant and highly conserved genes tend to be more central in physical protein-protein and regulatory networks.[24]

HIGHLIGHTED GENES

Genes involved recently in other disorders

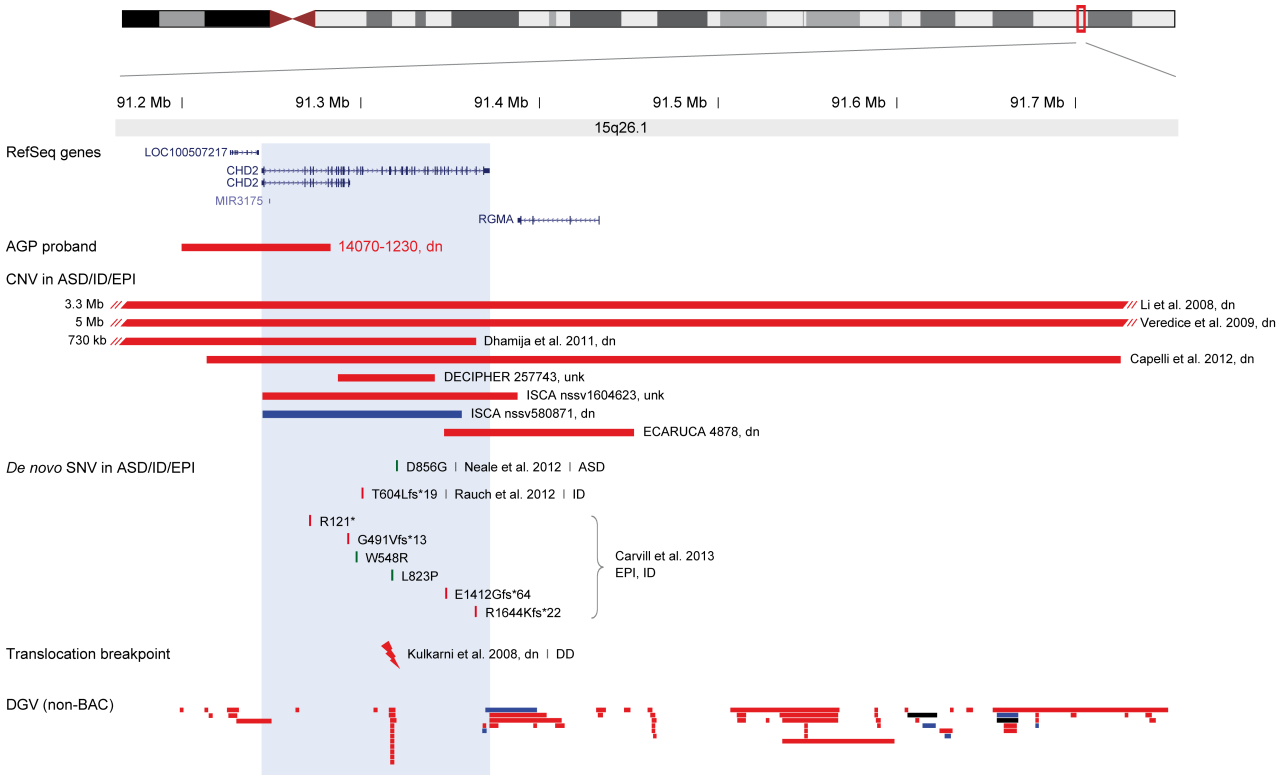


Figure S6. CNV and SNV identified in *CHD2* (chromodomain helicase DNA binding protein 2) in chromosome 15q26.1

Chr15:91,165,000-91,750,000 (hg18). Abbreviations: ASD, autism spectrum disorder; dn, *de novo*; DD, developmental delay; EPI, epilepsy; ID, intellectual disability; unk, unknown inheritance.

1) AGP

- Proband 14070-1230, male, *CHD2* exonic deletion (chr15:91200007-91283004), *de novo*. The deletion also involves *LOC100507217* and *MIR3175*. The affected younger brother also carries the deletion, absent from both parents. SNP genotyping indicated that the deletion occurred in the paternal chromosome, suggesting paternal germline mosaicism. Phenotype: index case, 14 yo, autism (meets criteria both on ADI-R and ADOS), walked at 14 mo, first words 24 mo, phrases 26 mo, currently verbal, ID (Griffiths: language DQ 68, performance DQ 46, global DQ 53); normal growth, normal neurological exam; minor dysmorphic features: micrognatia, protruding ears; no epilepsy; brain MRI: altered angular gyrus (normal variant, unknown pathological significance). Brother, 12 yo, mild autism on ADOS, walked at 15 mo, first words 24 mo, phrases 36 mo, verbal, ID (Griffiths: language DQ 65, performance DQ 61, global DQ 51); normal growth, normal neurological exam; minor dysmorphic features similar to his brother (micrognatia, protruding ears); epilepsy, onset 9 y, partial with secondary generalization, difficult to control with carbamazepine; EEG showed paroxistic activity in the left temporal lobe.
- No *CHD2* deletions among 4,768 controls and 4,875 parents.

2) Other human genetic evidence

CNV:

- Li et al.[25]: Male with t(15;22)(q26.1;q11.2) translocation and a 3.3 Mb deletion encompassing *CHD2* (chr15:89197342-92489641) at the 15q breakpoint, *de novo*. Phenotype: developmental delay, walked at 20 mo, language delay, developmental motor coordination disorder, height and head circumference consistently ≤3rd centile, weight ≤5th centile, delayed bone age (6-mo-old bone age at 18 mo of chronological age, and 18-mo-old bone age at 36 mo chronological age), mild dysmorphic features (anteverted nares, unilateral auricular pit, fetal fingertip pads, low posterior hairline, and back hirsutism), left eye amblyopia corrected surgically at age 2. He had two episodes of febrile seizures. *IGF1R* was translocated to chromosome 22 and showed 50%

reduction of expression, which could be responsible for the growth deficiency. Angelman syndrome was ruled out by methylation analysis.

- Veredice et al.[26]: Female, 5 Mb deletion encompassing *CHD2* (chr15:87796000-92700000), *de novo*. Phenotype: severe intractable myoclonic epilepsy with photosensitivity, with onset at 6 mo, ID, growth delay, peculiar facial features and minor physical anomalies. Born at 37 weeks, weight 3rd centile, length 10th centile, head circumference 2nd centile. At 21 mo her weight was 3rd centile, with length and head circumference <<3rd centile. She had congenital hypothyroidism, bicuspid aortic valve, diffuse hypotonia, ligamentous laxity, and dysmorphic features (upward slanting eyes, epicanthal fold, depressed nasal bridge, full cheeks, prominent lips, downturned corners of the mouth, protruding tongue, large ears with anteverted lobe, single palmar crease, increased 1st–2nd interdigital space and redundant nuchal skin). Her language was limited to a few words and she had mild developmental delay (Griffith's developmental quotient 67). Brain MRI showed cerebellar vermis hypoplasia with mega-cisterna magna.
- Dhamija et al.[27]: Female, deletion encompassing *CHD2*, with a minimum size of 731 kb (chr15:90633409-91364628) and a maximum size of 936 kb (chr15:90530351-91466733), *de novo*. Phenotype: 9 yo girl with developmental delay, ASD, growth delay, and intractable generalized epilepsy, with onset at 3.5 y. Initially the seizures were partial complex but later became generalized; most prominent were absence seizures, occurring many times a day. She had delayed motor development (walked at 24 mo) and delayed language development with echolalia and mild generalized cognitive delay upon formal testing. On examination, she had mild dysmorphic features (widely set eyes, bilateral pits on the helix of the ears, and crowded teeth with prominent incisors), short stature and head circumference <3rd centile. Brain MRI was normal.
- Capelli et al.[28]: Female, 511 kb deletion encompassing 4 genes, *CHD2*, *LOC100507217*, *MIR3175* and *RGMA* (chr15:91213864-91724860), *de novo*. *RGMA* (repulsive guidance molecule, member A) exerts a negative control on axon growth and could also potentially contribute to the phenotype. Phenotype: when examined at the age of 6 y, she presented global developmental delay, epilepsy, autistic behavior, severe speech impairment with minimal use of words, short attention span, facial dysmorphisms (prognathia, wide mouth, short and widely spaced teeth, strabismus), gait ataxia with uplifted arms and hand flapping, slight hypotonia and hyperactive deep tendon reflexes. Weight, height and head circumference were at the 50th, 75th and 10th centiles respectively. Angelman syndrome was suspected but the methylation pattern at *SNRPN* was normal. Seizures started at the age of 24 mo and were partially controlled with valproic acid. (Corresponds to subject 249888 in DECIPHER.)
- DECIPHER[29]: Subject 257743, male, *CHD2* intragenic deletion (chr15:91287269-91341234), inheritance unknown. Phenotype: autism, macrocephaly, proportionate short stature, scoliosis. This individual also has a 16p11.2 deletion syndrome, inheritance unknown.
- ISCA[30]: Subject nssv1604623, unknown gender, deletion encompassing 3 genes, *CHD2*, *MIR3175* and *RGMA* (chr15:91245098-91387818), inheritance unknown. Phenotype: seizures.
- ISCA: Subject nssv580871, unknown gender, intragenic *CHD2* duplication (chr15:91245028-91356680), *de novo*. Phenotype: developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing.
- ECARUCA[31]: Subject 4878, male, deletion encompassing only 2 genes, *CHD2* and *RGMA* (chr15:91346585-91452586), *de novo*. Phenotype: ID, truncal obesity, large ear lobules, bulbous nasal tip, low posterior/trident hairline, cryptorchid testes, small hands and feet, tapering fingers, hypotonia, seizures, abnormal EEG.

Translocation breakpoint:

- Kulkarni et al.[32]: Female proband with a balanced *de novo* translocation t(X;15)(p22.2;q26.1); the 15q26.1 breakpoint disrupts *CHD2*, whereas the Xp22.2 breakpoint lies in a gene-free region. Phenotype: developmental delay, learning problems, scoliosis, hirsutism, high arched palate, normal face (including eyes and ears), syndactyly of the toes, height <30th centile and head circumference <25th centile. No seizures reported at age 17 y.

SNV:

- Carvill et al.[33] (targeted sequencing in epileptic encephalopathies): 6 individuals (4 males and 2 females) with *de novo* mutations in *CHD2* (3 frameshift, 1 stop, 2 missense). Phenotype: 2 had myoclonic atonic epilepsy, 3 had epileptic encephalopathy not otherwise specified and 1 had Lennox-Gastaut syndrome. Seizure onset occurred between 1-3 y; myoclonic seizures were present in all six, photosensitivity in three. All had ID (2 moderate, 4 severe), 1 had ASD.
- Neale et al.[5] (exome sequencing in ASD): Subject 10C100480, male, *CHD2* missense variant, *de novo*. The variant is predicted to be either benign by PolyPhen2[34] and PANTHER,[35] or damaging by SIFT,[36] SNAP,[37] and Mutation Taster;[38] the affected residue is highly conserved: GERP 5.48, [39] ConSurf 6/9.[40] Phenotype: ASD (no other information provided).
- Rauch et al.[41] (exome sequencing in nonsyndromic sporadic ID): Subject MS134, female, *CHD2* frameshift mutation, *de novo*. Phenotype: severe nonsyndromic sporadic ID (IQ 50-69), absence seizures (onset at 5 y), Duane anomaly, no ASD.

3) Functional evidence

- *CHD2* encodes a member of the chromodomain helicase DNA-binding (CHD) family of proteins, characterized by the presence of chromodomains (chromatin organization modifier) and SNF2-related helicase/ATPase domains. CHD proteins alter gene expression by modification of chromatin structure, playing critical roles during development.

- Expressed in the brain.
- *Chd2* knockout in mice results in embryonic lethality; heterozygous mice have variable postnatal lethality and growth retardation, with reduced body fat, pronounced lordokyphosis, renal disease and anemia in adults.[42,43]

4) Other evidence

- % haploinsufficiency (HI) = 15.8 (likely to be haploinsufficient).[12]
- Mutations in *CHD7* cause CHARGE (coloboma, heart defects, atresia of the choanae, retardation of growth and developmental, genital and/or urinary defects and ear abnormalities) syndrome, associated with syndromic ID and sometimes ASD.[44,45]
- Nine *de novo* SNV in *CHD8* reported in subjects with ASD (3 nonsense, 4 frameshift, 1 splice, and 1 single amino acid deletion).[6,46]

5) Comment

CHD2 mutations were implicated in epileptic encephalopathies (Carvill et al. 2013) while this manuscript was in preparation. The deletion in the two siblings with ASD identified in our study, together with the review of other cases, also implicates this gene in ASD. *CHD2* haploinsufficiency is highly penetrant (all cases with known inheritance are *de novo*), associated with ID and variably with epilepsy and ASD. Although the phenotype description is missing in several cases and is incomplete in others, precluding complete analyses, among 18 cases (the missense variant reported by Neale et al. not included), ID is reported in 16, epilepsy in 13 and autism/ASD in 6 cases. Language delay and/or limited speech are described in 5. Eight individuals are reported with dysmorphic features, usually mild. Features reported in 2 or more individuals include: strabismus or amblyopia (n=3), ear pits (n=2), hirsutism (n=2), and low posterior hairline (n=2). At least 4 had growth delay/short stature, and 2 have scoliosis, two phenotypes reported in heterozygous *Chd2* mice.

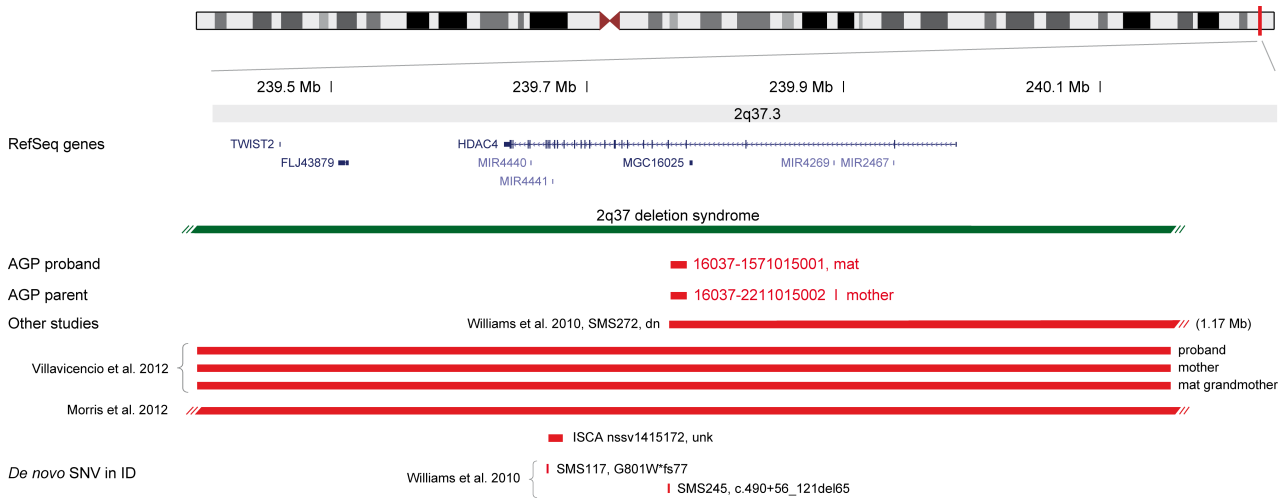


Figure S7. CNV and SNV identified in *HDAC4* (histone deacetylase 4) in chromosome 2q37.3

Chr2:239,350,000-240,180,000 (hg18). Abbreviations: dn, *de novo*; ID, intellectual disability; mat, maternal; unk, unknown inheritance

1) AGP

- Proband 16037-1571015001, male, 12 kb intragenic deletion of *HDAC4,* removing exon 4 (chr2:239766528-239778481), maternal. Phenotype: autism (ADI-R and ADOS positive), language delay (first words and phrases 42 mo), functional language, WISC-III at 21 y: verbal IQ 72, performance IQ 87, full scale IQ 77. Born at 37 weeks of gestation, no congenital malformations, normal general exam. Had seizures at age 4-5 y, treated with Depakote; treatment stopped for 2-3 y without seizure activity. Unfortunately, we were unable to have the proband re-evaluated by a clinical geneticist to assess the presence of mild dysmorphic features and hand or feet abnormalities. The father has schizophrenia; the mother has depression, no other information is available about her phenotype.
- No *HDAC4* deletions among 4,768 controls and 4,874 parents (1 deletion in the mother of proband 16037-1571015001).

2) Other human genetic evidence

- Over 120 individuals with 2q37 terminal deletions (2q37 deletion syndrome, also known as brachydactyly mental retardation syndrome) have been reported. Common features include developmental delay/ID, ASD, hypotonia, mild facial dysmorphism (frontal bossing, round face, depressed nasal bridge, abnormal or prominent ears, deep-set eyes, anteverted nares, and thin upper lip), short stature, obesity and brachymetaphalangy of digits 3-5 (>50%).[47] In 2010, haploinsufficiency of *HDAC4* was shown

to result in brachydactyly mental retardation syndrome. Williams et al.[48] described 5 individuals with 2q37 deletion syndrome in whom the smallest region of overlap contained only *HDAC4*. Sequencing of *HDAC4* identified *de novo* loss-of-function mutations in two additional subjects. Although *HDAC4* mutations may be causative for most of the features of the 2q37 microdeletion syndrome, other genes might also be involved, since individuals with distal deletions not including *HDAC4* have been reported with ID, ASD and seizures. Because no *HDAC4* mutations or single gene deletions had been reported previously in ASD, this gene was not yet considered as a cause of ASD (i.e. we had included it in the list of ID genes, not in the ASD list).

- 2q37 deletions as well as the 2 reported *HDAC4* mutations occur *de novo*, so the maternal transmission observed in the AGP proband would appear a highly unusual finding. However, a recent case report[49] described for the first time a deletion overlapping *HDAC4* inherited from a mildly affected parent (see below).

CNV:

As noted above, there are many large 2q37 deletions reported in the literature and in the databases. Here we review only cases with small deletions overlapping *HDAC4* and two recently reported familial cases.

- Villavicencio-Lorini et al.[49]: Report of the first three generation familial case of brachydactyly mental retardation syndrome with an interstitial 2q37.3 microdeletion of 758 kb (chr2:239395957-240154599). Subject 1: female index case, only child, motor development and growth delays during the first year of life, with subsequent catch up. When evaluated at 2 y 8 mo, she had midface hypoplasia, deep set eyes, posteriorly rotated and low-set ears, thin upper lip and pointed chin. Speech development was normal; she exhibited aggressive tantrums and sleep difficulties. Brachydactyly was excluded clinically and radiologically. Subject 2: 45-year-old mother of the index case, history of developmental and growth delays during childhood, but she was later able to attend normal school. She reported reduced spatial orientation and memory deficits. Examination revealed coarse facial appearance with broad and depressed nasal bridge, highly arched eyebrows, deep set eyes and narrow palpebral fissures; growth parameters were normal. Her hands and feet appeared normal, and brachydactyly was excluded radiologically. Subject 3: 68-year-old grandmother of the index case, she is the mother of subject 2, her only child. Family history revealed that her sister's daughter had ID with hydrocephalus and paraplegia of unknown cause. Subject 3 had severe osteoarthritis, dysmorphic facial features similar to those of her daughter with highly arched eyebrows, narrow palpebral fissures and everted full lips, with normal growth parameters. She communicated in simple sentences and her intellectual skills appeared to be lower than normal. Clinically her hands were normal.
- Morris et al.[50]: Report of a familial case of brachydactyly mental retardation syndrome, including a parent with mild symptoms and a child exhibiting a more severe phenotype. Cytogenetic testing showed a cryptic balanced translocation in the mother, t(2;10)(q37;q26), which resulted in a 9.84 Mb deletion on chromosome 2q37.1 (chr2:232810566-242654701) and a 10q26.1 duplication (chr10:131931089–135253240) in her son. *HDAC4* was deleted in the child but present and translocated in his mother. Interestingly, *HDAC4* expression in lymphocytes was 67% in the mother and 23% in the son compared to normal controls. Since the predicted expression after loss of one copy of *HDAC4* is ~50%, these findings suggest that there is an additional unknown mechanism decreasing *HDAC4* dosage, both in the mother and the child.
  Subject 1: male proband, only child; upon examination at the age of 15 y he was overweight (BMI 28), with short stature (–3 SD), facial dysmorphism with round face, prominent forehead, highly arched eyebrows, sparse hair, low-set ears, downslanted palpebral fissures, depressed nasal bridge, thin upper lip, and high palate. He also presented type E brachydactyly of fourth fingers and toes, and syndactyly of the second and third toes. Brain MRI showed abnormal gyration of the frontal lobes. He was schooled in a specialized institute. Subject 2: mother of subject 1, she had one previous termination of pregnancy for occipital meningoencephalocele. Family history showed a maternal aunt and three maternal cousins with ID. She presented with a similar dysmorphic features as her son (round face, prominent forehead, highly arched eyebrows, low-set ears and thin upper lip), short stature, obesity (BMI 45), and brachydactyly of the fourth finger of the right hand, and third and fourth fingers on the left. She had no obvious ID.
- ISCA: Subject nssv1415172, unknown gender, *HDAC4* intragenic deletion of 10 kb (chr2:239670063-239680639), inheritance unknown. Phenotype: developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing.

3) Functional evidence

- Histones play a critical role in transcriptional regulation, cell cycle progression, and development. Histone acetylation/deacetylation alters chromosome structure and affects transcription factor access to DNA. *HDAC4* encodes a histone deacetylase that shuttles between the nucleus and cytoplasm in response to calcium-regulated signals. HDAC4 has been shown to regulate a transcriptional program essential for synaptic plasticity and memory; by repressing genes encoding synaptic proteins, it affects the strength and architecture of excitatory synapses.[51] Conditional deletion of *Hdac4* in mouse forebrain neurons leads to impairments in spatial learning and memory and long-term synaptic plasticity.[52]
- Highly expressed in the brain.

4) Comment

Although many cases of 2q37 deletion syndrome with ASD have been reported in the literature, no deletion involving only *HDAC4* or mutation in this gene had been reported in individuals with ASD. The identification of an intragenic exonic deletion of the *HDAC4* gene in one AGP proband implicates haploinsufficiency of this gene as a cause for autism.

Figure S8. Recurrent duplications at Xq28 including *GDI1* (GDP dissociation inhibitor 1)

ChrX:153,085,000-153,640,000 (hg18). Abbreviations: ASD, autism spectrum disorder; dn, *de novo*; EPI, epilepsy; ID, intellectual disability; mat, maternal; unk, unknown inheritance

1) AGP

- Proband 20013-1075001, male, Xq28 duplication of 292 kb (chrX:153222048-153514311), *de novo*. The duplication encompasses 21 genes, including 3 genes involved in ID (*FLNA, GDI1, IKBKG*) and is flanked by segmental duplications. Phenotype: sporadic autism (ADI-R and ADOS positive), no language delay, normal IQ (WISC-III at 16 y: verbal IQ 93, performance IQ 94, full scale IQ 93), no dysmorphic features, head circumference -1.4 SD, normal neurological exam, no seizures.
- Proband 14216-3470, male, Xq28 duplication of 211 kb (chrX:153263157-153474401), maternally inherited. The duplication overlaps 19 genes, including 2 genes involved in ID (*GDI1, IKBKG*) and is flanked by segmental duplications. Phenotype: autism (ADI-R and ADOS positive), neurodevelopmental delay with onset at 2 y (first words 18 mo, first phrases 48 mo), mild ID (verbal IQ 61, performance IQ 65, full scale IQ 72), no dysmorphic features, sleep problems, no epilepsy. A brother with developmental delay also carries the duplication. He has a confirmed learning disorder and probably mild intellectual disability (he is an adult now and was not evaluated formally for this study).
- No similar duplications among 2,022 male controls and 2,441 fathers.

2) Other human genetic evidence

- Mutations in *GDI1* are a rare cause of nonsyndromic X-linked ID[53,54] and have not yet been reported in ASD.

CNV:

- Vandewalle et al.[55]: Four unrelated families with X-linked ID with recurrent 0.3 Mb Xq28 copy number gain (153.20-153.54 Mb) mediated by segmental duplications. Only males are affected, carrier mothers show skewed X chromosome inactivation. The copy-number gain is variable, ranging from 2 to 5 copies, and includes *GDI1, FLNA* and *IKBKG*, involved in ID through mutations. The authors suggest *GDI1* is the most likely candidate gene; it is highly expressed in the brain and its expression in blood is correlated with the severity of the phenotype. *FLNA* duplications have been reported in four males with intestinal dysfunction, without ID[56] and in one male control in the Database of Genomic Variants (DGV); a duplication including *FLNA* and *IKBKG* was reported in a healthy male.[55]
  Family 1: four affected males in two generations, Xq28 copy number gain (chrX:153218000-153535000). Phenotype: non-syndromic moderate ID. Family 2: two affected brothers, Xq28 copy number gain (chrX:153218000-153542000). Phenotype: severe ID, epilepsy in the elder brother. Family 3: three affected males, Xq28 copy number gain (chrX:153218000-153530000). Phenotype: the index case presented with delayed speech, learning disabilities, and mild ID. All affected male subjects had moderate ID, a peculiar face with dysmorphic features, and macrocephaly. Reported previously as case 6 in Madrigal et al.[57] Family 4: male (sporadic case), Xq28 copy number gain (chrX:153218000-153535000). Phenotype: global psychomotor delay, mild dysmorphism.
- Sakai et al.[58]: Subject 11092, male, Xq28 duplication (chrX:153229170-153433332), *de novo*. Phenotype: ASD, history of seizures, normal IQ (verbal IQ 108, performance IQ 125, full scale IQ 109).

- DECIPHER: Subject 262772, male, Xq28 duplication (chrX:153219789-153490319), maternal. No phenotype information; the subject carries two other duplications (chr3:685234-1166340 and chr17:31862-164024), both inherited.
- DECIPHER: Subject 252424, male, Xq28 duplication (chrX:153230084-153475911), *de novo*. No phenotype information; the subject carries an additional CNV (chr1:91820025-92502808 duplication, *de novo*).
- DECIPHER: Subject 253549, male, Xq28 duplication (chrX:153230084-153475911), maternal. No phenotype information; only CNV reported in the subject.
- DECIPHER: Subject 265379, male, Xq28 duplication (chrX:153230109-153485889), unknown inheritance. Phenotype: ID, hypotonia, sleep apnea. Only CNV reported in the subject.
- ECARUCA: Subject 4882, male, Xq28 duplication (chrX:153051459-153436637), maternal. Phenotype: ID, spasticity/increased tendon reflexes, macrocephaly, constipation, short stature.

  (Three ISCA individuals with duplications of this region —2 with developmental delay, 1 with ASD— were not added because their gender is unknown.)

3) Functional evidence

- *GDI1* encodes the GDP-dissociation inhibitor alpha (αGDI), which regulates the activity of small GTPases of the Rab family involved in intracellular vesicular trafficking. *Gdi1*-deficient mice exhibit membrane accumulation of Rab GTPases and decrease in the reserve pool of synaptic vesicles in the hippocampus, leading to altered synaptic plasticity and short-term memory deficits.[59]
- *GDI1* is highly expressed in the brain.

- **4) Comment**

  We identified two AGP probands with a recurrent Xq28 duplication corresponding to a genomic disorder recently described in ID. The identification of these two independent ASD cases, together with an ASD subject reported previously[58], implicates duplications of this region in ASD. Among the genes contained in the region, aberrant gene dosage of *GDI1* is likely to be responsible for the neurodevelopmental phenotype of subjects carrying this CNV since mutations in this gene are involved in nonsyndromic X-linked ID[53,54] and the expression of *GDI1* in the blood of individuals with the Xq28 duplication correlated with the severity of the phenotype.[55] Finally, *GDI1* is highly expressed in the brain and its function in intracellular vesicular trafficking together with the synaptic phenotype of the *Gdi1*-deficient mice make this gene particularly interesting with regard to the ASD/ID phenotypes.

Candidate genes affected by *de novo* CNV in AGP probands

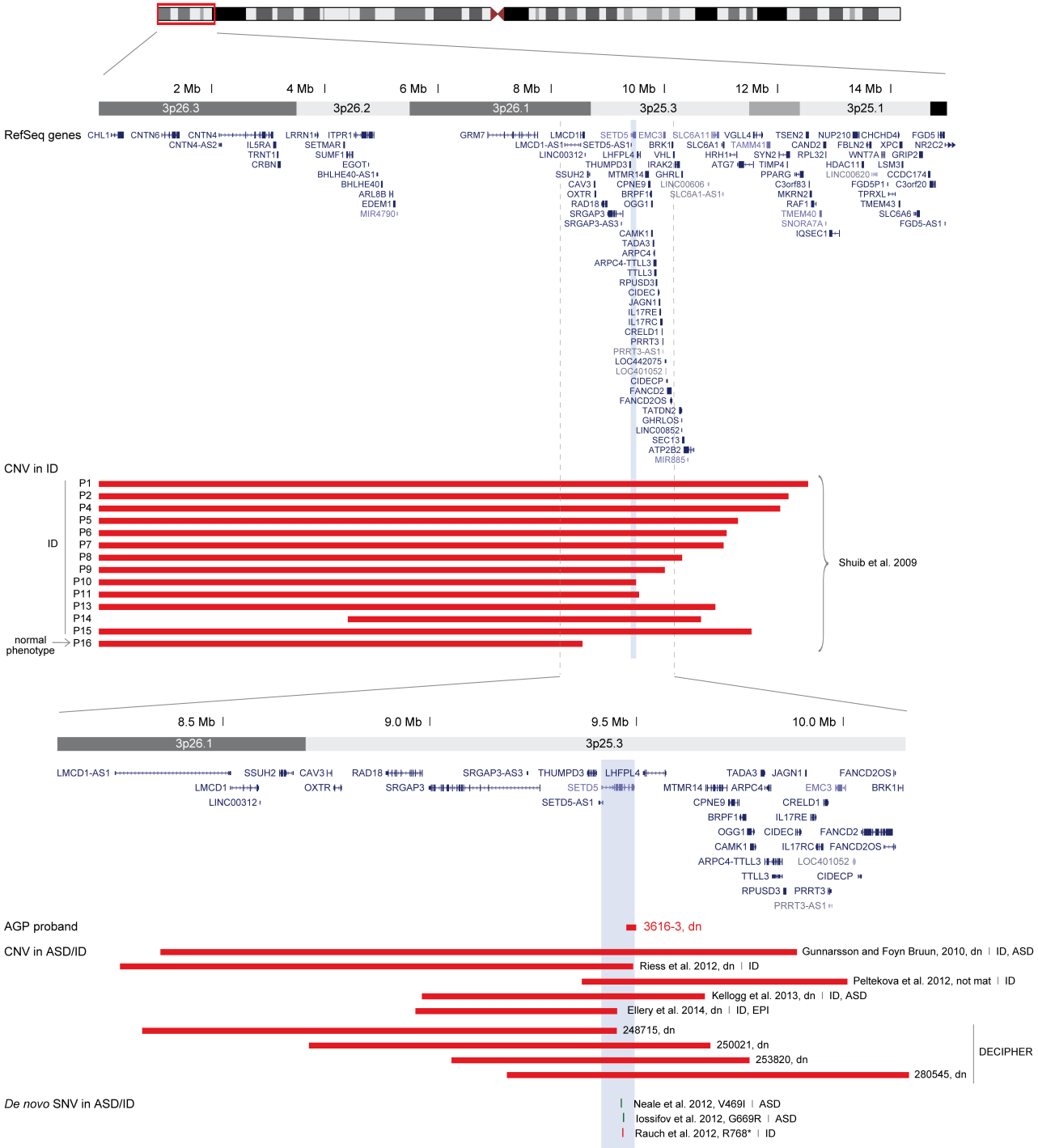


Figure S9. CNV and SNV identified in *SETD5* (SET domain containing 5) in chromosome 3p25.3

Chr3:1-15,000,000 (hg18). Abbreviations: ASD, autism spectrum disorder; dn, *de novo*; EPI, epilepsy; ID, intellectual disability; mat, maternal

1) AGP

- Proband 3616_3, male, 24 kb deletion including only *SETD5* (chr3:9474320-9498362), *de novo*. Phenotype: Born at term after normal pregnancy and delivery. No feeding or sleep problems; body length and head circumference during the first year of life

were consistently between 0 to –1 SD; walked at 15 mo, language delay (first words 24 mo), functional language, autism (meets criteria for autism on both ADI-R and ADOS), borderline IQ (WISC-III: verbal IQ 64, performance IQ 82, full scale IQ 70). Concerns about ADHD-like behavior in childhood, but did not fulfill criteria for ADHD and there are no concerns about hyperactivity at present. Normal vision and hearing; had sensory-integration training because of over and undersensitivity; asthma and hay fever, no other medical issues, no seizures. At 16 y, normal height and weight (50th centile), no obvious dysmorphic features (no ptosis, normal ears, nose and mouth), normal fingers and toes.

- No *SETD5* deletions among 4,768 controls and 4,875 parents.

2) Other human genetic evidence

- 3p deletion syndrome is a rare contiguous gene disorder resulting from terminal and interstitial deletions involving chromosome 3p25-p26. Common features include ID, growth retardation, microcephaly, dysmorphic facial features, and ptosis.[60] Rare individuals with a 3p26-p25 deletion and normal intelligence or only mild abnormalities have been reported, including several cases inherited from seemingly unaffected parents, indicating variable penetrance. There are over 30 individuals with 3p25-p26 deletions reported, the majority with large cytogenetically visible rearrangements, making delineation of the critical region and identification of candidate genes difficult. Shuib et al.[60] reported 14 individuals with cytogenetically visible deletions of 3p25, all with marked ID except one subject (P16) with a normal phenotype, ascertained incidentally because of recurrent miscarriages. Notably, the deletions in subjects with ID all overlapped *SETD5* whereas the subject without ID had the smallest terminal deletion, with a breakpoint proximal to *SETD5* (**Figure S9**). Here we review recently described cases with clinical features of distal 3p deletion syndrome carrying smaller interstitial deletions.

CNV:

- Gunnarsson and Foyn Bruun[61]: Female, 1.6 Mb 3p26.1-p25.3 deletion of 24 genes including *SETD5* (chr3:8305426-9885334), *de novo*. The girl was born at 37 weeks, with birth measures below –2 SD. She was noted to have dysmorphic facial features (hypertelorism, ptosis, strabism, flat and broad nasal root, long philtrum, downturned corner of the mouth, low set ears), bilateral overlap of the 2nd and 4th toes over the 3rd and 5th toes, atrial and ventricular septal defects, hypotonia, and severe developmental delay, with absent language at the age of 4 y. She showed autistic behavior; she smiled and laughed, but had poor eye contact and did not interact with other children, focusing on objects in her close vicinity. Vision, hearing, and growth were normal. At the age of 4, she had transient seizures, documented with EEG; a MRI showed asymmetry of thalamus. (Corresponds to subject 253231 in DECIPHER.)
- Riess et al.[62]: Female, 1.24 Mb 3p26.1-p25.3 deletion of 12 genes including *SETD5* (chr3:8250541-9491586), *de novo*. Normal birth measures, hypotonia during the first year of life, delayed psychomotor development. She started to walk and talk at the age of 2 y. On examination at the age of 3, she had large head (75th centile) compared to her height (10th centile), bilateral strabismus, large fontanelle, prominent forehead, depressed nasal bridge, thin upper lip and prominent philtrum. Ultrasound examinations of the heart and abdominal organs were normal.
- Peltekova et al.[63]: Female, 643 kb 3p25.3 deletion of 23 genes including *SETD5* (chr3:9367274-10010209), inheritance unknown (not maternal, father's DNA not available). The proband was born via cesarean section for breech presentation at 37 weeks gestation; birth weight 3rd–15th centile, length 5th centile, and head circumference 25th centile. She presented polydactyly in all four extremities, cleft palate, atrial septal defect, and bowel malrotation. At 1 year of age she developed tonic–clonic seizures and had episodes of aspiration pneumonia requiring placement of a gastrostomy tube. She was non-verbal as an adult, but smiled, laughed and was interactive. At 22 y her height and weight were <3rd centile, and head circumference was <5th centile. She had low anterior hairline, short forehead, thick eyebrows with synophrys, microphthalmia, downslanting palpebral fissures, blepharophimosis with mild ptosis, hypotelorism, esotropia, low set and cupped ears, and scoliosis. Hands and feet were small, with bilateral syndactyly of the 2nd and 3rd toes. Brain MRI showed parenchymal volume loss and atrophy of all structures except the brain stem.
- Kellogg et al.[64]: Female, 684 kb 3p25.3 deletion of 7 genes including *SETD5* (chr3:8980098-9664733), *de novo*. The proband was born at full term, with normal birth weight (50th centile) and length (90th centile). She had strabismus (corrected surgically) and developmental delay. Obsessive-compulsive disorder, with repetitively smelling of various objects, was noted at 10 y. On examination at age 11, she had dysmorphic features, including prominent ear lobes, right ear pit, depressed nasal bridge, anteverted nares, long philtrum, and proximally placed thumbs. Height, weight and head circumference were 50th, 10th and 50th centile, respectively; the brain MRI was normal. Assessment at 5 y 8 mo placed her in the borderline to intellectually disabled range (Stanford-Binet IV), and she met criteria for autism spectrum on ADOS.

  The authors compared this individual with three previously reported cases with interstitial 3p25 deletions with ID and characteristic facial features[61-63] and identified a region of overlap including only three genes: *THUMPD3, SETD5* and *SETD5-AS1*, which could play a critical role in the neurocognitive phenotype of the 3p deletion syndrome.
- Ellery et al.[65]: Male, 486 kb 3p25.3 deletion of 6 genes including *SETD5* (chr3:8965201-9450984), *de novo*. The proband was born at term with a birth weight of 4600 g, after a pregnancy marked by gestational diabetes. Bilateral postaxial polydactyly, a single palmar crease, right preauricular pit, mild hypertelorism, anteverted nostrils and micrognathia were noted at birth. Hypotonia, feeding difficulties and developmental delay became evident afterwards. A diagnosis of Simpson–Golabi–Behmel syndrome was considered. He walked independently at 4 y but tired easily; at 6 y his vocabulary was limited to two-dozen words and occasional

short sentences. He developed grand mal seizures at 7 y, treated with carbamazepine. He had a disturbed sleep pattern with frequent waking. On examination at 8 y of age, he had epicanthic folds, mild hypertelorism, high palate, myopathic facies with coarse features, and pectus excavatum. He had poor muscle bulk and showed a partial Gower's sign. These features persisted into adulthood.

- DECIPHER: Subject 248715, female, 1.1 Mb 3p26.1-p25.3 deletion of 12 genes including *SETD5* (chr3:8330935-9450984), *de novo*. Phenotype: ID/DD, macrocephaly, coarse facial features, synophrys, gum hypertrophy, low posterior/trident hairline, general abnormalities of hair texture. Only CNV reported in the subject.
- DECIPHER: Subject 250021, male, 946 kb 3p25.3 deletion of 10 genes including *SETD5* (chr3:8724500-9671040), *de novo*. No phenotype information; only CNV reported in the subject.
- DECIPHER: Subject 253820, male, 700 kb 3p25.3 deletion of 9 genes including *SETD5* (chr3:9061621-9763580), *de novo*. Phenotype: ID, hypertelorism, micrognathia, low-set ears, hydrocephalus, horseshoe kidney. Only CNV reported in the subject.
- DECIPHER: Subject 280545, male, 1.1 Mb 3p25.3 deletion of 33 genes including *SETD5* (chr3:9186364-10290795), *de novo*. Phenotype: global developmental delay, stereotypic behavior, absent speech, abnormality of the corpus callosum, abnormality of the hypothenar eminence, small thenar eminence, facial asymmetry and blepharophimosis. Only CNV reported in the subject.
- Note that several *SETD5* exonic deletions are reported in controls from DGV among non-BAC based studies, all from Shaik et al.;[66] these deletions are likely to be a study-specific artifact.

SNV:

- Neale et al.[5] (exome sequencing in ASD): Subject 09C98906, male, *SETD5* missense variant, *de novo*. The variant is predicted to be either benign (PolyPhen2, SNAP, Mutation Taster) or damaging (SIFT); the substituted residue is variable according to ConSurf (1/9); GERP score 1.91 (constrained >2). Phenotype: ASD, verbal IQ 93, performance IQ 112, full scale IQ 97.
- Iossifov et al.[4] (exome sequencing in ASD): Subject 13576, female, *SETD5* missense variant (gene listed as *KIAA1757*), *de novo*. The variant is predicted to be damaging (PolyPhen2, SNAP and Mutation Taster; not scored by SIFT). The substituted residue is variable according to ConSurf (1/9); GERP score 5.69. Phenotype: ASD (no other information provided).
- Rauch et al.[41] (exome sequencing in nonsyndromic sporadic ID): Subject ER14209, female, *SETD5* nonsense mutation, *de novo*. Phenotype: Born full term, weight –0.95 SD, head circumference –1.46 SD, sitting at 8 mo, walking at 20 mo, first words at 48 mo, IQ 70, mild attention deficit disorder, no ASD, no seizures, strabism, recurrent infections, constipation, prominent finger joints, facial dysmorphisms. When last evaluated at 9 y, height –0.78 SD, head circumference –1.16 SD, spoke in fluent sentences.

3) Functional evidence

- The gene is predicted to be a methyltransferase; several other genes encoding methyltransferases have been shown to be altered in ASD/ID (see examples below).
- Expressed in the brain.

4) Other evidence

- % HI = 21.3 (0%-10%: likely to be haploinsufficient; 90%-100%: not likely to be haploinsufficient)
- Haploinsufficiency of *NSD1* (nuclear receptor binding SET domain protein 1), encoding a histone methyltransferase, causes Sotos syndrome, a neurodevelopmental disorder characterized by overgrowth, distinctive craniofacial appearance, and variable ID, sometimes associated with ASD.[67,68]
- Another gene encoding a histone methyltransferase, *EHMT1* (euchromatic histone-lysine N-methyltransferase), is involved in Kleefstra syndrome through deletions (9q34.3 deletion syndrome) or mutations. *EHMT1* haploinsufficiency is associated with ID and sometimes with ASD.[69] An AGP proband with a 9q34.3 deletion encompassing *EHMT1* was identified in the present study (proband 6259-3, **Table S7B**).

5) Comment

The minimal region of overlap between our proband and those of 9 other cases with deletions involving 3p25.3 reviewed here contains only *SETD5*. Remarkably, all deletions are *de novo*. This finding, together with the recent report of a *de novo* loss-of-function mutation in a subject with language delay and borderline IQ (70, like in our proband)[41] suggest that *SETD5* is involved in cognitive, social, and language development. Therefore, *SETD5* could be associated with these features in 3p deletion syndrome. Two other *de novo* variants, identified in whole exome studies in ASD,[4,5] are missense changes and their functional effect is difficult to predict.

Interestingly, the AGP proband with the intragenic *SETD5* deletion, as well as the subject with the *SETD5* loss-of-function mutation, have a borderline IQ (70), whereas other individuals with larger deletions have ID ranging from severe to moderate, suggesting that other genes in the region contribute to the cognitive deficit associated with 3p deletions. Similarly, our proband lacks some of the characteristic but variable features associated with the 3p deletion syndrome, including growth retardation and dysmorphic features, such as ptosis and depressed nasal bridge.

