Copy Number Variant Analysis from Exome Data in 349 Patients with Epileptic Encephalopathy

Epilepsy Phenome/Genome Project &
Epi4K Consortium*

Infantile spasms (IS) and Lennox–Gastaut syndrome (LGS) are epileptic encephalopathies characterized by early onset, intractable seizures, and poor developmental outcomes. De novo sequence mutations and copy number variants (CNVs) are causative in a subset of cases. We used exome sequence data in 349 trios with IS or LGS to identify putative de novo CNVs. We confirm 18 de novo CNVs in 17 patients (4.8%), 10 of which are likely pathogenic, giving a firm genetic diagnosis for 2.9% of patients. Confirmation of exome-predicted CNVs by array-based methods is still required due to false-positive rates of prediction algorithms. Our exome-based results are consistent with recent array-based studies in similar cohorts and highlight novel candidate genes for IS and LGS.

ANN NEUROL 2015;78:323–328

The epileptic encephalopathies (EEs) are a devastating group of epilepsies in which epileptic activity and seizures contribute to cognitive impairment or regression.[1] Most EEs begin in infancy or early childhood and are associated with poor developmental outcome. Although the cause is unknown in the majority of cases, recent studies confirm that de novo mutations and copy number variants (CNVs) play an important role.[2,3] We recently reported exome sequencing data in 264 parent–proband trios with infantile spasms (n = 149) or Lennox–Gastaut syndrome (LGS; n = 115) without syndromic features or magnetic resonance imaging (MRI) abnormalities from the Epilepsy Phenome/Genome Project (EPGP) cohort, identifying likely pathogenic, de novo sequence changes in >10% of patients.[2] Here we report results of copy number analysis derived from the exome data of this cohort and 85 additional patients to further elucidate the genetic architecture of these paradigmatic EEs. Our exome-based CNV calling yields similar results to array-based studies for confirmed, de novo, likely pathogenic CNVs.

Patients and Methods

Samples

Probands and family members were collected as part of the EPGP cohort (Supplementary Table 1) as described previously[2,4] with approval by site-specific institutional review boards; 1,047 individuals comprising 349 parent–child trios were included in the present analysis. Of these, 264 were previously analyzed for de novo single nucleotide variants (SNVs).[2] Prior clinical CNV testing included chromosome microarray and/or karyotype analysis in 131 of 349 patients (38% of cohort). Detailed inclusion criteria have been published[4]; briefly, participants were required to have electroencephalographic findings consistent with LGS (slow or disorganized background, and slow spike and wave < 2.7Hz or generalized paroxysmal fast activity) or infantile spasms (IS; hypsarrhythmia, hypsarrhythmia variant, or electrodecremental discharge).[4] Exclusion criteria included evidence of a known genetic syndrome or chromosome abnormality. Extensive phenotype analysis of patients enrolled in the study are published elsewhere[5] (and Madou et al, manuscript in preparation). All available clinical records were re-reviewed for those patients found to have a de novo CNV, and evidence of syndromic features was often noted upon reexamination of the medical records.

CNV Calling and Validation

CNVs were detected by analyzing exome data using the CoNIFER pipeline, a depth-of-coverage–based algorithm using the conifer-tools package, which implements DNACopy.[6,7] Briefly, exome reads were mapped to exons and 300bp flanking sequence using mrsFAST. RPKM (mapped reads per kilobase per million reads) values were calculated, and the first 30 singular value decomposition (SVD) components were removed to minimize systematic noise and bias. The following filtering criteria were applied: CNVs of 3 to 5 probes average SVD-transformed signal > 1; CNVs of ≥6 probes, average signal > 0.5. CNVs >50% in repetitive or duplicated genomic space were removed. CNV calls were manually curated, and curated calls were compared to control CNV data sets to filter out common CNVs present in >1% of the general population. Control CNV data sets included (1) CNV calls from the ARIC (Atherosclerosis Risk in Communities) study (n = 11,305)

TABLE 1. De Novo CNVs in 349 Trios

Trio	CNV	Size	Candidate or Known Epilepsy Genes or Known Disease Association	De Novo SNV Calls from Exome[2]	Validation Platform	Gene(s) Enriched in CNVs Found in Patients with Neurodevelopmental Phenotypes[a]	Age at Onset	Seizure Types[b]
Likely pathogenic CNVs								
fx	2q24 dup	7.5Mb	*SCN2A, SCN1A*	*SMG9* (mis), *EPHB1* (synon)	CGH	*SCN1A, SCN2A*	7 mo	IS
iq	2q24 del	296kb	*SCN1A*	None	CGH, SNP	*SCN1A*	<1 yr	GTC, aA
hj	5p15 del	3.8Mb	*SEMA5A, CTNND2*	*SDCBP2* (mis)	CGH	*TAS2R1, FAM173B, CCT5, MTRR*	6 mo	FS, focal, GTC, aA, SE
cy	7q11 del	11.4Mb	*MAGI2, YWHAG, HIP1*	*ZNF12* (UTR), *FPGT-TNNI3K* (mis), *FAM50A*	CGH, SNP	*HIP1*	3 mo	IS, aA
aia	9p ter del	8.7Mb	9p deletion syndrome	None	SNP	*DMRT2, DMRT3*	5 mo	IS
iz	14q23 del	585kb	*GPHN*	*HRG* (mis), *PCDHB13* (mis)	CGH	—[c]	2.5 yr	FS+SE, T, drop
eh	15q11 dup	5.0Mb	15q11q13 dup syndrome; *GABRB3*	*PAQR8* (synon)	CGH	*GABRB3*	2 wk	IS, multiple other
ag	15q11 dup[d]	12.0Mb	15q11q13 dup syndrome; *GABRB3*	*MLL4* (mis)	CGH, karyo	*GABRB3*	8 mo	IS
gq^	15q11 dup	8.4Mb	15q11q13 dup syndrome; *GABRB3*	None	CGH, SNP	*GABRB3*	8 mo	GTC, T, atonic
fu	t(15;16)[d]	1.8Mb del, 16.3Mb dup	Large unbalanced translocation	None	CGH, karyo	*SNRPA1, FANCA*	8 mo	IS
CNVs of uncertain clinical significance								
ig	1p22 dup	140kb	1 gene: *ZNF644*	*IQSEC1* (mis)	CGH, SNP	—[c]	2 yr	A, GTC, M, T, drop
ad	1q21 dup	249kb	TAR region dup	*NFE2L1* (mis)	SNP	*LIX1L*	8 mo	IS

