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, et al.

HAL Id: inserm-00986225

http://www.hal.inserm.fr/inserm-00986225

Submitted on 1 May 2014

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.)

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 = 1×10^{-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 $\geq$ 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 = 5.7×10^{-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 = 2.74×10^{-15}), and pathogenic deletions showed a striking estimated OR of 23.13 (95% CI = 5.57–96.08, p = 2.6×10^{-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 = 1.22×10^{-5}) carried pathogenic CNVs, whereas 3.0% (35/1,168) in stage 2 (OR = 6.47, p = 2.89×10^{-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
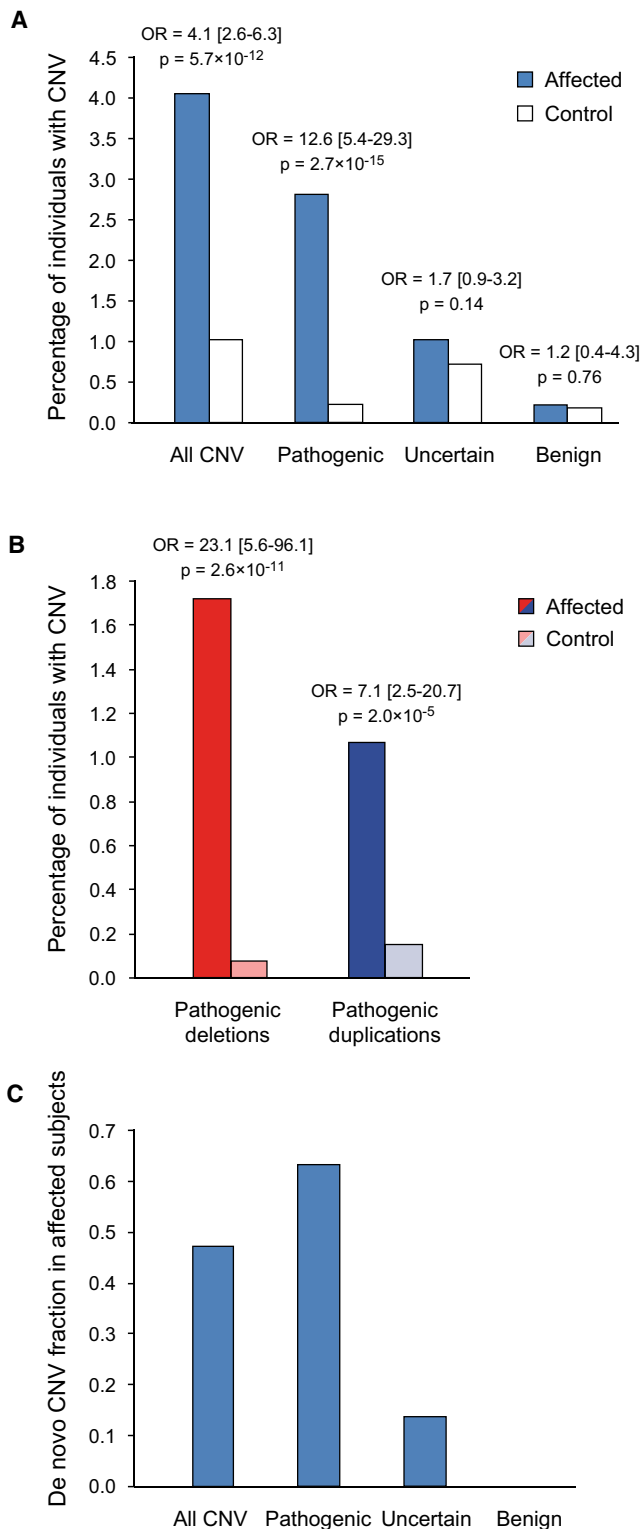
Figure 1. CNV Burden in Genes and Loci Implicated in ASD and/or ID

