Evaluation of copy number variations reveals novel candidate genes in autism spectrum disorder-associated pathways

Anthony J. Griswold1,†, Deqiong Ma1,2,†, Holly N. Cukier1,†, Laura D. Nations1, Mike A. Schmidt1,2, Ren-Hua Chung1,2, James M. Jaworski1, Daria Salyakina1, Ioanna Konidari1, Patrice L. Whitehead1, Harry H. Wright3, Ruth K. Abramson3, Scott M. Williams4, Ramkumar Menon5, Eden R. Martin1,2, Jonathan L. Haines4, John R. Gilbert1,2, Michael L. Cuccaro1,2 and Margaret A. Pericak-Vance1,2,*

1John P. Hussman Institute for Human Genomics and 2Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL, USA, 3Department of Neuropsychiatry and Behavioral Science, School of Medicine, University of South Carolina, Columbia, SC, USA, 4Center for Human Genetics Research, Vanderbilt University, Nashville, TN, USA and 5Division of Maternal-Fetal Medicine Perinatal Research, Department of Obstetrics and Gynecology, The University of Texas Medical Branch at Galveston, Galveston, TX, USA

Received January 17, 2012; Revised March 9, 2012; Accepted April 20, 2012

Autism spectrum disorders (ASDs) are highly heritable, yet relatively few associated genetic loci have been replicated. Copy number variations (CNVs) have been implicated in autism; however, the majority of loci contribute to <1% of the disease population. Therefore, independent studies are important to refine associated CNV regions and discover novel susceptibility genes. In this study, a genome-wide SNP array was utilized for CNV detection by two distinct algorithms in a European ancestry case–control data set. We identify a significantly higher burden in the number and size of deletions, and disrupting more genes in ASD cases. Moreover, 18 deletions larger than 1 Mb were detected exclusively in cases, implicating novel regions at 2q22.1, 3p26.3, 4q12 and 14q23. Case-specific CNVs provided further evidence for pathways previously implicated in ASDs, revealing new candidate genes within the GABAergic signaling and neural development pathways. These include DBI, an allosteric binder of GABA receptors, GABARAPL1, the GABA receptor-associated protein, and SLC6A11, a postsynaptic GABA transporter. We also identified CNVs in COBL, deletions of which cause defects in neuronal cytoskeleton morphogenesis in model vertebrates, and DNER, a neuron-specific Notch ligand required for cerebellar development. Moreover, we found evidence of genetic overlap between ASDs and other neurodevelopmental and neuropsychiatric diseases. These genes include glutamate receptors (GRID1, GRIK2 and GRIK4), synaptic regulators (NRXN3, SLC6A8 and SYN3), transcription factor (ZNF804A) and RNA-binding protein FMR1. Taken together, these CNVs may be a few of the missing pieces of ASD heritability and lead to discovering novel etiological mechanisms.

†The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

© The Author 2012. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
INTRODUCTION

Autism spectrum disorders (ASDs) are among the most heritable of neurodevelopmental disorders (1–4). Given their strong genetic component, gene discovery in ASDs has been the focus of research for the past decade. To date, no single gene has been shown to account for a majority of ASD susceptibility. The limited success of gene discovery has been attributed to the genetic heterogeneity and phenotypic variability of ASDs. In fact, hundreds of candidate genes have been implicated. These candidates include rare variations with large effect sizes in syndromic forms of ASD, including tuberous sclerosis (TSC1 and TSC2) (5,6), Rett syndrome (MECP2) (7) and Phelan–McDermid syndrome (SHANK3) (8). In addition, despite fruitful genome-wide association studies (GWAS) in many other genetic disorders (9), this approach in studying ASDs has revealed only a few associated common alleles with small effect sizes, many of which have not yet been replicated in independent studies (10–14). Therefore, studying low frequency variants with variable expression or reduced penetrance may be necessary to identify predisposing alleles.

Among aberrations of interest are copy number variations (CNVs). They represent a significant source of genetic diversity and contribute to disease susceptibility for several neuro-behavioral phenotypes, including ASDs, mental retardation and schizophrenia (15–22). The accelerated discovery of CNVs coinciding with ASDs has provided evidence of their involvement in ASD etiology, although individual CNVs generally account for <1% each of attributable risk (23). Some CNVs have pleiotropic effects that associate with susceptibility to several neuropsychiatric disorders (12,20,24–32). Therefore, careful characterization of CNVs in ASD patients represents an important aspect of genetic research.