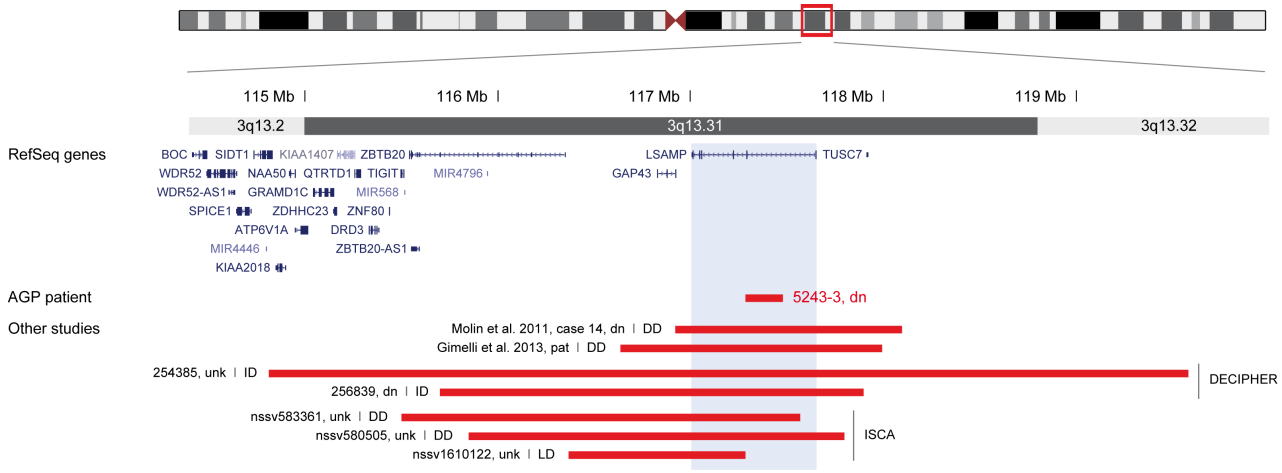


Figure S10. CNV identified in *LSAMP* (limbic system-associated membrane protein) in chromosome 3q13.31

Chr3:114,400,000-120,000,000 (hg18). Abbreviations: dn, *de novo*; DD, developmental delay; ID, intellectual disability; LD, learning disability; pat, paternal; unk, unknown inheritance

1) AGP

- Proband 5245_3, male, 192 kb deletion including only *LSAMP* (chr3:117285007-117477191), *de novo*. The deletion was shown to be mosaic; the percentage of mosaicism was estimated at 50% of deleted cells using a formula based on the deviation of the B allele frequency distribution.[70] Phenotype: born at 29 wks, intraventricular hemorrhage, mild cerebral palsy; autism (based on ADI-R and ADOS), low non-verbal IQ (<1st centile), language delay (1st centile), apraxia, abnormal sleep EEG without seizures; alopecia areata, no dysmorphic features. Multiplex family, a sister with ASD doesn't carry the CNV.
- No exonic deletions of *LSAMP* among 4,768 controls and 4,875 parents (1 control and 1 mother have intronic deletions).

2) Other human genetic evidence

CNV:

- Molin et al.[71]: Subject 14, male, 1.18 Mb 3q13.31 deletion (chr3:116922662-118098190) containing three genes, *GAP43*, *LSAMP* and *TUSC7*. Uncertain inheritance (absent in the mother and father not tested). Five year old boy with developmental delay; no other phenotype information provided. *GAP43* is found in growth cones of extending axons in the central nervous system. (Corresponds to subject 252520 in DECIPHER.)
- Gimelli et al.[72]: Girl with 1.36 Mb 3q13.31 deletion (chr3:116640577-118002810) containing three genes, *GAP43, LSAMP,* and *TUSC7*. The deletion was inherited from the father, who had slightly delayed psychomotor development but his cognitive level was not tested. The proband had developmental delay, clumsiness and attention deficit, associated with renal, vascular and skeletal anomalies.
- DECIPHER: Subject 254385, female, 4.8 Mb 3q13.2-q13.32 deletion involving *LSAMP* (chr3:114813585-119579912), inheritance unknown. Phenotype: ID, muscular hypotonia, downslanted palpebral fissures, joint laxity, open mouth. Only CNV reported in the subject.
- DECIPHER: Subject 256839, female, 2.2 Mb 3q13.31 deletion involving *LSAMP* (chr3:115701890-117895614), *de novo*. Phenotype: ID, broad nasal tip, short nose, anteverted nares. Only CNV reported in the subject.
- ISCA: Subject nssv583361, unknown gender, 2 Mb 3q13.31 deletion involving *LSAMP* (chr3:115499855-117567970), inheritance unknown. Phenotype: developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing.
- ISCA: Subject nssv580505, unknown gender, 1.9 Mb 3q13.31 deletion involving *LSAMP* (chr3:115850559-117797499), inheritance unknown. Phenotype: global developmental delay.
- ISCA: Subject nssv1610122, unknown gender, 918 kb 3q13.31 deletion involving *LSAMP* and *GAP43* (chr3:116367264-117285252), inheritance unknown. Phenotype: specific learning disability.

3) Functional evidence

- *LSAMP* encodes a neuronal surface glycoprotein belonging to the cell adhesion molecule (CAM) family, found in cortical and subcortical regions of the limbic system. During development of the limbic system, it is found on the surface of axonal membranes and growth cones, where it acts as a selective homophilic adhesion molecule, and guides the development of specific

patterns of neuronal connections. LSAMP mediates selective neuronal growth and axon targeting and also contributes to the guidance of developing axons and remodeling of mature circuits in the limbic system. This protein is essential for normal growth of the hippocampal mossy fiber projection.[73] *Lsamp*$^{-/-}$ mice have normal gross neuroanatomical organization but display heightened reactivity to novelty, reduced anxiety-like behaviors, impaired synaptic plasticity, and spatial memory deficit.[74,75] *LSAMP* has also been shown to function as a tumor suppressor gene.

- Brain expression: *LSAMP* is expressed on limbic regions but also, less intensely, in midbrain and hindbrain regions.

4) Other evidence

- % HI = 10.5 (likely to be haploinsufficient)

5) Comment

The minimal region of overlap between our proband and seven other cases with 3q13.31 deletions reviewed here contains only the *LSAMP* gene. All cases with known inheritance are *de novo* and would thus support involvement of deletions in this region in neurodevelopmental disorders. Other cases with deletions involving only *LSAMP* or deleterious mutations are needed to implicate this gene in ASD and ID.

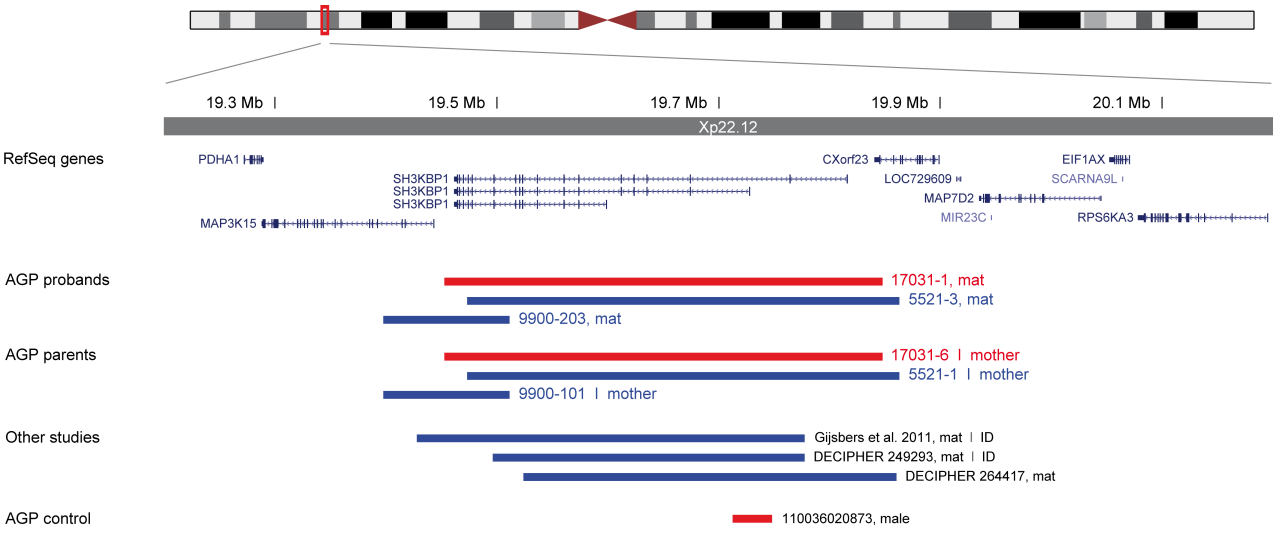


Figure S11. CNV identified in *SH3KBP1* (SH3-domain kinase binding protein 1) in chromosome Xp22.12

ChrX:19,200,000-20,200,000 (hg18). Abbreviations: ID, intellectual disability; mat, maternal

1) AGP

- Proband 17031_1, male, deletion involving *SH3KBP1* and *CXorf23* (chrX:19450969-19845766), maternal. *CXorf23* encodes a protein of unknown function. Phenotype: autism (based on ADI-R and ADOS), language delay, functional language, verbal IQ 79; physical examination at 5 y 8 mo revealed no dysmorphic features, neurological examination was normal except for a deficit in coordination and gross and fine motor development; no seizures, normal sleep EEG. Sporadic case, mother unaffected, negative family history of neuropsychiatric disorders.
- Proband 5521_3, male, partial duplication of *SH3KBP1* and *CXorf23* (chrX:19471138-19861338), maternal. Phenotype: autism (based on ADI-R and ADOS), below average IQ (<1st centile), nonverbal, seizure disorder, coarse facial features. Sporadic case, mother unaffected.
- Proband 9900_203, male, partial duplication of *SH3KBP1* and *MAP3K15* (chrX:19396116-19509785), maternal. Expression study in cell lines showed no increased expression of *SH3KBP1*. The CNV also involves *MAP3K15*, encoding a mitogen-activated protein (MAP) kinase expressed in the brain. Phenotype: autism (meets criteria on ADI-R and ADOS), language delay (first words 18 mo, first phrases 42 mo), limited language, moderate ID (Bayley Scales of Infant Development II, mental developmental index 45), strabismus (like his mother), normal physical exam, no dysmorphic features, no epilepsy. Sporadic case, mother unaffected.
- 1 of 2,022 male controls carries a *SH3KBP1* deletion overlapping only exon 1 of transcript variant 2, not present in other isoforms (chrX:19710729-19746114); no deletions among 2,441 fathers.

2) Other human genetic evidence

CNV:

- Gijsbers et al.[76]: Male, 349 kb duplication overlapping partially 2 genes, *SH3KBP1* and *MAP3K15* (chrX:19425768-19775308), maternal. Phenotype: growth retardation, severe ID, absent or delayed speech, stereotypic movement of head and hands, bitemporal narrowing, narrow palpebral fissures, deep-set eyes, large mouth, widely spaced teeth. The healthy mother and grandmother carried the same Xp22.12 duplication and showed skewed X inactivation. The authors interpreted the CNV as potentially pathogenic.
- DECIPHER: Subject 249293, male, intragenic *SH3KBP1* duplication (chrX:19494754-19775308), maternal. Phenotype: ID, narrow forehead, abnormality of the eyebrow, deep set eyes, abnormality of the mouth, widely spaced teeth, proportionate short stature, abnormal CNS myelination. Only CNV reported in the subject.
- DECIPHER: Subject 264417, male, duplication overlapping partially 2 genes, *SH3KBP1* and *CXorf23* (chrX:19521667-19858019), maternal. No phenotype information; only CNV reported in the subject.

3) Functional evidence

- The *SH3KBP1* gene encodes CIN85, an endocytic scaffold protein that facilitates protein-protein interactions and has been implicated in numerous cellular processes including apoptosis, cytoskeletal rearrangement, cell adhesion and clathrin-dependent endocytosis. CIN85 plays a role in receptor internalization, including dopamine receptors and epidermal growth factor receptor.[77,78] CIN85 localizes at synapses and interacts with the scaffold protein S-SCAM via dendrin.[79]
- Mice lacking CIN85 exon 2 (present in both isoforms expressed in the central nervous system) show hyperactivity and increased exploratory behavior, but no alterations in synaptic plasticity or learning and memory. These mice show increased dopamine and dopamine D2 receptors in the striatum, due to impaired endocytic internalization of D2 receptors.[77]
- Expressed in the brain.

4) Comment

The deletion in proband 17031_1 results in a *SH3KBP1* null allele. The intragenic duplication in DECIPHER subject 249293 is also likely to disrupt the gene, acting as a deletion. In contrast, the functional consequence of the partial duplications observed in AGP probands 5521-3 and 9900-203, and in the two other subjects ([76] and DECIPHER 264417) are difficult to predict and thus can not be counted as evidence in favor of the implication of *SH3KBP1* alterations in neurodevelopmental disorders, in the absence of expression studies. In one proband (9900_203) in whom we evaluated *SH3KBP1* mRNA expression in cell lines, no alteration was observed. Other affected males with deletions, intragenic duplications or deleterious mutations of *SH3KBP1* are required to implicate loss of function of this gene in ASD and ID.

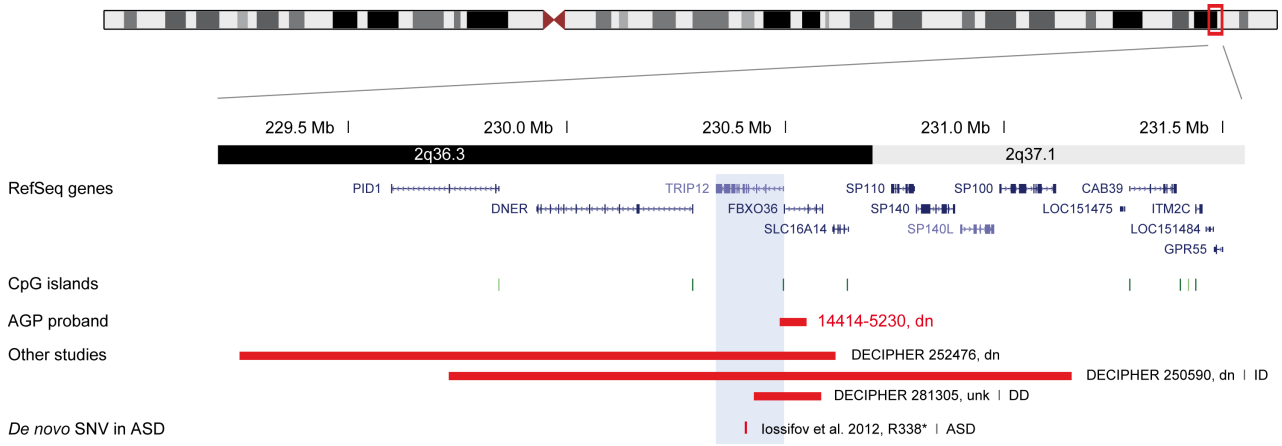


Figure S12. CNV and SNV identified in *TRIP12* (thyroid hormone receptor interactor 12) in chromosome 2q36.3

Chr2:229,200,000-231,550,000 (hg18). Abbreviations: ASD, autism spectrum disorder; DD, developmental delay; dn, de *novo*; ID, intellectual disability; unk, unknown inheritance

1) AGP

- Proband 14414_5230, male, 60 kb deletion involving 2 genes, *TRIP12* and *FBXO36* (chr2:230486629-230547253), *de novo*. The deletion only involves the first exon of *TRIP12*, which is non-coding, but seems to be part of the promoter of the gene since a CpG island is located in this region. The other deleted gene, *FBXO36*, encodes a F-box protein that plays a role in ubiquitination and is

not likely to be haploinsufficient (% HI = 52.9). Phenotype: autism (ADI-R and ADOS positive), no language delay (first words 12 mo, first phrases 24 mo), verbal, Griffiths at 5 y 9 mo: language DQ 87, performance DQ 78, global DQ 79.

- No *TRIP12* deletions among 4,768 controls and 4,875 parents.

2) Other human genetic evidence

CNV:

- DECIPHER: Subject 252476, female, 1.4 Mb deletion involving *TRIP12* (chr2:229250832-230614988), *de novo*. No phenotype information; only CNV reported in the subject.
- DECIPHER: Subject 250590, female, 1.4 Mb deletion involving *TRIP12* (chr2:229728861-231153046), *de novo*. Phenotype: ID, delayed speech and language development, epicanthus, hypermetropia, low hanging columella, palpebral edema, broad philtrum, thin upper lip vermilion, wide mouth. Only CNV reported in the subject.
- DECIPHER: Subject 281305, male, 144 kb deletion involving *TRIP12* (chr2:230432282-230576372), inheritance unknown. Phenotype: global developmental delay and cystic renal dysplasia. Only CNV reported in the subject.

SNV:

- Iossifov et al.[4] (exome sequencing in ASD): Subject 12867, female, *TRIP12 de novo* nonsense mutation. Phenotype: ASD (no other information provided).

3) Functional evidence

- *TRIP12* encodes a HECT-type E3 ubiquitin-ligase, which plays a role in degradation of ubiquitin fusion substrates and can regulate chromatin function to maintain genome integrity.[80]
- Expressed in the brain.

4) Other evidence

- % HI = 5.3 (highly likely to be haploinsufficient)
- *HUWE1*, mutated in X-linked ID, also encodes a HECT-type E3 ubiquitin-ligase involved in the ubiquitin-fusion degradation (UFD) pathway. Double knock-down of *HUWE1* and *TRIP12* results in additive stabilization of an UFD substrate, suggesting functional redundancy between both proteins.[81]

5) Comment

The evidence from the AGP subject is not very strong; although the *TRIP12* deletion is *de novo*, only the first non-coding exon is involved and the effect on the protein is difficult to predict. Expression studies are necessary to interpret the clinical significance of this CNV. The support from other human genetic studies is limited, but the *de novo* loss-of function *TRIP12* mutation in a female with ASD identified in an exome sequencing study, together with the function of the protein, make this gene an interesting candidate for ASD.

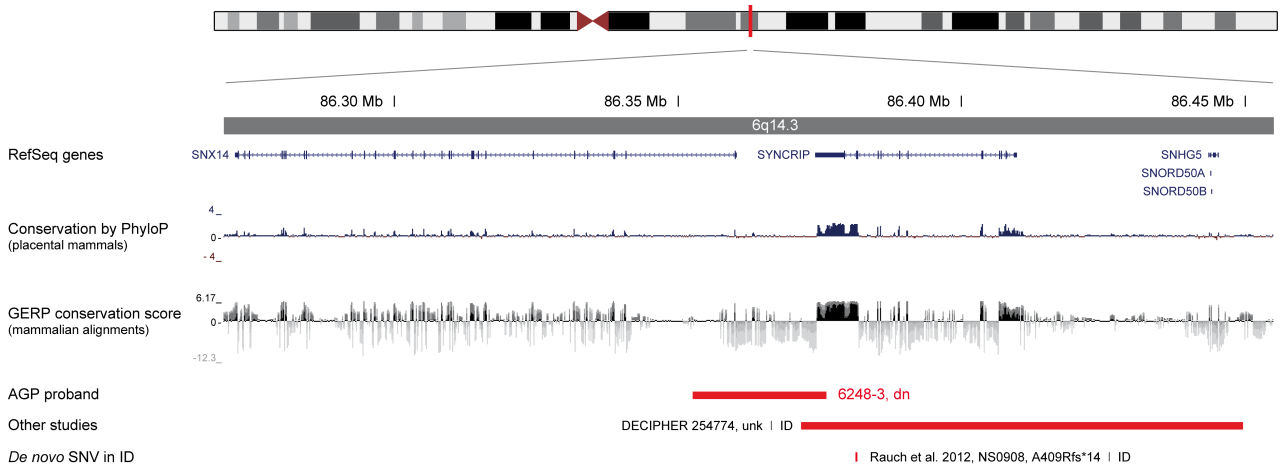


Figure S13. CNV and SNV identified in *SYNCRIP* (synaptotagmin binding, cytoplasmic RNA interacting protein) in chromosome 6q14.3

Chr6:86,270,000-86,455,000 (hg18). Abbreviations: dn, *de novo*; ID, intellectual disability; unk, unknown inheritance

1) AGP

- Proband 6248_3, male, 23 kb deletion involving 2 genes, *SYNCRIP* and *SNX14* (chr6:86352577-86376159), *de novo*. The deletion only involves the 3' untranslated region (UTR) of *SYNCRIP*; this region might contain regulatory elements that are crucial for gene expression. *SNX14* encodes a member of the sorting nexin family that are involved in intracellular trafficking; this gene is not predicted to be haploinsufficient (% HI = 58.6), but its contribution to the phenotype can not be ruled out. Phenotype: autism, severe ID, no language; born by C-section with increased height and head circumference. One absence seizure at 11 y, with abnormal EEG, probably related to medications taken at that time, no other seizures afterwards. At 15 y: height >+3 SD, macrocephaly (head circumference >+3 DS) (his unaffected father and 2 brothers are also very tall and have macrocephaly); long and narrow hands and feet, 2 café-au-lait spots, normal neurological exam.
- No *SYNCRIP* deletions among 4,768 controls; 1 carries a partial *SYNCRIP* duplication (chr6:86382351-87814038); no CNV overlapping *SYNCRIP* in 4,875 parents.

2) Other human genetic evidence

CNV:

- DECIPHER: Subject 254774, male, 78 kb deletion including *SYNCRIP* (chr6:86371713-86449627), inheritance unknown. The deletion also involves *SNHG5* (snoRNA host gene) as well as *SNORD50A* and *SNORD50B* (snoRNAs). Only CNV reported in the subject. Phenotype: ID.

  Note that there are several large deletions (>5 Mb) overlapping this gene in DECIPHER and ISCA, not reviewed here.

SNV:

- Rauch et al.[41] (exome sequencing in nonsyndromic sporadic ID): Subject NS0908, female, *SYNCRIP* frameshift mutation, *de novo*. Phenotype: severe nonsyndromic sporadic ID (IQ <50), myoclonic astatic seizures (onset at 13 mo), no ASD, MRI at 24 mo: prominent lateral ventricles.

3) Functional evidence

- *SYNCRIP* encodes a nuclear ribonucleoprotein implicated in mRNA processing mechanisms including mRNA stability and transport, RNA editing and splicing and localized mRNA translation. SYNCRIP is a component of mRNA transport granules in dendrites.[82] Selective mRNA transport, local translation and subsequent protein synthesis in neuronal dendrites are part of the fundamental mechanisms involved in synaptic plasticity, learning and memory.
- Other RNA binding proteins have been implicated in ID and ASD, including FMRP, involved in fragile X syndrome, the most frequent monogenic cause of ID and ASD, and ZC3H14, involved in recessive non-syndromic ID.[83]
- Expressed in the brain.

4) Other evidence

- % HI = 1.5 (highly likely to be haploinsufficient)

5) Comment

The AGP proband carries a *de novo SYNCRIP* deletion that only affects the highly conserved 3'UTR of the gene. Expression studies are necessary to determine the effects of the deletion. Although this gene does not have support from many other CNV studies, the *de novo* loss-of function *SYNCRIP* mutation reported in an individual with ID and the function of the encoded protein make this gene a strong candidate for involvement in neurodevelopmental disorders. Additional cases with small deletions affecting *SYNCRIP* and/or with deleterious mutations are required to implicate this gene definitely in ASD, ID and other neurodevelopmental phenotypes.

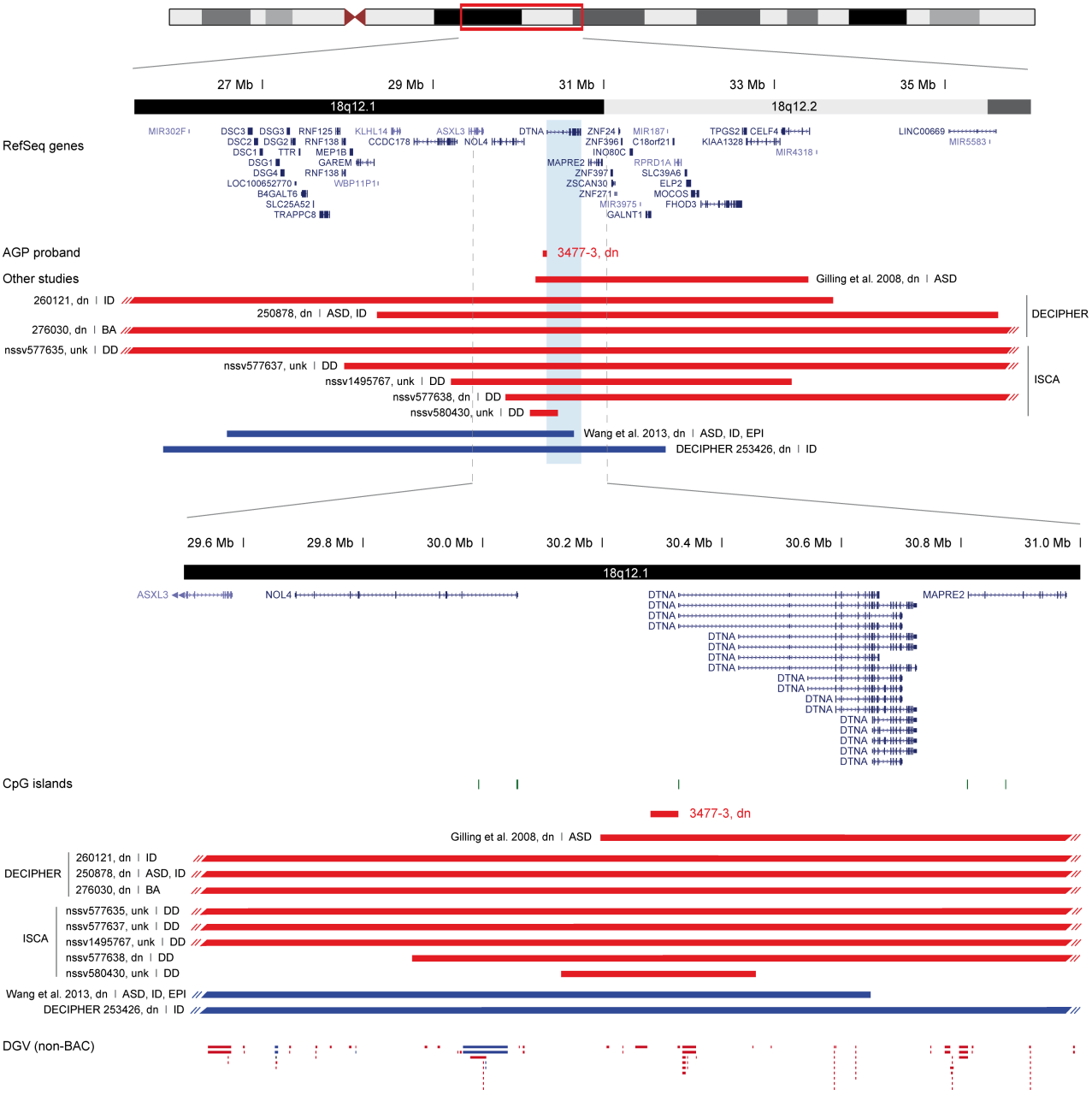


Figure S14. CNV identified in *DTNA* (dystrobrevin alpha) in chromosome 18q12.1

Chr18:25,500,000-36,000,000 (hg18). Abbreviations: ASD, autism spectrum disorder; BA, behavioral/psychiatric abnormality; dn, *de novo*; DD, developmental delay; EPI, epilepsy; ID, intellectual disability; unk, unknown inheritance

1) AGP

- Proband 3477_3, male, 47 kb deletion involving *DTNA* (chr18:30280260-30327512), *de novo*. Only the first non-coding exon of the longer isoforms of *DTNA* (2, 5, 7 and 17) is disrupted by the deletion. It may be part of the promoter of the gene since a CpG island is located in the region. Phenotype: ASD (ASD on ADI-R, autism on ADOS), verbal; WISC-R at 5 y 7 mo: verbal IQ 88, performance IQ 88, full scale IQ 86; no cardiovascular or neuromuscular abnormalities.
- No *DTNA* deletions among 4,768 controls and 4,875 parents.

2) Other human genetic evidence

CNV:

- Gilling et al.[84]: Female, *de novo* translocation t(5;18)(q34;q12.2), with a 3.2 Mb deletion at the 18q breakpoint encompassing 20 genes including *DTNA* (chr18:30197000-33392000); the breakpoint on 5q did not contain any known genes. She also carried a

1.27 Mb deletion on chromosome 4q35, containing two genes (*MTNR1A* and *FAT*); inherited from her father. Phenotype: born at term, prolonged delivery with asphyxia noted at birth. She presented mild cerebral palsy, language delay, autism (met criteria on ADI-R and ADOS), no ID (WAIS-R at 34 y: verbal IQ 78, performance IQ 105, full scale IQ 88), high-grade myopia, no dysmorphism, hyperflexible joints, no seizures.

- DECIPHER: Subject 260121, female, 13.4 Mb 18q12.1 deletion (chr18:20286120-33684898), *de novo*. Phenotype: ID, abnormality of the face, malformation of the heart and great vessels. Only CNV reported in the subject.
- DECIPHER: Subject 250878, male, 7.3 Mb deletion including *DTNA* (chr18:28338083-35619727), *de novo*. Phenotype: ID (full scale IQ 40-50), autism, delayed speech and language development, flat occiput, hypotelorism, narrow nasal bridge, narrow nares, narrow mouth, high palate, wide intermamillary distance, absent nipples, proximal placement of thumb, hypotonia, tall stature, abnormality of the male genitalia. This individual also carries a duplication involving 3 genes, *HAO1*, *TMX4* and *PLCB1* (chr20:7022125-8482355), inherited from a normal parent.
- DECIPHER: Subject 276030, male, 14.5 Mb deletion including *DTNA* (chr18:25269813-39765123), *de novo*. The deletion is mosaic (percentage of mosaicism not indicated). Phenotype: behavioral/psychiatric abnormality, autoagression and motor delay. Only CNV reported in the subject.
- ISCA: Subject nssv577635, unknown gender, 10.9 Mb 18q12.1 deletion including *DTNA* (chr18:25278473-36237614), inheritance unknown. Phenotype: developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing.
- ISCA: Subject nssv577637, unknown gender, 11.9 Mb 18q12.1 deletion including *DTNA* (chr18:27945491-39904182), inheritance unknown. Phenotype: global developmental delay, muscular hypotonia, short stature.
- ISCA: Subject nssv1495767, unknown gender, 4 Mb deletion including *DTNA* (chr18:29207760-33198709), inheritance unknown. Phenotype: developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing.
- ISCA: Subject nssv577638, unknown gender, 7.3 Mb 18q12.1 deletion including *DTNA* (chr18:29881080-37228316), *de novo*. Phenotype: global developmental delay, strabismus, stridor.
- ISCA: Subject nssv580430, unknown gender, deletion involving only *DTNA* (chr18:30131834-30456329), inheritance unknown. Only the first non-coding exon of the longer isoforms (2, 5, 7 and 17) is disrupted by the deletion, which is very similar to the one observed in the AGP proband 3477-3. Phenotype: developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing.
- Wang et al.[85]: Female, 4.1 Mb duplication including *DTNA* (chr18:26587700-30649280), *de novo*. Phenotype: autistic disorder (met criteria on ADI-R and ADOS), language delay, non-verbal, mild ID, focal epilepsy, short stature (5th centile), no dysmorphic features, mild myopia.
- DECIPHER: Subject 253426, male, 5.9 Mb duplication including *DTNA* (chr18:25834124-31721175), *de novo*. Phenotype: moderate ID (WAIS-IV full scale IQ 51), no ASD, recurrent seizures, facial dysmorphism, short stature, Chiari malformation, cryptorchidism, strabismus. Only CNV reported in the subject.

The smallest region of overlap among these 11 cases contains only the *DTNA* gene.

SNV:

- Ichida et al.[86]: Heterozygous missense mutation in *DTNA* in a large pedigree with left ventricular non compaction (family NVLNC-09), no neurological/behavioral phenotype described. No other *DTNA* mutations reported in the literature.

3) Functional evidence

- *DTNA* belongs to the dystrobrevin subfamily of the dystrophin family, like *DMD*, involved in Duchenne's/Becker muscular dystrophy, sometimes associated with ID/ASD.[87,88] Dystrobrevin alpha is part of the transmembrane dystrobrevin-associated protein complex (DPC), which participates in synaptic transmission at the neuromuscular junction, long-term memory and synaptic plasticity.[89]
- α-dystrobrevin knockout mice exhibit mild muscular dystrophy but show no obvious CNS defects, likely reflecting coexpression of the homolog β-dystrobrevin, which is predominantly expressed in the brain.[90] Double mutants lacking both α-dystrobrevin and β-dystrobrevin exhibit synaptic and behavioral defects similar to those seen in dystrophin-deficient mice.[91] Both dystrobrevin isoforms are required for the maturation and function of a subset of inhibitory synapses in the cerebellum and for correct execution of motor behaviors that depend on cerebellar integrity.[91]
- Expressed in the brain.

4) Other evidence

- % HI = 3.2 (highly likely to be haploinsufficient)
- Upregulation of *DTNA* has been reported in the temporal cortex of subjects with autism[92] and in the prefrontal cortex of individuals with bipolar disorder.[93]

- Dystrobrevin immunostaining is severely reduced at the sarcolemma of individuals with Duchenne muscular dystrophy and to a lesser extent in individuals with Becker muscular dystrophy.[94]

5) Comment

The gene function and the description of overlapping CNV would make this an excellent candidate gene, potentially pathogenic. However, the deletion in the AGP proband only affects the first non-coding exon of the longer isoforms, so it is difficult to know if it is deleterious. Expression studies are necessary to assess the effect on mRNA in this individual. Furthermore, description of other cases with small CNV overlapping *DTNA* or SNV are required to implicate this gene specifically. Indeed, given that the majority of the overlapping CNV are very large and contain many genes, it is difficult to ascribe pathogenicity to alterations of a single gene in the interval.

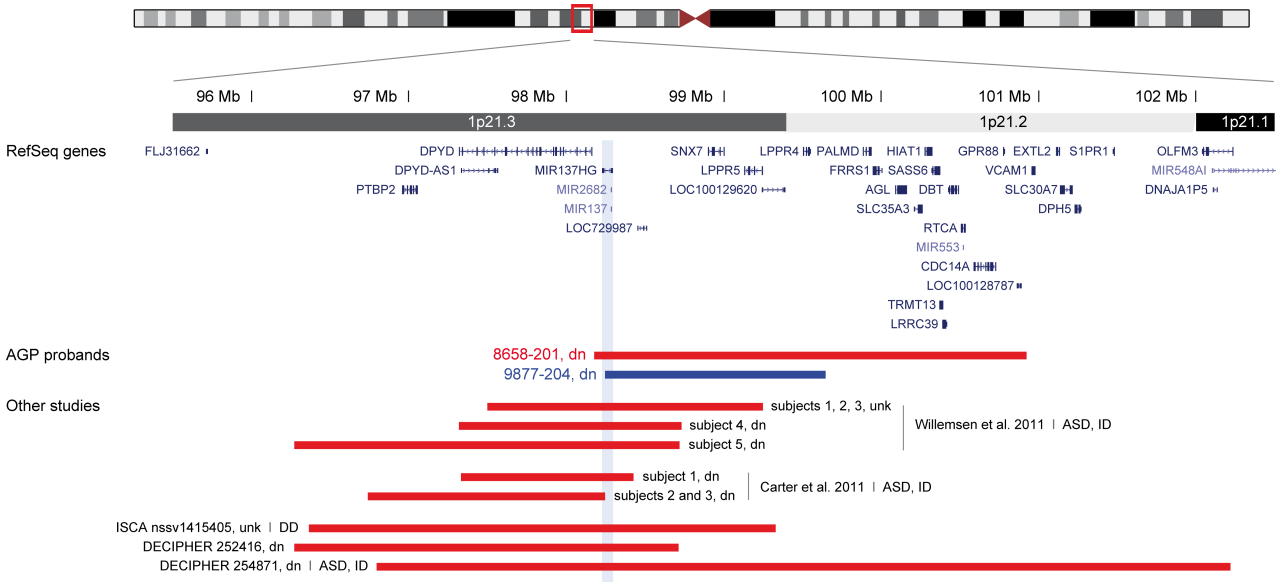


Figure S15. CNV overlapping *MIR137* (microRNA 137) in chromosome 1p21.3

Chr1:95,500,000-102,500,000 (hg18). Abbreviations: ASD, autism spectrum disorder; dn, *de novo*; DD, developmental delay; ID, intellectual disability; unk, unknown inheritance

1) AGP

- Proband 8658_201, female, 2.7 Mb 1p21.3-p21.2 deletion containing 22 genes including *MIR137* (chr1:98175622-100923952), *de novo*. Phenotype: autism on ADI and ADOS, comorbid ADHD, no language delay; low average IQ (WASI at 21 y: verbal IQ 78, performance IQ 88, full scale IQ 81); overweight, height and head circumference 50th centile, high pain tolerance, no epilepsy.
- Proband 9877_204, male, 1.4 Mb 1p21.3-p21.2 duplication containing 8 genes including *MIR137* (chr1:98247355-99645560), *de novo*. Phenotype: autism on ADI-R and ADOS, language delay (first words 36 mo, first phrases 60 mo), functional language, mild ID (WISC-R at 11 y: verbal IQ 51, performance IQ 80, full scale IQ 64), normal height and head circumference, weight -1.6 SD, no dysmorphic features, normal physical exam, no epilepsy.

  Both CNV in AGP probands also involve the *LPPR4* and *LPPR5* genes (lipid phosphate phosphatase-related proteins, types 4 and 5). *LPPR4* is specifically expressed in neurons and involved in axonal outgrowth during development and regenerative sprouting;[95] no haploinsufficiency score available (no information for *LPPR5*).
- No *MIR137* deletions among 4,768 controls; 1 carries a *MIR137* duplication (chr1:97673140-98319409); no CNV in 4,875 parents.

2) Other human genetic evidence

CNV:

- Willemsen et al.[96]: Chromosome 1p21.3 microdeletions comprising *DPYD* and *MIR137* associated with ID in 3 sibs and 2 unrelated subjects; the minimal region of overlap includes only *DPYD* and *MIR137*. The individuals displayed decreased expression of both precursor and mature miR-137 levels, as well as increased expression of the downstream targets *MITF*, *EZH2*, and *KLF4*. *DPYD* is involved in autosomal recessive dihydropyrimidine dehydrogenase deficiency; the significance of a defect in only one allele is uncertain.

Subjects 1, 2 and 3: siblings carrying a 1.75 Mb 1p21.3 deletion (chr1:97500000-99250000), inheritance unknown (parents deceased). Subject 1: male, borderline-mild ID (verbal IQ 65, performance IQ 90, full scale IQ 73), features of ASD, tendency to overeat, remarkably shy and friendly behavior, weight 90th centile, deep set eyes, broad nasal tip, long ears, thick lower lip, myopia. Subject 2: male, mild-moderate ID (verbal IQ < performance IQ 70, full scale IQ 52), features of ASD, tendency to overeat, self mutilation, aggressive outbursts, remarkably shy and friendly behavior, speech deficits, weight >98th centile, broad nasal tip, long ears, macrostomia, thick lower lip, astigmatism. Subject 3: female, mild-moderate ID (no other information available).
Subject 4: male, 1.41 Mb 1p21.3 deletion (chr1:97320000-98730000), *de novo*. Phenotype: mild ID (verbal IQ 59, performance IQ 71, global IQ 62), features of ASD, tendency to overeat, remarkably shy and friendly behavior, speech deficits, weight 98th centile, deep set eyes, astigmatism, myopia, broad nasal tip, full cheeks, thick lower lip, micrognathia and long ears.
Subject 5: female, 2.45 Mb 1p21.3 deletion (chr1:96270000-98720000), *de novo*. Phenotype: mild ID (global IQ 66), aggressive outbursts, remarkably shy and friendly behavior, weight >98th centile, full cheeks, long ears, thick lower lip.

- Carter et al.[97]: Subject 1: male, 1.1 Mb 1p21.3 deletion (chr1:97332167-98424667), *de novo*. Phenotype: severely delayed language, ID, autism. The individual also carries a balanced translocation t(9;21)(p13.3;q22.1) and has a *PTCHD1* missense mutation. Both abnormalities are inherited from the mother. The translocation was also transmitted to a healthy sister. Subjects 2 and 3: siblings with a 1.5 Mb 1p21.3 deletion (chr1:96742150-98243813), *de novo*. In their mother, the deleted region from chromosome 1p21.3 was inserted into chromosome 10. Subject 2: female, severe language delay, adaptive skills severely delayed, autism. Subject 3: male, language delay, no ID (full scale IQ 99), ASD. Both sibs had mild dysmorphic features, including upslanting palpebral fissures and small joint hypermobility.
- ISCA: Subject nssv1415405, unknown gender, 3 Mb 1p21.3 deletion (chr1:96362589-99332669), inheritance unknown. Phenotype: developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing.
- DECIPHER: Subject 252416, female, 2.4 Mb 1p21.3 deletion (chr1:96274145-98715464), *de novo*. No phenotype information; only CNV reported in the subject.
- DECIPHER: Subject 254871, male, 5.4 Mb 1p21.3-p21.1 deletion (chr1:96792350-102220124), *de novo*. Phenotype: autism, ID, spotty hyperpigmentation, precocious puberty. Only CNV reported in the subject.

3) Functional evidence

- The mature microRNA transcript miR-137 regulates neuronal maturation: overexpression of miR-137 inhibits dendritic and spine morphogenesis in newborn cells in the adult hippocampus and in cultured hippocampal neurons, whereas a reduction in miR-137 had opposite effects,[98] miR-137 has also been shown to modulate neurogenesis in adult neural stem cells.[99] Significant enrichment of miR-137 at the synapses of cortical and hippocampal neurons suggests a role in regulating local synaptic protein synthesis machinery.[96]
- Expressed in the brain, enriched in neurons, at the synaptic compartment.

4) Other evidence

- An intronic SNP in *MIR137* was strongly associated with schizophrenia in a mega-analysis combining genome-wide association study data from over 40,000 individuals.[100]

5) Comment

1p21.3 deletions of variable sizes overlapping *MIR137* reported in 11 individuals with ASD and/or ID. All cases in which inheritance is known originated *de novo*. Interestingly, both overexpression and inhibition of miR-137 had significant but opposite effects on dendritic development of hippocampal neurons, suggesting that the *MIR137* gene may be dosage sensitive, and that both the deletion and duplication observed in AGP probands could interfere with neuronal maturation.

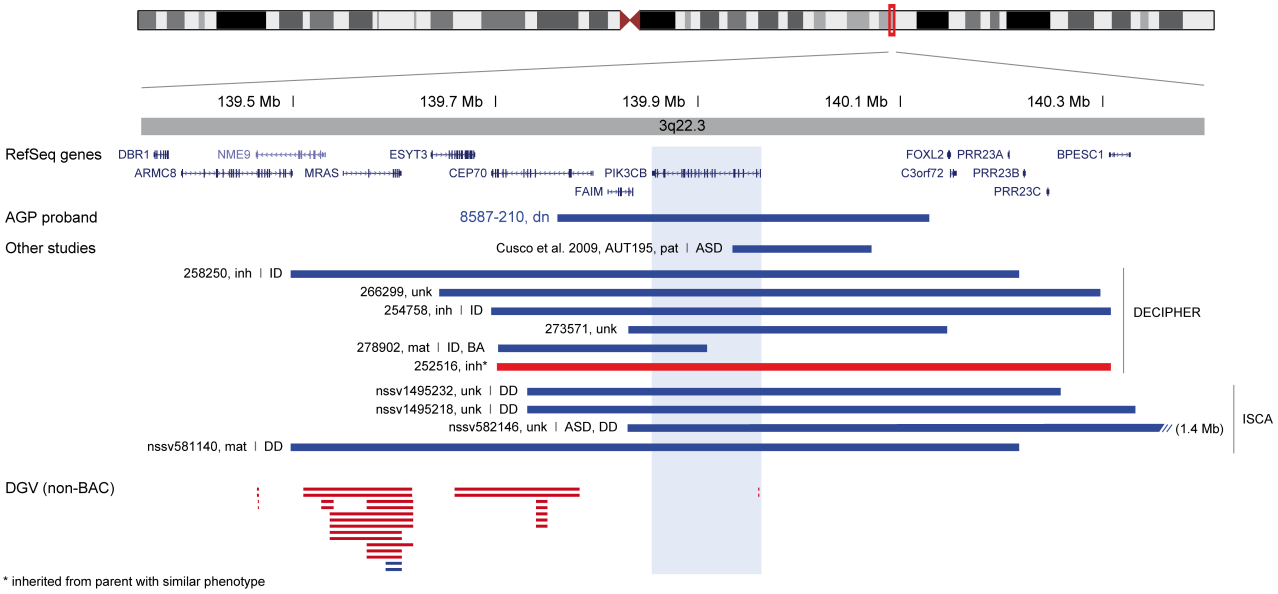


Figure S16. CNV identified in *PIK3CB* (phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit beta) in chromosome 3q22.3

Chr3:139,350,000-140,400,000 (hg18). Abbreviations: ASD, autism spectrum disorder; BA, behavioral/psychiatric abnormality; dn, *de novo*; DD, developmental delay; inh, inherited; ID, intellectual disability; mat, maternal; pat, paternal; unk, unknown inheritance

1) AGP

- Proband 8587_210, female, 366 kb duplication involving *PIK3CB, CEP70* and *FAIM* (chr3:139760896-140127703), *de novo*. The entire *PIK3CB* gene is duplicated, which could result in increased expression, leading to excessive phosphatidylinositol 3-kinase (PI3K) activity. Phenotype: autism (ADI-R and ADOS positive), no language delay (first words 10 mo, first phrases 12 mo), verbal, WISC-III at 13 y 9 mo: verbal IQ 92, performance IQ 78, full scale IQ 84.
- No deletions or duplications of *PIK3CB* among 4,768 controls and 4,875 parents.

