REPORT

Cryptic and Complex Chromosomal Aberrations in Early-Onset Neuropsychiatric Disorders

Harrison Brand,[1,2,9] Vamsee Pillalamarri,[1,9] Ryan L. Collins,[1] Stacey Eggert,[1,2,3] Colm O'Dushlaine,[4] Ellen B. Braaten,[5] Matthew R. Stone,[1] Kimberly Chambert,[4] Nathan D. Doty,[5] Carrie Hanscom,[1] Jill A. Rosenfeld,[6,10] Hillary Ditmars,[1] Jessica Blais,[1] Ryan Mills,[7] Charles Lee,[8] James F. Gusella,[1,2,3,4] Steven McCarroll,[3,4] Jordan W. Smoller,[1,4,5] Michael E. Talkowski,[1,2,4,5,*] and Alysa E. Doyle[1,4,5,*]

Structural variation (SV) is a significant component of the genetic etiology of both neurodevelopmental and psychiatric disorders; however, routine guidelines for clinical genetic screening have been established only in the former category. Genome-wide chromosomal microarray (CMA) can detect genomic imbalances such as copy-number variants (CNVs), but balanced chromosomal abnormalities (BCAs) still require karyotyping for clinical detection. Moreover, submicroscopic BCAs and subarray threshold CNVs are intractable, or cryptic, to both CMA and karyotyping. Here, we performed whole-genome sequencing using large-insert jumping libraries to delineate both cytogenetically visible and cryptic SVs in a single test among 30 clinically referred youth representing a range of severe neuropsychiatric conditions. We detected 96 SVs per person on average that passed filtering criteria above our highest-confidence resolution (6,305 bp) and an additional 111 SVs per genome below this resolution. These SVs rearranged 3.8 Mb of genomic sequence and resulted in 42 putative loss-of-function (LoF) or gain-of-function mutations per person. We estimate that 80% of the LoF variants were cryptic to clinical CMA. We found myriad complex and cryptic rearrangements, including a "paired" duplication (360 kb, 169 kb) that flanks a 5.25 Mb inversion that appears in 7 additional cases from clinical CNV data among 47,562 individuals. Following convergent genomic profiling of these independent clinical CNV data, we interpreted three SVs to be of potential clinical significance. These data indicate that sequence-based delineation of the full SV mutational spectrum warrants exploration in youth referred for neuropsychiatric evaluation and clinical diagnostic SV screening more broadly.

Structural variation (SV) is a major component of the genetic etiology of neurodevelopmental disorders. In recent years, enrichment of large, de novo copy-number variants (CNVs) and balanced chromosomal abnormalities (BCAs) has been reported and replicated in youth with autism spectrum disorder (ASD [MIM 209850]), developmental delay (DD), and intellectual disability (ID).[1–5] At present, genetic testing is frequently included in diagnostic evaluation of such youth, with chromosomal microarray (CMA) serving as the recommended first-tier genetic screen since 2010 based on a consensus statement in this journal.[6–8] For ASD, the use of CMA reflects the recognition that, in addition to the subset of cases with clinical features that can indicate a known genetic syndrome (e.g., Fragile X [MIM 300624]), nonsyndromic cases may benefit from genome-wide CNV evaluation.[9] Nonetheless, despite recommendations that extend across the full autism spectrum, genetic testing is not pursued for all individuals.[10] A significantly increased burden of large CNVs has also been observed in psychiatric disorders, including attention deficit hyperactivity disorder (ADHD [MIM 143465]), Tourette syndrome (MIM 137580), schizophrenia (MIM 181500), and early-onset psychosis and bipolar disorder (MIM 125480).[11–15] Notably, psychiatric and neurodevelopmental conditions often co-occur,[16] and findings for both rare SVs and common polymorphic risk variants suggest an overlapping etiology.[17,18] There is no current consensus on CMA or even general genetic testing for psychiatric disorders, although its potential benefit in this population has been discussed.[19,20]

Array-based technologies such as CMA can capture relative dosage imbalances that are a consequence of aneuploidy, CNV, and unbalanced translocation. In developmental disorders, the implementation of CMA as a first-tier genome-wide screen has significantly improved diagnostic yield over conventional karyotyping or gene-based mutation screening. One study of 6,539 consecutive referrals to Signature Genomics identified at least one clinically significant CNV in 17.6%–22.5% of cases, depending on the resolution of the array test performed (whole-genome BAC versus oligonucleotide).[21] In referrals for whom no causal genetic lesion is detected, however, additional SV testing is rarely pursued for mutations that are cryptic to CMA (defined herein as below the resolution

[1]Psychiatric and Neurodevelopmental Genetics Unit, Molecular Neurogenetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA; [2]Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA; [3]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA; [4]Stanley Center for Psychiatric Research and Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02141, USA; [5]Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA; [6]Signature Genomic Laboratories, PerkinElmer, Inc., Spokane, WA 99207, USA; [7]Departments of Computational Medicine & Bioinformatics and Human Genetics, University of Michigan Medical Center, Ann Arbor, MI 48109, USA; [8]The Jackson Laboratory for Genomic Medicine, Farmington, CT 06030, USA
[9]These authors contributed equally to this work
[10]Present address: Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA
*Correspondence: talkowski@chgr.mgh.harvard.edu (M.E.T.), doylea@helix.mgh.harvard.edu (A.E.D.)
http://dx.doi.org/10.1016/j.ajhg.2014.09.005. ©2014 by The American Society of Human Genetics. All rights reserved.

