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. Individuals with an ASD vary greatly in cognitive development, which can range from above average to intellectual disability. Although ASDs are known to be highly heritable (~90%), the underlying genetic determinants are still largely unknown. Here we analysed the genome-wide characteristics of rare (~<1% frequency) copy number variation 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 copy number variants (CNVs) (1.19 fold, $P = 0.012$), especially so for loci previously implicated in either ASD and/or intellectual disability (1.69 fold, $P = 3.4 \times 10^{-4}$). Among the CNVs there were numerous de novo and inherited events, sometimes in combination in a given family, implicating many novel ASD genes such as SHANK2, SYNGAP1, DLGAP2 and the X-linked DDX35–PTCHD1 locus. We also discovered an enrichment of CNVs disrupting functional gene sets involved in cellular proliferation, projection and motility, and GTPase/Ras signalling. 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 (for example, fragile X syndrome) and chromosomal rearrangements (for example, maternal duplication of 15q11-q13). Rare mutations have been identified in synaptic genes, including NLGN3, NLGN4X (ref. 4) and SHANK3 (ref. 5), and microarray studies have revealed copy number variation (CNV) as risk factors. CNV examples include de novo events observed in 5–10% of ASD cases $^{2, 4}$, de novo or inherited hemizygous deletions and duplications of 16p11.2 (refs 9–11) and NRXN1 (ref. 7), and exceptionally rare homozygous deletions in consanguineous families. Genome-wide association studies using single nucleotide polymorphisms (SNPs) have highlighted two potential ASD risk loci at 5p14.1 (ref. 13) and 5p15.2 (ref. 14), but these data indicate that common variation will account for only a small proportion of the heritability in ASD.

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

Two CNV prediction algorithms (QuantiSNP and iPatter (unpublished data)) and additional extensive quality control procedures were used to establish a stringent data set of non-redundant CNVs called by both algorithms in an individual (Fig. 1, Supplementary Tables 1–3 and Supplementary Table 4). This stringent data set of 5,478 rare CNVs in 996 cases and 1,287 controls of European ancestry had the following characteristics: (1) CNVs present at <1% frequency in the total sample (cases and controls); (2) CNV ≥30 kb in size (because >95% of these could be confirmed); and (3) all CNVs further verified using combined evidence from the PennCNV algorithm 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)\(^\text{16}\). 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 \(P = 8.0\times 10^{-5}\)). 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 (that is, after removing CNVs located on chromosome X) also resulted in a consistent enriched gene count in ASD cases compared to controls. Single-occurrence CNV deletions had increased rates in ASD cases over controls, indicating that some could be pathogenic.

We then examined parent–child transmission and confirmed that 5.7\% (50 out of 876) of ASD cases had at least one \textit{de novo} CNV with \(>0.6\%\) carrying two or more \textit{de novo} events (Supplementary Tables 4a, 6 and 7). The \textit{de novo} CNV rate in our simplex and multiplex families was 5.6\% (22 out of 393) and 5.5\% (19 out of 348), respectively, in contrast with previous studies showing a higher rate in simplex families\(^\text{5,6}\). A total of 226 validated \textit{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 \textit{SHANK2}, \textit{SYNGAP1} and \textit{DLGAP2} were identified on the basis of the observation that \textit{de novo} CNV affects these genes in cases but not controls (Supplementary Table 6). The relatedness of \textit{SHANK2} to the causal ASD gene \textit{SHANK3} (ref. 5), involvement of \textit{SYNGAP1} in intellectual disability\(^\text{19}\), and interaction of \textit{DLGAP} family proteins with \textit{SHANK} proteins\(^\text{20}\) further support their role in ASDs. Maternally inherited X-linked deletions at \textit{DDX53–PTCHD1} (7 cases) implicate this locus in ASD. We tested an additional 3,677 European ancestry controls (Fig. 1) and again found no CNV at these genes, and \textit{DDX53–PTCHD1} emerged as a significant ASD risk factor \((P = 3.1 \times 10^{-4})\) with the initial 1,287 controls; \(P = 3.6 \times 10^{-6}\) with combined controls; Supplementary Fig. 4).

