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

Dalila Pinto, Alistair T. Pagnamenta, Lambertus Klei, Richard Anney, Daniele Merico, Regina Regan, Judith Conroy, Tiago R. Magalhaes, Catarina Correia, Brett S. Abrahams, et al.

▶ **To cite this version:**

Dalila Pinto, Alistair T. Pagnamenta, Lambertus Klei, Richard Anney, Daniele Merico, et al.. Functional impact of global rare copy number variation in autism spectrum disorders.. Nature, 2010, 466 (7304), pp.368-72. 10.1038/nature09146 . inserm-00521387

HAL Id: inserm-00521387

https://inserm.hal.science/inserm-00521387

Submitted on 10 Dec 2010

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.

Functional Impact of Global Rare Copy Number Variation in Autism Spectrum Disorder

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

[1]The Centre for Applied Genomics and Program in Genetics and Genomic Biology, The Hospital for Sick Children, Toronto, Ontario, M5G 1L7, Canada. [2]Wellcome Trust Centre for Human Genetics, University of Oxford, OX3 7BN, UK. [3] Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh 15213, Pennsylvania, USA. [4]Autism Genetics Group, Department of Psychiatry, School of Medicine, Trinity College Dublin 8, Ireland. [5]Banting and Best Department of Medical Research, Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto. [6]School of Medicine and Medical Science University College, Dublin 4, Ireland. [7]Instituto Nacional de Saude Dr Ricardo Jorge and Instituto Gulbenkian de Cîencia Lisbon, Portugal. [8]BioFIG – Center for Biodiversity, Functional and Integrative Genomics, Campus da FCUL, C2.2.12, Campo Grande, 1749-016 Lisboa, Portugal. [9]Department of Neurology, University of California - Los Angeles School of Medicine, Los Angeles, California 90095, USA. [10]Hospital Pediatrico de Coimbra, Coimbra, Portugal. [11]Department of Biology, University of Bologna, 40126 Bologna, Italy. [12]Department of Psychiatry, University of Oxford, Warneford Hospital, Headington, Oxford, OX3 7JX, UK. [13]Newcomen Centre, Guy's Hospital, London, SE1 9RT, UK. [14]Stella Maris Institute for Child and Adolescent Neuropsychiatry, 56128 Calambrone (Pisa), Italy. [15]Child and Adolescent Mental Health, University of

Newcastle, Sir James Spence Institute, Newcastle upon Tyne, NE1 4LP, UK. [16]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Goethe University Frankfurt, 60528 Frankfurt, Germany. [17]Department of Child and Adolescent Psychiatry, Institute of Psychiatry, London, SE5 8AF, UK. [18]Human Genetics and Cognitive Functions, Institut Pasteur; University Paris Diderot-Paris 7, CNRS URA 2182, Fondation FondaMental, 75015 Paris, France. [19]Autism Research Unit, The Hospital for Sick Children and Bloorview Kids Rehabilitation, University of Toronto, Toronto, Ontario, M5G 1Z8, Canada. [20]Department of Pediatrics and Psychology, Dalhousie University, Halifax, Nova Scotia, B3K 6R8, Canada. [21]Autism and Communicative Disorders Centre, University of Michigan, Ann Arbor, Michigan, USA. [22]Department of Molecular Physiology and Biophysics, Vanderbilt Kennedy Center, and Centers for Human Genetics Research and Molecular Neuroscience, Vanderbilt University, Nashville, Tennessee 37232, USA. [23]Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. [24]Autism Speaks, USA. [25]Department of Psychiatry, University of North Carolina, Chapel Hill, North Carolina, 27599-3366, USA. [26]Department of Child Psychiatry, University Medical Center, Utrecht, The Netherlands. [27]INSERM U 955, Fondation FondaMental, APHP, Hôpital Robert Debré, Child and Adolescent Psychiatry, 75019 Paris, France. [28]Department of Speech and Hearing Sciences, University of Washington, Seattle, Washington 98195, USA. [29]Disciplines of Genetics and Medicine, Memorial University of Newfoundland, St. John's Newfoundland, Canada. [30]The John Hussman Institute for Human Genomics, University of Miami, Miami, Florida 33101, USA. [31]Division of Psychiatry, McGill University, Montreal, Quebec, Canada. [32]Department of Child and Adolescent Psychiatry, Göteborg University, Göteborg, S41345, Sweden. [33]The Center for Applied Genomics, Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. [34]Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Ontario, L8N 3Z5, Canada. [35]Academic Department of Child Psychiatry, Booth Hall of Children's Hospital, Blackley, Manchester, M9 7AA, UK. [36]Institute for Juvenile Research, Department of Psychiatry, University of Illinois at Chicago, Chicago, Illinois, USA. [37]Department of Pediatrics, Children's Hospital of Philadelphia, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA [38]Division of Molecular Genome Analysis, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany. [39]The Seaver Autism Center for Research and Treatment, Department of Psychiatry, Mount Sinai School of Medicine, New York 10029, USA. [40]Department of Medicine, University of Washington, Seattle, Washington 98195, USA. [41]Autism Genetic Resource Exchange, Autism Speaks, Los Angeles, California 90036-4234, USA. [42]Centre for Integrated Genomic Medical Research, University of Manchester, Manchester, M13 9PT, UK. [43]INSERM U995, Department of Psychiatry, Groupe Hospitalier Henri Mondor-Albert Chenevier, AP-HP; University Paris 12, Fondation FondaMental, Créteil, France. [44]Nathan Kline Institute for Psychiatric Research (NKI), 140 Old Orangeburg Road, Orangeburg, NY 10962, USA. [45]Department of Child and Adolescent Psychiatry, New York University and NYU Child Study Center, 550 First Avenue, NY, NY 10016, USA.[46]Department of Psychiatry, Division of Child and Adolescent Psychiatry and Child Development, Stanford University School of Medicine, Stanford, California 94304, USA. [47]Department of Pediatrics, McMaster University, Hamilton, Ontario, L8N 3Z5, Canada. [48]Centre d'Etudes et de Recherches en Psychopathologie, University de Toulouse Le Mirail, Toulouse 31200, France. [49]Department of Psychiatry, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. [50]Psychiatry Department, University of Utah Medical School, Salt Lake City, Utah 84108, USA. [51]Departments of Psychiatry and Neurology, University of Pittsburgh School of Medicine, 15213. [52]Department of Psychiatry and Behavioural Sciences, University of Washington, Seattle, Washington 98195, USA. [53]Department of Human Genetics, University of California - Los Angeles School of Medicine, Los Angeles, California 90095, USA. [54]Centre for Addiction and Mental Health, Clarke Institute and Department of Psychiatry, University of Toronto, Toronto, Ontario M5G 1X8, Canada. [55]University Department of Child Psychiatry, Athens University, Medical School, Agia Sophia Children's Hospital, 115 27 Athens, Greece. [56]Insitutes of Neuroscience and Health and Society, Newcastle University, Newcastle Upon Tyne, NE1 7RU, UK. [57]Department of Medicine, School of Epidemiology and Health Science, University of Manchester, Manchester, M13 9PT, UK. [58]INSERM U952 and CNRS UMR 7224 and UPMC Univ Paris 06, Paris 75005, France. [59]Carolina Institute for Developmental Disabilities, University of North Carolina at Chapel Hill, North Carolina 27599-3366, USA. [60]MRC Functional Genomics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, United Kingdom. [61]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, London, SE5 8AF, UK. [62]Department of Psychiatry, Washington University in St. Louis, School of Medicine, St. Louis, Missouri 63130, USA. [63]Department of Pediatrics

