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.

To cite this version:
Dalila Pinto, Elsa Delaby, Daniele Merico, Mafalda Barbosa, Alison Merikangas, et al.. Convergence of Genes and Cellular Pathways Dysregulated in Autism Spectrum Disorders.. American Journal of Human Genetics, Elsevier (Cell Press), 2014, 94 (5), pp.677-694. <10.1016/j.ajhg.2014.03.018>.<inserm-00986225>

HAL Id: inserm-00986225
http://www.hal.inserm.fr/inserm-00986225
Submitted on 1 May 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Convergence of Genes and Cellular Pathways Dysregulated in Autism Spectrum Disorders


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, MIR317, 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.

1Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 2Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 3Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 4The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 5The Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 6Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 7Discipline of Psychiatry, School of Medicine, Trinity College Dublin, Dublin 8, Ireland; 8Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA; 9Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, 3584CX Utrecht, the Netherlands; 10Department of Psychiatry and Behavioural Neurosciences, Oﬀord Centre for Child Studies, McMaster University, Hamilton, ON L8S 4K1, Canada; 11National Children’s Research Centre, Our Lady’s Children’s Hospital, Dublin 12, Ireland; 12Academic Centre on Rare Diseases, School of Medicine, University College Dublin, Dublin 4, Ireland; 13Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK; 14Instituto Nacional de Saúde Doutor Ricardo Jorge, 1649-016 Lisboa, Portugal; 15Center for Biodiversity, Functional & Integrative Genomics, Faculty of Sciences, University of Lisbon, 1749-016 Lisboa, Portugal; 16McLaughlin Centre, University of Toronto, Toronto, ON M5S 1A1, Canada; 17Convergence of Genes and Cellular Pathways

The American Journal of Human Genetics 94, 677–694, May 1, 2014 677
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., NRXN112 [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 ASD10,15 and an additional 5% conferred by rare inherited recessive or X-linked loss-of-function (LoF) SNVs.1,16 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

