

# Critical exon indexing improves clinical interpretation of copy number variants in neurodevelopmental disorders

E. Robert Wassman, MD,\* Karen S. Ho, PhD,\* Diana Bertrand, MS, Kyle W. Davis, ScM, Megan M. Martin, MS, Stephanie Page, BS, Andreas Peiffer, MD, PhD, Aparna Prasad, PhD, Moises A. Serrano, PhD, Hope Twede, BS, Rena Vanzo, MS, Stephen W. Scherer, PhD, Mohammed Uddin, PhD, and Charles H. Hensel, PhD\*

**Correspondence**  
Dr. Wassman  
bwassman@lineagen.com

*Neurol Genet* 2019;5:e378. doi:10.1212/NXG.0000000000000378

## Abstract

### Objective

To evaluate a new tool to aid interpretation of copy number variants (CNVs) in individuals with neurodevelopmental disabilities.

### Methods

Critical exon indexing (CEI) was used to identify genes with critical exons (CEGs) from clinically reported CNVs, which may contribute to neurodevelopmental disorders (NDDs). The 742 pathogenic CNVs and 1,363 variants of unknown significance (VUS) identified by chromosomal microarray analysis in 5,487 individuals with NDDs were subjected to CEI to identify CEGs. CEGs identified in a subsequent random series of VUS were evaluated for relevance to CNV interpretation.

### Results

CEI identified a total of 2,492 unique CEGs in pathogenic CNVs and 953 in VUS compared with 259 CEGs in 6,965 CNVs from 873 controls. These differences are highly significant ( $p < 0.00001$ ) whether compared as frequency, average, or normalized by CNV size. Twenty-one percent of VUS CEGs were not represented in Online Mendelian Inheritance in Man, highlighting limitations of existing resources for identifying potentially impactful genes within CNVs. CEGs were highly correlated with other indices and known pathways of relevance. Separately, 136 random VUS reports were reevaluated, and 76% of CEGs had not been commented on. In multiple cases, further investigation yielded additional relevant literature aiding interpretation. As one specific example, we discuss *GTF2I* as a CEG, which likely alters interpretation of several reported duplication VUS in the Williams-Beuren region.

### Conclusions

Application of CEI to CNVs in individuals with NDDs can identify genes of potential clinical relevance, aid laboratories in effectively searching the clinical literature, and support the clinical reporting of poorly annotated VUS.

---

\*These authors contributed equally to this work.

From Lineagen Inc. (E.R.W., K.S.H., D.B., K.W.D., M.M.M., S.P., A. Peiffer, A. Prasad, M.A.S., H.T., R.V., C.H.H.); Life Designs Ventures (E.R.W.), Park City, UT; Department of Pediatrics (K.S.H., A. Peiffer), University of Utah; The Centre for Applied Genomics (S.W.S., M.U.), The Hospital for Sick Children, Toronto, Ontario, Canada; Program in Genetics and Genome Biology (S.W.S.), The Hospital for Sick Children; McLaughlin Centre (S.W.S.), University of Toronto, Toronto, Ontario, Canada; and Department of Molecular Genetics (S.W.S.), University of Toronto, Toronto, Ontario, Canada.

Go to [Neurology.org/NG](http://Neurology.org/NG) for full disclosures. Funding information is provided at the end of the article.

The Article Processing Charge was funded by Lineagen, Inc.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND), which permits downloading and sharing the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

## Glossary

**ASD** = autism spectrum disorder; **CEG** = critical exon gene; **CEI** = critical exon indexing; **CMA** = chromosomal microarray analysis; **CNV** = copy number variant; **DD** = developmental delay; **FMRP** = fragile X mental retardation protein; **ID** = intellectual disability; **IPA** = ingenuity pathway analysis; **Mbp** = megabase pairs; **MCA** = multiple congenital anomaly; **NDD** = neurodevelopmental disorder; **OMIM** = Online Mendelian Inheritance in Man; **pLI** = probability of loss of function index; **VUS** = variants of unknown significance; **WBS** = Williams-Beuren syndrome.

Neurodevelopmental disorders (NDDs) including developmental delay (DD), intellectual disability (ID), and autism spectrum disorder (ASD) are often caused by copy number variants (CNVs) in the genome.<sup>1–6</sup> Chromosomal microarray analysis (CMA) for the detection of CNVs is the first-tier clinical diagnostic test for the evaluation of NDDs.<sup>2–6</sup> CMA frequently identifies variants of uncertain significance (VUS), which limit clinical utility in such cases. Interpretation of VUS is relevant to clinical care of these children and important to their families.<sup>7,8</sup>

Clinical application of an ultra-high-resolution whole-genome chromosomal microarray optimized for the detection of genetic changes associated with a variety of NDDs provides an increased diagnostic yield, particularly due to careful attention to VUS of potential clinical relevance, constituting up to 67% of reported CNVs.<sup>9</sup> These VUS could not be definitively linked to patients' neurodevelopmental traits despite extensive analysis of the extant literature using standard approaches, which include analysis of relevant gene content.<sup>10</sup>

The recently described<sup>11–14</sup> identification of “critical” exons with high levels of brain expression and low mutation burden recognizes genes (e.g., *OTUD7A*)<sup>14</sup> that might be NDD candidate genes. This was adapted to create a categorical classification tool that focuses analysis on specific genes among many contained in CNVs, which are likely to contribute to interpretation. This tool retrospectively identified many genes that have relevant clinical literature, but had not been specifically detailed in the clinical reports previously issued. Prospective application of this tool may aid in the classification of VUS and affect diagnosis and management of a wide range of NDDs.

