Please cite this article in press as: Talkowski et al., Sequencing Chromosomal Abnormalities Reveals Neurodevelopmental Loci that Confer Risk across Diagnostic Boundaries, Cell (2012), doi:10.1016/j.cell.2012.03.028

Cell

Sequencing Chromosomal Abnormalities Reveals Neurodevelopmental Loci that Confer Risk across Diagnostic Boundaries

Michael E. Talkowski,[1,5,7] Jill A. Rosenfeld,[8] Ian Blumenthal,[1] Vamsee Pillalamarri,[1] Colby Chiang,[1] Adrian Heilbut,[1] Carl Ernst,[1] Carrie Hanscom,[1] Elizabeth Rossin,[1,2,7] Amelia M. Lindgren,[9] Shahrin Pereira,[9] Douglas Ruderfer,[1,7] Andrew Kirby,[1,2,7] Stephan Ripke,[1,2,7] David J. Harris,[10] Ji-Hyun Lee,[1] Kyungsoo Ha,[12] Hyung-Goo Kim,[13] Benjamin D. Solomon,[14] Andrea L. Gropman,[15,16] Diane Lucente,[1] Katherine Sims,[1] Toshiro K. Ohsumi,[3] Mark L. Borowsky,[3] Stephanie Loranger,[17] Bradley Quade,[4] Kasper Lage,[2,7,18,19,20] Judith Miles,[21] Bai-Lin Wu,[4,11,22] Yiping Shen,[1,4,11,23] Benjamin Neale,[1,2,7] Lisa G. Shaffer,[8] Mark J. Daly,[1,2,7,17] Cynthia C. Morton,[7,4,9] and James F. Gusella[1,6,7,17,*]

[1]Center for Human Genetic Research
[2]Analytical and Translational Genetics Unit
[3]Department of Molecular Biology
[4]Department of Pathology
Massachusetts General Hospital, Boston, MA 02114, USA
[5]Department of Neurology
[6]Department of Genetics
Harvard Medical School, Boston, MA 02115, USA
[7]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02143, USA
[8]Signature Genomic Laboratories, PerkinElmer, Inc., Spokane, WA 99207, USA
[9]Departments of Obstetrics, Gynecology and Reproductive Biology, Brigham and Women's Hospital, Boston, MA 02115, USA
[10]Division of Clinical Genetics
[11]Department of Laboratory Medicine
Children's Hospital Boston, Boston, MA 02115, USA
[12]Cancer Research Center
[13]Department of Obstetrics and Gynecology, Institute of Molecular Medicine and Genetics
Georgia Health Sciences University, Augusta, GA 30912, USA
[14]Medical Genetics Branch, National Human Genome Research Institute, Bethesda, MD 20892, USA
[15]Department of Neurology, Children's National Medical Center, Washington, DC 20010, USA
[16]Department of Neurology, George Washington University of Health Sciences, Washington, DC 20052, USA
[17]Autism Consortium of Boston, Boston, MA 02115, USA
[18]Pediatric Surgical Research Laboratories, MassGeneral Hospital for Children, Massachusetts General Hospital, Boston, MA 02114, USA
[19]Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[20]Center for Protein Research, University of Copenhagen, 1165 Copenhagen, Denmark
[21]Departments of Pediatrics, Medical Genetics and Pathology, The Thompson Center for Autism and Neurodevelopmental Disorders, University of Missouri Hospitals and Clinics, Columbia, MO 65201, USA
[22]Children's Hospital and Institutes of Biomedical Science, Fudan University, Shanghai 200032, China
[23]Shanghai Children's Medical Center, Shanghai Jiaotong University School of Medicine, Shanghai 200025, China
*Correspondence: gusella@helix.mgh.harvard.edu
DOI 10.1016/j.cell.2012.03.028

SUMMARY

Balanced chromosomal abnormalities (BCAs) represent a relatively untapped reservoir of single-gene disruptions in neurodevelopmental disorders (NDDs). We sequenced BCAs in patients with autism or related NDDs, revealing disruption of 33 loci in four general categories: (1) genes previously associated with abnormal neurodevelopment (e.g., *AUTS2*, *FOXP1*, and *CDKL5*), (2) single-gene contributors to microdeletion syndromes (*MBD5*, *SATB2*, *EHMT1*, and *SNURF-SNRPN*), (3) novel risk loci (e.g., *CHD8*, *KIRREL3*, and *ZNF507*), and (4) genes associated with later-onset psychiatric disorders (e.g., *TCF4*, *ZNF804A*, *PDE10A*, *GRIN2B*, and *ANK3*). We also discovered among neurodevelopmental cases a profoundly increased burden of copy-number variants from these 33 loci and a significant enrichment of polygenic risk alleles from genome-wide association studies of autism and schizophrenia. Our findings suggest a polygenic risk model of autism and reveal that some neurodevelopmental genes are sensitive to perturbation by multiple mutational mechanisms, leading to variable phenotypic outcomes that manifest at different life stages.

INTRODUCTION

Mounting evidence indicates that genomic structural variants (SVs) collectively play a substantial role in susceptibility to autism spectrum disorders (ASDs) and other neurodevelopmental disorders (NDDs). However, the near ubiquitous use of chromosomal microarrays in research and clinical diagnostics has limited the assessment of those classes of SV that do not involve large gains and losses of genetic material, such as inversions, excision/insertions, and translocations, which together constitute balanced chromosomal abnormalities (BCAs). These events are typically defined clinically at karyotypic resolution and implicate only broad chromosomal regions, without the gene-level sequence specificity that would permit informative interpretation. Even in the research setting, there is often a failure to consider or assess BCAs, consequently bypassing a meaningful proportion of subjects with genetic events that may mark a single locus of potentially large effect. These balanced events offer a unique route, complementary to conventional approaches, for identifying individual genes or functional sequences that contribute to otherwise genetically complex human disorders. At cytogenetic resolution, the estimated frequency of BCAs in ASD is 1.3% (Marshall et al., 2008), an approximately 6-fold increase over that observed in more than 10,000 reproductively normal controls (Ravel et al., 2006). This ratio is almost certainly a lower bound for relative risk, given the resolution of available techniques and the inability to survey submicroscopic balanced alterations. Thus, BCAs have a meaningful impact in ASD and represent a fertile area for high-resolution study to identify functional sequences that contribute to human neurodevelopment.

We previously described innovations in the molecular approach to massively parallel sequencing and tailored bioinformatics applicable to the rapid, high-resolution discovery of chromosomal rearrangement breakpoints (Talkowski et al., 2011a). These and conceptually similar methods were recently used to derive potential mechanisms of chromosomal rearrangements in B-lymphocytes (Chiarle et al., 2011; Klein et al., 2011), to delineate complex chromosomal rearrangements and chromothripsis in cancer cells (Stephens et al., 2011), and to document balanced chromothripsis and predominant nonhomologous repair in the human germline and transgenic animals (Chiang et al., 2012). Here, with nucleotide resolution, we precisely characterize karyotypically defined human constitutional chromosomal rearrangements: 36 de novo BCAs and two inherited rearrangements that were transmitted from an affected parent. The results of our sequencing analyses, coupled with extensive secondary genomics support, indicate that disruption of genes from a wide range of biological pathways can contribute to ASD. In many instances, the same genes also confer risk, sometimes via different mutational mechanisms, to a range of NDDs and psychiatric disorders in both children and adults.

RESULTS

Identification of Genes Contributing to Neurodevelopmental Abnormalities

Using a series of previously developed next-generation sequencing techniques ranging from high-depth whole-genome sequencing to a targeted capture of breakpoints approach (Talkowski et al., 2011a), we delineated BCAs in 38 subjects with neurodevelopmental abnormalities; these subjects include two monozygotic twin pairs (36 independent probands). All harbored a BCA that appeared balanced at karyotypic resolution and was interpreted as pathogenic by the clinical geneticist; 36 aberrations arose de novo, whereas two alterations were transmitted from an affected parent and thus segregated with the phenotype. Extensive clinical data were collected for all subjects and affected parents, as described in the Supplemental Information (available online). If a structured diagnostic interview was performed, or a patient was formally diagnosed with autism or an ASD by the referring clinician according to DSM-IV criteria, they are referred to herein as ASD (50% of subjects); otherwise the disorder was classified as NDD, although many such subjects also displayed clinical features consistent with ASD. For complete clinical information for each subject and each BCA breakpoint, confirmed by PCR and capillary sequencing, see Data S1: Phenotypic and Sequencing Information on Individual Patients. Previously performed genetic testing was also obtained, and all results were unremarkable with 244,000 or 1 million feature array comparative genomic hybridization (aCGH) unless otherwise described in the Supplemental Information.

This BCA sequencing approach uncovered genes that conformed to four general classifications, and there was meaningful overlap between categories: (1) genes previously implicated in ASD or NDD and confirmed here by their heterozygous inactivation, (2) genes discovered to be single-locus contributors to microdeletion syndromes, (3) genes not previously implicated individually in ASD or NDD, and (4) genes previously associated with adolescent- and adult-onset psychiatric disorders by common variant genome-wide association studies (GWASs) or other approaches; all but three of these represent novel ASD or NDD loci. Alterations in gene expression were also assessed for all subjects where a lymphoblastoid cell line could be obtained (33 of 38 subjects; Figure S1). If a gene was not directly disrupted, positional effects on expression were evaluated for genes in proximity to the breakpoint. All genes and sequences disrupted by BCAs are presented in Table S1 along with results of mRNA expression studies, whereas the subset of genes, supported by secondary analyses, is presented below and in Table 1.