and Howard Hughes Medical Institute Carver College of Medicine, University of Iowa, Iowa City, Iowa 52242, USA. [64]Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital and The Ohio State University, Columbus, Ohio 43205, USA. [65]Neuropsichiatria Infantile, Ospedale Santa Croce, 61032 Fano, Italy. [66]Child Study Centre, Yale University, New Haven, Connecticut 06520, USA. [67]Department of Psychiatry, Carver College of Medicine, Iowa City, Iowa 52242, USA. [68]Department of Pediatrics, University of Alberta, Edmonton, Alberta, T6G 2J3, Canada. [69]Center for Human Genetics Research, Vanderbilt University Medical Centre, Nashville, Tennessee 37232, USA. [70]Pathology and Laboratory Medicine, University of Pennsylvania, Pennsylvania 19104, USA. [71]Departments of Biostatistics and Medicine, University of Washington, Seattle, Washington 98195, USA. [72]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, M5S 1A1, Canada

*Lead Autism Genome Project Consortium (AGP) investigators who contributed equally to this project.
[X] Deceased
Correspondence should be addressed to S.W.S. (stephen.scherer@sickkids.ca)

The autism spectrum disorders (ASDs) are a group of conditions characterized by impairments in reciprocal social interaction and communication, and the presence of restricted and repetitive behaviors[1]. Individuals with an ASD vary greatly in cognitive development, which can range from above average to intellectual disability (ID)[2]. While ASDs are known to be highly heritable (~90%)[3], the underlying genetic determinants are still largely unknown. Here, we analyzed the genome-wide characteristics of rare (<1% frequency) copy number variation (CNV) in ASD using dense genotyping arrays. When comparing 996 ASD individuals of European ancestry to 1,287 matched controls, cases were found to carry a higher global burden of rare, genic CNVs (1.19 fold, *P*= 0.012), especially so for loci previously implicated in either ASD and/or intellectual disability (1.69 fold, *P*= 3.4x10^{-4}). Among the CNVs, there were numerous *de novo* and inherited events, sometimes in combination in a given family, implicating many novel ASD genes like *SHANK2*, *SYNGAP1*, *DLGAP2* and the X-linked *DDX53-PTCHD1* locus. We also discovered an enrichment of CNVs disrupting functional gene-sets involved in cellular proliferation, projection and motility, and GTPase/Ras signaling. Our results reveal many new genetic and functional targets in ASD that may lead to final connected pathways.