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-associ-

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

ated 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.
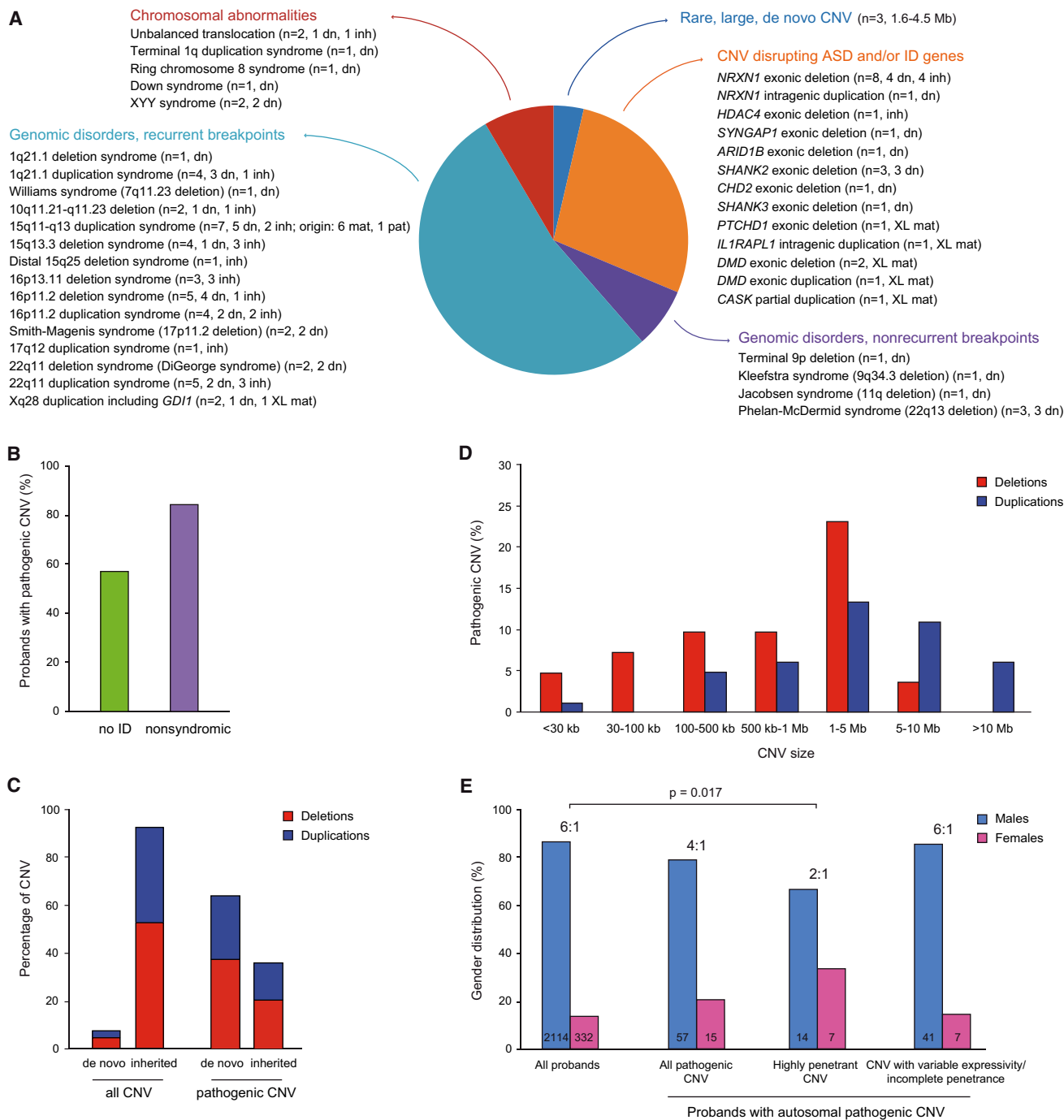
Figure 2. All Pathogenic CNVs Identified in Affected Subjects

