

Copy Number Variants in Extended Autism Spectrum Disorder Families Reveal Candidates Potentially Involved in Autism Risk

Daria Salyakina^{1,9}, Holly N. Cukier^{1,9}, Joycelyn M. Lee¹, Stephanie Sacharow^{1,2}, Laura D. Nations¹, Deqiong Ma^{1,2}, James M. Jaworski¹, Ioanna Konidari¹, Patrice L. Whitehead¹, Harry H. Wright³, Ruth K. Abramson³, Scott M. Williams⁴, Ramkumar Menon⁵, Jonathan L. Haines⁴, John R. Gilbert^{1,2}, Michael L. Cuccaro^{1,2}, Margaret A. Pericak-Vance^{1,2*}

1 John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, Florida, United States of America, **2** Dr. John T. Macdonald Foundation Department of Human Genetics, Miller School of Medicine, University of Miami, Miami, Florida, United States of America, **3** Department of Neuropsychiatry, University of South Carolina School of Medicine, Columbia, South Carolina, United States of America, **4** Center for Human Genetics Research, Vanderbilt University, Nashville, Tennessee, United States of America, **5** Department of Epidemiology and Department of Obstetrics and Gynecology, Rollins School of Public Health, Emory University, Atlanta, Georgia, United States of America

Abstract

Copy number variations (CNVs) are a major cause of genetic disruption in the human genome with far more nucleotides being altered by duplications and deletions than by single nucleotide polymorphisms (SNPs). In the multifaceted etiology of autism spectrum disorders (ASDs), CNVs appear to contribute significantly to our understanding of the pathogenesis of this complex disease. A unique resource of 42 extended ASD families was genotyped for over 1 million SNPs to detect CNVs that may contribute to ASD susceptibility. Each family has at least one avuncular or cousin pair with ASD. Families were then evaluated for co-segregation of CNVs in ASD patients. We identified a total of five deletions and seven duplications in eleven families that co-segregated with ASD. Two of the CNVs overlap with regions on 7p21.3 and 15q24.1 that have been previously reported in ASD individuals and two additional CNVs on 3p26.3 and 12q24.32 occur near regions associated with schizophrenia. These findings provide further evidence for the involvement of *ICA1* and *NXPH1* on 7p21.3 in ASD susceptibility and highlight novel ASD candidates, including *CHL1*, *FGFBP3* and *POUF41*. These studies highlight the power of using extended families for gene discovery in traits with a complex etiology.

Citation: Salyakina D, Cukier HN, Lee JM, Sacharow S, Nations LD, et al. (2011) Copy Number Variants in Extended Autism Spectrum Disorder Families Reveal Candidates Potentially Involved in Autism Risk. PLoS ONE 6(10): e26049. doi:10.1371/journal.pone.0026049

Editor: Michael Edward Zwick, Emory University School Of Medicine, United States of America

Received: June 24, 2011; **Accepted:** September 16, 2011; **Published:** October 7, 2011

Copyright: © 2011 Salyakina et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was supported by grants from the National Institutes of Health (R01MH080647), www.nih.gov; the National Institute of Neurological Disorders and Stroke (P01NS026630), www.ninds.nih.gov; and by a gift from the Hussman Foundation, www.hussmanfoundation.org. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: mpericak@med.miami.edu

9 These authors contributed equally to this work.

Introduction

Autism spectrum disorders (ASDs) encompass a range of abnormal neurodevelopmental phenotypes characterized by a core set of symptoms including social, communicative and behavioral impairments [1]. The most recent prevalence studies suggest that ASDs affect as many as 1 in 110 children [2–4]. While ASD etiology is multifaceted, a significant portion is known to stem from genetic causes given the increased concordance rate in monozygotic twins (~91%) as compared to dizygotic twins (~10%) [5–8]. Genetic studies over the past decade have demonstrated that the underlying components are complex with multiple genes acting either independently or perhaps synergistically to result in ASD phenotypes. While hundreds of candidate genes have been published, linkage and association studies have only recognized a small number of consensus regions [9]. Additionally, genome-wide association studies thus far have

identified very few regions that may contain a common genetic cause, such as the regions on 5p14.1 and 5p15 [10–12]. Therefore, it appears that rare and private variations may represent a considerable portion of the genetic components involved in ASDs.

A growing number of studies have demonstrated the potential contribution of copy number variations (CNVs) in the etiology of many complex disorders including autism, Alzheimer's disease, epilepsy, and schizophrenia [13,14]. To date, CNVs believed to be involved in ASDs occur at a relatively low frequency and are often private, familial alterations, reflecting the genetic heterogeneity of this disorder [15]. Furthermore, these alterations have implicated a variety of genes within the glutamatergic, ubiquitin, and cell adhesion pathways [16,17]. However, only a limited number of findings have been replicated between research groups including CNVs in *CNTN4*, *NRXN1*, 15q11-q13, 16p11.2, and 22q11.21 [17–20]. In addition, there appears to be variability in the clinical consequences of CNVs. For instance, patients with CNVs on

16p11.2 have been identified with a range of features including autism, developmental delay, intellectual disabilities and schizophrenia [19,21–23]. Furthermore, there is evidence of phenotypic heterogeneity within families carrying a 16p11.2 CNV including siblings discordant for autism and transmission of a CNV from asymptomatic parents [24,25]. This demonstrates that while a CNV may provide a strong susceptibility to a phenotype, it may act in concert with other genetic and environmental factors.

Several studies suggest that familial and sporadic cases of ASD are genetically distinct. Extended, multiplex families are more likely to carry heritable risk factors while, in contrast, sporadic ASD families have demonstrated a higher rate of *de novo* CNVs [26,27]. In this study, we examine 42 unique extended ASD families and identify rare structural alterations that are shared between multiple affected individuals. We hypothesized that that we would unearth novel ASD susceptibility genes with a moderate to strong effect by identifying CNVs that segregate in distantly affected relatives. Intriguingly, many CNVs carriers, while not being clinically diagnosed with an ASD, nonetheless present with a history of psychiatric and developmental conditions. Our study demonstrates the importance of decoding the various genetic components involved in ASDs.