Twin and family studies indicate a predominantly genetic basis for ASD susceptibility and provide support for considering these disorders as a clinical spectrum. Some 5-15% of individuals with an ASD have an identifiable genetic aetiology corresponding to known rare single-gene disorders (e.g., fragile X syndrome) and chromosomal rearrangements (e.g., maternal duplication of 15q11-q13). Rare mutations have been identified in synaptic genes, including *NLGN3*, *NLGN4X*[4] and SHANK3[5], and microarray studies have revealed copy number variation (CNV) as risk factors[6]. CNV examples include *de novo* events observed in 5-10% of ASD cases[7-9], *de novo* or inherited hemizygous deletions and duplications of 16p11.2[9-11] and *NRXN1*[7], and exceptionally rare homozygous deletions in consanguineous families[12]. Genome-wide

association studies using single nucleotide polymorphisms (SNPs) have highlighted two potential ASD risk loci at 5p14.1[13] and 5p15.2[14], but these data suggest common variation will account for only a small proportion of the heritability in ASD.

To further delineate the contribution of rare genomic variants to autism we genotyped 1,275 ASD cases and their parents using the Illumina Infinium 1M-single SNP-microarray (Fig.1). A set of 1,981 controls used for comparison studies was genotyped on the same platform[15] and both data sets were subjected to the same quality control (QC) procedures. Ultimately, we analyzed 996 ASD cases (876 trios) and 1,287 controls of European ancestry (EA) to minimize confounds due to population differences (Supplementary Fig. 1-2 and Supplementary Table 1)[16].

Two CNV prediction algorithms (QuantiSNP[17] and iPattern (unpublished)) and additional extensive QC were used to establish a stringent dataset of non-redundant CNVs called by both algorithms in an individual (Fig. 1, Supplementary Tables 1-3 and Supplementary Fig. 3). This stringent dataset of 5,478 rare CNVs in 996 cases and 1,287 controls of EA (Supplementary Table 4) had the following characteristics: (i) CNV present at <1% frequency in the total sample (cases and controls), (ii) CNV ≥30 kb in size (because >95% of these could be confirmed) and (iii) all CNVs further verified using combined evidence from the PennCNV algorithm[18] and child-parent intensity fold-changes, genotype proportions (to verify deletions) and visual inspection (for chromosome-X).

We assessed the impact of rare CNV in cases compared to controls using three primary measures of CNV burden: the number of CNVs per individual, the estimated CNV size, and the number of genes affected by CNVs (Table 1). No significant difference was found in the former two measures (Supplementary Tables 4a and 5), even after controlling for fine-level ancestry differences by pair-matching cases and controls (Supplementary Information)[16]. In contrast, we discovered a significant increase in the number of genes intersected by rare CNV in cases when focusing on gene-containing segments (1.19-fold increase, empirical P= 0.012). This ASD association with genic CNV was stronger for deletions (1.26-fold increase, empirical P= 8.0×10^{-3}). These differences remained after we further controlled for potential case-control differences that could be present due to biological differences or technical biases. Restricting our analysis to autosomal CNVs (ie. after removing CNVs located on chromosome X) resulted in a consistent enriched gene count in ASD cases compared to controls. Single-occurrence CNV deletions had increased rates in ASD over controls, suggesting some could be pathogenic.

We then examined parent-child transmission and confirmed that 5.7% (50/876) of ASD cases had at least one *de novo* CNV with >0.6% carrying two or more *de novo* events (Supplementary Tables 4a, 6 and 7). The *de novo* CNV rate in our simplex and multiplex families was 5.6% (22/393) and 5.5% (19/348), respectively, in contrast with previous studies showing a higher rate in simplex families[8,9]. A total of 226 validated *de*

novo (7) and inherited (219) CNVs not observed in controls and affecting single genes were found (Supplementary Table 8).

Numerous novel candidate ASD loci such as *SHANK2*, *SYNGAP1*, and *DLGAP2*, were identified based on the observation that *de novo* CNV affects these genes in cases and not controls (Supplementary Table 6). The relatedness of *SHANK2* to the causal ASD gene *SHANK3*[5], involvement of *SYNGAP1* in ID[19], and interaction of DLGAP family proteins with SHANK proteins[20] further support their role in ASDs. Maternally-inherited X-linked deletions at *DDX53/PTCHD1* (7 cases) implicated this locus in ASD. We tested an additional 3,677 EA controls (Fig. 1) and again found no CNV at these genes, and *DDX53/PTCHD1* emerged as a significant ASD risk factor (P= 3.1x10^{-3} with the initial 1,287 controls; P= 3.6x10^{-6} with combined controls; Supplementary Fig. 4).

Association studies of individual rare CNV often have insufficient power to discriminate benign from disease-causing variants. Here, we assessed whether genes and CNVs previously associated with ASD and/or ID were enriched in cases compared with controls, in order to help identify pathogenic events. We defined three gene-lists based on evidence from previous studies of their involvement in ASDs (Supplementary Table 9): (i) 'ASD implicated' list consisting of 36 disease genes and 10 loci strongly implicated in ASD and identified in subjects with ASD or ASD and ID; (ii) 'ID' consisting of 110 disease genes and 17 loci implicated in ID but not yet in ASD; and, (iii) 'ASD candidates' including 103 genes from previous studies of common and rare variants.