## Methods

### Patient data

CMA results were reviewed from 1,602 cases with clinically relevant findings from a previously described retrospective series of 5,487 consecutive patients with indications for testing for broader NDDs that includes ASD, DD, ID, multiple congenital anomaly (MCA), and seizures or combinations of these.<sup>9</sup> Medical records and indications for testing provided by referring providers were used to collect phenotype information. Subsequent to this series' analysis for critical exon genes (CEGs) prevalence, a separate group of 136 randomly selected reports of VUS was assessed to determine the impact critical exon indexing (CEI) might have on searching for the relevant literature and revised clinical reporting.

### Standard protocol approvals, registrations, and patient consents

Written informed consent was obtained for all participants' data in this study. Deidentified data were collected under Western Institutional Review Board Protocol #20162032.

### Chromosomal microarray and reporting

Between September 2012 and December 2015, 5,487 genomic DNAs were extracted with either a buccal swab or blood Genra Puregene kit (Qiagen, Redwood City, CA), fragmented, labeled, and hybridized to a previously described microarray optimized for NDDs<sup>9</sup> (FSDX; Lineagen, Salt Lake City, UT) as specified by the manufacturer (Affymetrix).<sup>15</sup> Washed arrays were scanned and raw data files analyzed as CYCHP files using Chromosome Analysis Suite version 2.0.1 (Affymetrix, Santa Clara, CA) with reference genome from the Genome Reference Consortium: human build 37/human genome 19 (GRCh37/hg19). American Board of Medical Genetics and Genomics-certified cytogeneticists used established interpretation standards.<sup>10</sup> In brief, CNVs were classified pathogenic if multiple peer-reviewed publications, representing at least 2 families indicate clinical features result from sensitivity of the region or gene(s) to dosage.<sup>10</sup> If there was preliminary but insufficient evidence for pathogenicity, CNVs were classified as variants of unknown significance (VUS). CNVs are classified as benign and not reported if the region had been well characterized as a common variant in the general population by multiple peer-reviewed publications or by curated databases (e.g., the Database of Genomic Variants<sup>16</sup>).

### Critical exon indexing

Identification of CEGs by CEI was adapted from previously described methods<sup>11,12</sup> The mutation burden of rare (<0.05)<sup>11</sup> missense and loss-of-function mutations for each exon was calculated from the 1000 Genomes Project (internationalgenome.org). The BrainSpan Atlas provided spatio-temporal exon level RNA-seq expression from 388 tissue samples from 16 brain regions of prenatal, early childhood, and adult brains from 32 postmortem donors.<sup>17</sup> Our cutoffs, which were modified for high stringency clinical translation, considered an exon critical if expressed at >90th percentile throughout all 3 developmental stages, and has a mutation burden <10th percentile (normalized by exon size). “Critical exon” enrichment between CNVs classified as pathogenic, VUS, or normal was tested with the Fisher exact test, Welch *t* test, or  $\chi^2$  analysis depending on the nature of the variables.

Online Mendelian Inheritance in Man (OMIM, omim.org) and PubMed (ncbi.nlm.nih.gov/pubmed/) were used to search for published data and disorder associations for each CEG.

Ingenuity Pathway Analysis (IPA; Qiagen, Redwood City, CA)<sup>18</sup> was used to determine the association of CEGs within CNVs with known pathways based on literature-curated ontologies for genes, their functions, and their interactions. The number of CEG genes overlapping the list of curated genes identified in each pathway by IPA compared with CEGs not in a given pathway was tested statistically using a 1-tailed  $\chi^2$  test for each pathway.

### Data availability

The authors take full responsibility for conduct of the research, its data, analyses, and interpretation. We have full access to the data and rights to publish any and all data, separate and apart from sponsor guidance.

## Results

CMA results on 5,487 individuals with NDD (1,558 females and 3,929 males) on a custom microarray in a Clinical Laboratories Improvement Act certified laboratory from September 2012 to December 2015 were previously reported.<sup>9</sup> A total of 2,105 rare CNVs (742 pathogenic and 1363 VUS) were clinically reported based on American College of Medical Genetics and Genomics Guidelines<sup>9,10</sup> (table 1).

Analysis of these CNVs for CEGs yielded 2,492 unique CEGs in pathogenic CNVs and 953 in VUS, including 469 genes

found in both pathogenic CNVs and VUS alike. The presence of at least 1 CEG in a CNV was significantly higher in pathogenic CNVs (69.9%) compared with either VUS (39.6%) ( $p < 0.00001$ ) or a control population<sup>13</sup> (11.4%) ( $p < 0.00001$ ). On average, there were significantly more CEGs per pathogenic variant (12) compared with VUS (1) ( $p < 0.00001$ ), and both contained significantly more than control's CNVs (0.17) ( $p < 0.00001$ ) (table 1). Because pathogenic CNVs are typically larger than VUS,<sup>9</sup> these data were normalized for CNV size, resulting in slightly more similar CEG/megabase pairs (Mbp) in pathogenic and VUS with 2.5 and 2.2 CEG/Mbp, respectively, but still significantly higher for pathogenic ( $p < 0.0006$ ) (table 1). Both are significantly greater than the 0.6 CEG/Mbp observed in the control population<sup>13</sup> ( $p < 0.00001$ ) (table 1).