Gene-set enrichment analyses from both GWAS (33) and CNV studies (20,31) have revealed functional gene families and molecular pathways likely to contribute to the pathogenesis of ASDs. However, the gene-set enrichment statistical framework, particularly as applied to CNVs, is limited and potentially confounded (34). A joint analysis of two autism GWAS data sets implicated neuronal growth and axon guidance as well as cytoskeletal regulation and neuronal cell signaling (33). Among CNV gene-set analyses, the ubiquitin pathway and genes involved in neuronal cell adhesion have been consistently implicated in ASDs (20,31). A recent CNV study from the Autism Genome Project (AGP) identified an enrichment of CNVs disrupting GTPase/Ras signaling genes, and confirmed the involvement of genes responsible for cellular proliferation, projection and motility (20). Moreover, a recent study utilizing pathway analysis of de novo CNVs has implicated genes directing synapse development, axon targeting and neuron motility in autism risk (35,36). Therefore, given the heterogeneous nature of ASDs and their potential genetic causes, disruption of additional genes in these and related molecular pathways are expected to be involved in ASDs.

Further studies are required not only to refine our understanding of the roles of previously isolated genes and regions in ASDs but also to identify novel candidates within implicated molecular pathways. In this study, we used a combined family and case–control Caucasian data set to support previously associated ASD CNVs and identify unique, large and potentially pathogenic deletions as well as novel genes within molecular pathways previously connected to ASDs.

RESULTS

Global CNV burden analysis

In an attempt to recognize overall differences in CNVs between cases and controls, high confidence CNVs from 813 unrelated ASD cases and 592 control Caucasian samples (Table 1) were included in the CNV burden analysis following extensive CNV QC filtering (Table 2). To eliminate potential biases from various DNA sources and gender between cases and controls, burden analysis for deletions and duplications was performed separately and only for autosomes (see details in Materials and Methods).

We detected a significantly higher burden in the number and size of deletions in cases for both the total and rare (<1% frequency) CNV sets (Table 2) (all deletions: size: \( P = 9.90 \times 10^{-6} \); number: \( P = 2.00 \times 10^{-5} \); rare deletions: size: \( P = 0.04 \); number: \( P = 9.90 \times 10^{-6} \) ). In addition, the average number of genes disrupted by deletions per individual was significantly greater in cases when compared with controls (the number of genes disrupted: all deletions: \( P = 9.90 \times 10^{-6} \); rare deletions: \( P = 9.90 \times 10^{-6} \) ). We also observed an increase in the number of rare genic duplications (2.7 per case versus 2.2 per control, \( P = 0.01 \) ), although the average size of the duplications was not different (176.6 kb in cases versus 201.1 kb controls, \( P = 0.95 \) ).

CNVs identified in previously implicated ASD regions

To identify specific CNVs which may contribute to autism risk, we first examined CNVs in eight well-established ASD associated regions (1q21, 7q11.23, 15q13, 15q11–13, 16p11.2, 17p11.2, 22q13.3 and 22q11.21). We detected and validated nine CNVs from these regions that overlap at least 50% of their length with regions previously implicated in autism risk within our data set, including reciprocal duplications and a deletion at 16p11.2. To determine whether these CNVs were concordant with disease, we determined the inheritance status in all families and segregation of variants within multiplex ASD families using the genotyping data from available family members and quantitative polymerase chain reaction (qPCR) (Table 3, Supplementary Material, Fig. S1).
Among these validated CNVs (Supplementary Material, Fig. S2) were two 5 Mb duplication at 15q11–13, one maternally inherited and the other de novo and a de novo 4 Mb deletion located at 22q13.3 which removes SHANK3 and results in Phelan-McDermid syndrome. Other CNVs included an 1.4 Mb duplication at 7q11.23, an 1 Mb duplication at 22q11.21, two 900 kb duplications of 1q21.1, a 3 Mb deletion at 17p11.2, two 550 kb duplications and a reciprocal deletion of 16p11.2 and two 500 kb deletions at 15q13.3.