CrossMark

of detection for a given technology), and the impact of cryptic rearrangements is therefore unknown in this population. With respect to BCAs, karyotyping remains the only conventional diagnostic method capable of surveying their presence, as illustrated in a recent prenatal diagnostic comparison of CMA and karyotyping.[22] Karyotyping transformed human genetics more than half a century ago by opening access to gross chromosomal changes through microscopic visualization of alterations to chromosome banding patterns. The method is limited to a resolution of ~5–10 Mb, depending on banding patterns within the breakpoint regions, and has yet to be supplanted in diagnostic practice by higher-resolution methods for BCA detection. Moreover, cryptic BCAs are intractable to all conventional clinical genetic screening. These BCAs are not measurable at karyotype resolution or by CMA, and have not been delineated by whole-exome sequencing (WES) or low-depth whole-genome sequencing from the 1000 Genomes Project to date.[23,24] The mutational spectrum of cryptic SVs (submicroscopic BCAs and small CNVs) therefore represents a largely uncharacterized source of potential loss-of-function (LoF) mutations in biomedical research and a blind spot in genetic diagnostics.

In a series of previous studies, we have shown that whole-genome sequencing (WGS) using large inserts of several kilobases (referred to herein as jumping libraries) can delineate cytogenetically visible BCAs in both a research capacity and prenatal diagnostic practice.[5,25–28] These methods provide a single technology capable of detecting both CMA-resolution CNVs and karyotype-resolution BCAs, as well as cryptic SVs. This approach thus allows whole-genome detection of the full SV mutational spectrum at a time and cost comparable to CMA and karyotype.[26] In the current study, we used this jumping library sequencing approach to evaluate the presence and potential impact of both cytogenetically visible and cryptic chromosomal aberrations in a clinically referred sample of children and adolescents. Specifically, we sequenced youth with a range of severe neuropsychiatric disorders (NPDs; i.e., neurodevelopmental and psychiatric conditions) whom we hypothesize are enriched for LoF variation.

Subjects were obtained through the Longitudinal Study of Genetic Influences on Cognition (LOGIC), which collects deep cognitive and psychiatric phenotyping and DNA on youth referred for neuropsychiatric evaluation. The study also collects abbreviated phenotypes and DNA on first-degree relatives where possible. Our goal for the current proof-of-concept project was the sequencing of genomes from 30 youth referred for clinical neuropsychiatric evaluation. We selected 29 probands (ages 4–19) as well as an affected sibling, an affected father, and a healthy mother from a four-member multiplex family (32 total subjects). Specifically, we selected the 29 consecutive cases at the time of analysis who had provided DNA via whole blood (as opposed to saliva) and who manifested particular diagnoses in order of our priorities. First, we prioritized individuals with severe early-onset psychiatric presentations reflecting the psychosis or mood disorder spectrum, regardless of their comorbidities (n = 18; 62%). Second, we included youth with other neuropsychiatric disorders (i.e., autism spectrum and ADHD) and some evidence of severe presentation (i.e., comorbidity, prior psychiatric hospitalization). As shown in Table S1 available online, a total of 25 youths met full diagnostic criteria for one or more psychiatric conditions, and 7 of these met criteria for a comorbid neurodevelopmental disorder. Thus, more than half of the sample (55%) had an exclusively psychiatric disorder (see Table S1 for complete details). We note that only one sample had previously undergone CMA analysis with no significant variants detected, and no samples had been previously referred for targeted gene panel testing. All subjects provided informed consent, and this study was approved by the Institutional Review Board of Partners HealthCare.

Large-insert jumping libraries were generated using our previously published protocols, which are provided in complete detail in Hanscom and Talkowski.[27] The method generates genomic libraries in which short end reads (25 base pairs in this study) are separated by long inserts (targeted to 2.5 kb in this study), yielding very high coverage of mapped inserts spanning the genome for minimal sequencing cost. Following library preparation, sequencing was performed on all samples on an Illumina HiSeq 2000, generating a median insert size of 2.6 kb and median insert coverage of 62× per library.[27] Analysis of large-insert jumping libraries leverages spatial relationships of mated reads to trace distinctive breakpoint signatures rather than relying upon coverage from the actual nucleotides sequenced.[25] Expanding upon our previous methods to delineate karyotypically visible BCAs,[25,28] for this study we developed a SV classifier for WGS using jumping libraries with a targeted emphasis on reducing type I (false-positive) errors that can present a major barrier to interpretation (see Figure S1 for details). In brief, we clustered anomalously mapping read pairs across all samples using BamStat and ReadPairCluster.[25] We next computed a set of metrics for each cluster based on its constituent reads and properties of the genomic region spanned by the cluster (see Figure S1). Each cluster was classified based on thresholds calculated from a training set of PCR and Sanger sequencing validated SVs. We executed this process within a joint calling framework to mitigate false positive variant classifications that are a consequence of reference misassembly or systematic short read alignment errors in regions with alignment biases (e.g., highly repetitive regions such as segmental duplications). Across the genomes of our 29 probands, we identified 98 deletion, 43 tandem duplication, 99 inversion, and 112 interchromosomal insertion clusters that occurred in 90% or more of probands, most of which appear to be systematic mapping errors in complex genomic regions.