Materials and Methods

Ethics statement

We ascertained individuals at the John P. Hussman Institute for Human Genomics (HIHG) at the University of Miami Miller School of Medicine (Miami, Florida), the University of South Carolina (Columbia, South Carolina), and the Center for Human Genetics Research at Vanderbilt University (Nashville, Tennessee). Written informed consent was obtained from parents for all minor children and those who were unable to give consent. In addition, we obtained assent from all participants of the appropriate developmental and chronological age. All participants were ascertained using the protocol approved by the appropriate Institutional Review Boards. Patients were collected for this study for over a decade, with protocols and amendments being approved at each stage. Oversight of the study at the University of Miami falls under the UM Medical IRB B committee whose members includes Ofelia Alvarez, M.D., Abdul Mian, Ph.D., Maria Alcaide, M.D., Michelle Dunaj Lucking, J.D., Jean Jose, D.O., Howard Landy, M.D., Bruce Nolan, M.D., FACP, FAASM, Emilio Weiss, Pharm.D., and Cecilia Grano de Oro, B.A.

Ascertainment and sample description

Our dataset consisted of 410 individuals from 42 extended ASD families. Extended families were defined as families with at least two affected cousins. These families included either two ($n = 28$), three ($n = 8$) or four ($n = 6$) cousins or more distant relatives with ASDs. We collected DNA from all affected and unaffected family members wishing to participate.

ASD families were enrolled and recruited via support groups, advertisements, and from clinical and educational settings. Core inclusion criteria for ASD individuals were as follows: (1) chronological age between 3 and 21 years of age, (2) a presumptive clinical diagnosis of ASD, (3) an expert clinical determination of ASD diagnosis using DSM-IV criteria supported by the Autism Diagnostic Interview (ADI-R) [28,29], and (4) an IQ equivalent > 35 or developmental level > 18 months as determined by the Vineland Adaptive Behavior Scale (VABS) [30]. Diagnostic determination was based on review by a panel consisting of experienced clinical psychologists and a pediatric medical geneticist. In those instances where an ADI-R was not available

($n = 18$), a best-estimate diagnosis was assigned using all available clinical information including clinician summaries, a caregiver report, and medical records. 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 identified metabolic, genetic, or progressive neurological disorders. Family history and pedigree information was collected in a standard semi-structured interview from a knowledgeable informant, most frequently the mother. Additional clinical data were also collected by reviewing available medical and psychiatric records of affected individuals.

Ancestry was identified by self-report and subsequently confirmed by genotyping data plotted in EIGENSTRAT and compared with HapMap populations [31,32]. One family, 7606, was of African American descent and 41 families were of European descent. Thirty-one families had only male ASD individuals while the remaining eleven families had both male and female affected individuals. Genomic DNA used for CNV identification was purified from three different sources, whole blood ($n = 276$), saliva ($n = 8$) and cell lines ($n = 126$), using Puregene chemistry on the Qiagen Autopure LS according to standard protocols. Sample concentration was assessed via the NanoDrop 8000 Spectrophotometer (Thermo Scientific) and all samples were normalized to 50 ng/ul prior to genotyping.

Control individuals

Eight hundred and ninety-nine healthy children were recruited by two centers. The first cohort was recruited by the Hussman Institute for Human Genomics (HIHG) at the University of Miami and included 565 children between the ages of 4 and 21 years. Participants with developmental, behavioral, or neurological conditions were excluded, as well as those with first degree relatives with such disorders. We obtained consent from all participants or, in the case of minors, their parents. Participants meeting these criteria provided a saliva sample. A knowledgeable informant, usually the mother, completed the Social Communication Questionnaire to screen for potential ASDs [33]. The remaining 334 control individuals were part of a preterm birth study from Nashville (Nashville Birth Cohort, NBC). These samples were collected from the Centennial Medical Center in Nashville, Tennessee from the cord blood of term pregnancies (> 37 weeks gestation) and were not screened for ASDs. All samples were collected from live, singleton births delivered by mothers between the ages of 18 and 40.

Race was identified in both control cohorts by parental self-report and verified by comparing the genotyping data to HapMap populations in EIGENSTRAT [31,32]. In rare cases of inconsistency, the genetically determined ancestry was used. In total, 112 controls were classified as African American and 787 of European descent.

Genotyping and copy number variation discovery

Genotyping of extended family members was performed with either the Illumina Human 1Mv1_C BeadChip ($n = 349$) or the 1M-DuoV3 BeadChip ($n = 61$). Samples were processed using the Infinium II assay. After genotyping, we carried out a series of quality control (QC) tests. Reported and genetic sex were examined using X-chromosome linked SNPs. Relatedness between samples and potential sample contaminations were evaluated using genome-wide identity by descent and identity by state estimation through genotyped SNPs in PLINK (<http://pngu.mgh.harvard.edu/~purcell/plink/ibdibs.shtml>). We tested for Mendelian inconsistencies for all SNPs in the WASP/PLATO software (<https://chgr.mc.vanderbilt.edu/plato>). Samples with

gender inconsistencies, elevated Mendelian error rates and genotyping call rates below 95% were removed. All controls were genotyped on the Human 1 M-DuoV3 BeadChips and underwent the same QC steps as the family cohort with exception of testing for Mendelian errors.

All subjects that passed QC were entered into the CNV analysis. The CNV calls were generated via the PennCNV algorithm (August 2009 version) using Log R ratio (LRR) and B allele frequency [34]. Individual CNV calling was adjusted for intensity wave artifacts (GC-waves) across the genome. A minimum of three SNPs were used to define each CNV call. For members within each family processed on the same chip, we were able to further refine boundaries using the PennCNV trio/quartet-based calling algorithm. Samples with a standard deviation of normalized intensity (SDLRR) <0.4 and not more than 300 detected CNVs were included in the analysis. Samples with $0.3 < \text{SDLRR} < 0.4$ and a number of detected CNVs between 100 and 300 ($n = 34$) were visually examined for mosaicism and extremely elevated heterozygosity in BeadStudio V3 software. Due to hemizyosity in males and X-chromosome inactivation in females, the CNV analysis was restricted to autosomes. We used the Kruskal-Wallis rank sum test to compare group specific CNV numbers. Associations in contingency tables were determined using Fisher's exact test.