**Identification of large, novel deletions**

Since no deletions >1 Mb were found in controls, we examined this subset of large and possibly pathogenic structural variations identified exclusively in cases. Eighteen such deletions ranging from 1.02 to 7.31 Mb were detected in 10 simplex families and 6 multiplex families (Table 4). Familial genotyping and CNV data as well as qPCR (Supplementary Material, Fig. S3) validation revealed that 7 of the 18 deletions were de novo and 10 were inherited (5 paternally, 5 maternally), while the inheritance status for one CNV could not be determined. Five of the 18 deletions overlapped with well-established ASD regions reported in Table 3. Five others have been identified in autistic patients in studies by the AGP (20,37). Independent autism CNV studies identified deletions at 5q35.3 (38), 10p15.2 (39), 10q11.22–23 (29) and Xp22.31 (40) that physically overlap at least half of a CNV found in our data. This left four novel regions, each found in a single affected individual from a simplex family, including a paternally inherited deletion at 3p26.3 and de novo deletions at 2q22.1, 4q12 and 14q23 that do not overlap >20% of their length with any deletion previously cataloged in the Database of Genomic Variants (http://projects.tcag.ca/variation/) or the Autism CNV Database (http://projects.tcag.ca/autism_500k/).

**Identification of potentially pathogenic CNVs**

Pathway enrichment analysis of case-specific CNVs did not reveal any gene ontology or KEGG pathways which were significantly overrepresented in our CNV set. Therefore, we evaluated our data to enrich for potentially pathogenic CNVs in ASDs. To do this, case-unique CNVs were selected. A total of 3855 CNVs met this criterion. Genes located within these CNV regions were compiled and their functions assessed by the literature review. First, CNVs in several genes previously associated with the autism risk, including deletions in A2BP1, APBA2, GRIN2B and NRXN3 were identified in this way. Secondly, several novel candidate genes in molecular pathways or genetic networks implicated in autism etiology were identified, including genes in the GABAergic signaling pathway (DBI, GABARAPL1 and SLC6A11) and genes acting in neural development and function (COBL, DNER and...
DISCUSSION

In the present study, we identified several potentially pathogenic candidate regions and genes that may contribute to ASD etiology. At the global level, our results replicate a recent study (28) that indicated an overall increase in the burden of rare genic deletions and more overall deletions (20). These differences may reflect data set-specific variation in factors, such as the ASD diagnostic criteria, ratios of simplex to multiplex families, regional or other sampling biases as well as array platform and DNA source-origin biases.

Beyond identifying a higher overall burden of deletions in ASD cases compared with controls, we detected and validated CNVs at eight regions that have been previously associated with autistic traits; 1q21, 7q11.23, 15q13, 15q11–13, 17p11.2, 17p11.2, 22q13.3 and 22q11.21 (28,41–47). While all of these CNVs have been previously implicated by classical cytogenetic techniques, recent large-scale CNV studies using both genotyping and array CGH have also isolated these same CNVs as associated with autism (35,48,49). In our data, we identified two cases overlapping a classic cytogenetic aberration associated with autism, the 5 Mb maternally inherited 15q11–13 duplication. This duplication, which occurs in ~1% of all autism cases (44,50), falls between low copy repeats on chromosome 15 and duplicates many genes, including three GABA receptors. We also detected CNVs within regions known to be syndromic causes of autism; two duplications and one de novo deletion of 550 kb on 16p11.2 (28,30,51) and two duplications of 900 kb on 1q21.1 (41), both of which are recurrent CNVs due to their location between segmental duplications. In addition, a 3 Mb de novo deletion on 17p11.2, which deletes RAII, has been previously implicated in both Smith-Magenis syndrome and an autistic syndrome (52). Additionally, one case carrying a complex 1.4 Mb duplication between segmental duplications at 7q11.23 partially overlaps the region deleted in Williams–Beuren syndrome in which duplications have previously been associated with autism (42,53). Finally, another individual carries the same 4 Mb deletion on 22q11.2 that causes Phelan–McDermid syndrome and contains the autism gene SHANK3 (54,55). Moreover, several large CNVs that affect previously reported candidate genes were also identified. A 1 Mb duplication at 22q11.21 encompasses the entire COMT gene that has been implicated in developmental delay and autism (45,47,56,57). Two patients carry deletions of 15q13.1–3, thereby removing either the entire gene or several exons of CHRNA7, deletions of which have been shown to be involved in developmental delay, mental retardation and epilepsy (43,58,59). While all of these CNVs have strong evidence for being causative of ASDs, discordance within multiplex families has been previously noted for duplications and deletions on 16p11.2 (30,60) and duplications of 22q11.21 (61). Given such evidence, the lack of perfect concordance between these established CNVs and disease in some of our multiplex families (Supplementary Material, Fig. S1) suggests that there are probably additional genetic components contributing to disease in these families.