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 × 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 × 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 × 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 × 10^{-5} and p = 4.7 × 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
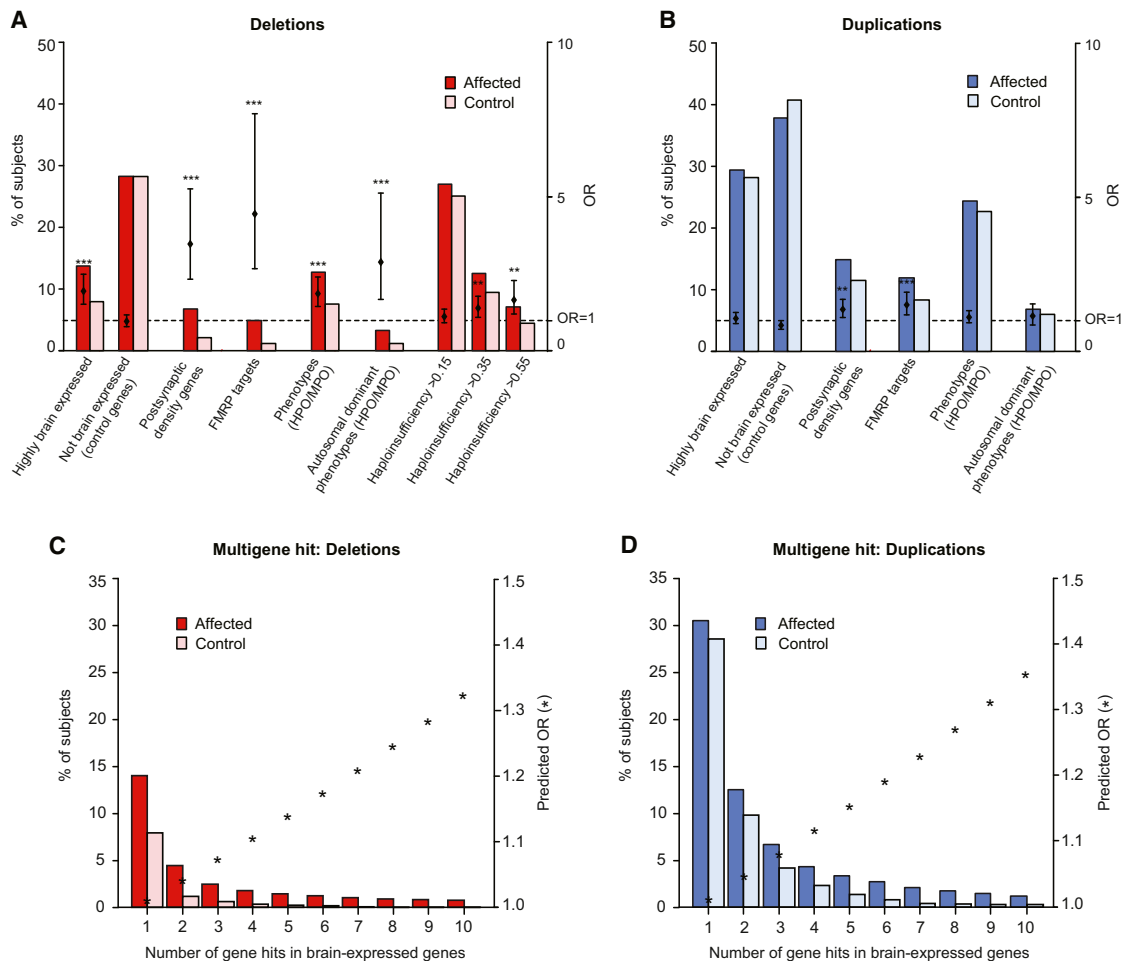
Figure 3. Enrichment of Functional Gene Sets Affected by Rare Exonic CNVs in Affected versus Control Subjects