We observed a higher proportion of cases with rare CNVs overlapping 'ASD implicated' disease genes compared to controls (4.3% versus 2.3%, Fisher exact test P= 5.4x10^{-3}; Fig. 2a), corresponding to a significant enrichment for genes in this set (OR= 1.8; 95% CI 1.3–2.6, empirical P= 2.6x10^{-3}; Fig. 2b, see also Supplementary Information). This effect was stronger for duplications, which may also disrupt genes (OR= 2.3; 95% CI 1.4–3.8, empirical P= 9.4x10^{-4}). Enrichment was also found for rare CNVs overlapping ID genes, more notably for deletions (OR= 2.1; 95% CI 1.1–4.2, empirical P= 5.3x10^{-2}). In contrast, there was no evidence of enrichment among case-CNVs compared to control-CNVs for genes in the 'ASD candidates' set (empirical P >0.3). When the two disease gene-sets 'ASD implicated' and 'ID' were combined, we observed 7.6% of cases with rare CNVs preferentially affecting ASD/ID genes compared to 4.5% in controls (Fisher exact test P= 1.2x10^{-3}, Fig.2a). The observed enrichments did not change when potential case-control genome-wide differences for CNV rate and size were considered.

Our global analyses of these putative pathogenic loci use somewhat subjective boundaries for CNV overlap. Manual inspection of the data yields more accurate results. After eliminating CNVs that are less likely to have an aetiological role (heterozygous CNVs that disrupt autosomal recessive loci, events outside the critical region of overlap of genomic disorders, X-linked genes in females inherited from non-ASD fathers, duplications inherited from non-ASD parents, and intronic CNVs in *NRXN1*), 25 CNVs remained in

the ASD group, compared to only four in the controls (P= 3.6x10^{-6}; Supplementary Table 10). Moreover, the latter four CNVs were all duplications at 1q21.1, 16p11.2 or 22q11.2, loci known to exhibit incomplete penetrance and variable expressivity[6]. The population attributable risk provided by the combination of all ASD-CNVs that overlap ASDs and/or ID genes is estimated to be 3.3% (Supplementary Table 11). We also identified rare *de novo* chromosomal abnormalities and large CNVs likely to be aetiologic (Supplementary Table 10).

We then tested for functional enrichment of gene-sets among those genes affected by CNVs to identify biological processes involved in ASD (Fig. 3). Here, the term gene-set refers to groups of genes that share a common function or operate in the same pathway. Such a functional enrichment mapping approach can combine single-gene effects into meaningful groups[21].

We compiled comprehensive collections of gene-sets (Supplementary Table 12) and used the Fisher's exact test to assess which gene-sets were more frequently affected by rare CNV events in ASD cases compared to controls. An estimate of the false-discovery rate (FDR) at each gene-set was obtained by random permutation of case and control labels (Supplementary Information). To visualize enriched gene-sets, overlap scores were used to graphically organize these sets into a functional enrichment map (or network) using Cytoscape[22]. We identified the 'seed' genes for the network at an FDR q-value of 5% and further relaxed the thresholds to 12.5% to better capture the network topology[23].

Using these criteria only deletions were found significantly enriched in gene-sets in cases over controls (Supplementary Fig. 5), consistent with the global burden results (Table 1). Specifically, 76 gene-sets affected by deletions (2.18% of sets tested) were found enriched and used to construct a functional map (Figure 3a, Supplementary Fig. 6-7). We tested for possible bias, including measures of CNV size and number for cases versus controls per gene-set, as well as genome proximity, but no differences were found that might explain the observed enrichments (Supplementary Fig. 8-9).

We identified enrichments in gene-sets known to be involved in ASDs and also discovered new candidate ASD pathways (Fig. 3a, Supplementary Table 13). For example, gene-sets involved in cell and neuronal development and function (including projection, motility, and proliferation) previously reported in ASD-associated phenotypes, were identified[24]. Novel observations included gene-sets involved in GTPase/Ras signaling, with component Rho GTPases known to be involved in regulating dendrite and spine plasticity and associated with ID. We also found a tentative link to sets in the kinase activity/regulation functional group where only minorities of these sets meet a stringent 5% FDR q-value threshold (Supplementary Fig. 10).

We further assessed the relationship of our functional enrichment map with known ASD/ID genes (Fig. 3b, Supplementary Fig. 11) and found genes enriched in sets linked to microtubule cytoskeleton, glycosylation and CNS development/adhesion[25]. The two groups of genes found enriched in deletions (Fig.

3a) also displayed connectivity to the ASD/ID disease gene-sets, either directly or through intermediates (Fig 3b, Supplementary Fig. 12). Although ASD genes appear to be enriched in different subsets of genes compared to ID-only genes, we cannot discount the possibility that this is the result of selection bias, and we expect that more ID genes may yet be linked to ASD.

Our findings provide strong support for the involvement of multiple rare genic CNVs, both genome-wide and at specific loci, in ASD. These findings, similar to those recently described in schizophrenia[26], suggest that at least some of these ASD-CNVs (and the genes that they affect) are under purifying selection[27]. Genes previously implicated in ASD by rare variant findings have pointed to functional themes in ASD pathophysiology[6,28]. Molecules such as NRXN1, NLGN3/4X and SHANK3, localized presynaptically or at the post-synaptic density (PSD), highlight maturation and function of glutamatergic synapses. Our data reveal *SHANK2*, *SYNGAP1* and *DLGAP2* as new ASD loci, which also encode proteins in the PSD. We also found ID genes to be important in ASD[29]. Furthermore, our functional enrichment map identifies new groups such as GTPase/Ras, effectively expanding both the number and connectivity of modules that may be involved in ASD. The next steps will be to relate defects or patterns of alterations in these groups to ASD endophenotypes. The combined identification of higher-penetrance rare variants and new biological pathways, including those identified in this study, may broaden the targets amenable to genetic testing and therapeutic intervention.