Besides the 5 large deletions discussed above (two at 15q13.1, 15q13.1–3, 17p11.2 and 22q12.31-33), 13 other...
Table 5. Case-unique CNVs in ASD candidate genes

<table>
<thead>
<tr>
<th>Position (hg 18)</th>
<th>Cytoband</th>
<th>Genes</th>
<th>CNV type</th>
<th>Inheritance</th>
<th>Family</th>
<th>Affected siblings with CNV</th>
<th>Unaffected siblings with CNV</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr2:184744259–186437079</td>
<td>2q32.1</td>
<td>FLJ44048,FSIP2,ZNF804A</td>
<td>Duplication</td>
<td>Paternal</td>
<td>7678</td>
<td>2/2</td>
<td>0/0</td>
</tr>
<tr>
<td>chr2:119648273–119882278</td>
<td>2q14.2</td>
<td>C2orf76,DL1,STEAP3</td>
<td>Duplication</td>
<td>De novo</td>
<td>17113</td>
<td>1/1</td>
<td>0/0</td>
</tr>
<tr>
<td>chr2:230078358–230181261</td>
<td>2q36.3</td>
<td>DNER</td>
<td>Deletion</td>
<td>Paternal</td>
<td>7171</td>
<td>1/1</td>
<td>0/1</td>
</tr>
<tr>
<td>chr3:1083410–1094317</td>
<td>3p25.3</td>
<td>SLC6A11</td>
<td>Deletion</td>
<td>Paternal</td>
<td>7804</td>
<td>2/2</td>
<td>0/0</td>
</tr>
<tr>
<td>chr10:101260386–102979782</td>
<td>6q16.3</td>
<td>ASCC3,GRIK2</td>
<td>Deletion</td>
<td>Paternal</td>
<td>17182</td>
<td>2/2</td>
<td>0/0</td>
</tr>
<tr>
<td>chr7:51155279–51263037</td>
<td>7p12.1</td>
<td>COBL</td>
<td>Deletion</td>
<td>Paternal</td>
<td>39391</td>
<td>1/1</td>
<td>1/3</td>
</tr>
<tr>
<td>chr10:87754758–87839034</td>
<td>10q23.1</td>
<td>GRID1</td>
<td>Deletion</td>
<td>Paternal</td>
<td>37276</td>
<td>1/1</td>
<td>1/1</td>
</tr>
<tr>
<td>chr11:118685526–121040148</td>
<td>11q23.3–q24.1</td>
<td>ARHGDEF12,C1QTNTF5,GRIK4,MCAM,</td>
<td>Duplication</td>
<td>De novo</td>
<td>37341</td>
<td>1/1</td>
<td>0/1</td>
</tr>
<tr>
<td>chr12:10248041–10332152</td>
<td>12p13.2</td>
<td>GABARAP1</td>
<td>Deletion</td>
<td>Paternal</td>
<td>37240</td>
<td>1/1</td>
<td>1/1</td>
</tr>
<tr>
<td>chr12:13753741–13788728</td>
<td>12p13.1</td>
<td>GRIN2B</td>
<td>Deletion</td>
<td>Maternal</td>
<td>7614</td>
<td>1/1</td>
<td>0/0</td>
</tr>
<tr>
<td>chr14:78373505–78506223</td>
<td>14q24.3–q31.1</td>
<td>NRXN3</td>
<td>Deletion</td>
<td>Maternal</td>
<td>7484</td>
<td>1/1</td>
<td>1/2</td>
</tr>
<tr>
<td>chr15:26849239–28157206</td>
<td>15q13.1–q13.2</td>
<td>APBA2,KAIA0574,NDNL2,TP1</td>
<td>Deletion</td>
<td>Maternal</td>
<td>17630</td>
<td>2/2</td>
<td>0/1</td>
</tr>
<tr>
<td>chr16:6866262–7042998</td>
<td>16p13.2</td>
<td>ASBP1</td>
<td>Deletion</td>
<td>Paternal</td>
<td>7636</td>
<td>2/2</td>
<td>0/0</td>
</tr>
<tr>
<td>chr17:231499975–31930759</td>
<td>22q12.3</td>
<td>SYN3,TIMP3</td>
<td>Duplication</td>
<td>Paternal</td>
<td>39377</td>
<td>1/1</td>
<td>0/0</td>
</tr>
<tr>
<td>chrX:114667441–146930522</td>
<td>Xq27.3–q28</td>
<td>ASFMRI,FMRI1,FMRI1,FMRI1N8</td>
<td>Deletion</td>
<td>Paternal</td>
<td>18072</td>
<td>2/2</td>
<td>0/0</td>
</tr>
</tbody>
</table>

