

ORIGINAL ARTICLE

Rare structural variation of synapse and neurotransmission genes in autism

X Gai¹, HM Xie¹, JC Perin¹, N Takahashi², K Murphy¹, AS Wenocur¹, M D'arcy¹, RJ O'Hara¹, E Goldmuntz^{3,6}, DE Grice⁴, TH Shaikh⁵, H Hakonarson^{6,7,8}, JD Buxbaum², J Elia^{9,10} and PS White^{1,6,11}

¹Center for Biomedical Informatics, Children's Hospital of Philadelphia, Philadelphia, PA, USA; ²Seaver Autism Center and Department of Psychiatry, Mt Sinai School of Medicine, New York, NY, USA; ³Division of Cardiology, Children's Hospital of Philadelphia, Philadelphia, PA, USA; ⁴Department of Child and Adolescent Psychiatry, Columbia University, New York, NY, USA; ⁵Department of Pediatrics, University of Colorado School of Medicine, Denver, CO, USA; ⁶Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA, USA; ⁷Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, USA; ⁸Division of Pulmonary Medicine, Children's Hospital of Philadelphia, Philadelphia, PA, USA; ⁹Department of Child and Adolescent Psychiatry, Children's Hospital of Philadelphia, Philadelphia, PA, USA; ¹⁰Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA, USA and ¹¹Division of Oncology, Children's Hospital of Philadelphia, Philadelphia, PA, USA

Autism spectrum disorders (ASDs) comprise a constellation of highly heritable neuropsychiatric disorders. Genome-wide studies of autistic individuals have implicated numerous minor risk alleles but few common variants, suggesting a complex genetic model with many contributing loci. To assess commonality of biological function among rare risk alleles, we compared functional knowledge of genes overlapping inherited structural variants in idiopathic ASD subjects relative to healthy controls. In this study we show that biological processes associated with synapse function and neurotransmission are significantly enriched, with replication, in ASD subjects versus controls. Analysis of phenotypes observed for mouse models of copy-variant genes established significant and replicated enrichment of observable phenotypes consistent with ASD behaviors. Most functional terms retained significance after excluding previously reported ASD loci. These results implicate several new variants that involve synaptic function and glutamatergic signaling processes as important contributors of ASD pathophysiology and suggest a sizable pool of additional potential ASD risk loci.

Molecular Psychiatry (2012) 17, 402–411; doi:10.1038/mp.2011.10; published online 1 March 2011

Keywords: autism; copy-number variation; glutamatergic signaling; neurotransmission; structural variation; synaptic transmission

Introduction

Autism and autism spectrum disorders (ASDs) are characterized by deficits in social interactions, communication difficulties and the presence of restricted, repetitive and stereotyped patterns of behavior.^{1,2} Autism is co-occurrent in disorders with known cytogenetic etiologies such as Fragile X, suggesting a role for genome structural aberrations; however, these account for <10% of cases to date.³ The remainder, often referred to as 'idiopathic' autism, are considered highly heritable with a 5–10% recurrence rate in siblings and a 60–90% concordance rate in monozygotic twins.^{4,5} To date, genetic linkage and association studies have had limited success in identifying genetic risk loci for ASDs.^{6–11} Most recently, genome

copy-number variation (CNV) analyses have identified several candidate loci, including 16p11.2 and *NRXN1* (*neurexin 1*).^{10,12} The lack of common predisposition genes has led to the suggestion that rare variants constitute the majority of ASD risk, and that these variants collectively affect a limited number of common functional pathways. In particular, genetic and functional studies of candidate genes suggest the involvement of synaptic transmission in ASD etiology.^{13,14} This evidence served as the basis for our hypothesis that inherited rare structural variants in genes from a number of closely related neurological pathways collectively contribute to a significant portion of disease risk in idiopathic ASDs.

Materials and methods

Study cohorts

The Autism Genetic Resource Exchange (AGRE) collection includes a total of 5431 affected and parental samples from 1000 families. These samples

Correspondence: Dr PS White, Children's Hospital of Philadelphia, 34th Street and Civic Center Boulevard, Room 1407 CHOP North, Philadelphia, PA 19104-4318, USA.

E-mail: whitep@email.chop.edu

Received 4 May 2010; revised 6 January 2011; accepted 19 January 2011; published online 1 March 2011

are grouped into four different sets based on the time when they were recruited: set 1 (660), set 2 (1373), set 3 (598) and set 4 (2800). Pedigree and affected status information on 4411 subjects are provided in the AGRE pedigree file at the AGRE website. Approximately 95% of these families are multiplex. The autism discovery and replication cohorts were derived from AGRE set 4 and sets 1–3, respectively. Cases designated as ‘autism’ and corresponding parents were included; cases designated as ‘spectrum’ and ‘not quite autism’ were not considered. Cases not of European descent and cases with known genetic syndromes or other non-idiopathic autistic causes were excluded. After applying these exclusion criteria and our genotyping and CNV detection quality control criteria (see below), the final discovery cohort included 1793 subjects (631 cases and 1162 parental samples), and the replication cohort included 1702 subjects (593 cases and 1109 parents). A number of parents (90 parents in the discovery cohort and 77 parents in the replication cohort) were excluded for failing to meet genotype quality measures. For the discovery cohort, 262 of the 335 families have multiple siblings included, for a total of 558 siblings. For the replication cohort, 218 of the 354 families have multiple siblings included, for a total of 457 siblings.

Healthy control samples were recruited from well-child visits conducted within the healthcare network of the CHOP (Children’s Hospital of Philadelphia). Criteria for discovery control cohort inclusion were: age within age 3–18 years; screened for having no chronic illness as well as autism; and an extensive and continuous electronic health record with no indications of chronic health issues by diagnosis, prescribed medications, diagnostic tests or subspecialty encounters. All controls were unrelated as determined by health record and genotype assessment. After quality metrics used for the autism cohort were applied, the discovery control cohort comprised 1005 individuals of European descent, 723 of African descent and 47 of Asian descent. The replication control cohort consisted of 2026 healthy individuals comprising the CHOP CNV resource, as previously described.¹⁵ Gene-specific CNV analyses included healthy control individuals confined to European descent, whereas functional analyses included all ethnicities in order to identify and remove a larger constellation of nonpathogenic variants. The congenital heart defect cohort used as a control comparison consisted of 610 subjects with conotruncal defects and 896 parents of European descent who were uniformly assessed, genotyped and analyzed at CHOP. The congenital heart defect cohort was processed using the same protocols, quality measures and CNV size thresholds described for the ASD cohort, including restricting consideration of CNVs to those with unequivocal inheritance from a parent.