We searched for overlaps between our list of CEGs present in clinically reported CNVs and (1) genes containing de novo variants identified in individuals with ASD<sup>19,20</sup> and (2) fragile X (fragile X mental retardation protein [FMRP]) target genes.<sup>21</sup> ASD candidate genes were highly represented in the list of CEGs identified in our CNVs (1,023 unique out of 2,228 candidates,  $p = 0.0039$ , Fisher exact test). Of the observed CEGs from both pathogenic and VUS, 38% overlap with known FMRP targets ( $p < 0.00001$ , Fisher exact test).

Of the observed CEGs, 18% interact significantly in pathways with gene families that have key neurodevelopmental functions ( $p$  values  $< 1.15E-05$  to  $4.87E-12$ , 1-tailed  $\chi^2$ ). (table 2).

When the CEGs in our cohort are cross-referenced to the OMIM database (omim.org) commonly used to aid

**Table 1** Observed distribution of CEGs in CNVs from neurodevelopmental disorder and controls

	Pathogenic	VUS	Normal control <sup>a</sup>
Total CNVs	742	1,363	6,965
CNVs with at least 1 CEG (% of total)	519 (69.9%)	540 (39.6%)	665 (11.4%)
<sup>b</sup> <i>p</i> value (% of total) vs pathogenic		$p < 0.00001$	$p < 0.00001$
<sup>b</sup> <i>p</i> value (% of total) vs VUS	—		$p < 0.00001$
Total number of CEGs identified	9,168	1,335	798
Unique CEGs identified	2,492	953	259
Average number CEG/CNV	12	1	0.11
<sup>c</sup> <i>p</i> value vs pathogenic		$p < 0.00001$	$p < 0.00001$
<sup>c</sup> <i>p</i> value vs VUS	—		$p < 0.00001$
Average CEG/Mbp	2.5	2.2	0.7
<sup>d</sup> <i>p</i> value vs pathogenic		$p < 0.0006$	$p < 0.00001$
<sup>d</sup> <i>p</i> value vs VUS	—		$p < 0.00001$

Abbreviations: CNV = copy number variant; VUS = variants of unknown significance; Mbp = megabase pairs.

<sup>a</sup> Ontario Population Genomics Platform (OPGP)<sup>3</sup>.

<sup>b</sup> Fisher exact test.

<sup>c</sup> Welch *t* test.

<sup>d</sup> Chi-square test.

interpretations, 6.3% (158/2,492) of the unique CEGs from pathogenic CNVs and 21.0% (200/953) of those from VUS were not yet represented in OMIM.

### CEGs in VUS reports

These population-based observations of enrichment of CEGs in NDDs and their enrichment in relevant pathways led us to evaluate the utility of translating CEI to previous clinical reports of VUS, i.e., CNVs where board-certified analysts using external and internal clinical databases and established clinical guidelines<sup>10</sup> failed to find clearly pathogenic genes. We selected a random subset of 136 VUS reports to evaluate the potential impact of CEI on clinical reports and compared how gene content was discussed. Of these reports, 21% of VUS CEGs were not represented in OMIM, highlighting the limitations of existing resources for identifying potentially impactful genes in CNVs. Although some well-known CEGs were mentioned in reports as being potentially causal, in 76% of clinical reports (103/136), at least 1 CEG was not described specifically in the report. Targeted PubMed searches on 49 CEGs not previously commented on in clinical reports yielded 25 articles (51%) relevant for NDDs, which might aid the interpretation of these VUS.

### CEGs within subregions of large pathogenic regions

VUS can occur as smaller CNVs overlapping or completely within known pathogenic CNVs. We hypothesized that CEGs within such VUS may indicate a critical region for pathogenicity in both the VUS and the larger known pathogenic CNV. As an example, several cases with small duplications interpreted clinically as VUS overlap the 7q11.23 Williams-Beuren syndrome (WBS) region, but from patients with only ASD clinically (table 3). One duplication VUS (table 3, patient 20) affects part of a single CEG, *GTF2I*, which encodes a multifunctional transcription factor.<sup>22,23</sup> It also slightly overlaps a noncoding gene, *LOC10192694.3*, and *NCF1*, a gene expressed primarily in neutrophils. Two brothers (table 3, patients 23 and 24) shared a slightly larger overlapping VUS that completely duplicates *GTF2I* and partially including another CEG, *CLIP2*, and a nonCEG, *GTF2IRD1*. Patient 21 (table 3) has a VUS that affects *GTF2I*, *GTF2IRD1*, and *LOC10192694.3*. Another individual (table 3, patient 22) has a partial duplication of *GTF2I* that extends 1.9 Mbp proximally and includes several additional genes. Patients 25 and 26 have overlapping VUS of intermediate length. CNV coordinates and testing indications of each patient are shown in table 3

**Table 2** Top pathways represented in clinical copy number variants with statistical enrichment observed by segregation based on CEGs