After excluding all reference variation and alignment artifacts, we tested the precision of our classification algorithm by investigating the inheritance of SV calls among

Table 1. **Counts of All Structural Variants**[a]

Event Classifications	All SV Observations	SV Count	Private SV Count	Polymorphic SV Count
Deletions	3,234	622	318	304
Tandem duplications	383	119	72	47
Inversions	888	112	26	86
Interchromosomal insertions	701	80	21	59
Intrachromosomal insertions	633	55	7	48
Complex chromosomal rearrangements	170	16	6	10
Total	6,009	1,004	450	554

[a]Structural-variant counts from large-insert sequencing after filtering for mapping and reference artifacts.

our four-member family. We performed the initial analyses at a resolution approximately equal to the maximum insert size of all samples that the BWA aligner will designate as reads having a proper alignment ($\geq$6,305 bp in this study). Subthreshold CNV detection is therefore unreliable as deletions are indistinguishable from proper pairs below this high-confidence resolution.[29] In total, we identified 190 high-confidence SVs between both children, 188 (98.9%) of which exhibited normal Mendelian inheritance upon inspection of corresponding SV clusters in parents. We observed two events in the proband and affected sibling (a 6.5 kb interchromosomal insertion of chr9 into chr22 and a 90 kb deletion at 14q11.2) that were not in either parent, but detected in at least one other unrelated sample in our larger cohort. This suggests that these polymorphic variants were not captured or filtered in the parent (false negatives). We next considered balanced SVs smaller than our high-confidence resolution of 6.3 kb. We identified an additional 243 balanced SVs in the affected siblings. Of these 243 subthreshold SVs, 95.5% followed normal Mendelian inheritance, suggesting that our methods retain a high level of precision below 6.3 kb resolution. The potential for type II errors (false negatives) is the major limitation of exploring SV below this size resolution, as SVs localized entirely within a fragment insert cannot be detected by our methods, and small BCAs can be missed based on stochastic fluctuations in coverage. Therefore, while we do detect many small balanced SV events, this is likely to be a significant underestimate of the total SVs present in a given genome below our highest-confidence resolution.

Sequencing in all probands identified a total of 650 variants larger than the 6.3 kb high-confidence resolution threshold (362 private, 288 polymorphic), resulting in a median of 96 SVs per proband. We found an additional 354 variants (88 private, 266 polymorphic) below our high-confidence resolution, resulting in an additional 111 SVs per proband. In sum, these events correspond to 112 deletions, 13 tandem duplications, 31 balanced inversions, 24 interchromosomal insertions, 21 intrachromosomal insertions, and 6 complex chromosomal rearrangements (CCRs; rearrangements with three or more breakpoints) per average proband (Table 1). The majority of the CCRs were cryptic to karyotyping and CMA, suggesting that CCRs may be a more common phenomenon in the viable human germline than previously appreciated. For genome annotation, we investigated both LoF and gain-of-function (GoF) mutations. Absent molecular characterization of all variants, we adhered to the interpretation of previous exome-sequencing studies that classify a LoF mutation as any variant that disrupts the protein coding sequence. While haploinsufficiency of RNA cannot be confirmed without molecular evaluation, we have demonstrated in our previous BCA sequencing and 16p11.2 transcriptome sequencing studies that dosage compensation of a gene in which a single copy is disrupted is rare.[5,30] We found an average of 25 LoF mutations per proband from all SVs in this study that meet this definition. For the purposes of this study and without access to material for molecular confirmation, we nominally defined GoF in this study as a copy gain of the entire gene, recognizing that an increase in mRNA may not correspond to a molecular gain of function; we observed this 17 times per proband. Importantly, the vast majority of observed LoF mutation (80%) would be cryptic to conventional CMA. In these analyses, we used 100 kb as a conservative estimate of CMA resolution, although many clinical CMA studies have lower resolution on the genomic backbone (e.g., it is approximately 240 kb on the SignatureChip Oligo Solution array) and higher resolution in targeted regions of known genomic disorders.[31] On average, the SVs detected result in 3.8 Mb of rearranged DNA per person (0.1% of haploid genome).

A limitation of WGS as compared to CMA is the ability to detect CNVs in repetitive or misassembled genomic regions, such as segmental duplications or microsatellites. Given that several common genomic disorders are a consequence of segmental-duplication-mediated nonhomologous allelic recombination (NAHR) (e.g., 16p11.2 deletion/duplication, 22q11.2 deletion syndrome[32]), we performed a complementary focal insert-depth analysis to survey genomic imbalance in these regions. Notably, BCAs localized to highly repetitive regions remain undetectable by this and all short-read sequencing methods. We modified a pre-existing WGS read-depth algorithm (cn.MOPS[33]) to accommodate insert coverage rather than read depth. This modified method modeled normalized insert depth alterations simultaneously across all samples in 1, 3, 10, and 30 kb bins. We applied this insert-depth calling approach to 88 genomic regions corresponding to established genomic disorders already being investigated in prenatal diagnostic testing[22] and uncovered a 15q13.3 microduplication flanked by segmental duplications—thus intractable to paired-end clustering alone—and previously identified as a duplication syndrome (Figure 1).[34] We confirmed this variant through microarray analysis of the