---

22Department of Pharmacy and Biotechnology, University of Bologna, 40126 Bologna, Italy; 23Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON M5G 1X5, Canada; 24Cancer Care Ontario, Toronto, ON M5G 2L7, Canada; 25Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA; 26Department of Pediatrics, University of Alberta, Edmonton, AB T6G 2H3, Canada; 27School of Education, University of Birmingham, Birmingham B15 2TT, UK; 28Department of Psychiatry, University of Oxford and Warneford Hospital, Oxford OX3 7JX, UK; 29Unité de Recherche Interdisciplinaire Octogone, Centre d’Etudes et de Recherches en Psychopathologie, Toulouse 2 University, 31058 Toulouse, France; 30Autism Research Unit, The Hospital for Sick Children, Toronto, ON M5G 1X8, Canada; 31Unidade de Neurodesenvolvimento e Autismo do Serviço do Centro de Desenvolvimento da Criança e Centro de Investigação e Formação Clínica, Pediatric Hospital, Centro Hospitalar e Universitário de Coimbra, 3000-602 Coimbra, Portugal; 32University Clinic of Pediatrics and Institute for Biomedical Imaging and Life Science, Faculty of Medicine, University of Coimbra, 3000-354 Coimbra, Portugal; 33Department of Pediatrics, Vanderbilt University, Nashville, TN 37232, USA; 34NewYork-Presbyterian/Weill Cornell Medical Center, New York, NY 10065, USA; 35Fondation Scientifique de l’Assurance Maladie du Travail (FSM), 94010 Créteil, France; 41Institut National de la Santé et de la Recherche U955, Psychiatrie Génétique, 94010 Créteil, France; 42Faculté de Médecine, Université Paris Est, 94010 Créteil, France; 43Department of Psychiatry, Henri Mondor-André Chanvrier Hospital, Assistance Publique – Hôpitaux de Paris, 94010 Créteil, France; 44Institute of Health and Society, Newcastle University, Newcastle upon Tyne NE1 4LP, UK; 45Child Developmental and Behavioral Unit, Hospital Nacional de Niños Dr. Sáenz Herrera, Caja Costarricense de Seguro Social, San José, Costa Rica; 46Institute for juvenile Research, Department of Psychiatry, University of Illinois at Chicago, Chicago, IL 60608, USA; 47Institute of Translational Neuroscience and Department of Psychiatry, University of Minnesota, Minneapolis, MN 55455, USA; 48Institute of Brain, Behaviour, and Mental Health, University of Manchester, Manchester M13 9PL, UK; 49Manchester Academic Health Sciences Centre, Manchester M13 9NT, UK; 50National Centre for Medical Genetics, Our Lady’s Children’s Hospital, Dublin 12, Ireland; 51Gillberg Neuropsychiatry Centre, University of Gothenburg, 41119 Gothenburg, Sweden; 52Discipline of Genetics, Faculty of Medicine, Memorial University of Newfoundland, St. John’s, NL A1B 3V6, Canada; 53Human Genetics and Cognitive Functions Unit, Institut Pasteur, 75015 Paris, France; 54Centre National de la Recherche Scientifique URA 2182 (Genes, Synapses, and Cognition), Institut Pasteur, 75015 Paris, France; 55Department of Child and Adolescent Psychiatry, Robert Debré Hospital, Assistance Publique – Hôpitaux de Paris, 75019 Paris, France; 56Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, NC 27710, USA; 57University Paris Diderot, Sorbonne Paris Cité, 75013 Paris, France; 58Institute of Psychiatry, King’s College London, London SE1 9RJ, UK; 59South London & Maudsley Biomedical Research Centre for Mental Health, London SE5 8AF, UK; 60Department of Child and Adolescent Psychiatry, Psychosomatics, and Psychotherapy, University of Frankfurt, 60528 Frankfurt, Germany; 61Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA 98195, USA; 62Paediatric Neurodisability, King’s Health Partners, King’s College London, London WC2R 2LS, UK; 63Bloorview Research Institute, University of Toronto, Toronto, ON M4G 1R8, Canada; 64Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA 98195, USA; 65Department of Biostatistics, University of Washington, Seattle, WA 98195, USA; 66Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children’s Hospital, Columbus, OH 43205, USA; 67Dalla Lana School of Public Health, Toronto, ON M5T 3M7, Canada; 68Institute of Neuroscience, Newcastle University, Newcastle upon Tyne NE2 4HH, UK; 69Institute of Psychiatric Research, Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN 46202, USA; 70Department of Medical and Molecular Genetics and Program in Medical Neuroscience, Indiana University School of Medicine, Indianapolis, IN 46202, USA; 71Office of the President, Tufts University, Medford, MA 02155, USA; 72Division of Molecular Genome Analysis, German Cancer Research Center (Deutsches Krebsforschungszentrum), 69120 Heidelberg, Germany; 73Center for Applied Genomics, The Children’s Hospital of Philadelphia, Philadelphia, PA 19104, USA; 74Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; 75Division of Child and Adolescent Psychiatry, Department of Psychiatry, Miller School of Medicine, University of Miami, Miami, FL 33136, USA; 76Utah Autism Research Program, Department of Psychiatry, University of Utah School of Medicine, Salt Lake City, UT 84108, USA; 77Stella Maris Clinical Research Institute for Child and Adolescent Neuropsychiatry, 56128 Calabrnna, Pisa, Italy; 78Department of Psychiatry, Stanford University Medical School, Stanford, CA 94305, USA; 79Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
80Present address: International Laboratory for Brain, Music, and Sound Research, University of Montreal and McGill University, Montreal, QC H3C 3J7, Canada
81Present address: Department of Psychiatry, Rush University Medical Center, Chicago, IL 60612, USA
82Present address: Department of Psychiatry, Kaiser Permanente, San Francisco, CA 94118, USA
83Present address: Department of Women’s and Children’s Health, Center of Neurodevelopmental Disorders, Karolinska Institutet, 11330 Stockholm, Sweden
84Present address: Department of Psychiatry, University of British Columbia, Vancouver, BC V6T 2A1, Canada
85These 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, ©2014 The Authors
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).
might still perturb connected biological pathways.\textsuperscript{17} Although numerous ASD-associated loci have been recognized to date,\textsuperscript{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\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{5} and Anney et al.\textsuperscript{20,21} Many of the subjects at the recruiting sites were tested for fragile X syndrome (FXS [MIM 300624]) and assessed for chromosomal rearrangements in 1,168 families with at least two ASD-affected individuals.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{5} and Anney et al.\textsuperscript{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.\textsuperscript{21} All diagnostic, clinical, and cognitive assessments were carried out at each contributing site. All data were gathered at a central coordinator 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)\textsuperscript{22} and the Autism Diagnostic Interview, Revised (ADI-R).\textsuperscript{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.\textsuperscript{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),\textsuperscript{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),\textsuperscript{26} the Ontario Colorectal Cancer Case-Control Study,\textsuperscript{27,28} and Health, Aging, and Body Composition (HABC) (Table S1B).\textsuperscript{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\textsuperscript{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)\textsuperscript{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 control subject of two published data sets: (1) 17 de novo CNVs detected in 57 out of 1,023 individuals from the Simons Simplex Collection and (2) 59 de novo CNVs detected in 57 out of 2,623 Icelandic control trios. The clinical relevance of CNVs was interpreted according to the American College of Medical Genetics guidelines 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), 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; 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.