Identification of copy number variations concordant with autism spectrum disorders

To identify CNVs that potentially contribute to ASD risk, we performed a multi-step procedure of CNV filtering in the extended families. First, we sought to recognize identical by descent variants by screening for CNVs that segregate with all affected individuals available for testing within the extended family. Next, we excluded CNVs overrepresented in unaffected family members by filtering out CNVs carried by more than one married-in unaffected individual or unaffected child with no clear segregation with affection status, since those CNVs likely segregate in affected individuals by chance alone. To identify CNVs present in the general population that are unlikely to contribute to ASDs, we evaluated each region of interest to determine whether these areas also contained CNVs in our 899 pediatric control samples as well as the Database of Genomic Variants v9 (DGV, <http://projects.tcag.ca/variation/>). For this assessment, we considered deletions or duplications overlapping by at least 50% of their lengths as the same CNV.

Validation by real-time PCR copy number assays

Pre-designed TaqMan real-time polymerase chain reaction (PCR) copy number assays (Applied Biosystems) were selected within the CNVs chosen for validation using the GeneAssist™ Copy number Assay Workflow Builder (Table S1). Patients carrying CNVs of interest as well as relatives negative for the CNV were evaluated by real-time PCR. Each individual was tested in quadruplicate according to standard protocols. Analysis was performed in the CopyCaller Software v1.0 (Applied Biosystems).

Results

Copy number variants identified

In total, 382 samples (97 ASD patients and 285 relatives without a formal ASD diagnosis) were used in the final CNV analysis of the families. Thirteen samples were excluded due to QC failure and an additional fifteen samples were removed because their family

branches became uninformative. 838 pediatric control samples survived QC and were also evaluated for CNVs.

Altogether, 23,263 autosomal CNVs at 5,653 distinct genomic locations were identified in the 382 individuals comprising the ASD family cohort. This includes 1,447 homozygous deletions (6.2%), 14,212 single copy deletions (61.0%), 7,534 single copy duplications (32.3%) and 170 homozygous duplications (0.7%). CNVs were detected with an average size of 51.3 kilobases (kb).

We observed a difference in the average number of CNVs between samples originating from different DNA sources (whole blood, saliva, or lymphoblast cell line) as well as variation in samples genotyped on each version of the Illumina BeadChips (Table S2). The genotyping chip design contributes to this variation because the number of sites used for CNV calling is substantially greater on the Human 1 M-DuoV3 than on the Human 1Mv1 C Beadchip (1,199,188 vs. 1,072,821). However, since CNVs of interest were verified for transmission in family members using real-time PCR, these observed differences did not result in an increase of false positive findings in our study.

Copy number variants concordant with autism spectrum disorders

To determine which CNVs may play a role in ASD risk, we screened for CNVs segregating in all affected individuals within each extended family. We hypothesized that these families would carry a heterozygous genetic component that, while having a moderate to strong influence on ASD, would exhibit incomplete penetrance. We identified overlapping CNV segments shared by every affected individual in each pedigree and examined the transmission from the common ancestor to affected individuals, excluding all recurrent CNVs that did not segregate with affection status. This strategy reduced the initial thousands of copy number variants to a manageable few given that first cousins share 12.5% of their genetic information by chance alone and second cousins share 3.125%.

Twelve CNVs of interest within eleven families of European ancestry were identified and successfully validated using TaqMan real-time PCR copy number assays (Table 1, Figure S1). Our findings include five deletions on 1p34.1, 2p24.2-p24.1, 6q11.1, 12q24.32, and 13q31.1 (Figure 1A–E), along with seven duplications on 3p26.3, 4p16.3, 7p21.3, 10p12.31, 10q23.32, 15q11.2 and 15q24.1 (Figure 1F–K). Five of these CNVs at 7p21.3, 10p12.31, 10q23.32, 13q31.1 and 15q24.1 were not identified in our control samples or in the publicly available DGV database (Figure 2). The microduplications on 7p21.3 and 15q24.1 both segregate in the same four individuals in family 17122 (Figure 1H). While these two duplications were absent from our control dataset and the DGV, they each overlap with ASD regions reported in the literature [35]. Further investigation into this family permitted us to refine the breakpoints for both duplications and sequence the 7p21.3 breakpoints, thereby determining that this CNV was duplicated in tandem [36].

Six of the twelve CNVs fall across the following known genes: *BTAFL1*, *FGFPB3*, *ICAI*, *KHDRBS2*, *NXPPI1*, *TNIP2*, *UBL7*, and *ZSWIM5* (Table 1, Figure 2, Figure 3). The deletions on 1p34.1 and 6q11.1 are relatively small (2.7 and 12.8 kb) and located within the intronic regions of *ZSWIM5* and *KHDRBS2*, respectively (Figure 3A,B). In contrast, the four duplications (4p16.3, 7p21.3, 10q 23.2, and 15q24.1) are, for the most part, larger (11.0–352.4 kb) and either span whole genes, *FGFPB3* and *TNIP2*, or include exonic portions of *BTAFL1*, *ICAI*, *NXPPI1*, and *UBL7* (Figure 3C–F).

Table 1. Validated Copy Number Variations, Nearest Genes and Control Frequencies.