Convergent Data from Molecular Diagnostics

Although the presence of a BCA is conservatively associated with an ~6-fold increased risk of ASD, any individual BCA is rare and generally nonrecurrent, precluding assessment of false discovery by replication; consequently, we analyzed copy-number variants (CNVs) to test whether specific locus hemizygosity contributes to genetic risk of neurodevelopmental abnormalities. We curated and analyzed a large collection of 33,573 cases from molecular diagnostic facilities and identified that subset of cases referred for a variety of neurodevelopmental abnormalities (n = 19,556), as well as that comprising cases without a referral indication of NDD (n = 14,017). We also surveyed CNV data from 13,991 controls screened for absence of a reported developmental or psychiatric phenotype. Although

Table 1. Genes Disrupted by Chromosomal Rearrangements[a]

Cat	ID	Dx	ChrA	ChrB	Disrupted	Fisher's Exact p[b]	Function
1	DGAP201	ASD	7q11.22	7q36.3	*AUTS2*	5.6×10^{-4}	unknown
1 and 4	NDR27031	NDD	3q13.32	18q21.2	*TCF4*	6.2×10^{-4}	transcription factor
1	DGAP093	NDD	Xp22.13	19p13.3	*CDKL5*	7.2×10^{-2}	protein kinase
1	DGAP157	NDD	3p13	10q21.2	*FOXP1*	4.5×10^{-2}	transcription factor
1 and 4	NDR25941	ASD	12p13.1	12q21.31	*GRIN2B*	7.9×10^{-2}	glutamate receptor
1	DGAP189	NDD	11p13	12p12.1	*SOX5*	8.4×10^{-2}	transcription factor in embryonic development
2	DGAP232	ASD	9p11.2	15q11.2	*SNURF-SNRPN*	1.1×10^{-13}	genomic imprinting in angelman – pws region
2 and 4	DGAP155	ASD	9q34.3	11p11.2	*EHMT1*	3.3×10^{-7}	histone methyltransferase
2	DGAP142	ASD	2q23.1	22q13	*MBD5*	3.1×10^{-5}	methylation binding
2	DGAP211	ASD	2q33.1	6q16.3	*SATB2*	1.1×10^{-3}	transcriptional regulation and chromatin remodeling
3	DGAP148	NDD	Xp11.4	11q24.2	*KIRREL3*	1.6×10^{-4}	cell adhesion
3	DGAP154	NDD	Xq22	17p13.3	*SMG6*	5.9×10^{-4}	nonsense-mediated decay
3	NDR26867	ASD	3q25.31	14q11.2	*CHD8*	2.4×10^{-2}	chromatin remodeling
3	DGAP125	NDD	7q32.1	19q13.11	*ZNF507*	8.0×10^{-2}	zinc finger
3	DGAP132[c]	NDD	5q12.2	7q21.3	*PON3*	1.5×10^{-1}	lactonase
3	AC02-0053	ASD	6q16.1	9q21.13	*GNA14*	2.7×10^{-1}	g-protein signaling
3	DGAP131	NDD	1p22.3	5q33	*ZNHIT6*	2.7×10^{-1}	zinc finger protein
3	DGAP193	ASD	2p22.3	2q31.3	*SPAST*	2.7×10^{-1}	membrane trafficking
3 and 4	DGAP143	NDD	6q22.1	6q27	*PDE10A*	5.2×10^{-3}	phosphodiesterase
3 and 4	DGAP171	NDD	17p13.2	18p11.21	*C18orf1*	3.2×10^{-2}	unknown
3 and 4	DGAP180[c]	NDD	2q32	11q14	*ZNF804A*	4.7×10^{-2}	zinc finger protein

The following abbreviations are used: Cat, disruption category; Dx, diagnosis; ASD, autism spectrum disorder; NDD, other neurodevelopmental disorders; ChrA and ChrB = sequenced chromosomal sub-band containing the BCA. For the entire data set used to generate this table, see also Tables S1, S2, and S3.

[a]BCA-disrupted genes individually implicated by case-control CNV burden at uncorrected $p < 0.10$ or by a minimum of 3 CNVs in cases with none in controls are provided. See Table S1 and Supplemental Information for all subjects and phenotypes and Table S2 for CNV counts on all subjects.

[b]Fisher's exact test p value from comparison of CNV burden between NDD cases and controls.

[c]BCA inherited from similarly affected parent.

almost 25% of the 19,556 cases with neurodevelopmental abnormalities were referred for testing with an indication of autism or ASD, we did not have direct access to clinical records to confirm this diagnosis so we conservatively describe these cases by using the broader term NDD. Our findings indicate that genes disrupted by BCAs in ASD individuals frequently show increased CNV burden across the diagnostic cases referred for ASD as well as those referred for other neurodevelopmental abnormalities, suggesting that the genes discovered can contribute to phenotypic outcomes that are not constrained by ASD diagnostic criteria.

An example of the convergent genomic information collected in this study is depicted in Figure 1; sequencing of monozygotic twins with extensive clinical information revealed a translocation disrupting *TCF4,* a gene previously implicated in both Pitt-Hopkins syndrome and schizophrenia, with concomitantly decreased mRNA expression in both subjects and a significant impact of CNV burden across the clinical diagnostic samples (14 case CNVs, 0 control CNVs; $p = 0.0006$). Additional follow-up analyses revealed a significant GWAS signal from common SNPs at *TCF4* in both autism and schizophrenia (see analyses of polygenic risk from GWAS studies below and in Table S2).

Across all loci disrupted by BCAs, a dramatic increase in overall CNV burden was observed in cases compared to controls (CNV burden across 33 loci: $p = 2.07 \times 10^{-47}$, odds ratio [OR] = 5.12, 95% CI = 3.92–6.79), and this result remained robust to subset analyses and 1 million random simulations to assess empirical significance (see Table S2 for CNV results and Extended Experimental Procedures for analysis details). Comparison of the NDD cases to the 14,017 diagnostic cases referred for a primary indication other than NDD and analyzed on identical platforms also showed an increased CNV burden across these genes ($p = 1.8 \times 10^{-5}$), a result that again exceeded the significance of 1 million random simulations despite the previously established enrichment of large CNVs in the latter cases, which we replicate here. Notably, restricting CNV analyses to only those genes disrupted by a BCA in an individual with a confirmed diagnosis of ASD was also highly significant ($p = 2.76 \times 10^{-28}$; Table 1). Individually, increased CNV burden was nominally significant for 14 of the genes in this study, with three nonsignificant trends ($p < 0.10$), whereas five additional genes were disrupted by CNVs in three or more independent cases but never altered in controls. For the latter, the rarity of dosage alteration in cases and absence of alterations in controls limit statistical power but are consistent with a strong deleterious effect. In each category discussed further below, the genes supported by secondary CNVs are briefly mentioned and presented in Table 1, whereas the full list of genes disrupted by BCAs is

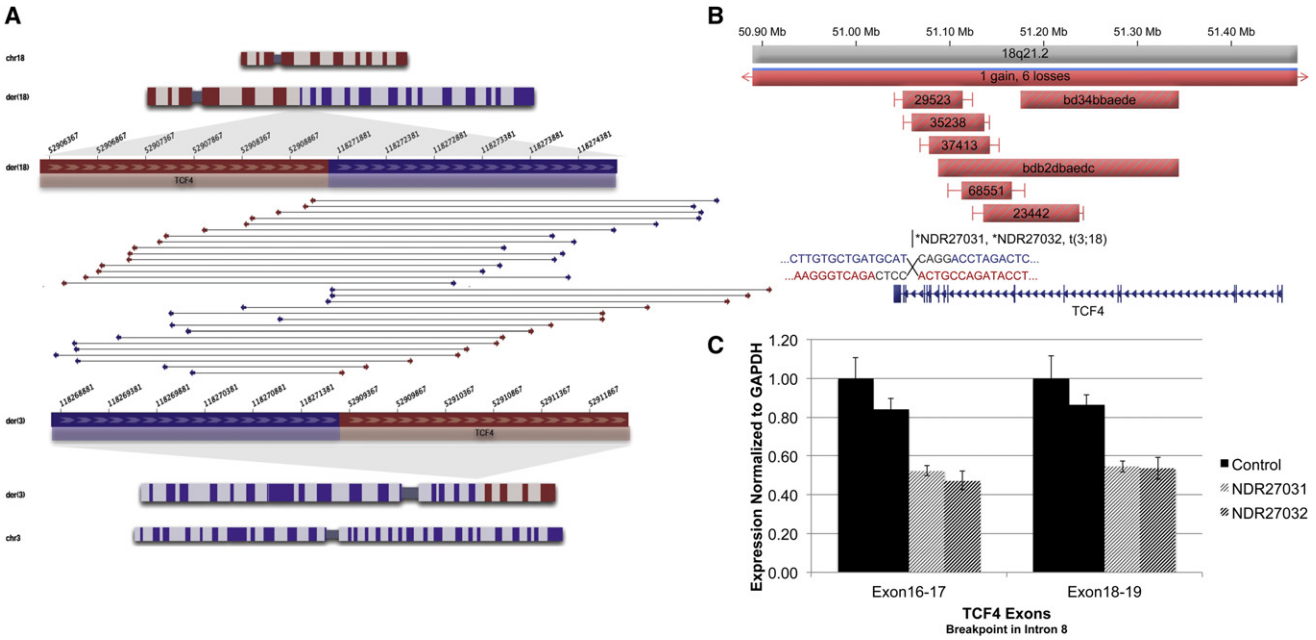

Figure 1. Convergent Genomic Evidence Implicates *TCF4* in Neurodevelopment

(A) A t(3;18)(q13.32;q21.2) translocation was sequenced with our custom jumping library protocol in monozygotic twin boys with multiple developmental abnormalities, directly disrupting *TCF4* in intron 8, a gene previously associated with Pitt-Hopkins syndrome (PHS) by mutation analysis and with schizophrenia by GWAS.