Methods Summary

Cases were classified using the Autism Diagnostic Interview-Revised (ADI-R) and Autism Diagnostic Observation Schedule (ADOS) instruments and those with known karyotypic abnormalities or genetic disorders were excluded. Informed consent was obtained from all families and procedures had approval from institutional review boards. DNA was obtained from blood or buccal-swabs (73% of cases; 75% of controls) or cell-lines (22% of cases; 25% of controls) (in 5% of cases the DNA source was not identified). The 1,287 EA controls passing all QC-filters included 1,261 individuals recruited as controls for the study of addiction (SAGE)[15] and 26 HapMap samples (from Illumina). An additional 3,677 EA controls from three separate studies genotyped on other platforms were also used. Raw data from ASD family (Accession pending) and SAGE control (Accession: phs000092.v1.p1) genotyping are at NCBI dbGAP. CNVs were analysed using PLINK v1.07[30], R stats and custom scripts. Primary analyses were robust to potential systematic measurement differences between cases and controls; it was not possible to control for site but we controlled for the overall extent and number of CNVs for all burden comparisons, and obtained a consistent enriched gene count in ASD cases compared to controls.

References

1 Veenstra-Vanderweele, J., Christian, S.L., & Cook, E.H., Jr., Autism as a paradigmatic complex genetic disorder. *Annu Rev Genomics Hum Genet* 5, 379-405 (2004).

2 Chakrabarti, S. & Fombonne, E., Pervasive developmental disorders in preschool children: confirmation of high prevalence. *Am J Psychiatry* 162 (6), 1133-1141 (2005).

3 Bailey, A. *et al.*, Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med* 25 (1), 63-77 (1995).

4 Jamain, S. *et al.*, Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet* 34 (1), 27-29 (2003).

5 Durand, C.M. *et al.*, Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat Genet* 39 (1), 25-27 (2007).

6 Cook, E.H., Jr. & Scherer, S.W., Copy-number variations associated with neuropsychiatric conditions. *Nature* 455 (7215), 919-923 (2008).

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

8 Sebat, J. *et al.*, Strong association of de novo copy number mutations with autism. *Science* 316 (5823), 445-449 (2007).

9 Marshall, C.R. *et al.*, Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* 82 (2), 477-488 (2008).

10 Weiss, L.A. *et al.*, Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* 358 (7), 667-675 (2008).

11 Kumar, R.A. *et al.*, Recurrent 16p11.2 microdeletions in autism. *Hum Mol Genet* 17 (4), 628-638 (2008).

12 Morrow, E.M. *et al.*, Identifying autism loci and genes by tracing recent shared ancestry. *Science* 321 (5886), 218-223 (2008).

13 Wang, K. *et al.*, Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* 459 (7246), 528-533 (2009).

14 Weiss, L.A., Arking, D.E., Daly, M.J., & Chakravarti, A., A genome-wide linkage and association scan reveals novel loci for autism. *Nature* 461 (7265), 802-808 (2009).

15 Bierut, L.J. *et al.*, A genome-wide association study of alcohol dependence. *Proc Natl Acad Sci U S A* 107 (11), 5082-5087 (2010).

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

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

18 Wang, K. *et al.*, PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 17 (11), 1665-1674 (2007).

19 Hamdan, F.F. *et al.*, Mutations in SYNGAP1 in autosomal nonsyndromic mental retardation. *N Engl J Med* 360 (6), 599-605 (2009).

20 Romorini, S. *et al.*, A functional role of postsynaptic density-95-guanylate kinase-associated protein complex in regulating Shank assembly and stability to synapses. *J Neurosci* 24 (42), 9391-9404 (2004).

21 O'Dushlaine, C. *et al.*, Molecular pathways involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia and bipolar disorder susceptibility. *Mol Psychiatry* [Epub ahead of print] (2010).

22 Cline, M.S. *et al.*, Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* 2 (10), 2366-2382 (2007).

23 Subramanian, A. *et al.*, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102 (43), 15545-15550 (2005).
24 Meechan, D.W., Tucker, E.S., Maynard, T.M., & LaMantia, A.S., Diminished dosage of 22q11 genes disrupts neurogenesis and cortical development in a mouse model of 22q11 deletion/DiGeorge syndrome. *Proc Natl Acad Sci U S A* 106 (38), 16434-16445 (2009).
25 Wegiel, J. *et al.*, The neuropathology of autism: defects of neurogenesis and neuronal migration, and dysplastic changes. *Acta Neuropathol* [Epub ahead of print] (2010).
26 International Schizophrenia Consortium, Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 455 (7210), 237-241 (2008).
27 Conrad, D.F. *et al.*, Origins and functional impact of copy number variation in the human genome. *Nature* 464 (7289), 704-712 (2010).
28 Glessner, J.T. *et al.*, Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 459 (7246), 569-573 (2009).
29 Skuse, D.H., Rethinking the nature of genetic vulnerability to autistic spectrum disorders. *Trends Genet* 23 (8), 387-395 (2007).
30 Purcell, S. *et al.*, PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81 (3), 559-575 (2007).