2) Other human genetic evidence

CNV:

- Cusco et al.[101]: Subject AUT195, male, 3q22.3 *PIK3CB* partial duplication (chr3:139934042-140070771), paternal. Phenotype: autism, mild ID, unilateral sensorineural deafness, no dysmorphism, no seizures. The functional consequences of a partial duplication are unknown.
- DECIPHER: Subject 258250, female, 720 kb 3q22.3 duplication involving *PIK3CB* (chr3:139497567-140217021), inherited from normal parent. Phenotype: cognitive impairment, Dandy-Walker malformation. Only CNV reported in the subject.
- DECIPHER: Subject 266299, male, 650 kb 3q22.3 duplication involving *PIK3CB* (chr3:139644292-140297350), inheritance unknown. Phenotype: microtia. This individual carries another duplication (chr3:160280710-160878458), inheritance unknown.
- DECIPHER: Subject 254758, male, 610 kb 3q22.3 duplication involving *PIK3CB* (chr3:139695765-140307747), inherited from normal parent. Phenotype: ID, short attention span, microcephaly, long face, high anterior hairline, downslanted palpebral fissures, depressed nasal tip, prominent nasal bridge, macrotia, micrognathia, abnormality of the pinna, prominent ears, slender build, scoliosis, mitral regurgitation, atrial septum defect, cryptorchidism, large hands. Only CNV reported in the subject.
- DECIPHER: Subject 273571, male, 315 kb 3q22.3 duplication involving *PIK3CB* (chr3:139830639-140145866), inheritance unknown. No phenotype information; only CNV reported in the subject.
- DECIPHER: Subject 252516, female, 606 kb 3q22.3 deletion involving *PIK3CB* (chr3:139701603-140307606), inherited from parent with similar phenotype. No phenotype information; only CNV reported in the subject.
- DECIPHER: Subject 278902, female, 207 kb 3q22.3 duplication involving *PIK3CB* (chr3:139701603-139908842), maternal. Phenotype: moderate ID and behavioral/psychiatric abnormality. The phenotype of the mother is unknown. Only CNV reported in the subject.

- ISCA: Subject nssv1495232, unknown gender, 527 kb 3q22.3 duplication involving *PIK3CB* (chr3:139730880-140257978), inheritance unknown. Phenotype: developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing.
- ISCA: Subject nssv1495218, unknown gender, 600 kb 3q22.3 duplication involving *PIK3CB* (chr3:139730880-140331637), inheritance unknown. Phenotype: developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing.
- ISCA: Subject nssv582146, unknown gender, 1.4 Mb 3q22.3-q23 duplication involving *PIK3CB* (chr3:139830647-141256588), inheritance unknown. Phenotype: autism, developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing.
- ISCA: Subject nssv581140, unknown gender, 719 kb 3q22.3 duplication involving *PIK3CB* (chr3:139497567-140217080), maternal. Phenotype: developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing.

3) Functional evidence

- *PIK3CB* encodes an isoform of the catalytic subunit PI3Kbeta of phosphoinositide 3-kinase (PI3K). These signaling molecules activate a wide range of downstream targets that regulate multiple cellular processes, including intracellular trafficking of proteins, cell proliferation, migration and survival.
- Expressed in the brain.

4) Other evidence

- % HI = 0.5 (highly likely to be haploinsufficient)
- PI3K is regulated by the fragile X mental retardation protein (FMRP),[10] and is elevated in fragile X syndrome Fmr1-knockout mice,[102,103] suggesting that dysregulated PI3K signaling may underlie the synaptic impairments in fragile X syndrome. Accordingly, PI3K antagonists rescue fragile X syndrome phenotypes, including dysregulated synaptic protein synthesis, excess AMPA receptor internalization, and increased spine density.[102]

5) Comment

The only report identified in the literature is a partial *PIK3CB* duplication in a subject with ASD, paternal;[101] the functional consequences of a partial duplication are difficult to predict. Eight cases of whole *PIK3CB* duplication are described in DECIPHER and ISCA, 4 are inherited and for the other 4 the inheritance is unknown. Thus, the duplication of *PIK3CB* identified in the AGP proband 8587-210 is the only one *de novo*. No similar duplications overlapping this gene were found among AGP controls, parents or population controls in DGV. Taken together these findings suggest that *PIK3CB* duplications could represent a risk factor for ASD/ID, associated with incomplete penetrance and/or variable expressivity. Further studies in larger samples of cases and controls are necessary to confirm this hypothesis.

Examples of candidate genes affected by inherited CNV in AGP probands

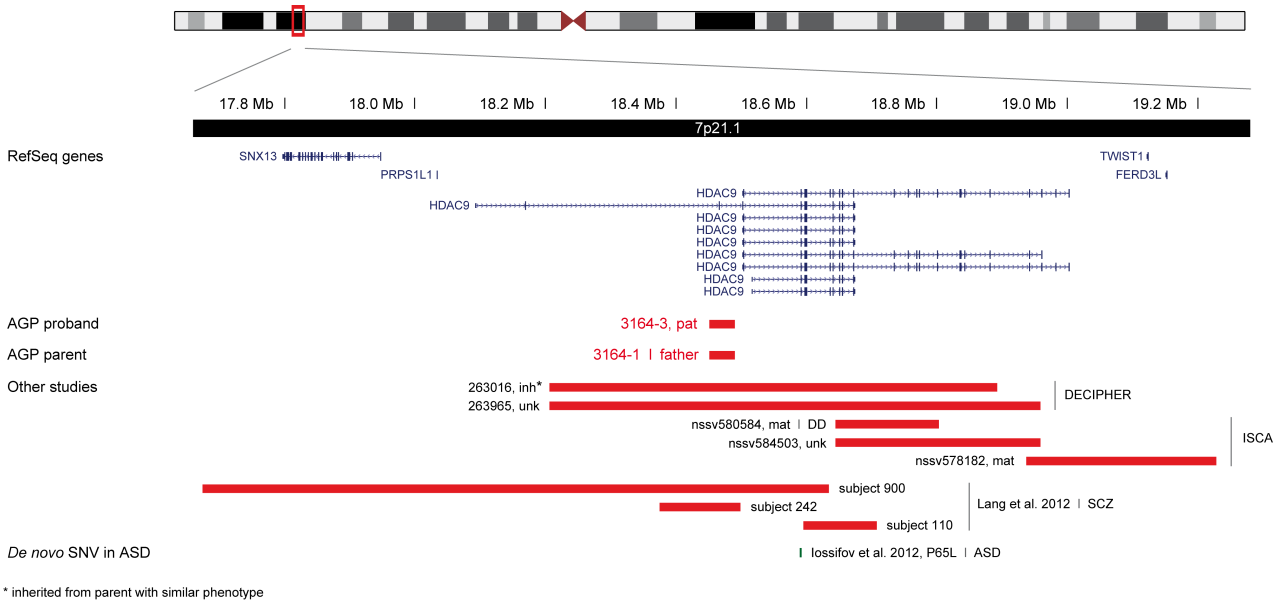


Figure S17. CNV and SNV identified in *HDAC9* (histone deacetylase 9) in chromosome 7p21.1

Chr7:17,660,000-19,280,000 (hg18). Abbreviations: ASD, autism spectrum disorder; dn, *de novo*; DD, developmental delay; inh, inherited; mat, maternal; pat, paternal; SCZ, schizophrenia; unk, unknown inheritance

1) AGP

- Proband 3164_3, male, rare 40 kb deletion involving only *HDAC9* (chr7:18450792-18490822), paternal. Phenotype: autism (ADI-R and ADOS positive), language delay (first words 54 mo, first sentences 60 mo), verbal, mild ID (PPVT-III at 6 y 10 m: verbal IQ 67); multiplex family, affected sibling not yet tested. Phenotype information about father not available.
- No *HDAC9* exonic deletions among 4,768 controls and 4,874 parents (1 deletion in the father of proband 3164-003).

2) Other human genetic evidence

CNV:

- DECIPHER: Subject 263016, female, *HDAC9* deletion (chr7:18206711-18892382), inherited from parent with similar phenotype. No phenotype information; only CNV reported in the subject.
- DECIPHER: Subject 263965, female, *HDAC9* deletion (chr7:18206712-18958442), inheritance unknown. No phenotype information; only CNV reported in the subject.
- ISCA: Subject nssv580584, unknown gender, *HDAC9* deletion (chr7:18644447-18803445), maternal. Phenotype: global developmental delay, seizures.
- ISCA: Subject nssv584503, unknown gender, *HDAC9* deletion (chr7:18644447-18958471), inheritance unknown. Phenotype: abnormal facial shape, facial asymmetry.
- ISCA: Subject nssv578182, unknown gender, 7p21.1 deletion of 3 genes: *HDAC9, TWIST1* and *FERD3L* (chr7:18937678-19227544), maternal. Phenotype: craniosynostosis (defects in *TWIST1* cause autosomal dominant craniosynostosis type 1).
- Lang et al.[104]: Three schizophrenia individuals with exonic *HDAC9* deletions among 3391 cases (inheritance unknown); no *HDAC9* deletion in 3181 controls.

SNV:

- Iossifov et al.[4] (exome sequencing in ASD): Subject 13076, male, *HDAC9* missense variant, *de novo*. The variant appears to be damaging (PolyPhen2, SIFT, PANTHER), affecting a highly conserved residue (GERP 5.93, ConSurf 9/9). Phenotype: ASD (no other information provided).

3) Functional evidence

- *HDAC9* encodes a histone deacetylase, expressed in the brain. Histones play a critical role in transcriptional regulation, cell cycle progression, and development.
- Expressed in the brain.

4) Other evidence

- % HI = 2.9 (highly likely to be haploinsufficient)
- The *HDAC* family of genes has already been involved in ASD/ID through *HDAC4* (involved in brachydactyly-mental retardation syndrome) and *HDAC8* (X-linked ID, mutations are responsible for Cornelia de Lange syndrome).

5) Comment

In addition to the *HDAC9* paternal deletion identified in AGP proband 3164-003, we identified 5 other overlapping deletions in subjects in DECIPHER and ISCA, as well as three deletions in schizophrenia. The deletion was inherited in 3 subjects; no information was available for the others. No *HDAC9* exonic deletions were observed among the AGP 4,768 controls and 4,874 parents; in addition, no deletions overlapping this gene were found in DGV. Taken together these findings suggest that *HDAC9* deletions could represent a risk factor for ASD, ID and schizophrenia, associated with incomplete penetrance/variable expressivity. Further studies in larger samples of cases and controls are necessary to confirm this hypothesis.

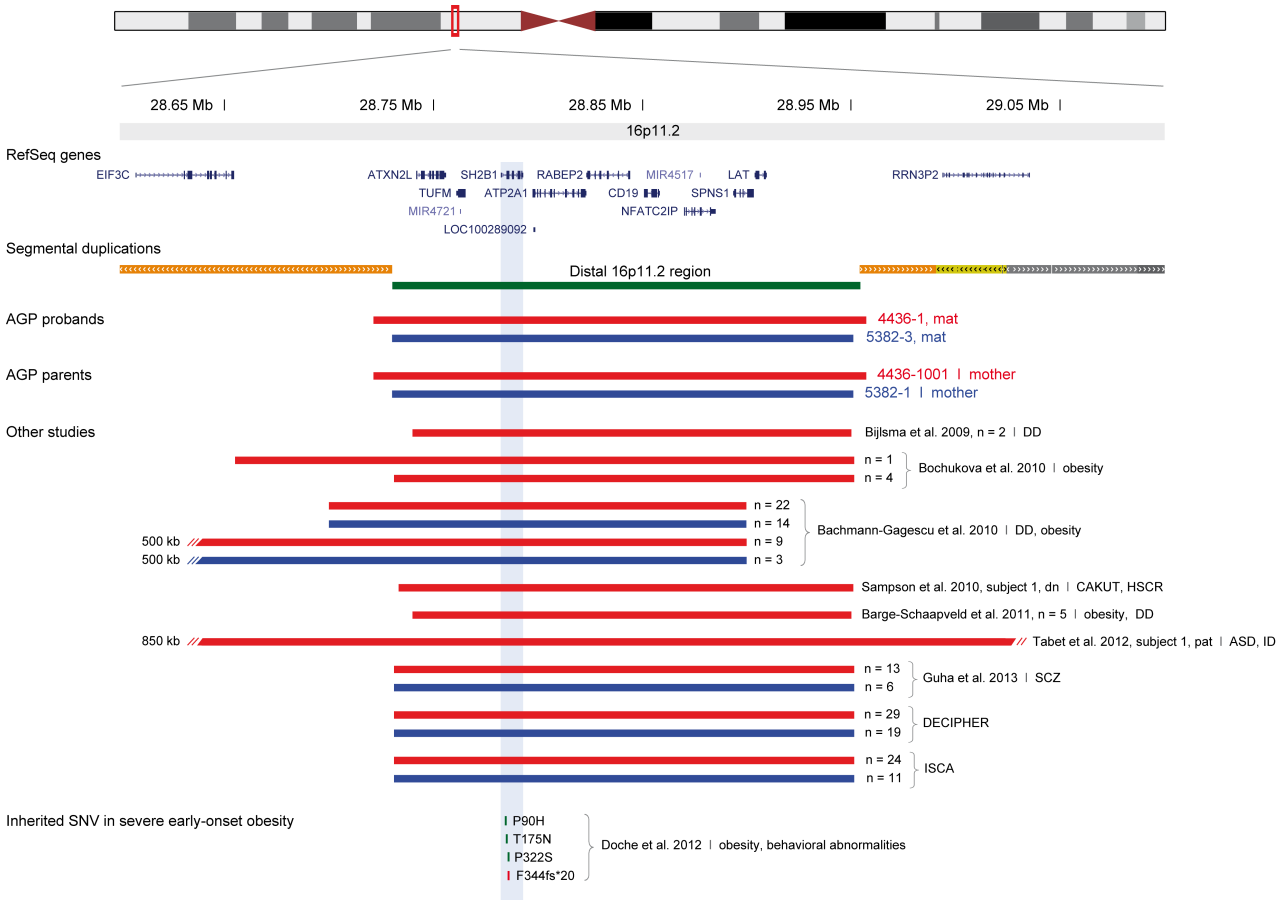


Figure S18. CNV identified in the distal 16p11.2 region containing *SH2B1* (SH2B adaptor protein 1)

Chr16:28,600,000-29,100,000 (hg18). Abbreviations: ASD, autism spectrum disorder; CAKUT, congenital anomalies of the kidney and urinary tract; DD, developmental delay; HSCR, Hirschsprung disease; ID, intellectual disability; mat, maternal; pat, paternal; SCZ, schizophrenia.

Longer CNV encompassing the distal 16p11.2 region and the proximal 16p11.2 region (29.5-30.1 Mb) involved in ASD, ID, and regulation of body mass index are not shown here.

1) AGP

- Proband 4436_1, male, distal 16p11.2 deletion including *SH2B1* (chr16:28721599-28957155), maternal. Phenotype: ASD diagnosis, no language delay (first words and phrases 15 mo), verbal, WISC-IV at 8 y 8 mo: verbal IQ 95, performance IQ 67, full scale IQ 76. Sporadic case.

- Proband 5382_3, male, distal 16p11.2 duplication including *SH2B1* (chr16:28730274-28950951), maternal. Phenotype: autism (ADI-R and ADOS positive), language delay (first words 42 mo, first phrases 48 mo), verbal; PPVT-IV verbal IQ 82, Leiter-R Brief performance IQ 97 (both at 11 y 10 m). Sporadic case.
- 1 distal 16p11.2 deletion among 4,768 controls and 1 deletion among 4,875 parents; no reciprocal duplications in controls or parents (the mother of proband 5382_3 was excluded from the microarray analyses after quality checks, her duplication was identified during qPCR validation of the CNV in her son).

2) Other human genetic evidence

CNV:

- Deletions at distal 16p11.2, with a minimal common region of 220 kb (28.73–28.95 Mb), have been implicated in early-onset obesity and developmental delay,[105-108] and in other variable phenotypes, including behavioral problems such as ASD and ADHD, anomalies of the kidney and urinary tract and Hirschsprung disease.[109,110] Whereas deletions appeared to be significantly enriched in populations with early-onset obesity or with developmental delay,[105,108] reciprocal duplications were not enriched in cases compared to controls.[105]
- A recent meta-analysis in large clinical cohorts with developmental delay, ID, ASD and congenital malformations referred for genetic testing found deletions at distal 16p11.2 in 23 of 31516 cases and in 2 of 13696 controls (OR 5, *P* = 0.01).[111] Reciprocal duplications were found in 25 of 31516 cases and in 3 of 13696 controls (OR 3.62, *P* = 0.02). Analysis of three ASD cohorts (AGP, SSC, and AGRE; n = 3955) found 1 deletion (OR 1.73, *P* = 0.53) and 1 duplication (OR 1.15, *P* = 1). The lack of a significant effect of these CNV in ASD was suggested to be due to the relatively small sample size.[111]
- Tabet et al.[110]: Male, 847 kb 16p11.2 distal deletion containing *SH2B1* (chr16:28401454-29249055), paternal. Autism, severely delayed speech, childhood-onset obesity, IQ 47. At age 19, he was tall (+2.5 SD), with troncular obesity (+4 SD). The father was described as being non talkative, introverted and having few social relationships.
- Guha et al.[112]: Deletions at distal 16p11.2 were reported in schizophrenia, in 13 of 13850 cases (0.094%) and 3 of 19954 controls (0.015%) (OR 6.25 [95% CI, 1.78-21.93]; *P* = 0.001). The rate of duplications in the region was not significantly different between cases and controls: 6 of 13850 cases (0.043%) vs 13 of 19954 controls (0.065%).

SNV:

- Doche et al.[113]: The minimal deleted interval contains nine genes, including *SH2B1*, which plays a role in the regulation of body weight and glucose metabolism in mice (see below). Mutation screening of *SH2B1* in 300 individuals with severe early-onset obesity revealed five mutations, one frameshift and three missense (including one found in two subjects). Mutation carriers exhibited childhood-onset obesity, hyperphagia, insulin resistance and short stature as adults. Neurobehavioral phenotypes included social isolation, speech and language delay and aggression. All mutations were inherited from overweight/obese parents reported to also have variable behavioral abnormalities. The mutations were absent from 500 controls.

3) Functional evidence

- *SH2B1* encodes an adaptor protein that binds to a large range of receptor tyrosine kinases and is thus involved in multiple biological pathways, including leptin and insulin signaling. The widely expressed scaffold protein SH2B1 binds to the receptors for nerve growth factor, insulin and insulin-growth factor 1, and has been implicated in neuronal differentiation and neurite outgrowth.[114,115]
- Expressed in the brain.
- *Sh2b1* deficient mice develop obesity and diabetes, a phenotype rescued by neuron-specific expression of *SH2B1*.[116]

4) Other evidence

- % HI = 18.7 (likely to be haploinsufficient)

5) Comment

SH2B1 haploinsufficiency is clearly implicated in early-onset obesity. Recent evidence suggests that distal 16p11.2 deletions could also be involved in neurodevelopmental phenotypes, associated with incomplete penetrance and variable expressivity. Although a significant enrichment has been reported in samples with developmental delay/ID, the risk effect appears to be weak compared to other recurrent CNV.[111] The implication of deletions at distal 16p11.2 in schizophrenia[112] and the description of maladaptive behaviors in individuals carrying *SH2B1* mutations,[113] lend further support to their role as risk factors. The involvement of distal 16p11.2 duplications in ID and ASD is difficult to assess at present, since they have not been found to be consistently enriched in cases. Further studies, comparing the frequency of distal 16p11.2 deletions and duplications in larger samples of cases and controls, are needed to clarify the impact of these CNV in neurodevelopmental disorders.

SUPPLEMENTAL TABLES

Table S1A. Autism strict and spectrum classifications

ASD diagnostic category	Phenotype classifications	
	ADI-R	ADOS
Strict	Autism	Autism
Spectrum	Autism	NA
	Autism	ASD
	ASD	Autism
	NA	Autism

NA, not available or not administered

Table S1B. Quality control – Family and control sample breakdown

Quality control filters	Initial	Filter 1 Low call rate	Filter 2 Mendelian errors	Filter 3 Gender-mismatch	Filter 4 Duplicates	Filter 5 High LRR/ BAF SD	Filter 6 Excess calls	Filter 7 Excess *de novos*	Filter 8 Peri-centrom.	Filter 9 Large chrom. abnorm.	Filter 10 Incomplete phenotype data	Filter 11 All ancestries	Filter 12 European-only
# Single probands	51	56	55	55	55	90	106	106	106	106	106	106	102
# Proband + mother duo	10	30	30	30	30	100	125	125	125	131	126	126	106
# Proband + father duo	12	31	31	31	31	113	127	127	127	142	137	137	119
# Complete trios*	2772	2677	2620	2613	2606	2268	2161	2158	2155	2126	2077	2077*	1820
Total # families	**2845**	**2794**	**2736**	**2729**	**2722**	**2571**	**2519**	**2516**	**2513**	**2505**	**2446**	**2446**	**2147**
# Father + mother only (with or without relatives)	258	277	273	273	273	309	324	324	326	332	380	199	196
# Fathers only	12	21	22	22	22	60	69	69	69	68	73	28	28
# Mothers only	16	27	27	27	27	52	60	60	61	62	67	32	32
# Relatives only	0	0	0	0	0	0	0	0	0	0	0	4	4
# Technical controls	9	9	9	8	8	8	8	8	8	8	8	8	8
Control datasets													
SAGE (dbGaP)[a]	1847	1847	–	1847	1829	1815	1793	–	1792	1769	–	1769	1166[#]
OC (European-only)[b#]	511	509	–	509	501	475	440	–	437	433	–	433	234[#]
HABC (dbGaP)[c]	2860	2860	–	2857	2809	2658	2571	–	2570	2566	–	2566	1240

CNV detection and quality control evaluation: For samples that passed the SNP and intensity QC, genome-wide CNVs were detected using a multiple-algorithm approach to maximize sensitivity and specificity of CNV calling. For a detailed description see the Supplemental information of Pinto et al.[117] Briefly, CNVs were identified by using QuantiSNP,[118] iPattern,[119] and PennCNV;[120] the family-based CNV detection option of PennCNV was used to confirm inheritance.

We excluded CNVs when they failed stringent QC criteria: <5 probes and low confidence score (QuantiSNP log Bayes factor <15); if CNVs resided in regions of extreme GC content (>70%); or if they were within centromere proximal cytobands. CNVs detected by QuantiSNP and iPattern in one individual with a minimum of 5 consecutive probes covering at least 5 kb of sequence were merged using outer probe boundaries (i.e., union of the 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 the presence of large abnormalities in chromosomes X and Y (that is, in addition to the algorithm calling). As a final step, we joined CNVs that appeared to be artificially split by either of the calling algorithms and also removed CNVs that spanned known large assembly gaps in hg18 (greater than 200 kb).

Filter descriptions: A total of 9,050 individuals from 2,845 ASD families were genotyped as part of Stages 1([117]) and 2 and those passing QC filters were used in the rare CNV analysis. Incomplete families, where proband-father/mother duos passed QC filters were also analysed for CNVs. The number and the composition of families remaining after each filtering step is indicated. Counts may increase or decrease after each step, as removal of individuals in some instances will break complete trios into proband+parent duos or single probands. **Filter 1**: Low call rate or high missingness. **Filter 2**: High Mendelian error rate (with or without proband gender mismatch); families with unresolved gender mismatches were excluded; families where both parents had a gender mismatch without high Mendelian error rate were kept and the parents' gender swapped. Families with high Mendelian rate for one parent only were excluded. **Filter 3**: Proband with gender mismatch but no Mendelian error (i.e., another sib of the same family was genotyped instead) resulted in exclusion of the whole family. Technical controls with gender mismatches were excluded. **Filter 4**: Any duplicate samples/families had one sample/family excluded. **Filter 5**: Samples with high standard deviation (SD) of log R ratio of intensities (LRR) and/or B allele frequency (BAF), or showing extreme/wide intensities. **Filter 6**: Samples with excess of CNV calls by at least one of the algorithms, except those with fragmented calls due to large chromosome abnormalities. **Filter 7**: Samples with excess of *de novos* (which were confirmed to be false positives with experimental validation). **Filter 8**: Lack of CNV calls after three filters (removing pericentromeric calls, <30 kb size, 50% overlap with segmental duplication

blocks). **Filter 9**: Large chromosomal abnormalities >7.5 Mb. A list of chromosome abnormalities detected in probands can be found in **Table S1C**, and for parents and controls in **Table S17B**. Subjects were removed if one or more CNVs were found to be cell-line artifacts after experimental validation. If a parent failed QC at this step, his/her family was not excluded. **Filter 10**: No phenotype data in database or proband did not meet full criteria for ASD. **Filter 11**: Parents that passed QC but are parents of probands that failed QC. **Filter 12**: European ancestry only.

Ancestry: Ancestry for each of the four samples (AGP cases, and SAGE, HABC, and OC controls) was inferred by eigenvector decomposition and clustering. To identify European subjects from the Ontario controls, we used the multidimensional scaling (MDS) function of PLINK to cluster the OC subjects with HapMap-CEU. The remaining three samples had been described and analyzed for ancestry in previous studies.[117,121,122] We used the results from those published studies to identify subjects of inferred European ancestry for this study. In each of the published studies, ancestry was inferred by using SpectralGem[123] to analyze thousands of high quality SNPs genotyped for all subjects. After the clustering step from SpectralGem, it was simple to identify groups of European ancestry because a substantial fraction of the contributing AGP sites were European.

Data from 2,446 families passed all QC steps (13% of subjects excluded), adding 1,359 new families to the combined analysis. Of the new families, 1,168 families were European and 191 were of other ancestries. As described in detail in Pinto et al.[117], to avoid confounding by ancestry, all downstream CNV analyses used European-only cases (n=2,147) and controls (n=2,640). For the analyses presented in **Figure S1**, we extended the number of controls to include 1,843 subjects from other ancestries —517 SAGE and 1,326 HABC non-Europeans controls that passed QC— giving a total of 4,768 control subjects (2,022 males and 2,746 females) from all ancestries to be compared to ASD cases of all ancestries.

***** *de novo* CNVs were detected and confirmed in a total of 2077 complete trios passing array QC plus 19 families in which at least one of the parents failed initial array QC but additional experimental validation in both parents confirmed the presence of *de novo* CNVs, giving a total of 2096 complete trios of all ancestries (1838 European trios = 1820 + 18) studied.

[a] **Study of Addiction: Genetics and Environment (SAGE) cohort**: Both raw intensities and genotypes were obtained from 1,847 SAGE control subjects from NHGRI-dbGaP (accession: phs000092.v1.p1) as part of the larger SAGE case-control study[124]. The consented sample included 31% males and 69% females, with a mean age of 39.2 y (SD 9.1); 73% of subjects self-identified as European-American, 26% as African-American and 1% as other (http://zork.wustl.edu/gei /study_description.html). Subjects may have had exposure to alcohol (and possibly to other drugs), but did not meet criteria for alcohol or other drug dependence. The subset of control dataset used in the specific CNV analyses in this paper is composed of 1,166 unrelated European control samples that passed all quality control filters (75% had DNA extracted from whole blood and 25% from cell lines), composed of 370 males and 796 females.

[b] **Ontario Colorectal Cancer case-control study cohort (OC)**: 433 unrelated European control subjects from the population-based Colorectal Cancer case-control study, recruited randomly from the province of Ontario in Canada (Ontario Familial Colorectal Cancer Registry, OFCCR) as described elsewhere[125] and genotyped with the Illumina 1M single array.[126] The OC control sample consisted of 199 females and 234 males with mean age of 61.8 y (range: 27-78); all subjects were self-identified as non-Hispanic whites and estimated to be of European ancestry from their genotypes. All DNA samples were extracted from whole blood. The GenomeStudio v. 2010.3, with the clustering algorithm GenTrain2 and a GenCall cutoff of 0.15 was used to generate genotypes. The same quality control procedures applied to the ASD family samples and SAGE controls were used here. For the main CNV analysis, we used only the 234 OC males.

[c] **HealthABC (HABC)**: 1,240 unrelated European control subjects from the whole-genome-association study of visceral adiposity in the Health, Aging, and Body Composition (HealthABC or HABC)[127] were used in the main analyses. The HABC cohort studied the factors that contribute to disability and the decline in function of healthier older persons, with a particular emphasis on changes in body composition. The HealthABC study recruited 3,075 70-79 year-old community-dwelling adults (41% African-American, remainder were white-European), who were initially free of mobility and activities of daily living disability. Genotyping was performed by the Center for Inherited Disease Research (CIDR) using the Illumina Human1Mv3 (duo) BeadChip system like most of the AGP data, providing excellent comparability with the case dataset. Both raw intensities and genotypes were obtained from JAAMH-dbGaP (accession: phs000169.v1.p1) for 2,860 samples, resulting in 1,240 European (637 males and 603 females) and 1,326 non-European samples after QC. Samples were excluded from the dataset in case of sample failure, genotypic sex mismatch, or first-degree relative of an included individual based on genotype data.

[#] To try to balance the number of male and female controls used in the various analyses (by stage and by platform), we only included male samples from the OC dataset (n=234), and excluded 86 female SAGE samples of European ancestry (i.e., all males [n=370] and 796 females were included).

Abbreviations: chrom. abnorm., chromosomal abnormalitities; pericentrom., pericentromeric; M, males.

Table S1C. Quality control – Chromosome abnormalities detected in probands

(Chromosome abnormalities in parents and controls are listed in **Table S17B**)

Stage	Sample ID	Sex	Family type	DNA source	Cytoband	Anomaly type	Karyotype	Comment
Confirmed chromosomal abnormalities								
1	5467_3	M	SPX	Blood	1q42.3-q44	High intensity, BAF split	N/A	13.7 Mb 1q42-q44 duplication (1:233476547-247165725), *de novo*; confirmed by qPCR
1	14270_3930	F	SPX	Blood	6q25.3-q27	High intensity, BAF split	46,XX.ish der(22)t(6;22)(6q25.3;p11.2)pat(6qtel+)	10 Mb 6q25.3-q27 duplication (6:160773919-170761395); confirmed by FISH, resulting from a balanced translocation in the father
1	13137_1543	F	UNK	Blood	8p12-8q12.1	8p-q duplication	47,XX,+r[10]/46,XX[70]	27 Mb 8p12-8q12.1 duplication (8:31928590-58996070); karyotype: mosaic supernumerary ring chromosome
1	5420_3	M	UNK	CL	Whole chr 21	High chr 21 intensity, BAF split	47,XY+21	Down syndrome, confirmed by karyotype
1	5266_4	F	UNK	Blood	Whole chr 2	Run of homozygosity entire chr 2	N/A	*De novo* uniparental disomy chr 2, confirmed in blood DNA
2	21020_1	M	SPX	CL (array), blood (qPCR & FISH)	4p16	4p16.3p16.1 DUP	46,XY, der(8)t(4;8)(4p16.1→ter; 8p23.1→ter)	*De novo* unbalanced translocation t(4p16;8p23) leading to 4p16.3p16.1 duplication (4:53403-9016339) and 8p23.3p23.1 deletion; maternal origin by microsatellite markers. Normal karyotype, translocation validated by FISH and qPCR; the 4p terminal duplication (8.9 Mb) includes the Wolf-Hirschhorn syndrome region
2	21020_1	M	SPX	(see above)	8p23	8p23.3p23.1 DEL	(see above)	*De novo* unbalanced translocation t(4p16;8p23) (see above). The 6.8 Mb 8p23.3p23.1 terminal deletion (8:154984-6994825) overlaps numerous deletions described in individuals with ID
2	2175_1	M	SPX	Blood	22q13.2-qter	22q13.2-qter BAF split (no visible LRR deflection)	N/A	*De novo* 10 Mb 22q13.2-qter uniparental disomy (22:39729380-49582267), mosaicism in approximately 24% of cells; normal MLPA
2	4316_1	F	UNK	Blood	Whole chr X	Whole chr BAF split (no visible LRR deflection)	N/A	Triple X syndrome, *de novo*, mosaic
1	5257_3	M	SPX	Blood	Whole chr Y	Male, high chr Y intensity	47,XYY	XYY syndrome, *de novo*, confirmed by karyotype
1	5515_3	M	SPX	Blood	Whole chr Y	Male, high chr Y intensity	47,XYY	XYY syndrome, *de novo*, confirmed by karyotype
Other abnormalities resulting from cell line artifacts or not validated in blood								
1	5010_3	M	MPX	CL	1q43-qter DEL, 9q13-qter DUP	1q43-q44 low intensity & BAF split; 9q arm high intensity & little BAF split	47,XY	1q43-qter 9.4 Mb DEL, mosaic (1:237816283-247249719), 9q13-qter 69.8 Mb DUP, mosaic (9:70400178-140273252); found in the same DNA batch by Affy 500K-EA; normal karyotype; parents normal; AffyCytoScanHD on blood DNA showed normal results: cell line artifact
1	5321_3	M	SPX	CL	Whole chr 4	Run of homozygosity entire chr 4	N/A	Uniparental disomy chr 4, confirmed in cell line; no blood DNA available so we can't exclude a cell line artifact
1	6379_4	M	SPX	CL	Whole chr 14	High intensity, BAF split	46,XY	Whole chr 14 duplication by one algorithm only; karyotype excluded a chr14 trisomy: cell line artifact

Samples that passed QC filters but showed CNVs by any algorithm larger than 7.5 Mb, long runs of homozygosity, or CNVs encompassing/or close to the centromere on any of the autosomes or chromosome X were further inspected manually by plotting their log2 ratio intensities as well as allelic genotype ratios. A cutoff of >7.5 Mb was selected to be consistent with large cytogenetically visible chromosome abnormalities. Samples containing such alterations were excluded from the main burden analyses, except for two AGP probands with XYY syndrome that were retained since chromosome Y markers were not used in the CNV analysis, and two probands with uniparental disomy (proband 2175_1 with a segmental uniparental disomy of chr 22q was excluded during QC because of excessive number of calls).

Abbreviations: BAF, B allele frequency; CL, cell line; DEL, deletion; DUP, duplication; EA: early access (Affy500K-EA vs. Affy500K); F, female; LRR; log R ratio; M, male; MPX, multiplex; N/A, not available; SPX, simplex; UNK, unknown family type (extended family not evaluated for ASD).

Table S2A. Sample characteristics

Classification		Sex	Stage 1 (n=1087)		Stage 2 (new cases, n=1359)		Combined (n=2446)	
			European n=979[a]	Other n=108	European n=1168	Other n=191	European n=2147	Other n=299
ASD[b]		M	826	90	1027	171	1853	261
		F	153	18	141	20	294	38
Developmental impairment[c]	Without DI	M	433	51	569	65	1002	116
		F	67	11	63	9	130	20
	With DI	M	361	35	413	103	774	138
		F	84	7	72	11	156	18
Family type[d]	Simplex	M	366	33	660	132	1026	165
		F	62	3	78	14	140	17
	Multiplex	M	330	53	208	30	538	83
		F	59	14	43	5	102	19
	Unknown	M	131	4	159	9	290	13
		F	31	1	20	1	51	2

[a] 17 of the 996 (1.7%) European cases used in Pinto et al.[117] were excluded from the combined sample after applying additional QC steps.

[b] Subjects met criteria for strict autism or autism spectrum according to the ADI-R and/or ADOS (see Methods for detailed description).

[c] Developmental impairment is a hierarchical classification based on scores on full-scale IQ, performance IQ, verbal IQ and the Vineland Adaptive Behavior Scales composite score (see Methods for detailed description). A cut-off score of 70 was applied on all measures. Some data are missing on this item, ranging from 2%-4%.

[d] Family-history reports were taken to inform on the family type. Multiplex families had at least two individuals receiving validated ASD diagnoses who were first to third degree relatives (for third degree, only cousins were considered). Simplex families had only one known individual with ASD in first to third (cousin) degree relatives. Families that did not fall into the multiplex or simplex criteria above were classified as unknown.

The ratio of simplex to multiplex families was 1.82 (1348:742), with 356 (14.5%) families of unknown status. The ratio of males to females was 6.4:1; 46% of cases (n=1,086 out of 2,354) showed developmental impairment (DI). There were 1.43 times more females than males with DI compared to no-DI (M:F for DI is 5.2 vs. 7.45 for no-DI, 95% CI 1.12-1.82, chi square p=0.003). When considering family type, there were 1.47 times more males than females in simplex compared to multiplex families (95% CI 1.13-1.91, chi square p=0.0039), and simplex cases had 1.57 fold more DI compared to multiplex cases (95% CI 1.30-1.90, chi-square p=0.004).

Simplex males: 1026+165=1191
Simplex females: 140+17=157
Simplex M:F ratio: 1191/157= 7.59:1

Multiplex males: 538+83=621
Multiplex females: 102+19=121
Multiplex M:F ratio: 621/121= 5.13:1

Table S2B. Sample characteristics (*continued*)

Cases

	European	European Males	European Females	Other	Total
Stage 1	979	826	153	108	1,087
Stage 2	1,168	1,027	141	191	1,359
Stage 1 + 2 (all cases)	2,147	1,853	294	299	2,446

Complete Trios

	European	European Males	European Females	Other	Total (1)	Total (2)
Stage 1	862	733	129	91	953	n/a
Stage 2	958	845	113	166	1,124	n/a
Stage 1 + 2 (all trios)	1,820	1,578	242	257	2,077	2,096

(1) trios with array data after QC available from both parents

(2) trios in (1) plus additional trios with laboratory validation data from both parents;
n/a: not applicable

Cases, breakdowns

	Stage 1+2 Males		Stage 1+2 Females		Stage 1+2 All	
	All ancestries	European	All ancestries	European	All ancestries	European
Single probands	92	88	14	14	106	102
Proband+mother duo	101	83	25	23	126	106
Proband+father duo	120	104	17	15	137	119
Complete trios	1,801	1,578	276	242	2,077	1,820
Total probands	2,114	1,853	332	294	2,446	2,147

Cases with rare CNVs >30 kb

	European (all)	European (genic)*	European (exonic)
Stage 1	868	747	691
Stage 2	1,046	871	795
Stage 1 + 2 (all cases)	1,914	1,618	1,486
Stage 1 + 2 (all controls)	2,359	1,971	1,820
Total sample (cases + controls)	4,273	3,589	3,306

* with 10 kb-flanking

Controls

PRIMARY controls	European	European Males	European Females
Stage 1: 1166 SAGE	1,166	370	796
Stage 2: 1240 HABC + 234 OC	1,474	871	603
Totals	2,640	1,241	1,399

Additional controls

SECONDARY controls	Other ancestries	Other ancestries Males	Other ancestries Females
Stage 1: SAGE	517	185	332
Stage 2: HABC	1,326	596	730
Totals	1,843	781	1,062
Controls All ancestries - totals	4,483	2,022	2,461

Tables S3A-S3C. CNV burden

Table S3A. Platform comparison

CNV rate

Type	Group	1M single: 1382 cases / 1400 controls # rare CNVs	Baseline CNV rate (Ctrl)	Case/ Ctrl ratio	P value	1M duo: 765 cases / 1240 controls # rare CNVs	Baseline CNV rate (Ctrl)	Case/ Ctrl ratio	P value
All	All	6,506	2.32	1.02	0.27621	4,407	2.16	1.05	0.06357
DEL	All	3,353	1.17	1.07	0.03401	2,248	1.10	1.05	0.13466
DUP	All	3,153	1.15	0.96	0.83521	2,159	1.06	1.05	0.14956
All	30 – 500 kb	6,088	2.18	1.01	0.39300	4,201	2.08	1.02	0.25742
	≥ 500 kb	418	0.14	1.15	0.09388	206	0.08	1.79	0.00005
	≥ 1 Mb	127	0.04	1.24	0.13127	76	0.03	2.00	0.00214
DEL	30 – 500 kb	3,235	1.12	1.07	0.03287	2,187	1.08	1.02	0.34081
	≥ 500 kb	118	0.04	1.01	0.50979	61	0.02	2.87	0.00005
	≥ 1 Mb	42	0.01	1.65	0.08491	24	0.00	4.86	0.00026
DUP	30 – 500 kb	2,853	1.06	0.94	0.93625	2,014	1.00	1.02	0.31434
	≥ 500 kb	300	0.10	1.20	0.06567	145	0.06	1.47	0.01390
	≥ 1 Mb	85	0.03	1.09	0.39456	52	0.02	1.39	0.15035

Total CNV size (kb)

Type	Group	1M single: 1382 cases / 1400 controls Baseline size (Ctrl)	Case/ Ctrl ratio	P value	1M duo: 765 cases / 1240 controls Baseline size (Ctrl)	Case/ Ctrl ratio	P value
All	All	442.7	1.10	0.03171	365.1	1.39	0.00001
DEL	All	222.2	1.12	0.06035	190.1	1.35	0.00011
DUP	All	362.8	1.08	0.10991	305.1	1.41	0.00005
All	30 – 500 kb	298.9	0.99	0.56721	285.8	1.06	0.07551
	≥ 500 kb	995.1	1.24	0.00201	998.7	1.44	0.00534
	≥ 1 Mb	1,621.0	1.27	0.00651	1,519.0	1.59	0.00514
DEL	30 – 500 kb	167.9	1.02	0.28813	168.7	0.97	0.70923
	≥ 500 kb	948.0	1.45	0.00841	964.5	1.40	0.08070
	≥ 1 Mb	1,551.0	1.51	0.02177	1,411.0	1.42	0.12728
DUP	30 – 500 kb	233.9	0.98	0.72926	223.3	1.11	0.01821
	≥ 500 kb	985.7	1.12	0.08485	980.7	1.45	0.01585
	≥ 1 Mb	1,649.0	1.11	0.15156	1,487.0	1.77	0.00600

Gene count

Type	Group	1M single: 1382 cases / 1400 controls # Rare genic CNVs	# Genes inters. by rare CNVs	Baseline gene rate (Ctrl)	Case/ Ctrl gene ratio	P value	Pcorr	1M duo: 765 cases / 1240 controls # Rare genic CNVs	# Genes inters. by rare CNVs	Baseline gene rate (Ctrl)	Case/ Ctrl gene ratio	P value	Pcorr
All	All	4,180	5,210	4.11	1.22	0.00140	0.00120	2,681	3,587	3.12	1.64	0.00001	0.00001
DEL	All	1,806	2,040	1.37	1.33	0.00122	0.01068	1,164	1,469	1.16	1.40	0.00146	0.05123
DUP	All	2,374	4,002	2.74	1.17	0.03831	0.00824	1,517	2,597	1.95	1.78	0.00001	0.00004
All	30 – 500 kb	3,800	3,795	3.12	1.03	0.30338	0.33000	2,496	2,780	2.71	1.12	0.02734	0.02645
	≥ 500 kb	380	2,011	0.99	1.84	0.00034	0.00234	185	1,039	0.41	5.07	0.00001	0.00019
	≥ 1 Mb	122	1,075	0.37	2.65	0.00096	0.00231	71	641	0.22	7.10	0.00001	0.00329
DEL	30 – 500 kb	1,700	1,407	1.09	1.15	0.02952	0.10020	1,113	1,028	1.08	0.92	0.84269	0.86370
	≥ 500 kb	106	732	0.28	2.05	0.00992	0.01546	51	485	0.08	7.80	0.00001	0.00589
	≥ 1 Mb	41	504	0.13	2.80	0.01122	0.11560	23	339	0.04	10.00	0.00005	0.05489
DUP	30 – 500 kb	2,100	2,849	2.03	0.96	0.70648	0.43300	1,383	2,045	1.63	1.26	0.00239	0.00397
	≥ 500 kb	274	1,561	0.72	1.76	0.00545	0.01727	134	697	0.33	4.38	0.00008	0.01109
	≥ 1 Mb	81	721	0.24	2.57	0.01518	0.00677	48	391	0.17	6.38	0.00086	0.08588

Baseline CNV rate (Ctrl): average number of CNVs per control subject
Baseline size (Ctrl): average total size per control subject in kb
Baseline gene rate (Ctrl): average number of genes intersected by CNVs per control subject

Pcorr: corrected for global differences in CNV size and rate
100,000 permutations

Global burden analyses for rare CNVs were performed using PLINK v1.0730, R stats and custom scripts, as previously described in detail.[117] We tested for global increased burden in a combined set of 2,147 European ASD cases compared to 2,640 European controls for three measures: CNV rate, CNV size (**Tables S3A-S3B**) and the average number of genes affected by CNVs (gene-count) (**Table 1**, main text). We observed a significantly increased burden in the number of genes affected by rare CNVs in cases over controls (**Table 1**). This enrichment for rare genic CNVs was apparent for both deletions and duplications, and remained after we further controlled for potential case-control differences that could be present due to biological differences or technical biases (**Table 1**). Similar findings were obtained when data were broken down by array type (**Table S3A**), or when each stage was considered separately (**Table S3B**).