(B) Analysis of CNVs revealed disruption of coding and noncoding exons in 14 independent CNVs among 19,556 NDD cases compared to no CNVs observed in 13,991 controls ($p = 0.0006$), implicating hemizygosity of *TCF4* as a highly penetrant locus in NDD, PHS, and ASD (see also Table S2 and Rosenfeld et al., 2009b).

(C) Reduced mRNA expression of *TCF4* was detected in both cases from EBV-transformed lymphoblastoid cell lines compared to two gender-matched controls with two independent primers spanning exons 16-17 and exons 18-19. For related supplemental data, see also Figures S1 and S2, Tables S1 and S2, and Data S1: Phenotypic and Sequencing Information on Individual Patients.

delineated in Table S1 and discussed in the Supplemental Information.

Category 1: Balanced Alterations Confirming Loci Previously Implicated in Neurodevelopment

The power of sequencing BCAs as a discovery tool for loci contributing to disease etiology is illustrated in our independent identification of several genes previously suggested as candidates in ASD or NDD. We found direct disruption of *AUTS2* and *CDKL5,* two established neurodevelopmental loci (Bakkaloglu et al., 2008; Sultana et al., 2002). We also found disruption of the fork-head transcription factor *FOXP1* and the glutamate receptor *GRIN2B,* genes identified from de novo mutations in ASD by an exome-sequencing study (O'Roak et al., 2011). We also uncovered disruptions in *TCF4*, a gene mutated in Pitt-Hopkins syndrome as well as other genes implicated by CNV analysis of an autism cohort (the transcription factor *SOX5* and the dystrophin regulator *SNTG2)* (Rosenfeld et al., 2010). Consistent with their status as previously recognized contributors to neurodevelopment, the collective CNV burden of these genes is significant ($p = 7.74 \times 10^{-20}$; OR = 3.6) (Table 2).

Category 2: Individual Genes Contributing to Microdeletion Syndromes

Microdeletion syndromes, which contribute to neurodevelopmental and psychiatric disorders, typically involve hemizygosity of large genomic regions where the difficulty of defining specific genes responsible for core phenotypes has been an obstacle for clinical genetics, predictive diagnostics, and the study of disease pathogenesis. BCA sequencing in three ASD subjects pinpointed three individual gene contributors to microdeletion syndromes, each of which is involved in transcriptional and/or epigenetic regulation, in the 2q23.1, 2q33.1, and 9q34.3 microdeletion syndrome regions, respectively. In 2q23.1, a translocation disrupted *MBD5*, a member of the methyl-CpG binding domain protein family defined by a highly conserved methyl binding domain and including *MeCP2*, a causal locus in Rett syndrome. An international consortium follow-up study recently found 65 structural variations spanning the 2q23.1 microdeletion region in cases with syndromic features, including ASD, seizures, and intellectual disability, establishing *MBD5* as a necessary, sufficient, and predictive locus for a majority of the phenotypic features of the 2q23.1 microdeletion syndrome (Talkowski et al., 2011b). In the 2q33.1 and 9q34.3 regions, we respectively identified disruption of *SATB2,* a gene involved in transcriptional regulation and chromatin remodeling (Rosenfeld et al., 2009a), and of *EHMT1,* encoding a histone methyltransferase (Kleefstra et al., 2006). In a fourth region (15q11-13), the nested genes *SNURF-SNRPN* were disrupted at 15q11.2 in a subject with ASD and multiple developmental abnormalities, including sensory integration disorder, but without Angelman or Prader-Willi syndromes, both of which result from imprinting

Table 2. Analysis of Copy-Number Variants in Independent Samples

Gene Categories	NDD Count[a]	Control Count[a]	p Value	OR	95% CI
All Genes	443	63	2.1×10^{-47}	5.1	3.9–6.8
Category 1	114	23	7.7×10^{-20}	3.6	2.3–5.9
Category 2	169	12	1.6×10^{-26}	10.2	5.7–20.1
Category 3	160	28	2.2×10^{-15}	4.1	2.7–6.4
Category 4	111	12	5.1×10^{-15}	6.7	3.7–13.3
Categories 1, 3, and 4	274	51	5.2×10^{-24}	3.9	2.9–5.3

All Genes = all 33 loci disrupted by BCAs and included in the CNV analyses. Categories 1, 3, and 4 = analyses of all genes not localized to microdeletion syndromes known to be associated with increased CNV burden. See Table S2 for all genes in each category.
[a]Counts of CNVs from 19,556 NDD cases and 13,991 controls.

within the region. Our data argue for increased resolution and interpretation from molecular diagnostic testing of regional disorders, particularly for loci in which individual gene disruptions can yield phenotypes that are similar to or indistinguishable from that defining the syndrome. Given their localization to syndromic regions, it is not surprising that CNV analysis of genes in this category reflected a strong impact in neurodevelopment ($p = 1.64 \times 10^{-26}$; OR = 10.2) (Table 2 and Figure 3).

Category 3: Novel Genes Conferring Risk to Autism and Neurodevelopment

BCA sequencing yielded 22 novel ASD/NDD candidate genes that, like the genes in categories 1 and 2, are also likely to contribute to the neurodevelopmental phenotype in the corresponding subjects. The collective increase in CNV burden for these candidates was highly significant ($p = 2.21 \times 10^{-15}$, OR = 4.1) (Table 2).

The most significant individual genes are two loci localized to previously identified genomic disorder regions (*KIRREL3* and *SMG6*) and one encoding a DNA helicase (*CHD8*). Similar to the category 2 genes, we observed dysregulation by BCA and secondary CNV support for novel genes in regions associated with classic terminal deletion disorders: *KIRREL3* in 11q24.1 (Jacobsen syndrome) and *SMG6* in 17p13.3 (Miller-Dieker syndrome). *KIRREL3* encodes a cell adhesion molecule of the immunoglobulin family expressed in developing and adult brain of mouse and developing sensory pathways (Morikawa et al., 2007; Serizawa et al., 2006; Tamura et al., 2005). The locus was disrupted by a BCA 39.6 kb upstream of the mRNA coding region that altered both mRNA and protein levels (Table 1 and Figure S2). Disruption of *SMG6* nominates nonsense-mediated decay as another pathway in ASD and NDD, but not necessarily in the lissencephaly phenotype commonly seen in Miller-Dieker syndrome (see Supplemental Information and Figure S3). *CHD8* encodes a DNA helicase that remodels chromatin structure (Thompson et al., 2008). Although it has never been individually linked to a human disorder, it represents a strong autism and NDD candidate locus. It was disrupted by a BCA in a subject diagnosed with ASD (Table 1 and Data S1: Phenotypic and Sequencing Information on Individual Patients), was supported by our CNV analyses (Table S2, see also genes sensitive to dosage dysregulation section below), and was among the loci within the minimal region of overlap from previous analyses of de novo microdeletions (Zahir et al., 2007). CHD8 is a transcriptional repressor that also interacts with genes implicated in NDD such as *CHD7*, a causal locus in CHARGE syndrome. (Batsukh et al., 2010; Nishiyama et al., 2009; Rodríguez-Paredes et al., 2009).

Additionally, in seven subjects no annotated gene was disrupted but several expressed sequence tags (ESTs), conserved sequences, and regions with predicted regulatory effects were impacted by breakpoints (Table S1), suggesting that BCAs may also provide a novel entrée into such regions. In these subjects, we considered both loci in proximity to the breakpoint for positional effects (e.g., *KIRREL3*), as well as disrupted but functionally unannotated sequences themselves, such as the highly conserved 6q16.3 sequence of unknown function disrupted by a BCA in an ASD subject (denoted as "High Cons" in Table S1) and the noncoding RNA *LOC401324*.

Category 4: Neurodevelopmental Loci Associated with Risk that Crosses Traditional Diagnostic Boundaries

A remarkable number of genes implicated in ASD or NDD by single-gene disruption from BCAs in our study have also been recently associated with a spectrum of developmental, psychiatric, and behavioral phenotypes by other strategies, such as GWAS and mutation screening, including *TCF4* (Pitt-Hopkins syndrome, intellectual disability, schizophrenia) (Amiel et al., 2007; Blake et al., 2010; Rosenfeld et al., 2009b), *GRIN2B* (schizophrenia, bipolar disorder, and neurodevelopment) (Endele et al., 2010), *EHMT1* (schizophrenia) (Kirov et al., 2012), and four novel neurodevelopmental genes that overlap with category 3: *ZNF804A* (schizophrenia, psychosis, and cognitive function) (O'Donovan et al., 2008; Walters et al., 2010), *ANK3* (bipolar disorder and schizophrenia) (Ferreira et al., 2008; Williams et al., 2011), *C18orf1* (schizophrenia) (Meerabux et al., 2009), and *PDE10A* (schizophrenia) (Kehler and Nielsen, 2011). All loci with the exception of *ANK3* are supported by secondary CNV analyses (Table 1). Of these category 4 candidates not previously associated with ASD or NDD, only *PDE10A* has an established function, encoding a phosphodiesterase suggested as a biological candidate in schizophrenia due to its high tissue-specific expression in the caudate nucleus. Specific *PDE10A* inhibitors provide a potential therapeutic approach to schizophrenia due to their regulation of cAMP and cGMP, thereby altering dopamine D1 and D2 receptor activity (Kehler and Nielsen, 2011; Lakics et al., 2010).