Supplementary Information is linked to the online version of the paper.

Acknowledgements The authors gratefully acknowledge the families participating in the study and the main funders of the Autism Genome Project Consortium (AGP): Autism Speaks (USA), the Health Research Board (HRB; Ireland), The Medical Research Council (MRC; UK), Genome Canada/Ontario Genomics Institute, and the Hilibrand Foundation (USA). Additional support for individual groups was provided by the US National Institutes of Health (NIH grants: HD055751, HD055782, HD055784, HD35465, MH52708, MH55284, MH57881, MH061009, MH06359, MH066673, MH080647, MH081754, MH66766, NS026630, NS042165, NS049261), the Canadian Institutes for Health Research (CIHR), Assistance Publique-Hôpitaux de Paris (France), Autism Speaks UK, Canada Foundation for Innovation/Ontario Innovation Trust, Deutsche Forschungsgemeinschaft (grant: Po 255/17-4) (Germany), EC Sixth FP AUTISM MOLGEN, Fundação Calouste Gulbenkian (Portugal), Fondation de France, Fondation FondaMental (France), Fondation Orange (France), Fondation pour la Recherche Médicale (France), Fundação para a Ciência e Tecnologia (Portugal), the Hospital for Sick Children Foundation and University of Toronto (Canada), INSERM (France), Institut Pasteur (France), the Italian Ministry of Health (convention 181 of 19.10.2001), the John P Hussman Foundation (USA), McLaughlin Centre (Canada), Ontario Ministry of Research and Innovation (Canada), the Seaver Foundation (USA), the Swedish Science Council, The Centre for Applied Genomics (Canada), the Utah Autism Foundation (USA) and the Wellcome Trust core award 075491/Z/04 (UK). D.P. is supported by fellowships from the Royal Netherlands Academy of Arts and Sciences (TMF/DA/5801) and the Netherlands Organization for Scientific Research (Rubicon 825.06.031).

Author contributions D.P., J.D.B., R.M.C., E.H.C., H.C., M.C., B.D., S.E., L.G., D.H.G., M.G., J.L.H., J.H., J.M., A.P.M., J.I.N.Jr., A.D.P., M.A.P.-V., G.D.S., P.S., A.M.V., V.J.V., E.M.W., J.S.S., C.B. and S.W.S. were leading contributors in the design, analysis and writing of this study.

Correspondence and requests for materials should be addressed to S.W.S. (stephen.scherer@sickkids.ca).

Table 1. Global burden of genic rare CNVs in cases vs. controls[1]

Type	Classification	Total CNVs (n)	*P*	Case/Ctrl. ratio	Baseline rate (Ctrl.)	P_{corr}
	None					
All	All	5,478	**0.012**	1.19	3.59	**0.003**
Deletions only	All	2,757	**0.008**	1.26	1.08	**0.006**
Duplications only	All	2,721	0.072	1.16	2.51	**0.015**
	CNV frequency					
All	2-6x	1,831	0.058	1.26	1.03	0.062
	1x	1,419	0.375	1.03	0.83	0.180
Deletions only	2-6x	1,094	**0.004**	1.57	0.43	**0.011**
	1x	880	**0.036**	1.26	0.30	**0.036**
Duplications only	2-6x	977	0.203	1.16	1.05	0.118
	1x	896	0.749	0.92	0.72	0.442
	CNV size					
All	30 – 500 kb	5,086	0.313	1.03	2.72	0.181
	≥500 kb	392	**0.005**	1.69	0.88	**0.005**
Deletions only	30 – 500 kb	2,645	**0.004**	1.24	0.85	**0.004**
	≥500 kb	112	0.209	1.32	0.23	0.143
Duplications only	30 – 500 kb	2,441	0.801	0.93	1.86	0.518
	≥500 kb	280	**0.007**	1.82	0.65	**0.012**

[1]Gene-count in cases vs. controls. Total burden for genes intersected by CNVs in ASD cases (n = 996) compared to controls (n = 1,287). Gene coordinates were defined by the RefSeq boundaries plus a 10 kb region on either side. All genomic analyses used NCBI Build 36. 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. Significant differences ($P \leq 0.05$) are bold. Analyses were further stratified according to CNV type ('Deletions only' and 'Duplications only') and frequency (single occurrences (1x) or CNVs observed two to six times in the total sample (2-6x)). Note that for the two lower-frequency groups, the 'Deletions only' and 'Duplications only' counts are not expected to sum to the 'All' count (see ref. [26] and Supplementary Information).