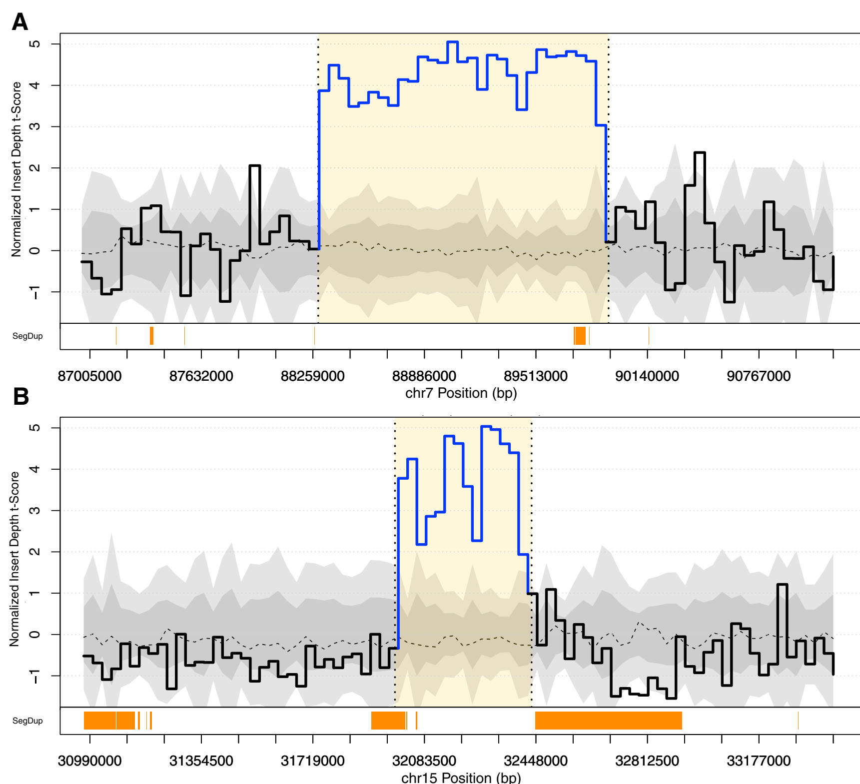

Figure 1. Focal Insert-Depth Comparison across All 33 Libraries Detects CNVs Mediated by Segmental Duplication
Focal insert-depth analysis successfully delineated CNVs in both the presence and absence of paired-end cluster support. Shown are a 1.55 Mb duplication with paired-end clustering support (A) and a 432 kb duplication flanked by segmental duplications (B) (orange) without pair-end support. Insert depth was scaled by chromosome-specific coverage within each library before bin-wise normalization across all libraries, yielding a t-score representative of relative enrichment or depletion in insert depth for each sample as compared to the entire cohort (plotted in thick line above). Blue highlights bins that achieved nominal significance ($p \leq 0.05$). Light yellow shading corresponds to a cn.MOPS duplication call. Gray shaded regions reflect variability of insert depth by position (dashed line: median t-score; dark gray: t-score MAD; light gray: 2 * t-score MAD). Bins were analyzed at multiple sizes (1, 3, 10, and 30 kb).

same case (Birdsuite[35] LOD score 545.6). It was present in a 14-year-old male with average range intellectual functioning, pervasive developmental disorder not otherwise specified (PDD-NOS), and clinically impairing but fluctuating anxiety (anxiety disorder, not otherwise specified). The 15q13.3 microduplication syndrome has been associated with these psychiatric conditions in addition to major depressive disorder (MDD [MIM 608516]), ADHD, obsessive-compulsive disorder (OCD [MIM 164230]), and alcohol abuse (MIM 103780)[34,36] and is not fully penetrant in the population.[36,37]

We found that a significant subset of the duplications observed involve insertion of a copy at a distant site, rather than being tandem duplications (9.9% above our high-confidence threshold in this cohort). One striking example involves two independent duplications on the same chromosome in a 19-year-old with ASD and ID, detectable by CMA, on 15q25 (360.3 kb) and 15q26 (169 kb) that flank a 5.25 Mb inversion, disrupting *AKAP13* (MIM 604686) and *IQGAP1* (MIM 603379) (Figure 2). We investigated duplication signatures at 15q25 and 15q26 in array CNV data from the Database of Genomic Variation[38] (DGV) and an internal CNV database of 33,573 independent cases sourced from clinical diagnostic laboratories and 13,989 controls from GWAS microarrays.[5] We detected 7 individuals with a similar duplication signature as our subject, suggesting that these subjects may harbor the same cryptic inversion.

Interpretation of rare, private mutation remains a significant challenge in human genetic research and clinical diagnostics. To annotate detected genic disruptions in the absence of molecular characterization, we performed two stages of interpretation. In the first, we used the criteria established by Wapner et al.[22] for clinically significant SVs in prenatal diagnosis. This analysis required an SV to meet at least one of the following criteria to be considered pathogenic: (1) SV within or overlapping a targeted region from an established genomic disorder, (2) SV of 1 Mb or greater size not inherited from an unaffected parent, or (3) SV that disrupts a gene implicated in an autosomal-dominant Mendelian or X-linked disorder. Notably, these analyses encompass many of the criteria described for CMA testing of ASD or ID. Based upon these criteria, we found one likely pathogenic SV (the 15q13.3 duplication described above). We did not interpret the 5.3 Mb inversion flanked by "paired" duplications as pathogenic as it was inherited from an unaffected mother, and its molecular significance therefore warrants further investigation.