Genotyping

AGRE and control samples were assayed on the Illumina Infinium II Human-Hap550 BeadChip

(Illumina, San Diego, CA, USA) as previously described.¹⁶ The standard Illumina cluster file was used for the analysis, which is generated at Illumina by genotyping 120 HapMap samples, running the BeadStudio clustering algorithm and reviewing single-nucleotide polymorphisms (SNPs) with poor performance statistics, including call frequency, cluster separation and Hardy–Weinberg equilibrium. For both control sets, ancestry informative markers available on the Human-Hap550 BeadChip¹⁷ were used to evaluate eligible subjects to determine ethnicity. Genotypes that failed quality measures for accurate CNV detection (call rate >98%) were excluded.

CNV detection and analysis

The Illumina BeadStudio 3.0. software package was used for initial CNV detection analysis. Log R ratios and B allele frequencies were first exported from BeadStudio. Log R ratio values were used as an additional sample-wide genotype quality control measure, and log R ratios with the s.d. above 0.35 were excluded from the study. CNV predictions were generated using CNV Workshop as described previously.^{15,18,19} To reduce the possibility of type I error, only deletions spanning ≥ 10 consecutive SNPs and duplications spanning ≥ 20 consecutive SNPs were included. These CNV size thresholds were selected based upon previous studies from our group^{15,18,19} and experience with samples undergoing array-based clinical diagnostics at our institution.²⁰ Specifically, in these previous studies, our CNV detection method was shown to have a negligible false-positive rate using these thresholds. An increased threshold for duplications was used because of the difference in the degree of expected signal intensity variation between copy states for each probe: a twofold decrease for heterozygous deletion versus only a 1.5-fold increase for duplication. For comparative analysis purposes, two CNVs were considered equivalent if their genomic regions overlapped at least 80% reciprocally. All CNVs meeting quality and experimental design metrics were independently assessed by visual inspection with BeadStudio. For pathway analyses, all olfactory receptor genes were removed before consideration, as this gene family is rapidly evolving^{21–23} and tends to be associated with gene clusters that can over-represent such genes in functional enrichment studies. Careful analysis of enriched gene sets indicated no genomic region-specific bias from other functional gene clusters. For single gene analyses, apparently identical CNVs (>80% reciprocal overlap in genomic regions) from related siblings were removed before analysis. Significance testing for individual gene analyses used Fisher’s exact test.

DNA for all AGRE and congenital heart defect samples was derived from immortalized cell lines, whereas all healthy control DNAs were derived from blood. Although acknowledging that the difference in DNA origin between cases and controls could theoretically have an effect on the results, we consider this

to be highly unlikely, for the following reasons: (1) with the exception of the increased frequency of autism subject CNVs detected in the IgH-juxtaposed RNA gene *LOC10013346*, we saw no significant difference in genome-wide CNV rates in autism cases, and no other single gene was significantly enriched in cases; (2) real and presumptive *de novo* CNVs are excluded from our functional analyses; (3) our functional analyses would likely not be affected by increased cell line artifacts, as each gene is counted in a frequency-independent manner in these analyses.

Gene ontology analysis

Gene ontology analysis was carried out using the Explain Analysis System 2.4.2 (BIOBASE, Wolfenbüttel, Germany). RefSeq genes overlapping rare and inherited CNVs observed in autism cases were extracted from the UCSC (University of California, Santa Cruz) Genome Browser and input into the requisite pathway analysis models using default settings. Fisher's exact test was used when testing for enrichment of any gene ontology terms in the input gene lists. Mouse gene annotations were used for analysis because of a richer annotation set than for humans. All analyses were repeated with human annotations and yielded comparable results, although significance values were lessened because of the paucity of annotations for some genes. A Benjamini-Hochberg correction for multiple testing was applied for all functional enrichment analyses using the *p.adjust* function with parameter BH in the statistical package R.

Mouse phenotype ontology analysis

Mouse phenotype analysis was performed using Mouse Phenotype Ontology (MP) term annotations from the MGI (Mouse Genome Informatics) resource (Jackson Laboratory; Bar Harbor, ME, USA). Assigned MP terms for mouse models of genes overlapping rare and inherited autism CNVs were extracted from MGI. If a MP term was associated with a gene, all parental MP terms were assigned to the gene as well. Fisher's exact test was employed to measure the enrichment of autism CNV genes annotated with a given MP term compared with all genes annotated with a MP term. The *P*-values were used to rank the significance of enrichments.

CNV validation by quantitative PCR

Universal Probe Library (UPL; Roche, Indianapolis, IN, USA) probes were selected using the ProbeFinder v2.41 software (Roche). Probes and primer sequences selected for each CNV region are listed in Supplementary Table 12. Quantitative PCR was performed on an ABI 7500 Real Time PCR Instrument or on an ABI Prism 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA, USA). Each sample was analyzed in quadruplicate in a 10 μ l reaction mixture (100 nM probe, 200 nM each primer, 1 \times Platinum Quantitative PCR SuperMix-Uracil-

DNA-Glycosylase with ROX (Invitrogen, Carlsbad, CA, USA) and 25 ng genomic DNA). Values were evaluated using Sequence Detection Software v2.2.1 (Applied Biosystems). Data analysis was further performed using either the $\Delta\Delta C_T$ method or qBase.²⁴ Reference genes, chosen from *COBL*, *GUSB*, *PPIA* and *SNCA*, were included based on the minimal coefficient of variation. Data were subsequently normalized by setting a normal control to a value of 1.