Genes in this category appear to contribute to pleiotropic effects ranging from early-onset autism and intellectual disability to adult-onset psychosis, often through different mutational mechanisms. Several were previously associated with psychiatric disorders by unbiased GWAS and/or by candidate gene studies of common variants, which are thought to reflect a more subtle effect on gene regulation than the outright inactivation caused by the BCA disruption. One compelling example of different mutational mechanisms is *TCF4,* where rare mutations are recognized as causing NDD, sometimes with a diagnosis of Pitt-Hopkins syndrome, but common TCF4

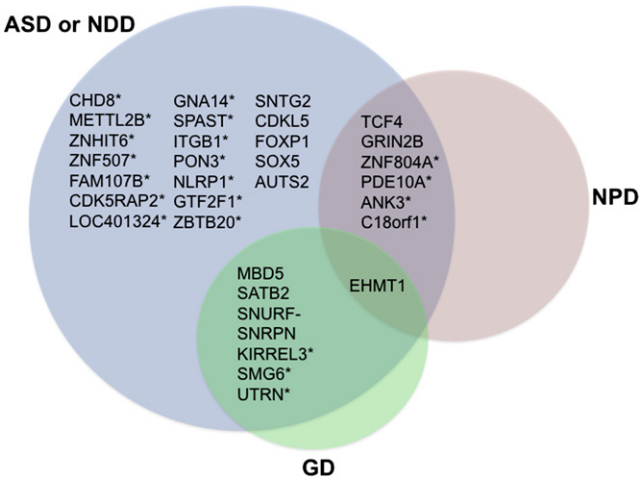

Figure 2. Genes Disrupted by Chromosomal Abnormalities Confer Risk across Diagnostic Groups

All genes disrupted by a BCA and analyzed in the CNV analyses are shown. Although all genes are implicated in ASD or NDD by BCA disruption in this study, some loci also represented single-gene contributors to previously recognized genomic disorder (GD) regions (three microdeletion syndromes, two terminal deletion syndromes, and one duplication syndrome). There were also genes discovered in ASD or NDD in this study that had been previously linked to adolescent- or adult-onset neuropsychiatric disorders (NPD) by common variation association studies. The asterisk (*) denotes a gene not previously implicated in ASD or NDD (category 3). See also Table 1 and Table S2 for CNV and GWAS support for each locus.

variation has recently emerged as a significant risk factor for schizophrenia. Taken together, our findings support the long-hypothesized notion of a neurodevelopmental component to adult-onset neuropsychiatric disorders such as schizophrenia (Murray and Lewis, 1987; Owen et al., 2011; Weinberger, 1986). However, our data extend this hypothesis to suggest that differing mutational impact on the same sets of genes constitutes a significant overlap in the genetic etiology of autism, schizophrenia, psychosis, bipolar disorder, and intellectual disability, comprising at least a subset of the total genetic variance for each of these disorders. The collective CNV burden for these genes decisively supports the fundamental hypothesis that some psychiatric disease-associated genes are important in neurodevelopment ($p = 5.1 \times 10^{-15}$; OR = 6.7) (Table 2).

Polygenic Risk from Genome-wide Association Studies

Our identification of genes disrupted by BCA in ASD or NDD with an increased CNV burden among diagnostic cases with neurodevelopmental abnormalities suggests that these are relatively penetrant alterations in human development, consistent with polygenic risk factors of modest to large effect. However, the discovery of the genes in category 4 suggests that for some loci, an accumulation of subtle genetic effects associated with common polymorphisms could have a pleiotropic impact across a spectrum of early childhood and adult-onset psychiatric disorders. To test this hypothesis, we performed gene-set enrichment analyses in data sets from GWASs of autism and schizophrenia (Ripke et al., 2011; Wang et al., 2009; Weiss et al., 2009) by using an established method in which each linkage disequilibrium block across the genome is scored with the maximum Z score achieved in the block (Rossin et al., 2011). Analysis of an initial autism study revealed a highly significant enrichment of risk alleles across the gene set (empirical p = 0.0018), a result that persisted in the second GWAS of autism (p = 0.068). Moreover, we discovered a significant enrichment of associated alleles from the largest GWAS meta-analysis of schizophrenia to date (empirical p = 0.0009). Struck by these results, we evaluated the potential for any unforeseen confounding variables by performing identical enrichment analyses in phenotype-permuted data sets from the meta-analysis of schizophrenia and autism studies (p = 0.444 and p = 0.518, respectively), in a well-powered GWAS study of Crohn's disease (p = 0.819) (Franke et al., 2010), and from GWAS data for seven other unrelated traits (p values ranged from p = 0.06 to p = 0.917, fitting nicely to the expected null distribution). These data indicate an unusually strong enrichment of subtle effects from common polygenic risk loci in autism and schizophrenia among the genes identified by our BCA sequencing and further support the hypothesis that diverse mutational mechanisms at these loci can confer pleiotropic effects across conventional diagnostic classifications (Figure 2).

Genes Sensitive to Dosage Dysregulation

The finding that some genes associated with ASD or NDD due to inactivation of one allele may also contribute to abnormal phenotypes when more subtly disrupted suggests that some ASD or NDD genes require tight control of their expression for appropriate neurodevelopmental function. In such circumstances, disruption by increased dosage might also produce an NDD phenotype. This is highlighted in more detailed examination of the CNV analyses presented in Figures 1, 3, and 4, which indicate that for some genes, the CNV data predict that both deletion and duplication are risk factors for abnormal neurodevelopment, whereas for other loci the mechanism of disruption appears to be dosage specific. For example, previously established NDD risk loci (*TCF4*, Figure 1; *SATB2* and *MBD5*, Figure 3) almost exclusively display deletion among the CNV cases, clearly supporting a similar mechanism of dysregulation to that seen for the BCA disruption. This is also true for NDD candidates such as *PDE10A* and *KIRREL3*. However, there are a number of genes similar to the three instances shown here for which CNV analysis supported genetic risk from both deletion and duplication, approaching a near 50:50 balance, including two well-established loci (*AUTS2* and *CDKL5*) (Figure 4A and Table S3). Interestingly, CNV analysis also supported genetic risk predominantly resulting from duplication of a locus (e.g., *CHD8*, *GRIN2B*, and *FOXP1*, Figure 4B), suggesting similar phenotypic outcomes from both dosage increase and heterozygous inactivation by BCA disruption, as in this study, or de novo mutations in a previous study (for *GRIN2B* and *FOXP1*) (O'Roak et al., 2011). These data build upon previous findings in recurrent rearrangement regions, such as the common recurrent 16p11.2 microdeletion/microduplication (Weiss et al., 2008), where both deletion and duplication increase risk for autism to different degrees and have disparate impacts (including reciprocal phenotypes) for other disorders and physiological traits.