The second stage of interpretation investigated only LoF variants with disruptions of at least one coding exon in a gene previously associated with NPDs and required a nominally significant CNV burden from cases with neurodevelopmental disorders obtained from clinical diagnostic laboratories compared to controls (total n = 47,562). This convergent genomic approach was previously described in detail (see Talkowski et al.).[5] We PCR-validated all putatively significant SVs in probands and available parental DNA; as above, we required any LoF variant passing these criteria to not be inherited from an unaffected parent. Considering these criteria, we found two additional private SVs to be potential NPD risk factors based upon disruption of a gene previously associated with NPDs and analyses that revealed a nominally significant increased CNV burden

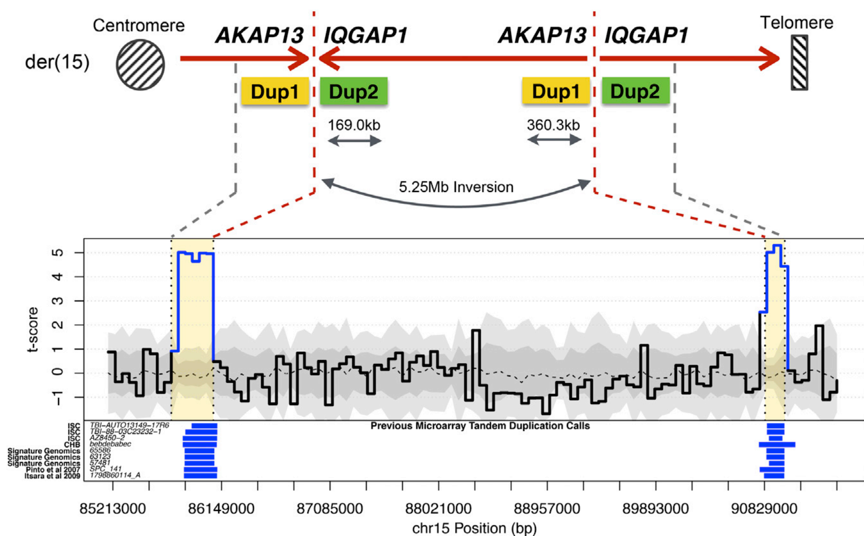

Figure 2. Inversion Signature Marked by Two Tandem Duplications
SV sequencing revealed clusters of inverted read-pairs 5.25 Mb apart (top). The inverted segment contains duplicated regions at each breakpoint (360 kb, 169 kb). Subsequent analysis of insert depth (at bottom) revealed a duplication signature easily detectable by CMA. We investigated a collection of CNV data from 47,562 individuals and identified 7 additional cases who display this same "paired" duplications signature (blue bars), suggesting the presence of an inverted segment between the two CNVs in these cases.

(Table 2). The first was a deletion of *NKD2* (MIM 607852; CNV burden $p = 0.03$) in a 12-year-old female diagnosed with MDD, anxiety disorder not otherwise specified, and borderline intellectual functioning. This variant has been previously associated with ASD.[39] The second was disruption of *UBE2F* (CNV burden $p = 8.3 \times 10^{-4}$) in a 12-year-old male who met diagnostic criteria for multiple conditions (i.e., bipolar disorder with psychotic features, Asperger syndrome, generalized anxiety disorder, and ADHD, variant previously associated with ASD[40]). Both of these youths had a history of psychiatric hospitalization and parents with anxiety and mood diagnoses. Taken together with the 15q13.3 duplication, these represent three likely pathogenic variants among the 29 probands in our study (10.3%), two of which were cryptic to CMA. We also discovered three private variants that represent potentially novel NPD loci and warrant further study. One is a 130.9 kb duplication of chromosome 1p34.2 with a nominally significant increased CNV burden. A second is a 44 kb deletion on chromosome 19 that deletes *ZNF57* and the first three exons of *ZNF77* (MIM 194551). The final variant detected is a complete deletion of exon 15 of *SLC23A2* (MIM 603791), which encodes sodium-ascorbate cotransporter 2, a key cellular transporter of vitamin C[41] that has not previously been implicated in NPDs. Neurons deficient for SLC23A2 exhibit reduced activity and neurite growth,[42] a finding further supported by animal studies that have indicated vitamin C modulates both learning and memory.[43] Another variant detected lies in the promoter region of *RBFOX1* (MIM 605104), which encodes a neuron-specific splicing factor involved with neuronal excitation[44] and has been implicated in a range of NPDs including intellectual disability,[45] ASD,[46,47] and ADHD.[48] This 58 kb deletion was present in all affected members of our four-member family and thus appeared to segregate with NPDs. However, when we scrutinized available convergent genomic data for *RBFOX1* from our clinical diagnostic CNV cohort, we found no evidence that dosage alterations of this locus represent an NPD risk factor ($p \leq 0.995$). Similar results have been previously observed in control cohorts.[49] Of these pathogenic or notable variants detected in our SV screen, four of the six SVs (66.7%) would be cryptic to CMA.

There are limitations to interpretation of these findings. The sample size for this proof-of-concept study is small. Much larger cohorts are required to robustly establish estimates of cryptic SVs and clinical diagnostic yield from such populations. The jumping library method, as with all short-read sequencing approaches, is limited to detection of the portion of the human genome that can be uniquely aligned by short reads; SVs in highly repetitive regions are largely intractable to these methods and are not accounted for in this study. Further, our particular approach applied stringent filtering of SV clusters based on thresholds derived from validated variants in our training set. These data suggest that the method performed well in reducing false-positive SVs, but this was probably at a cost of increased false-negative results. Future studies in much larger samples and access to very deep WGS using long-read technologies will greatly improve upon these initial estimates of cryptic SVs.