Results

We conducted a modified case-control CNV study, restricting inclusion to probands and siblings of European descent with a diagnosis of idiopathic autism. We focused our analysis on rare inherited structural variants in the autistic subjects and potential enrichment of any functional categories assigned to genes overlapping these CNVs. Genotypes of autistic subjects were obtained from the AGRE²⁵ (see Materials and methods) and were divided into discovery and replication cohorts based upon the time of recruitment.

In the discovery stage of the study, 631 autism subjects and 1162 parents of these subjects were compared with a healthy control set of 1775 children. We observed no significant differences in the frequency of CNV rate ($P=0.109$), CNV size ($P=0.136$), deletions ($P=0.987$) or duplications ($P=0.144$) between the subject and control groups. A total of 395 inherited CNVs (135 duplications, 256 heterozygous deletions and 4 hemizygous deletions of the X chromosome) were identified in 286 autism subjects that were not present in healthy controls (Supplementary Table 1). CNV transmission to probands was independent of parent gender (185/395 with paternal inheritance; $P=1$). For the purpose of this study, we considered and refer to these CNVs as 'rare', as the estimated prevalence of each is $<0.05\%$ in an ethnically matched control population. These CNVs overlapped a total of 382 genes excluding olfactory receptor genes. In all, 14 genes had CNVs in 5 unrelated individuals, and the remaining 368 genes had CNVs in ≤ 4 individuals. This autism-enriched, CNV-associated gene set (autism gene set) was analyzed to determine if it was enriched for particular biological or disease processes.

Gene ontology analysis revealed significant enrichment of the autism gene set for a number of biological processes (BPs) after correction for multiple testing (Table 1 and Supplementary Table 2). Many of these processes were associated with synaptic function and neurotransmission, including *cell-cell signaling* ($P=1.97 \times 10^{-7}$), *transmission of nerve impulse* ($P=7.06 \times 10^{-5}$), *synaptic transmission* ($P=8.05 \times 10^{-5}$), *neuron adhesion* ($P=2.06 \times 10^{-3}$), *central nervous system development* ($P=4.30 \times 10^{-3}$), *regulation of neurotransmitter levels* ($P=4.88 \times 10^{-3}$), *neurotransmitter secretion* ($P=1.30 \times 10^{-2}$) and *neurotransmitter transport* ($P=2.95 \times 10^{-2}$). Several cellular component (CC) terms were similarly enriched in the autism gene set,

Table 1 GO biological process terms significantly enriched in genes overlapping inherited rare autism CNVs

GO term	Discovery cohort			Replication cohort			Combined cohort		
	Genes in CNVs ^a	Significance ^b	Adjusted significance ^c	Genes in CNVs ^a	Significance ^b	Adjusted significance ^c	Genes in CNVs ^a	Significance ^b	Adjusted significance ^c
Cell–cell signaling	37	2.08×10^{-9}	1.97×10^{-7}	22	1.32×10^{-3}	3.53×10^{-2}	52	3.88×10^{-9}	6.14×10^{-7}
Transmission of nerve impulse	23	2.23×10^{-6}	7.06×10^{-5}	19	4.01×10^{-5}	8.05×10^{-3}	36	5.60×10^{-8}	4.98×10^{-6}
Synaptic transmission	21	3.06×10^{-6}	8.05×10^{-5}	18	2.17×10^{-5}	8.05×10^{-3}	33	6.31×10^{-8}	4.98×10^{-6}
Neuron adhesion	3	1.29×10^{-4}	2.06×10^{-3}	3	8.80×10^{-5}	8.82×10^{-3}	3	6.69×10^{-4}	1.17×10^{-2}
Central nervous system development	16	3.80×10^{-4}	4.30×10^{-3}	13	2.63×10^{-3}	4.73×10^{-2}	24	1.63×10^{-4}	3.22×10^{-2}
Regulation of neurotransmitter levels	8	4.95×10^{-4}	4.88×10^{-3}	7	1.15×10^{-3}	3.28×10^{-2}	12	8.60×10^{-5}	2.09×10^{-3}
Developmental process	80	7.50×10^{-4}	7.08×10^{-3}	72	6.45×10^{-4}	2.87×10^{-2}	137	4.21×10^{-5}	1.33×10^{-3}
Multicellular organismal development	64	9.36×10^{-4}	8.37×10^{-3}	59	4.05×10^{-4}	2.70×10^{-2}	109	6.64×10^{-5}	1.78×10^{-3}
Neurotransmitter secretion	6	1.62×10^{-3}	1.30×10^{-2}	7	1.19×10^{-4}	9.54×10^{-3}	10	6.77×10^{-5}	1.78×10^{-3}
Post-translational protein modification	35	3.58×10^{-3}	2.27×10^{-2}	32	2.83×10^{-3}	4.73×10^{-2}	58	8.80×10^{-4}	1.50×10^{-2}
Protein modification process	41	4.03×10^{-3}	2.45×10^{-2}	43	7.34×10^{-5}	8.82×10^{-3}	75	2.99×10^{-5}	1.05×10^{-3}
Neurotransmitter transport	7	5.29×10^{-3}	2.95×10^{-2}	8	5.53×10^{-4}	2.87×10^{-2}	12	3.36×10^{-4}	6.43×10^{-3}

Abbreviations: CNV, copy-number variation; GO, gene ontology.

^aNumber of genes overlapping CNVs exclusively in the autism cohort assigned to the specified GO term.

^bSingle-test *P*-values.

^cMultiple test-corrected *P*-values using the Benjamini–Hochberg correction; only categories reaching a corrected statistical significance of *P* < 0.05 in both the discovery and replication cohorts are shown. Bold values indicate an adjusted significance of < 0.05.

including *presynaptic membrane* ($P=5.65 \times 10^{-4}$) and *synapse* ($P=1.11 \times 10^{-3}$) (Table 2 and Supplementary Table 3). Both the BP and CC terms of significance were represented by sizable and only modestly overlapping gene sets (BP: $n=114$ unique genes, mean = 3.2 terms/gene; CC: $n=18$ unique genes, mean = 1.4 terms/gene), indicating considerable independence in implication of individual terms (Supplementary Figure 1). Autism CNV genes frequently associated with terms related to synapse function and neurotransmission included previous candidates *CNTN4*, *NRXN1* and *PARK2*, as well as genes *CNTN6*, *CTNND2*, *GRIN2A*, *NCAM2*, *GRM7*, *GRM8* and *RCAN1*.