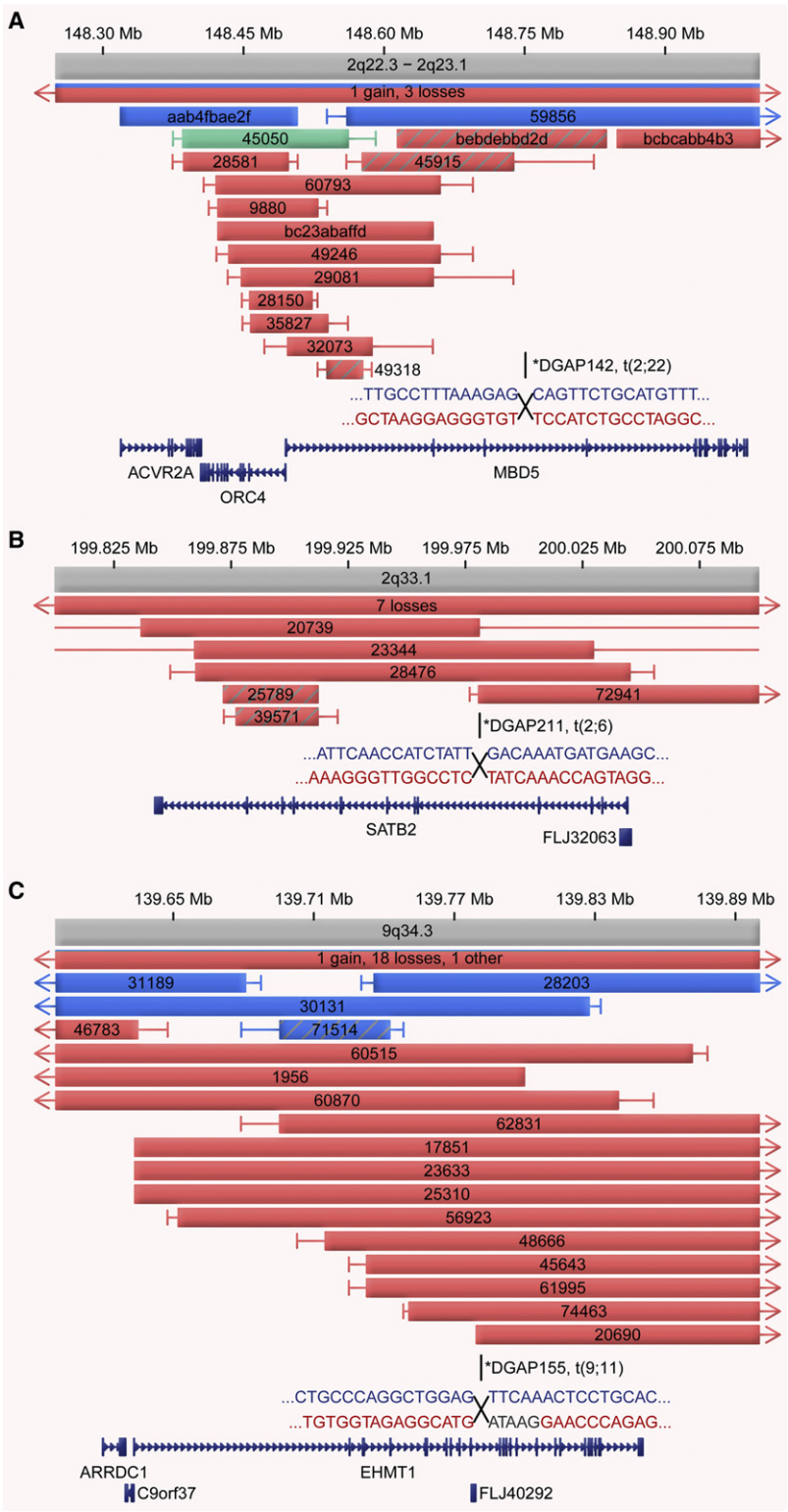

Figure 3. Disruption of Individual Genes in Microdeletion Syndromes Regions

(A–C) Gene-specific view of sequenced BCAs and CNVs implicating individual genes involved in transcriptional or epigenetic regulation and localized to previously described microdeletion syndromes, each of which implicates a single gene by the combined BCA and CNV analyses. Haploinsufficiency of mRNA was confirmed for each gene (Figure S1). A breakpoint disrupting a two-gene locus (*SNURF* and *SNRPN*) was detected in a fourth microdeletion syndrome, 15q11-13 (not shown). Red bars indicate copy loss, blue bars are copy gains, and striped bars are gene-specific alterations. Green bars represent "Other" alterations (balanced or unbalanced translocation, inversion, or complex alteration involving both deletion and duplication). The bar at the top provides the count of gains and losses that extend beyond the window and is color coded to reflect the relative proportion of each variant type. Confidence intervals provide the distance to the next probe without a detected dosage imbalance. A single RefSeq transcript is provided. Breakpoint sequence from each derivative is provided with text color coded by originating chromosome or nontemplated inserted sequence in gray. A subset of the *SATB2* patients included here were previously reported by Rosenfeld et al. (2009), and all cases reported here for *MBD5*, as well as 44 additional structural variants, are reported in a follow-up consortium study (Talkowski et al., 2011b). See also Table S3 and Figure S3.

The clear distinction between neurodevelopmental loci associated primarily with deletion or duplication and those displaying similarly abnormal neurodevelopment from either event emphasizes the need for detailed experimental annotation of the genome with respect to dosage-sensitive loci and phenotypic prediction.

Disrupted Genes Do Not Converge on a Single Pathway in Neurodevelopment

The genomics approach in this study enabled direct interpretation of locus specificity for further downstream analyses. We evaluated the networks in which these genes may participate and whether any biological pathways emerged as significantly enriched for genes disrupted by these BCAs. A qualitative network analysis based on interactions from PubMed abstracts and the use of a natural language processing algorithm identified a network of 429 interacting genes (Figure 5 and Figure S4). Fourteen of the original genes, many of which are involved in transcriptional regulation, were found to interact indirectly in a large, interconnected network, and *TCF4*, *SNRPN*, *CHD8*, and *GTF2F1* were confirmed as interacting partners of the RNA Polymerase II complex. Analysis of gene ontology (GO) terms found enrichment of transcription factors, phosphoproteins, and protein heterodimerization activity ($p < 0.005$). A quantitative network assessment did not find statistically significant first- or second-order interactions compared to chance expectations (given the size and composition of the given networks). The nominally significant

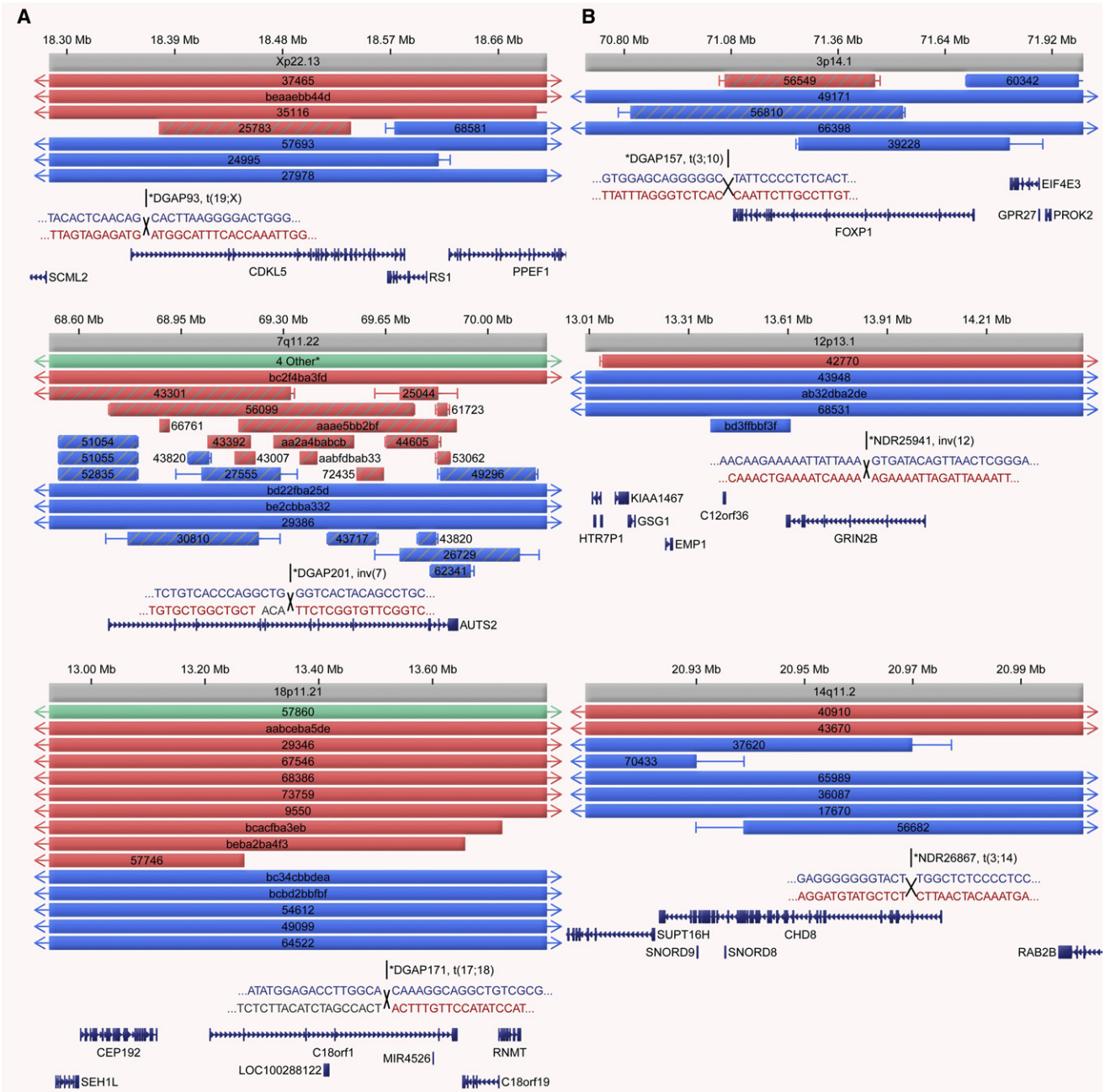

Figure 4. Genes Sensitive to Dosage Dysregulation

Provided is an example of genes illustrated by this study that appear to be sensitive to a spectrum of mutational mechanisms reciprocally altering their gene dosage. In Figures 1 and 2 are examples of genes such as *TCF4* and *SATB2* that are disrupted exclusively by deletions.

(A) Genes sensitive to dosage dysregulation from which both deletion and duplication appear to confer risk for similar phenotypes (*AUTS2*, *CDKL5*, *C18orf1*).

(B) Three loci, *FOXP1*, *GRIN2B*, and *CHD8*, are recognized for duplication of the locus as the predominant mutational mechanism, yet all three loci are implicated by heterozygous inactivation from BCA disruption in this study. Two of these genes, *FOXP1* and *GRIN2B* are also supported by de novo mutation in ASD families (O'Roak et al., 2011). *CHD8* is a completely novel candidate not previously implicated in ASD or any human abnormality. These data suggest annotation of the human genome is insufficient to reliably predict the impact of various mutational mechanisms of specific loci involved in neurodevelopment. See also Table S3.

networks (statistical enrichment of pathways at $p < 0.05$) included shared interactions between *SNTG2*, *UTRN*, *GNA14*, and *CDKRAP2* ($p = 0.01$) as well as *KCND2* and *GRIN2B* ($p = 0.03$) (Figure S4). The *SNURF-SNRPN* complex was nominally significant in both analyses and in an assessment of physical interactions (Supplemental Information). No results were significant after correction for multiple testing. There was also no convergence on one or a few pathways involved in