5. Genes found to interact with fragile X mental retardation protein (FMRP).

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) 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 Analyses**

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 a 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, Bioconductor, NETBAG, and DAPPLE.

**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) 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 to detect rare CNVs 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. We used the same pipeline to analyze 2,640 control individuals of European ancestry who were genotyped with the same array platforms. 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, prompting us to assess this further. After screening 2,096 trios (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%. 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) and in control trios (0.55 Mb, p = 0.01). 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 (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...
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\(^3\)\(^,\)\(^3\(^1\)) (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 S5 A–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\(^1\)) 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,\(^10\) ID,\(^40\) and schizophrenia.\(^31\) 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).

### 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

<table>
<thead>
<tr>
<th>Type</th>
<th>Group Size</th>
<th>No. of Rare Genic CNVs</th>
<th>No. of Genes Intersected by Rare CNVs</th>
<th>Baseline Gene Rate (Control) ( ^a )</th>
<th>Case-Control Gene Ratio</th>
<th>( p_{corr} ) b</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>all</td>
<td>6,859</td>
<td>6,745</td>
<td>3.55</td>
<td>1.41</td>
<td>0.000001*</td>
</tr>
<tr>
<td>Deletions</td>
<td>all</td>
<td>2,946</td>
<td>2,804</td>
<td>1.23</td>
<td>1.40</td>
<td>0.00049*</td>
</tr>
<tr>
<td>Duplications</td>
<td>all</td>
<td>3,913</td>
<td>5,217</td>
<td>2.32</td>
<td>1.41</td>
<td>0.00001*</td>
</tr>
<tr>
<td>All 30–500 kb</td>
<td>all</td>
<td>6,307</td>
<td>5,163</td>
<td>2.89</td>
<td>1.07</td>
<td>0.03628*</td>
</tr>
<tr>
<td>&gt;500 kb</td>
<td>all</td>
<td>552</td>
<td>2,491</td>
<td>0.66</td>
<td>2.88</td>
<td>0.00001*</td>
</tr>
<tr>
<td>&gt;1 Mb</td>
<td>all</td>
<td>187</td>
<td>1,337</td>
<td>0.26</td>
<td>4.48</td>
<td>0.00001*</td>
</tr>
<tr>
<td>Deletions 30–500 kb</td>
<td></td>
<td>2,795</td>
<td>2,014</td>
<td>1.07</td>
<td>1.07</td>
<td>0.20110</td>
</tr>
<tr>
<td>&gt;500 kb</td>
<td>all</td>
<td>151</td>
<td>947</td>
<td>0.16</td>
<td>3.60</td>
<td>0.00051*</td>
</tr>
<tr>
<td>&gt;1 Mb</td>
<td>all</td>
<td>63</td>
<td>647</td>
<td>0.08</td>
<td>4.58</td>
<td>0.02289*</td>
</tr>
<tr>
<td>Duplications 30–500 kb</td>
<td></td>
<td>3,512</td>
<td>3,934</td>
<td>1.83</td>
<td>1.08</td>
<td>0.03750*</td>
</tr>
<tr>
<td>&gt;500 kb</td>
<td>all</td>
<td>401</td>
<td>1,896</td>
<td>0.50</td>
<td>2.64</td>
<td>0.00026*</td>
</tr>
<tr>
<td>&gt;1 Mb</td>
<td>all</td>
<td>124</td>
<td>890</td>
<td>0.18</td>
<td>4.43</td>
<td>0.00036*</td>
</tr>
</tbody>
</table>