Table S3B. Stage 1 (Pinto et al.) versus Stage 2 (new cases) versus Combined (all 2147 European cases)

CNV rate

		Stage 1: 979 cases / 1166 controls				Stage 2: 1168 cases / 1474 controls				Combined stage 1 + stage 2: 2147 cases / 2640 controls			
Type	Group	# rare CNVs	Baseline CNV rate (Ctrl)	Case/ Ctrl ratio	P value	# rare CNVs	Baseline CNV rate (Ctrl)	Case/ Ctrl ratio	P value	# rare CNVs	Baseline CNV rate (Ctrl)	Case/ Ctrl ratio	P value
All	All	5,153	2.44	0.97	0.85304	5,923	2.16	1.09	0.00066	11,076	2.28	1.03	0.05804
DEL	All	2,664	1.23	1.02	0.29184	3,063	1.11	1.09	0.00554	5,727	1.16	1.06	0.01238
DUP	All	2,489	1.21	0.91	0.98261	2,860	1.05	1.08	0.02640	5,349	1.12	1.00	0.49532
All	30 – 500 kb	4,830	2.29	0.96	0.90681	5,634	2.08	1.06	0.01827	10,464	2.17	1.01	0.27705
	≥ 500 kb	323	0.14	1.11	0.19004	289	0.08	1.88	0.00002	612	0.11	1.43	0.00003
	≥ 1 Mb	91	0.04	1.02	0.50026	106	0.02	2.45	0.00003	197	0.03	1.62	0.00066
DEL	30 – 500 kb	2,575	1.19	1.03	0.27934	2,978	1.09	1.07	0.03504	5,553	1.14	1.05	0.04012
	≥ 500 kb	89	0.04	0.97	0.59513	85	0.02	2.71	0.00002	174	0.03	1.59	0.00173
	≥ 1 Mb	27	0.01	1.28	0.32376	38	0.01	4.73	0.00005	65	0.01	2.58	0.00024
DUP	30 – 500 kb	2,255	1.11	0.89	0.99523	2,656	0.99	1.05	0.13460	4,911	1.04	0.97	0.82515
	≥ 500 kb	234	0.10	1.17	0.13809	204	0.06	1.63	0.00037	438	0.08	1.37	0.00084
	≥ 1 Mb	64	0.03	0.93	0.66292	68	0.02	1.80	0.01192	132	0.02	1.31	0.07407

Total CNV size (kb)

		Stage 1: 979 cases / 1166 controls			Stage 2: 1168 cases / 1474 controls			Combined stage 1 + stage 2: 2147 cases / 2640 controls		
Type	Group	Baseline size (Ctrl)	Case/ Ctrl ratio	P value	Baseline size (Ctrl)	Case/ Ctrl ratio	P value	Baseline size (Ctrl)	Case/ Ctrl ratio	P value
All	All	453.1	1.03	0.26520	368.9	1.41	0.00001	406	1.22	0.00001
DEL	All	222.5	1.02	0.40872	196.4	1.38	0.00006	207.9	1.21	0.00028
DUP	All	365.2	1.06	0.20712	306.1	1.36	0.00001	333.3	1.21	0.00007
All	30 – 500 kb	304.4	0.97	0.78403	290.9	1.06	0.04013	296.8	1.02	0.22737
	≥ 500 kb	1009.0	1.14	0.04687	963.5	1.49	0.00027	990.1	1.32	0.00005
	≥ 1 Mb	1632.0	1.18	0.04903	1,533.0	1.54	0.00144	1,591.0	1.38	0.00013
DEL	30 – 500 kb	168.3	0.99	0.56454	173.4	0.99	0.63485	171.2	0.99	0.64394
	≥ 500 kb	952.3	1.12	0.19717	967.3	1.64	0.01571	957.4	1.42	0.00371
	≥ 1 Mb	1560.0	1.17	0.16605	1,483.0	1.61	0.03369	1,531.0	1.44	0.01435
DUP	30 – 500 kb	233.9	0.97	0.75751	226.2	1.08	0.03083	229.7	1.03	0.18388
	≥ 500 kb	997.0	1.11	0.13966	940.2	1.37	0.00872	972.4	1.24	0.00429
	≥ 1 Mb	1658.0	1.15	0.14467	1,493.0	1.52	0.00919	1,586.0	1.34	0.00380

Gene count

		Stage 1: 979 cases / 1166 controls					Stage 2: 1168 cases / 1474 controls					Combined stage 1 + stage 2: 2147 cases / 2640 controls				
Type	Group	# Rare genic CNVs	# Genes inters. by rare CNVs	Baseline gene rate (Ctrl)	Case/ Ctrl gene ratio	P value	Pcorr	# Rare genic CNVs	# Genes inters. by rare CNVs	Baseline gene rate (Ctrl)	Case/ Ctrl gene ratio	P value	Pcorr	# Rare genic CNVs	# Genes inters. by rare CNVs	Baseline gene rate (Ctrl)	Case/ Ctrl gene ratio	P value	Pcorr
All	All	3,252	4,319	4.16	1.17	0.02076	0.00462	3,607	4,514	3.08	1.66	0.00001	0.00001	6,859	6,745	3.55	1.41	0.00001	0.00001
DEL	All	1,378	1,505	1.32	1.25	0.01603	0.01603	1,568	1,924	1.16	1.55	0.00001	0.01498	2,946	2,804	1.23	1.40	0.00001	0.00049
DUP	All	1,874	3,417	2.84	1.14	0.10573	0.01342	2,039	3,244	1.92	1.73	0.00001	0.00001	3,913	5,217	2.32	1.41	0.00001	0.00001
All	30 – 500 kb	2,962	3,211	3.14	0.98	0.60840	0.29800	3,345	3,414	2.70	1.15	0.00384	0.02582	6,307	5,163	2.89	1.07	0.03752	0.03628
	≥ 500 kb	290	1,549	1.02	1.75	0.00488	0.00700	262	1,485	0.38	5.29	0.00001	0.00001	552	2,491	0.66	2.88	0.00001	0.00001
	≥ 1 Mb	86	805	0.38	2.39	0.01245	0.00648	101	865	0.17	8.12	0.00001	0.00001	187	1,337	0.26	4.48	0.00001	0.00001
DEL	30 – 500 kb	1,300	1,137	1.07	1.16	0.03994	0.03134	1,495	1,296	1.06	0.99	0.55163	0.79270	2,795	2,014	1.07	1.07	0.11389	0.20110
	≥ 500 kb	78	454	0.25	1.63	0.09459	0.08104	73	696	0.09	7.91	0.00001	0.00051	151	947	0.16	3.60	0.00001	0.00051
	≥ 1 Mb	26	278	0.11	2.22	0.08343	0.13570	37	492	0.07	7.84	0.00002	0.10500	63	647	0.08	4.58	0.00004	0.02289
DUP	30 – 500 kb	1,662	2,411	2.07	0.9	0.92581	0.62180	1,850	2,491	1.63	1.26	0.00094	0.00364	3,512	3,934	1.83	1.08	0.08690	0.03750
	≥ 500 kb	212	1,325	0.77	1.79	0.01117	0.01837	189	979	0.28	4.43	0.00001	0.00026	401	1,896	0.50	2.64	0.00001	0.00026
	≥ 1 Mb	60	632	0.27	2.46	0.03874	0.01247	64	471	0.11	8.30	0.00006	0.00176	124	890	0.18	4.43	0.00004	0.00036

Baseline CNV rate (Ctrl): average number of CNVs per control subject
Baseline size (Ctrl): average total size per control subject in kb
Baseline gene rate (Ctrl): average number of genes intersected by CNVs per control subject

Pcorr: corrected for global differences in CNV size and rate
100,000 permutations

Table S3C. Characteristics of rare CNVs in 2,147 European ASD probands and 2,640 European controls

	ASD probands, European (n = 2,147 probands; 1,838 trios)				Controls, European (n = 2,640)		
	Stringent CNVs[a] (all sizes)	**Stringent CNVs**[b] (≥30 kb size, no peri-centromeric + no segdup)	**Rare CNVs**[c]	**Rare *de novo* CNVs**[d]	**Stringent CNVs**[a] (all sizes)	**Stringent CNV**[b] (≥30 kb size, no peri-centromeric + no segdup)	**Rare CNVs**[c]
Samples	2,147	2,147	1,941	86	2,640	2,640	2,359
# CNVs	36,034	15,423	5,054	89	46,101	20,341	6,022
Mean / median CNVs per genome[e]	16.8 / 16	7.18 / 7	2.64 / 2	1.03 / 1	17.5 / 17	7.70 / 8	2.55 / 2
Mean / median CNV size (kb)	107.0 / 44.8	162.5 / 94.5	187.9 / 88.4	1,244.5 / 477.4	91.0 / 39.3	150.6 / 89.9	159.1 / 87.0
% Gain / loss	26.4 / 73.6	39.6 / 60.4	47.9 / 52.1	33.7 / 66.3	24.8 / 75.2	36.9 / 63.1	49.3 / 50.7
# Recurrent/overlapping CNVs (%) / # loci[f]	33,803 (93.8%) / 1,536	13,812 (89.5%) / 912	3,431 (67.9%) / 822	34 (38.2%) / 11	43,905 (95.2%) / 1,613	18,641 (91.6%) / 1,019	4,321 (71.8%) / 936
# CNVs >1 Mb (%)	365 (1.0%)	277 (1.8%)	112 (2.2%)	28 (31.5%)	264 (0.6%)	226 (1.1%)	85 (1.4%)
# CNVs >100-999 kb (%)	9,782 (27.1%)	7,019 (45.5%)	2,445 (48.4%)	42 (47.2%)	11,945 (25.9%)	9,341 (45.9%)	2,537 (42.1%)

[a] Stringent CNVs are CNV called by two or more algorithms. CNVs detected in the same individual by at least two algorithms were merged with the outside probes used as boundaries. All sizes, no filter applied.

[b] Stringent CNVs ≥30 kb and filtered for pericentromeric calls as well as calls overlapped by segmental duplications (segdup) for >50% of their length.

[c] Rare stringent CNVs ≥30 kb (filtered for pericentromeric and segdup calls) present at a frequency <1% in the total sample of 2,147 European cases and 2,640 European controls**.**

[d] Inheritance state was estimated for CNVs detected in 1,838 European probands from complete trios, including 1,820 complete trios with array data after QC available from both parents plus 18 families in which at least one of the parents failed initial array QC but additional laboratory validation was obtained for both parents. Laboratory validation confirmed that at least 4.7% (86/1,838) of European families carried at least one *de novo* CNV (average of 1 verified *de novo* CNV/sample). Similar rates were obtained for families of all ancestries, with 4.7% (98/2,096) carrying at least one *de novo* event.

[e] Probands with CNVs larger than 7.5 Mb are listed in **Table S1C**; they were excluded from the main burden analyses.

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

Table S4. Rare *de novo* CNVs in probands confirmed experimentally

Sample ID	Sex	Ancestry	CNV type	Chr#:start-end (hg18)	Size	Cytoband	RefSeq gene(s)	Location
14310_4270	M	Eur	Gain	1:105623065-105795181	172,117	1p21.1	—	intergenic
14310_4270	M	Eur	Gain	1:105896243-106480692	584,450	1p21.1	—	intergenic
16035_1571013001	M	Eur	Gain	1:144482933-146325557	1,842,625	1q21.1	15 genes (includes *CHD1L, GJA5, GJA8*)	gene/exonic
1952_301	M	Other	Gain	1:144500467-146336720	1,836,254	1q21.1	15 genes (includes *CHD1L, GJA5, GJA8*)	gene/exonic
8635_201	M	Eur	Loss	1:144500467-146377870	1,877,404	1q21.1	17 genes (includes *CHD1L, GJA5, GJA8*)	gene/exonic
13135_1523	F	Eur	Gain	1:144838594-146308287	1,469,694	1q21.1	14 genes (includes *CHD1L, GJA5, GJA8*)	gene/exonic
6356_5	M	Eur	Loss	1:2118508-2325536	207,029	1p36.33-p36.32	*MORN1, SKI, LOC100129534, RER1, C1orf86*	gene/exonic
8658_201	F	Eur	Loss	1:98175622-100923952	2,748,331	1p21.3-p21.2	18 genes (includes *CDC14A, MIR137, RTCD1, SNX7*)	gene/exonic
9877_204	M	Other	Gain	1:98247355-99645560	1,398,206	1p21.3-p21.2	*LOC100129620, LPPR4, LPPR5, MIR137, SNX7*	gene/exonic
5089_5	M	Other	Loss	2:102292943-102345460	52,518	2q12.1	*IL1RL1*	gene/exonic
13082_963	M	Eur	Gain	2:11712589-11741036	28,448	2p25.1	*NTSR2*	gene/exonic
4260_1	M	Eur	Gain	2:119603876-119882278	278,403	2q14.2	*C1QL2, C2orf76, DBI, STEAP3*	gene/exonic
13027_353	F	Eur	Gain	2:144837809-145315383	477,575	2q22.3	*DKFZp686O1327, ZEB2*	gene/exonic
14414_5230	M	Eur	Loss	2:230486629-230547253	60,625	2q36.3	*FBXO36, TRIP12*	gene/exonic
14068_1180	M	Eur	Gain	2:50493827-50677835	184,009	2p16.3	*NRXN1*	gene/exonic
13017_223	F	Eur	Loss	2:50539877-50730546	190,670	2p16.3	*NRXN1*	gene/exonic
13216_2383	M	Eur	Loss	2:50968208-51214171	245,964	2p16.3	*NRXN1*	gene/exonic
13153_1703	M	Eur	Loss	2:50990306-51222043	231,738	2p16.3	*NRXN1*	gene/exonic
13037_463	M	Eur	Loss	2:51002576-51157742	155,167	2p16.3	*NRXN1*	gene/exonic
3616_3	M	Eur	Loss	3:9474320-9498362	24,043	3p25.3	*SETD5*	gene/exonic
5220_3	F	Eur	Gain	3:19127998-19640299	512,302	3p24.3	*KCNH8*	gene/exonic
3174_3	M	Eur	Loss	3:19921061-20096832	175,772	3p24.3	*C3orf48, EFHB, KAT2B, RAB5A*	gene/exonic
5245_3	M	Eur	Loss	3:117285007-117477191	192,185	3q13.31	*LSAMP* (50% mosaic)	gene/exonic
8587_210	F	Eur	Gain	3:139760896-140127703	366,808	3q22.3	*CEP70, FAIM, PIK3CB*	gene/exonic
8588_201	M	Eur	Loss	3:195971510-197675831	1,704,322	3q29	20 genes (includes *PCYT1A, TFRC, TNK2*)	gene/exonic
8677_201	M	Other	Gain	4:94659049-95289667	630,619	4q22.2	*ATOH1, GRID2*	gene/exonic
6402_3	M	Eur	Gain	4:98781694-99627129	845,436	4q22.3-q23	*C4orf37, RAP1GDS1, TSPAN5*	gene/exonic
8709_201	F	Eur	Loss	5:88588170-89505509	917,340	5q14.3	*MIR3660*	gene/exonic
5386_3	M	Other	Loss	6:156785155-158489874	1,704,720	6q25.3	*ARID1B, MIR3692, SERAC1, SNX9, SYNJ2, ZDHHC14*	gene/exonic
6164_3	F	Eur	Gain	6:160023074-160081618	58,545	6q25.3	*SOD2, WTAP*	gene/exonic
8612_201	M	Eur	Loss	6:162588879-162629938	41,060	6q26	*PARK2*	gene/exonic
8404_201	M	Eur	Loss	6:169136788-170761395	1,624,608	6q27	14 genes (includes *PSMB1, TBP, TCTE3, THBS2, WDR27*)	gene/exonic
5353_3	F	Eur	Loss	6:33399849-33512042	112,194	6p21.32	*CUTA, KIFC1, PHF1, SYNGAP1*	gene/exonic
18133_302	M	Eur	Gain	6:69540429-73375020	3,834,592	6q12-q13	*B3GAT2, BAI3, C6orf155, C6orf57, COL19A1, COL9A1, FAM135A, LMBRD1, MIR30A, MIR30C2, OGFRL1, RIMS1, SMAP1*	gene/exonic
6248_3	M	Eur	Loss	6:86352577-86376159	23,583	6q14.3	*SNX14, SYNCRIP*	gene/exonic
1960_301	F	Eur	Loss	7:102699832-102798745	98,914	7q22.1	*DNAJC2, DPY19L2P2, PMPCB, PSMC2, SLC26A5*	gene/exonic
5370_3	M	Eur	Loss	7:153775586-153844747	69,162	7q36.2	*DPP6*	gene/exonic
8446_201	M	Eur	Loss	7:72344426-73782113	1,437,688	7q11.23	28 genes (includes *GTF2I, GTF2IRD1, LAT2, LIMK1, STX1A*)	gene/exonic
1142_4	F	Eur	Gain	8:48631388-48802529	171,142	8q11.21	*KIAA0146*	gene/exonic
6321_3	M	Eur	Loss	8:65354366-66254869	900,504	8q12.3-q13.1	*BHLHE22, CYP7B1, LOC100130155, LOC401463, MIR124-2*	gene/exonic
5290_3	M	Other	Gain	8:704383-1521910	817,528	8p23.3	*DLGAP2, LOC286083*	gene/exonic
8500_201	M	Eur	Loss	8:74692898-74967280	274,383	8q21.11	*STAU2, UBE2W*	gene/exonic
14237_2650	M	Eur	Loss	9:103456776-107140273	3,683,498	9q31.1	*ABCA1, CYLC2, GRIN3A, LOC286367, NIPSNAP3A, NIPSNAP3B, OR13C2, OR13C3, OR13C4, OR13C5, OR13C8, OR13C9, OR13D1, OR13F1, SLC44A1, SMC2*	gene/exonic
14417_5260	M	Eur	Loss	9:117311405-117727764	416,360	9q33.1	*C9orf27* (or *LINC00474*)	gene/exonic
9756_201	M	Other	Loss	9:132572435-132597937	25,503	9q34.12	*ABL1*	gene/exonic
6259_3	M	Eur	Loss	9:139516033-140208462	692,430	9q34.3	*ARRDC1, C9orf37, CACNA1B, EHMT1, FLJ40292* (or *EHMT1-IT1*) , *MIR602, MRPL41, PNPLA7, TUBBP5, WDR85, ZMYND19*	gene/exonic
6246_4	M	Eur	Loss	9:9399606-9631169	231,564	9p23	*PTPRD*	gene/exonic
5032_4	M	Eur	Loss	9:98998-334508	235,511	9p24.3	*C9orf66, CBWD1, DOCK8, FOXD4*	gene/exonic
13123_1403	F	Eur	Loss	9:98998-3682923	3,583,926	9p24.3-p24.2	*C9orf66, CBWD1, DMRT1, DMRT2, DMRT3, DOCK8, FLJ35024, FOXD4, KANK1, KCNV2, KIAA0020, RFX3, SMARCA2, VLDLR*	gene/exonic
18172_302	M	Other	Gain	10:34963905-35536952	573,048	10p11.21	*CREM, CUL2, PARD3*	gene/exonic
8534_201	M	Eur	Loss	10:45633089-51564756	5,931,668	10q11.21-q11.23	56 genes (includes *CHAT, ERCC6, MAPK8, SLC18A3*)	gene/exonic

Sample ID	Sex	Ancestry	CNV type	Chr#:start-end (hg18)	Size	Cytoband	RefSeq gene(s)	Location
14393_5020	M	Eur	Loss	10:74801551-75278951	477,401	10q22.2	*AGAP5, ANXA7, BMS1P4, CAMK2G, CHCHD1, FUT11, KIAA0913, MYOZ1, NDST2, PPP3CB, SEC24C, SYNPO2L, USP54, ZMYND17*	gene/exonic
6240_4	M	Eur	Loss	11:126633939-132060374	5,426,436	11q24.2-q25	20 genes (includes *ARHGAP32, SNX19*)	gene/exonic
6325_3	M	Eur	Loss	11:70077507-70506315	428,809	11q13.3	*MIR3664, SHANK2*	gene/exonic
6319_3	M	Eur	Loss	11:70119917-70187872	67,956	11q13.3	*SHANK2*	gene/exonic
5237_3	M	Eur	Loss	11:70154458-70220632	66,175	11q13.3	*SHANK2*	gene/exonic
6053_3	M	Eur	Gain	12:54218922-58779615	4,560,694	12q13.2-q14.1	101 genes (includes *CDK2, CDK4, SMARCC2*)	gene/exonic
5272_3	M	Other	Loss	12:98445422-98540678	95,257	12q23.1	*ANKS1B*	gene/intronic
1050_3	F	Eur	Gain	14:20279711-20345174	65,464	14q11.2	*EDDM3A, EDDM3B, RNASE1, RNASE6*	gene/exonic
8638_201	M	Eur	Loss	14:35203374-35369853	166,480	14q13.2	*BRMS1L, RALGAPA1*	gene/exonic
4272_1	M	Eur	Loss	14:63824114-65347410	1,523,297	14q23.2-q23.3	18 genes (includes *ESR2, MAX, MTHFD1, SPTB*)	gene/exonic
17009_1	M	Eur	Loss	14:78575296-78596793	21,498	14q31.1	*NRXN3*	gene/intronic
20187_1464001	M	Eur	Gain	15:18811937-26209270	7,397,334	15q11.2-q13.1	125 genes (includes *CYFIP1, UBE3A, HERC2*)	gene/exonic
8630_201	M	Eur	Gain	15:19800798-26209270	6,408,473	15q11.2-q13.1	117 genes (includes *CYFIP1, UBE3A, HERC2*)	gene/exonic
20069_1328001	M	Eur	Gain	15:20203578-26209270	6,005,693	15q11.2-q13.1	112 genes (includes *CYFIP1, UBE3A, HERC2*)	gene/exonic
17035_1	F	Eur	Gain	15:20274130-26120360	5,846,231	15q11.2-q13.1	111 genes (includes *CYFIP1, UBE3A, HERC2*)	gene/exonic
8430_204	M	Other	Loss	15:20301669-20777695	476,027	15q11.2	*CYFIP1, NIPA1, NIPA2, TUBGCP5, WHAMML1*	gene/exonic
13050_593	M	Eur	Gain	15:21190624-26203954	5,013,331	15q11.2-q13.1	101 genes (includes *UBE3A, HERC2*)	gene/exonic
16040_1571029001	M	Eur	Loss	15:28450423-30303265	1,852,843	15q13.2-q13.3	*ARHGAP11B, CHRFAM7A, CHRNA7, FAM7A1, FAM7A2, FAN1, KLF13, LOC100288637, MIR211, MTMR10, OTUD7A, TRPM1*	gene/exonic
6101_4	M	Other	Loss	15:74735339-74929817	194,479	15q24.3	*SCAPER*	gene/exonic
8695_201	M	Eur	Loss	15:81728085-82623936	895,852	15q25.2	*ADAMTSL3, BNC1, LOC648809, SH3GL3*	gene/exonic
14181_2940	M	Eur	Gain	15:82906265-82985247	78,983	15q25.2	*LOC100506874, SCAND2, UBE2Q2P1, ZSCAN2*	gene/exonic
14070_1230	M	Eur	Loss	15:91200007-91283004	82,998	15q26.1	*CHD2, LOC100507217, MIR3175*	gene/exonic
2204_1	M	Eur	Loss	16:29466569-30147029	680,461	16p11.2	42 genes (includes *KCTD13, MAPK3, SEZ6L2*)	gene/exonic
20089_1391001	M	Eur	Loss	16:29502984-30107306	604,323	16p11.2	29 genes (includes *KCTD13, MAPK3, SEZ6L2*)	gene/exonic
5068_3	F	Eur	Loss	16:29502984-30127026	624,043	16p11.2	40 genes (includes *KCTD13, MAPK3, SEZ6L2*)	gene/exonic
5262_4	M	Eur	Gain	16:29502984-30210849	707,866	16p11.2	43 genes (includes *KCTD13, MAPK3, SEZ6L2*)	gene/exonic
4030_1	M	Eur	Gain	16:29554843-30107306	552,464	16p11.2	28 genes (includes *KCTD13, MAPK3, SEZ6L2*)	gene/exonic
5359_4	M	Eur	Loss	16:29554843-30195224	640,382	16p11.2	42 genes (includes *KCTD13, MAPK3, SEZ6L2*)	gene/exonic
3439_3	M	Eur	Loss	17:17156307-18262979	1,106,673	17p11.2	22 genes (includes *RAI1, SREBF1, LLGL1, TOP3A*)	gene/exonic
2211_1	F	Eur	Loss	17:17169258-20101517	2,932,260	17p11.2	53 genes (includes *MAPK7, RAI1, SREBF1, FAM83G, LLGL1, TOP3A*)	gene/exonic
5056_4	M	Eur	Gain	17:34612208-34732327	120,120	17q12	*FBXL20, RPL19, STAC2*	gene/exonic
8463_202	M	Eur	Loss	17:52774693-52895975	121,283	17q22	*MSI2*	gene/exonic
5444_3	M	Eur	Gain	17:76953064-77782267	829,204	17q25.3	38 genes (includes *ACTG1, ARHGDIA, FASN, RAC3, HGS, DUS1L*)	gene/exonic
5444_3	M	Eur	Loss	17:77785939-77849717	63,779	17q25.3	*CSNK1D, SLC16A3*	gene/exonic
3477_3	M	Eur	Loss	18:30280260-30327512	47,253	18q12.1	*DTNA*	gene/exonic
14331_4450	M	Eur	Loss	18:72085223-73670156	1,584,934	18q23	*GALR1, LOC284276, MBP, ZNF236, ZNF516*	gene/exonic
6358_6	M	Eur	Loss	19:4548413-5287389	738,977	19p13.3	*ARRDC5, C19orf10, DPP9, FEM1A, KDM4B, MIR7-3, NCRNA00306, PLIN3, PTPRS, TICAM1, TNFAIP8L1, UHRF1*	gene/exonic
5335_3	M	Eur	Loss	20:14545734-14948785	403,052	20p12.1	*MACROD2, MACROD2-AS1*	gene/exonic
5046_3	M	Other	Loss	20:8607242-8637441	30,200	20p12.3	*PLCB1*	gene/exonic
20180_1704001	M	Eur	Gain	21:43018846-43444308	425,463	21q22.3	*CBS, NDUFV3, PDE9A, PKNOX1, U2AF1, WDR4*	gene/exonic
3183_7	M	Eur	Loss	22:17241748-19819918	2,578,171	22q11.21	63 genes (includes *TBX1, CRKL, TSSK2*)	gene/exonic
8627_201	M	Eur	Gain	22:17257787-19793730	2,535,944	22q11.21	63 genes (includes *TBX1, CRKL, TSSK2*)	gene/exonic
17015_1	M	Eur	Loss	22:17257787-19795780	2,537,994	22q11.21	63 genes (includes *TBX1, CRKL, TSSK2*)	gene/exonic
4271_1	M	Eur	Gain	22:17257787-19795780	2,537,994	22q11.21	63 genes (includes *TBX1, CRKL, TSSK2*)	gene/exonic
8630_201	M	Eur	Loss	22:32346124-32413987	67,864	22q12.3	*LARGE*	gene/exonic
2072_1	M	Eur	Loss	22:45159185-49582267	4,423,083	22q13.31-q13.33	47 genes (includes *SHANK3*)	gene/exonic
6130_4	F	Eur	Loss	22:47996161-49512530	1,516,370	22q13.32-q13.33	37 genes (includes *SHANK3*)	gene/exonic
14291_4120	F	Eur	Loss	22:49468716-49485255	16,540	22q13.33	*SHANK3*	gene/exonic
16079_1571066001	M	Eur	Loss	22:49470371-49567383	97,013	22q13.33	*ACR, RABL2B, RPL23AP82, SHANK3*	gene/exonic
20013_1075001	M	Eur	Gain	X:153239048-153521797	282,750	Xq28	20 genes (includes *EMD, FLNA, GDI1, IKBKG, PLXNA3, RPL10*)	gene/exonic
16076_1571045001	M	Eur	Loss	X:22768481-23133948	365,468	Xp22.11	*DDX53* (includes *PTCHD1-AS*)	gene/exonic

We identified 102 rare *de novo* CNVs in 99 cases; three individuals (5444_3, 8630_201 and 14310_4270) have 2 *de novo* CNV each. Of the 99 subjects with *de novo* events, 60 (60%) are simplex, 26 (27%) multiplex, and 13 (13%) of unknown status. Eleven *de novo* chromosomal abnormalities >7.5 Mb identified in probands are listed in **Table S1C**. Abbreviations: Eur, European; F, female; M, male.

Table S5A. Parent of origin for rare *de novo* validated CNVs in probands

Sample ID	Chr#:start-end (hg18)	CNV type	Family type	Parent of origin	SNPs: |#paternal|#maternal|total
14310_4270	1:105623065-105795181	Gain	SPX	Paternal	SNPs: |20|0|41
14310_4270	1:105896243-106480692	Gain	SPX	Paternal	Custom-designed microsatellite (14310b)
16035_1571013001	1:144482933-146325557	Gain	MPX	Unknown	No informative SNPs or microsatellites
1952_301	1:144500467-146336720	Gain	MPX	Paternal	SNPs: |24|0|534
8635_201	1:144500467-146377870	Loss	MPX	Paternal	SNPs: |73|0|617
13135_1523	1:144838594-146308287	Gain	UNK	Paternal	SNPs: |121|56|667
6356_5	1:2118508-2325536	Loss	SPX	Maternal	SNPs: |0|11|92
8658_201	1:98175622-100923952	Loss	SPX	Maternal	SNPs: |0|199|1018
9877_204	1:98247355-99645560	Gain	SPX	Unknown	Ambiguous, SNPs: |74|27|488; microsat ?D1S2739
5089_5	2:102292943-102345460	Loss	MPX	Unknown	No informative SNPs or microsatellites
13082_963	2:11712589-11741036	Gain	UNK	Maternal	Custom-designed microsatellite (13082b)
4260_1	2:119603876-119882278	Gain	SPX	Maternal	SNPs: |1|26|118 (Griswold et al. 2012)[128]
13027_353	2:144837809-145315383	Gain	UNK	Maternal	SNPs: |0|40|124
14414_5230	2:230486629-230547253	Loss	SPX	Unknown	No informative SNPs
14068_1180	2:50493827-50677835	Gain	SPX	Maternal	SNPs: |0|15|75
13017_223	2:50539877-50730546	Loss	UNK	Paternal	SNPs: |19|0|89
13216_2383	2:50968208-51214171	Loss	UNK	Maternal	SNPs: |0|10|88
13153_1703	2:50990306-51222043	Loss	UNK	Paternal	Custom-designed microsatellite (13153a)
13037_463	2:51002576-51157742	Loss	SPX	Paternal	SNPs: |10|0|47
5245_3	3:117285007-117477191	Loss	MPX	Unknown	No informative SNPs
8587_210	3:139760896-140127703	Gain	UNK	Paternal	SNPs: |5|0|105
5220_3	3:19127998-19640299	Gain	SPX	Paternal	SNPs: |38|0|205
8588_201	3:195971510-197675831	Loss	UNK	Maternal	SNPs: |0|35|633
3174_3	3:19921061-20096832	Loss	MPX	Maternal	SNPs: |3|0|72
3616_3	3:9474320-9498362	Loss	UNK	Unknown	No informative SNPs
8677_201	4:94659049-95289667	Gain	SPX	Paternal	SNPs: |28|1|195
6402_3	4:98781694-99627129	Gain	SPX	Maternal	SNPs: |0|28|230
8709_201	5:88588170-89505509	Loss	SPX	Paternal	SNPs: |48|0|249
5386_3	6:156785155-158489874	Loss	MPX	Paternal	SNPs: |115|0|605
6164_3	6:160023074-160081618	Gain	SPX	Unknown	Ambiguous, SNPs: |2|5|23 *
8612_201	6:162588879-162629938	Loss	UNK	Paternal	Custom-designed microsatellite (8612a)
8404_201	6:169136788-170761395	Loss	MPX	Maternal	SNPs: |0|104|648
5353_3	6:33399849-33512042	Loss	SPX	Paternal	SNPs: |1|0|63 **
18133_302	6:69540429-73375020	Gain	SPX	Maternal	SNPs: |0|182|1516
6248_3	6:86352577-86376159	Loss	SPX	Unknown	No informative SNPs
1960_301	7:102699832-102798745	Loss	MPX	Paternal	D7S2509
5370_3	7:153775586-153844747	Loss	SPX	Maternal	SNPs: |0|6|43
8446_201	7:72344426-73782113	Loss	SPX	Maternal	SNPs: |0|66|615
1142_4	8:48631388-48802529	Gain	MPX	Paternal	SNPs: |3|0|27
6321_3	8:65354366-66254869	Loss	SPX	Unknown	No informative SNPs or microsatellites
5290_3	8:704383-1521910	Gain	SPX	Paternal	SNPs: |85|37|361
8500_201	8:74692898-74967280	Loss	MPX	Maternal	SNPs: |0|22|91
14417_5260	9:117311405-117727764	Loss	SPX	Paternal	SNPs: |30|0|141
9756_201	9:132572435-132597937	Loss	SPX	Unknown	No informative SNPs
6259_3	9:139516033-140208462	Loss	SPX	Paternal	SNPs: |23|0|316
6246_4	9:9399606-9631169	Loss	SPX	Unknown	No informative SNPs; no informative microsatellites***
5032_4	9:98998-334508	Loss	MPX	Paternal	SNPs: |15|0|128
13123_1403	9:98998-3682923	Loss	UNK	Paternal	SNPs: |204|0|1992
18172_302	10:34963905-35536952	Gain	SPX	Paternal	SNPs: |61|0|179
8534_201	10:45633089-51564756	Loss	SPX	Maternal	SNPs: |0|161|1197
14393_5020	10:74801551-75278951	Loss	SPX	Paternal	SNPs: |1|0|225 and patD10S188
6240_4	11:126633939-132060374	Loss	SPX	Paternal	SNPs: |351|0|2144
6325_3	11:70077507-70506315	Loss	SPX	Maternal	SNPs: |0|15|138 (Leblond et al 2012)[129]
6319_3	11:70119917-70187872	Loss	SPX	Paternal	SNPs: |3|0|24
5237_3	11:70154458-70220632	Loss	SPX	Maternal	SNPs: |0|3|15

Sample ID	Chr#:start-end (hg18)	CNV type	Family type	Parent of origin	SNPs: |#paternal|#maternal|total
6053_3	12:54218922-58779615	Gain	MPX	Maternal	SNPs: |0|241|1766
5272_3	12:98445422-98540678	Loss	MPX	Maternal	SNPs: |0|5|32
1050_3	14:20279711-20345174	Gain	MPX	Paternal	SNPs: |14|0|49
8638_201	14:35203374-35369853	Loss	UNK	Paternal	Custom designed microsatellite (8638b)
4272_1	14:63824114-65347410	Loss	SPX	Maternal	SNPs: |0|47|563 (Griswold et al. 2012)[128]
17009_1	14:78575296-78596793	Loss	SPX	Unknown	No informative SNPs or microsatellites
20187_1464001	15:18811937-26209270	Gain	SPX	Maternal	D15S1002, D15S128
8630_201	15:19800798-26209270	Gain	MPX	Paternal	SNPs: |328|9|2265
20069_1328001	15:20203578-26209270	Gain	SPX	Maternal	SNPs: |0|467|2230
17035_1	15:20274130-26120360	Gain	SPX	Maternal	D15S1002, D15S128
8430_204	15:20301669-20777695	Loss	SPX	Paternal	SNPs: |16|0|187
13050_593	15:21190624-26203954	Gain	UNK	Maternal	D15S128
16040_1571029001	15:28450423-30303265	Loss	MPX	Paternal	SNPs: |56|0|481
6101_4	15:74735339-74929817	Loss	MPX	Paternal	SNPs: |5|0|36
8695_201	15:81728085-82623936	Loss	SPX	Maternal	SNPs: |0|27|299
14181_2940	15:82906265-82985247	Gain	SPX	Maternal	SNPs: |0|4|18
14070_1230	15:91200007-91283004	Loss	MPX	Paternal	SNPs: |1|0|22
2204_1	16:29466569-30147029	Loss	SPX	Maternal	SNPs: |0|28|248
20089_1391001	16:29502984-30107306	Loss	SPX	Maternal	SNPs: |0|21|306
5068_3	16:29502984-30127026	Loss	MPX	Paternal	SNPs: |5|0|249
5262_4	16:29502984-30210849	Gain	SPX	Paternal	SNPs: |69|4|249
4030_1	16:29554843-30107306	Gain	MPX	Unknown	Ambiguous, SNPs: |22|22|253
5359_4	16:29554843-30195224	Loss	SPX	Maternal	SNPs: |0|1|248; no informative microsatellites
3439_3	17:17156307-18262979	Loss	SPX	Paternal	SNPs: 184|0|439
2211_1	17:17169258-20101517	Loss	SPX	Paternal	SNPs: |79|0|927
5056_4	17:34612208-34732327	Gain	MPX	Maternal	SNPs: |0|1|57; no informative microsatellites
8463_202	17:52774693-52895975	Loss	MPX	Paternal	SNPs: |5|0|66
5444_3	17:76953064-77782267	Gain	SPX	Paternal	SNPs: |20|5|276
5444_3	17:77785939-77849717	Loss	SPX	Paternal	Custom-designed microsatellite
3477_3	18:30280260-30327512	Loss	SPX	Maternal	Custom-designed microsatellite (20xAC, 22xAC)
14331_4450	18:72085223-73670156	Loss	SPX	Paternal	SNPs: |168|0|900
6358_6	19:4548413-5287389	Loss	SPX	Maternal	SNPs: |0|57|310
5335_3	20:14545734-14948785	Loss	SPX	Maternal	SNPs: |0|7|151
5046_3	20:8607242-8637441	Loss	MPX	Maternal	Custom-designed microsatellite (5046)
20180_1704001	21:43018846-43444308	Gain	SPX	Maternal	SNPs: |7|33|225
3183_7	22:17241748-19819918	Loss	MPX	Maternal	SNPs: |0|184|1251
8627_201	22:17257787-19793730	Gain	SPX	Paternal	SNPs:|292|94|1350
17015_1	22:17257787-19795780	Loss	SPX	Maternal	SNPs: |0|190|1444
4271_1	22:17257787-19795780	Gain	SPX	Maternal	SNPs: |105|249|1311
8630_201	22:32346124-32413987	Loss	MPX	Unknown	Ambiguous, SNPs: |21|25|291; no informative microsatellites
2072_1	22:45159185-49582267	Loss	MPX	Maternal	Gonadal mosaicism, two affected sibs (Moessner et al. 2008)[130]
6130_4	22:47996161-49512530	Loss	SPX	Paternal	SNPs: |170|0|875
14291_4120	22:49468716-49485255	Loss	SPX	Paternal	SNPs: |1|0|5
16079_1571066001	22:49470371-49567383	Loss	SPX	Paternal	SNPs: |7|0|39
20013_1075001	X:153239048-153521797	Gain	SPX	Unknown	—
16076_1571045001	X:22768481-23133948	Loss	SPX	Unknown	—

For 85 of 102 *de novo* events it was possible to determine the parent-of-origin from SNPs or microsatellite genotypes. Of the 85 *de novo* CNVs where parental origin could be assigned, 45 (53%) were paternally-derived and 40 (47%) were maternally-derived. These findings do not confirm the results from Hehir-Kwa et al.[131] of an increased rate of paternally-derived *de novo* events. Furthermore, paternal age was not found to be increased in fathers of ASD probands (with or without *de novo* CNVs) when compared to fathers of controls from that study (**Table S5H**). Parental bias was also not observed when we separately considered *de novo* events by type of CNV (pathogenic, uncertain significance, intergenic/intronic), mechanism of causality (mediated/flanked by segmental duplications or not) and type of family (simplex or multiplex).

* Similar results when using a larger 500 kb-flanking SNP window: ambiguous (SNPs: |8|17|203)
** Similar results when using a larger 500 kb-flanking SNP window: pat (SNPs: |1|0|369)
*** When using a 500 kb-flanking window: ambiguous (SNPs: |0|1|208)
Abbreviations: microsat, microsatellite marker; MPX, multiplex; SPX, simplex; UNK, unknown family type.

Tables S5B-S5H. Parent of origin of *de novo* CNVs – breakdown by family type and CNV characteristics

Table S5B. Breakdown by type of CNV

	Paternal	Maternal	Total
Pathogenic	23	20	43
Uncertain	20	19	39
Intergenic/ intronic	2	1	3
Total	45	40	85

Classification according to **Tables S5A** and **S7A**

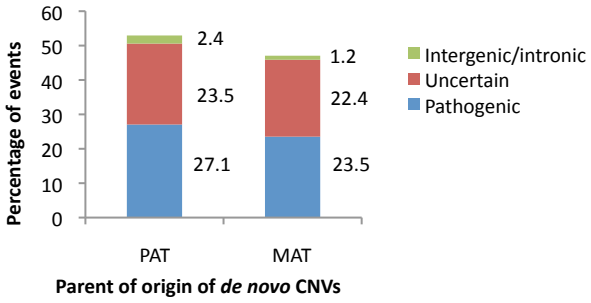


Table S5C. Breakdown by type of CNV, including large chromosomal abnormalities

	Paternal	Maternal	Total
Pathogenic	25	22	47
Uncertain	20	19	39
Intergenic/ intronic	2	1	3
Total	47	42	88

Classification according to **Tables S5A** and **S7A**

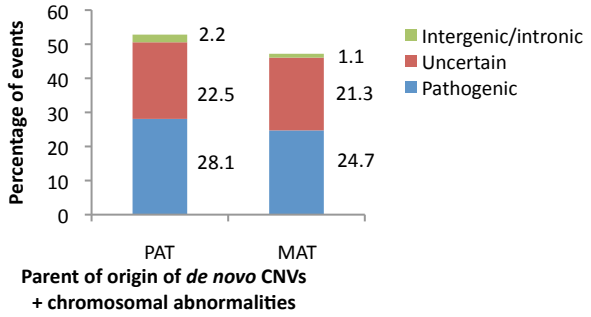


Table S5D. Breakdown by mechanism (CNV flanked or not by segmental duplications)

	Paternal		Maternal		Total
	SD	no SD	SD	no SD	
Pathogenic	10	13	13	7	43
Uncertain	1	19	1	18	39
Intergenic/ intronic	0	2	0	1	3
Total	11	34	14	26	85

Classification according to **Tables S5A** and **S7A**
SD, segmental duplication

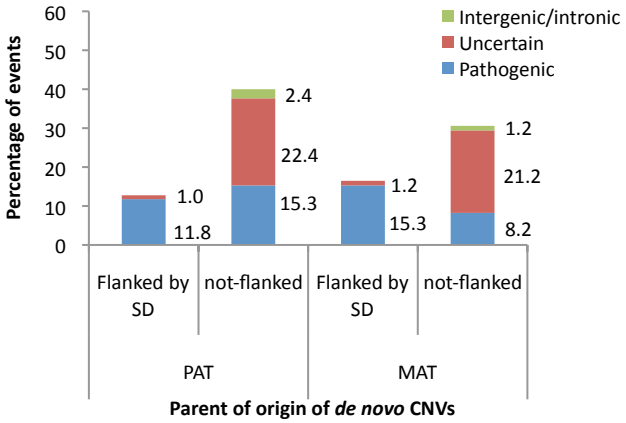


Table S5E. Breakdown of CNVs by type of family

	# Probands without de novo events	# Probands with *de novo* events	Total	%
AGP SPX	1142	59	1201	4.9
AGP MPX	600	26	626	4.2
AGP Unknown	256	13	269	4.8
Total	1998	98	2096	4.7
SSC	1059	65	1124	5.8

Number of simplex (SPX) and multiplex (MPX) families after QC is 1201 and 626, respectively. SSC, Simons Simplex Collection[132]

Table S5F. Breakdown of *de novo* events by type of family

	Paternal	Maternal	Total
AGP SPX	23	26	49
AGP MPX	13	9	22
SSC	27	34	61
Total	63	69	132

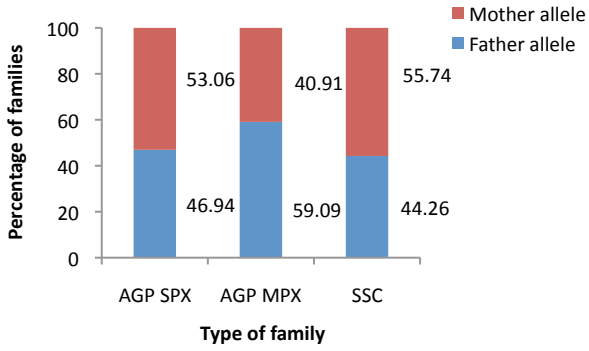


Table S5G. Breakdown of *de novo* CNVs by size in different family types

	<30 kb	30 kb-500 kb	500 kb-1 Mb	>1 Mb
AGP SPX	4	23	15	17
AGP MPX	0	14	2	10
Total	4	37	17	27

Average CNV size is 1,199,751 bp for SPX, 1,180,129 bp for MPX, and 1,041,851 bp for families classified as unknown.