Taken together, these data provide insight into the landscape of cryptic SVs in clinically referred youth with severe NPDs, including youth with purely psychiatric conditions and youth with developmental disorders who have not previously been referred for genetic testing. We find that balanced and cryptic SVs represent an important and yet uncharacterized component of the genetic architecture of NPDs and warrant exploration in standard evaluation of severe psychiatric presentations in clinically referred youth. Moreover, our analyses establish that innovative sequencing approaches are capable of detecting pathogenic SVs that are currently detectable by CMA, as well as those that are cryptic to conventional technologies yet represent an important fraction of LoF mutation in a given genome. These data suggest that cost-effective WGS may represent a higher-resolution alternative testing modality to CMA or karyotyping for genome-wide SV assessment in clinical practice.

Table 2. Variants with Nominally Significant Support from Convergent Genomic Data Sets

	Position				Phenotype[a]		CNV Burden		
CNV Type	Chr	Start (Mb)	Stop (Mb)	Size (kb)	Proband	Gene(s) Disrupted	NDD	Control	p Value[b]
Del	5	1.00	1.05	48.4	MDD, ANX-NOS, bIQ	*NKD2*	23	7	0.03
Del	19	2.90	2.94	44.2	BPD, ADHD	*ZNF57, ZNF77*[c]	6	0	0.04
Dup	2	238.87	238.88	10.7	BPD w/ psychotic features, AS, GAD, ADHD	*UBE2F*	17	1	0.001
Dup	1	40.18	40.31	130.9	MDD, ANX-NOS	*BMP8B,*[d] *OXCT2,*[e] *PPIE, TRIT1*	15	3	0.02
Del	20	4.93	4.97	40.8	ID, features of ADHD and ASD	*SLC23A2*	12	1	0.01

[a]Phenotype abbreviations: All diagnoses are based on criteria for the Diagnostic and Statistical Manual of Mental Disorders – Fourth Edition (DSM-IV). MDD, major depressive disorder; ANX-NOS, anxiety disorder, not otherwise specified; bIQ, borderline IQ; BPD, bipolar disorder; AS, Asperger syndrome; GAD, generalized anxiety disorder; ADHD, attention-deficit/hyperactivity disorder; ID, intellectual disability; ASD, autism spectrum disorder.
[b]Fisher's exact test p values are provided for CNV burden.
[c]MIM 602284.
[d]Gene with reported CNV burden; MIM 610289.
[e]MIM 602435.

Supplemental Data

Supplemental Data include one figure and 1 table and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2014.09.005.

Acknowledgments

The genomic studies were supported by funds from the National Institute of Mental Health (MH095867), the Simons Foundation for Autism Research, the Nancy Lurie Marks Family Foundation, the March of Dimes, the Charles Hood Foundation, and the Brain and Behavioral Research Foundation (NARSAD) to M.E.T. The LOGIC study and affiliated personnel were supported by funds from the Stanley Center for Psychiatric Research and the David Judah Foundation to A.E.D. J.W.M. is supported in part by NIMH grant K24MH094614.

Received: August 8, 2014
Accepted: September 11, 2014
Published: October 2, 2014

Web Resources

The URLs for data presented herein are as follows:

Burrows-Wheeler Aligner, http://bio-bwa.sourceforge.net/
cn.MOPS, http://www.bioinf.jku.at/software/cnmops/
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/
R statistical software, http://www.r-project.org/
RefSeq, http://www.ncbi.nlm.nih.gov/RefSeq
UCSC Genome Browser, http://genome.ucsc.edu

Accession Numbers

All sequencing data reported in these 32 subjects have been deposited in the National Database for Autism Research (NDAR accession: http://ndar.nih.gov/study.html?id=357).

References

1. Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., et al. (2007). Strong association of de novo copy number mutations with autism. Science *316*, 445–449.
2. Pinto, D., Pagnamenta, A.T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S., et al. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. Nature *466*, 368–372.
3. 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.
4. 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.
5. Talkowski, M.E., Rosenfeld, J.A., Blumenthal, I., Pillalamarri, V., Chiang, C., Heilbut, A., Ernst, C., Hanscom, C., Rossin, E., Lindgren, A.M., et al. (2012). Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. Cell *149*, 525–537.
6. Miller, D.T., Adam, M.P., Aradhya, S., Biesecker, L.G., Brothman, A.R., Carter, N.P., Church, D.M., Crolla, J.A., Eichler, E.E., Epstein, C.J., et al. (2010). Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am. J. Hum. Genet. *86*, 749–764.
7. Sullivan, P.F., Daly, M.J., and O'Donovan, M. (2012). Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nat. Rev. Genet. *13*, 537–551.
8. Kearney, H.M., South, S.T., Wolff, D.J., Lamb, A., Hamosh, A., and Rao, K.W.; Working Group of the American College of Medical Genetics (2011). American College of Medical Genetics recommendations for the design and performance expectations for clinical genomic copy number microarrays intended for use in the postnatal setting for detection of constitutional abnormalities. Genet. Med. *13*, 676–679.
9. Schaefer, G.B., and Mendelsohn; Professional Practice and Guidelines Committee. (2013). Clinical genetics evaluation in identifying the etiology of autism spectrum disorders: 2013 guideline revisions. Genet. Med. *15*, 399–407.
10. Shea, L., Newschaffer, C.J., Xie, M., Myers, S.M., and Mandell, D.S. (2014). Genetic testing and genetic counseling among

Medicaid-enrolled children with autism spectrum disorder in 2001 and 2007. Hum. Genet. *133*, 111–116.