aNumber of affected siblings in the family with the CNV/total affected siblings.
bNumber of unaffected siblings in the family with the CNV/total unaffected siblings.
cUnaffected with CNV has speech delay.
dUnaffected with CNV has a questionable autism diagnosis.

delictions >1 Mb were observed exclusively in cases. These are of particular interest due to their possible pathogenic effect in neuropsychiatric diseases (62). Nine of these 12 have been identified as ASD candidate regions in previous CNV studies (20,29,37–40). The detection of deletions in these regions (5q23.1, 5q35.3, 10p15.2, 10q11.22-23, 13q12.12, 13q33.1, two at 17p12, and Xp22.31) lend further support to their harboring ASD pathogenic variations. Moreover, the mechanism of formation of these deletions, with the exception of 5q23.1, is predicted to be non-allelic homologous recombination due to the presence of flanking segmental duplications. Interestingly, four large deletions (2q22.1, 3p26.3, 4q12 and 14q23) are not flanked by known genomic instability elements and represent potentially novel contributors to autism etiology. A de novo deletion on 2q22.1 includes HNMT, a gene which regulates histamine metabolism in the brain and has been previously implicated in anxiety and alcoholism (63). We also identify a paternal deletion on 3p26.3 that deletes the neural adhesion molecules CHL1 and CNTN6, which is near a previously reported autism-associated CNV (64). Furthermore, genes involved in circadian rhythms have been implicated in autism (65) and a new candidate, CLOCK, was removed in a de novo deletion on 4q12. Finally, pleiotropic effects of these novel, large deletions were found in an autistic patient with pherocytosis carrying a de novo 14q23 deletion that removes several neurodevelopmental genes, including CHURC1, MTHFD1 and PLEKHG3 as well as the red blood cell structural gene SPTB (66). As such, these substantial deletions point toward dysfunction of neuron cell structure and adhesion as well as circadian regulation as contributors to autism.

In an attempt to find potentially pathogenic genes contributing to autism risk, we first used a pathway enrichment strategy to identify functional molecular networks which were overly represented by CNVs in cases. However, we did not identify a statistical overrepresentation of any category. While large-scale CNV studies report enrichment in their data sets (36,49), other studies have suggested that since the causes of autism are quite diverse, no singular mechanism or pathway is likely to explain the entire spectrum of risk (35). Therefore, we chose to search for novel candidate genes or gene clusters in rare, case-unique CNVs. This analysis revealed several genes previously associated with autism risk, including A2BP1 (67), APBA2 (68), GRIN2B (69,70), NRXN3 (71) and SHANK3 (72,73). Of particular interest among these genes is the splice factor A2BP1 which has not only been associated with autism but regulates the expression of many genes misregulated in post-mortem autism brains (74). Novel members of pathways previously implicated in autism include three genes in the GABAergic signaling pathway (75–77). Due to the prior evidence of the importance of GABAergic genes in autism, it is possible that duplication of DI, an allosteric binder of GABRA receptors (79), deletion of the GABRA receptor-associated protein GABARAP1 (80) or deletion of the first eight exons of SLC6A11(GAT-3), a postsynaptic GABA transporter (81), could affect signaling and lead to deficits in neuronal function. The role of each of these genes in ASDs warrants further investigation.
Beyond the GABAergic system, rare CNVs disrupting genes which play a role in neural development have been implicated in autism etiology (20). We identified case-unique CNVs that traverse genes affecting these processes. Our findings further implicate genes involved in neuronal cytoskeletal formation as we find a CNV which deletes exons 2–5 of COBL, knockouts of which show severe defects in neural cell cytoskeleton morphogenesis (82). In addition, our CNV analyses implicated signaling factors involved in neural development. We detected a duplication of exons 2–6 of DNER, a neuron-specific Notch ligand required for cerebellar development (83,84), containing a marker previously associated with autism (33), and deletion of the entire GDNF gene, a growth factor which fosters neuronal cell survival (85). The discovery of case-unique CNVs within these novel genes in previously associated molecular networks supports the importance of these pathways in autism as well as implicating new potential genetic risk factors, making them attractive candidates for follow-up.