To demonstrate the specificity of the approach, the same CNV inclusion criteria and analysis strategies were re-applied using a large set of parent–child trios for congenital heart disease. The enriched pathway spectrum was almost entirely dissimilar in the cardiovascular cohort, and no function classes consistent with neuropsychiatric function were found significant, even without application of a multiple test correction (HM Xie *et al.*, in preparation).

To replicate the functional analyses, we applied the same CNV inclusion criteria and methods as used for the discovery cohort to 593 additional idiopathic autism subjects of European descent, and 1109 corresponding parental samples, from AGRE. These were compared with CNVs from 2026 healthy individuals comprising the CHOP healthy control cohort.¹⁵ A total of 392 inherited CNVs (131 duplications, 256 heterozygous deletions, 1 homozygous deletion and 4 hemizygous X chromosome deletions) were exclusive to 270 of the autistic subjects (Supplementary Table 1). CNV transmission was again independent of parent gender (205/392 with paternal inheritance; $P=1$). Excluding olfactory receptors, these CNVs overlapped 387 unique genes, of which only 43 (11.1%) were present in the discovery autism CNV gene set.

A total of 12 BP and 3 CC terms reaching corrected significance in the discovery cohort were also enriched in the replication cohort after multiple test correction. As with the discovery cohort, the genes with CNVs in these enriched terms showed only moderate overlap between terms (Supplementary Figure 1). Furthermore, the autism CNV gene sets representing the replicated enriched terms demonstrated only a small overlap between the two cohorts, with only 23.7% (19/99) and 29.4% (5/17) of replication genes also present in the discovery cohort for BP and CC, respectively. For example, of 21 and 18 genes assigned to the BP term *synaptic transmission* with autism-specific CNVs in the discovery and replication cohorts, respectively, only 6 genes had CNVs in both cohorts. Among the genes with autism-specific CNVs found in both cohorts were known autism candidate genes *CTNND2*, *CNTN4*, *NRXN1* and *PARK2*, along with glutamate receptors *GRIN2A*, *GRIN2C* and *GRM7* (Table 5).

The discovery and replication cohorts were then combined into a single set of 1224 samples. Analysis

Table 2 GO cellular component terms significantly enriched in genes overlapping inherited rare autism CNVs

GO term	Discovery cohort		Replication cohort		Combined cohort	
	Genes in CNVs ^a	Adjusted significance ^b	Genes in CNVs ^a	Adjusted significance ^b	Genes in CNVs ^a	Adjusted significance ^b
Presynaptic membrane	6	7.61×10^{-6}	4	7.96×10^{-4}	8	1.40×10^{-6}
Synapse	16	2.85×10^{-5}	14	8.52×10^{-5}	27	9.45×10^{-8}
Dystrophin-associated glycoprotein complex	4	1.04×10^{-4}	3	1.33×10^{-3}	5	5.96×10^{-5}

Abbreviations: CNV, copy-number variation; GO, gene ontology.

^aNumber of genes overlapping CNVs exclusively in the autism cohort assigned to the specified GO term.

^bSingle-test *P*-values.

^cMultiple test-corrected *P*-values using the Benjamini–Hochberg correction; only categories reaching a corrected statistical significance of $P < 0.05$ in both the discovery and replication cohorts are shown. Bold values indicate an adjusted significance of < 0.05 .

of this combined cohort yielded results that were consistent with the initial findings. Autism candidate genes *NRXN1* ($P = 1.46 \times 10^{-2}$) and *CNTNAP2* ($P = 3.51 \times 10^{-3}$) demonstrated individual gene significance but failed to retain significance after multiple test correction (Table 4). In functional enrichment analysis, all BP and CC terms enriched in the discovery cohort retained multiple test-corrected significance in the combined cohort. The most significant functions were BP terms *cell–cell signaling* ($P = 6.14 \times 10^{-7}$), *transmission of nerve impulse* ($P = 4.98 \times 10^{-6}$) and *synaptic transmission* ($P = 4.98 \times 10^{-6}$) (Table 1 and Supplementary Table 2); and CC terms *synapse* ($P = 1.49 \times 10^{-5}$), *presynaptic membrane* ($P = 1.10 \times 10^{-4}$) and *dystrophin-associated glycoprotein complex* ($P = 2.36 \times 10^{-3}$) (Table 2 and Supplementary Table 3). There was again broad representation of genes among significant terms (Supplementary Figure 1).

As a complementary approach, we determined whether our autism CNV gene sets were preferentially associated with particular phenotypes assigned to mouse transgenic and null mice previously developed for these genes. For this analysis, we used MP term assignments from the MGI group²⁶ that have been reported for 113 of the 382 discovery cohort autism CNV genes, and 98 of 387 such genes in the replicate cohort. Of 6406 functional terms, 12 were significantly enriched in the autism gene set both for the discovery and replication cohorts, and subsequently in the combined cohort (Table 3 and Supplementary Table 4). Many of the terms were highly consistent with an autistic phenotype, including *abnormal (\pm CNS) synaptic transmission*, *abnormal motor capabilities/coordination/movement*, *abnormal motor coordination/balance*, *reduced NMDA-mediated synaptic currents*, *abnormal impulse conducting system conduction*, *decreased startle reflex*, *abnormal miniature excitatory postsynaptic currents* and *abnormal spatial learning*. After correction for multiple testing, two closely associated terms retained significance in the replication and combined cohorts (*abnormal synaptic transmission*, combined $P = 2.52 \times 10^{-2}$; *abnormal CNS synaptic transmission*, combined $P = 2.67 \times 10^{-2}$). The generally lower significance values than observed for the BP and CC analyses likely reflect the relatively small number of genes with well-annotated phenotypes; however, the general trend of enrichment for synaptic function and neurotransmission was consistent with our previous analyses. Only 5 of 33 (15.2%) autism CNV genes within the 12 significant MP assignments were represented in both the discovery and replication cohorts, again suggesting widespread genetic heterogeneity.