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).
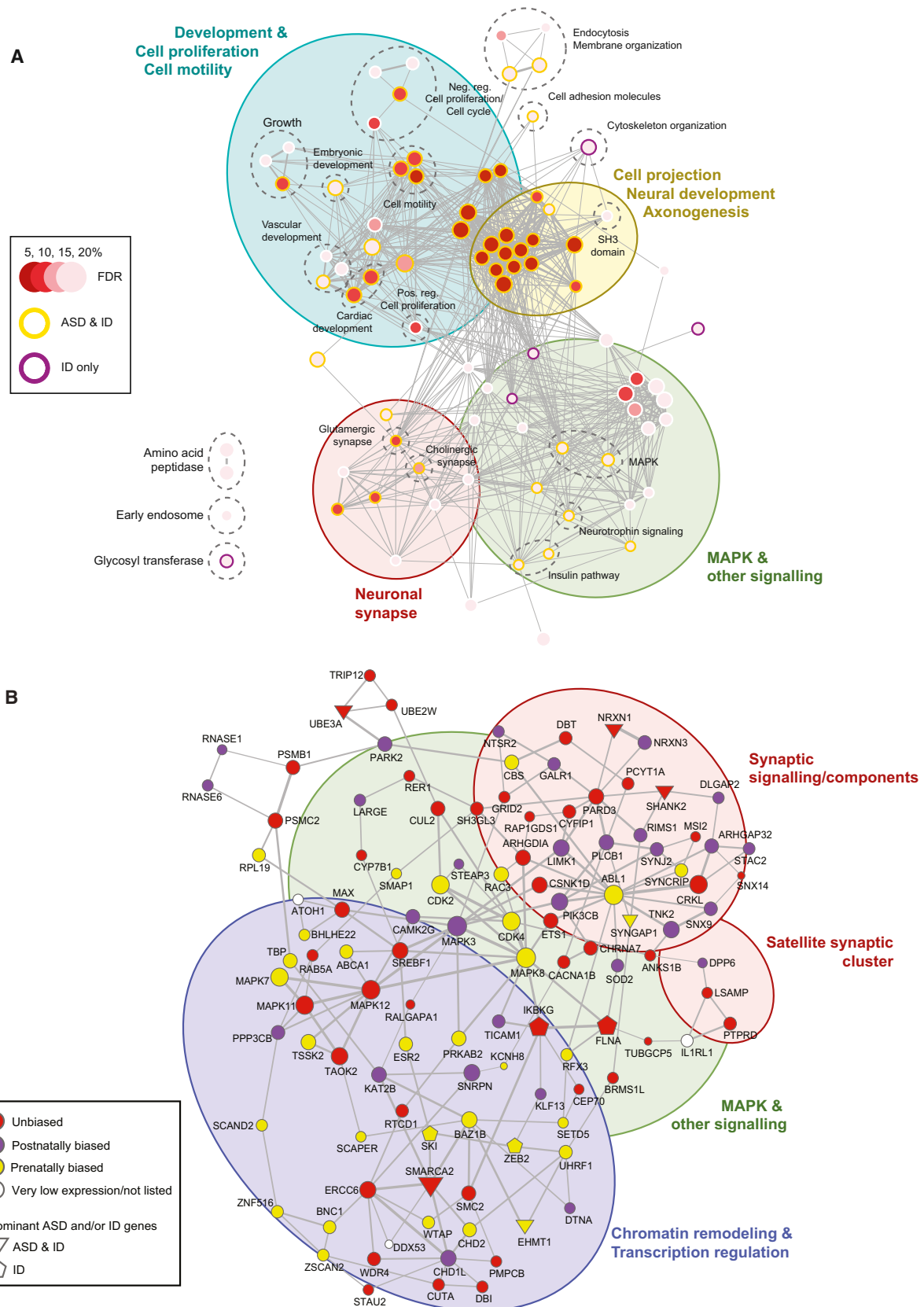
(C–D) Pattern of increased burden as the number of brain-expressed genes affected by deletions (C) or duplications (D) increased. The percentages of affected and control subjects with CNVs overlapping genes are shown for deletions and duplications separately. For estimating the expected ORs (stars), 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 ≤ 