Top pathways	p Value	Overlap with pathway members % (fraction)
<b>CE genes in all reported findings</b>		
Mitochondrial dysfunction	4.87E-12	34.5 (57/165)
Regulation of eIF4 and p70S6K signaling	1.22E-11	35.9 (51/142)
EIF2 signaling	1.05E-10	32.6 (56/172)
mTOR signaling	3.90E-10	31.3 (57/182)
Huntington disease signaling	4.70E-10	29.2 (66/226)
<b>CE genes in pathogenic findings</b>		
Huntington disease signaling	2.30E-08	174 (42/241)
HIF1 signaling	3.12E-06	20.0 (23/115)
Breast cancer regulation by stathmin1	3.51E-06	16.3 (33/203)
Tight junction signaling	1.04E-05	16.8 (28/167)
CDK5 signaling	1.15E-05	20.2 (20/99)
<b>CE genes in VUS</b>		
Regulation of eIF4 and p70S6K signaling	4.32E-09	21.0 (33/157)
Huntington disease signaling	1.37E-08	17.4 (42/241)
EIF2 signaling	2.80E-08	18.6 (36/194)
p70S6K signaling	4.32E-08	21.4 (28/131)
mTOR signaling	5.51E-08	18.1 (36/199)

Abbreviations: CE = critical exon; VOS = variants of unknown significance.

**Table 3** Observed variants of unknown significance copy number variants overlapping the WBS region

Patient	ISCN	Size	CEGs	Other genes included	Clinical symptoms
20	7q11.23 (74,143,059–74,209,864) x3	67kb	<i>GTF2I</i>	<i>NCF1</i> and <i>LOC10192694.3</i>	ASD, seizures, and staring spells
21	7q11.23 (74,009,922–74,136,633) x3	127kb	<i>GTF2I</i>	<i>GTF2IRD1</i> and <i>LOC10192694.3</i>	ASD, speech and language deficits, and delayed milestones
22	7q11.23 (74,143,059–76,007,380) x3	1.86Mb	<i>GTF2I</i> , <i>RHBDD2</i> , <i>POR</i> , <i>TMEM120A</i> , <i>MDH2</i> , <i>HSPB1</i> , <i>YWHAG</i>	<i>LOC10192694.3</i> , <i>NCF1</i> , <i>GTF2IRD2</i> , <i>WBSCR16</i> , <i>GTF2IRD2B</i> , <i>GATSL2</i> , <i>TRIM73</i> , <i>POM121C</i> , <i>SPDYE5</i> , <i>HIP1</i> , <i>CCL26</i> , <i>CCL24</i> , <i>POR</i> , <i>MIR4651</i> , <i>SNORA14A</i> , <i>STYXL1</i> , and <i>SRRM3</i>	ASD and unique physical features (prominent forehead, congenital facial compression, asymmetric nipples, and misplaced right ear)
23 <sup>a</sup>	7q11.23 (73,802,871–74,177,758) x3	375kb	<i>GTF2I</i> , <i>CLIP2</i>	<i>GTF2IRD1</i> and <i>LOC10192694.3</i>	ASD
24 <sup>a</sup>	7q11.23 (73,802,871–74,177,899) x3	375kb	<i>GTF2I</i> , <i>CLIP2</i>	<i>GTF2IRD1</i> and <i>LOC10192694.3</i>	ASD
25	7q11.23 (74,142,189–74,400,363) x3	258kb	<i>GTF2I</i>	<i>LOC10192694.3</i> , <i>NCF1</i> , <i>GTF2IRD2</i> , <i>STAG3L2</i> , and <i>PMS2P5</i>	Delayed motor skills, speech and language deficits, repetitive movements, and possible ASD features
26	7q11.23 (74,165,273–74,628,840) x3	464kb	<i>GTF2I</i>	<i>NCF1</i> , <i>GTF2IRD2</i> , <i>STAG3L2</i> , <i>PMS2P5</i> , <i>RCC1L</i> , <i>GTF2IRD2B</i> , <i>NCF1C</i> , <i>GTF2IP1</i> , and <i>GTF2IP4</i>	ASD, ADHD, temper tantrums, and ID

Abbreviations: ADHD = attention deficit hyperactivity disorder; ASD = autism spectrum disorder; ID = intellectual disability; ISCA = International Standards for Cytogenomic Arrays Consortium; WBS = Williams-Beuren syndrome.

<sup>a</sup> Nonidentical siblings.

and figure. The only gene contained, wholly or partially, in all of these VUS from individuals with ASD is *GTF2I*, which has been previously associated with neurobehavioral and cognitive features of WBS.<sup>22,23</sup>