Our data also identify several genes with neuronal functions that have been previously connected to other neuropsychiatric diseases. Such genetic overlap has been noted and previously discussed (86). We found CNVs in genes associated with schizophrenia, including two glutamate receptors, the first deleting the entire GRIK2 gene (87,88) and the second being an intronic deletion in GRID1 (89), a duplication of the first five exons of the synaptic protein SYN3 (90), a deletion of two exons of the neuronal adhesion molecule NRXN3 and duplication of the zinc finger ZNF804A (91). We also found CNVs in genes involved in other neurological disorders. The neurotransmitter SLC6A8 has been implicated in mental retardation (92) and we identified a deletion of its first eight exons. Expansions in the 5′ untranslated region of FMR1 are known to cause both Fragile X syndrome and Fragile X-associated tremor/ataxia syndrome (93) and we found a deletion of this entire gene. The overlap of these genes with other neuropsychiatric diseases suggests an etiological connection between these disorders and ASDs and points to the disruption of these genes as having broad neurological phenotypes.

In conclusion, we found there to be a significantly higher burden in the number and size of deletions carried by ASD individuals when compared with controls. Among the CNVs identified were several that overlapped with well-established autism-associated regions and candidate genes, further supporting their roles in autism etiology. Moreover, we isolated four large, novel deletions on 2q22.1, 3p26.3, 4q12 and 14q23 that include new genes and regions linked to ASDs. These findings provide further evidence for the involvement of previously molecular pathways, including the GABAergic signaling and neurodevelopmental pathways, while revealing several novel genes that may lead to new etiological mechanisms.

**MATERIALS AND METHODS**

**Ascertainment of ASD families**

This study is part of the Collaborative Autism Project (CAP), for which a total of 1084 simplex and multiplex autism families were ascertained with a range of self-reported ethnicities through clinical groups at the John P. Hussman Institute for Human Genomics at the University of Miami (HIHG, Miami, FL, USA), the University of South Carolina (Columbia, SC, USA) and the Center for Human Genetics Research at Vanderbilt University (Nashville, TN, USA). Participating families were enrolled via support groups, advertisements and clinical and educational settings. All participants and families were ascertained using a standard protocol approved by the appropriate Institutional Review Boards. Written informed consent was obtained from parents as well as from minors who were able to give informed consent; for individuals unable to give consent due to age or developmental problems, assent was obtained whenever possible. Core inclusion criteria were as follows: (i) chronological age between 3 and 21 years of age, (ii) a presumptive clinical diagnosis of autism, (iii) expert clinical determination of an autism diagnosis using DSM-IV criteria supported by the Autism Diagnostic Interview-Revised (ADI-R) in the majority of cases and all available clinical information (94), (iv) minimal developmental level of 18 months as determined by the Vineland Adaptive Behavior Scale (VABS) (95) or the VABS-II (96) or IQ equivalent > 35. These minimal developmental levels assure that the ADI-R results are valid and reduce the likelihood of including individuals with severe mental retardation only. We excluded participants with severe sensory problems (e.g., visual impairment or hearing loss), significant motor impairments (e.g., failure to sit by 12 months or walk by 24 months) or who had a previously identified metabolic, genetic or progressive neurological disorders.