**Figure 2 | CNV burden in known ASD and/or intellectual disability genes.** a, Proportion of samples with CNVs overlapping genes and loci known to be associated in ASD with or without intellectual disability (ID) or intellectual disability 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 \(\geq 50\%\) of their length. Fisher’s exact test \(P\)-values for significant differences \((P \leq 0.05\), one tailed) are shown. NS, not significant. b, Enrichment analysis for genes overlapped by rare CNVs in cases compared to controls for the three gene sets in a, relative to the whole genome. Odds ratio and 95% confidence intervals are given for each gene set. Empirical \(P\)-values for gene-set enrichment are indicated above each odds ratio. All \(P\)-values <0.1 are listed.
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 intellectual disability 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): (1) 'ASD implicated' list consisting of 36 disease genes and 10 loci strongly implicated in ASD and identified in subjects with ASD or ASD and intellectual disability; (2) 'intellectual disability' consisting of 110 disease genes and 17 loci implicated in intellectual disability but not yet in ASD; and (3) '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 \( P = 5.4 \times 10^{-3} \); Fig. 2a), corresponding to a significant enrichment for genes in this set (odds ratio (OR) = 1.8; 95% confidence interval (CI) 1.3–2.6, empirical \( P = 2.6 \times 10^{-3} \); 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 \( P = 9.4 \times 10^{-4} \)). Enrichment was also found for rare CNVs overlapping intellectual disability genes, more notably for deletions (OR = 2.1; 95% CI 1.1–4.2, empirical \( P = 0.053 \)). 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 'intellectual disability' were combined, we observed 7.6% of cases with rare CNVs preferentially affecting ASD/intellectual disability genes compared to 4.5% in controls (Fisher exact test \( P = 1.2 \times 10^{-3} \); 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 used subjective boundary conditions 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 (\( P = 3.6 \times 10^{-6} \); 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. The population attributable risk provided by the combination of all ASD CNVs that overlap ASDs and/or intellectual disability genes is estimated to be 3.3% (Supplementary Table 11). We also identified rare de novo chromosomal abnormalities and large CNVs likely to be aetiological (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 biologically meaningful groups.

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 organize these sets graphically into a functional enrichment map (or network) using Cytoscape. We identified the 'seed' gene sets for the network at an FDR \( q \)-value of 5% and further relaxed the thresholds to 12.5% to better capture the network topology.

Using these criteria only deletions were found to be 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 to be enriched and used to construct a functional map (Fig. 3a and Supplementary Figs 6 and 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 Figs 8 and 9).

We identified enrichments in gene sets known to be involved in autism spectrum disorders (ASDs) and also discovered novel candidate ASD pathways (Fig. 3a and 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. Novel observations included gene sets involved in GTPase/Ras signalling, with component Rho GTPases known to be involved in regulating dendrite and spine plasticity and associated with intellectual disability. 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/intellectual disability genes (Fig. 3b and Supplementary Fig. 11) and found genes enriched in sets linked to microtuble cytoskeleton, glycosylation and CNS development/adhesion. The two groups of genes found to be enriched in deletions (Fig. 3a) also displayed connectivity to the ASD/intellectual disability disease gene sets, either directly or through intermediates (Fig. 3b and Supplementary Fig. 12). Although ASD genes seem to be enriched in different subsets of genes compared to intellectual-disability-only genes, we cannot discount the possibility that this is the result of selection bias, and we expect that both intellectual disability genes may yet be linked to ASD.

Our findings provide strong support for the involvement of multiple rare genetic CNVs, both genome-wide and at specific loci, in ASD. These findings, similar to those recently described in schizophrenia, suggest that at least some of these ASD CNVs (and the genes that they affect) are under purifying selection. Genes previously implicated in ASD by rare variant findings have pointed to functional themes in ASD pathophysiology. 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 that SHANK2, SYNGAP1 and DLGAP2 are new ASD loci that also encode proteins in the PSD. We also found intellectual disability genes to be important in ASD. 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 step 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**

Raw data from ASD family (accession phs000267.v1.p1) and SAGE control (Accession: phs000092.v1.p1) genotyping are at NCBI dbGAP. CNVs were analysed using PLINK v1.07, R stats and custom scripts. See Supplementary Information for details. A list of all CNVs passing quality control is available in Supplementary Table 8.

Received 3 December 2009; accepted 7 May 2010.

**Published online 9 June 2010.**


**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements**

The authors 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 Hillstrand Foundation (USA). Additional support for individual groups was provided by the US National Institutes of Health (NIH grants HD055751, HD05782, HD055784, HD355465, MH25708, MH55284, MH57881, MH061009, MH06359, MH066673, MH080647, MH081754, MH66676, NS026630, NS042165, NS049261), the Canadian Institute for Advanced Research (CIFAR), the Canadian Institutes for Health Research (CIHR), Assistance Publique–Hôpitaux de Paris (France), Autistica, 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), Fundación para una Ciencia e Tecnología (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/03/UK (D.P.) is supported by fellowships from the Royal Netherlands Academy of Arts and