## Discussion

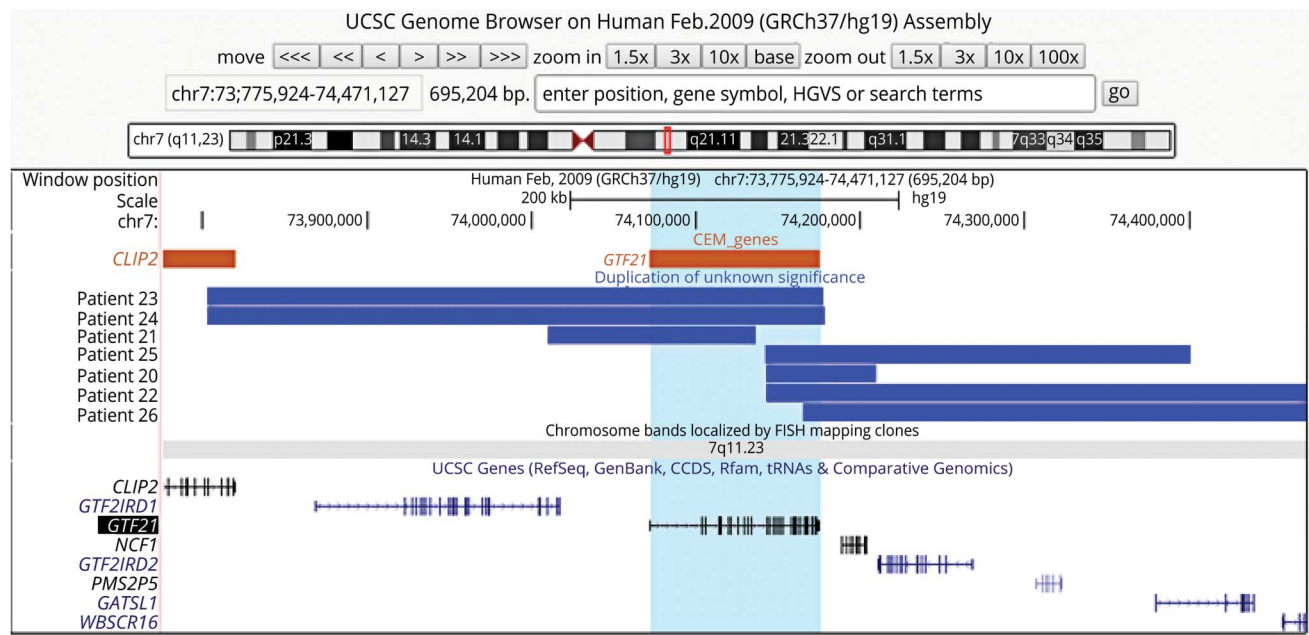
The presence of critical genomic content within CNVs is central to clinical interpretation.<sup>10</sup> The recognition that gene expression and conservation at the exon level opens new understanding of candidate genes also showed CEGs correlate significantly with patients with ASD and DD compared with their control siblings.<sup>11,12</sup> CEI led to the identification of the *OTUD7A* gene as a key phenotype regulator for 15q13.3 microdeletion syndrome<sup>14</sup> and has been cited in multiple publications across a variety of neurodevelopmental and behavioral disabilities including epilepsies, ASD, cerebral palsy, schizophrenia, and obsessive-compulsive disorder.<sup>24–30</sup> The application of CEI by clinical laboratories may assist clarifying pathogenicity of CNVs and expands where gene content is factored into interpretation and discussion of results with referring clinicians and families.

VUS remain a major challenge in clinical interpretation. VUS tend to be ultra-rare, relatively small, and encompass fewer genes than known pathogenic deletions and duplications.

Improved CMAs make analytical detection of very small CNVs more robust,<sup>9,31,32</sup> increasing the number of VUS requiring analysis. Interpretation is dependent on the coincidence of multiple rare cases and their successful alignment from the literature and databases such as OMIM. Without compelling clinical case experience compared with control individuals, functional evidence, or unambiguous genotype-phenotype correlations, inferences of pathogenicity are difficult and laborious processes for the standard clinical laboratory.

This observational study on the prevalence of CEGs identified by CEI was conducted on CNVs found in consecutive individuals presenting with NDDs and/or MCA by CMA. In total, 2,976 unique CEGs were identified within reported CNVs. It is not particularly surprising that the CEG content of pathogenic CNVs is significantly higher than in VUS or CNVs in controls, given the literature on CEI to date,<sup>11,12,14,24–30</sup> (table 1); however, it is of interest that whether looked at in terms of CNVs with at least one CEG ( $p < 0.00001$ ), average number per CNV ( $p < 0.00001$ ), or CEG/Mbp basis ( $p < 0.00001$ ) in VUS is also significantly higher than in controls, too. It is perhaps noteworthy that when adjusted for size differences in VUSs, the CEGs/Mbp are more similar with 2.5 vs 2.2 per Mbp ( $p = 0.0006$ ) for pathogenic and VUS, respectively, than when compared in terms of overall prevalence ( $p < 0.0001$ ). This supports

**Figure** Map of VUS overlapping CEG GTF2I from individuals with ASD phenotypes



possible clinical relevance in some VUS, as does the overlapping presence of 469 specific CEGs (18.8% of 2,492) in both pathogenic and VUS CNVs. We suggest that this supports their importance in the VUS potential for pathogenicity and may therefore aid interpretation.

CEI on a subsequent random set of clinical reports issued by the same clinical laboratory showed that 76% (103/136) failed to mention at least one of these potentially important genes. Of CEGs from VUS, 21% (200/953) were not represented in OMIM at the time but may well be in the future as the clinical literature evolves. Further targeted PubMed searches of VUS CEGs yielded articles relevant for NDDs in many. This gap between clinical reporting of rare events and genes that segregate in terms of frequency between affected and control populations highlights the value of this CEI to assessment of VUS.

Of the observed CEGs, 18% interact in pathways with gene families that have key neurodevelopmental functions ( $p$  values:  $<1.15E-05$  to  $4.87E-12$ ,  $\chi^2$  test) (table 2). Of the observed CEGs from both pathogenic and VUS CNVs, 38% overlap with known FMRP targets ( $p < 0.00001$ ), which are believed to be associated with NDDs, and some are now targets of directed drug development.