11. Cook, E.H., Jr., and Scherer, S.W. (2008). Copy-number variations associated with neuropsychiatric conditions. Nature *455*, 919–923.
12. International Schizophrenia Consortium (2008). Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature *455*, 237–241.
13. Nag, A., Bochukova, E.G., Kremeyer, B., Campbell, D.D., Muller, H., Valencia-Duarte, A.V., Cardona, J., Rivas, I.C., Mesa, S.C., Cuartas, M., et al.; Tourette Syndrome Association International Consortium for Genetics (2013). CNV analysis in Tourette syndrome implicates large genomic rearrangements in COL8A1 and NRXN1. PLoS ONE *8*, e59061.
14. Grozeva, D., Kirov, G., Ivanov, D., Jones, I.R., Jones, L., Green, E.K., St Clair, D.M., Young, A.H., Ferrier, N., Farmer, A.E., et al.; Wellcome Trust Case Control Consortium (2010). Rare copy number variants: a point of rarity in genetic risk for bipolar disorder and schizophrenia. Arch. Gen. Psychiatry *67*, 318–327.
15. Malhotra, D., McCarthy, S., Michaelson, J.J., Vacic, V., Burdick, K.E., Yoon, S., Cichon, S., Corvin, A., Gary, S., Gershon, E.S., et al. (2011). High frequencies of de novo CNVs in bipolar disorder and schizophrenia. Neuron *72*, 951–963.
16. Doshi-Velez, F., Ge, Y., and Kohane, I. (2014). Comorbidity clusters in autism spectrum disorders: an electronic health record time-series analysis. Pediatrics *133*, e54–e63.
17. Malhotra, D., and Sebat, J. (2012). CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell *148*, 1223–1241.
18. Lee, S.H., Ripke, S., Neale, B.M., Faraone, S.V., Purcell, S.M., Perlis, R.H., Mowry, B.J., Thapar, A., Goddard, M.E., Witte, J.S., et al.; Cross-Disorder Group of the Psychiatric Genomics Consortium; International Inflammatory Bowel Disease Genetics Consortium (IIBDGC) (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat. Genet. *45*, 984–994.
19. Gershon, E.S., and Alliey-Rodriguez, N. (2013). New ethical issues for genetic counseling in common mental disorders. Am. J. Psychiatry *170*, 968–976.
20. Costain, G., Lionel, A.C., Merico, D., Forsythe, P., Russell, K., Lowther, C., Yuen, T., Husted, J., Stavropoulos, D.J., Speevak, M., et al. (2013). Pathogenic rare copy number variants in community-based schizophrenia suggest a potential role for clinical microarrays. Hum. Mol. Genet. *22*, 4485–4501.
21. Neill, N.J., Torchia, B.S., Bejjani, B.A., Shaffer, L.G., and Ballif, B.C. (2010). Comparative analysis of copy number detection by whole-genome BAC and oligonucleotide array CGH. Mol. Cytogenet. *3*, 11.
22. Wapner, R.J., Martin, C.L., Levy, B., Ballif, B.C., Eng, C.M., Zachary, J.M., Savage, M., Platt, L.D., Saltzman, D., Grobman, W.A., et al. (2012). Chromosomal microarray versus karyotyping for prenatal diagnosis. N. Engl. J. Med. *367*, 2175–2184.
23. Mills, R.E., Walter, K., Stewart, C., Handsaker, R.E., Chen, K., Alkan, C., Abyzov, A., Yoon, S.C., Ye, K., Cheetham, R.K., et al.; 1000 Genomes Project (2011). Mapping copy number variation by population-scale genome sequencing. Nature *470*, 59–65.
24. Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A.; 1000 Genomes Project Consortium (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061–1073.
25. Talkowski, M.E., Ernst, C., Heilbut, A., Chiang, C., Hanscom, C., Lindgren, A., Kirby, A., Liu, S., Muddukrishna, B., Ohsumi, T.K., et al. (2011). Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. Am. J. Hum. Genet. *88*, 469–481.
26. Talkowski, M.E., Ordulu, Z., Pillalamarri, V., Benson, C.B., Blumenthal, I., Connolly, S., Hanscom, C., Hussain, N., Pereira, S., Picker, J., et al. (2012). Clinical diagnosis by whole-genome sequencing of a prenatal sample. N. Engl. J. Med. *367*, 2226–2232.
27. Hanscom, C., and Talkowski, M. (2014). Design of large-insert jumping libraries for structural variant detection using illumina sequencing. Curr. Protoc. Hum. Genet. *80*, 1–9.
28. 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, S1.
29. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.
30. Blumenthal, I., Ragavendran, A., Erdin, S., Klei, L., Sugathan, A., Guide, J.R., Manavalan, P., Zhou, J.Q., Wheeler, V.C., Levin, J.Z., et al. (2014). Transcriptional consequences of 16p11.2 deletion and duplication in mouse cortex and multiplex autism families. Am. J. Hum. Genet. *94*, 870–883.
31. 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.
32. Dittwald, P., Gambin, T., Szafranski, P., Li, J., Amato, S., Divon, M.Y., Rodríguez Rojas, L.X., Elton, L.E., Scott, D.A., Schaaf, C.P., et al. (2013). NAHR-mediated copy-number variants in a clinical population: mechanistic insights into both genomic disorders and Mendelizing traits. Genome Res. *23*, 1395–1409.
33. Klambauer, G., Schwarzbauer, K., Mayr, A., Clevert, D.A., Mitterecker, A., Bodenhofer, U., and Hochreiter, S. (2012). cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. Nucleic Acids Res. *40*, e69.
34. Miller, D.T., Shen, Y., Weiss, L.A., Korn, J., Anselm, I., Bridgemohan, C., Cox, G.F., Dickinson, H., Gentile, J., Harris, D.J., et al. (2009). Microdeletion/duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders. J. Med. Genet. *46*, 242–248.
35. Korn, J.M., Kuruvilla, F.G., McCarroll, S.A., Wysoker, A., Nemesh, J., Cawley, S., Hubbell, E., Veitch, J., Collins, P.J., Darvishi, K., et al. (2008). Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nat. Genet. *40*, 1253–1260.
36. Szafranski, P., Schaaf, C.P., Person, R.E., Gibson, I.B., Xia, Z., Mahadevan, S., Wiszniewska, J., Bacino, C.A., Lalani, S., Potocki, L., et al. (2010). Structures and molecular mechanisms for common 15q13.3 microduplications involving CHRNA7: benign or pathological? Hum. Mutat. *31*, 840–850.
37. van Bon, B.W., Mefford, H.C., Menten, B., Koolen, D.A., Sharp, A.J., Nillesen, W.M., Innis, J.W., de Ravel, T.J., Mercer, C.L., Fichera, M., et al. (2009). Further delineation of the