**Ascertainment of pediatric controls**

Eight hundred ninety-six healthy children, including 664 self-reported Caucasians, were recruited as pediatric controls from two centers. Of these, 565 individuals aged 4 to 21 years old were sampled through the John. P. Hussman Institute for Human Genomics (Miami, FL, USA). Participants were screened for eligibility by a series of preliminary questions to determine whether the individual either had been diagnosed with or had a parent or sibling with a developmental, behavioral, neurological or other disability or physical conditions. If none of those conditions were present, parents of young children or the participants were informed and signed the informed consent and completed the Social Communication Questionnaire (97) to screen for potential ASDs. The remaining 331 controls were part of a preterm birth study. Cord blood was collected at the Centennial Medical Center (Nashville, TN, USA) from women aged from 18 to 40 years old with term pregnancies (>37 weeks gestation), live, singleton births.

**Genotyping**

Genomic DNA was purified from whole blood, cell line and saliva samples using Puregene chemistry on the Qiagen Autopure LS according to standard automated Qiagen protocols (Qiagen, Valencia, CA, USA). DNA samples were quantitated via the ND-8000 spectrophotometer and DNA quality was evaluated by gel electrophoresis on a 0.8% agarose gel. The concentration for all qualified samples was normalized.
to 50 ng/μl and samples were arrayed in Matrix 0.5 ml 2D bar-coded tubes in racks of 96. Sample identity was confirmed by genotyping eight SNPs using TaqMan allelic discrimination assays (Applied Biosystems, Foster City, CA, USA) and assessing for concordance with historical data. All recruited participants, available siblings and parents and recruited controls were genotyped using either Illumina’s Human 1M-v1_C Beadchip or the 1M-DuoV3 Beadchip. The samples were processed according to Illumina Procedures for the Infinium II® assay (Illumina Inc., San Diego, CA, USA). The protocol was automated using the Tecan EVO-1 to further enhance efficiency and consistency (Tecan Group Ltd, Männedorf, Switzerland). A single quality-control DNA sample was tested during each run to ensure reproducibility of results between runs. Data were extracted with the Illumina® Beadstudio software from data files created in the Illumina BeadArray reader. Samples and markers with call rates below 95% were excluded from analysis and a GenCall cutoff score of 0.15 was used for all Infinium II® products.

Genotyping and CNV quality control (QC)

Out of 1084 autistic families and 896 pediatric controls, samples with a genotype call rate >97% and only cases without excessive Mendelian inconsistencies within the pedigree (>2% among all tested SNPs) were processed further for CNV QC. Following this family-based sample QC, 1063 families, encompassing 1301 ASD cases, remained for CNV calling. The signal intensity QC used the parameters generated for CNV QC. Following this family-based sample QC, 1063 families. In addition, two samples were taken out due to analysis, but later used to examine concordance in multiplex families. Therefore, 184 affected siblings were removed from the initial CNV analysis stratification checking. To maintain independence of this significant portion of our sample.

Next, unrelated cases and controls were selected for population stratification checking. To maintain independence of this data set, only one ASD case per family was analyzed. Therefore, 184 affected siblings were removed from the initial CNV analysis, but later used to examine concordance in multiplex families. In addition, two samples were taken out due to overlap with the previously published AGP data set (20). One trisomy X female and one duplicated Xp male were also removed. These steps left 986 cases and 771 controls fit the above criteria suggesting that overall data quality would not remove a significant portion of our sample.

Next, unrelated cases and controls were selected for population stratification checking. To maintain independence of this data set, only one ASD case per family was analyzed. Therefore, 184 affected siblings were removed from the initial CNV analysis, but later used to examine concordance in multiplex families. In addition, two samples were taken out due to overlap with the previously published AGP data set (20). One trisomy X female and one duplicated Xp male were also removed. These steps left 986 cases and 771 controls fit the above criteria suggesting that overall data quality would not remove a significant portion of our sample.

Finally, high confidence CNV calls were defined as those called by both algorithms in a person and overlapped for at least 20% of their length.