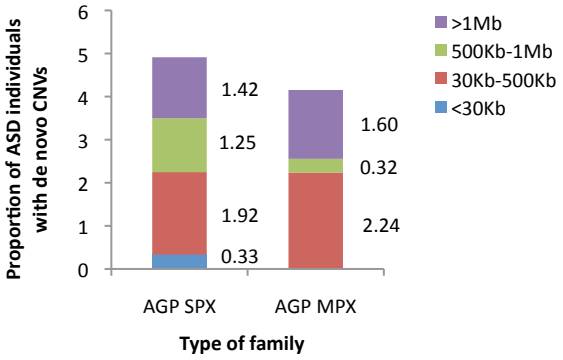


Table S5H. Breakdown of *de novo* CNVs by paternal age

	# Fathers in group*	Median paternal age ± SD
ASD without *de novo* CNVs	1659	32.33 ± 5.74
ASD with *de novo* CNVs	76	32.17 ± 6.57
ASD with *de novo* CNVs, paternal in origin, with flanking SD	6	33.75 ± 6.03
ASD with *de novo* CNVs, paternal in origin, without flanking SD	24	33.71 ± 6.08

*45 *de novo* CNVs were paternally derived from 43 fathers (2 affected individuals had 2 paternally-derived *de novo* CNVs each), and parental age was available for 30 of them. The average maternal and paternal age at childbirth in our group of rare *de novo* CNVs was 29.6 and 32.2 years, respectively (76/99 had information on parental age). These age ranges were similar to parental ages at childbirth in the control cohort used by Hehir-Kwa et al.[131] (31.4 and 32.1 years, respectively).

Tables S6A-S6D. List of genes and loci implicated in ASD and ID

Table S6A. Genes implicated in ASD

	Gene	Chr. band	Disorder	Inheritance pattern
1	*POMGNT1*	1p34.1	Muscle-eye-brain disease	AR
2	*RPE65*	1p31.3	Leber congenital amaurosis	AR
3	*DPYD*	1p21.3	Dihydropyrimidine dehydrogenase deficiency	AR
4	*NRXN1*	2p16.3	Disrupted in ASD, ID, schizophrenia (dominant); Pitt-Hopkins-like syndrome 2 (recessive)	AD/AR
5	*NPHP1*	2q13	Joubert syndrome type 4, nephronophthisis	AR
6	*MBD5*	2q23.1	Autosomal dominant ID, 2q23.1 microdeletion syndrome	AD
7	*SCN1A*	2q24.3	Severe myoclonic epilepsy of infancy (Dravet syndrome)	AD
8	*SCN2A*	2q24.3	Benign familial neonatal-infantile seizures, intractable childhood epilepsy	AD
9	*SATB2*	2q33.1	Cleft palate and ID; implicated in the 2q33.1 microdeletion syndrome	AD
10	*BTD*	3p24.3	Biotinidase deficiency	AR
11	*FOXP1*	3p14.1	Non-syndromic ID with language impairment and ASD	AD
12	*PRSS12*	4q26	Autosomal recessive non-syndromic ID	AR
13	*NIPBL*	5p13.2	Cornelia de Lange syndrome	AD
14	*MEF2C*	5q14.3	ID, stereotypic movements, epilepsy, and/or cerebral malformations; 5q14.3 microdeletion syndrome	AD
15	*ALDH7A1*	5q23.2	Pyridoxine-dependent epilepsy	AR
16	*NSD1*	5q35.2-q35.3	Sotos syndrome	AD
17	*ALDH5A1*	6p22.2	Succinic semialdehyde dehydrogenase deficiency	AR
18	*SYNGAP1*	6p21.32	Autosomal dominant ID	AD
19	*AHI1*	6q23.3	Joubert syndrome 3	AR
20	*PEX7*	6q23.3	Refsum disease, Rhizomelic chondrodysplasia punctata, type 1	AR
21	*ARID1B*	6q25.3	ID, speech impairment, minor anomalies and variable corpus callosum abnormalities; Coffin-Siris syndrome	AD
22	*HOXA1*	7p15.2	HOXA1 syndrome, Bosley-Salih-Alorainy variant	AR
23	*BRAF*	7q34	Cardio-facio-cutaneous syndrome	AD
24	*CNTNAP2*	7q35-q36.1	Cortical dysplasia-focal epilepsy syndrome, Pitt-Hopkins-like syndrome 1 (recessive); the clinical significance of the disruption of 1 allele is unknown	AR
25	*HGSNAT*	8p11.21	Mucopolysaccharidosis type IIIC (Sanfilippo syndrome C)	AR
26	*TUSC3*	8p22	Autosomal recessive non-syndromic ID	AR
27	*CHD7*	8q12.2	CHARGE syndrome	AD
28	*VPS13B*	8q22.2	Cohen syndrome	AR
29	*SMARCA2*	9p24.3	Nicolaides-Baraitser syndrome; Coffin-Siris syndrome	AD
30	*STXBP1*	9q34.11	Non-syndromic epilepsy, ID and autism; early infantile epileptic encephalopathy	AD
31	*POMT1*	9q34.13	Limb-girdle muscular dystrophy with ID; Walker-Warburg syndrome	AR
32	*TSC1*	9q34.13	Tuberous sclerosis	AD
33	*EHMT1*	9q34.3	Kleefstra syndrome (9q subtelomeric deletion syndrome)	AD
34	*PTEN*	10q23.31	PTEN hamartoma-tumor syndrome, ID and ASD with macrocephaly	AD
35	*FGFR2*	10q26.13	Apert syndrome	AD
36	*HRAS*	11p15.5	Costello syndrome	AD
37	*IGF2*	11p15.5	Aberrant imprinting of *IGF2* is associated with Beckwith–Wiedermann syndrome and Silver–Russell syndrome	AD
38	*KCNJ11*	11p15.1	DEND syndrome (developmental delay, epilepsy, and neonatal diabetes)	AD
39	*SHANK2*	11q13.3	Non-syndromic ID and ASD	AD
40	*DHCR7*	11q13.4	Smith-Lemli-Opitz syndrome	AR
41	*FOLR1*	11q13.4	Cerebral folate transport deficiency	AR
42	*HEPACAM*	11q24.2	Megalencephalic leukoencephalopathy with subcortical cysts (recessive); leukodystrophy and macrocephaly (dominant)	AR/AD
43	*CACNA1C*	12p13.33	Timothy syndrome	AD
44	*GRIN2B*	12p13.1	Autosomal dominant ID	AD
45	*KRAS*	12p12.1	Cardio-facio-cutaneous syndrome	AD
46	*SCN8A*	12q13.13	Early infantile epileptic encephalopathy	AD
47	*GNS*	12q14.3	Mucopolysaccharidosis type IIID (Sanfilippo disease D)	AR
48	*BBS10*	12q21.2	Bardet-Biedl syndrome	AR
49	*CEP290*	12q21.32	Joubert syndrome 5, Leber congenital amaurosis, Bardet-Biedl syndrome, Meckel syndrome	AR
50	*PAH*	12q23.2	Phenylketonuria	AR
51	*PTPN11*	12q24.13	Noonan syndrome	AD
52	*CHD8*	14q11.2	Autosomal dominant ASD and ID	AD
53	*FOXG1*	14q12	Congenital variant of Rett syndrome	AD
54	*L2HGDH*	14q22.1	L-2-hydroxyglutaric aciduria	AR
55	*UBE3A*	15q11.2	Angelman syndrome	AD
56	*SPRED1*	15q14	Neurofibromatosis type 1-like syndrome (Legius syndrome)	AD
57	*GATM*	15q21.1	Arginine:glycine amidinotransferase (AGAT) deficiency	AR
58	*MAP2K1*	15q22.31	Cardio-facio-cutaneous syndrome	AD
59	*TSC2*	16p13.3	Tuberous sclerosis	AD
60	*CREBBP*	16p13.3	Rubinstein-Taybi syndrome	AD
61	*SRCAP*	16p11.2	Floating-Harbor syndrome	AD
62	*BCKDK*	16p11.2	Autosomal recessive autism, ID & epilepsy/abnormal EEG	AR
63	*RPGRIP1L*	16q12.2	Joubert syndrome 7, Meckel syndrome, COACH syndrome	AR

	Gene	Chr. band	Disorder	Inheritance pattern
64	*ANKRD11*	16q24.3	KBG syndrome; 16q24.3 microdeletion syndrome	AD
65	*YWHAE*	17p13.3	Miller-Dieker syndrome	AD
66	*PAFAH1B1*	17p13.3	Isolated lissencephaly, Miller-Dieker syndrome	AD
67	*GUCY2D*	17p13.1	Leber congenital amaurosis	AR
68	*RAI1*	17p11.2	Smith-Magenis syndrome (deletion, mutation), Potocki-Lupski syndrome (duplication)	AD
69	*RNF135*	17q11.2	Overgrowth syndrome; haploinsufficiency of *RNF135* contributes to the phenotype of the NF1 microdeletion syndrome	AD
70	*NF1*	17q11.2	Neurofibromatosis type 1	AD
71	*NAGLU*	17q21.31	Mucopolysaccharidosis type IIIB (Sanfilippo syndrome B)	AR
72	*SGSH*	17q25.3	Sanfilippo syndrome A (mucopolysaccharidosis III A)	AR
73	*SETBP1*	18q12.3	Haploinsufficiency of *SETBP1* causes the core clinical features of the del(18)(q12.2q21.1) syndrome	AD
74	*SMAD4*	18q21.2	Myhre syndrome	AD
75	*TCF4*	18q21.2	Pitt-Hopkins syndrome	AD
76	*NFIX*	19p13.13	Sotos-like overgrowth syndrome, Marshall-Smith syndrome	AD
77	*GAMT*	19p13.3	Guanidine acetate methyltransferase (GAMT) deficiency	AR
78	*DMPK*	19q13.32	Myotonic dystrophy type 1 (Steinert disease)	AD
79	*MKKS*	20p12.2	Bardet-Biedl syndrome	AR
80	*DYRK1A*	21q22.13	Autosomal dominant ID	AD
81	*ADSL*	22q13.1	Adenylosuccinate lyase deficiency	AR
82	*SHANK3*	22q13.33	Phelan-mcdermid syndrome (22q13 deletion syndrome)	AD
83	*NLGN4X*	Xp22.31-p22.32	Non-syndromic X-linked ID and ASD	XLR
84	*MID1*	Xp22.2	Opitz syndrome (Opitz/BBB syndrome)	XLR
85	*AP1S2*	Xp22.2	Syndromic X-linked ID, Fried type; non-syndromic X-linked ID	XLR
86	*NHS*	Xp22.13	Nance-Horan syndrome	XLD
87	*CDKL5*	Xp22.13	Early infantile epileptic encephalopathy	XLD
88	*PTCHD1*	Xp22.11	Non-syndromic X-linked ID and ASD	XLR
89	*ARX*	Xp21.3	X-linked lissencephaly and abnormal genitalia, West syndrome, Partington syndrome, non-syndromic X-linked ID	XLR
90	*IL1RAPL1*	Xp21.2-p21.3	Non-syndromic X-linked ID and ASD	XLR
91	*DMD*	Xp21.1-21.2	Muscular dystrophy, Duchenne and Becker types	XLR
92	*OTC*	Xp11.4	Ornithine transcarbamylase deficiency	XLD/XLR
93	*CASK*	Xp11.4	Syndromic and non-syndromic X-linked ID	XLD
94	*NDP*	Xp11.3	Norrie disease	XLR
95	*KDM6A*	Xp11.3	Kabuki syndrome	XLD
96	*SYN1*	Xp11.23	X-linked epilepsy and ID	XLR
97	*FTSJ1*	Xp11.23	Non-syndromic X-linked ID	XLR
98	*PQBP1*	Xq11.23	Renpenning syndrome, non-syndromic ID	XLR
99	*CACNA1F*	Xp11.23	X-linked incomplete congenital stationary night blindness	XLR
100	*KDM5C*	Xp11.22	Syndromic X-linked ID, Claes-Jensen type; non-syndromic X-linked ID	XLR
101	*IQSEC2*	Xp11.22	Non-syndromic X-linked ID	XLD
102	*SMC1A*	Xp11.22	Cornelia de Lange syndrome	XLD
103	*HSD17B10*	Xp11.22	17-beta-hydroxysteroid dehydrogenase X deficiency, X-linked syndromic ID with choreoathetosis and abnormal behavior	XLD
104	*PHF8*	Xp11.22	Siderius-Hamel syndrome	XLR
105	*FGD1*	Xp11.22	Aarskog-Scott syndrome, non-syndromic X-linked ID	XLR
106	*OPHN1*	Xq12	X-linked ID with cerebellar hypoplasia and distinctive facial appearance	XLR
107	*MED12*	Xq13.1	Lujan-Fryns syndrome, Opitz-Kaveggia syndrome	XLR
108	*NLGN3*	Xq13.1	Non-syndromic X-linked ID and ASD	XLR
109	*ATRX*	Xq21.1	Alpha-thalassemia/mental retardation syndrome, non-syndromic X-linked ID	XLD
110	*PCDH19*	Xq22.1	Female-limited epilepsy with ID, early infantile epileptic encephalopathy	XL - affected females
111	*ACSL4*	Xq22.3	Non-syndromic X-linked ID	XLD
112	*DCX*	Xq22.3	Type 1 lissencephaly	XLD
113	*UPF3B*	Xq24	Non-syndromic X-linked ID, Opitz-Kaveggia/Lujan-Fryns phenotype	XLR
114	*LAMP2*	Xq24	Danon disease	XLD
115	*GRIA3*	Xq25	Syndromic X-linked ID, Wu type; non-syndromic X-linked ID	XLR
116	*OCRL*	Xq25	Lowe syndrome	XLR
117	*PHF6*	Xq26.2	Borjeson-Forssman-Lehmann syndrome	XLR
118	*SLC9A6*	Xq26.3	Syndromic X-linked ID, Christianson type	XLD?
119	*FMR1*	Xq27.3	Fragile X syndrome	XLD
120	*AFF2*	Xq28	Fragile X mental retardation 2	XLR
121	*SLC6A8*	Xq28	Creatine deficiency syndrome, non-syndromic X-linked ID	XLR
122	*L1CAM*	Xq28	MASA (mental retardation, aphasia, shuffling gait, and adducted thumbs) syndrome	XLR
123	*MECP2*	Xq28	Rett syndrome, non-syndromic X-linked ID (mutation, deletion; XL dominant); *MECP2* duplication syndrome (XL recessive)	XLD/XLR
124	*RAB39B*	Xq28	Non-syndromic X-linked ID	XLR

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; ASD, autism spectrum disorder; ID, intellectual disability; XL, X-linked; XLD, X-linked dominant; XLR, X-linked recessive

Table S6B. Loci implicated in ASD

	Disorder	Chr. band	Start (hg18)	End (hg18)	Genes involved	Inheritance pattern
1	1p36 deletion syndrome	1p36.32-p36.33	823,946	5,308,621	contiguous gene syndrome	AD
2	1q21.1 deletion/duplication syndrome	1q21.1	145,044,110	145,861,130	?	AD
3	1q41q42 microdeletion syndrome	1q41q42	219,500,000	223,000,000	?	AD
4	1q43q44 microdeletion syndrome	1q43q44	238,000,000	247,249,718	?	AD
5	2p16.1p15 microdeletion syndrome	2p16.1p15	57,595,300	61,591,838	?	AD
6	2q13 deletion/duplication	2q13	111,158,601	112,782,250	?	AD
7	2q23.1 microdeletion syndrome	2q23.1	148,932,242	148,987,514	*MBD5*	AD
8	2q33.1 deletion syndrome	2q32.3-q33.2	196,633,334	204,915,185	*SATB2* (maybe other gene(s) contribute too)	AD
9	2q37 deletion syndrome (brachydactyly-mental retardation syndrome)	2q37.3	239,619,630	242,579,273	*HDAC4* (other distal gene(s) contribute too)	AD
10	3q13.31 microdeletion syndrome	3q12.3q21.3	115,335,356	115,916,848	?	AD
11	3q29 microdeletion/microduplication syndrome	3q29	197,240,451	198,829,062	? (candidates: *PAK2* and *DLG1*)	AD
12	Wolf-Hirschhorn syndrome	4p16.3	62,448	2,297,002	contiguous gene syndrome	AD
13	4q21 microdeletion syndrome	4q21.21-q21.22	82,228,875	83,182,488	?	AD
14	Cri du Chat syndrome (5p deletion)	5p15.2-p15.33	90,693	11,400,262	?	AD
15	5q14.3 microdeletion syndrome	5q14.3	88,049,814	88,235,678	*MEF2C*	AD
16	Sotos syndrome (deletion), 5q35.2q35.3 duplication	5q35.2-q35.3	175,661,584	176,946,567	*NSD1*	AD
17	6p subtelomere deletion syndrome	6p25	100,000	3,000,000	? (*FOXC1* involved in ophthalmologic findings)	AD
18	Williams syndrome (deletion), 7q11.23 duplication syndrome	7q11.23	72,382,390	73,780,449	contiguous gene syndrome	AD
19	8p23.1 deletion/duplication syndrome	8p23.1	8,156,705	11,803,128	? (*GATA4* involved in heart defects)	AD
20	8q21.11 microdeletion syndrome	8q21.11	77,389,019	77,928,794	?	AD
21	Kleefstra syndrome (9q subtelomeric deletion syndrome)	9q34.3	139,523,184	140,273,252	*EHMT1*	AD
22	10p14p15 deletion syndrome	10p14-p15.1	4,700,001	10,600,000	? *(GATA3* involved in hypoparathyroidism, deafness, renal disease)	AD
23	10p12p11 microdeletion	10p12.31p11.21	28,833,195	29,138,742	?	AD
24	10q22-q23 deletion	10q22.3-q23.2	81,682,644	88,931,994	?	AD
25	Distal 10q deletion syndrome	10q26.2-q26.3	128,000,000	135,374,737	?	AD
26	11p15.5 duplication: Beckwith-Wiedemann/Silver-Russell syndromes	11p15.4-p15.5	1,970,000	2,870,000	*H19, IGF2*	AD
27	WAGR syndrome (11p13 deletion syndrome)	11p13	31,760,085	32,467,564	?	AD
28	Potocki-Shaffer syndrome (11p11.2 deletion syndrome)	11p11.2	43,941,853	46,021,136	*PHF21A (EXT2* and *ALX4* involved in bone defects)	AD
29	Jacobsen syndrome (11q deletion syndrome)	11q23.3-qter	115,400,001	134,452,384	?	AD
30	12q14 microdeletion syndrome	12q14	63,358,186	66,931,792	? (*HMGA2* involved in short stature)	AD
31	Terminal deletion 14q syndrome	14q32.31-q32.33	101,500,000	105,000,000	?	AD
32	Angelman syndrome (maternal deletion), Prader-Willi syndrome (paternal deletion), 15q11-q13 duplication syndrome	15q11.2-q13.1	21,309,483	26,230,781	Angelman: maternal *UBE3A*; Prader-Willi: paternally expressed genes (HBII-85 snoRNA cluster; *SNURF-SNRPN, NDN* and *MAGEL2*?)	AD
33	15q13.3 deletion syndrome (duplication = uncertain significance)	15q13.2-q13.3	28,924,396	30,232,700	*CHRNA7*	AD
34	15q24 microdeletion syndrome	15q24.1-q24.2	72,164,227	73,949,332	?	AD
35	Distal 15q25.2q25.3 microdeletion	15q25.2q25.3	82,944,098	83,484,862	?	AD
36	15q26 overgrowth syndrome	15q26.3	97,175,493	100,218,756	*IGF1R*	AD
37	Rubinstein-Taybi syndrome (deletion), 16p13.3 duplication syndrome	16p13.3	3,715,057	3,870,122	*CREBBP*	AD
38	16p13.11 microdeletion syndrome (duplication = uncertain significance)	16p13.11	15,411,955	16,199,769	?	AD
39	16p11.2-p12.2 microdeletion/microduplication syndrome	16p11.2-p12.2	21,521,457	28,949,693	?	AD
40	16p11.2 microdeletion/microduplication syndrome	16p11.2	29,557,497	30,107,356	?	AD
41	Miller-Dieker syndrome/isolated lissencephaly (deletion), 17p13.3 microduplication	17p13.3	1,129,706	2,535,659	*PAFAH1B1, YWHAE*	AD
42	Smith-Magenis syndrome (deletion), Potocki-Lupski syndrome (duplication)	17p11.2	16,697,836	20,160,243	*RAI1*	AD
43	NF1 microdeletion/microduplication syndrome	17q11.2	26,186,948	27,242,780	*NF1* (*RNF135* contributes to the overgrowth, facial dysmorphism and ID)	AD
44	17q12 deletion syndrome (renal cysts and diabetes syndrome), 17q12 duplication syndrome	17q12	31,930,169	33,323,031	? (*HNF1B* involved in renal cysts and diabetes syndrome)	AD
45	Koolen-De Vries syndrome (17q21.31 microdeletion syndrome), 17q21.31 microduplication syndrome	17q21.31	41,060,949	41,544,225	*KANSL1*	AD
46	del(18)(q12.2q21.1) syndrome	18q12.2q21.1	39,890,000	41,700,000	*SETBP1*	AD
47	Down syndrome (trisomy 21)	whole chr 21	1	46,944,323	contiguous gene syndrome	AD

	Disorder	Chr. band	Start (hg18)	End (hg18)	Genes involved	Inheritance pattern
48	22q11.2 deletion syndrome (velocardiofacial/DiGeorge syndrome), 22q11.2 duplication syndrome	22q11.21-q11.22	17,041,725	18,691,904	contiguous gene syndrome, *TBX1* is responsible for most of the physical malformations	AD
49	Phelan-McDermid syndrome (22q13 deletion syndrome), 22q13 duplication	22q13.33	49,392,382	49,525,811	*SHANK3*	AD
50	Xq28 duplication syndrome (*MECP2* duplication syndrome)	Xq28	152,403,094	153,044,193	*MECP2*	XLR
51	Turner syndrome (X0)	whole chr X	1	154,913,754	contiguous gene syndrome	
52	Klinefelter syndrome (XXY)	whole chr X	1	154,913,754	contiguous gene syndrome	
53	XYY syndrome	whole chr Y	1	57,772,954	contiguous gene syndrome	
54	XXYY syndrome	whole chr X-Y			contiguous gene syndrome	
55	45,X/46,XY mosaicism	whole chr X	1	154,913,754	contiguous gene syndrome	

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; ASD, autism spectrum disorder; ID, intellectual disability; XLR, X-linked recessive

Table S6C. Genes implicated in ID

	Gene	Chr. band	Disorder	Inheritance pattern
1	*SKI*	1p36.33	Shprintzen-Goldberg syndrome	AD
2	*GALE*	1p36.11	Galactose epimerase deficiency (galactosemia III)	AR
3	*FUCA1*	1p36.11	Fucosidosis	AR
4	*ARID1A*	1p36.11	Coffin-Siris syndrome	AD
5	*PIGV*	1p36.11	Hyperphosphatasia mental retardation syndrome	AR
6	*SLC2A1*	1p34.2	Glucose transport defect	AD
7	*ST3GAL3*	1p34.3	Autosomal recessive non-syndromic ID	AR
8	*STIL*	1p33	Primary microcephaly	AR
9	*ALG6*	1p31.3	Congenital disorder of glycosylation, type Ic	AR
10	*DBT*	1p21.2	Maple syrup urine disease, type II	AR
11	*AP4B1*	1p13.2	Autosomal recessive ID with spastic paraplegia	AR
12	*NRAS*	1p13.2	Noonan syndrome	AD
13	*GATAD2B*	1q21.3	Autosomal dominant ID	AD
14	*KCNJ10*	1q23.2	SESAME syndrome (seizures, sensorineural deafness, ataxia, mental retardation, and electrolyte imbalance)	AR
15	*ASPM*	1q31	Microcephaly and ID	AR
16	*SYT14*	1q32.2	ID with adult-onset spinocerebellar ataxia	AR
17	*RAB3GAP2*	1q41	Martsolf syndrome (congenital cataracts, hypogonadism, and ID)	AR
18	*TBCE*	1q42.3	Hypoparathyroidism-retardation-dysmorphism syndrome	AR
19	*FH*	1q43	Fumarase deficiency	AR
20	*MYCN*	2p24.3	Feingold syndrome (microcephaly-oculo-digito-esophageal-duodenal syndrome), Microcephaly and digital abnormalities with normal intelligence	AD
21	*SOS1*	2p22.1	Noonan syndrome	AD
22	*ERCC3*	2q14.3	Trichothiodystrophy; Xeroderma pigmentosum, group B	AR
23	*RAB3GAP1*	2q21.3	Warburg Micro syndrome 1	AR
24	*ZEB2*	2q22.3	Mowat-Wilson syndrome (Hirschsprung disease-mental retardation syndrome)	AD
25	*BBS5*	2q31.1	Bardet-Biedl syndrome	AR
26	*GAD1*	2q31.1	Spastic quadriplegic cerebral palsy	AR
27	*TMEM237*	2q33.1	Joubert syndrome 14	AR
28	*HDAC4*	2q37.3	Brachydactyly mental retardation syndrome (2q37 deletion syndrome)	AD
29	*D2HGDH*	2q37.3	D-2-hydroxyglutaric aciduria	AR
30	*CRBN*	3p26.2	Autosomal recessive non-syndromic ID	AR
31	*SUMF1*	3p26.2	Multiple sulfatase deficiency	AR
32	*TSEN2*	3p25.1	Pontocerebellar hypoplasia type 2B	AR
33	*RAF1*	3p25.1	Noonan syndrome, LEOPARD syndrome	AD
34	*TGFBR2*	3p24.1	Loeys–Dietz syndrome	AD
35	*GLB1*	3p22.3	GM1-gangliosidosis, Mucopolysaccharidosis IVB	AR
36	*CTNNB1*	3p22.1	Autosomal dominant ID	AD
37	*GTDC2*	3p22.1	Walker-Warburg syndrome	AR
38	*LZTFL1*	3p21.31	Bardet-Biedl syndrome with situs inversus and polydactyly	AR
39	*DAG1*	3p21.31	Limb-girdle muscular dystrophy	AR
40	*ARL13B*	3q11.2	Joubert syndrome 8	AR
41	*ARL6*	3q11.2	Bardet-Biedl syndrome	AR
42	*CEP63*	3q22.2	Primary microcephaly	AR
43	*ATR*	3q23	Seckel syndrome	AR
44	*ALG3*	3q27.1	Congenital disorder of glycosylation, type Id	AR
45	*KIAA0226*	3q29	Syndromic ID with ataxia, dysarthria and epilepsy	AR
46	*IDUA*	4p16.3	Mucopolysaccharidosis Ih (Hurler syndrome); mucopolysaccharidosis Is (Scheie syndrome)	AR
47	*CC2D2A*	4p15.3	Joubert syndrome 9, Meckel syndrome, COACH syndrome	AR
48	*QDPR*	4p15.32	Hyperphenylalaninemia due to dihydropteridine reductase deficiency	AR
49	*SRD5A3*	4q12	Kahrizi syndrome; congenital disorder of glycosylation, type Iq	AR
50	*TMEM165*	4q12	Congenital disorder of glycosylation, type iik	AR
51	*CEP135*	4q12	Autosomal-recessive primary microcephaly	AR

	Gene	Chr. band	Disorder	Inheritance pattern
52	*SLC4A4*	4q13.3	Renal tubular acidosis, proximal, with ocular abnormalities	AR
53	*BBS7*	4q27	Bardet-Biedl syndrome	AR
54	*BBS12*	4q27	Bardet-Biedl syndrome	AR
55	*AGA*	4q34.3	Aspartylglucosaminuria	AR
56	*NSUN2*	5p15.31	Autosomal-recessive syndromic ID, Dubowitz syndrome	AR
57	*ANKH*	5p15.2	Chondrocalcinosis 2, Craniometaphyseal dysplasia	AR
58	*C5orf42*	5p13.2	Joubert syndrome 17	AR
59	*MOCS2*	5q11.2	Molybdenum cofactor deficiency, type B	AR
60	*ERCC8*	5q12.1	Cockayne syndrome type A	AR
61	*SIL1*	5q31.2	Marinesco-Sjogren syndrome	AR
62	*TUBB2B*	6p25.2	Asymmetric polymicrogyria	AD
63	*NEU1*	6p21.3	Sialidosis type I, Sialidosis type II	AR
64	*MOCS1*	6p21.2	Molybdenum cofactor deficiency, type A	AR
65	*SLC17A5*	6q13	Salla disease; Sialic acid storage disorder, infantile	AR
66	*ELOVL4*	6q14.1	Ichthyosis, spastic quadriplegia, and mental retardation (recessive); Macular dystrophy (dominant)	AR/AD
67	*BCKDHB*	6q14.1	Maple syrup urine disease, type Ib	AR
68	*RARS2*	6q15	Pontocerebellar hypoplasia	AR
69	*GRIK2*	6q16.3	Autosomal recessive non-syndromic ID	AR
70	*SOBP*	6q21	Autosomal recessive syndromic and non-syndromic ID	AR
71	*LAMA2*	6q22.33	Merosin-deficient congenital muscular dystrophy type 1A	AR
72	*ARG1*	6q23.2	Argininemia	AR
73	*MED23*	6q23.2	Autosomal recessive non-syndromic ID	AR
74	*GTF2H5*	6q25.3	Trichothiodystrophy	AR
75	*ACTB*	7p22.1	Baraitser-Winter syndrome	AD
76	*FAM126A*	7p15.3	Hypomyelinating leukodystrophy	AR
77	*BBS9*	7p14.3	Bardet-Biedl syndrome	AR
78	*C7orf11*	7p14.1	Trichothiodystrophy	AR
79	*GUSB*	7q11.21	Mucopolysaccharidosis VII	AR
80	*AP4M1*	7q22.1	Autosomal recessive tetraplegic cerebral palsy with ID	AR
81	*RELN*	7q22	Lissencephaly	AR
82	*DLD*	7q31.1	Maple syrup urine disease, type III	AR
83	*CEP41*	7q32.2	Joubert syndrome 15	AR
84	*TPK1*	7q35	Thiamine metabolism dysfunction syndrome (episodic encephalopathy type)	AD
85	*EZH2*	7q36.1	Weaver syndrome	AD
86	*MCPH1*	8p23	Microcephaly and ID	AR
87	*ERLIN2*	8p12	Autosomal recessive ID, motor dysfunction and multiple joint contractures	AR
88	*CA8*	8q12.1	Cerebellar ataxia, quadrupedal locomotion and ID	AR
89	*TMEM67*	8q21	Joubert syndrome 6, Meckel-Gruber syndrome	AR
90	*RAD21*	8q24.11	Cornelia de Lange syndrome	AD
91	*KCNK9*	8q24.3	Birk-Barel mental retardation dysmorphism syndrome (imprinting defect)	AD
92	*TRAPPC9*	8q24.3	Autosomal recessive non-syndromic ID	AR
93	*RECQL4*	8q24.3	Baller-Gerold syndrome, Rothmund-Thomson syndrome and RAPADILINO syndrome	AR
94	*VLDLR*	9p24.2	Cerebellar ataxia and ID	AR
95	*SLC1A1*	9p24.2	Dicarboxylic aminoaciduria	AR
96	*PIGO*	9p13.3	Hyperphosphatasia with mental retardation syndrome	AR
97	*EXOSC3*	9p13.2	Pontocerebellar hypoplasia and spinal motor neuron degeneration	AR
98	*TGFBR1*	9q22.33	Loeys–Dietz syndrome	AD
99	*FKTN*	9q31.2	Fukuyama congenital muscular dystrophy with type 2 lissencephaly, Walker-Warburg syndrome	AR
100	*TRIM32*	9q33.1	Bardet-Biedl syndrome	AR
101	*CDK5RAP2*	9q33.2	Microcephaly vera	AR
102	*SPTAN1*	9q34.11	West syndrome with severe cerebral hypomyelination, spastic quadriplegia and ID	AD
103	*KCNT1*	9q34.3	Malignant migrating partial seizures of infancy; nocturnal frontal lobe epilepsy, ID	AD
104	*INPP5E*	9q34.3	Joubert syndrome 1	AR
105	*MAN1B1*	9q34.3	Autosomal recessive non-syndromic ID	AR
106	*RAB18*	10p12.1	Warburg micro syndrome	AR
107	*ERCC6*	10q11.23	Cockayne syndrome type B, Cerebro-oculo-facio-skeletal syndrome	AR
108	*KIAA1279*	10q21.3	Goldberg-Shprintzen megacolon syndrome	AR
109	*ADK*	10q22.2	Hypermethioninemia due to adenosine kinase deficiency	AR
110	*KAT6B*	10q22.2	Say-Barber-Biesecker-Young-Simpson syndrome (SBBYSS syndrome)	AD
111	*POLR3A*	10q22.3	Hypomyelinating leukodystrophy with or without oligodontia and/or hypogonadotropic hypogonadism	AR
112	*KIF11*	10q23.33	Microcephaly variably associated with congenital lymphedema, chorioretinopathy and learning difficulties	AD
113	*TCTN3*	10q23.33	Joubert syndrome 18, orofaciodigital syndrome IV	AR
114	*SMC3*	10q25.2	Cornelia de Lange syndrome	AD
115	*SHOC2*	10q25.2	Noonan syndrome	AD
116	*SLC25A22*	11p15.5	Autosomal recessive neonatal epileptic encephalopathy	AR
117	*PAX6*	11p13	Isolated and syndromic aniridia, including Gillespie syndrome (aniridia, cerebellar ataxia and ID)	AD
118	*SLC35C1*	11p11.2	Congenital disorder of glycosylation, type iic	AR
119	*PHF21A*	11p11.2	Potocki-Shaffer syndrome (11p11.2 deletion)	AD
120	*TMEM138*	11q12.2	Joubert syndrome 16	AR
121	*TMEM216*	11q12.2	Joubert syndrome 2	AR
122	*BBS1*	11q13.1	Bardet-Biedl syndrome	AR

	Gene	Chr. band	Disorder	Inheritance pattern
123	*ALG8*	11q14.1	Congenital disorder of glycosylation type Ih	AR
124	*MED17*	11q21	Primary microcephaly of postnatal onset, spasticity, epilepsy, and ID	AR
125	*ALG9*	11q23.1	Congenital disorder of glycosylation, type Il	AR
126	*CBL*	11q23.3	Noonan syndrome-like disorder	AD
127	*PVRL1*	11q23.3	Cleft lip/palate ectodermal dysplasia syndrome	AR
128	*MLL2*	12q13.12	Kabuki syndrome	AD
129	*TUBA1A*	12q13.12	Lissencephaly	AD
130	*DIP2B*	12q13.13	Mental retardation, FRA12A type	AD
131	*SUOX*	12q13.2	Sulfite oxidase deficiency	AR
132	*GNPTAB*	12q23.2	Mucolipidosis III alpha/beta	AR
133	*IGF1*	12q23.2	Growth retardation with deafness and mental retardation due to IGF1 deficiency	AR
134	*POLR3B*	12q23.3	Leukodystrophy, hypomyelinating, 8, with or without oligodontia and/or hypogonadotropic hypogonadism	AR
135	*ATP6V0A2*	12q24.31	Cutis laxa with epilepsy and mental retardation	AR
136	*CENPJ*	13q12.12	Microcephaly vera, Seckel syndrome	AR
137	*SLC25A15*	13q14.11	Orinithine translocase deficiency	AR
138	*MIR17HG*	13q31.3	Feingold syndrome	AD
139	*ERCC5*	13q33.1	Cockayne syndrome, Cerebro-oculo-facio-skeletal syndrome	AR
140	*COL4A1*	13q34	Porencephaly	AD
141	*AP4S1*	14q12	Autosomal recessive ID with spastic paraplegia	AR
142	*MGAT2*	14q22.1	Congenital disorder of glycosylation, type iia	AR
143	*GCH1*	14q22.2-q22.3	Tetrahydrobiopterin-deficient hyperphenylalaninemia B	AR
144	*POMT2*	14q24.3	Walker-Warburg syndrome	AR
145	*GALC*	14q31.3	Krabbe disease	AR
146	*ZC3H14*	14q31.3	Autosomal recessive non-syndromic ID	AR
147	*TTC8*	14q31.3	Bardet-Biedl syndrome	AR
148	*VRK1*	14q32.2	Pontocerebellar hypoplasia type 1	AR
149	*DYNC1H1*	14q32.31	Severe ID with neuronal migration defects; Charcot-Marie-Tooth disease, axonal, type 20	AD
150	*SLC12A6*	15q14	Andermann syndrome	AR
151	*CEP152*	15q21.1	Primary microcephaly	AR
152	*AP4E1*	15q21.2	Spastic paraplegia 51, autosomal recessive	AR
153	*BBS4*	15q24.1	Bardet-Biedl syndrome	AR
154	*KIF7*	15q26.1	Joubert syndrome 12	AR
155	*IDH2*	15q26.1	D-2-hydroxyglutaric aciduria 2	AD
156	*GNPTG*	16p13.3	Mucolipidosis III gamma	AR
157	*TBC1D24*	16p13.3	Autosomal recessive syndrome of focal epilepsy, dysarthria, and ID	AR
158	*ZNF423*	16q12.1	Joubert syndrome 19 (dominant); nephronophthisis (recessive)	AD/AR
159	*TUBB3*	16q34.3	Cortical dysplasia, complex, with other brain malformations	AD
160	*PMM2*	16p13.2	Congenital disorder of glycosylation, type Ia	AR
161	*GRIN2A*	16p13.2	Variable neurodevelopmental phenotypes, including ID and/or epilepsy	AD
162	*NDE1*	16p13.11	Lissencephaly	AR
163	*BBS2*	16q13	Bardet-Biedl syndrome	AR
164	*GPR56*	16q13	Autosomal recessive bilateral frontoparietal polymicrogyria	AR
165	*COG8*	16q22.1	Congenital disorder of glycosylation, type iih	AR
166	*TMEM231*	16q23.1	Joubert syndrome 20	AR
167	*CDH15*	16q24.3	Autosomal dominant non-syndromic ID	AD
168	*CHMP1A*	16q24.3	Pontocerebellar hypoplasia, microcephaly, ID	AR
169	*WDR81*	17p13.3	Cerebellar hypoplasia, quadrupedal locomotion and ID	AR
170	*MPDU1*	17p13.1	Congenital disorder of glycosylation, type If	AR
171	*PIGL*	17p11.2	CHIME syndrome (colobomas, heart defects, ichthyosiform dermatosis, mental retardation, and ear anomalies, including conductive hearing loss)	AR
172	*ALDH3A2*	17p11.2	Sjogren-Larsson syndrome	AR
173	*SLC46A1*	17q11.2	Folate malabsorption	AR
174	*SMARCE1*	17q21.2	Coffin-Siris syndrome	AD
175	*EFTUD2*	17q21.31	Mandibulofacial dysostosis with microcephaly	AD
176	*GFAP*	17q21.31	Alexander disease	AD
177	*MKS1*	17q22	Bardet-Biedl syndrome, Meckel syndrome	AR
178	*COG1*	17q25.1	Congenital disorder of glycosylation, type iig	AR
179	*TSEN54*	17q25.1	Pontocerebellar hypoplasia type 2A	AR
180	*ACTG1*	17q25.3	Baraitser-Winter syndrome	AD
181	*RBBP8*	18q11.2	Seckel syndrome, Jawad syndrome	AR
182	*IER3IP1*	18q21.1	Microcephaly with simplified gyration, epilepsy, and infantile diabetes	AR
183	*RTTN*	18q22.2	Polymicrogyria, ID	AR
184	*MAP2K2*	19p13.3	Cardio-facio-cutaneous syndrome	AD
185	*MCOLN1*	19p13.2	Mucolipidosis IV	AR
186	*SMARCA4*	19p13.2	Coffin-Siris syndrome	AD
187	*CC2D1A*	19p13.12	Autosomal recessive non-syndromic ID	AR
188	*GPSN2*	19p13.12	Autosomal recessive non-syndromic ID	AR
189	*WDR62*	19q13.12	Severe brain malformations, including microcephaly, pachygyria and hypoplasia of the corpus callosum	AR
190	*BCKDHA*	19q13.2	Maple syrup urine disease, type Ia	AR
191	*ERCC2*	19q13.32	Cockayne syndrome, Trichothiodystrophy, Cerebro-oculo-facio-skeletal syndrome	AR
192	*ERCC1*	19q13.32	Cerebro-oculo-facio-skeletal syndrome	AR

	Gene	Chr. band	Disorder	Inheritance pattern
193	*FKRP*	19q13.32	Congenital muscular dystrophy 1C, limb-girdle muscular dystrophy type 2I, muscle-eye-brain disease, Walker-Warburg syndrome	AR
194	*PNKP*	19q13.33	Microcephaly, seizures and defects in DNA repair	AR
195	*ASXL1*	20q11.21	Bohring-Opitz syndrome	AD
196	*DNMT3B*	20q11.2	Immunodeficiency-centromeric instability-facial anomalies syndrome 1	AR
197	*CTSA*	20q13.12	Galactosialidosis	AR
198	*ARFGEF2*	20q13.13	Autosomal recessive periventricular heterotopia with microcephaly	AR
199	*DPM1*	20q13.13	Congenital disorder of glycosylation, type Ie	AR
200	*CBS*	21q22.3	Homocystinuria	AR
201	*PCNT*	21q22.3	Seckel syndrome, Majewski osteodysplastic primordial dwarfism type II	AR
202	*SNAP29*	22q11.21	Cerebral dysgenesis, neuropathy, ichthyosis, and palmoplantar keratoderma syndrome	AR
203	*SMARCB1*	22q11.23	Coffin-Siris syndrome	AD
204	*LARGE*	22q12.3	Congenital muscular dystrophy	AR
205	*EP300*	22q13.2	Rubinstein-Taybi syndrome	AD
206	*CYB5R3*	22q13.2	Methemoglobinemia type II	AR
207	*ALG12*	22q13.33	Congenital disorder of glycosylation, type Ig	AR
208	*CHKB*	22q13.33	Congenital muscular dystrophy, mitochondrial structural abnormalities and ID	AR
209	*HCCS*	Xp22.2	Microphthalmia with linear skin defects syndrome	XLD
210	*OFD1*	Xp22.2	Oral-facial-digital syndrome type I (XL dominant), Simpson-Golabi-Behmel syndrome, type 2 (XL recessive), Joubert syndrome 10	XLD/XLR
211	*FANCB*	Xp22.2	VACTERL with hydrocephalus, Fanconi anemia of complementation group B	XLR
212	*PDHA1*	Xp22.12	Pyruvate decarboxylase deficiency	XLD
213	*RPS6KA3*	Xp22.12	Coffin-Lowry syndrome, non-syndromic ID	XLD
214	*MBTPS2*	Xp22.12	Ichthyosis follicularis, atrichia and photophobia syndrome	XLR
215	*SMS*	Xp22.11	X-linked ID, Snyder-Robinson type	XLR
216	*GK*	Xp21.2	Glycerol kinase deficiency	XLD
217	*TSPAN7*	Xp11.4	Non-syndromic X-linked ID	XLR
218	*BCOR*	Xp11.4	Syndromic Lenz microphthalmia-2, oculofaciocardiodental syndrome	XLD
219	*ATP6AP2*	Xp11.4	X-linked ID with epilepsy	XLR
220	*MAOA*	Xp11.3	Brunner syndrome (Monoamine oxidase A deficiency)	XLR
221	*PORCN*	Xp11.23	Focal dermal hypoplasia	XLD
222	*SYP*	Xp11.23	Non-syndromic X-linked ID	XLD
223	*SHROOM4*	Xp11.22	Stocco dos Santos X-linked ID syndrome, non-syndromic X-linked ID	XLR
224	*HUWE1*	Xp11.22	Non-syndromic X-linked ID (duplications, XL recessive); syndromic X-linked ID, Turner type (point mutations, XL dominant)	XLD/XLR
225	*ARHGEF9*	Xq11.1	Syndromic X-linked ID, hyperekplexia and epilepsy	XLR
226	*DLG3*	Xq13.1	Non-syndromic X-linked ID	XLR
227	*HDAC8*	Xq13.1	Cornelia de Lange syndrome; Wilson-Turner syndrome	XLD
228	*SLC16A2*	Xq13.2	T3 transporter deficiency; syndromic and non-syndromic ID	XLD
229	*ATP7A*	Xq21.1	Menkes disease, occipital horn syndrome	XLR
230	*PGK1*	Xq21.1	Phosphoglycerate kinase-1 deficiency	XLR
231	*BRWD3*	Xq21.1	Non-syndromic X-linked ID	XLR
232	*ZNF711*	Xq21.1	Non-syndromic X-linked ID	XLR
233	*TIMM8A*	Xq22.1	Mohr-Tranebjaerg syndrome, Jensen syndrome	XLR
234	*RAB40AL*	Xq22.1	Syndromic X-linked ID, Martin-Probst type	XLR
235	*PLP1*	Xq22.2	Pelizaeus-Merzbacher disease	XLR
236	*PRPS1*	Xq22.3	Phosphoribosylpyrophosphate synthetase superactivity, Arts syndrome	XLR
237	*PAK3*	Xq22.3	Non-syndromic X-linked ID	XLR
238	*UBE2A*	Xq24	Syndromic X-linked ID, Nascimento-type	XLR
239	*NDUFA1*	Xq24	Mitochondrial complex I deficiency (syndromic X-linked ID)	XLD
240	*CUL4B*	Xq24	Syndromic X-linked ID, Cabezas type	XLR
241	*ZDHHC9*	Xq25	Syndromic X-linked ID, Raymond type; non-syndromic X-linked ID	XLD
242	*GPC3*	Xq26.2	Simpson-Golabi-Behmel syndrome type 1	XLR
243	*HPRT1*	Xq26.2	Lesch-Nyhan syndrome	XLR
244	*SOX3*	Xq27.1	Isolated GH deficiency, short stature and ID	XLR
245	*IDS*	Xq28	Mucopolysaccharidosis II (Hunter syndrome)	XLR
246	*NSDHL*	Xq28	CK syndrome (XL recessive); CHILD syndrome (Congenital Hemidysplasia with Ichthyosiform Nevus and Limb Defects) (XL dominant)	XLR/XLD
247	*ABCD1*	Xq28	Adrenoleukodystrophy	XLD
248	*HCFC1*	Xq28	Non-syndromic X-linked ID	XLR
249	*FLNA*	Xq28	Bilateral periventricular nodular heterotopia, otopalatodigital syndrome, frontometaphyseal dysplasia	XLD
250	*GDI1*	Xq28	Non-syndromic X-linked ID	XLD
251	*IKBKG*	Xq28	Incontinentia pigmenti	XLD
252	*DKC1*	Xq28	Dyskeratosis congenita	XLR