Comparison to genes containing de novo variants identified in individuals with ASD<sup>19</sup> showed a statistically significant representation of the CEGs in the clinically reported CNVs ( $p = 0.0039$ ). As the research literature suggests, these overlapping genes could represent causal genes lying within the larger pathogenic CNVs.<sup>14,24–30</sup>

Clinical genomic literature does not always include discussion of VUS, and clinical reporting is inconsistent within and across laboratories as guidelines<sup>10</sup> focus on clinically definitive findings. In the context of genomic testing of affected individuals, the a priori significance of VUS is different than in the predictive setting, e.g., cancer predisposition, and may be relevant to both clinician's management and family perceptions.<sup>7,8</sup> Over time, further clinical experience will eventually allow for reclassification of many VUS as either benign or pathogenic; however, the rarity of these sometimes "private" genetic findings delays their clinical utility. As an independent and nonoverlapping resource, CEI can focus the efforts of cytogeneticists and others interpreting test results on specific genes within CNVs rather than the affected region in total or the comprehensive list of the genes affected, and this can aid in the determination of pathogenicity of a VUS.

Another strategy using a model trained on nonpathogenic deletions in the general population has been applied to attempt to estimate the effect size of pathogenic deletions and correlate quantitatively the effects of haploinsufficiency on Intelligence Quotient (IQ). This approach may also help predict the impact of VUS in NDD; however, studies to date have exclusively evaluated deletions and largely focus on size alone<sup>33</sup> rather than specific genomic content as prescribed by interpretive guidelines.<sup>10</sup> The databases available to clinical laboratories to interpret genomic content, including CNVs, are constantly improving and expanding. Notably, the Exome Aggregation Consortium created a metric based on the probability of loss of function (pLI) of the protein products of genes in that database for similar filtering of candidate disease-causing variants. A pLI  $\geq 0.9$  is considered "high" and serves as a cutoff to separate

genes of sufficient length into loss of function intolerant,<sup>34</sup> and we are studying the application of “high pLI” now as a non-redundant resource along with CEI for clinical interpretation.

Finally, we describe a specific example in this series where several clinically reported VUS overlap a single well-known CEG, *GTF2I*. This gene is located within the Williams-Beuren 7q11.23 microdeletion-duplication syndrome region, which contains nearly 100 other genes.<sup>22</sup> However, our VUS cases show a remarkable phenotype-genotype consistency (ASD) distinct from these well-described syndromes. Dosage of the *GTF2I* gene has been proposed to be responsible for the differing neurobehavioral phenotypes of WBS and the 7q11.23 microduplication syndrome.<sup>22,23,35–39</sup> Individuals with partial WBS deletions that do not include *GTF2I* are reported without ID, but typically have the WBS-associated visual-spatial cognition impairment.<sup>39</sup> *GTF2I* mediates oxytocin reactivity, and individuals with WBS have elevated oxytocin levels, whereas unaffected individuals with different polymorphisms in *GTF2I* manifest varying oxytocin levels, which correspond with their self-reported levels of social anxiety.<sup>36,37</sup> *GTF2I* encodes a multifunctional transcription factor, a gene class previously associated with NDDs,<sup>35–39</sup> and specific single nucleotide polymorphisms are associated with ASD.<sup>38</sup> *GTF2I* is the only gene contained, wholly or partially, in all of these VUS, suggesting that its partial or complete duplication might be responsible for ASD features noted. The additional cases identified here by application of CEI merit reassessment of these CNVs and correlating phenotypes for pathogenicity.

CEI pinpointed clinically relevant genes in VUS and streamlined the identification of citations relevant to the patients’ indications for testing. Such relevant literature adds to VUS reports and keys in on potentially critical genes that may define pathogenicity. CEI also has the potential to identify networks and pathways of relevant genes shared within one CNV or across several CNVs in a single individual or in multiple individuals. By narrowing down the list of genes within the breakpoints of CNVs, CEI could aid assessment of “multiple-hit” interactions. This could target potential drug development for some individuals with NDDs who share common causal networks and pathways, which otherwise might not be evident.

Of course, there are limitations and caveats to this approach. CEI as described defines CEGs as those that contain exons with high brain expression or genes with at least 1 splice isoform that is highly expressed in the brain. Genes that do not display this exact expression profile, i.e., ubiquitously expressed in the body but still perform an essential brain development function, would not be identified as a CEG despite potential importance to the etiology of NDDs, e.g., *MECP2* (Rett syndrome). These could represent false-negative results; however, the CEI tool is always used in conjunction with other methods, and well-described genes should not in fact be missed in analysis. Similarly, some CNVs common in the general population do also harbor CEGs, such as *BTRC* and *ZNF90*, both CEGs, but apparently commonly

duplicated or deleted. Whether these CEGs are simply not dose sensitive or whether they only affect NDD etiology when combined with “second hits” (single nucleotide variants in other genes, rare unreported CNVs, or structural variants not detected by CMA) remains to be determined. The cross-referencing of databases should identify these as exceptions and not cause a false-positive misinterpretation.<sup>12,13,16</sup>

As whole-exome/genome sequencing becomes more commonly used clinically,<sup>40</sup> and the analytical capability of these tools to accurately detect CNVs improves<sup>41</sup> the challenge of interpreting VUS and apparent “multi-hit” situations grows and should similarly be aided by CEI. Further assessment of this tool is therefore warranted, while cautiously applying its demonstrated strengths in enhancing the clinical interpretation of rare CNVs.