To determine the relative contribution of the unreported autism candidate genes to the enriched functional groups, each functional analysis was repeated after excluding a set of 205 genes either previously reported as autism risk candidates, or that reside in a genomic rearrangement associated with

Table 3 Mouse phenotype ontology terms significantly enriched in genes overlapping inherited rare autism CNVs

Phenotype term	Discovery cohort			Replication cohort			Combined cohort		
	Genes in CNVs ^a	Significance ^b	Adjusted significance ^c	Genes in CNVs ^a	Significance ^b	Adjusted significance ^c	Genes in CNVs ^a	Significance ^b	Adjusted significance ^c
Abnormal synaptic transmission	17	7.19×10^{-3}	0.37	21	2.26×10^{-5}	1.97×10^{-2}	33	1.76×10^{-5}	2.52×10^{-2}
Abnormal CNS synaptic transmission	15	8.72×10^{-3}	0.40	18	9.17×10^{-5}	4.00×10^{-2}	29	3.73×10^{-5}	2.67×10^{-2}
Abnormal motor capabilities/coordination/movement	41	3.30×10^{-4}	0.17	29	4.34×10^{-2}	0.47	63	2.30×10^{-4}	0.11
Small hair follicles	2	3.59×10^{-3}	0.37	2	2.73×10^{-3}	0.26	3	6.12×10^{-4}	0.22
Abnormal motor coordination/balance	24	5.54×10^{-4}	0.17	16	4.51×10^{-2}	0.47	34	1.31×10^{-3}	0.37
Reduced NMDA-mediated synaptic currents	3	5.49×10^{-4}	0.17	2	7.39×10^{-3}	0.26	3	2.39×10^{-3}	0.39
Abnormal impulse conducting system conduction	5	1.19×10^{-2}	0.40	4	3.00×10^{-2}	0.43	8	2.44×10^{-3}	0.39
Abnormal QT interval	3	7.34×10^{-3}	0.37	2	4.40×10^{-2}	0.47	4	4.84×10^{-3}	0.40
Decreased startle reflex	4	1.36×10^{-2}	0.40	4	8.41×10^{-3}	0.28	6	5.04×10^{-3}	0.40
Abnormal nervous system physiology	32	4.61×10^{-2}	0.47	31	1.09×10^{-2}	0.31	57	3.15×10^{-3}	0.40
Abnormal miniature excitatory postsynaptic currents	3	4.45×10^{-2}	0.47	3	3.13×10^{-2}	0.43	5	1.17×10^{-2}	0.44
Abnormal spatial learning	7	3.33×10^{-2}	0.47	6	4.88×10^{-2}	0.47	9	7.25×10^{-2}	0.57

Abbreviations: CNS, central nervous system; CNV, copy-number variation; NMDA, N-methyl D-aspartate.

^aNumber of genes overlapping CNVs exclusively in the autism cohort assigned to the specified phenotype term.

^bSingle-test *P*-values.

^cMultiple test-corrected *P*-values using the Benjamini–Hochberg correction; the 12 most significant terms in the combined cohort are listed. Bold values indicate an adjusted significance of < 0.05.

autism susceptibility (Supplementary Table 5). Despite the reduced power of these analyses, most BP and all CC and MP terms retained multiple test corrected significance (BP and CC) or uncorrected significance (MP) in both the discovery and replication cohorts (Supplementary Tables 6–11). Several additional terms consistent with autism pathophysiology were newly significant, including *cell–cell adhesion* (BP), *response to mechanical stimulus* (BP), *cell junction* (CC) and *cytoplasmic vesicle membrane* (CC). These results support representation of previously unattributed autism risk in the enriched autism gene sets.

To search for common variants, we determined the extent to which any single genes were represented by CNVs in autism cases relative to matched healthy individuals. We first compared our CNV set with findings from previous autism genomic studies, including a recent study by Glessner and colleagues^{11,27,28} that reported CNV findings for a set of autism spectrum disorder subjects partially overlapping our cohort. Combining the two cohorts, 25 inherited CNVs overlapped one or more previously implicated candidate genes (*NRXN1*, *CNTN4*, *SUMF1*, *DISC1*, *PARK2*, *AUTS2* and *CNTNAP2*) or genomic regions (Table 4 and Supplementary Table 5). No gene had evidence of over-representation in the ASD cases, although formal significance testing was not possible because of the unavailability of inheritance status of CNVs in the control groups. The majority of CNVs detected in these genes did not substantially overlap CNVs detected in healthy controls.

When expanding this approach genome-wide and considering all ASD subject CNVs regardless of inheritance status and presence in controls, only a single RNA gene on 14q32, *LOC10013346*, was significantly enriched for CNVs in ASD cases (27 overall CNVs versus none in the discovery control cohort, and 29 versus 4 in the replication control cohort). However, CNVs disrupting this gene in autism cases also spanned the adjoining immunoglobulin heavy chain locus, suggesting that these CNVs were likely cell culture artifacts. Moreover, only one of these CNVs in the discovery cohort and two in the replication cohort were present in parental samples. No other individual genes were significantly over-represented in either autism cohort.

CNVs overlapping five genomic regions (*CTNND2*, *GRIN2A*, *NCAM1*, *KCNE2/FAM165B/KCNE1/RCAN1* and *ICA1/NXPH1*) were selected for experimental validation of CNVs based upon the following criteria: higher frequency seen in autism cases relative to controls; evidence for involvement in neurotransmission processes; and little or no previous implication for autism risk. As a validation control, we included two autism probands with known *CNTN4* deletions and one proband with a known *CNTN4* duplication.²⁸ All 17 CNVs identified in probands and parental samples in these loci were experimentally validated (Supplementary Figure 2).