Note that all the ASD genes listed in **Table S6A** are also involved in ID.
Abbreviations: AD, autosomal dominant; AR, autosomal recessive; ID, intellectual disability; XL, X-linked; XLD, X-linked dominant; XLR, X-linked recessive

Table S6D. Loci implicated in ID

	Disorder	Chr. band	Start (hg18)	End (hg18)	Genes involved	Inheritance pattern
1	Thrombocytopenia-absent radius (TAR) syndrome	1q21.1	144,110,432	144,305,571	*RBM8A*	AR
2	3p deletion syndrome	3p26.3p25.3	1	9,000,000	contiguous gene syndrome?	AD
3	Proximal 15q25.2 microdeletion	15q25	80,451,495	82,719,635	?	AD
4	ATR-16 syndrome (alpha-thalassemia/mental retardation syndrome)	16p13.3	30,843	774,373	?	AD
5	17p13.1 microdeletion syndrome	17p13.1	7,429,371	7,937,620	?	AD
6	19p13.13 microdeletion/microduplication syndrome	19p13.13	12,793,474	13,104,643	?	AD
7	Cat-Eye syndrome	22p13-22q11.21	15,772,953	16,971,860	?	AD
8	22q11.2 distal deletion syndrome	22q11.2	20,445,848	22,026,229	?	AD
9	Pelizaeus-Merzbacher disease	Xq22.2	102,918,095	102,934,203	*PLP1*	XLR

Note that all the ASD loci listed in **Table S6B** are also involved in ID.
Abbreviations: AD, autosomal dominant; AR, autosomal recessive; ID, intellectual disability; XLR, X-linked recessive

Table S7A. CNVs overlapping ASD or ID genes and loci in affected and control subjects (all ancestries)

Cases

AGP ID	Sex	Chr#:start-end (hg18)	Size	CNV type	Inheri-tance	Classification	Validation	CNV description
16035_1571013001	M	1:144482933-146325557	1,842,625	Gain	*dn*	Pathogenic	qPCR	1q21.1 duplication syndrome
8635_201	M	1:144500467-146377870	1,877,404	Loss	*dn*	Pathogenic	qPCR	1q21.1 deletion syndrome
1952_301	M	1:144500467-146336720	1,836,254	Gain	*dn*	Pathogenic	qPCR	1q21.1 duplication syndrome
13135_1523	F	1:144838594-146308287	1,469,694	Gain	*dn*	Pathogenic	qPCR	1q21.1 duplication syndrome
4291_1	M	1:144967972-146317915	1,349,944	Gain	mat	Pathogenic	qPCR	1q21.1 duplication syndrome
16074_1571042001	M	2:50037898-50407550	369,653	Loss	mat	Pathogenic	qPCR	*NRXN1* exonic deletion
14068_1180	M	2:50493827-50677835	184,009	Gain	*dn*	Pathogenic	qPCR	*NRXN1* intragenic duplication, predicted to result in a premature truncated protein
13017_223	F	2:50539877-50730546	190,670	Loss	*dn*	Pathogenic	qPCR	*NRXN1* exonic deletion
13216_2383	M	2:50968208-51214171	245,964	Loss	*dn*	Pathogenic	qPCR	*NRXN1* exonic deletion
13153_1703	M	2:50990306-51222043	231,738	Loss	*dn*	Pathogenic	qPCR	*NRXN1* exonic deletion
13037_463	M	2:51002576-51157742	155,167	Loss	*dn*	Pathogenic	qPCR	*NRXN1* exonic deletion
17027_1	M	2:51076611-51147600	70,990	Loss	pat	Pathogenic	qPCR	*NRXN1* exonic deletion
5353_3	F	6:33399849-33512042	112,194	Loss	*dn*	Pathogenic	qPCR	*SYNGAP1* exonic deletion
5386_3	M	6:156785155-158489874	1,704,720	Loss	*dn*	Pathogenic	qPCR, Illumina 1M	*ARID1B* exonic deletion
8446_201	M	7:72344426-73782113	1,437,688	Loss	*dn*	Pathogenic	qPCR	Williams syndrome (7q11.23 deletion)
13123_1403	F	9:98998-3682923	3,583,926	Loss	*dn*	Pathogenic	qPCR	Terminal 9p deletion, 3.58 Mb (14 genes)
6259_3	M	9:139516033-140208462	692,430	Loss	*dn*	Pathogenic	qPCR	Kleefstra syndrome (9q34.3 deletion including *EHMT1*)
6325_3	M	11:70077507-70506315	428,809	Loss	*dn*	Pathogenic	qPCR	*SHANK2* exonic deletion
6319_3	M	11:70119917-70187872	67,956	Loss	*dn*	Pathogenic	qPCR	*SHANK2* exonic deletion
5237_3	M	11:70154458-70220632	66,175	Loss	*dn*	Pathogenic	qPCR, Agilent 1M	*SHANK2* exonic deletion
6240_4	M	11:126633939-132060374	5,426,436	Loss	*dn*	Pathogenic	qPCR	Chromosome 11q deletion syndrome (Jacobsen syndrome)
20187_1464001	M	15:18811937-26209270	7,397,334	Gain	*dn*	Pathogenic	qPCR	15q11-q13 duplication syndrome, maternally derived
8630_201	M	15:19800798-26209270	6,408,473	Gain	*dn*	Pathogenic	qPCR	15q11-q13 duplication syndrome, paternally derived
20069_1328001	M	15:20203578-26209270	6,005,693	Gain	*dn*	Pathogenic	qPCR	15q11-q13 duplication syndrome, maternally derived
17035_1	F	15:20274130-26120360	5,846,231	Gain	*dn*	Pathogenic	qPCR	15q11-q13 duplication syndrome, maternally derived
8117_202	M	15:21168391-26217954	5,049,564	Gain	mat	Pathogenic	qPCR	15q11-q13 duplication syndrome, maternally derived
8741_201	F	15:21168391-26315093	5,146,703	Gain	mat	Pathogenic	qPCR	15q11-q13 duplication syndrome, maternally derived
13050_593	M	15:21190624-26203954	5,013,331	Gain	*dn*	Pathogenic	qPCR	15q11-q13 duplication syndrome, maternally derived
1950_301	M	15:26762141-30436163	3,674,023	Loss	mat	Pathogenic	Illumina 550K	15q13.3 deletion syndrome
16040_1571029001	M	15:28450423-30303265	1,852,843	Loss	*dn*	Pathogenic	qPCR	15q13.3 deletion syndrome
14167_2720	M	15:28705540-30436163	1,730,624	Loss	pat	Pathogenic	qPCR	15q13.3 deletion syndrome
18100_302	M	15:28714502-30303265	1,588,764	Loss	pat	Pathogenic	qPCR	15q13.3 deletion syndrome
5537_3	M	15:82722026-83529838	807,813	Loss	mat	Pathogenic	Affy 500K, Illumina 2.5M, qPCR	Distal 15q25.2 deletion syndrome
14283_4060	M	16:14771033-16307313	1,536,281	Loss	mat	Pathogenic	qPCR	16p13.11 deletion syndrome
3441_3	M	16:14808871-16215852	1,406,982	Loss	pat	Pathogenic	qPCR	16p13.11 deletion syndrome
14412_5210	M	16:14960247-16307313	1,347,067	Loss	mat	Pathogenic	qPCR, LR-PCR	16p13.11 deletion syndrome
2204_1	M	16:29466569-30147029	680,461	Loss	*dn*	Pathogenic	qPCR	16p11.2 deletion syndrome
3544_3	M	16:29499858-30107306	607,449	Gain	mat	Pathogenic	—	16p11.2 duplication syndrome
5068_3	F	16:29502984-30127026	624,043	Loss	*dn*	Pathogenic	qPCR, Affy 500K, Agilent 1M	16p11.2 deletion syndrome, 50% mosaicism
20089_1391001	M	16:29502984-30107306	604,323	Loss	*dn*	Pathogenic	qPCR	16p11.2 deletion syndrome
3211_3	M	16:29502984-30127026	624,043	Gain	mat	Pathogenic	qPCR	16p11.2 duplication syndrome
5262_4	M	16:29502984-30210849	707,866	Gain	*dn*	Pathogenic	qPCR, Affy 500K	16p11.2 duplication syndrome
5359_4	M	16:29554843-30195224	640,382	Loss	*dn*	Pathogenic	qPCR, Affy 500K, Agilent 1M	16p11.2 deletion syndrome
20127_4014001	M	16:29554843-30130862	576,020	Loss	pat	Pathogenic	qPCR	16p11.2 deletion syndrome
4030_1	M	16:29554843-30107306	552,464	Gain	*dn*	Pathogenic	qPCR	16p11.2 duplication syndrome
3439_3	M	17:17156307-18262979	1,106,673	Loss	*dn*	Pathogenic	qPCR	Smith-Magenis syndrome (17p11.2 deletion including *RAI1*)
2211_1	F	17:17169258-20101517	2,932,260	Loss	*dn*	Pathogenic	qPCR	Smith-Magenis syndrome (17p11.2 deletion including *RAI1*)
14315_4320	M	17:31621634-33323919	1,702,286	Gain	mat	Pathogenic	qPCR	17q12 duplication syndrome
3183_7	M	22:17241748-19819918	2,578,171	Loss	*dn*	Pathogenic	qPCR	22q11.2 deletion syndrome
17015_1	M	22:17257787-19795780	2,537,994	Loss	*dn*	Pathogenic	qPCR	22q11.2 deletion syndrome
3127_4	M	22:17257787-19793730	2,535,944	Gain	pat	Pathogenic	qPCR	22q11.2 duplication syndrome
4271_1	M	22:17257787-19795780	2,537,994	Gain	*dn*	Pathogenic	qPCR	22q11.2 duplication syndrome
5261_4	F	22:17257787-19795780	2,537,994	Gain	pat	Pathogenic	qPCR, Illumina 1M	22q11.2 duplication syndrome
16074_1571042001	M	22:17257787-19793730	2,535,944	Gain	pat	Pathogenic	qPCR	22q11.2 duplication syndrome
8627_201	M	22:17257787-19793730	2,535,944	Gain	*dn*	Pathogenic	qPCR	22q11.2 duplication syndrome
2072_1	M	22:45159185-49582267	4,423,083	Loss	*dn*	Pathogenic	qPCR	Phelan-McDermid syndrome (22q13 deletion including *SHANK3*)
6130_4	F	22:47996161-49512530	1,516,370	Loss	*dn*	Pathogenic	qPCR, MLPA	Phelan-McDermid syndrome (22q13 deletion including *SHANK3*)
16079_1571066001	M	22:49470371-49567383	97,013	Loss	*dn*	Pathogenic	qPCR	Phelan-McDermid syndrome (22q13 deletion including *SHANK3*)
5240_4	M	X:23116188-23280628	164,441	Loss	mat	Pathogenic	qPCR, Illumina 1M	*PTCHD1* exonic deletion
5126_4	M	X:28931559-29478966	547,408	Gain	mat	Pathogenic	qPCR, Agilent 1M	*IL1RAPL1* intragenic duplication of exons 3-5
8597_201	M	X:31303978-32025062	721,085	Loss	mat	Pathogenic	qPCR	*DMD* deletion of exons 45-60
3019_3	M	X:32100618-32315937	215,320	Gain	mat	Pathogenic	qPCR	*DMD* duplication of exons 31-44
20013_1075001	M	X:153239048-153521797	282,750	Gain	*dn*	Pathogenic	qPCR, LR-PCR	Xq28 duplication encompassing 20 genes, including 3 involved in ID: *FLNA, GDI1, IKBKG*. Corresponds to the recurrent Xq28 duplication reported in XLID families;[55] *GDI1* is the most likely candidate gene (sequencing coordinates: chrX:153222048-153514311, size 292 kb)
14216_3470	M	X:153263157-153474401	211,245	Gain	mat	Pathogenic	qPCR	Xq28 duplication encompassing 18 genes, including 2 involved in ID: *GDI1* and *IKBKG.* Recurrent Xq28 duplication involved in XLID (see above)
6356_5	M	1:2118508-2325536	207,029	Loss	*dn*	Uncertain	qPCR	Deletion encompassing 5 genes, including *SKI,* involved in Shprintzen-Goldberg syndrome through dominant negative mutations
6317_5	M	1:26767083-26941756	174,674	Gain	pat	Uncertain	qPCR	*ARID1A* partial duplication; *ARID1A* mutations reported recently in Coffin-Siris syndrome
13027_353	F	2:144837809-145315383	477,575	Gain	*dn*	Uncertain	qPCR	*ZEB2* duplication; mutations and deletions in *ZEB2* cause Mowat-Wilson syndrome

AGP ID	Sex	Chr#:start-end (hg18)	Size	CNV type	Inheri-tance	Classification	Validation	CNV description
3424_3	M	2:148881443-149078468	197,026	Gain	mat	Uncertain	qPCR	*MBD5* partial duplication; *MBD5* is implicated in autosomal dominant ID through deletions (2q23.1 deletion syndrome) and mutations
3599_3	F	6:155906594-157336808	1,430,215	Gain	pat	Uncertain-likely benign	qPCR	*ARID1B* partial duplication; haploinsufficiency of *ARID1B* causes ID and Coffin-Siris syndrome
13037_463	M	9:137682721-137840339	157,619	Loss	mat	Uncertain	—	Deletion encompassing 4 genes, including *KCNT1*; gain-of-function *KCNT1* mutations cause epilepsy
16072_1571036001	M	11:381049-1019320	638,272	Gain	pat	Uncertain	qPCR	Duplication encompassing 35 genes, including *HRAS* (involved in Costello syndrome through activating mutations); although duplication of *HRAS* is not expected to be pathogenic, the contribution of other genes in the interval is unknown
20070_1331001	M	16:14808871-15935225	1,126,355	Gain	mat	Uncertain	qPCR	16p13.11 microduplication, significance unknown. Recurrent 16p13.11 deletions are associated with variable phenotype and incomplete penetrance, and have been reported in subjects with diverse neuropsychiatric disorders, including ID, ASD, schizophrenia and epilepsy, sometimes inherited from unaffected parents. Duplications have been described in neurodevelopmental disorders and in controls, with studies reporting either no enrichment or a small but significant enrichment in cases. Further studies in larger case populations and controls are needed to clarify their role as risk factors
9766_202	M	16:15032942-16199484	1,166,543	Gain	mat	Uncertain	qPCR	16p13.11 microduplication (see above)
14142_2400	M	16:15387380-16256106	868,727	Gain	pat	Uncertain	qPCR	16p13.11 microduplication (see above)
5258_3	M	16:15387380-16270740	883,361	Gain	pat	Uncertain	qPCR, Illumina 1M	16p13.11 microduplication (see above)
4182_1	M	16:15387380-16199484	812,105	Gain	mat	Uncertain	—	16p13.11 microduplication (see above)
2265_1	M	16:15387380-18176669	2,789,290	Gain	mat	Uncertain	qPCR	16p13.11 microduplication (see above)
8703_201	M	17:1092256-1249222	156,967	Gain	pat	Uncertain-likely benign	qPCR	Duplication of 3 genes, partially overlapping *YWHAE*; whole duplications are pathogenic (17p13.3 duplication syndrome); there is an AGP control with a *YWHAE* partial duplication; several partial duplications reported in DGV
5444_3	M	17:76953064-77782267	829,204	Gain	*dn*	Uncertain	qPCR	829 kb duplication encompassing 38 genes, including *ACTG1* (involved in Baraitser-Winter syndrome through dominant-negative or gain-of-function missense mutations); although duplication of *ACTG1* is not expected to be pathogenic, the contribution of the other genes in the interval is unknown
14242_3660	F	22:21995356-22598120	602,765	Gain	mat	Uncertain	—	602 kb duplication encompassing 15 genes, including *SMARCB1*; *SMARCB1* mutations reported recently in Coffin-Siris syndrome
6381_3	M	X:30521738-30789831	268,094	Gain	mat	Uncertain	qPCR	Duplication of 2 genes, partially overlapping *GK*; *GK* mutations and deletions cause XLID
5126_4	M	X:32948977-33330592	381,616	Gain	mat	Uncertain	Agilent 1M	*DMD* duplication of exon 1 of the Dp427 transcripts; the effect of this duplication is difficult to predict; experimental evidence at the RNA level is required to interpret the significance
4356_1	M	X:38001148-38346471	345,324	Gain	mat	Uncertain-likely benign	—	Duplication of *RPGR, OTC,* and *TSPAN7*; similar duplication found in an AGP male control. *OTC* mutations and deletions cause ornithine transcarbamylase deficiency; *TSPAN7* duplications are found in healthy controls
4152_1	M	X:40260354-40372806	112,453	Gain	mat	Uncertain	—	Duplication encompassing 3 genes, including *ATP6AP2*; *ATP6AP2* silent mutation affecting splicing described in XLID with epilepsy
3240_3	M	X:44706702-44919064	212,363	Gain	mat	Uncertain	—	Duplication of 2 genes, partially overlapping *KDM6A* partial duplication; *KDM6A* mutations and deletions cause X-linked Kabuki syndrome
14314_4310	M	X:70865247-71509736	644,490	Gain	mat	Uncertain	qPCR	644 kb duplication encompassing 13 genes, partially overlapping*HDAC8*; *HDAC8* mutations described recently in Cornelia de Lange syndrome and XLID resembling Wilson-Turner syndrome; similar partial duplication of *HDAC8* observed in a male control
1348_301	M	X:147163528-147758700	595,173	Gain	mat	Uncertain	qPCR	*AFF2* partial duplication; involved in XLID through trinucleotide expansion or deletion
5036_4	M	X:148075334-148617551	542,218	Gain	mat	Uncertain-likely benign	—	542 kb duplication encompassing 9 genes, including *IDS*; mutations and deletions cause mucopolysaccharidosis type II; duplication reported in unaffected males. Although duplication of *IDS* is not expected to be pathogenic, the contribution of the other genes in the interval is unknown. The duplication is also present in the proband's unaffected brother, suggesting this CNV is unlikely to be pathogenic
4354_1	M	X:148344051-148707925	363,875	Gain	mat	Uncertain-likely benign	—	363 kb duplication encompassing 9 genes, including *IDS*; mutations and deletions cause mucopolysaccharidosis type II; duplication reported in unaffected males (see above)
4166_1	M	X:151620401-151805387	184,987	Gain	mat	Uncertain	—	Xq28 duplication of 10 genes, including *NSDHL*; mutations and deletions cause syndromic XLID (CK syndrome and CHILD syndrome); similar duplication including *NSDHL* in a male AGP control
9901_201	M	3:12608293-12786824	178,532	Gain	pat	Benign	—	Duplication of 2 genes, partially overlapping *RAF1*; *RAF1* is involved in Noonan syndrome through activating mutations; similar duplication present in 2 AGP controls
20087_1386001	M	3:12610706-12786824	176,119	Gain	mat	Benign	—	Duplication of 2 genes, partially overlapping *RAF1*; *RAF1* is involved in Noonan syndrome through activating mutations; similar duplication present in 2 AGP controls
4457_1	M	X:38375788-38515190	139,403	Gain	mat	Benign	—	*TSPAN7* partial duplication; mutations and deletions cause XLID, duplications are found in healthy controls
5524_3	M	X:38375788-38515190	139,403	Gain	mat	Benign	—	*TSPAN7* partial duplication (see above)
6034_3	M	X:38375788-38515190	139,403	Gain	mat	Benign	—	*TSPAN7* partial duplication (see above)
20033_1227001	M	X:38375788-38515190	139,403	Gain	mat	Benign	—	*TSPAN7* partial duplication (see above)
20141_1396001	M	X:38375788-38515190	139,403	Gain	mat	Benign	qPCR	*TSPAN7* partial duplication (see above)
5089_5	M	X:53555568-53640902	85,335	Gain	mat	Benign	—	*HUWE1* partial duplication. Whole gene duplications cause non-syndromic XLID; similar recurrent partial duplications of *HUWE1* reported recently, considered polymorphic[133]
17018_1	M	X:53568262-53640902	72,641	Gain	mat	Benign	qPCR	*HUWE1* partial duplication (see above)

Controls

ID	Sex	Chr#:start-end (hg18)	Size	CNV type	Classification	CNV description
HABC_902399_902399	F	1:144614719-146470277	1,855,559	Loss	"Pathogenic"	1q21.1 deletion syndrome
B436528_1007852654	M	1:144627859-146546371	1,918,513	Loss	"Pathogenic"	1q21.1 deletion syndrome
B984152_1007842480	F	1:144800611-145863421	1,062,811	Gain	"Pathogenic"	1q21.1 duplication syndrome
B618929_1007875266	M	1:144933825-145518117	584,293	Gain	"Pathogenic"	1q21.1 duplication syndrome
B666224_1007871687	M	1:144967972-145863421	895,450	Gain	"Pathogenic"	1q21.1 duplication syndrome
HABC_902895_902895	F	2:50829989-51064129	234,141	Loss	"Pathogenic"	*NRXN1* exonic deletion
B964957_1007872180	F	7:72344426-73782113	1,437,688	Gain	"Pathogenic"	7q11.23 duplication syndrome (Williams syndrome region)
B914224_1007874975	M	10:89620404-89723400	102,997	Loss	"Pathogenic"	*PTEN* exonic deletion
HABC_902475_902475	M	15:36401817-36434987	33,171	Loss	"Pathogenic"	*SPRED1* deletion; gene involved in Legius syndrome (phenotypic overlap with neurofibromatosis 1); most individuals don't have ID and only present dermatologic findings
B675955_1007841005	F	16:29415871-30239704	823,834	Gain	"Pathogenic"	16p11.2 duplication syndrome
B416484_1007875540	M	16:29502984-30127026	624,043	Loss	"Pathogenic"	16p11.2 deletion syndrome
B879700_1007854073	F	16:29502984-30127026	624,043	Gain	"Pathogenic"	16p11.2 duplication syndrome
HABC_900681_900681	M	16:29554843-30180288	625,446	Gain	"Pathogenic"	16p11.2 duplication syndrome

ID	Sex	Chr#:start-end (hg18)	Size	CNV type	Classification	CNV description
B121881_1007874637	M	17:31621634-33323919	1,702,286	Gain	"Pathogenic"	17q12 duplication syndrome
HABC_901636_901636	M	22:17248170-19795780	2,547,611	Gain	"Pathogenic"	22q11.2 duplication syndrome
B928258_1007854097	F	22:17257787-18693299	1,435,513	Gain	"Pathogenic"	22q11.2 duplication syndrome
110036016178_	M	2:148753482-148819104	65,623	Loss	Uncertain	*MBD5* exonic deletion; deletions of *MBD5* cause autosomal dominant ID, but this one overlaps only the long isoform, which has not been fully characterized and contains 5 additional non-coding exons (this deletion overlaps 1 of the non-coding exons)
B183736_1007853714	F	8:117910382-117948637	38,256	Gain	Uncertain-likely benign	*RAD21* partial duplication; mutations and deletions cause Cornelia de Lange syndrome
HABC_900405_900405	M	8:117910403-117948935	38,533	Gain	Uncertain-likely benign	*RAD21* partial duplication; mutations and deletions cause Cornelia de Lange syndrome
HABC_900854_900854	M	9:2018757-2080718	61,962	Loss	Uncertain-likely benign	*SMARCA2* deletion of exons 2-19, removes the translation start site in exon 2; mutations resulting in Nicolaides-Baraitser syndrome are thought to act through a dominant–negative or gain-of-function manner and cluster in exons 15–25; deletions encompassing *SMARCA2* do not cause this syndrome, except for one reported in-frame deletion overlapping the mutation-clustering region. The deletion in the AGP control involves exons 2-19 and removes the translation start site in exon 2.
B116679_1007853952	M	9:129475725-129875601	399,877	Gain	Uncertain	Duplication encompassing 19 genes, partially overlapping *STXBP1*; mutations and deletions of *STXBP1* cause nonsyndromic ID with epilepsy and infantile epileptic encephalopathy
B246752_1007872634	F	9:139616009-139725155	109,147	Gain	Uncertain	Duplication encompassing 3 genes, partially overlapping *EHMT1*; *EHMT1* mutations and deletions cause Kleefstra syndrome
B252606_1007874475	M	12:1889583-2538831	649,249	Gain	Uncertain	Duplication encompassing 4 genes, partially overlapping *CACNA1C*, involved in Timothy syndrome through activating mutations
B936611_1007853579	F	12:25087733-25286046	198,314	Gain	Uncertain	Duplication encompassing 4 genes, partially overlapping *KRAS*; *KRAS* is involved in cardio-facio-cutaneous syndrome through activating mutations
B978305_1007874920	M	13:109744730-110078003	333,274	Gain	Uncertain	Duplication encompassing 4 genes, partially overlapping *COL4A1*; only *COL4A1* missense mutations reported thus far, no deletions or duplications
B777599_1007853701	F	14:101564897-101608061	43165	Gain	Uncertain	*DYNC1H1* partial duplication; all mutations identified so far (ID with neuronal migration defects and motor neuropathies) are heterozygous missense mutations, suggesting a dominant-negative effect
HABC_900744_900744	F	15:28723577-30232287	1,508,711	Gain	Uncertain	15q13.3 microduplication
HABC_901557_901557	F	15:28730804-30389965	1,659,162	Gain	Uncertain	15q13.3 microduplication
B833125_0057060983	F	16:1995854-2052977	57,124	Gain	Uncertain	Duplication encompassing 5 genes, partially overlapping *TSC2*; *TSC2* mutations and deletions cause tuberous sclerosis
HABC_902940_902940	M	16:14831165-16199484	1,368,320	Gain	Uncertain	16p13.11 microduplication
HABC_901197_901197	F	16:14882793-16199484	1,316,692	Gain	Uncertain	16p13.11 microduplication
HABC_902864_902864	M	16:14977368-16190572	1,213,205	Gain	Uncertain	16p13.11 microduplication
HABC_902897_902897	M	16:15387380-18075924	2,688,545	Gain	Uncertain	16p13.11 microduplication
HABC_901863_901863	F	16:87658525-87896679	238,155	Gain	Uncertain	*CDH15* duplication and *ANKRD11* partial duplication; *CDH15* is involved in non syndromic ID through mutations and *ANKRD11* in KBG syndrome through mutations and deletions
HABC_901829_901829	F	16:87729306-87962759	233,454	Gain	Uncertain	*CDH15* duplication and *ANKRD11* partial duplication (see above)
HABC_902145_902145	F	17:1136077-1211814	75,738	Gain	Uncertain-likely benign	Duplication of 2 genes, partially overlapping *YWHAE*; *YWHAE* whole gene duplications are pathogenic; several partial duplications in DGV
HABC_902046_902046	M	17:2437871-2476338	38,468	Gain	Uncertain	*PAFAH1B1* partial duplication; whole gene duplications are pathogenic
HABC_901640_901640	M	22:21320275-23360745	2,040,471	Gain	Uncertain	2 Mb duplication encompassing 44 genes, including *SMARCB1*; *SMARCB1* mutations reported in Coffin-Siris syndrome
B278753_1007874641	M	22:21995356-22676385	681,030	Gain	Uncertain	681 kb duplication encompassing 20 genes, including *SMARCB1*; *SMARCB1* mutations reported in Coffin-Siris syndrome
HABC_900402_900402	F	X:11038333-11069582	31,250	Gain	Uncertain-likely benign	Duplication of 2 genes, including *HCCS*; *HCCS* mutations and deletions cause a syndromic form of XLID
HABC_902971_902971	F	X:13633067-13700254	67,188	Gain	Uncertain-likely benign	Duplication encompassing 4 genes, including *OFD1*; mutations and deletions cause syndromic XLID
HABC_901634_901634	M	X:17239813-17435795	195,983	Gain	Uncertain-likely benign	*NHS* partial duplication; mutations and deletions cause Nance-Horan syndrome
HABC_902647_902647	M	X:18574793-18780863	206,071	Gain	Uncertain-likely benign	Xp22.13 duplication encompassing 3 genes, partially overlapping *CDKL5*; *CDKL5* mutations and deletions cause epileptic encephalopathy
HABC_900416_900416	M	X:38013482-38643203	629,722	Gain	Uncertain-likely benign	Duplication encompassing 4 genes, including *OTC* and *TSPAN7*. *OTC* mutations and deletions cause ornithine transcarbamylase deficiency; no cases have been reported with a duplication. *TSPAN7* mutations and deletions cause XLID, duplications are found in healthy controls
B345605_1007844543	M	X:71239825-71490721	250,897	Gain	Uncertain-likely benign	Duplication encompassing 8 genes, partially overlapping *HDAC8*; *HDAC8* mutations described recently in Cornelia de Lange syndrome and XLID
HABC_900366_900366	F	X:76924341-77030430	106,090	Gain	Uncertain-likely benign	Partial duplication of *ATRX* and *MAGT1*; *ATRX* is involved in XLID through mutations or deletions
HABC_900333_900333	F	X:134903813-135264655	360,843	Gain	Uncertain	Xq26.3 duplication encompassing 4 genes, partially overlapping *SLC9A6*; mutations and deletions cause syndromic XLID (Christianson syndrome)
B418695_1007840289	M	X:151644548-151839695	195,148	Gain	Uncertain	Xq28 duplication encompassing 10 genes, including *NSDHL*; *NSDHL* mutations and deletions cause XLID (CK syndrome and CHILD syndrome); several overlapping duplications reported in DGV
HABC_902725_902725	F	X:153340432-153435070	94,639	Gain	Uncertain-likely benign	95 kb duplication encompassing 7 genes, partially overlapping *IKBKG*; only mutations, deletions and intragenic duplications described
HABC_902313_902313	M	X:153440007-153622054	182,048	Gain	Uncertain-likely benign	182 kb duplication encompassing 7 genes, partially overlapping *IKBKG*; only mutations, deletions and intragenic duplications described
HABC_901807_901807	M	3:12608293-12781123	172,831	Gain	Benign	Duplication encompassing 2 genes, partially overlapping *RAF1*; *RAF1* is involved in Noonan syndrome through activating mutations
B260038_1007841400	F	3:12610706-12781123	170,418	Gain	Benign	Duplication encompassing 2 genes, partially overlapping *RAF1*; *RAF1* is involved in Noonan syndrome through activating mutations
B630497_1007872229	F	12:2085709-2127756	42,048	Loss	Benign	*CACNA1C* deletion; *CACNA1C* causes Timothy syndrome through activating mutations
B131548_1007842659	F	12:2663790-2714138	50,349	Gain	Benign	*CACNA1C* partial duplication; *CACNA1C* causes Timothy syndrome through activating mutations
HABC_902939_902939	F	12:2675893-2717980	42,088	Gain	Benign	*CACNA1C* partial duplication; *CACNA1C* causes Timothy syndrome through activating mutations
110036016517_	M	X:38375788-38515190	139,403	Gain	Benign	*TSPAN7* partial duplication; mutations and deletions cause XLID, duplications are found in healthy controls
B818627_1007854359	M	X:38375788-38515190	139,403	Gain	Benign	*TSPAN7* partial duplication; mutations and deletions cause XLID, duplications are found in healthy controls

This table shows the CNVs overlapping ASD/ID genes and loci interpreted as pathogenic, uncertain or benign in cases and controls of all ancestries. Phenotype information and CNV segregation in siblings can be found in **Table S8**. In order to compare the burden of CNVs overlapping ASD/ID genes and loci in cases and controls (**Figures 1** and **S1**), CNVs were interpreted irrespective of affected status. A small number of CNVs that would have been considered pathogenic in an affected individual were identified among controls; all were CNVs known to be associated with incomplete penetrance/variable expressivity.

Abbreviations: DGV, Database of Genomic Variants; dn, *de novo*; F, female; ID, intellectual disability; LR-PCR, long range PCR; M, male; mat, maternal; MLPA, multiplex ligation-dependent probe amplification; pat, paternal; qPCR, quantitative PCR; XLID, X-linked intellectual disability. —, no validation attempted, CNV confirmed by visual inspection.

Table S7B. Pathogenic CNVs in affected subjects (all ancestries)

Pathogenic CNVs overlapping ASD/ID genes or loci (stringent CNV, >30 kb) n = 64

AGP ID	Sex	Family type	Cytoband	Chr#:start-end (hg18)	Size	CNV type	Inheritance	CNV description	Classification penetrance/ expressivity[§]
16035_1571013001	M	Familial	1q21.1	1:144482933-146325557	1,842,625	Gain	*De novo*	1q21.1 duplication syndrome	VE/IP
1952_301	M	Familial	1q21.1	1:144500467-146336720	1,836,254	Gain	*De novo*	1q21.1 duplication syndrome	VE/IP
8635_201	M	Familial	1q21.1	1:144500467-146377870	1,877,404	Loss	*De novo*	1q21.1 deletion syndrome	VE/IP
13135_1523	F	Unknown	1q21.1	1:144838594-146308287	1,469,694	Gain	*De novo*	1q21.1 duplication syndrome	VE/IP
4291_1	M	Familial	1q21.1	1:144967972-146317915	1,349,944	Gain	Maternal	1q21.1 duplication syndrome	VE/IP
16074_1571042001†	M	Sporadic	2p16.3	2:50037898-50407550	369,653	Loss	Maternal	*NRXN1* exonic deletion	VE/IP
14068_1180	M	Sporadic	2p16.3	2:50493827-50677835	184,009	Gain	*De novo*	*NRXN1* intragenic duplication, predicted to result in a premature truncated protein	VE/IP
13017_223	F	Sporadic	2p16.3	2:50539877-50730546	190,670	Loss	*De novo*	*NRXN1* exonic deletion	VE/IP
13216_2383	M	Unknown	2p16.3	2:50968208-51214171	245,964	Loss	*De novo*	*NRXN1* exonic deletion	VE/IP
13153_1703	M	Sporadic	2p16.3	2:50990306-51222043	231,738	Loss	*De novo*	*NRXN1* exonic deletion	VE/IP
13037_463	M	Sporadic	2p16.3	2:51002576-51157742	155,167	Loss	*De novo*	*NRXN1* exonic deletion	VE/IP
17027_1	M	Familial	2p16.3	2:51076611-51147600	70,990	Loss	Paternal	*NRXN1* exonic deletion	VE/IP
5353_3	F	Sporadic	6p21.32	6:33399849-33512042	112,194	Loss	*De novo*	*SYNGAP1* exonic deletion	HP
5386_3	M	Familial	6q25.3	6:156785155-158489874	1,704,720	Loss	*De novo*	*ARID1B* exonic deletion	HP
8446_201	M	Sporadic	7q11.23	7:72344426-73782113	1,437,688	Loss	*De novo*	Williams syndrome (7q11.23 deletion)	HP
13123_1403	F	Sporadic	9p24.3-p24.2	9:98998-3682923	3,583,926	Loss	*De novo*	Terminal 9p deletion, 3.58 Mb (14 genes)	HP
6259_3	M	Sporadic	9q34.3	9:139516033-140208462	692,430	Loss	*De novo*	Kleefstra syndrome (9q34.3 deletion including *EHMT1*)	HP
6325_3	M	Sporadic	11q13.3	11:70077507-70506315	428,809	Loss	*De novo*	*SHANK2* exonic deletion[129]	HP
6319_3	M	Sporadic	11q13.3	11:70119917-70187872	67,956	Loss	*De novo*	*SHANK2* exonic deletion[117,129]	HP
5237_3	M	Sporadic	11q13.3-q13.4	11:70154458-70220632	66,175	Loss	*De novo*	*SHANK2* exonic deletion[117,134]	HP
6240_4	M	Sporadic	11q24.2-q25	11:126633939-132060374	5,426,436	Loss	*De novo*	Chromosome 11q deletion syndrome (Jacobsen syndrome)	HP
20187_1464001	M	Sporadic	15q11.2-q13.1	15:18811937-26209270	7,397,334	Gain	*De novo*	15q11-q13 duplication syndrome, maternally derived	VE/IP
8630_201	M	Familial	15q11.2-q13.1	15:19800798-26209270	6,408,473	Gain	*De novo*	15q11-q13 duplication syndrome, paternally derived	VE/IP
20069_1328001	M	Sporadic	15q11.2-q13.1	15:20203578-26209270	6,005,693	Gain	*De novo*	15q11-q13 duplication syndrome, maternally derived	VE/IP
17035_1	F	Sporadic	15q11.2-q13.1	15:20274130-26120360	5,846,231	Gain	*De novo*	15q11-q13 duplication syndrome, maternally derived	VE/IP
8117_202	M	Familial	15q11.2-q13.1	15:21168391-26217954	5,049,564	Gain	Maternal	15q11-q13 duplication syndrome, maternally derived	VE/IP
8741_201	F	Familial	15q11.2-q13.1	15:21168391-26315093	5,146,703	Gain	Maternal	15q11-q13 duplication syndrome, maternally derived	VE/IP
13050_593	M	Sporadic	15q11.2-q13.1	15:21190624-26203954	5,013,331	Gain	*De novo*	15q11-q13 duplication syndrome, maternally derived	VE/IP
1950_301	M	Familial	15q13.1-q13.3	15:26762141-30436163	3,674,023	Loss	Maternal	15q13.3 deletion syndrome	VE/IP
16040_1571029001	M	Familial	15q13.2-q13.3	15:28450423-30303265	1,852,843	Loss	*De novo*	15q13.3 deletion syndrome	VE/IP
14167_2720	M	Sporadic	15q13.2-q13.3	15:28705540-30436163	1,730,624	Loss	Paternal	15q13.3 deletion syndrome	VE/IP
18100_302	M	Familial	15q13.2-q13.3	15:28714502-30303265	1,588,764	Loss	Paternal	15q13.3 deletion syndrome	VE/IP
5537_3	M	Familial	15q13.2-q13.3	15:82722026-83529838	807,813	Loss	Maternal	Distal 15q25.2 deletion syndrome	VE/IP
14283_4060	M	Sporadic	16p13.11	16:14771033-16307313	1,536,281	Loss	Maternal	16p13.11 deletion syndrome	VE/IP
3441_3	M	Sporadic	16p13.11	16:14808871-16215852	1,406,982	Loss	Paternal	16p13.11 deletion syndrome	VE/IP
14412_5210	M	Sporadic	16p13.11	16:14960247-16307313	1,347,067	Loss	Maternal	16p13.11 deletion syndrome	VE/IP
2204_1	M	Sporadic	16p11.2	16:29466569-30147029	680,461	Loss	*De novo*	16p11.2 deletion syndrome	VE/IP
3544_3	M	Familial	16p11.2	16:29499858-30107306	607,449	Gain	Maternal	16p11.2 duplication syndrome	VE/IP
20089_1391001	M	Sporadic	16p11.2	16:29502984-30107306	604,323	Loss	*De novo*	16p11.2 deletion syndrome	VE/IP
5068_3	F	Familial	16p11.2	16:29502984-30127026	624,043	Loss	*De novo*	16p11.2 deletion syndrome, 50% mosaicism	VE/IP
5262_4	M	Sporadic	16p11.2	16:29502984-30210849	707,866	Gain	*De novo*	16p11.2 duplication syndrome	VE/IP
3211_3	M	Familial	16p11.2	16:29502984-30127026	624,043	Gain	Maternal	16p11.2 duplication syndrome	VE/IP
5359_4	M	Sporadic	16p11.2	16:29554843-30195224	640,382	Loss	*De novo*	16p11.2 deletion syndrome	VE/IP
20127_4014001	M	Sporadic	16p11.2	16:29554843-30130862	576,020	Loss	Paternal	16p11.2 deletion syndrome	VE/IP
4030_1	M	Familial	16p11.2	16:29554843-30107306	552,464	Gain	*De novo*	16p11.2 duplication syndrome	VE/IP
3439_3	M	Sporadic	17p11.2	17:17156307-18262979	1,106,673	Loss	*De novo*	Smith-Magenis syndrome (17p11.2 deletion including *RAI1*)	HP
2211_1	F	Sporadic	17p11.2	17:17169258-20101517	2,932,260	Loss	*De novo*	Smith-Magenis syndrome (17p11.2 deletion including *RAI1*)	HP
14315_4320	M	Sporadic	17q12	17:31621634-33323919	1,702,286	Gain	Maternal	17q12 duplication syndrome	VE/IP
3183_7	M	Familial	22q11.21	22:17241748-19819918	2,578,171	Loss	*De novo*	22q11.2 deletion syndrome	VE/IP
17015_1	M	Sporadic	22q11.21	22:17257787-19795780	2,537,994	Loss	*De novo*	22q11.2 deletion syndrome	VE/IP
8627_201	M	Sporadic	22q11.21	22:17257787-19793730	2,535,944	Gain	*De novo*	22q11.2 duplication syndrome	VE/IP
4271_1	M	Sporadic	22q11.21	22:17257787-19795780	2,537,994	Gain	*De novo*	22q11.2 duplication syndrome	VE/IP
16074_1571042001†	M	Sporadic	22q11.21	22:17257787-19793730	2,535,944	Gain	Paternal	22q11.2 duplication syndrome	VE/IP
3127_4	M	Familial	22q11.21	22:17257787-19793730	2,535,944	Gain	Paternal	22q11.2 duplication syndrome	VE/IP
5261_4	F	Familial	22q11.21	22:17257787-19795780	2,537,994	Gain	Paternal	22q11.2 duplication syndrome	VE/IP
2072_1	M	Familial	22q13.31-q13.33	22:45159185-49582267	4,423,083	Loss	*De novo*	Phelan-McDermid syndrome (22q13 deletion including *SHANK3*)[130,135]	HP
6130_4	F	Sporadic	22q13.32-q13.33	22:47996161-49512530	1,516,370	Loss	*De novo*	Phelan-McDermid syndrome (22q13 deletion including *SHANK3*)	HP
16079_1571066001	M	Sporadic	22q13.33	22:49470371-49567383	97,013	Loss	*De novo*	Phelan-McDermid syndrome (22q13 deletion including *SHANK3*)	HP
5240_4	M	Familial	Xp22.11	X:23116188-23280628	164,441	Loss	Maternal	*PTCHD1* exonic deletion[136,137]	— (sex chr)
5126_4	M	Familial	Xp21.3	X:28931559-29478966	547,408	Gain	Maternal	*IL1RAPL1* intragenic duplication of exons 3-5	— (sex chr)
8597_201	M	Sporadic	Xp21.2-p21.1	X:31303978-32025062	721,085	Loss	Maternal	*DMD* deletion of exons 45-60	— (sex chr)
3019_3	M	Familial	Xp21.1	X:32100618-32315937	215,320	Gain	Maternal	*DMD* duplication of exons 31-44 (sequencing coordinates: chrX:32097213-32321945, size: 224733 bp)[138]	— (sex chr)
20013_1075001	M	Sporadic	Xq28	X:153239048-153521797	282,750	Gain	*De novo*	Xq28 duplication encompassing 20 genes, including 3 involved in ID: *FLNA, GDI1*, *IKBKG*. Recurrent Xq28 duplication implicated in XLID; GDI1 is the most likely candidate gene[55] (sequencing coordinates: chrX:153,222,048-153,514,311, size 292 kb)	— (sex chr)
14216_3470	M	Sporadic	Xq28	X:153263157-153474401	211,245	Gain	Maternal	Xq28 duplication encompassing 18 genes, including 2 involved in ID: *GDI1* & *IKBKG.* Recurrent Xq28 duplication implicated in XLID; *GDI1* is the most likely candidate gene	— (sex chr)