15q13 microdeletion and duplication syndromes: a clinical spectrum varying from non-pathogenic to a severe outcome. J. Med. Genet. *46*, 511–523.

38. MacDonald, J.R., Ziman, R., Yuen, R.K., Feuk, L., and Scherer, S.W. (2014). The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res. *42* (Database issue), D986–D992.
39. Noh, H.J., Ponting, C.P., Boulding, H.C., Meader, S., Betancur, C., Buxbaum, J.D., Pinto, D., Marshall, C.R., Lionel, A.C., Scherer, S.W., and Webber, C. (2013). Network topologies and convergent aetiologies arising from deletions and duplications observed in individuals with autism. PLoS Genet. *9*, e1003523.
40. Piven, J., Vieland, V.J., Parlier, M., Thompson, A., O'Conner, I., Woodbury-Smith, M., Huang, Y., Walters, K.A., Fernandez, B., and Szatmari, P. (2013). A molecular genetic study of autism and related phenotypes in extended pedigrees. J. Neurodev. Disord. *5*, 30.
41. Savini, I., Rossi, A., Pierro, C., Avigliano, L., and Catani, M.V. (2008). SVCT1 and SVCT2: key proteins for vitamin C uptake. Amino Acids *34*, 347–355.
42. Qiu, S., Li, L., Weeber, E.J., and May, J.M. (2007). Ascorbate transport by primary cultured neurons and its role in neuronal function and protection against excitotoxicity. J. Neurosci. Res. *85*, 1046–1056.
43. Harrison, F.E., and May, J.M. (2009). Vitamin C function in the brain: vital role of the ascorbate transporter SVCT2. Free Radic. Biol. Med. *46*, 719–730.
44. Gehman, L.T., Stoilov, P., Maguire, J., Damianov, A., Lin, C.H., Shiue, L., Ares, M., Jr., Mody, I., and Black, D.L. (2011). The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. Nat. Genet. *43*, 706–711.
45. Bhalla, K., Phillips, H.A., Crawford, J., McKenzie, O.L., Mulley, J.C., Eyre, H., Gardner, A.E., Kremmidiotis, G., and Callen, D.F. (2004). The de novo chromosome 16 translocations of two patients with abnormal phenotypes (mental retardation and epilepsy) disrupt the A2BP1 gene. J. Hum. Genet. *49*, 308–311.
46. Davis, L.K., Maltman, N., Mosconi, M.W., Macmillan, C., Schmitt, L., Moore, K., Francis, S.M., Jacob, S., Sweeney, J.A., and Cook, E.H. (2012). Rare inherited A2BP1 deletion in a proband with autism and developmental hemiparesis. Am. J. Med. Genet. A. *158A*, 1654–1661.
47. Zhao, W.W. (2013). Intragenic deletion of RBFOX1 associated with neurodevelopmental/neuropsychiatric disorders and possibly other clinical presentations. Mol. Cytogenet. *6*, 26.
48. Elia, J., Gai, X., Xie, H.M., Perin, J.C., Geiger, E., Glessner, J.T., D'arcy, M., deBerardinis, R., Frackelton, E., Kim, C., et al. (2010). Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. Mol. Psychiatry *15*, 637–646.
49. Kamien, B., Lionel, A.C., Bain, N., Scherer, S.W., and Hunter, M. (2014). Outfoxed by RBFOX1-a caution about ascertainment bias. Am. J. Med. Genet. A. *164A*, 1411–1418.