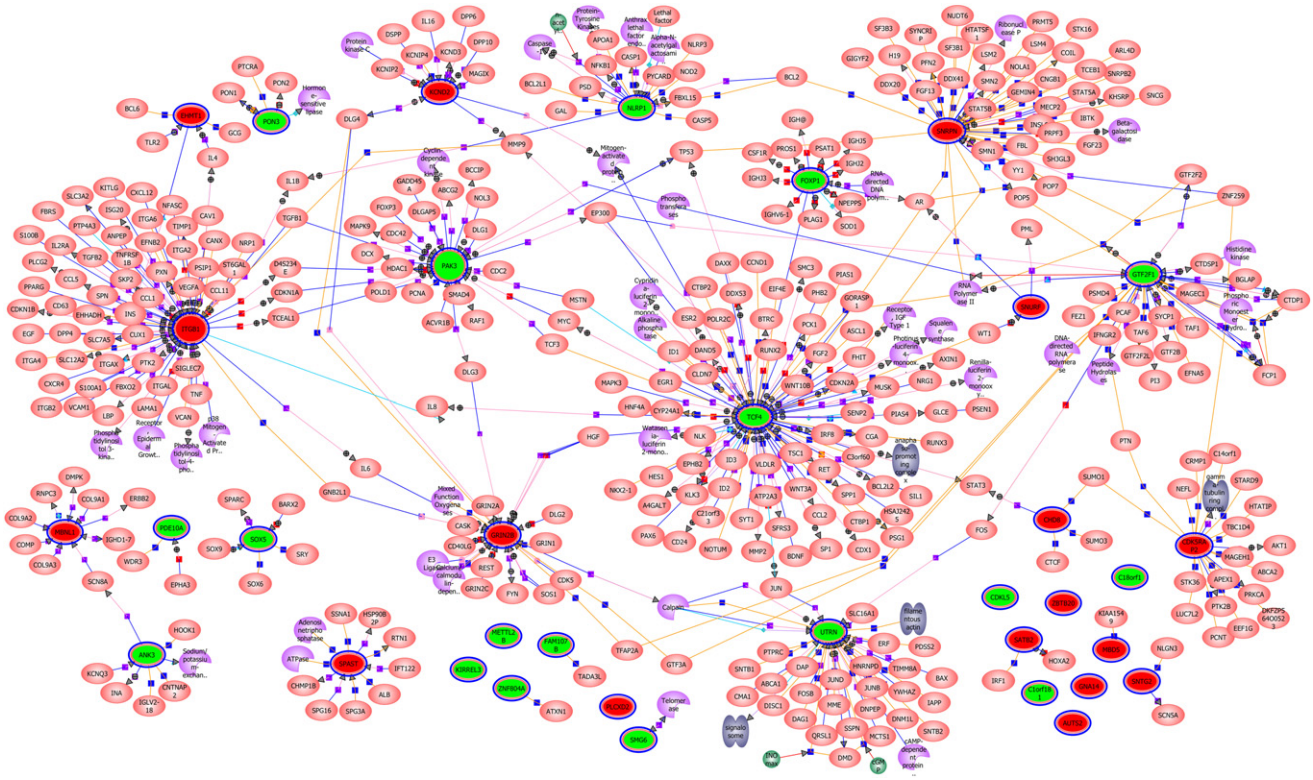

Figure 5. Network Analysis of Genes Implicated in Autism or Neurodevelopment in This Study

A large network of genes disrupted by BCAs in this study are connected by first-, second-, or higher-order interactions. No networks were significantly enriched for genes disrupted by BCAs after correction for multiple comparisons, though a number of loci have limited functional annotation available or remain of unknown function. See also Figure S4.

neurodevelopment, probably a reflection of the modest number of genes studied and the unknown function of many of the novel genes. These results could also signify that interaction of a diverse range of functional networks at many levels is critical to normal human neurodevelopment and that there is ample opportunity for genetic lesions to disrupt different functional pathways while still leading to similar neurodevelopmental outcomes.

DISCUSSION

Direct sequencing of BCA breakpoints, followed by targeted assessment of molecular diagnostic CNV findings in independent subjects, proved to be an efficient strategy for individual gene discovery in abnormal neurodevelopment. As this approach begins with sequence resolution of individual BCAs, it is not subject to notable limitations of other de novo mutation studies that rely purely on CNVs or on exome sequencing (i.e., primarily identification of large multigene regions in the former and failure to assess most noncoding sequence in the latter) and therefore provides an effective complement to these approaches. The yield of this study is considerable; 22 novel loci are disrupted by BCAs and 14 loci are supported from secondary analyses either by a statistical enrichment of CNV burden or disruption in multiple cases and absence of dosage alterations in controls, with several additional genes and sequences whose potential for contribution merits further examination. These findings not only extend specific knowledge of neurodevelopmental genes implicated by microdeletion syndromes, GWA, CNV, and exome sequencing studies but also reveal entirely unsuspected genes in ASD and NDD. This work also presents direct evidence for a complex genetic architecture that connects neurodevelopmental and adult-onset psychiatric disorders by implicating robustly associated schizophrenia loci as contributors to neurodevelopmental abnormalities.

Our analyses were based upon the premise that alone, a de novo BCA disruption in a single case, CNVs spanning a given region locus, or the presence of a de novo mutation within the coding region of an interesting biological candidate does not by itself represent compelling evidence that a gene contributes to neurodevelopment. Instead, we sought convergent genomic evidence combining disruption by a de novo BCA (or in two subjects a BCA that cosegregated with the phenotype) with association by CNV or mutation in independent and similarly affected cases, followed by comparison with risk loci from GWAS studies. The strength of evidence varied between loci, as shown in Table 1 and Table S2. In some instances, genes were statistically supported by secondary CNV analyses, in others, the statistical support was inconclusive, and in several the CNV data do not nominate the locus as an NDD candidate (e.g., *ZBTB20*; see

Table 1, Table S2, and Table S3). Additionally, like any genetic study, we cannot discount a potential contribution from altered expression of other genes at a distance from the site of the disruption. Although such an effect could result from the transposition of functional elements, the comprehensive testing of this possibility is problematic because it could involve dysregulation that is tissue-specific, that occurs due to the potentially altered nuclear organization of the rearranged chromosomes, or that is a secondary physiological consequence of the primary gene disruption. Nonetheless, secondary CNV analyses from independent NDD cases indicate a profound collective contribution of the disrupted genes on neurodevelopment. Simulations and subset analyses to evaluate empirical significance established that the increased CNV burden of the loci disrupted by BCAs was robust; unusually specific to neurodevelopmental disorders compared to other phenotypic presentations referred for molecular diagnostic testing; and not driven by any single gene, individual disruption category, cluster of symptoms, or discrete diagnostic category. Rather, it was an accumulation of risk factors from all four categories that collectively contributed to the significant burden observed (see Supplemental Information). For some genes, developmental abnormalities were predominantly associated with dosage alteration in only one direction, whereas for others the increased CNV burden involved both deletions and duplications. These data illustrate the immature state of annotation of the human genome with respect to dosage sensitivity and to prediction of phenotypic outcomes from genetic lesions, a limitation that may be alleviated at least in part by the growing capacity to sequence BCAs in a relatively high-throughput manner.

A surprising number of genes previously associated with adolescent- or adult-onset psychiatric disorders were disrupted in children with autism and NDD in this study. The concept of schizophrenia as a neurodevelopmental disorder has long been proposed (Murray and Lewis, 1987; Weinberger, 1986), and a growing consensus in the recent literature suggests that there are shared risk factors across what are viewed clinically as distinct phenotypic classifications, although few of these have been described at the individual gene level (Owen et al., 2011). Our study supports a shared genetic etiology for at least a portion of the phenotypic spectrum of schizophrenia, autism, and the neurodevelopmental abnormalities studied here. We find unambiguous gene disruption by BCAs in ASD and NDD subjects, coupled with an increased CNV burden and a substantial overrepresentation of polygenic risk compared to null expectations from GWAS of schizophrenia. Individual genes may thus show a differential risk depending on the nature of the genetic lesion, with heterozygous inactivation from BCA, CNV, or deactivating point mutation being a relatively penetrant contributor to ASD or NDD, whereas subtle effects from common variants contribute to later-onset disorders. However, contrary to this simple hypothesis, we were surprised to find persistent enrichment of these genes also among common variant associations from autism GWAS studies, suggesting that even subtle perturbation of genes important in normal human development can contribute to abnormal outcomes across the lifespan, presumably in interaction with other genetic and environmental influences.

This initial sequence-based delineation of a large collection of subjects harboring chromosomal aberrations in autism and related neurodevelopmental disorders establishes an approach that can be exploited for efficient discovery of individual genetic factors contributing to otherwise complex disorders. Each individual gene identified provides a new, specific hypothesis concerning the disease to be tested with further genetic and biological study. If supported, each then represents a foundation for investigations into the role that the biochemical activity and regulation of its product play in pathogenesis and into the potential for treatment through their manipulation. Ultimately, such data will also provide invaluable annotation of the human genome and profoundly impact the clinical interpretation of genomic events in subjects referred to diagnostic laboratories for autism and other developmental abnormalities.

EXPERIMENTAL PROCEDURES

Patients

Subjects were obtained from the Developmental Genome Anatomy Project (DGAP) (Higgins et al., 2008), the Autism Consortium of Boston, the Center for Human Genetic Research (CHGR) Neurodevelopmental Repository, and the Autism Genome Resource Exchange (AGRE). These studies were approved by the institutional review board of Partners HealthCare System. Clinical information was obtained by direct questionnaires, medical records, or structured clinical interviews (see Data S1: Phenotypic and Sequencing Information on Individual Patients).

Sequencing

Sequencing was performed on the Illumina platform (Illumina). Libraries were created by four different methods optimized for delineating BCAs, including (1) Illumina standard insert paired-end sequencing, (2) Illumina mate-pair sequencing (long 2,000–4,000 bp inserts), (3) our customized jumping libraries (long 3,000–4,500 bp inserts), and (4) our capture of breakpoints method (CapBP) for rearrangements previously localized (see Supplemental Information with complete protocols in Talkowski et al. [2011a]). Interestingly, some rearrangements proved to be far more complex than suspected by karyotyping, and in at least two subjects this complexity represented balanced germline chromothripsis similar to but distinct from previously described events in cancer cells (DGAP127 and DGAP203; see Table S4 and (Chiang et al., 2012)). In complex rearrangements where more than two genes were disrupted by BCA breakpoints, we conservatively excluded these cases from interpretation in our secondary CNV/GWAS analyses.

Bioinformatics

Sequencing reads were aligned via publicly available alignment programs and custom scripts, followed by processing of BAM files with a C++ program, Bamstat, to tabulate mapping statistics and output lists of anomalous read-pairs (defined as having ends that map to two different chromosomes, having an abnormal insert size, and/or unexpected strand orientations) (Talkowski et al., 2011a). Anomalous pairs were clustered by their mapped mates with a program that performs a single-linkage clustering of paired reads if corresponding ends map within a specified distance. All junction fragments predicted from paired-end sequencing were PCR amplified from genomic DNA, independent of the libraries and sequencing reads, and all breakpoints presented in this study were confirmed by capillary sequencing (Table S1 and also Table S5).