CNV filtering

PLINK (1.0.7 version) (101) was employed to remove CNVs that overlapped for at least 50% of their length with segmental duplication regions (downloaded from the UCSC genome browser, http://genome.ucsc.edu), peri-centromeric regions (recommended by PennCNV), pseudoautosomal regions on the X chromosome and CNVs containing >70% GC content. CNVs were only selected for analysis that spanned at least five SNPs and had a length of >30 kb as these filtering criteria in a previous large-scale study achieved a validation rate of >90% (49). Rare CNVs were defined as CNVs with a frequency of <1% in the entire case–control data set. In addition, case-unique CNVs that did not overlap more than 50% of their length with any control CNVs identified in the CHOP CNV database (http://cnv.chop.edu/) were used to compile a list of potential candidate genes for validation.

Bias checking

Since the total number of CNVs per individual could be influenced by gender, genotyping platform or DNA source, several comparisons were made to evaluate these potential biases in cases and controls. In our data set, there is a distinct male to female bias in cases (707/106) when compared with controls (289/303). While females have significantly more CNVs identified on the X chromosome, no difference was found on autosomes (P = 0.903 for controls; P = 0.795 for cases). We also examined the potential difference between two genotyping arrays used because all control samples were genotyped on Illumina’s 1M-duo-v3 while 634 out of 813 cases were genotyped on Illumina’s 1M-v1-C chip. Data from cases indicate that more duplications were identified using the 1M-duo-v3 (P < 0.001). Moreover, potential biases arising from different DNA sources [blood (438/813 cases versus 299/592 controls), saliva (2/813 cases versus 293/592 controls), cell lines (49/813 cases) and DNA-origin source unknown (324 cases)] were assessed. Consistent with the findings from other studies, a potential difference was present in the number of duplications across the samples from various DNA sources [cases genotyped on 1M-v1-C: one-way ANOVA P = 0.002 (blood, cell-line and unknown); controls genotyped on 1M-duo-v3: t-test P = 0.001(blood, saliva)].
CNV validation

We selected a subset of candidate genes and regions for validation by real-time PCR. These included four distinct categories: (i) regions or genes previously reported to carry ASD specific CNVs, (ii) novel deletions over 1 Mb, (iii) CNVs containing genes with a known neuronal function, and (iv) CNVs that overlapped genes involved in other neuropsychiatric and neurodevelopmental disorders. For each gene or region of interest, the individual determined to carry the CNV as well as all available family members were selected for validation using Applied Biosystems’ TaqMan Copy Number Assays. An assay in every CNV region was selected using the GeneAssist software available at http://www5.appliedbiosystems.com/tools/cnv/ (102). Assays were run in quadruplicate, multiplexed 10 μl reactions containing 1X TaqMan Universal PCR Master Mix, VIC-labeled RNaseP reference assay and the FAM-labeled test assay. Reactions were performed in 384-well plates on the ABI 7900HT Real-Time PCR System and data were collected with Applied Biosystems SDSv2.3. Data were analyzed with a manual Ct threshold of 0.2 and an automatic baseline and copy numbers were determined using the CopyCaller software with the RNaseP level as the calibrator for copy number.

CNV analysis

Analysis was carried out in two data sets: (i) the entire CNV set and (ii) the rare CNV set (frequency < 1%). The genome-wide CNV burden was assessed in deletions and duplication separately due to the potential bias identified from genotyping array differences on duplications. In addition, the burden analysis was carried out in autosome CNVs only due to a skewed male to female ratio in cases when compared with controls. Significances of the mutational burden comparisons was determined via permutation (100 000 time, one-sided test) using PLINK (version 1.0.7). Pathway enrichment for CNV sets was performed using WebGestalt (http://bioinfo.vanderbilt.edu/webgestalt/). The CNV lists were uploaded using the Illumina Omni-1 Quad SNP gene list as the background gene list. Enrichment was performed using a hypergeometric test requiring at least two genes in a pathway with multiple test correction proposed by Benjamini and Hochberg (103).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS

We thank all participants and their family members in this study and personnel at the John P. Hussman Institute for Human Genomics (HIHG) including staff at the HIHG Biorepository and HIHG autism ascertainment team as well as the members of the Vanderbilt Center for Human Genetics Research. A subset of the participants was ascertained while Dr Pericak-Vance was a faculty member at Duke University.

FUNDING

This work was supported by the National Institutes of Health (R01MH080647, P01NS026630); and a gift from the Hussman Foundation.

REFERENCES


