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.

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