Other pathogenic CNVs, including large *de novo* CNVs as well as CNVs not included in the main analyses (chromosomal abnormalities, CNVs <30 kb, non stringent CNVs or CNVs identified with 1 algorithm only and validated), n = 20

AGP ID	Sex	Family type	Cytoband	Chr#:start-end (hg18)	Size	CNV type	Inheritance	CNV description	Classification penetrance/ expressivity§
8658_201	F	Sporadic	1p21.3-p21.2	1:98175622-100923952	2,748,331	Loss	*De novo*	2.7 Mb 1p21.3-p21.2 deletion, *de novo* (18 genes, including *DPYD* and *MIR137*). 1p21.3 microdeletions comprising *DPYD* and *MIR137* have been reported in subjects with ID	— (novel region)
5467_3	M	Sporadic	1q42.3-q44	1:233476547-247165725	13,689,179	Gain	*De novo*	13.7 Mb 1q42.3-q44 duplication, *de novo* (114 genes). This CNV overlaps the critical region of the 1q43-q44 deletion syndrome. Distal duplications of the long arm of chromosome 1 are rare, reported in few individuals in its pure form (chromosome abnormality)	HP
5236_3	F	Familial	2p16.3	2:50705521-50719594	14,074	Loss	Maternal	*NRXN1* exonic deletion (CNV is exonic according to RefSeq, but not to UCSC) (<30 kb, non stringent: QSNP|PCNV)	VE/IP
5328_3	M	Sporadic	2p16.3	2:51044181-51120644	76,464	Loss	Paternal	*NRXN1* exonic deletion (non stringent: QSNP|PCNV)	VE/IP
16037_1571015001	M	Sporadic	2q37.3	2:239765200-239777909	12,710	Loss	Maternal	*HDAC4* small intragenic deletion. *HDAC4* is responsible for some of the features of the 2q37 deletion syndrome (brachydactyly mental retardation syndrome). There are many 2q37 deletions reported in individuals with ASD, but no mutations or single gene deletions of *HDAC4* had been reported in ASD. (sequencing coordinates: chr2:239766528-239778481, size 11954 bp) (<30 kb)	VE/IP
21020_1‡	M	Sporadic	4p16.3-p16.1	4:53403-9016339	8,962,937	Gain	*De novo*	*De novo* unbalanced translocation leading to 4p16.3-p16.1 duplication and 8p23.3-p23.1 deletion. The 4p duplication spans 8.9 Mb and involves the region implicated in Wolf-Hirschhorn syndrome. Normal 46,XY karyotype, unbalanced translocation shown by FISH (chromosome abnormality)	HP
21020_1‡	M	Sporadic	8p23.3-p23.1	8:154984-6994825	6,839,842	Loss	*De novo*	*De novo* unbalanced translocation leading to 4p16.3-p16.1 duplication and 8p23.3-p23.1 deletion (see above). The 8p terminal deletion spans 6.8 Mb; numerous terminal deletions of 8p have been described in the literature in subjects with ID, behavioral issues and mild dysmorphic features (chromosome abnormality)	HP
14270_3930	F	Sporadic	6q25.3-q27	6:160773919-170761395	9,987,477	Gain	*De novo*	10 Mb duplication 6q25.3-q27, *de novo*, 46,XX.ish der(22)t(6;22)(6q25.3;p11.2)pat(6qtel+). "6q duplication syndrome" is a rare cytogenetic abnormality reported in few individuals in its pure form; the duplications can affect any part of the 6q arm and few overlap. Most of the reported cases are the result of abnormal segregation of a balanced translocation carried by a parent, like in this case (chromosome abnormality)	HP
8404_201	M	Familial	6q27	6:169136788-170761395	1,624,608	Loss	*De novo*	1.6 Mb *de novo* deletion in 6q27 region (14 genes). Terminal 6q deletion syndrome. The CNV contains *DLL1*, an ASD candidate gene: *de novo* frameshift variant identified in ASD exome study[4]	HP
13137_1543	F	Sporadic	8p12-8q12.1	8:31928590-58996070	27,067,481	Gain	*De novo*	27 Mb pericentromeric duplication 8p12-8q12.1, *de novo*. The karyotype revealed a mosaic supernumerary ring chromosome (47, XX, +r[10]/46, XX[70]) of unknown origin, shown to involve chr 8 by SNP array. Trisomy 8 syndrome is characterized by mild to severe mental and growth deficiency, facial dysmorphisms, and limb abnormalities. In cases of supernumerary marker chromosomes or supernumerary ring chromosomes derived from chr 8, mosaic or non-mosaic, the clinical presentation varies from normal phenotype to features overlapping the trisomy 8 syndrome. (chromosome abnormality)	HP
8534_201	M	Sporadic	10q11.21-q11.23	10:45633089-51564756	5,931,668	Loss	*De novo*	5.9 Mb 10q11.22-q11.23 deletion (56 genes), *de novo*. Recurrent deletions in this region are associated with variable clinical features, with ID as the only feature present in the majority of individuals; most deletions are inherited from apparently normal parents, indicating variable expressivity/incomplete penetrance[139]	VE/IP
4312_1	M	Sporadic	10q11.21-q11.23	10:45550419-51496386	5,945,968	Loss	Paternal	5.9 Mb 10q11.22-q11.23 deletion (56 genes), paternal. Recurrent deletion (see above)	VE/IP
6053_3	M	Familial	12q13.3-q14.1	12:54218922-58779615	4,560,694	Gain	*De novo*	4.5 Mb *de novo* duplication in 12q13 (101 genes); no similar CNV reported	— (novel region)
14070_1230	M	Familial	15q26.1	15:91200007-91283004	82,998	Loss	*De novo*	*CHD2* exonic deletion. *CHD2 de novo* mutations reported recently in epileptic encephalopathy and ID; several deletions reported in individuals with ID, ASD and epilepsy	HP
5420_3*	M	Sporadic	Whole chr 21	21:1-247249719	247,249,719	Gain	*De novo*	Down syndrome (47,XY+21) (chromosome abnormality)	HP
14291_4120	F	Sporadic	22q13.33	22:49470371-49480446	10,076	Loss	*De novo*	*SHANK3* deletion of exons 9-13 (<30 kb, called with 1 algorithm only: PCNV) (sequencing coordinates: 49468716-49485255, size 16540 bp)	HP
5241_3	M	Familial	Xp21.1	X:31793278-31822704	29,427	Loss	Maternal	*DMD* deletion of exon 48 (predicted to lead to an in-frame deletion) (<30 kb)	— (sex chr)
9861_202	M	Sporadic	Xp11.4	X: 41248675-41259467	10,793	Gain	Maternal	*CASK* partial duplication of the 5' UTR of exon 1, decreased mRNA expression in cell line (<30 kb, called with 1 algorithm only: PCNV)	— (sex chr)
5257_3*	M	Sporadic	Whole chr Y	Y:1-57772954	57,772,954	Gain	*De novo*	XYY syndrome (chromosome abnormality)	— (sex chr)
5515_3*	M	Sporadic	Whole chr Y	Y:1-57772954	57,772,954	Gain	*De novo*	XYY syndrome (chromosome abnormality)	— (sex chr)

This table includes pathogenic CNVs overlapping ASD/ID genes or loci as well as other pathogenic CNVs in cases of all ancestries. The latter included chromosome abnormalities (>7.5 Mb) that had been excluded from the main analyses, selected large rare *de novo* events, as well as experimentally validated smaller CNVs (<30 kb) or CNVs not considered stringent (called by at least one algotihm only).

With the exception of the 16p11.2 duplication found in proband 3544_3 (note tested), all the CNVs reported in this table were experimentally validated (see details in **Table S8**).

§ For the analysis shown in **Figure 2E**, autosomal pathogenic CNVs were classified as highly penetrant (HP) or associated with variable expressivity and/or incomplete penetrance (VE/IP). CNVs on sex chromosomes or affecting novel regions were not classified for this analysis.

† Proband 16074_1571042001 carries two pathogenic CNV: a *NRXN1* exonic deletion inherited from his mother and a 22q11.2 duplication inherited from his father.

‡ Proband 21020_1 has a *de novo* unbalanced translocation leading to 4p duplication and 8p deletion, both considered pathogenic.

* Three cases showed whole chromosome aneuploidies: one case with Down syndrome and two with XYY syndrome.

Abbreviations: F, female; ID, intellectual disability; M, male; UTR, untranslated region; XLID, X-linked intellectual disability

Table S8. Phenotypes in ASD subjects with pathogenic CNVs or with selected CNVs of uncertain significance

mmc2.xlsx (Excel workbook)

This file also contains information on CNV validation and segregation in siblings, when available.

Table S9. Meta-analysis of loci and genes affected by rare CNVs in large ASD cohorts

Locus/gene	AGP, Stages 1+2 [a] Cases (1853 M, 294 F) Controls (1241 M, 1399 F)				SSC [b] Cases (968 M, 156 F) Controls (403 M, 469 F)				AGRE [c] Cases (n=1835)* Cases (n=837; 755 M)**		Combined AGP + SSC + AGRE Cases (2821 M, 450 F + 1835* or 837; 755 M**) Controls (1644 M, 1868 F)				Combined AGP+SSC+AGRE Del	Combined AGP+SSC+AGRE Dup	Del P-value†	Dup P-value†	Del Freq %	Dup Freq %
	Cases		Controls		Cases		Controls		Cases		Cases		Controls		Cases/Controls	Cases/Controls				
	Del	Dup	Del	Dup	Del	Dup	Del	Dup	Del	Dup	Del	Dup	Del	Dup						
16p11.2	5 (4)	4 (2)	–	2	8 (7)	6 (4)	–	–	3 (3)*	2 (0) *	16 (14)	12 (6)	–	2	16/5106; 0/3512	12/5106; 2/3512	2.28E-04	3.52E-02	0.313	0.235
15q13.3 (BP3-BP5, BP4-BP5)#	4 (1)	–	–	1	2 (1)	1 (1)	–	–	2 (0)*	1 (0)*	8 (2)	2 (1)	–	1	8/5106; 0/3512	2/5106; 1/3512	1.52E-02	0.637	0.157	0.039
16p13.11#	3 (0)	5 (0)	–	2	1 (0)	2 (1)	1	3	3 (0)*	4 (0; 1 unk)**	7 (0)	11 (1; 1 unk)	1	5	7/5106; 1/3512	11/4108; 5/3512	9.86E-02	0.174	0.137	0.268
15q11q13 (BP1-BP3, BP2-BP3)$	–	6 (5)	–	–	–	1 (1)	–	–	–	6 (2)*	–	13 (8)	–	–	0	13/5106; 0/3512	–	1.10E-03	–	0.255
22q11.2 (DiGeorge syndrome)	2 (2)	5 (2)	–	1	1 (1)	–	–	–	–	3 (1)*	3 (3)	8 (3)	–	1	3/5106; 0/3512	8/5106; 1/3512	0.208	6.46E-02	0.059	0.157
17p11.2 (Smith-Magenis syndrome)	2 (2)	–	–	–	–	–	–	–	–	–	2 (2)	–	–	–	2/4108**; 0/3512	0	0.291	–	0.049	–
9q34.3 (Kleefstra syndrome)	1 (1)	–	–	–	1 (1)	–	–	–	–	–	2 (2)	–	–	–	2/4108**; 0/3512	0	0.291	–	0.049	–
1q21.1	1 (1)	3 (2)	–	1	–	3 (2)	–	–	1 (1)*	2 (0)*	2 (2)	8 (4)	–	1	2/5106; 0/3512	8/5106; 1/3512	0.351	6.46E-02	0.039	0.157
17q12	–	1 (0)	–	–	2 (1)	–	–	–	–	–	2 (1)	1 (0)	–	–	2/5106; 0/3512	1/5106; 0/3512	0.351	0.593	0.039	0.020
7q11.23 (Williams syndrome)	1 (1)	–	–	–	–	4 (4)	–	–	–	–	1(1)	4 (4)	–	–	1/4108; 0/3512	4/4108; 0/3512	0.539	8.44E-02	0.024	0.097
																	–	–	–	–
PTCHD1/PTCHD1AS (Xp22.11)‡+	8 (1)	–	1	–	3 (1)	–	1	–	1 (0)**	–	12 (2)	–	2	–	12/3576**; 2/1644 M	0	0.133	–	0.336	–
NRXN1 (2p16.3)	6 (4)	1 (1)§	1	–	3 (1)	–	1	–	4 (0)**	–	13 (5)	1 (1)	2	–	13/4108**; 2/3512	1/4108; 0/3512	8.50E-03	0.539	0.316	0.024
SHANK3 (22q13.33)&	4 (4)	–	–	–	–	–	–	–	–	–	4 (4)	–	–	–	4/4108**; 0/3512	0	8.44E-02	–	0.097	–
SHANK2 (11q13.3)	3 (3)	–	–	–	–	–	–	–	–	1 (0)**	3 (3)	1 (0)	–	–	3/4108**; 0/3512	1/4108; 0/3512	0.157	0.539	0.073	0.024
NLGN3 (Xq13.1) ‡	–	–	–	–	1 (0)	–	–	–	–	–	1 (0)	–	–	–	1/3576**; 0/1644 M	0	0.685	–	0.028	–
NLGN4X (Xp22.3) ‡	–	–	–	–	–	1 (0)	–	–	–	1 (1 unk)**	–	2 (1 unk)	–	–	0	2/3576; 0/1644 M	–	0.469	–	0.056

CNVs are ordered according to frequency. Numbers in parentheses indicate *de novo* events. For *NRXN1*, *SHANK2*, *SHANK3*, *NLGN3* and *NLGN4X* CNVs are only counted in cases and controls if they affect one or more exons; chromosome X events are only counted for males.

*Number of AGRE cases in Moreno-de-Luca et al.[111] (n=1835 cases). A few of the regions in this table were not listed in Moreno-de-Luca et al.'s paper; we obtained the counts of subjects with deletions/duplications in these additional regions directly from the authors - the total number of inspected cases after quality control for the additional regions was 837 unrelated probands from 1105 families (or a total of 1472 all affected): 161 families with multiple males and females affected children, 594 families with only one male affected child and 82 families with only one female affected child. The total number of AGRE families with at least one male affected child was 161+594=755, and the total number of AGRE families with at least one female affected child was 161+82=243. The combined number of cases used to get estimates for this table was: i) 5106 = (AGP+SSC) + AGRE= 3271+1835 or ii) 4108 (**) = 3271+837.

† Fisher exact test, one-sided P-value

[#] Duplications of the 15q13.3 and 16p13.11 regions are of uncertain clinical significance, because they have not been found to be consistently enriched in cases compared to controls.

[$] Of the 13 duplications of the 15q11q13 region, 5 were maternally inherited (1 AGP, 4 AGRE), and 8 were *de novo*. Of the 8 *de novo* duplications, 5 were of maternal origin (4 AGP, 1 SSC), two were of paternal origin (AGP: 8630_201 and AGRE: AU1135202), and one without information (1 AGRE: AU1042303).

[§] *NRXN1* intragenic duplication, predicted to result in a premature truncated protein.

‡ Only males counted. The combined number of males used to get estimates for this table was 3576 = (AGP+SSC) + AGRE= 2821+755

+ The two *de novo* deletions in this locus were found in male probands from an SSC family (12561) and AGP family (subject 5240_4)

[&] Includes *SHANK3* gene exonic deletions as well as 22q13 deletion syndrome (Phelan-McDermid syndrome).

Abbreviations: AGP, Autism Genome Project; AGRE, Autism Genetic Resource Exchange; BP, breakpoint; Del, deletion; Dup, duplication; F, female; Freq, frequency; M, male; SSC, Simons Simplex Collection; unk, unknown.

[a] Ref [117] + Stage 2 (this study) (only subjects of European ancestry included)

[b] Ref [132]

[c] Ref [111]

Table S10. FMRP targets affected by deletions in probands and not yet implicated in ASD or ID

FMRP targets not yet implicated in ASD or ID	
ABR	*MAPK8IP3*
AGTPBP1	*MBP*
ARHGAP32	*NBEA*
BAI1	*NRXN3*
BRSK1	*PCDH9*
CAMSAP1	*PLCB1*
DTNA	*PLEC*
EML2	*PRKACB*
EP400	*PTPRD*
FAM115A	*PTPRT*
FAM21A	*R3HDM1*
FAM91A1	*RALGAPA1*
FAT4	*RALGDS*
GRM5	*RPRD2*
KCNH1	*SHANK1*
KIAA0430	*SPTB*
KIAA0913	*TCF25*
LLGL1	*TRIP12*
LPHN3	*TRPM3*
LPPR4	*ULK2*
LYNX1	*UNC13C*
MADD	

Of the FMRP targets[9] (n=842, one without correspondence in our annotation file) affected by exonic deletions in ASD subjects, 58 genes correspond to known ASD/ID genes (**Tables S6A-S6D**). This table lists the remaining 43 FMRP targets not yet implicated in ASD or ID, considered ASD candidate genes.

Tables S11A-S11E. Multigene analyses – various models

Table S11A. Primary analysis

	Estimate	Std. Error	z value	Pr(>|z|)
Model treating CNVs within the same individual as independent (pseudo R^2 0.007)				
(Intercept)	-0.195271	0.027722	-7.044	1.87e-12***
factor(CNV)Dup	-0.120601	0.039779	-3.032	0.00243**
ngene	0.030912	0.005768	5.359	8.37e-08***
RPKM	0.057893	0.024207	2.392	0.01677*
Model treating CNVs within the same individual as independent, and analyzing deletions only (pseudo R^2 0.007)				
(Intercept)	-0.21920	0.02979	-7.359	1.85e-13***
ngene	0.02857	0.01030	2.773	0.005549**
RPKM	0.14595	0.04153	3.515	0.000441***
Model treating CNVs within the same individual as independent, and analyzing duplications only (pseudo R^2 0.008)				
(Intercept)	-0.282964	0.036126	-7.833	4.78e-15***
ngene	0.031361	0.006981	4.492	7.04e-06***
RPKM	0.011110	0.029966	0.371	0.711
Model with the number of genes and the average RPKM nested within deletion or duplication status (pseudo R^2 0.008)				
(Intercept)	-0.219196	0.029785	-7.359	1.85e-13***
factor(CNV)Dup	-0.063768	0.046821	-1.362	0.173217
factor(CNV)Del:ngene	0.028566	0.010301	2.773	0.005549**
factor(CNV)Dup:ngene	0.031361	0.006981	4.492	7.04e-06***
factor(CNV)Del:RPKM	0.145950	0.041528	3.515	0.000441***
factor(CNV)Dup:RPKM	0.011110	0.029966	0.371	0.710813

Abbreviations: ngene, number of genes; RPKM, reads per kb per million reads

In our primary analysis we treated CNVs falling within the same individual as independent (main **Figures 3C-D** and **Table S11A**). **Figures 3C-D** show the pattern of increased burden with the increased number of brain-expressed genes affected (1-hit, 2-hit, 3-hit, 4- to 10-hit) by deletions or duplications. The percentage of cases and controls with CNVs overlapping genes is shown for deletions and duplications separately. To account for the fact that some individuals carry a multiplicity of rare CNVs, we also performed analyses counting all genes in all rare CNVs per individual, as well as the brain expression of those genes, and entered those in the model (**Table S11B)**; and we evaluated the impact of the largest CNV and its gene expression. Results from these alternative models were similar, showing that the dependence due to CNVs occurring in the same individual has only a minor impact on inference. For each model, Nagelkerke's pseudo-R^2 was calculated to provide a description of how well the model fitted the data.

In **Figure 3C-D** (main text), the expected odds ratios (depicted by stars) were estimated by fitting a logit model of the case status (case/control) with CNV type and three covariates, the deletion/duplication status, the number of genes covered by each CNV and their average brain expression value for the genes covered by the CNV. Gene level expression values for the neocortex were obtained from the BrainSpan RNA sequencing resource, transformed to log(1+RPKM), and the average of the transformed RPKM value for the genes covered by the CNV was used in the model (i.e. total RPKM of all genes covered by the CNV divided by the total number of genes within the CNV). Differences between the effect of the number of genes and expression values between deletion and duplication CNVs were further evaluated by analyzing each subset of the data separately.

In **Table S11A**, we fit a logit model with the following covariates: deletion/duplication status of the CNV, the number of genes covered by each CNV and the average of the transformed RPKM value (i.e. RPKM value transformed to log(1+RPKM)) for the genes covered by the CNV (total RPKM of all genes covered by the CNV divided by the total number of genes). We ignored the fact that individuals could have multiple CNVs. All effects in the model are significant. There is higher risk for autism when the number of genes covered by the CNV is larger and when the average (transformed) RPKM value is higher. Also notice that the risk for autism is slightly lower for duplications as opposed to deletions.

We also checked whether there were differences between the effect of the number of genes and RPKM between deletions and duplications by analyzing each subset of the data separately. Notice that the increased risk due to the number of covered genes for autism is fairly similar for duplications and deletions, while the increased risk due to an increase of the average RPKM is significant for deletions but not for duplications.

Another way to assess the differences between duplications and deletions would be to fit the effect of the number of genes and the average RPKM nested within deletion and duplication CNV status. The results are virtually the same for this model as the previous split data analysis.

Table S11B. Models accounting for the fact that some individuals carry a multiplicity of rare CNVs

	Estimate	Std. Error	z value	Pr(>|z|)
Model considering the total number of genes covered by all deletions or duplications in an individual (ngeneDel, ngeneDup) and the average RPKM value for the genes in deleted or duplicated regions (mRPKMDel, mRPKMDup) (pseudo R^2 0.016)				
(Intercept)	-0.37032	0.04326	-8.560	<2e-16***
ngeneDel	0.03292	0.01006	3.273	0.001066**
ngeneDup	0.03506	0.01010	3.471	0.000519***
mRPKMDel	0.17159	0.05014	3.422	0.000621***
mRPKMDup	0.01307	0.03796	0.344	0.730597
Model when selecting CNVs within individuals based on the maximum number of genes covered by the CNVs (pseudo R^2 0.018)				
(Intercept)	-0.365583	0.042510	-8.600	<2e-16***
maxGeneDel	0.035244	0.010939	3.222	0.001273**
maxGeneDup	0.027772	0.006957	3.992	6.55e-05***
RPKMDel	0.171068	0.048600	3.520	0.000432***
RPKMDup	-0.000702	0.036458	-0.019	0.984637
Model when selecting the CNVs based on the maximum average standardized RPKM (pseudo R^2 0.015)				
(Intercept)	-0.36725	0.04326	-8.490	<2e-16***
nGeneDel	0.03289	0.01173	2.805	0.00503**
nGeneDup	0.02543	0.00795	3.199	0.00138**
maxRPKMDel	0.17640	0.04392	4.016	5.91e-05***
maxRPKMDup	0.01741	0.03205	0.543	0.58705

Abbreviations: ngene, number of genes; RPKM, reads per kb per million reads

To take into account the effect of multiple CNVs per individual we decided to re-analyze the data using a model in which we fit the total number of genes covered by all deletions in an individual (ngeneDel), the total number of genes covered by duplications in an individual (ngeneDup) as well as the average RPKM value for these genes in a deletion region (mRPKMDel) and the average RPKM for those in a duplication region (mRPKMDup) **(upper panel)**. All effects, except the one for the average RPKM of genes covered by duplications (mRPKMDup) were significant. Notice the similarity of the results of this model with the one in which we analyzed the duplications and deletions separately. In both analyses the risk due to the number of genes is similar in duplications and deletions. However, only higher mean RPKM values for deletions increase risk for ASD.

Instead of looking at the average number of genes and average RPKM covered by deletions and duplications, one can also identify the CNVs for which the maximum value is obtained. This table shows the results of the generalized linear model when CNVs are selected within individuals based on the maximum number of genes covered by the CNVs **(middle panel)**, as well as the results obtained when selecting the CNVs based on the maximum average standardized RPKM **(lower panel)**. Notice the great similarity between these two analyses as well as the analysis in which we took the average of multiple CNVs per individual. In all three cases the effect of the number of genes covered by the CNVs is significant and the magnitude of the risk is very similar. Also, for all three analyses, the value of the average RPKM is only significant for deletions.

Table S11C. Removing cases with validated *de novo* CNVs

	Estimate	Std. Error	z value	Pr(>|z|)
Model considering the average number of genes and the average RPKM of genes covered by the CNVs in the remaining individuals (pseudo R^2 0.005)				
(Intercept)	-0.343950	0.043740	-7.863	3.74e-15***
ngeneDel	0.010159	0.011975	0.848	0.39625
ngeneDup	0.024238	0.010586	2.290	0.02204*
mRPKMDel	0.137438	0.051433	2.672	0.00754**
mRPKMDup	0.006056	0.038418	0.158	0.87474
Model considering the CNVs with the maximum number of genes covered, in the remaining individuals (pseudo R^2 0.006)				
(Intercept)	-0.343072	0.043122	-7.956	1.78e-15***
maxGeneDel	0.010173	0.013176	0.772	0.44005
maxGeneDup	0.020914	0.007687	2.721	0.00651**
RPKMDel	0.139397	0.049723	2.803	0.00506**
RPKMDup	-0.004947	0.036953	-0.134	0.89351
Model considering the CNVs with the maximum RPKM, in the remaining individuals (pseudo R^2 0.006)				
(Intercept)	-0.342025	0.043733	-7.821	5.25e-15***
nGeneDel	0.009579	0.014296	0.670	0.50285
nGeneDup	0.020248	0.008438	2.399	0.01642*
maxRPKMDel	0.131945	0.045620	2.892	0.00382**
maxRPKMDup	0.001850	0.032626	0.057	0.95477

Abbreviations: ngene, number of genes; RPKM, reads per kb per million reads

To assess the impact of validated *de novo* CNVs we removed 77 cases that had at least one validated *de novo* CNV. (Note that the total dataset contained 90 validated *de novo* CNVs in 87 cases of European ancestry; of these, 85 *de novo* CNVs in 82 unique European cases had sizes ≥30 kb; individuals with chromosomal abnormalities were excluded from the main analyses). Because of the nature of the control data, *de novo* CNV status was not available in the controls.

Results for the analysis of the average number of genes and the average RPKM of genes covered by the CNVs in the remaining individuals showed that the average number of genes covered by deletions is no longer significant **(upper panel)**. In this analysis, only the average number of genes covered by duplications and the average RPKM of genes covered by deletions are significant. A similar pattern of results is found when analyzing the CNVs with the maximum number of genes covered **(middle panel)**, and the CNVs selected based on the maximum RPKM **(lower panel)**. The results in this table show that removing only 77 out of 1,914 cases, or 4% of the case sample, results in a large decrease in signal, suggesting that most of the risk traces to *de novo* CNVs.

Table S11D. Removing subjects with CNVs considered pathogenic

	Estimate	Std. Error	z value	Pr(>|z|)
Model considering the average number of genes and the average RPKM of genes covered by the CNVs in the remaining individuals (pseudo R^2 0.004)				
(Intercept)	-0.329996	0.043973	-7.505	6.17e-14***
ngeneDel	0.006021	0.012452	0.484	0.62871
ngeneDup	0.015398	0.011366	1.355	0.17551
mRPKMDel	0.159530	0.051078	3.123	0.00179**
mRPKMDup	0.013352	0.038353	0.348	0.72774
Model considering the CNVs with the maximum number of genes covered, in the remaining individuals (pseudo R^2 0.005)				
(Intercept)	-0.326948	0.043399	-7.533	4.94e-14***
maxGeneDel	0.003602	0.013877	0.260	0.795200
maxGeneDup	0.015129	0.008447	1.791	0.073290 .
RPKMDel	0.164659	0.049298	3.340	0.000838***
RPKMDup	-0.001081	0.037003	-0.029	0.976694
Model considering the CNVs with the maximum RPKM, in the remaining individuals (pseudo R^2 0.005)				
(Intercept)	-0.330966	0.044004	-7.521	5.43e-14***
nGeneDel	0.004337	0.014825	0.293	0.769840
nGeneDup	0.011546	0.009555	1.208	0.226890
maxRPKMDel	0.156664	0.045120	3.472	0.000516***
maxRPKMDup	0.011904	0.032612	0.365	0.715084

Abbreviations: ngene, number of genes; RPKM, reads per kb per million reads

The AGP list of CNVs was curated to identify a subset that could be considered pathogenic (**Table S7B**). These CNVs were carried by 82 unique individuals, and of these 69 were of European ancestry and thus included in these analyses (missing individuals include other ancestries as well as individuals with chromosomal abnormalities [**Table S1C**], excluded from the main analyses).

After removing the individuals with pathogenic CNVs we reran the same three analyses as before. In general, these analyses show good agreement with the results from the comparable models obtained by removing the validated *de novo* CNVs only. This is not surprising because the two sets of events overlap fairly substantially. Here again it is shown that the value of RPKM for deletion CNVs is the most important irrespective of CNV status (i.e., even after removing both *de novo* and inherited pathogenic CNVs), significant in all three analyses. The magnitude of the risk associated with the deletion RPKMs is again very consistent to what was found in the previous analysis. Altogether these results are consistent with a genetic interaction model where imbalance of multiple genes intersected by rare *de novo* and inherited pathogenic CNVs contributes to risk. Furthermore, our findings are also in line with a previous report suggesting that deletions have larger effects on transcriptional level and contained more genes with altered expression compared to duplications.[140]

Table S11E. Model including sex

	Estimate	Std. Error	z value	Pr(>|z|)
Males (pseudo R^2 0.011)				
(Intercept)	0.23934	0.05445	4.396	1.1e-05***
nGeneDel	0.03734	0.01417	2.635	0.00842**
nGeneDup	0.03118	0.01300	2.398	0.01649*
maxRPKMDel	0.12490	0.06376	1.959	0.05010 .
maxRPKMDup	0.03293	0.04898	0.672	0.50133
Females (pseudo R^2 0.010)				
(Intercept)	-1.69525	0.09428	-17.980	<2e-16***
nGeneDel	0.04850	0.01853	2.617	0.00886**
nGeneDup	0.02393	0.01777	1.347	0.17803
maxRPKMDel	0.18177	0.10508	1.730	0.08366 .
maxRPKMDup	0.01009	0.07995	0.126	0.89953
Model selecting the CNV covering the largest number of genes in males (pseudo R^2 0.013)				
(Intercept)	0.237196	0.053568	4.428	9.51e-06***
maxGeneDel	0.033687	0.014671	2.296	0.02167*
maxGeneDup	0.027932	0.009389	2.975	0.00293**
RPKMDel	0.132569	0.061994	2.138	0.03248*
RPKMDup	0.027329	0.047174	0.579	0.56238
Model selecting the CNV covering the largest number of genes in females (pseudo R^2 0.012)				
(Intercept)	-1.68903	0.09260	-18.240	<2e-16***
maxGeneDel	0.05820	0.02053	2.835	0.00458**
maxGeneDup	0.01797	0.01216	1.477	0.13955
RPKMDel	0.19642	0.10003	1.964	0.04957*
RPKMDup	-0.01833	0.07737	-0.237	0.81267
Model selecting the CNV with the largest average standardized RPKM in males (pseudo R^2 0.011)				
(Intercept)	0.23799	0.05448	4.369	1.25e-05***
maxGeneDel	0.02668	0.01536	1.737	0.08239 .
maxGeneDup	0.02680	0.01086	2.467	0.01361*
maxRPKMDel	0.16081	0.05596	2.874	0.00406**
maxRPKMDup	0.03207	0.04131	0.776	0.43749
Model selecting the CNV with the largest average standardized RPKM in females (pseudo R^2 0.011)				
(Intercept)	-1.684587	0.094701	-17.788	<2e-16***
maxGeneDel	0.065273	0.021786	2.996	0.00273**
maxGeneDup	0.008598	0.017921	0.480	0.63137
maxRPKMDel	0.147739	0.094098	1.570	0.11640
maxRPKMDup	0.019179	0.068047	0.282	0.77806

Abbreviations: ngene, number of genes; RPKM, reads per kb per million reads

We ran the same three models, except this time the data were separated by sex: 1,641 cases and 1,102 controls were male, while 273 cases and 1,257 controls were female. There is general agreement in the parameter estimates between males and females for all three sets of analyses. We also ran tests for significant differences in predictors when they interacted with sex, but none of these interactions were significant (data not shown).

Tables S12A-S12D. GO terms, pathways, and MPO enrichment in affecteds versus control subjects

mmc3.xlsx (Excel workbook)

Table S13A. Gene Ontology terms and pathways used to generate a list of neurodevelopmental functions

Gs ID	Gs Name
GO:0007399	Nervous system development
GO:0050877	Neurological system process
GO:0043025	Neuronal cell body
GO:0043005	Neuron projection
KEGG:04725	Cholinergic synapse
KEGG:04724	Glutamatergic synapse
KEGG:04728	Dopaminergic synapse
KEGG:04727	GABAergic synapse
KEGG:04726	Serotonergic synapse
KEGG:04721	Synaptic vesicle cycle
KEGG:04723	Retrograde endocannabinoid signaling
KEGG:05030	Cocaine addiction
KEGG:05031	Amphetamine addiction
KEGG:05032	Morphine addiction
KEGG:05033	Nicotine addiction
KEGG:04722	Neurotrophin signaling pathway
REACT:708	REACT: Neuronal system
REACT:675	REACT: NCAM signaling for neurite out-growth
REACT:138	REACT: Axon guidance
NCI:142	NCI: netrin pathway
NCI:180	NCI: reelin pathway
KEGG:04360	Axon guidance
KEGG:04720	Long-term potentiation
KEGG:04730	Long-term depression

Table S13B. Effect size for neurobiological-related clusters

B1. FDR 15%

Logistic regression-FDR 15%[a]	Combined (% all subjects)	Stage 1	Stage 2
Cases	89 (4.70%)	57 (4.65%)	32 (4.79%)
Controls	35 (1.49%)	17 (1.37%)	18 (1.63%)
OR	3.15	3.39	2.94
Cases (max 10 genes/CNV)	56 (2.96%)	37 (3.02%)	19 (2.84%)
Controls (max 10 genes/CNV)	31 (1.32%)	14 (1.13%)	17 (1.54%)
OR (max 10 genes/CNV)	2.24	2.67	1.85

[a]REACT: Neuronal system; REACT: Transmission across chemical synapses; KEGG: Glutamatergic synapse; KEGG: Cholinergic synapse; GO: Generation of neurons; GO: Neuron projection morphogenesis; GO: Neuron differentiation; GO: Neuron development; GO: Cell morphogenesis involved in neuron differentiation; GO: Axonogenesis; GO: Neuron projection development; GO: Axon guidance.
FDR, false-discovery rate; OR, odds ratio; max-10 genes-CNVs, estimates considering CNVs with maximum 10 genes.

B2. FDR 20%

Logistic regression-FDR 20%[a]	Combined (% all subjects)	Stage 1	Stage 2
Cases	99 (5.23%)	64 (5.23%)	35 (5.24%)
Controls	49 (2.09%)	26 (2.10%)	23 (2.08%)
OR	2.50	3.39	2.51
Cases (max 10 genes/CNV)	66 (3.49%)	44 (3.59%)	22 (3.29%)
Controls (max 10 genes/CNV)	45 (1.92%)	23 (1.86%)	22 (1.99%)
OR (max 10 genes/CNV)	1.82	1.93	1.65

[a]REACT: Neuronal system; REACT: Transmission across chemical synapses; KEGG: Glutamatergic synapse; KEGG: Cholinergic synapse; GO: Generation of neurons; GO: Neuron projection morphogenesis; GO: Neuron differentiation; GO: Neuron development; GO: Cell morphogenesis involved in neuron differentiation; GO: Axonogenesis; GO: Neuron projection development; GO: Axon guidance; REACT: Neurotransmitter receptor binding and downstream transmission in the postsynaptic cell; KEGG: Retrograde endocannabinoid signaling; REACT: Axon guidance; KEGG: Dopaminergic synapse; KEGG: Neurotrophin signaling pathway; GO: Learning or memory.
FDR, false-discovery rate; OR, odds ratio; max-10 genes-CNVs, estimates considering CNVs with maximum 10 genes.

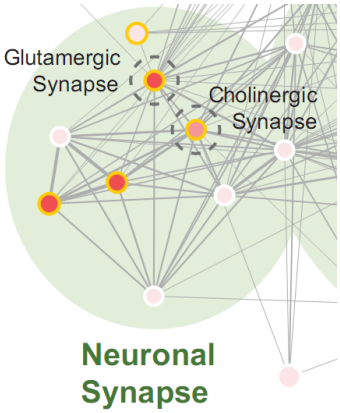
Table S13C. Neuronal synapse main cluster

C1. Statistics for all subjects with exonic CNVs (deletions and duplications) at FDR 15%

	Subject N (% all subjects)	Stage 1	Stage 2
Cases	36 (1.96%)	20 (1.63%)	16 (2.40%)
Controls	9 (0.38%)	5 (0.40%)	4 (0.36%)
OR [95% CI]	**4.93 [2.31-11.72]**	4.04	6.62

36 cases (25 supporting genes) *vs*. 9 controls (9 supporting genes). Stage 2 is more enriched, but Stage 1 also displays a respectable signal.

FDR, false-discovery rate; CI, confidence interval; OR: odds ratio



C2. Detailed statistics for pathways included within the Neuronal synapse cluster at FDR 15%

	All Cases	All Controls	Stage 1 Cases	Stage 1 Controls	Stage 2 Cases	Stage 2 Controls
KEGG Glutamatergic	19 (1.00%)	2 (0.09%)	10 (0.82%)	1 (0.08%)	9 (1.35%)	1 (0.09%)
KEGG Cholinergic	17 (0.90%)	2 (0.09%)	9 (0.74%)	2 (0.16%)	8 (1.20%)	0 (0.00%)
REACT Neuronal System	26 (1.37%)	6 (0.26%)	14 (1.14%)	3 (0.24%)	12 (1.80%)	3 (0.27%)
REACT Synaptic Transmission	21 (1.11%)	3 (0.13%)	11 (0.90%)	2 (0.16%)	10 (1.50%)	1 (0.09%)

Statistics for all subjects with exonic CNVs (deletions and duplications included). Calculation of % of subjects is based on subjects with at least one genic exonic loss; for max-10 this is limited to subjects with a maximum of 10 deleted genes.

Table S13D. Genes and subjects represented in the enriched cholinergic and glutamergic synapse subclusters

D1. Genes

	Cholinergic		Both		Glutamergic	
	Gene name	**# cases; # controls**	**Gene name**	**# cases; # controls**	**Gene name**	**# cases; # controls**
Neurotransmitters (receptors and metabolism)	*CHAT**	2; 0 (2*)	–	–	*GRIK2*	1; 0
	CHRNA7	4; 0	–	–	*GRM5*	1; 0
	KCNJ12	1; 0	–	–	–	–
	*SLC18A3 (VAChT)**	2; 0 (2*)	–	–	–	–
Scaffolds (synaptic organizers)	–	–	–	–	*SHANK1*	1; 0
	–	–	–	–	*SHANK2*	3; 0
	–	–	–	–	*SHANK3*	3; 0
Signalling (downstream neurotransmitter receptors and ion channels)	*CAMK2G***	1; 0 (1**)	*GNG2****	2; 0 (1***)	*PPP3CB***	1; 0 (1**)
	–	–	*GNG13*	1; 0	–	–
	–	–	*MAPK3****	5; 0 (1***)	–	–
	–	–	*PRKACB*	1; 0	–	–
	–	–	*PLCB1*	1; 0	–	–

26 cases (16 supporting genes) *vs.* 3 controls (3 genes). Numbers between parentheses represent the number of *de novo* events.
** Two genes within the same event/same sample (multigene); *** two genes in two different events in same sample (double-CNV-hit).

D2. ASD subjects

Sample ID	Gene(s) in CNV	CNV inheritance	Biological function	Functional sub-cluster
8534_201	*CHAT*; SLC18A3**	dn (5.9 Mb del)	Neurotransmitters (receptors and metabolism)	Cholinergic
4312_1	*CHAT*; SLC18A3**	pat (5.9 Mb del)		Cholinergic
8465_202	*GRIK2*	pat (39 kb del)		Glutamergic
8549_201	*GRM5*	pat (1.98 Mb del)		Glutamergic
3567_4	*KCNJ12*	pat (324 kb del)		Cholinergic
18100_302	*CHRNA7*	pat (1.6 Mb del)		Cholinergic
1950_301	*CHRNA7*	mat (1.7 Mb del)		Cholinergic
14167_2720	*CHRNA7*	pat (1.7 Mb del)		Cholinergic
16040_157102900	*CHRNA7*	dn (1.8 Mb del)		Cholinergic
14393_5020	*CAMK2G**; PPP3CB**	dn (477 kb del)	Signalling (downstream neurotransmitter receptors and ion channels)	Cholinergic; Glutamergic
13204_883	*GNG13*	unk[$] (81 kb del)		Both
2204_1	*GNG2****	mat (101 kb del)		Both
	*MAPK3****	dn (680 kb del)		Both
20057_1290002	*GNG2*	pat (102 kb del)		Both
20089_1391001	*MAPK3*	dn (680 kb del)		Both
20127_4014001	*MAPK3*	pat (680 kb del)		Both
5359_4	*MAPK3*	dn (680 kb del)		Both
5068_3	*MAPK3*	dn (680 kb del)		Both
5451_3	*PRKACB*	dn (80 kb del)		Both
5046_3	*PLCB1*	dn (30 kb del)		Both
5237_3	*SHANK2**	dn del	Scaffolds (synaptic organizers)	Glutamergic
6319_3	*SHANK2**	dn del		Glutamergic
6325_3	*SHANK2**	dn del		Glutamergic
16079_1571066001	*SHANK3**	dn del		Glutamergic
2072_1	*SHANK3**	dn del		Glutamergic
6130_4	*SHANK3**	dn del		Glutamergic
5340_3	*SHANK1*	mat del		Glutamergic

* Same gene in two or more samples (recurrent gene); ** two genes within the same deletion/same sample (multigene); *** two genes in two different events in same sample (double-CNV-hit).
[$] Both parental samples failed array QC.
del, deletion; dn, *de novo*; mat; maternal; pat, paternal; unk, inheritance unknown.