Table 4 Analysis of inherited CNVs identified in genes previously implicated in autism

Gene	Location	Evidence for implication ^a	Discovery		Replication		Combined	
			Cases ^b	Controls	Cases ^b	Controls	Cases ^b	Controls
<i>NRXN1</i>	2p16.3	Rare variant	4	2	1	6	5	8
<i>CNTN4</i>	3p26.3	Rare variant	1	1	1	1	2	2
<i>SUMF1</i>	3p26.1	Rare variant	0	0	1	0	1	0
<i>PARK2</i>	6q26	Rare variant	5	15	7	17	12	32
<i>AUTS2</i>	7q11.22	Rare variant	1	0	0	1	1	1
<i>ASTN2</i>	9q33.1	Rare variant	0	1	1	4	1	5
<i>CNTNAP2</i>	7q36.1	Association	2	0	0	0	2	0
<i>DISC1</i>	1q42.2	Association	0	2	1	5	1	7

Abbreviation: CNV, copy-number variation.

^aRefer to Supplementary Table 1 for gene implication references.

^bNumber of cases containing a CNV overlapping a listed gene or region.

Discussion

Our functional enrichment analysis of the AGRE cohort identified a number of genes enriched for inherited CNVs in ASD subjects, several of which are known to have functions consistent with ASD etiology. Of particular interest were the ionotropic glutamate receptor *GRIN2A* and CNVs spanning both the islet cell autoantigen *ICA1* (*islet cell autoantigen 1*) and the adjacent α -neurexin ligand *NXPH1* (*neurexophilin-1*), as these genes were disrupted by CNVs in subjects from both autism cohorts and had no evidence of structural variation in healthy controls (Supplementary Figure 3). *ICA1* is involved in glutamate receptor-mediated transmission,²⁹ whereas the neurexophilin *NXPH1* is a ligand for the protein product of the autism candidate gene *NRXN1*.³⁰ Barnby *et al.*³¹ implicated *GRIN2A* (*glutamate receptor, ionotropic, N-methyl D-aspartate 2A*) as an autism candidate by gene-specific association, and further evidence suggests involvement in susceptibility to schizophrenia and attention deficit hyperactivity disorder.^{32–34} *GRIN2A*-null mice exhibit impairments in contextual memory and synaptic plasticity.³⁵

Also of special interest were 12 autism CNV genes associated with significantly enriched functions each in the BP, CC and MP analyses (Table 5). Three of these genes (*DOC2A*, *NRXN1* and *PARK2*) have previous support as autism susceptibility loci and validate our approach. The receptor tyrosine kinase *ERBB4* controls maturation and plasticity of glutamatergic synapses, myelination and dopaminergic function, and has been suggested as having a role in schizophrenia risk.³⁶ *GRIN2A* (described above), *GRIN2C* and *GRM7* function as glutamate receptors, whereas *UNC13A*-null mice demonstrate impairment of glutamatergic synaptic vesicle maturation.³⁷ The neurexin homolog *NRXN3* is not well characterized but has roles in synaptic plasticity and synaptic vesicle exocytosis.³⁸ *SYN3* functions in synaptogenesis and the modulation of neurotransmitter release,^{39,40} and *SYN3*-null mice exhibit deficits in

social learning, contextual and cued fear conditioning and fear-potentiated startle.⁴¹ *KCNMA1* encodes a calcium-activated potassium channel linked to glutamatergic synapses,⁴² whereas mice null for this gene exhibit cerebellar learning deficiencies.⁴³ Disruption of *KCNMA1* was identified in an autistic subject with a balanced translocation t(9;10)(q23;q22).⁴⁴

During review of this manuscript, Pinto *et al.*⁴⁵ reported an analysis of rare CNVs in ASD cases relative to healthy controls. Both studies are consistent with the presence of many novel and previously reported loci, each with only a small number of distinct CNVs observed in the subjects considered. In comparing our results with the study of Pinto *et al.*,⁴⁵ there is little overlap between the sets of individual genes suggested as ASD candidates. These differences may be because of our exclusive focus on inherited events or might further support the presence of a substantial pool of potential ASD risk loci. Encouragingly, pathway enrichment findings between the two studies show greater consistency, with enrichment found in cell adhesion, central nervous system development, synaptic function and glutamate-mediated neuronal signaling processes. A follow-up study comparing inherited and *de novo* CNV functional attributes would allow determination of the degree to which each CNV class is functionally distinct.

Our results are consistent with the hypothesis that inherited autism risk is genetically highly heterogeneous, both from our observations of a lack of autism-specific CNVs overlapping any single gene in our analysis, and the lack of overlap between gene sets represented by the same autism-enriched functional terms in our two cohorts. Better estimates of the number of genes that can potentially confer risk, and the number of risk alleles typically required to confer sufficient risk in an individual, await higher-resolution genomic studies in larger cohorts, as well as platforms that more fully consider common CNVs. Our method, which has yielded results similarly consistent with disease-specific pathophysiology for

Table 5 Autism CNV genes implicated in significant functional processes in both the discovery and replication cohorts

<i>Replicated biological process genes^{a,b}</i>	<i>Replicated cellular component genes^c</i>	<i>Replicated mouse phenotype genes^d</i>	<i>Genes implicated in all functional types^e</i>
<i>CALCR</i>	<i>GRIN2A</i>	<i>CTNND2</i>	<i>DOC2A</i>
<i>CDH13</i>	<u><i>NRXN1</i></u>	<i>GRIN2A</i>	<u><i>ERBB4</i></u>
<u><i>CNTN4</i></u>	<u><i>PARK2</i></u>	<u><i>NRXN1</i></u>	<i>GRIN2A</i>
<i>CNTN6</i>	<i>SGCZ</i>	<u><i>PARK2</i></u>	<i>GRIN2C</i>
<i>CTNND2</i>	<i>SNTG2</i>	<u><i>PTPRD</i></u>	<i>GRM7</i>
<i>GRIN2A</i>			<i>KCNMA1</i>
<i>ICA1</i>			<u><i>NRXN1</i></u>
<i>IMMP2L</i>			<i>NRXN3</i>
<i>KCNE1</i>			<u><i>PARK2</i></u>
<i>MYOM2</i>			<i>SYN3</i>
<i>NCAM2</i>			<i>UNC13A</i>
<u><i>NRXN1</i></u>			<i>UTRN</i>
<u><i>PARK2</i></u>			
<i>PCSK6</i>			
<i>PTPRD</i>			
<i>RAF1</i>			
<i>RCAN1</i>			
<i>SDK1</i>			
<i>SNTG2</i>			

Abbreviation: CNV, copy-number variation.