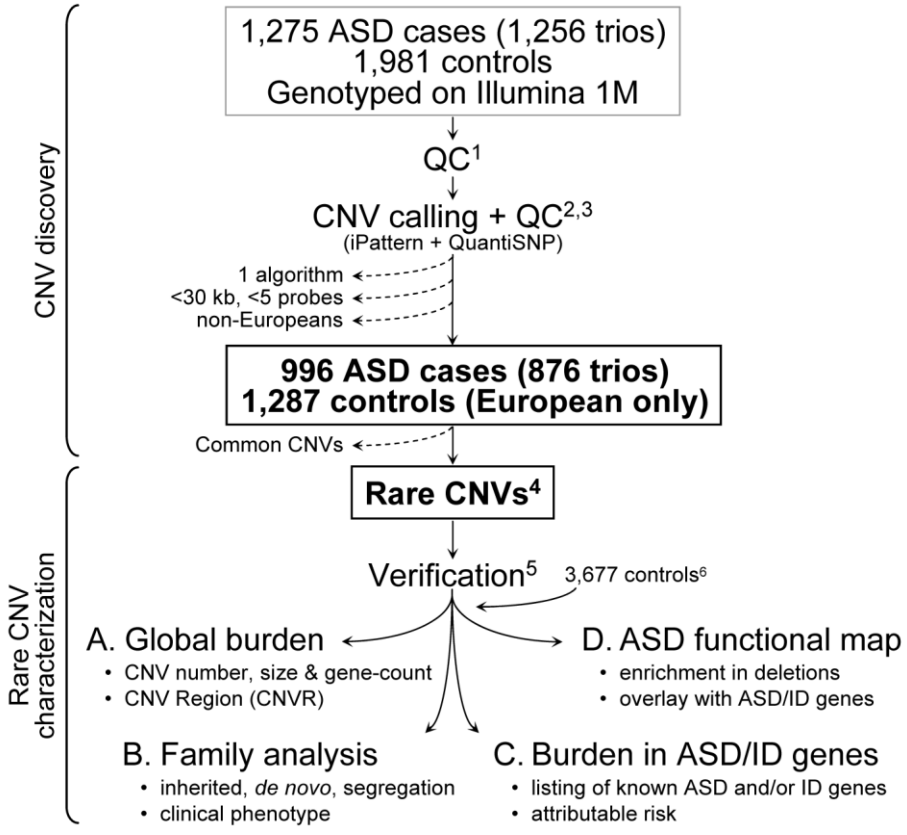

Figure 1: CNV discovery and characterization. Comprehensive procedures were used to identify the rare CNV dataset (boxed). Dashed arrows indicate CNVs not included in downstream analyses. [1]SNP and intensity quality control (QC) with ancestry estimation. [2]QC for CNV calls. [3]Pilot validation experiments using quantitative-PCR were used to evaluate the false discovery-rate. [4]Rare CNVs in samples of EA ancestry were defined as ≥30 kb in size and present in the total sample set at a frequency <1%. 70/996 (17%) of ASD cases were analyzed on different lower-resolution arrays in previous studies[9,10,28]. [5]All CNVs were computationally verified and at least 40% of case-CNVs were also experimentally validated by qPCR and/or independent Agilent or other SNP microarrays. [6]3,677 additional EA controls were used to test specific loci from the primary burden analyses. Additional details are in the Methods Summary and Supplementary Information

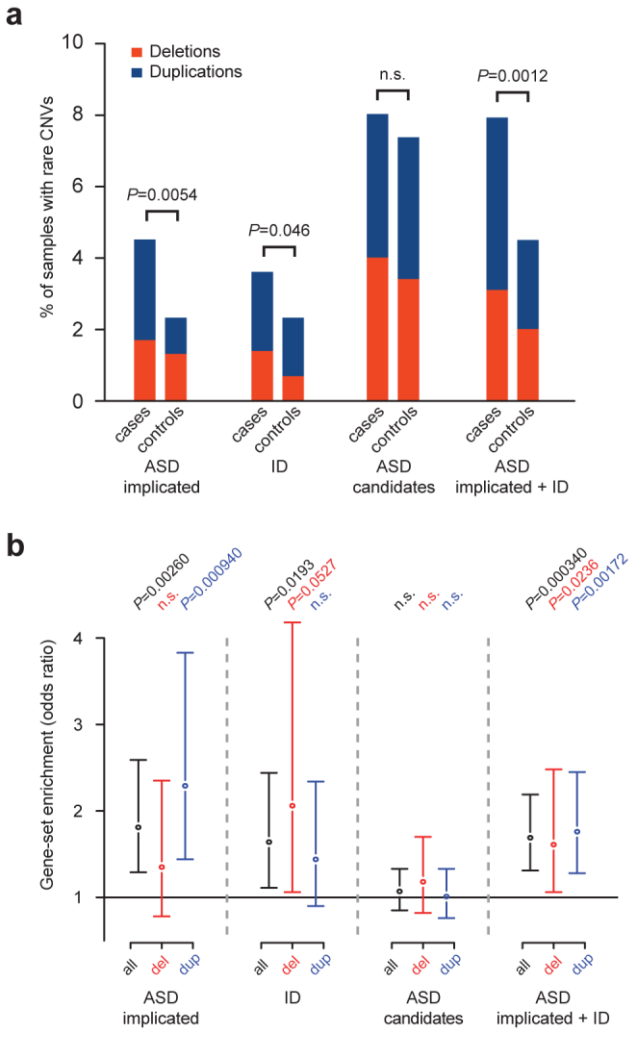

Figure 2. CNV burden in known ASD and/or ID genes. a, Proportion of samples with CNVs overlapping genes and loci known to be associated in ASD with or without ID or ID only, as well as published candidate genes and loci for ASD (Supplementary Table 9). To select for CNVs with maximal impact, they needed to intersect genes, and overlap the target loci by ≥50% of their length. Fisher's exact test *P*-values for significant differences ($P \leq 0.05$, one tailed) are shown. **b**, enrichment analysis for genes overlapped by rare CNVs in cases compared to controls for the three gene-sets in panel a, relative to the whole genome. Odds ratio (OR) and 95% confidence intervals are given for each gene set. Empirical *P*-values for gene-set enrichment are indicated above each OR. All *P*-values <0.1 are listed.