CNV	Cytoband	Boundaries (Build NCBI136/hg18)	Estimated length (bp)	Family	Gene Nearest to CNV	Frequency in Caucasian controls (n = 727) [HIHG/NBC] [*]	Frequency in African American controls (n = 111) [HIHG/NBC] [*]	Count in DGV (n > 9200)
del	1p34.1	chr1:45408389–45411073	2,684	17545	<i>ZSWIM5</i>	0.012 [8/1]	0.027 [3/0]	4
	2p24.2–p24.1	chr2:19065745–19100096	34,351	7745	<i>OSR1</i>	0	0	5
	6q11.1	chr6:62507037–62519883	12,846	17342	<i>KHDRBS2</i>	0	0	2
	12q24.32	chr12:125875006–125881162	6,156	17678	<i>LOC100128554</i>	0.007 [1/4]	0	18
	13q31.1	chr13:77912491–77917728	5,237	17142	<i>POU4F1</i>	0	0	0
dup	3p26.3	chr3:57010–103697	46,687	37994	<i>CHL1</i>	0.003 [2/0]	0	5
	4p16.3	chr4:2704290–2747426	43,136	7663	<i>TNIP2</i> [†]	0	0	0
	7p21.3	chr7:8144894–8497305	352,411	17122	<i>ICA1, NXPH1</i>	0	0	0
	10p12.31	chr10:21727129–21818381	91,252	37232	<i>C10orf114</i>	0	0	0
	10q23.32	chr10:93621936–93730409	108,473	37425	<i>BTA1, FGFBP3</i> [‡]	0	0	0
	15q11.2	chr15:18411624–18472812	61,188	17668	<i>GOLGA6L6</i>	0.004 [3/0]	0.009 [0/1]	>300
	15q24.1	chr15:72538187–72549215	11,028	17122	<i>UBL7</i>	0	0	0

[†]CNV boundaries correspond to the minimal size detected by Illumina's 1M genotyping array.

^{**}CNV boundaries sequenced in Cukier, et al, 2011.

^{***}CNV boundaries correspond to the minimal size refined in Cukier, et al, 2011.

^{*}CNV type reported in controls corresponds to the CNV type (del or dup) identified in ASD extended families.

Genes in bold fall within the CNV.

[†]Whole gene duplication.

HIHG - University of Miami Hussman Institute for Human Genomics control cohort.

NBC - Nashville Birth Cohort.

DGV - Database of Genomic Variants.

doi:10.1371/journal.pone.0026049.t001

Clinical information and CNVs transmission in extended families

To investigate the potential inheritance models of putatively pathogenic CNVs, we evaluated all available family history data (Table S3). Each of the segregating CNVs were found in families with only males affected, although a sister in family 17545 carries the 1p34.1 deletion and was described as having developmental delay and ASD symptoms (Figure 1A, individual 103, Table S3). Five of the CNVs were transmitted maternally while the remaining six were inherited both maternally and paternally. In every family, the CNV passes through individuals without a formal ASD diagnosis, supporting a dominant model of inheritance with incomplete penetrance. Furthermore, the fact that some CNVs are only transmitted maternally suggests possible parent of origin effects.

For all family members negative for a diagnosis of ASD and carrying a CNV of interest, half (19 of 38, 50.0%) have a history of psychiatric or developmental conditions, including ASD symptoms without a formal diagnosis, alcoholism, anxiety disorder, attention deficit and hyperactivity disorder (ADHD), bipolar disorder, major or postpartum depression, obsessive compulsive disorder, oppositional defiant disorder (ODD), and speech disorders. The complete clinical information for these individuals is shown in Table S3. To determine whether CNV carriers had an increased frequency of psychiatric or developmental disorders, we compared the clinical features of twenty CNV carriers to their spouses who did not carry the CNV (individuals specified on Table S3). Twelve of the CNV carriers were found to have one of the clinical diagnoses listed above, while only six of their spouses were diagnosed with one of the above features. When measured by Fisher's exact test, this difference trended towards but did not reach significance ($p = 0.111$). While we believe that the CNVs of

interest may confer risk for ASD, our data also suggest that these CNVs may play a more general role in a range of other neuropsychiatric conditions.

Discussion

Using a unique resource of ASD extended families, twelve loci were found that co-segregate with disease and may be involved in ASD susceptibility. In evaluating the potential pathogenicity of each region, we first prioritized rare structural variants not identified in controls, then proven ASD regions and genomic areas linked to neuropsychiatric disease, and, lastly, on CNVs containing genes with a known, neuronal function (Figure 2). Of the twelve structural variants identified, seven were observed at a low frequency in our control and the DGV datasets and two additional CNVs overlap with those identified in the literature, leaving three novel CNVs unique to our ASD families (Figure 2). Two CNVs fall within regions previously linked to ASD, while five more are appealing candidates due to their locations mapping across or near regions either correlated with neuropsychiatric disease or containing genes related to neuronal processes. Three of the duplications (3p26.3, 7p21.3 and 15q24) also provide further support for the hypothesis that the ubiquitin and cell adhesion pathways are involved in ASD risk [17].

Previous evidence from ASD families intersects with two of the structural variants that we identified: 7p21.3 and 15q24.1. On 7p21.3, duplications of approximately the same size and location have been reported in two additional ASD families [35]. The duplications in all three ASD families fall across *ICA1* (*islet cell autoantigen 1*) and *NXPH1* (*newexophilin 1*), both appealing candidate genes. Interestingly, *ICA1* localizes to dendrites and plays a role in glutamatergic signaling [37]. Furthermore, the *NXPH1* interacts with neuroligin 1, a protein that promotes adhesion between