^aPreviously implicated autism candidate genes are underlined.

^bAutism CNV genes for which at least one autism-specific CNV overlapping the gene was identified in both the discovery and replication cohorts, and which has also been assigned to a biological process functional class demonstrating significant enrichment in both the discovery and replication cohorts.

^cAutism CNV genes for which at least one autism-specific CNV overlapping the gene was identified in both the discovery and replication cohorts, and which has also been assigned to a cellular component functional class demonstrating significant enrichment in both the discovery and replication cohorts.

^dAutism CNV genes for which at least one autism-specific CNV overlapping the gene was identified in both the discovery and replication cohorts, and which has also been assigned to a mouse phenotype functional class demonstrating significant enrichment in the combined cohort.

^eAutism CNV genes for which at least one autism-specific CNV overlapping the gene was identified in the combined cohort, and which has also been assigned to at least one biological process, cellular component and mouse phenotype functional class demonstrating significant enrichment in the combined cohort.

attention deficit hyperactivity disorder,¹⁸ may be a viable approach for identifying risk signatures across functionally related processes. Caveats of this approach include insufficient specificity to identify individual candidate genes within moderate cohort sizes, and reliance upon incomplete and potentially biased gene knowledge. In addition, our approach only considered ASD variants not found in controls, which may oversimplify the biological role that certain variants might contribute to ASD etiology and is also dependent upon control cohort size. However, our results do consistently and independently implicate possible pathways, functions and gene sets that can be readily investigated by follow-up genomic and experimental investigation. Recent studies have implicated individual genes associated with synaptic function, including several neuroligands, as well as indications of glutamatergic pathway involvement.^{8,13} Our current results extend these findings and, more importantly, for the first time provide statistical evidence that synaptic function and glutamate-mediated neurotransmission are contributing factors in autism etiology.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

We acknowledge the resources provided by the Autism Genetic Resource Exchange (AGRE) consortium and participating AGRE families, and members of the Center for Applied Genomics at the Children's Hospital of Philadelphia for making genotype data available. The Autism Genetic Resource Exchange is a program of Autism Speaks and is supported, in part, by Grant 1U24MH081810 from the National Institute of Mental Health to Clara M Lajonchere (PI). This work was supported in part by Pennsylvania Department of Health Grant SAP 4100037707 (PSW), NIH grants GM081519 (THS) and P30HD026979 (MD and XG) and the Seaver Foundation (JDB). We thank Marcella Devoto for helpful comments. We also thank all the participating individuals and families of the AGRE study and for all the control families at the

Children's Hospital of Philadelphia for making this study possible.

Author contributions: THS, HH, JDB and BG collected samples and contributed phenotype data. XG, HMX, DDG, MD, JDB, JE and PSW designed the experiments. XG, HMX, JCP, KM, ASW, MD, RJO and PSW performed the experiments and analyzed the data. XG, JDB, JE and PSW reviewed the data, assisted with interpretation of the data and edited the manuscript.