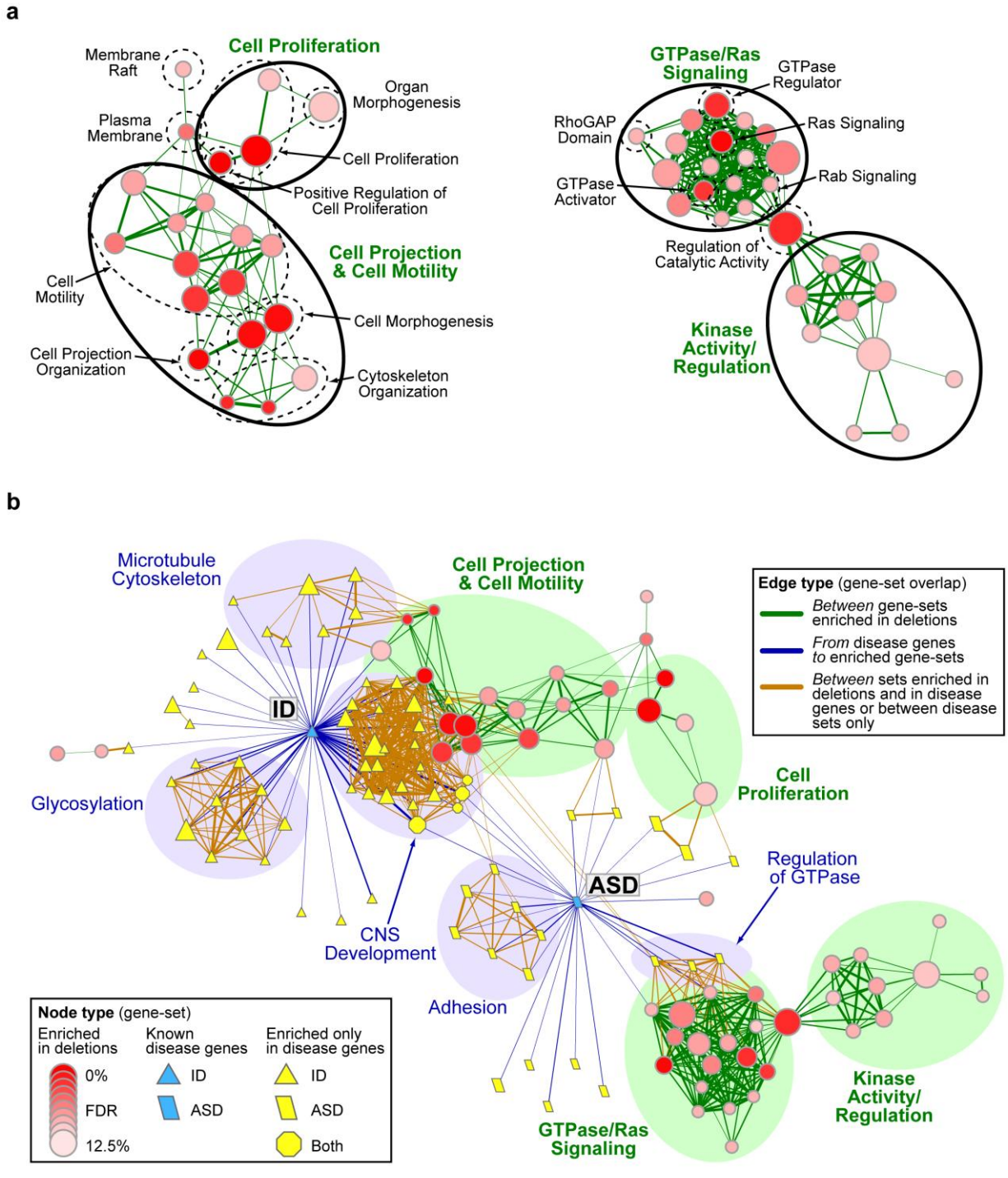

Figure 3. A functional map of ASD. Enrichment results were mapped into a network of gene-sets (nodes) related by mutual overlap (edges), where the color (red, blue, or yellow) indicates the class of gene-set. Node size is proportional to the total number of genes in each set and edge thickness represents the number of overlapping genes between sets. **a,** Gene-sets enriched for deletions are shown (red) with enrichment significance (FDR q-value) represented as a node color gradient. Groups of functionally related gene-sets are circled and labeled (groups, solid line; sub-groups, dashed line). **b,** An expanded enrichment map shows the relationship between gene-sets enriched in deletions (panel a) and sets of known ASD/ID genes. Node color hue represents the class of gene-set (i.e. enriched in deletions, red; known disease genes (ie. ASD and/or ID genes), blue; enriched only in disease genes, yellow). Edge color represents the overlap between gene-sets enriched in deletions (green), from disease genes to enriched sets (blue), and between sets enriched in deletions and in disease genes or between disease gene-sets only (orange). The major functional groups are highlighted by filled circles (enriched in deletions, green; enriched in ASD/ID, blue).