## Acknowledgment

The authors thank Dr. Patricia Mowery-Rushton, Dr. Colleen Bilancia, Dr. Ankita Patel, and the entire Lineagen clinical team for helpful discussions. SWS is supported by The Centre for Applied Genomics, the University of Toronto McLaughlin Centre, and the GlaxoSmithKline-CIHR Endowed Chair in Genome Sciences at SickKids and the University of Toronto. This work has been presented in part at the NSGC and ASHG annual meetings in 2017 and 2016.

## Study funding

Lineagen, Inc.; S.W. Scherer is supported by The Centre for Applied Genomics, the University of Toronto McLaughlin Centre, and the GlaxoSmithKline-CIHR Endowed Chair in Genome Sciences at SickKids and the University of Toronto.

## Disclosure

Disclosures available: [Neurology.org/NG](http://Neurology.org/NG).

## Publication history

Received by *Neurology: Genetics* August 19, 2018. Accepted in final form October 11, 2019.

## Appendix Authors

Name	Location	Role	Contribution
<b>E. Robert Wassman, MD</b>	Lineagen, Inc., dba Life Designs Ventures	Author	Oversaw all aspects of translation, study design, and conduct and primary author
<b>Karen S. Ho, PhD</b>	Lineagen, Inc., Clene Nanomedicine	Author	Directly transferred methodology to clinical database and analyzed and drafted the initial summary of work
<b>Diana Bertrand, MS</b>	Lineagen, Inc.	Author	Researched and coordinated additional clinical cases and drafted clinical synopses

Continued

## Appendix (continued)

Name	Location	Role	Contribution
<b>Kyle W. Davis, ScM</b>	Lineagen, Inc.	Author	Statistical analyses and revision review
<b>Megan M. Martin, MS</b>	Lineagen, Inc.	Author	Applied method to clinical case work and coordinated comparative reporting information
<b>Stephanie Page, BS</b>	Lineagen, Inc.	Author	Assembled clinical reporting data and maintained data
<b>Andreas Peiffer, MD, PhD</b>	Lineagen, Inc., University of Utah	Author	Contributed to interpretation of analyses and drafting
<b>Aparna Prasad, PhD</b>	Lineagen, Inc.	Author	Contributed to interpretation of analyses and drafting and statistical analyses of data
<b>Moises A. Serrano, PhD</b>	Lineagen, Inc.	Author	Contributed to interpretation of test results and impact of CEI
<b>Hope Twede, BS</b>	Lineagen, Inc.	Author	Assembled clinical reporting data and maintained data and statistical analyses of data
<b>Rena Vanzo, MS</b>	Lineagen, Inc.	Author	Contributed to drafting and oversight of clinical reports, data access, and integrity
<b>Stephen W. Scherer, PhD</b>	The Centre for Applied Genomics, the University of Toronto SickKids	Author	Study concept development, critical input to data interpretation, and contributed to drafting
<b>Mohammed Uddin, PhD</b>	The Centre for Applied Genomics, the University of Toronto Mohammed Bin Rashid University of Medicine and Health Sciences	Author	Initial method developer and tech transfer consultant, critical input to data interpretation, and contributed to drafting
<b>Charles H. Hensel, PhD</b>	Lineagen, Inc.	Author	Critical input to data interpretation, contributed to drafting, reviewed all data analysis in preparation for publication, and developer of proprietary CMA

## References

- Iafate AJ, Feuk L, Rivera MN, et al. Detection of large-scale variation in the human genome. *Nat Genet* 2004;36:949–951.
- Miller DT, Adam MP, Aradhya S, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet* 2010;86:749–764.
- Schaefer GB, Mendelsohn NJ. Clinical genetics evaluation in identifying the etiology of autism spectrum disorders: 2013 guideline revisions. *Genet Med* 2013;15:399–407.