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.

\( ^b \)Genomic 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.

**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
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 \( \text{SHANK3} \) [MIM 606230]), Smith-Magenis syndrome (MIM 182290, 17p11.2 deletion including \( \text{SHANK3} \) [MIM 606230]), and so on.

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 \( \geq 30 \) kb CNVs affecting an ASD- and/or ID-associ-
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.
RAII [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,44 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 pathogenic events (59/1,486) or 3% after exclusion of exonic events (p = 0.007). Notably, females were over-represented 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^{-5}; 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.44 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 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...
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 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 and among de novo events in the Simons Simplex Collection. 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., GN13 [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 telecephalon 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 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 FRMP targets or PSQ genes. Compared to all other genes within de novo CNVs (or deletion CNVs only), genes in the network exhibited a significantly higher pH1 (Wilcoxon rank sum \( p = 7.07 \times 10^{-8} \)), and 55% (59/107 [6 without information]) had a pH1 > 0.35. A similar NETBAG analysis of de novo CNVs in control subjects did not yield significant results.46

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 S5A) 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], DLD1 [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).
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; (3) 31 genes with hemizygous LoF SNVs on the X chromosome of male ASD subjects and not observed in male control subjects; and (4) 92 ASD-implicated genes previously described as autosomal dominant, X-linked dominant, or X-linked recessive in males.

(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...
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. Three other genes affected by de novo CNVs (also picked by NETBAG)—CHD2, SYNGAP1, and SYCRIP—are also affected by LoF de novo SNVs in ID, and two (SYNGAP1 and DPYD [MIM 612779]) are altered in schizophrenia. CHD2 (discussed above) and SYNGAP1 (encoding a Ras/Rap GTP-activating protein) are both involved in autosomal-dominant ID, ASD, and epilepsy.

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 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 CNVs, and 41/122 (33.6%) de novo LoF SNVs. 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 pH (median pH = 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...
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 times the interquartile range. “Genome” indicates all 16,781 genes with an available pH 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, and our foundational data support this general hypothesis. This same phenomenon has recently been shown for SHANK and the 16p13.11 CNV.

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, 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. Other chromatin regulator genes, such as MBDS (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, is elevated in FXS mouse knockouts 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.

Our findings implicate many ASD candidate genes altered by de novo, inherited, or X-linked CNVs (e.g., SETDS, 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.
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 Hilb rand 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-CHIR 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 Integenix 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 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
- DECIPHER, http://decipher.sanger.ac.uk
- European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations (ECARUCA), http://www.ecaruca.net
- MutationTaster, http://www.mutationtaster.org
- 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