Evaluation of CNVs in Independent Subjects

Complete details on the CNV data, statistical analyses, and simulations to determine empirical significance are fully described in Supplemental Information. We compiled CNV data from 33,573 cases from molecular diagnostic laboratories of Signature Genomic Laboratories (SG) and Children's Hospital Boston (CHB) analyzed by oligonucleotide aCGH. A subset of the SG data

was included in a recent CNV study (Cooper et al., 2011). We fully annotated both data sets and collapsed all subjects with an indication for study (but without additional clinical information to verify a specific diagnosis) of autism, ASD, or a related neurodevelopmental abnormality into a combined NDD cohort (19,556 cases) and those for which the clinical indication did not indicate an NDD (14,017 cases). Control data were obtained across multiple published resources as described in the Supplemental Information (n = 13,991 independent controls). Only 6,239 controls were analyzed for CNVs on the X chromosome, and this reduced comparison group was used for analyses of *CDKL5*. The resolution between the clinical aCGH and higher-resolution control SNP microarrays varied extensively. To overcome this disparity but retain the native resolution of each individual array platform, we analyzed all CNVs that disrupted any documented coding or noncoding exon from all transcripts available from multiple database sources (Table S1). Notably, post hoc analyses reveal that a size filter of 100 kb, or the resolution of the most sparse control arrays, would have resulted in an almost identical CNV burden test to our exon disruption model ($p = 1.18 \times 10^{-47}$), but would have omitted high-confidence CNV calls in controls that could point toward reduced penetrance or nonsignificant loci in this study such as *ZBTB20*. Empirical significance was established by simulations performed in MATLAB (Mathworks) with custom scripts with all methods and findings described in the Supplemental Information. In brief, 1 million gene sets of 33 loci were randomly selected from the genome, and case-control CNV burden tests were performed for each random gene set. These same analyses were performed for NDD cases compared to all other cases in the molecular diagnostic cohort. Neither simulation detected a random gene set exceeding the significance of the experimental gene set. Another set of experiments performed subset analyses, randomly selecting up to 1,000 gene sets of each possible number of k genes from the experimental gene set of 33 loci, ranging from k = 5 to k = 33, finding a minimum of 90.5% of all subsets at a level as low as k = 5 exceeded multiple testing correction and 100% of the 1,000 gene sets to be significant at k > 13 (see Supplemental Information).

Genome-wide Association Studies

We performed gene-set enrichment analysis from GWAS data by using published methods (Rossin et al., 2011) on a recent meta-analysis of schizophrenia (Ripke et al., 2011) and two GWASs of autism (Weiss et al., 2009; Wang et al., 2009). Briefly, linkage disequilibrium blocks across the genome are scored with the maximum Z score achieved in the block. That score is corrected for the number of tests across the block via linear regression and the residuals are then used as the new, corrected score for each block. Genes nominate scores based upon the unique set of blocks they overlap, and the nominated scores are compared to background scores from all genes in the genome via a one-tailed rank-sum test.

Network Analysis

A qualitative assessment of interacting genes and complexes for each of the genes in which functional information was available (little data existed for some of the novel genes) was initially performed with Natural Language Processing (NLP) from published abstracts (see Supplemental Information). Secondary analyses of GO and Kyoto Encyclopedia Of Genes And Genomes (KEGG) enrichments were then performed followed by quantitative network building of first- and second-order interactions of the set of proteins coded by the genes disrupted by BCAs. The significance of these networks was determined by permutation testing (Supplemental Information).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and five tables and can be found with this article online at doi:10.1016/j.cell.2012.03.028.

ACKNOWLEDGMENTS

We are grateful to all participating subjects and families and to the many healthcare professionals who have contributed to this study, including Mary-Alice Abbott, Darius J. Adams, Kwame Anyane-Yeboa, Stephen G. Bamforth, Tina Bartell, David P. Bick, Joann N. Bodurtha, Carol Clericuzio, Stephanie Cohen, Kristin Dalton, Maria Descartes, Joanne Milisa Drautz, Dawn L. Earl, Luis F. Escobar, Shannon Gerner, Edwin Guzman, Kenneth Handelman, Tim Heshka, Robert J. Hopkin, Micheil Innes, Debby Lambert, Emmanuelle Lemyre, Cynthia Lim, Livija Medne, Graciela Moya, Katie Rutledge, Wendy Smith, Mark Stephan, Darci Sternen, Katie Stoll, Paulien van Galen, Nancy J. Van Vranken, Erica Wahl, Susan E. Wiley, Amy L. White, Anne Woods, and Elaine H. Zackai. The invaluable control data for this study were provided by Pamela Sklar, Shaun Purcell, the International Schizophrenia Consortium, the Wellcome Trust Case Control Consortium, Evan Eichler, Bradley Coe, and Greg Cooper. We thank Tammy Gillis, Mary Anne Anderson, Jayla Ruliera, and Thon de Boer for technical assistance. We also thank Dennis Gurgul, Nilay Roy, and Brent Richter of Partners Research Computing at Massachusetts General Hospital, and contributing staff from Signature Genomic Laboratories and Children's Hospital Boston. This work was funded by grants GM061354 and HD065286 from the National Institutes of Health, the Simons Foundation Autism Research Initiative, and Autism Speaks. This research was also supported by the Division of Intramural Research, National Human Genome Research Institute, National Institutes of Health and Human Services. M.T. was supported by a National Institute of Mental Health National Research Service Award (MH087123) and an Massachusetts General Hospital Executive Committee on Research Fund for Medical Discovery Award. L.G.S. and J.A.R. are employees of Signature Genomic Laboratories, PerkinElmer.

Received: November 23, 2011
Revised: February 27, 2012
Accepted: March 28, 2012
Published online: April 19, 2012

REFERENCES

Amiel, J., Rio, M., de Pontual, L., Redon, R., Malan, V., Boddaert, N., Plouin, P., Carter, N.P., Lyonnet, S., Munnich, A., and Colleaux, L. (2007). Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. Am. J. Hum. Genet. *80*, 988–993.

Bakkaloglu, B., O'Roak, B.J., Louvi, A., Gupta, A.R., Abelson, J.F., Morgan, T.M., Chawarska, K., Klin, A., Ercan-Sencicek, A.G., Stillman, A.A., et al. (2008). Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. Am. J. Hum. Genet. *82*, 165–173.

Batsukh, T., Pieper, L., Koszucka, A.M., von Velsen, N., Hoyer-Fender, S., Elbracht, M., Bergman, J.E., Hoefsloot, L.H., and Pauli, S. (2010). CHD8 interacts with CHD7, a protein which is mutated in CHARGE syndrome. Hum. Mol. Genet. *19*, 2858–2866.

Blake, D.J., Forrest, M., Chapman, R.M., Tinsley, C.L., O'Donovan, M.C., and Owen, M.J. (2010). TCF4, schizophrenia, and Pitt-Hopkins Syndrome. Schizophr. Bull. *36*, 443–447.

Chiang, C., Jacobsen, J.C., Ernst, C., Hanscom, C., Heilbut, A., Blumenthal, I., Mills, R.E., Kirby, A., Lindgren, A.M., Rudiger, S.R., et al. (2012). Complex reorganization and predominant non-homologous repair following chromosomal breakage in karyotypically balanced germline rearrangements and transgenic integration. Nat. Genet. *44*, 390–397.

Chiarle, R., Zhang, Y., Frock, R.L., Lewis, S.M., Molinie, B., Ho, Y.J., Myers, D.R., Choi, V.W., Compagno, M., Malkin, D.J., et al. (2011). Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. Cell *147*, 107–119.

Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V., et al. (2011). A copy number variation morbidity map of developmental delay. Nat. Genet. *43*, 838–846.

Endele, S., Rosenberger, G., Geider, K., Popp, B., Tamer, C., Stefanova, I., Milh, M., Kortüm, F., Fritsch, A., Pientka, F.K., et al. (2010). Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. Nat. Genet. *42*, 1021–1026.

Ferreira, M.A., O'Donovan, M.C., Meng, Y.A., Jones, I.R., Ruderfer, D.M., Jones, L., Fan, J., Kirov, G., Perlis, R.H., Green, E.K., et al; Wellcome Trust Case Control Consortium. (2008). Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. Nat. Genet. *40*, 1056–1058.

Franke, A., McGovern, D.P., Barrett, J.C., Wang, K., Radford-Smith, G.L., Ahmad, T., Lees, C.W., Balschun, T., Lee, J., Roberts, R., et al. (2010). Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat. Genet. *42*, 1118–1125.

Higgins, A.W., Alkuraya, F.S., Bosco, A.F., Brown, K.K., Bruns, G.A., Donovan, D.J., Eisenman, R., Fan, Y., Farra, C.G., Ferguson, H.L., et al. (2008). Characterization of apparently balanced chromosomal rearrangements from the developmental genome anatomy project. Am. J. Hum. Genet. *82*, 712–722.

Kehler, J., and Nielsen, J. (2011). PDE10A inhibitors: novel therapeutic drugs for schizophrenia. Curr. Pharm. Des. *17*, 137–150.

Kirov, G., Pocklington, A.J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., Moran, J., Chambert, K., Toncheva, D., Georgieva, L., et al. (2012). De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol. Psychiatry *17*, 142–153.

Kleefstra, T., Brunner, H.G., Amiel, J., Oudakker, A.R., Nillesen, W.M., Magee, A., Geneviève, D., Cormier-Daire, V., van Esch, H., Fryns, J.P., et al. (2006). Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. Am. J. Hum. Genet. *79*, 370–377.