References

- Association AP. *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association: Washington, DC, 2004.
- Gupta AR, State MW. Recent advances in the genetics of autism. *Biol Psychiatry* 2007; **61**: 429–437.
- Muhle R, Trentacoste SV, Rapin I. The genetics of autism. *Pediatrics* 2004; **113**: e472–e486.
- Folstein S, Rutter M. Genetic influences and infantile autism. *Nature* 1977; **265**: 726–728.
- Ronald A, Happe F, Bolton P, Butcher LM, Price TS, Wheelwright S *et al*. Genetic heterogeneity between the three components of the autism spectrum: a twin study. *J Am Acad Child Adolesc Psychiatry* 2006; **45**: 691–699.
- Alarcon M, Abrahams BS, Stone JL, Duvall JA, Perederiy JV, Bomar JM *et al*. Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. *Am J Hum Genet* 2008; **82**: 150–159.
- Arking DE, Cutler DJ, Brune CW, Teslovich TM, West K, Ikeda M *et al*. A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism. *Am J Hum Genet* 2008; **82**: 160–164.
- Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu XQ *et al*. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* 2007; **39**: 319–328.
- Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS *et al*. Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* 2009; **459**: 528–533.
- Weiss LA. Autism genetics: emerging data from genome-wide copy-number and single nucleotide polymorphism scans. *Expert Rev Mol Diagn* 2009; **9**: 795–803.
- Weiss LA, Arking DE, Daly MJ, Chakravarti A. A genome-wide linkage and association scan reveals novel loci for autism. *Nature* 2009; **461**: 802–808.
- Merikangas AK, Corvin AP, Gallagher L. Copy-number variants in neurodevelopmental disorders: promises and challenges. *Trends Genet* 2009; **25**: 536–544.
- Betancur C, Sakurai T, Buxbaum JD. The emerging role of synaptic cell-adhesion pathways in the pathogenesis of autism spectrum disorders. *Trends Neurosci* 2009; **32**: 402–412.
- Zoghbi HY. Postnatal neurodevelopmental disorders: meeting at the synapse? *Science* 2003; **302**: 826–830.
- Shaikh TH, Gai X, Perin JC, Glessner JT, Xie H, Murphy K *et al*. High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. *Genome Res* 2009; **19**: 1682–1690.
- Hakonarson H, Grant SF, Bradfield JP, Marchand L, Kim CE, Glessner JT *et al*. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature* 2007; **448**: 591–594.
- Yang N, Li H, Criswell LA, Gregersen PK, Alarcon-Riquelme ME, Kittles R *et al*. Examination of ancestry and ethnic affiliation using highly informative diallelic DNA markers: application to diverse and admixed populations and implications for clinical epidemiology and forensic medicine. *Hum Genet* 2005; **118**: 382–392.
- Elia J, Gai X, Xie HM, Perin JC, Geiger E, Glessner JT *et al*. Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. *Mol Psychiatry* 2010; **15**: 637–646.
- Gai X, Perin JC, Murphy K, O'Hara R, D'Arcy M, Wenocur A *et al*. CNV Workshop: an integrated platform for high-throughput copy number variation discovery and clinical diagnostics. *BMC Bioinformatics* 2010; **11**: 74.
- Conlin LK, Thiel BD, Bonnemant CG, Medne L, Ernst LM, Zackai EH *et al*. Mechanisms of mosaicism, chimerism and uniparental disomy identified by single nucleotide polymorphism array analysis. *Hum Mol Genet* 2010; **19**: 1263–1275.
- Hasin Y, Olender T, Khen M, Gonzaga-Jauregui C, Kim PM, Urban AE *et al*. High-resolution copy-number variation map reflects human olfactory receptor diversity and evolution. *PLoS Genet* 2008; **4**: e1000249.
- Nozawa M, Nei M. Evolutionary dynamics of olfactory receptor genes in Drosophila species. *Proc Natl Acad Sci USA* 2007; **104**: 7122–7127.
- Young JM, Endicott RM, Parghi SS, Walker M, Kidd JM, Trask BJ. Extensive copy-number variation of the human olfactory receptor gene family. *Am J Hum Genet* 2008; **83**: 228–242.
- Hellemsans J, Mortier G, De Paepe A, Speleman F, Vandensompele J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* 2007; **8**: R19.
- Geschwind DH, Sowiński J, Lord C, Iversen P, Shestack J, Jones P *et al*. The autism genetic resource exchange: a resource for the study of autism and related neuropsychiatric conditions. *Am J Hum Genet* 2001; **69**: 463–466.
- Smith CL, Goldsmith CA, Eppig JT. The mammalian phenotype ontology as a tool for annotating, analyzing and comparing phenotypic information. *Genome Biol* 2005; **6**: R7.
- Bucan M, Abrahams BS, Wang K, Glessner JT, Herman EI, Sonnenblick LI *et al*. Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes. *PLoS Genet* 2009; **5**: e1000536.
- Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S *et al*. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 2009; **459**: 569–573.
- Cao M, Xu J, Shen C, Kam C, Hugarir RL, Xia J. PICK1-ICA69 heteromeric BAR domain complex regulates synaptic targeting and surface expression of AMPA receptors. *J Neurosci* 2007; **27**: 12945–12956.
- Missler M, Hammer RE, Sudhof TC. Neurexophilin binding to alpha-neurexins. A single LNS domain functions as an independently folding ligand-binding unit. *J Biol Chem* 1998; **273**: 34716–34723.
- Barnby G, Abbott A, Sykes N, Morris A, Weeks DE, Mott R *et al*. Candidate-gene screening and association analysis at the autism-susceptibility locus on chromosome 16p: evidence of association at GRIN2A and ABAT. *Am J Hum Genet* 2005; **76**: 950–966.
- Kebir O, Tabbane K, Sengupta S, Joobar R. Candidate genes and neuropsychological phenotypes in children with ADHD: review of association studies. *J Psychiatry Neurosci* 2009; **34**: 88–101.
- Lang UE, Puls I, Muller DJ, Strutz-Seebohm N, Gallinat J. Molecular mechanisms of schizophrenia. *Cell Physiol Biochem* 2007; **20**: 687–702.
- Tang J, Chen X, Xu X, Wu R, Zhao J, Hu Z *et al*. Significant linkage and association between a functional (GT)_n polymorphism in promoter of the N-methyl-D-aspartate receptor subunit gene (GRIN2A) and schizophrenia. *Neurosci Lett* 2006; **409**: 80–82.
- Sprengel R, Suchanek B, Amico C, Brusa R, Burnashev N, Rozov A *et al*. Importance of the intracellular domain of NR2 subunits for NMDA receptor function in vivo. *Cell* 1998; **92**: 279–289.
- Mei L, Xiong WC. Neuregulin 1 in neural development, synaptic plasticity and schizophrenia. *Nat Rev Neurosci* 2008; **9**: 437–452.
- Augustin I, Rosenmund C, Sudhof TC, Brose N. Munc13-1 is essential for fusion competence of glutamatergic synaptic vesicles. *Nature* 1999; **400**: 457–461.
- Missler M, Zhang W, Rohlmann A, Kattenstroth G, Hammer RE, Gottmann K *et al*. Alpha-neurexins couple Ca²⁺ channels to synaptic vesicle exocytosis. *Nature* 2003; **423**: 939–948.
- Kao HT, Porton B, Czernik AJ, Feng J, Yiu G, Haring M *et al*. A third member of the synapsin gene family. *Proc Natl Acad Sci USA* 1998; **95**: 4667–4672.

- 40 Feng J, Chi P, Blanpied TA, Xu Y, Magarinos AM, Ferreira A *et al*. Regulation of neurotransmitter release by synapsin III. *J Neurosci* 2002; **22**: 4372–4380.
- 41 Porton B, Rodriguiz RM, Phillips LE, Gilbert JWT, Feng J, Greengard P *et al*. Mice lacking synapsin III show abnormalities in explicit memory and conditioned fear. *Genes Brain Behav* 2010; **9**: 257–268.
- 42 Misonou H, Menegola M, Buchwalder L, Park EW, Meredith A, Rhodes KJ *et al*. Immunolocalization of the Ca²⁺-activated K⁺ channel Slo1 in axons and nerve terminals of mammalian brain and cultured neurons. *J Comp Neurol* 2006; **496**: 289–302.
- 43 Sausbier M, Hu H, Arntz C, Feil S, Kamm S, Adelsberger H *et al*. Cerebellar ataxia and Purkinje cell dysfunction caused by Ca²⁺-activated K⁺ channel deficiency. *Proc Natl Acad Sci USA* 2004; **101**: 9474–9478.
- 44 Laumonier F, Roger S, Guerin P, Molinari F, M'Rad R, Cahard D *et al*. Association of a functional deficit of the BKCa channel, a synaptic regulator of neuronal excitability, with autism and mental retardation. *Am J Psychiatry* 2006; **163**: 1622–1629.
- 45 Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R *et al*. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 2010; **466**: 368–372.



This work is licensed under the Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>

Supplementary Information accompanies the paper on the Molecular Psychiatry website (<http://www.nature.com/mp>)