Tables S14A-S14E. Characterization of genes selected by NETBAG

mmc4.xlsx (Excel workbook)

Table S15. Functional-group enrichment for DAPPLE results

mmc5.*xlsx* (Excel workbook)

Table S16. List of 97 high-confidence CNV/SNV genes

Gene symbol	Type	Gene symbol	Type	Gene symbol	Type	Gene symbol	Type
ABCA1	CNV	*DTNA**	CNV	*MDM2*	LoF-SNV	*SKI*	CNV
ABL1	CNV	*DRP2*	LoF XL-SNV in males	*MED13L*	LoF-SNV	*SMARCC2*	LoF-SNV
ANK2	LoF-SNV	*DST*	LoF-SNV	*NCKAP1*	LoF-SNV	*SMC2*	CNV
ARHGAP32	CNV	*EPHB2*	LoF-SNV	*NFIA*	LoF-SNV	*SNRPN*	CNV
ARHGDIA	CNV	*ERCC6*	CNV	*NRXN3*	CNV	*SNX9*	CNV
ATOH1	CNV	*ESR2*	CNV	*PARD3*	CNV	*SOD2*	CNV
ATP1B1	LoF-SNV	*ETS1*	CNV	*PARK2*	CNV	*SPAST*	LoF-SNV
BAZ1B	CNV	*FLNA*	CNV	*PAX5*	LoF-SNV	*SPATA13*	LoF-SNV
BCL11A	LoF-SNV	*GARNL1*	CNV	*PIK3CB**	CNV	*SREBF1*	CNV
BRWD1	LoF-SNV	*HAUS7*	LoF XL-SNV in males	*PIR*	LoF XL-SNV in males	*STAU2*	CNV
CACNA1B	CNV	*IKBKG*	CNV	*PLCB1*	CNV	*SVIL*	LoF-SNV
CAMK2G	CNV	*IQGAP2*	LoF-SNV	*PPM1D*	LoF-SNV	*SYNCRIP**	CNV
CBX4	LoF-SNV	*ITGA5*	LoF-SNV	*PPP3CB*	CNV	*SYNJ2*	CNV
CDK2	CNV	*KAT2B*	CNV	*PRKAB2*	CNV	*TAOK2*	CNV
CDK4	CNV	*KIAA0232*	LoF-SNV	*PSMB1*	CNV	*TBC1D23*	LoF-SNV
*CHD2**	CNV	*KLF13*	CNV	*PSMC2*	CNV	*TBP*	CNV
CNOT3	LoF-SNV	*LIMK1*	CNV	*PTPRD*	CNV	*TBR1*	LoF-SNV
CRKL	CNV	*LRP2*	LoF-SNV	*RAB5A*	CNV	*TCF3*	LoF-SNV
CSDE1	LoF-SNV	*MAPK11*	CNV	*RAC3*	CNV	*TNK2*	CNV
CUBN	LoF-SNV	*MAPK12*	CNV	*RAP1GDS1*	CNV	*VCP*	LoF-SNV
CUL2	CNV	*MAPK3*	CNV	*RER1*	CNV	*ZMYND11*	LoF-SNV
CUL3	LoF-SNV	*MAPK7*	CNV	*RIMS1*	CNV, LoF-SNV	*ZNF292*	LoF-SNV
CYFIP1	CNV	*MAPK8*	CNV	*RPS6KA3*	LoF-SNV		
DLGAP2	CNV	*MAX*	CNV	*SETD2*	LoF-SNV		
DLL1	LoF-SNV	*MCF2*	LoF XL-SNV in males	*SH3GL3*	CNV		

This table lists 97 CNV or SNV genes present in the DAPPLE network, but not yet implicated in ASD or ID (i.e. not yet in Tables S6A-S6D), considered high-confidence ASD candidate genes. *Genes in section 'Highlighted genes'.

Table S17A. Listing of CNV calls in affected subjects

mmc6.xlsx (Excel workbook)

Table S17B. Chromosome abnormalities in parents and control subjects

mmc7.xlsx (Excel workbook)

Chromosome abnormalities in probands are listed in **Table S1C**.

Table S17C. Experimentally validated CNVs

mmc8.xlsx (Excel workbook)

The 456 validated CNVs in **Table S17C** comprise all CNVs validated experimentally across the 2446 ASD cases, including 315 CNVs confirmed in stage 1 samples and 141 CNVs confirmed in stage 2 samples.

SUPPLEMENTAL DATAFILES

Listing of separate Excel workbooks.

Table S8. Phenotypes in ASD subjects with pathogenic CNVs or with selected CNVs of uncertain significance

mmc2.xlsx

Tables S12A-S12D. GO terms, pathways, and MPO enrichment in affecteds versus control subjects

mmc3.xlsx

Tables S14A-S14E. Characterization of genes selected by NETBAG

mmc4.xlsx

Table S15. Functional-group enrichment for DAPPLE results

mmc5.xlsx

Table S17A. Listing of CNV calls in affected subjects

mmc6.xlsx

Table S17B. Chromosome abnormalities in parents and control subjects

mmc7.xlsx

Table S17C. Experimentally validated CNVs

mmc8.xlsx

SUPPLEMENTAL ACKNOWLEDGEMENTS

Additional support for individual groups of the Autism Genome Project was provided by the US National Institutes of Health (NIH grants: HD34565, HD055751, HD055782, MH06359, MH52708, MH057881, MH061009, MH082121, MH089924, MH094293, MH094382, MH094400, NS043540, NS049261), Assistance Publique-Hôpitaux de Paris (France), Autism Speaks Canada, Autistica, Canada Foundation for Innovation, Canadian Institutes for Advanced Research, The Centre for Applied Genomics (Canada), The Center for Applied Genomics–Children's Hospital of Philadelphia (USA); Deutsche Forschungsgemeinschaft (Germany), Fondation de France, Fondation FondaMental (France), Fondation pour la Recherche Médicale (France), Foundation for Autism Research (USA), Fundação Calouste Gulbenkian (Portugal), Fundação para Ciência e Tecnologia (SFRH/BPD/74739/2010, SFRH/BPD/64281/2009, SFRH/BD/79081/2011, Portugal), the Gulbenkian Programme for Advanced Medical Education and the Portuguese Foundation for Science and Technology (SFRH/BDINT/51549/2011, Portugal), the Hospital for Sick Children Foundation (Canada), INSERM (France), The Margaret Q. Landenberger Foundation (USA), The Medical Charities Research Group (Ireland), The Mindich Child Health and Development Institute (USA), Nancy Lurie Marks Family Foundation (USA); The National Children's Research Centre (Ireland), National Institute of Health Research (UK), NeuroDevNet (Canada), the Ontario government and Ontario Brain Institute (Canada), Our Lady's Children's Hospital, Dublin (Ireland), the Pennsylvania Department of Health (4100047863), the Seaver Foundation (USA), the University of Toronto McLaughlin Centre (Canada), and the Wellcome Trust (core award 090532/Z/09/Z, UK).

Acknowledgement NIH grant No. U24 CA074783 to S. Gallinger: This work was made possible through collaboration and cooperative agreements with the Colon Cancer Family Registry and P.I.s.

This study makes use of data generated by the DECIPHER Consortium. A full list of centres who contributed to the generation of the data is available from http://decipher.sanger.ac.uk and via email from decipher@sanger.ac.uk. Funding for the project was provided by the Wellcome Trust.

Lead AGP investigators: D. Pinto, E. M. Wijsman, V. J. Vieland, A. M. Vicente, G. D. Schellenberg, M. Pericak-Vance, A. D. Paterson, J. R. Parr, G. Oliveira, J. I. Nurnberger, A. P. Monaco, E. Maestrini, S. M. Klauck, H. Hakonarson, J. L. Haines, D. H. Geschwind, C. M. Freitag, S. E. Folstein, S. Ennis, H. Coon, A. Battaglia, P. Szatmari, J. S. Sutcliffe, J. Hallmayer, M. Gill, E. H. Cook, J. D. Buxbaum, B. Devlin, L. Gallagher, C. Betancur, and S. W. Scherer.

SUPPLEMENTAL REFERENCES

1. Gilman, S.R., Iossifov, I., Levy, D., Ronemus, M., Wigler, M., and Vitkup, D. (2011). Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. Neuron *70*, 898-907.
2. Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., and Lewis, S. (2009). AmiGO: online access to ontology and annotation data. Bioinformatics *25*, 288-289.
3. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res *40*, D109-114.
4. Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y.H., Narzisi, G., Leotta, A., et al. (2012). De novo gene disruptions in children on the autistic spectrum. Neuron *74*, 285-299.
5. Neale, B.M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K.E., Sabo, A., Lin, C.F., Stevens, C., Wang, L.S., Makarov, V., et al. (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature *485*, 242-245.
6. O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D., et al. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature *485*, 246-250.
7. Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature *485*, 237-241.
8. Kielinen, M., Rantala, H., Timonen, E., Linna, S.L., and Moilanen, I. (2004). Associated medical disorders and disabilities in children with autistic disorder: a population-based study. Autism *8*, 49-60.
9. Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y., Mele, A., Fraser, C.E., Stone, E.F., Chen, C., Fak, J.J., Chi, S.W., et al. (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell *146*, 247-261.
10. Ascano, M., Jr., Mukherjee, N., Bandaru, P., Miller, J.B., Nusbaum, J.D., Corcoran, D.L., Langlois, C., Munschauer, M., Dewell, S., Hafner, M., et al. (2012). FMRP targets distinct mRNA sequence elements to regulate protein expression. Nature *492*, 382-386.
11. Bayes, A., van de Lagemaat, L.N., Collins, M.O., Croning, M.D., Whittle, I.R., Choudhary, J.S., and Grant, S.G. (2011). Characterization of the proteome, diseases and evolution of the human postsynaptic density. Nat Neurosci *14*, 19-21.
12. Huang, N., Lee, I., Marcotte, E.M., and Hurles, M.E. (2010). Characterising and predicting haploinsufficiency in the human genome. PLoS Genet *6*, e1001154.
13. Robinson, P.N., Kohler, S., Bauer, S., Seelow, D., Horn, D., and Mundlos, S. (2008). The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. Am J Hum Genet *83*, 610-615.
14. Eppig, J.T., Blake, J.A., Bult, C.J., Kadin, J.A., and Richardson, J.E. (2012). The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse. Nucleic Acids Res *40*, D881-886.
15. BrainSpan. (2012). Atlas of the Developing Human Brain, Technical white paper; Section: Measuring Expression. In. (
16. Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M., Pletikos, M., Meyer, K.A., Sedmak, G., et al. (2011). Spatio-temporal transcriptome of the human brain. Nature *478*, 483-489.
17. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods *5*, 621-628.
18. Xu, B., Ionita-Laza, I., Roos, J.L., Boone, B., Woodrick, S., Sun, Y., Levy, S., Gogos, J.A., and Karayiorgou, M. (2012). De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. Nat Genet *44*, 1365-1369.
19. Rossin, E.J., Lage, K., Raychaudhuri, S., Xavier, R.J., Tatar, D., Benita, Y., Cotsapas, C., and Daly, M.J. (2011). Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. PLoS Genet *7*, e1001273.
20. Lage, K., Karlberg, E.O., Storling, Z.M., Olason, P.I., Pedersen, A.G., Rigina, O., Hinsby, A.M., Tumer, Z., Pociot, F., Tommerup, N., et al. (2007). A human phenome-interactome network of protein complexes implicated in genetic disorders. Nat Biotechnol *25*, 309-316.
21. Betancur, C. (2011). Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. Brain Res *1380*, 42-77.

22. Lim, E.T., Raychaudhuri, S., Sanders, S.J., Stevens, C., Sabo, A., MacArthur, D.G., Neale, B.M., Kirby, A., Ruderfer, D.M., Fromer, M., et al. (2013). Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. Neuron *77*, 235-242.
23. Meyer, L.R., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Kuhn, R.M., Wong, M., Sloan, C.A., Rosenbloom, K.R., Roe, G., Rhead, B., et al. (2013). The UCSC Genome Browser database: extensions and updates 2013. Nucleic Acids Res *41*, D64-69.
24. Khurana, E., Fu, Y., Chen, J., and Gerstein, M. (2013). Interpretation of genomic variants using a unified biological network approach. PLoS Comput Biol *9*, e1002886.
25. Li, M.M., Nimmakayalu, M.A., Mercer, D., Andersson, H.C., and Emanuel, B.S. (2008). Characterization of a cryptic 3.3 Mb deletion in a patient with a "balanced t(15;22) translocation" using high density oligo array CGH and gene expression arrays. Am J Med Genet A *146*, 368-375.
26. Veredice, C., Bianco, F., Contaldo, I., Orteschi, D., Stefanini, M.C., Battaglia, D., Lettori, D., Guzzetta, F., and Zollino, M. (2009). Early onset myoclonic epilepsy and 15q26 microdeletion: observation of the first case. Epilepsia *50*, 1810-1815.
27. Dhamija, R., Breningstall, G., Wong-Kisiel, L., Dolan, M., Hirsch, B., and Wirrell, E. (2011). Microdeletion of chromosome 15q26.1 in a child with intractable generalized epilepsy. Pediatr Neurol *45*, 60-62.
28. Capelli, L.P., Krepischi, A.C., Gurgel-Giannetti, J., Mendes, M.F., Rodrigues, T., Varela, M.C., Koiffmann, C.P., and Rosenberg, C. (2012). Deletion of the RMGA and CHD2 genes in a child with epilepsy and mental deficiency. Eur J Med Genet *55*, 132-134.
29. Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M., and Carter, N.P. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet *84*, 524-533.
30. Kaminsky, E.B., Kaul, V., Paschall, J., Church, D.M., Bunke, B., Kunig, D., Moreno-De-Luca, D., Moreno-De-Luca, A., Mulle, J.G., Warren, S.T., et al. (2011). An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. Genet Med *13*, 777-784.
31. Feenstra, I., Fang, J., Koolen, D.A., Siezen, A., Evans, C., Winter, R.M., Lees, M.M., Riegel, M., de Vries, B.B., Van Ravenswaaij, C.M., et al. (2006). European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations (ECARUCA); an online database for rare chromosome abnormalities. Eur J Med Genet *49*, 279-291.
32. Korn, J.M., Kuruvilla, F.G., McCarroll, S.A., Wysoker, A., Nemesh, J., Cawley, S., Hubbell, E., Veitch, J., Collins, P.J., Darvishi, K., et al. (2008). Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nat Genet *40*, 1253-1260.
33. Carvill, G.L., Heavin, S.B., Yendle, S.C., McMahon, J.M., O'Roak, B.J., Cook, J., Khan, A., Dorschner, M.O., Weaver, M., Calvert, S., et al. (2013). Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. Nat Genet *45*, 825-830.
34. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat Methods *7*, 248-249.
35. Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., and Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. Genome Res *13*, 2129-2141.
36. Ng, P.C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res *31*, 3812-3814.
37. Bromberg, Y., and Rost, B. (2007). SNAP: predict effect of non-synonymous polymorphisms on function. Nucleic Acids Res *35*, 3823-3835.
38. Schwarz, J.M., Rodelsperger, C., Schuelke, M., and Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods *7*, 575-576.
39. Cooper, G.M., Stone, E.A., Asimenos, G., Green, E.D., Batzoglou, S., and Sidow, A. (2005). Distribution and intensity of constraint in mammalian genomic sequence. Genome Res *15*, 901-913.
40. Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res *38*, W529-533.
41. Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endele, S., Schwarzmayr, T., Albrecht, B., Bartholdi, D., Beygo, J., Di Donato, N., et al. (2012). Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet *380*, 1674-1682.
42. Kulkarni, S., Nagarajan, P., Wall, J., Donovan, D.J., Donell, R.L., Ligon, A.H., Venkatachalam, S., and Quade, B.J. (2008). Disruption of chromodomain helicase DNA binding protein 2 (CHD2) causes scoliosis. Am J Med Genet A *146A*, 1117-1127.
43. Marfella, C.G., Henninger, N., LeBlanc, S.E., Krishnan, N., Garlick, D.S., Holzman, L.B., and Imbalzano, A.N. (2008). A mutation in the mouse Chd2 chromatin remodeling enzyme results in a complex renal phenotype. Kidney Blood Press Res *31*, 421-432.
44. Vissers, L.E., van Ravenswaaij, C.M., Admiraal, R., Hurst, J.A., de Vries, B.B., Janssen, I.M., van der Vliet, W.A., Huys, E.H., de Jong, P.J., Hamel, B.C., et al. (2004). Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. Nat Genet *36*, 955-957.
45. Johansson, M., Rastam, M., Billstedt, E., Danielsson, S., Stromland, K., Miller, M., and Gillberg, C. (2006). Autism spectrum disorders and underlying brain pathology in CHARGE association. Dev Med Child Neurol *48*, 40-50.
46. O'Roak, B.J., Vives, L., Fu, W., Egertson, J.D., Stanaway, I.B., Phelps, I.G., Carvill, G., Kumar, A., Lee, C., Ankenman, K., et al. (2012). Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science *338*, 1619-1622.

47. Doherty, E.S., and Lacbawan, F. (2013). 2q37 microdeletion syndrome. GeneReviews, http://www.ncbi.nlm.nih.gov/books/NBK1158/.
48. Williams, S.R., Aldred, M.A., Der Kaloustian, V.M., Halal, F., Gowans, G., McLeod, D.R., Zondag, S., Toriello, H.V., Magenis, R.E., and Elsea, S.H. (2010). Haploinsufficiency of HDAC4 causes brachydactyly mental retardation syndrome, with brachydactyly type E, developmental delays, and behavioral problems. Am J Hum Genet *87*, 219-228.
49. Villavicencio-Lorini, P., Klopocki, E., Trimborn, M., Koll, R., Mundlos, S., and Horn, D. (2012). Phenotypic variant of Brachydactyly-mental retardation syndrome in a family with an inherited interstitial 2q37.3 microdeletion including HDAC4. Eur J Hum Genet *21*, 743-748.
50. Morris, B., Etoubleau, C., Bourthoumieu, S., Reynaud-Perrine, S., Laroche, C., Lebbar, A., Yardin, C., and Elsea, S.H. (2012). Dose dependent expression of HDAC4 causes variable expressivity in a novel inherited case of brachydactyly mental retardation syndrome. Am J Med Genet A *158A*, 2015-2020.
51. Sando, R., 3rd, Gounko, N., Pieraut, S., Liao, L., Yates, J., 3rd, and Maximov, A. (2012). HDAC4 governs a transcriptional program essential for synaptic plasticity and memory. Cell *151*, 821-834.
52. Kim, M.S., Akhtar, M.W., Adachi, M., Mahgoub, M., Bassel-Duby, R., Kavalali, E.T., Olson, E.N., and Monteggia, L.M. (2012). An essential role for histone deacetylase 4 in synaptic plasticity and memory formation. J Neurosci *32*, 10879-10886.
53. Bienvenu, T., des Portes, V., Saint Martin, A., McDonell, N., Billuart, P., Carrie, A., Vinet, M.C., Couvert, P., Toniolo, D., Ropers, H.H., et al. (1998). Non-specific X-linked semidominant mental retardation by mutations in a Rab GDP-dissociation inhibitor. Hum Mol Genet *7*, 1311-1315.
54. D'Adamo, P., Menegon, A., Lo Nigro, C., Grasso, M., Gulisano, M., Tamanini, F., Bienvenu, T., Gedeon, A.K., Oostra, B., Wu, S.K., et al. (1998). Mutations in GDI1 are responsible for X-linked non-specific mental retardation. Nat Genet *19*, 134-139.
55. Vandewalle, J., Van Esch, H., Govaerts, K., Verbeeck, J., Zweier, C., Madrigal, I., Mila, M., Pijkels, E., Fernandez, I., Kohlhase, J., et al. (2009). Dosage-dependent severity of the phenotype in patients with mental retardation due to a recurrent copy-number gain at Xq28 mediated by an unusual recombination. Am J Hum Genet *85*, 809-822.
56. Clayton-Smith, J., Walters, S., Hobson, E., Burkitt-Wright, E., Smith, R., Toutain, A., Amiel, J., Lyonnet, S., Mansour, S., Fitzpatrick, D., et al. (2009). Xq28 duplication presenting with intestinal and bladder dysfunction and a distinctive facial appearance. Eur J Hum Genet *17*, 434-443.
57. Madrigal, I., Rodriguez-Revenga, L., Badenas, C., Sanchez, A., Martinez, F., Fernandez, I., Fernandez-Burriel, M., and Mila, M. (2007). MLPA as first screening method for the detection of microduplications and microdeletions in patients with X-linked mental retardation. Genet Med *9*, 117-122.
58. Sakai, Y., Shaw, C.A., Dawson, B.C., Dugas, D.V., Al-Mohtaseb, Z., Hill, D.E., and Zoghbi, H.Y. (2011). Protein interactome reveals converging molecular pathways among autism disorders. Sci Transl Med *3*, 86ra49.
59. Bianchi, V., Farisello, P., Baldelli, P., Meskenaite, V., Milanese, M., Vecellio, M., Muhlemann, S., Lipp, H.P., Bonanno, G., Benfenati, F., et al. (2009). Cognitive impairment in Gdi1-deficient mice is associated with altered synaptic vesicle pools and short-term synaptic plasticity, and can be corrected by appropriate learning training. Hum Mol Genet *18*, 105-117.
60. Shuib, S., McMullan, D., Rattenberry, E., Barber, R.M., Rahman, F., Zatyka, M., Chapman, C., Macdonald, F., Latif, F., Davison, V., et al. (2009). Microarray based analysis of 3p25-p26 deletions (3p- syndrome). Am J Med Genet A *149A*, 2099-2105.
61. Gunnarsson, C., and Foyn Bruun, C. (2010). Molecular characterization and clinical features of a patient with an interstitial deletion of 3p25.3-p26.1. Am J Med Genet A *152A*, 3110-3114.
62. Riess, A., Grasshoff, U., Schaferhoff, K., Bonin, M., Riess, O., Horber, V., and Tzschach, A. (2012). Interstitial 3p25.3-p26.1 deletion in a patient with intellectual disability. Am J Med Genet A *158A*, 2587-2590.
63. Peltekova, I.T., Macdonald, A., and Armour, C.M. (2012). Microdeletion on 3p25 in a patient with features of 3p deletion syndrome. Am J Med Genet A *158A*, 2583-2586.
64. Kellogg, G., Sum, J., and Wallerstein, R. (2013). Deletion of 3p25.3 in a patient with intellectual disability and dysmorphic features with further definition of a critical region. Am J Med Genet A *161*, 1405-1408.
65. Ellery, P.M., Ellis, R.J., and Holder, S.E. (2014). Interstitial 3p25 deletion in a patient with features of 3p deletion syndrome: further evidence for the role of SRGAP3 in mental retardation. Clin Dysmorphol *23*, 29-31.
66. Shaikh, T.H., Gai, X., Perin, J.C., Glessner, J.T., Xie, H., Murphy, K., O'Hara, R., Casalunovo, T., Conlin, L.K., D'Arcy, M., et al. (2009). High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. Genome Res *19*, 1682-1690.
67. Schaefer, G.B., and Lutz, R.E. (2006). Diagnostic yield in the clinical genetic evaluation of autism spectrum disorders. Genet Med *8*, 549-556.
68. Ververi, A., Vargiami, E., Papadopoulou, V., Tryfonas, D., and Zafeiriou, D.I. (2012). Clinical and laboratory data in a sample of Greek children with autism spectrum disorders. J Autism Dev Disord *42*, 1470-1476.
69. Kleefstra, T., van Zelst-Stams, W.A., Nillesen, W.M., Cormier-Daire, V., Houge, G., Foulds, N., van Dooren, M., Willemsen, M.H., Pfundt, R., Turner, A., et al. (2009). Further clinical and molecular delineation of the 9q subtelomeric deletion syndrome supports a major contribution of EHMT1 haploinsufficiency to the core phenotype. J Med Genet *46*, 598-606.
70. Rodriguez-Santiago, B., Malats, N., Rothman, N., Armengol, L., Garcia-Closas, M., Kogevinas, M., Villa, O., Hutchinson, A., Earl, J., Marenne, G., et al. (2010). Mosaic uniparental disomies and aneuploidies as large structural variants of the human genome. Am J Hum Genet *87*, 129-138.

71. Molin, A.M., Andrieux, J., Koolen, D.A., Malan, V., Carella, M., Colleaux, L., Cormier-Daire, V., David, A., de Leeuw, N., Delobel, B., et al. (2012). A novel microdeletion syndrome at 3q13.31 characterised by developmental delay, postnatal overgrowth, hypoplastic male genitals, and characteristic facial features. J Med Genet *49*, 104-109.
72. Gimelli, S., Leoni, M., Di Rocco, M., Caridi, G., Porta, S., Cuoco, C., Gimelli, G., and Tassano, E. (2013). A rare 3q13.31 microdeletion including GAP43 and LSAMP genes. Mol Cytogenet *6*, 52.
73. Gil, O.D., Zhang, L., Chen, S., Ren, Y.Q., Pimenta, A., Zanazzi, G., Hillman, D., Levitt, P., and Salzer, J.L. (2002). Complementary expression and heterophilic interactions between IgLON family members neurotrimin and LAMP. J Neurobiol *51*, 190-204.
74. Catania, E.H., Pimenta, A., and Levitt, P. (2008). Genetic deletion of Lsamp causes exaggerated behavioral activation in novel environments. Behav Brain Res *188*, 380-390.
75. Qiu, S., Champagne, D.L., Peters, M., Catania, E.H., Weeber, E.J., Levitt, P., and Pimenta, A.F. (2010). Loss of limbic system-associated membrane protein leads to reduced hippocampal mineralocorticoid receptor expression, impaired synaptic plasticity, and spatial memory deficit. Biol Psychiatry *68*, 197-204.
76. Gijsbers, A.C., den Hollander, N.S., Helderman-van de Enden, A.T., Schuurs-Hoeijmakers, J.H., Vijfhuizen, L., Bijlsma, E.K., van Haeringen, A., Hansson, K.B., Bakker, E., Breuning, M.H., et al. (2011). X-chromosome duplications in males with mental retardation: pathogenic or benign variants? Clin Genet *79*, 71-78.
77. Shimokawa, N., Haglund, K., Holter, S.M., Grabbe, C., Kirkin, V., Koibuchi, N., Schultz, C., Rozman, J., Hoeller, D., Qiu, C.H., et al. (2010). CIN85 regulates dopamine receptor endocytosis and governs behaviour in mice. Embo J *29*, 2421-2432.
78. Kowanetz, K., Husnjak, K., Holler, D., Kowanetz, M., Soubeyran, P., Hirsch, D., Schmidt, M.H., Pavelic, K., De Camilli, P., Randazzo, P.A., et al. (2004). CIN85 associates with multiple effectors controlling intracellular trafficking of epidermal growth factor receptors. Mol Biol Cell *15*, 3155-3166.
79. Kawata, A., Iida, J., Ikeda, M., Sato, Y., Mori, H., Kansaku, A., Sumita, K., Fujiwara, N., Rokukawa, C., Hamano, M., et al. (2006). CIN85 is localized at synapses and forms a complex with S-SCAM via dendrin. J Biochem *139*, 931-939.
80. Gudjonsson, T., Altmeyer, M., Savic, V., Toledo, L., Dinant, C., Grofte, M., Bartkova, J., Poulsen, M., Oka, Y., Bekker-Jensen, S., et al. (2012). TRIP12 and UBR5 suppress spreading of chromatin ubiquitylation at damaged chromosomes. Cell *150*, 697-709.
81. Poulsen, E.G., Steinhauer, C., Lees, M., Lauridsen, A.M., Ellgaard, L., and Hartmann-Petersen, R. (2012). HUWE1 and TRIP12 collaborate in degradation of ubiquitin-fusion proteins and misframed ubiquitin. PLoS One *7*, e50548.
82. Bannai, H., Fukatsu, K., Mizutani, A., Natsume, T., Iemura, S., Ikegami, T., Inoue, T., and Mikoshiba, K. (2004). An RNA-interacting protein, SYNCRIP (heterogeneous nuclear ribonuclear protein Q1/NSAP1) is a component of mRNA granule transported with inositol 1,4,5-trisphosphate receptor type 1 mRNA in neuronal dendrites. J Biol Chem *279*, 53427-53434.
83. Pak, C., Garshasbi, M., Kahrizi, K., Gross, C., Apponi, L.H., Noto, J.J., Kelly, S.M., Leung, S.W., Tzschach, A., Behjati, F., et al. (2011). Mutation of the conserved polyadenosine RNA binding protein, ZC3H14/dNab2, impairs neural function in Drosophila and humans. Proc Natl Acad Sci U S A *108*, 12390-12395.
84. Piton, A., Michaud, J.L., Peng, H., Aradhya, S., Gauthier, J., Mottron, L., Champagne, N., Lafreniere, R.G., Hamdan, F.F., Joober, R., et al. (2008). Mutations in the calcium-related gene IL1RAPL1 are associated with autism. Hum Mol Genet *17*, 3965-3974.
85. Wang, P., Carrion, P., Qiao, Y., Tyson, C., Hrynchak, M., Calli, K., Lopez-Rangel, E., Andrieux, J., Delobel, B., Duban-Bedu, B., et al. (2013). Genotype-phenotype analysis of 18q12.1-q12.2 copy number variation in autism. Eur J Med Genet *(Epub ahead of print)*.
86. Ichida, F., Tsubata, S., Bowles, K.R., Haneda, N., Uese, K., Miyawaki, T., Dreyer, W.J., Messina, J., Li, H., Bowles, N.E., et al. (2001). Novel gene mutations in patients with left ventricular noncompaction or Barth syndrome. Circulation *103*, 1256-1263.
87. Hinton, V.J., Cyrulnik, S.E., Fee, R.J., Batchelder, A., Kiefel, J.M., Goldstein, E.M., Kaufmann, P., and De Vivo, D.C. (2009). Association of autistic spectrum disorders with dystrophinopathies. Pediatr Neurol *41*, 339-346.
88. Young, H.K., Barton, B.A., Waisbren, S., Portales Dale, L., Ryan, M.M., Webster, R.I., and North, K.N. (2008). Cognitive and psychological profile of males with Becker muscular dystrophy. J Child Neurol *23*, 155-162.
89. Pilgram, G.S., Potikanond, S., Baines, R.A., Fradkin, L.G., and Noordermeer, J.N. (2010). The roles of the dystrophin-associated glycoprotein complex at the synapse. Mol Neurobiol *41*, 1-21.
90. Grady, R.M., Grange, R.W., Lau, K.S., Maimone, M.M., Nichol, M.C., Stull, J.T., and Sanes, J.R. (1999). Role for alpha-dystrobrevin in the pathogenesis of dystrophin-dependent muscular dystrophies. Nat Cell Biol *1*, 215-220.
91. Grady, R.M., Wozniak, D.F., Ohlemiller, K.K., and Sanes, J.R. (2006). Cerebellar synaptic defects and abnormal motor behavior in mice lacking alpha- and beta-dystrobrevin. J Neurosci *26*, 2841-2851.
92. Garbett, K., Ebert, P.J., Mitchell, A., Lintas, C., Manzi, B., Mirnics, K., and Persico, A.M. (2008). Immune transcriptome alterations in the temporal cortex of subjects with autism. Neurobiol Dis *30*, 303-311.
93. Iwamoto, K., Kakiuchi, C., Bundo, M., Ikeda, K., and Kato, T. (2004). Molecular characterization of bipolar disorder by comparing gene expression profiles of postmortem brains of major mental disorders. Mol Psychiatry *9*, 406-416.
94. Metzinger, L., Blake, D.J., Squier, M.V., Anderson, L.V., Deconinck, A.E., Nawrotzki, R., Hilton-Jones, D., and Davies, K.E. (1997). Dystrobrevin deficiency at the sarcolemma of patients with muscular dystrophy. Hum Mol Genet *6*, 1185-1191.
95. Brauer, A.U., Savaskan, N.E., Kuhn, H., Prehn, S., Ninnemann, O., and Nitsch, R. (2003). A new phospholipid phosphatase, PRG-1, is involved in axon growth and regenerative sprouting. Nat Neurosci *6*, 572-578.
96. Willemsen, M.H., Valles, A., Kirkels, L.A., Mastebroek, M., Olde Loohuis, N., Kos, A., Wissink-Lindhout, W.M., de Brouwer, A.P., Nillesen, W.M., Pfundt, R., et al. (2011). Chromosome 1p21.3 microdeletions comprising DPYD and MIR137 are associated with intellectual disability. J Med Genet *48*, 810-818.

97. Carter, M.T., Nikkel, S.M., Fernandez, B.A., Marshall, C.R., Noor, A., Lionel, A.C., Prasad, A., Pinto, D., Joseph-George, A.M., Noakes, C., et al. (2011). Hemizygous deletions on chromosome 1p21.3 involving the DPYD gene in individuals with autism spectrum disorder. Clin Genet *80*, 435-443.
98. Smrt, R.D., Szulwach, K.E., Pfeiffer, R.L., Li, X., Guo, W., Pathania, M., Teng, Z.Q., Luo, Y., Peng, J., Bordey, A., et al. (2010). MicroRNA miR-137 regulates neuronal maturation by targeting ubiquitin ligase mind bomb-1. Stem Cells *28*, 1060-1070.
99. Szulwach, K.E., Li, X., Smrt, R.D., Li, Y., Luo, Y., Lin, L., Santistevan, N.J., Li, W., Zhao, X., and Jin, P. (2010). Cross talk between microRNA and epigenetic regulation in adult neurogenesis. J Cell Biol *189*, 127-141.
100. Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. (2011). Genome-wide association study identifies five new schizophrenia loci. Nat Genet *43*, 969-976.
101. Cusco, I., Medrano, A., Gener, B., Vilardell, M., Gallastegui, F., Villa, O., Gonzalez, E., Rodriguez-Santiago, B., Vilella, E., Del Campo, M., et al. (2009). Autism-specific copy number variants further implicate the phosphatidylinositol signaling pathway and the glutamatergic synapse in the etiology of the disorder. Hum Mol Genet *18*, 1795-1804.
102. Gross, C., Nakamoto, M., Yao, X., Chan, C.B., Yim, S.Y., Ye, K., Warren, S.T., and Bassell, G.J. (2010). Excess phosphoinositide 3-kinase subunit synthesis and activity as a novel therapeutic target in fragile X syndrome. J Neurosci *30*, 10624-10638.
103. Sharma, A., Hoeffer, C.A., Takayasu, Y., Miyawaki, T., McBride, S.M., Klann, E., and Zukin, R.S. (2010). Dysregulation of mTOR signaling in fragile X syndrome. J Neurosci *30*, 694-702.
104. Lang, B., Alrahbeni, T.M., Clair, D.S., Blackwood, D.H., McCaig, C.D., and Shen, S. (2012). HDAC9 is implicated in schizophrenia and expressed specifically in post-mitotic neurons but not in adult neural stem cells. Am J Stem Cells *1*, 31-41.
105. Bachmann-Gagescu, R., Mefford, H.C., Cowan, C., Glew, G.M., Hing, A.V., Wallace, S., Bader, P.I., Hamati, A., Reitnauer, P.J., Smith, R., et al. (2010). Recurrent 200-kb deletions of 16p11.2 that include the SH2B1 gene are associated with developmental delay and obesity. Genet Med *12*, 641-647.
106. Barge-Schaapveld, D.Q., Maas, S.M., Polstra, A., Knegt, L.C., and Hennekam, R.C. (2011). The atypical 16p11.2 deletion: a not so atypical microdeletion syndrome? Am J Med Genet A *155A*, 1066-1072.
107. Bijlsma, E.K., Gijsbers, A.C., Schuurs-Hoeijmakers, J.H., van Haeringen, A., Fransen van de Putte, D.E., Anderlid, B.M., Lundin, J., Lapunzina, P., Perez Jurado, L.A., Delle Chiaie, B., et al. (2009). Extending the phenotype of recurrent rearrangements of 16p11.2: deletions in mentally retarded patients without autism and in normal individuals. Eur J Med Genet *52*, 77-87.
108. Bochukova, E.G., Huang, N., Keogh, J., Henning, E., Purmann, C., Blaszczyk, K., Saeed, S., Hamilton-Shield, J., Clayton-Smith, J., O'Rahilly, S., et al. (2010). Large, rare chromosomal deletions associated with severe early-onset obesity. Nature *463*, 666-670.
109. Sampson, M.G., Coughlin, C.R., 2nd, Kaplan, P., Conlin, L.K., Meyers, K.E., Zackai, E.H., Spinner, N.B., and Copelovitch, L. (2010). Evidence for a recurrent microdeletion at chromosome 16p11.2 associated with congenital anomalies of the kidney and urinary tract (CAKUT) and Hirschsprung disease. Am J Med Genet A *152A*, 2618-2622.
110. Tabet, A.C., Pilorge, M., Delorme, R., Amsellem, F., Pinard, J.M., Leboyer, M., Verloes, A., Benzacken, B., and Betancur, C. (2012). Autism multiplex family with 16p11.2p12.2 microduplication syndrome in monozygotic twins and distal 16p11.2 deletion in their brother. Eur J Hum Genet *20*, 540-546.
111. Moreno-De-Luca, D., Sanders, S.J., Willsey, A.J., Mulle, J.G., Lowe, J.K., Geschwind, D.H., State, M.W., Martin, C.L., and Ledbetter, D.H. (2013). Using large clinical data sets to infer pathogenicity for rare copy number variants in autism cohorts. Mol Psychiatry *18*, 1090-1095.
112. Guha, S., Rees, E., Darvasi, A., Ivanov, D., Ikeda, M., Bergen, S.E., Magnusson, P.K., Cormican, P., Morris, D., Gill, M., et al. (2013). Implication of a rare deletion at distal 16p11.2 in schizophrenia. JAMA Psychiatry *70*, 253-260.
113. Doche, M.E., Bochukova, E.G., Su, H.W., Pearce, L.R., Keogh, J.M., Henning, E., Cline, J.M., Saeed, S., Dale, A., Cheetham, T., et al. (2012). Human SH2B1 mutations are associated with maladaptive behaviors and obesity. J Clin Invest *122*, 4732-4736.
114. Zhang, Y., Zhu, W., Wang, Y.G., Liu, X.J., Jiao, L., Liu, X., Zhang, Z.H., Lu, C.L., and He, C. (2006). Interaction of SH2-Bbeta with RET is involved in signaling of GDNF-induced neurite outgrowth. J Cell Sci *119*, 1666-1676.
115. Maures, T.J., Chen, L., and Carter-Su, C. (2009). Nucleocytoplasmic shuttling of the adapter protein SH2B1beta (SH2-Bbeta) is required for nerve growth factor (NGF)-dependent neurite outgrowth and enhancement of expression of a subset of NGF-responsive genes. Mol Endocrinol *23*, 1077-1091.
116. Ren, D., Zhou, Y., Morris, D., Li, M., Li, Z., and Rui, L. (2007). Neuronal SH2B1 is essential for controlling energy and glucose homeostasis. J Clin Invest *117*, 397-406.
117. Pinto, D., Pagnamenta, A.T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S., et al. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. Nature *466*, 368-372.
118. Colella, S., Yau, C., Taylor, J.M., Mirza, G., Butler, H., Clouston, P., Bassett, A.S., Seller, A., Holmes, C.C., and Ragoussis, J. (2007). QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res *35*, 2013-2025.
119. Pinto, D., Darvishi, K., Shi, X., Rajan, D., Rigler, D., Fitzgerald, T., Lionel, A.C., Thiruvahindrapuram, B., Macdonald, J.R., Mills, R., et al. (2011). Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. Nat Biotechnol *29*, 512-520.
120. Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F., Hakonarson, H., and Bucan, M. (2007). PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res *17*, 1665-1674.

121. Anney, R., Klei, L., Pinto, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S., Sykes, N., Pagnamenta, A.T., et al. (2010). A genome-wide scan for common alleles affecting risk for autism. Hum Mol Genet *19*, 4072-4082.
122. Anney, R., Klei, L., Pinto, D., Almeida, J., Bacchelli, E., Baird, G., Bolshakova, N., Bolte, S., Bolton, P.F., Bourgeron, T., et al. (2012). Individual common variants exert weak effects on the risk for autism spectrum disorders. Hum Mol Genet *21*, 4781-4792.
123. Klei, L., Kent, B.P., Melhem, N.M., Devlin, B., and Roeder, K. (2011). GemTools: a fast and efficient approach to estimating genetic ancestry. Statistics Applications *http://arxiv.org/PS_cache/arxiv/pdf/1104/1104.1162v1.pdf*.
124. Bierut, L.J., Agrawal, A., Bucholz, K.K., Doheny, K.F., Laurie, C., Pugh, E., Fisher, S., Fox, L., Howells, W., Bertelsen, S., et al. (2010). A genome-wide association study of alcohol dependence. Proc Natl Acad Sci U S A *107*, 5082-5087.
125. Newcomb, P.A., Baron, J., Cotterchio, M., Gallinger, S., Grove, J., Haile, R., Hall, D., Hopper, J.L., Jass, J., Le Marchand, L., et al. (2007). Colon Cancer Family Registry: an international resource for studies of the genetic epidemiology of colon cancer. Cancer Epidemiol Biomarkers Prev *16*, 2331-2343.
126. Figueiredo, J.C., Lewinger, J.P., Song, C., Campbell, P.T., Conti, D.V., Edlund, C.K., Duggan, D.J., Rangrej, J., Lemire, M., Hudson, T., et al. (2011). Genotype-environment interactions in microsatellite stable/microsatellite instability-low colorectal cancer: results from a genome-wide association study. Cancer Epidemiol Biomarkers Prev *20*, 758-766.
127. Fox, C.S., Liu, Y., White, C.C., Feitosa, M., Smith, A.V., Heard-Costa, N., Lohman, K., Johnson, A.D., Foster, M.C., Greenawalt, D.M., et al. (2012). Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women. PLoS Genet *8*, e1002695.
128. Griswold, A.J., Ma, D., Cukier, H.N., Nations, L.D., Schmidt, M.A., Chung, R.H., Jaworski, J.M., Salyakina, D., Konidari, I., Whitehead, P.L., et al. (2012). Evaluation of copy number variations reveals novel candidate genes in autism spectrum disorder-associated pathways. Hum Mol Genet *21*, 3513-3523.
129. Leblond, C.S., Heinrich, J., Delorme, R., Proepper, C., Betancur, C., Huguet, G., Konyukh, M., Chaste, P., Ey, E., Rastam, M., et al. (2012). Genetic and functional analyses of SHANK2 mutations suggest a multiple hit model of autism spectrum disorders. PLoS Genet *8*, e1002521.
130. Moessner, R., Marshall, C.R., Sutcliffe, J.S., Skaug, J., Pinto, D., Vincent, J., Zwaigenbaum, L., Fernandez, B., Roberts, W., Szatmari, P., et al. (2007). Contribution of SHANK3 mutations to autism spectrum disorder. Am J Hum Genet *81*, 1289-1297.
131. Hehir-Kwa, J.Y., Rodriguez-Santiago, B., Vissers, L.E., de Leeuw, N., Pfundt, R., Buitelaar, J.K., Perez-Jurado, L.A., and Veltman, J.A. (2011). De novo copy number variants associated with intellectual disability have a paternal origin and age bias. J Med Genet *48*, 776-778.
132. Sanders, S.J., Ercan-Sencicek, A.G., Hus, V., Luo, R., Murtha, M.T., Moreno-De-Luca, D., Chu, S.H., Moreau, M.P., Gupta, A.R., Thomson, S.A., et al. (2011). Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron *70*, 863-885.
133. Froyen, G., Belet, S., Martinez, F., Santos-Reboucas, C.B., Declercq, M., Verbeeck, J., Donckers, L., Berland, S., Mayo, S., Rosello, M., et al. (2012). Copy-number gains of HUWE1 due to replication- and recombination-based rearrangements. Am J Hum Genet *91*, 252-264.
134. Berkel, S., Marshall, C.R., Weiss, B., Howe, J., Roeth, R., Moog, U., Endris, V., Roberts, W., Szatmari, P., Pinto, D., et al. (2010). Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. Nat Genet *42*, 489-491.
135. Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., et al. (2007). Strong association of de novo copy number mutations with autism. Science *316*, 445-449.
136. Marshall, C.R., Noor, A., Vincent, J.B., Lionel, A.C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y., et al. (2008). Structural variation of chromosomes in autism spectrum disorder. Am J Hum Genet *82*, 477-488.
137. Noor, A., Whibley, A., Marshall, C.R., Gianakopoulos, P.J., Piton, A., Carson, A.R., Orlic-Milacic, M., Lionel, A.C., Sato, D., Pinto, D., et al. (2010). Disruption at the PTCHD1 Locus on Xp22.11 in Autism spectrum disorder and intellectual disability. Sci Transl Med *2*, 49ra68.
138. Pagnamenta, A.T., Holt, R., Yusuf, M., Pinto, D., Wing, K., Betancur, C., Scherer, S.W., Volpi, E.V., and Monaco, A.P. (2011). A family with autism and rare copy number variants disrupting the Duchenne/Becker muscular dystrophy gene DMD and TRPM3. J Neurodev Disord *3*, 124-131.
139. Stankiewicz, P., Kulkarni, S., Dharmadhikari, A.V., Sampath, S., Bhatt, S.S., Shaikh, T.H., Xia, Z., Pursley, A.N., Cooper, M.L., Shinawi, M., et al. (2012). Recurrent deletions and reciprocal duplications of 10q11.21q11.23 including CHAT and SLC18A3 are likely mediated by complex low-copy repeats. Hum Mutat *33*, 165-179.
140. Luo, R., Sanders, S.J., Tian, Y., Voineagu, I., Huang, N., Chu, S.H., Klei, L., Cai, C., Ou, J., Lowe, J.K., et al. (2012). Genome-wide transcriptome profiling reveals the functional impact of rare de novo and recurrent CNVs in autism spectrum disorders. Am J Hum Genet *91*, 38-55.