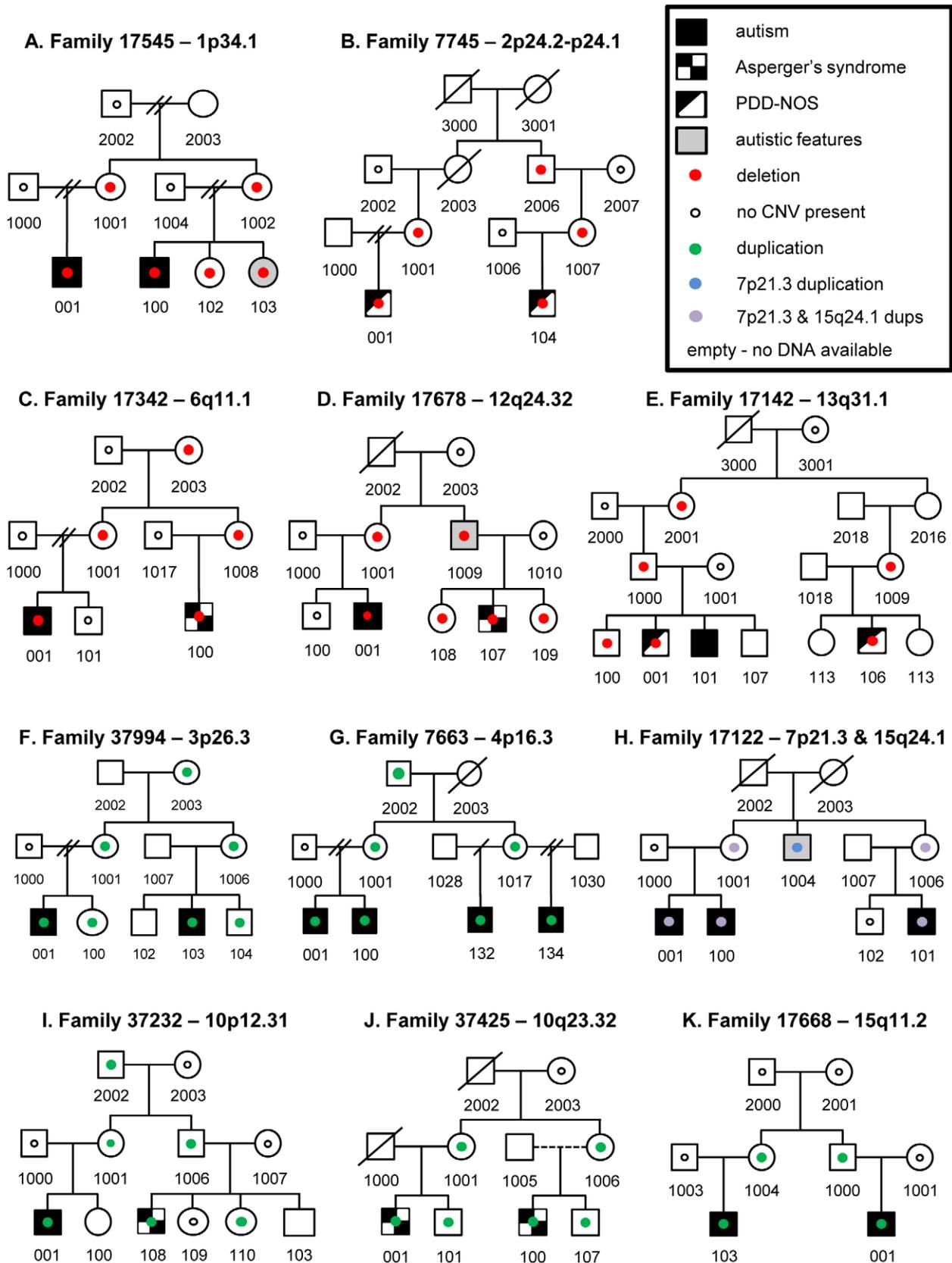


Figure 1. Copy number variations identified in 11 families. A–E) Five extended ASD families had deletions segregating with disease. F–K) Six ASD families carried duplications that segregated with ASD. Family 17122 had two duplications that were inherited in an identical manner (H). Only individuals marked with a small circle had DNA available for testing. doi:10.1371/journal.pone.0026049.g001

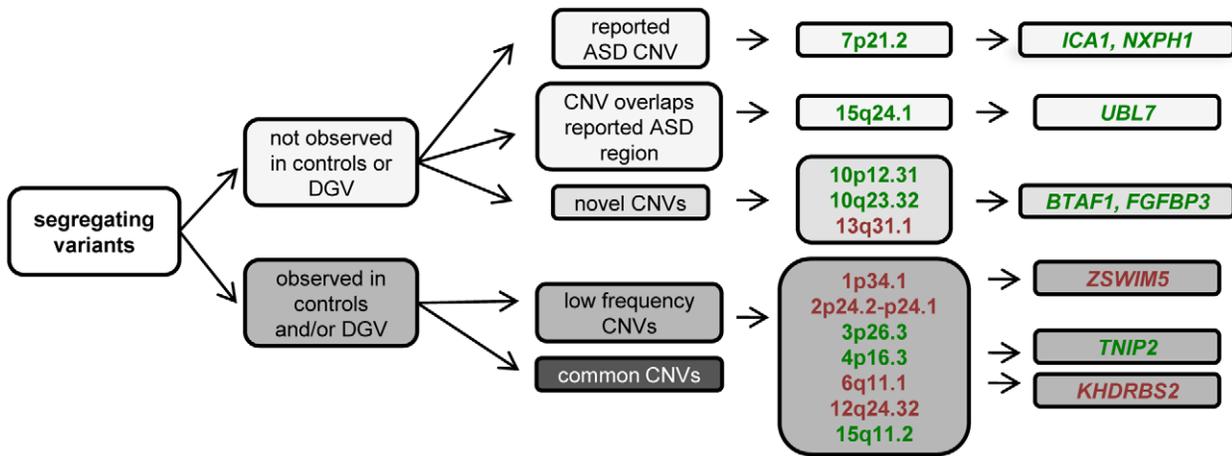


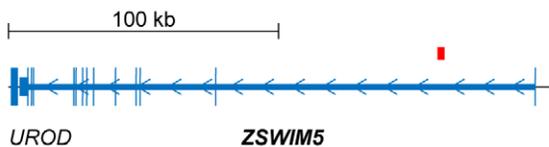
Figure 2. Prioritization of copy number variations of interest. Structural variants were first assessed to determine whether they were present in our control individuals or the Database of Genomic Variants. Literature searches were then performed to determine if the CNVs we identified were previously reported as ASD regions with similar breakpoints or overlay with larger ASD reported regions. Higher priority was given to variation absent in controls and those that fell within regions previously connected to ASDs. Low frequency CNVs were identified in our control HIHG and NBC dataset, but at a frequency below 1.5%. Duplications are in green and deletions are in red.
doi:10.1371/journal.pone.0026049.g002

dendrites and axons and has been identified to carry variants in autism patients [38,39]. The duplication on chromosome 15q24.1 falls within the 1.33 Mb critical region of a previously reported microdeletion and microduplication syndrome found in patients with autistic features, delayed development and mental retardation [36,40,41]. At approximately 10 kb, this CNV is much smaller than those previously reported in the region and only falls across a

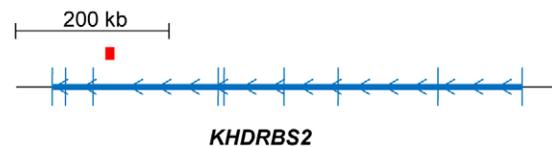
single gene, *UBL7* (*ubiquitin-like 7*) [36]. While *UBL7* is expressed in the brain, little is currently known about its specific function [42].