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 × 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 × 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.
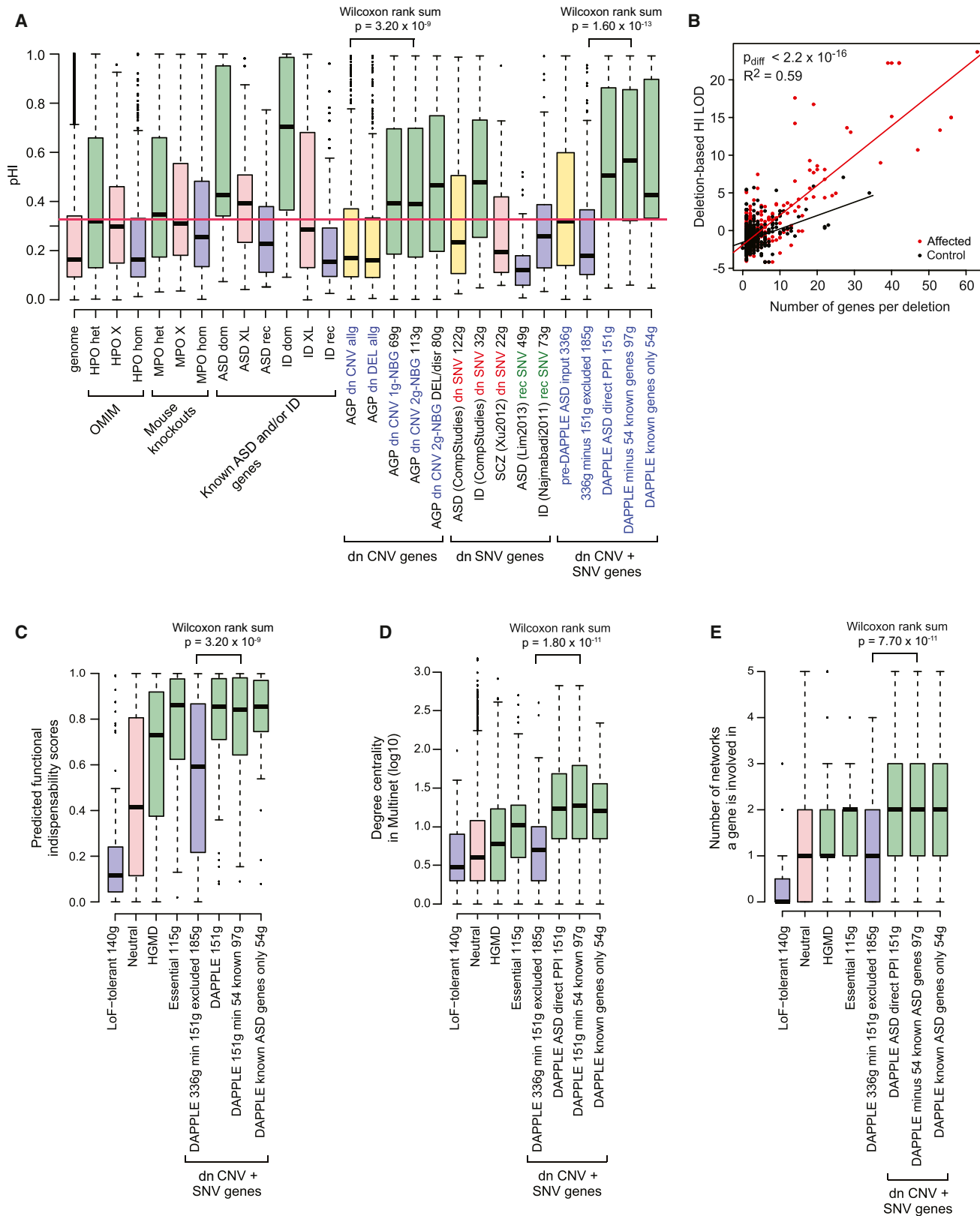
Figure 6. Functional Metrics for Various Gene Sets Derived from CNV and SNV Studies, as well as HI Scores for Genic Deletions

(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

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).

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.