Klein, I.A., Resch, W., Jankovic, M., Oliveira, T., Yamane, A., Nakahashi, H., Di Virgilio, M., Bothmer, A., Nussenzweig, A., Robbiani, D.F., et al. (2011). Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. Cell *147*, 95–106.

Lakics, V., Karran, E.H., and Boess, F.G. (2010). Quantitative comparison of phosphodiesterase mRNA distribution in human brain and peripheral tissues. Neuropharmacology *59*, 367–374.

Marshall, C.R., Noor, A., Vincent, J.B., Lionel, A.C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y., et al. (2008). Structural variation of chromosomes in autism spectrum disorder. Am. J. Hum. Genet. *82*, 477–488.

Meerabux, J.M., Ohba, H., Iwayama, Y., Maekawa, M., Detera-Wadleigh, S.D., DeLisi, L.E., and Yoshikawa, T. (2009). Analysis of a t(18;21)(p11.1;p11.1) translocation in a family with schizophrenia. J. Hum. Genet. *54*, 386–391.

Morikawa, Y., Komori, T., Hisaoka, T., Ueno, H., Kitamura, T., and Senba, E. (2007). Expression of mKirre in the developing sensory pathways: its close apposition to nephrin-expressing cells. Neuroscience *150*, 880–886.

Murray, R.M., and Lewis, S.W. (1987). Is schizophrenia a neurodevelopmental disorder? Br. Med. J. (Clin. Res. Ed.) *295*, 681–682.

Nishiyama, M., Oshikawa, K., Tsukada, Y., Nakagawa, T., Iemura, S., Natsume, T., Fan, Y., Kikuchi, A., Skoultchi, A.I., and Nakayama, K.I. (2009). CHD8 suppresses p53-mediated apoptosis through histone H1 recruitment during early embryogenesis. Nat. Cell Biol. *11*, 172–182.

O'Donovan, M.C., Craddock, N., Norton, N., Williams, H., Peirce, T., Moskvina, V., Nikolov, I., Hamshere, M., Carroll, L., Georgieva, L., et al; Molecular Genetics of Schizophrenia Collaboration. (2008). Identification of loci associated with schizophrenia by genome-wide association and follow-up. Nat. Genet. *40*, 1053–1055.

O'Roak, B.J., Deriziotis, P., Lee, C., Vives, L., Schwartz, J.J., Girirajan, S., Karakoc, E., Mackenzie, A.P., Ng, S.B., Baker, C., et al. (2011). Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat. Genet. *43*, 585–589.

Owen, M.J., O'Donovan, M.C., Thapar, A., and Craddock, N. (2011). Neurodevelopmental hypothesis of schizophrenia. Br. J. Psychiatry *198*, 173–175.

Ravel, C., Berthaut, I., Bresson, J.L., and Siffroi, J.P.; Genetics Commission of the French Federation of CECOS. (2006). Prevalence of chromosomal abnormalities in phenotypically normal and fertile adult males: large-scale survey of over 10,000 sperm donor karyotypes. Hum. Reprod. *21*, 1484–1489.

Ripke, S., Sanders, A.R., Kendler, K.S., Levinson, D.F., Sklar, P., Holmans, P.A., Lin, D.Y., Duan, J., Ophoff, R.A., Andreassen, O.A., et al; Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. (2011). Genome-wide association study identifies five new schizophrenia loci. Nat. Genet. *43*, 969–976.

Rodríguez-Paredes, M., Ceballos-Chávez, M., Esteller, M., García-Domínguez, M., and Reyes, J.C. (2009). The chromatin remodeling factor CHD8 interacts with elongating RNA polymerase II and controls expression of the cyclin E2 gene. Nucleic Acids Res. *37*, 2449–2460.

Rosenfeld, J.A., Ballif, B.C., Lucas, A., Spence, E.J., Powell, C., Aylsworth, A.S., Torchia, B.A., and Shaffer, L.G. (2009a). Small deletions of SATB2 cause some of the clinical features of the 2q33.1 microdeletion syndrome. PLoS ONE *4*, e6568.

Rosenfeld, J.A., Leppig, K., Ballif, B.C., Thiese, H., Erdie-Lalena, C., Bawle, E., Sastry, S., Spence, J.E., Bandholz, A., Surti, U., et al. (2009b). Genotype-phenotype analysis of TCF4 mutations causing Pitt-Hopkins syndrome shows increased seizure activity with missense mutations. Genet. Med. *11*, 797–805.

Rosenfeld, J.A., Ballif, B.C., Torchia, B.S., Sahoo, T., Ravnan, J.B., Schultz, R., Lamb, A., Bejjani, B.A., and Shaffer, L.G. (2010). Copy number variations associated with autism spectrum disorders contribute to a spectrum of neurodevelopmental disorders. Genet. Med. *12*, 694–702.

Rosenfeld, J.A., Ballif, B.C., Lucas, A., Spence, E.J., Powell, C., Aylsworth, A.S., Torchia, B.A., and Shaffer, L.G. (2009). Small deletions of SATB2 cause some of the clinical features of the 2q33.1 microdeletion syndrome. PLoS ONE *4*, e6568.

Rossin, E.J., Lage, K., Raychaudhuri, S., Xavier, R.J., Tatar, D., Benita, Y., Cotsapas, C., and Daly, M.J.; International Inflammatory Bowel Disease Genetics Constortium. (2011). Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. PLoS Genet. *7*, e1001273.

Serizawa, S., Miyamichi, K., Takeuchi, H., Yamagishi, Y., Suzuki, M., and Sakano, H. (2006). A neuronal identity code for the odorant receptor-specific and activity-dependent axon sorting. Cell *127*, 1057–1069.

Stephens, P.J., Greenman, C.D., Fu, B., Yang, F., Bignell, G.R., Mudie, L.J., Pleasance, E.D., Lau, K.W., Beare, D., Stebbings, L.A., et al. (2011). Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell *144*, 27–40.

Sultana, R., Yu, C.E., Yu, J., Munson, J., Chen, D., Hua, W., Estes, A., Cortes, F., de la Barra, F., Yu, D., et al. (2002). Identification of a novel gene on chromosome 7q11.2 interrupted by a translocation breakpoint in a pair of autistic twins. Genomics *80*, 129–134.

Talkowski, M.E., Ernst, C., Heilbut, A., Chiang, C., Hanscom, C., Lindgren, A., Kirby, A., Liu, S., Muddukrishna, B., Ohsumi, T.K., et al. (2011a). Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. Am. J. Hum. Genet. *88*, 469–481.

Talkowski, M.E., Mullegama, S.V., Rosenfeld, J.A., van Bon, B.W., Shen, Y., Repnikova, E.A., Gastier-Foster, J., Thrush, D.L., Kathiresan, S., Ruderfer, D.M., et al. (2011b). Assessment of 2q23.1 microdeletion syndrome implicates MBD5 as a single causal locus of intellectual disability, epilepsy, and autism spectrum disorder. Am. J. Hum. Genet. *89*, 551–563.

Tamura, S., Morikawa, Y., Hisaoka, T., Ueno, H., Kitamura, T., and Senba, E. (2005). Expression of mKirre, a mammalian homolog of Drosophila kirre, in the developing and adult mouse brain. Neuroscience *133*, 615–624.

Thompson, B.A., Tremblay, V., Lin, G., and Bochar, D.A. (2008). CHD8 is an ATP-dependent chromatin remodeling factor that regulates beta-catenin target genes. Mol. Cell. Biol. *28*, 3894–3904.

Walters, J.T., Corvin, A., Owen, M.J., Williams, H., Dragovic, M., Quinn, E.M., Judge, R., Smith, D.J., Norton, N., Giegling, I., et al. (2010). Psychosis susceptibility gene ZNF804A and cognitive performance in schizophrenia. Arch. Gen. Psychiatry *67*, 692–700.

Wang, K., Zhang, H., Ma, D., Bucan, M., Glessner, J.T., Abrahams, B.S., Salyakina, D., Imielinski, M., Bradfield, J.P., Sleiman, P.M., et al. (2009).

Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature *459*, 528–533.

Weinberger, D. (1986). The pathogenesis of schizophrenia: a neurodevelopmental theory. In The Neurology of Schizophrenia, R.N.D. Weinberger, ed. (Amsterdam: Elsevier), pp. 387–405.

Weiss, L.A., Shen, Y., Korn, J.M., Arking, D.E., Miller, D.T., Fossdal, R., Saemundsen, E., Stefansson, H., Ferreira, M.A., Green, T., et al; Autism Consortium. (2008). Association between microdeletion and microduplication at 16p11.2 and autism. N. Engl. J. Med. *358*, 667–675.

Weiss, L.A., Arking, D.E., Daly, M.J., and Chakravarti, A.; Gene Discovery Project of Johns Hopkins & the Autism Consortium. (2009). A genome-wide linkage and association scan reveals novel loci for autism. Nature *461*, 802–808.

Williams, H.J., Craddock, N., Russo, G., Hamshere, M.L., Moskvina, V., Dwyer, S., Smith, R.L., Green, E., Grozeva, D., Holmans, P., et al. (2011). Most genome-wide significant susceptibility loci for schizophrenia and bipolar disorder reported to date cross-traditional diagnostic boundaries. Hum. Mol. Genet. *20*, 387–391.

Zahir, F., Firth, H.V., Baross, A., Delaney, A.D., Eydoux, P., Gibson, W.T., Langlois, S., Martin, H., Willatt, L., Marra, M.A., and Friedman, J.M. (2007). Novel deletions of 14q11.2 associated with developmental delay, cognitive impairment and similar minor anomalies in three children. J. Med. Genet. *20*, 556–561.