Three additional loci we identified occur near areas connected to other neuropsychiatric diseases: 3p26.3, 4p16.3 and 12q24.32. Regions on 3p26.3 and 12q24.32 have each been implicated in schizophrenia, with 3p26.3 also being connected to intellectual disability [43–45]. While the 3p26.3 microduplication does not

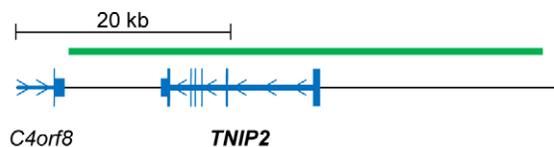
A. Family 17545 – 1p34.1



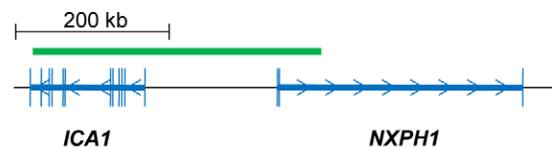
B. Family 17342 – 6q11.1



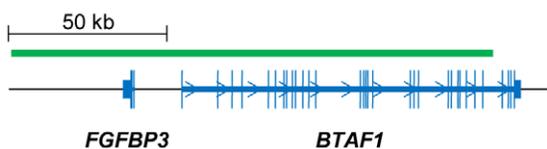
C. Family 7663 – 4p16.3



D. Family 17122 – 7p21.3



E. Family 37425 – 10q23.32



F. Family 17122 – 15q24.1

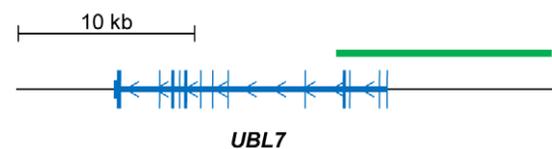


Figure 3. Copy number variations that intersect genes. A,B) Deletions on 1p34.1 and 6q11.1 fall within the intronic regions of *ZSWIM5* and *KHDRBS2*, respectively. C–F) Four duplications on 4p16.3, 7p21.3, 10q 23.2, and 15q24.1 completely encompass the *FGFBP3* and *TNIP2* genes and traverse portions of *BTAF1*, *ICA1*, *NXPH1*, and *UBL7*. Deletions are denoted with red bars, duplications marked by green bars and genes are annotated with blue lines. Genes that intersect CNVs are in bold.
doi:10.1371/journal.pone.0026049.g003

intersect any genes, it falls 109 kb upstream of *CHLI* (*cell adhesion molecule with homology to L1CAM*), an intriguing ASD candidate gene. *CHLI* has increased expression in presynaptic terminals and acts to regulate SNARE complexes [46]. Lastly, the 4p16.3 region has been connected to developmental delay, mental retardation, Huntington's disease and Wolf-Hirschhorn syndrome [47,48]. The variety of features reported in patients carrying CNVs within each region, combined with the pattern of transmission within our families, may demonstrate a variability in penetrance and clinical presentations, possibly in a gender specific manner. For example, there are four male ASD individuals in family 7663 who carry a 4p16.3 duplication, but all six family members with the duplication (individuals 001, 100, 132, 134, 1001, and 1017) were reported to have either a speech or language problem as a child (Table 1, Figure 1G).

The novel 10q23.32 microduplication encompasses a strong ASD candidate, the *fibroblast growth factor binding protein 3* (*FGFBP3*) gene. *FGFBP3* is expressed in fetal brain and binds to the FGFs, which are required throughout brain development, neuronal differentiation and survival [49,50]. Notably, a transgenic mouse model of a related FGF gene, *FGFR1*, mimics physiological features of schizophrenia such as increased levels of dopamine and sensory impairment [51,52].

One last intriguing CNV is the novel deletion on 13q31.1. This deletion does not contain a gene, but is located 153.5 kb upstream of the *POU domain, class 4, transcription factor 1* (*POU4F1*) gene, also known as *BRN3A*, an appealing ASD candidate. *POU4F1* plays a role in neuronal development and differentiation through careful timed repression of non-neuronal genes [53,54]. The 13q31.1 deletion, while not disrupting *POU4F1* directly, may affect the regulation of developmental timing or tissue specific regulation of this gene.

Our approach of using extended ASD families successfully identified rare variants that may be associated with ASD on 7p21.3 and 15q24.1, as well as structural aberrations near regions related to other neuropsychiatric disorders on 3p26.3, 4p16.3 and 12q24.32. These results demonstrate that rare, heritable CNVs play a significant role in ASD etiology. Our findings support the hypothesis that altering the dosage of a gene or its regulatory elements can lead to a severe clinical consequence and brings us one step closer to deciphering the genetic heterogeneity of autism spectrum disorders.