- Volkmar F, Siegel M, Woodbury-Smith M, King B, McCracken J, State M. Practice parameter for the assessment and treatment of children and adolescents with autism spectrum disorder. *J Am Acad Child Adolesc Psychiatry* 2014;53:237–257.
- Moeschler JB, Shevell M, Saul RA, et al. Comprehensive evaluation of the child with intellectual disability or global developmental delays. *Pediatrics* 2014;134:e903–18.
- Kearney HM, South ST, Wol DJ, Lamb A, Hamosh A, Rao KW. 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* 2011;13:676–679.
- Rei M, Bernhardt BA, Mulchandani S, et al. “What does it mean?”: uncertainties in understanding results of chromosomal microarray testing. *Genet Med* 2012;14:250–258.
- Jez S, Martin M, South S, Vanzo R, Rothwell E. Variants of unknown significance on chromosomal microarray analysis: parental perspectives. *J Community Genet* 2015;6:343–349.
- Ho KS, Twede H, Vanzo R, et al. Real-world clinical performance of an ultra-high resolution chromosomal microarray optimized for neurodevelopmental disorders. *Biomed Res Int* 2016;2016:3284534.
- South ST, Lee C, Lamb AN, Higgins AW, Kearney HM; Working Group for the American College of Medical Genetics and Genomics Laboratory Quality Assurance Committee. ACMG standards and guidelines for constitutional cytogenomic microarray analysis, including postnatal and prenatal applications: revision 2013. *Genet Med* 2013;15:901–909.
- Uddin M, Tammimies K, Pellicchia G, et al. Brain-expressed exons under purifying selection are enriched for de novo mutations in autism spectrum disorder. *Nat Genet* 2014;46:742–747.
- Uddin M, Pellicchia G, Thiruvahindrapuram B, et al. Indexing effects of copy number variation on genes involved in developmental delay. *Sci Rep* 2016;6:28663.
- Uddin M, Thiruvahindrapuram B, Walker S, et al. A high-resolution copy-number variation resource for clinical and population genetics. *Genet Med* 2015;17:747–752.
- Uddin M, Unda BK, Kwan V, et al. OTUD7A regulates neurodevelopmental phenotypes in the 15q13.3 microdeletion syndrome. *Am J Hum Genet* 2018;102:278–295.
- Mason-Suares H, Kim W, Grimmett L, et al. Density matters: comparison of array platforms for detection of copy-number variation and copy-neutral abnormalities. *Genet Med* 2013;15:706–712.
- MacDonald JR, Ziman R, Yuen RK, et al. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res* 2014;42(Database issue):D986–D992.
- Sunkin SM, Ng L, Lau C, et al. Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res* 2013;41(Database issue):D996–D1008.
- Krämer A, Green J, Pollard J, Tugendreich S. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* 2014;30:523–530.
- Yuen RKC, Merico D, Bookman M, et al. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat Neurosci* 2017;20:602–611.
- Iossifov I, O’Roak BJ, Sanders SJ, et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 2014;515:216–221.
- Darnell JC, Van Driesche SJ, Zhang C. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* 2011;146:247–261.
- Sakurai T, Dorr NP, Takahashi N, McInnes LA, Elder GA, Buxbaum JD. Haploinsufficiency of GTF2I, a gene deleted in Williams syndrome, leads to increases in social interactions. *Autism Res* 2011;4:28–39.
- Crespi BJ, Hurd PL. Cognitive-behavioral phenotypes of Williams syndrome are associated with genetic variation in the GTF2I gene, in a healthy population. *BMC Neurosci* 2014;15:127.
- Lal D, Ruppert AK, Trucks H, et al. Burden analysis of rare microdeletions suggests a strong impact of neurodevelopmental genes in genetic generalised epilepsies. *PLoS Genet* 2015;11:e1005226.
- Gazzellone MJ, Zhou X, Lionel AC, et al. Copy number variation in Han Chinese individuals with autism spectrum disorder. *J Neurodev Disord* 2014;6:34.
- Liu Y, Zhao D, Dong R, et al. De novo exon 1 deletion of AUTS2 gene in a patient with autism spectrum disorder and developmental delay: a case report and a brief literature review. *Am J Med Genet A* 2015;167:1381–1385.
- Lowther C, Speevak M, Armour CM, et al. Molecular characterization of NRXN1 deletions from 19,263 clinical microarray cases identifies exons important for neurodevelopmental disease expression. *Genet Med* 2017;19:53.
- Zarrei M, Fehlings DL, Mawjee K, et al. De novo and rare inherited copy-number variations in the hemiplegic form of cerebral palsy. *Genet Med* 2018;20:172–180.
- Wang Q, Li M, Yang Z, et al. Increased co-expression of genes harboring the damaging de novo mutations in Chinese schizophrenic patients during prenatal development. *Scientific Rep* 2015;5:18209.
- Gazzellone MJ, Zarrei M, Burton CL, et al. Uncovering obsessive-compulsive disorder risk genes in a pediatric cohort by high-resolution analysis of copy number variation. *J Neurodev Disord* 2016;8:36.
- Prasad A, Merico D, Thiruvahindrapuram B, et al. A discovery resource of rare copy number variations in individuals with autism spectrum disorder. *G3 (Bethesda)* 2012;2:1665–1685.
- Tucker T, Montpetit A, Chai D, et al. Comparison of genome-wide array genomic hybridization platforms for the detection of copy number variants in idiopathic mental retardation. *BMC Med Genomics* 2011;4:25–35.
- Huguet G, Schramm C, Douard E, et al. Measuring and estimating the effect sizes of copy number variants on general intelligence in community-based samples. *JAMA Psychiatry* 2018;75:447–457.



34. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536:285–291.
35. Borralleras C, Sahun I, Pérez-Jurado LA, Campuzano V. Intracisternal GTF2I gene therapy ameliorates deficits in cognition and synaptic plasticity of a mouse model of williams-Beuren syndrome. *Mol Ther* 2015;23:1691–1699.
36. Dai L, Carter CS, Ying J, Bellugi U, Pournajafi-Nazarloo H, Korenberg JR. Oxytocin and vasopressin are dysregulated in Williams syndrome, a genetic disorder affecting social behavior. *PLoS One* 2012;7:e38513.
37. Procyshyn TL, Spence J, Read S, Watson NV, Crespi BJ. The Williams syndrome prosociality gene *GTF2I* mediates oxytocin reactivity and social anxiety in a healthy population. *Biol Lett* 2017;13:20170051.
38. Malenfant P, Liu X, Hudson ML, et al. Association of GTF2I in the Williams-Beuren syndrome critical region with autism spectrum disorders. *J Autism Dev Disord* 2012;42:1459–1469.
39. Shirai Y, Li W, Suzuki T. Role of splice variants of GTF2I, a transcription factor localizing at postsynaptic sites, and its relation to neuropsychiatric diseases. *Int J Mol Sci* 2017;18:E411.
40. Srivastava S, Love-Nichols JA, Dies KA, et al. Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. *Genet Med* 2019;21:2413–2421.
41. Trost B, Walker S, Wang Z. A comprehensive workflow for read depth-based identification of copy-number variation from whole-genome sequence data. *Am J Hum Genet* 2018;102:142–155.