Supporting Information

Figure S1 Validation of copy number variations by real-time PCR. A-K) Results for one assay at each genomic location

References

- American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders (DSM-IV). Washington, D.C.: American Psychiatric Press, Inc.
- Fombonne E (2009) Epidemiology of pervasive developmental disorders. *Pediatr Res* 65(6): 591–598.
- Autism and Developmental Disabilities Monitoring Network Surveillance Year 2006 Principal Investigators, Centers for Disease Control and Prevention (CDC). (2009) Prevalence of autism spectrum disorders - autism and developmental disabilities monitoring network, united states, 2006. *MMWR Surveill Summ* 58(10): 1–20.
- Brugha TS, McManus S, Bankart J, Scott F, Purdon S, et al. (2011) Epidemiology of autism spectrum disorders in adults in the community in england. *Arch Gen Psychiatry* 68(5): 459–465.
- Folstein S, Rutter M (1977) Infantile autism: A genetic study of 21 twin pairs. *J Child Psychol Psychiatry* 18(4): 297–321.
- Ritvo ER, Freeman BJ, Mason-Brothers A, Mo A, Ritvo AM (1985) Concordance for the syndrome of autism in 40 pairs of afflicted twins. *Am J Psychiatry* 142: 74–77.
- Steffenburg S, Gillberg C, Hellgren L, Andersson L, Gillberg IC, et al. (1989) A twin study of autism in denmark, finland, iceland, norway, and sweden. *J Child Psychol Psychiatry* 30: 405–416.
- Piven J, Chase GA, Landa R, Wzorek M, Gayle J, et al. (1991) Psychiatric disorders in the parents of autistic individuals. *Journal of the American Academy of Child & Adolescent Psychiatry* 30: 471–418.
- Kumar RA, Christian SL (2009) Genetics of autism spectrum disorders. *Curr Neurol Neurosci Rep* 9(3): 188–197.
- Ma D, Salyakina D, Jaworski JM, Konidari I, Whitehead PL, et al. (2009) A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. *Ann Hum Genet* 73(Pt 3): 263–273.
- Wang K, Zhang H, Ma D, Bucan M, Glessner JT, et al. (2009) Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* 459(7246): 528–533.
- Weiss LA, Arking DE, Gene Discovery Project of Johns Hopkins & the Autism Consortium, Daly MJ, Chakravarti A (2009) A genome-wide linkage and association scan reveals novel loci for autism. *Nature* 461(7265): 802–808.
- Stankiewicz P, Lupski JR (2010) Structural variation in the human genome and its role in disease. *Annu Rev Med* 61: 437–455.
- Brouwers N, Van Cauwenbergh C, Engelborghs S, Lambert JC, Bettens K, et al. (2011) Alzheimer risk associated with a copy number variation in the complement receptor 1 increasing C3b/C4b binding sites. *Mol Psychiatry*.

that was validated. The X axis marks the individual within each family and the Y-axis shows the number of copies. The normal copy number is always two. For those families carrying deletions (A–E), individuals with a copy number of one were confirmed to carry the deletion. For families with duplications (F–G), a copy number of three identifies those individuals with the increase in copy number. Family 17122 (H) shows one assay for each of the duplications it carries on 7p21.3 and 15q24.

(PDF)

Table S1 TaqMan Real-Time PCR Copy Number Assays. At least one copy number variation assay was selected to validate each genomic region.

(XLSX)

Table S2 Number Variations identified in Autism Spectrum Disorder Family Cohort and Control Individuals. Summary of the number of variations identified, with case and control samples divided by DNA source (blood, cell lines, or saliva) and the type of Illumina SNP array chip used (either version 1Mv1_C or 1M-DuoV3).

(XLSX)

Table S3 Clinical Information for Families with Validated Copy Number Variations. This table provides phenotype information for all family members evaluated, with emphasis on ASD symptoms and developmental and neuropsychiatric diagnoses. Deletions and then duplications are listed, organized by chromosomal location.

(XLSX)

Acknowledgments

We thank the patients with autism and their family members who participated in this study and personnel at the John P. Hussman Institute for Human Genomics (HIHG) at the University of Miami, including the HIHG Center for Genome Technology, the HIHG autism ascertainment team, with special thanks to Daphne Champagne, members of the HIHG Biorepository, especially Sandra West, as well as collaborators at Vanderbilt University's Center for Human Genetics Research.

Author Contributions

Conceived and designed the experiments: DS HNC JLH JRG MLC MAPV. Performed the experiments: DS HNC IK PLW. Analyzed the data: DS HNC DM JMJ. Wrote the paper: DS HNC MAPV. Collected samples from ASD families: JML SS LDN HHW RKA JLH JRG MLC MAPV. Collected control samples: SMW RM JLH MLC MAPV.

15. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, et al. (2010) Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 15 466(7304): 368–372.
16. Cusco I, Medrano A, Gener B, Vilardell M, Gallastegui F, et al. (2009) Autism-specific copy number variants further implicate the phosphatidylinositol signaling pathway and the glutamatergic synapse in the etiology of the disorder. *Hum Mol Genet* 18(10): 1795–1804.
17. Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, et al. (2009) Autism genome-wide copy number variation reveals ubiquitous and neuronal genes. *Nature* 459(7246): 569–73.
18. Jacquemont ML, Sanlaville D, Redon R, Raoul O, Cormier-Daire V, et al. (2006) Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders. *J Med Genet* 43(11): 843–849.
19. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, et al. (2008) Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* 82(2): 477–488.
20. Bucan M, Abrahams BS, Wang K, Glessner JT, Herman EI, et al. (2009) Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes. *PLoS Genet* 5(6): e1000536.
21. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, et al. (2008) Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* 358(7): 667–675.
22. McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, et al. (2009) Microduplications of 16p11.2 are associated with schizophrenia. *Nat Genet* 41(11): 1223–7.
23. Fernandez BA, Roberts W, Chung B, Weksberg R, Meyn S, et al. (2010) Phenotypic spectrum associated with de novo and inherited deletions and duplications at 16p11.2 in individuals ascertained for diagnosis of autism spectrum disorder. *J Med Genet* 47(3): 195–203.
24. Bijlsma EK, Gijbbers ACJ, Schuurs-Hoeijmakers JHM, van Haeringen A, Franssen van de Putte DE, et al. (2009) Extending the phenotype of recurrent rearrangements of 16p11.2: Deletions in mentally retarded patients without autism and in normal individuals. *Eur J Med Genet* 52(2-3): 77–87.
25. Shen Y, Chen X, Wang L, Guo J, Shen J, et al. (2010) Intra-Family Phenotypic Heterogeneity of 16p11.2 Deletion Carriers in a Three-Generation Chinese Family. *Am J Med Genet B Neuropsychiatr Genet* 156(2): 225–232.
26. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, et al. (2007) Strong association of de novo copy number mutations with autism. *Science* 316(5823): 445–449.
27. Morrow EM, Yoo SY, Flavell SW, Kim TK, Lin Y, et al. (2008) Identifying autism loci and genes by tracing recent shared ancestry. *Science* 321(5886): 218–223.
28. Lord C, Rutter M, DiLavore P, Risi S Autism diagnostic observation schedule-WPS (WPS edition).
29. Rutter M, LeCouteur A, Lord C Autism diagnostic interview, revised (ADI-R).
30. Sparrow SS, Balla D, Cicchetti D (1984) Vineland adaptive behavior scales, interview edition. Circle Pines, MN: AGS Publishing.
31. Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2(1): e190.
32. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38(8): 904–909.
33. Rutter M, Bailey A, Berument SK, LeCouteur A, Lord C, et al. (2003) Social communication questionnaire (SCQ). Los Angeles, CA: Western Psychological Services.
34. Wang K, Li M, Hadley D, Liu R, Glessner J, et al. (2007) PennCNV: An integrated hidden markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 17(11): 1665–1674.
35. Gai X, Xie HM, Perin JC, Takahashi N, Murphy K, et al. (2011) Rare structural variation of synapse and neurotransmission genes in autism. *Mol Psychiatry*.
36. Cukier HN, Salyakina D, Blankstein SF, Robinson JL, Sacharow S, et al. (2011) Microduplications in an autism multiplex family narrow the region of susceptibility for developmental disorders on 15q24 and implicate 7p21. *Am J Med Genet B Neuropsychiatr Genet* 156(4): 493–501.
37. Cao M, Xu J, Shen C, Kam C, Haganir RL, et al. (2007) PICK1-ICA69 heteromeric BAR domain complex regulates synaptic targeting and surface expression of AMPA receptors. *J Neurosci* 27(47): 12945–12956.
38. Missler M, Sudhof TC (1998) Neurexophilins form a conserved family of neuropeptide-like glycoproteins. *J Neurosci* 18(10): 3630–3638.
39. Yan J, Noltner K, Feng J, Li W, Schroer R, et al. (2008) Neurexin Ialpha structural variants associated with autism. *Neurosci Lett* 438(3): 368–370.
40. Sharp AJ, Selzer RR, Veltman JA, Gimelli S, Gimelli G, et al. (2007) Characterization of a recurrent 15q24 microdeletion syndrome. *Hum Mol Genet* 16(5): 567–572.
41. El-Hattab AW, Smolarek TA, Walker ME, Schorry EK, Immken LL, et al. (2009) Redefined genomic architecture in 15q24 directed by patient deletion/duplication breakpoint mapping. *Hum Genet* 126(4): 589–602.
42. Liu S, Yu Y, An H, Li N, Lin N, et al. (2003) Cloning and identification of a novel ubiquitin-like protein, BMSC-UbP, from human bone marrow stromal cells. *Immunol Lett* 86(2): 169–175.
43. Frints SG, Marynen P, Hartmann D, Frys JP, Steyaert J, et al. (2003) CALL interrupted in a patient with non-specific mental retardation: Gene dosage-dependent alteration of murine brain development and behavior. *Hum Mol Genet* 12(13): 1463–1474.
44. Tam GW, van de Lagemaat LN, Redon R, Strathdee KE, Croning MD, et al. (2010) Confirmed rare copy number variants implicate novel genes in schizophrenia. *Biochem Soc Trans* 38(2): 445–451.
45. Lien YJ, Liu CM, Faraone SV, Tsuang MT, Hwu HG, et al. (2010) A genome-wide quantitative trait loci scan of neurocognitive performances in families with schizophrenia. *Genes Brain Behav* 9(7): 695–702.
46. Andreyeva A, Leshchynska I, Knepper M, Betzel C, Redecke L, et al. (2010) CHL1 is a selective organizer of the presynaptic machinery chaperoning the SNARE complex. *PLoS One* 5(8): e12018.
47. Rauch A, Schellmoser S, Kraus C, D'rr HG, Trautmann U, et al. (2001) First known microdeletion within the wolf-hirschhorn-syndrome critical region refines genotype-phenotype correlation. *Am J Med Genet* 99: 338–342.
48. Wu Y, Ji T, Wang J, Xiao J, Wang H, et al. (2010) Submicroscopic subtelomeric aberrations in chinese patients with unexplained developmental delay/mental retardation. *BMC Med Genet* 11: 72.
49. Dono R, Texido G, Dussel R, Ehmke H, Zeller R (1998) Impaired cerebral cortex development and blood pressure regulation in FGF-2-deficient mice. *EMBO J* 17(15): 4213–4225.
50. Sanchez-Calderon H, Milo M, Leon Y, Varela-Nieto I (2007) A network of growth and transcription factors controls neuronal differentiation and survival in the developing ear. *Int J Dev Biol* 51(6-7): 557–570.
51. Klejbor I, Myers JM, Hausknecht K, Corso TD, Gambino AS, et al. (2006) Fibroblast growth factor receptor signaling affects development and function of dopamine neurons - inhibition results in a schizophrenia-like syndrome in transgenic mice. *J Neurochem* 97(5): 1243–1258.
52. Klejbor I, Kucinski A, Wersinger SR, Corso T, Spodnik JH, et al. (2009) Serotonergic hyperinnervation and effective serotonin blockade in an FGF receptor developmental model of psychosis. *Schizophr Res* 113(2-3): 308–321.
53. Xiang M, Zhou L, Macke JP, Yoshioka T, Hendry SH, et al. (1995) The brn-3 family of POU-domain factors: Primary structure, binding specificity, and expression in subsets of retinal ganglion cells and somatosensory neurons. *J Neurosci* 15(7 Pt 1): 4762–4785.
54. Lanier J, Dykes IM, Nissen S, Eng SR, Turner EE (2009) Brn3a regulates the transition from neurogenesis to terminal differentiation and represses non-neural gene expression in the trigeminal ganglion. *Dev Dyn* 238(12): 3065–3079.