TABLE 1: Continued

Trio	CNV	Size	Candidate or Known Epilepsy Genes or Known Disease Association	De Novo SNV Calls from Exome[2]	Validation Platform	Gene(s) Enriched in CNVs Found in Patients with Neurodevelopmental Phenotypes[a]	Age at Onset	Seizure Types[b]
aib	2q37 del	154kb	4 genes: *PPP1R7, HDLBP, ANO7, SEPT2*	*CXXC11* (synon)	SNP	*PPP1R7*	5 mo	IS, T
gc	7q22 del	622kb	15 genes in region	*NR1H2* (mis)	CGH	*MUC17, MYL10, TRIM56*	8 mo	IS
ahp	7q31 dup	94kb	2 genes: *CPED1, ING3*	*ADAMSL4* (mis), *PPP6R2* (mis)	SNP	—[c]	7 mo	IS
le	8p23 del	140kb	2 genes: *MCPH1, AGTP2*	*DACH2* (mis)	CGH, SNP	*ANGPT2*	3 y 10 mo	GTC, drop, T, M, A, aA
bda	17q12 del	1.5Mb	15 genes in region	None	h.c.		8 mo	IS, M, SE, GTC

Additional information is available in Supplementary Table 7.
[a]Genes listed represent those with mean probability value < 0.05 for known disease gene(s) in region or peak probability value < 0.05 for novel regions as described by Cooper and colleagues.[21] See Supplementary Table 5 for details.
[b]Seizure types include all reported; first type listed was the initial seizure type.
[c]No gene within region with $p < 0.05$.
[d]Upon review of records, diagnosis made prior to enrollment.
A = absence; aA = atypical absence; CGH = comparative genome hybridization; CNV = copy number variant; FS = febrile seizures; GTC = generalized tonic clonic; h.c. = high-confidence CNV call by CoNIFER; IS = infantile spasms; M = myoclonic; SE = status epilepticus; SNP = single nucleotide polymorphism; SNV = single nucleotide variant; T = tonic; UTR = untranslated region.

analyzed using Affymetrix AFFY_6.0 single nucleotide polymorphism (SNP) microarray; and (2) CNV calls from the National Heart, Lung, and Blood Institute GO Exome Sequencing Project (ESP, n = 2,972) from CoNIFER analysis of exome sequence data. CoNIFER-predicted de novo CNVs and a subset of predicted large (>500kb), inherited CNVs were validated using oligonucleotide (Agilent Technologies, Santa Clara, CA) and/or SNP (HumanCore 12v1; Illumina, San Diego, CA; n = 295,393 probes) microarray. De novo CNVs were considered pathogenic if the CNV (or largely overlapping CNV) was previously associated with epilepsy or related neurodevelopmental disorders or contained a known epilepsy gene.

Results

CNV Discovery and Validation

As CNV detection from exome data is still an emerging technique, we initially performed comprehensive validation studies in 43 probands to estimate our overall validation rate for CoNIFER calls in this data set. We validated 53 of 80 (66%) predicted inherited CNVs, consistent with our previous studies[8] (Supplementary Table 2). Twenty-four were paternally inherited and 29 were maternally inherited, with a size range of 5.2kb to 8.8Mb (mean = 377kb). For the same 43 probands, we validated 5 of 21 (24%) predicted de novo CNVs (Supplementary Table 3). The lower validation rate is not unexpected, given that any false-positive call in a proband will appear to be de novo, whereas inherited CNV predictions are supported by the same predicted CNV in 2 individuals (proband and 1 parent).

As the majority of causative CNVs in this cohort were expected to be de novo, we targeted the remainder of our validation studies to predicted de novo CNVs. We confirmed a total of 18 de novo CNVs in 17 patients (Table 1). The de novo CNVs ranged in size from 94kb to 16Mb and involved 1 to 163 genes. Notably, none of these 17 individuals had clearly pathogenic de novo SNVs by exome sequencing. In 10 patients, the de novo CNV(s) was likely pathogenic based on size, previous association with epilepsy, or gene content.[9] One pathogenic CNV (15q11 dup) recurred in 3 cases. In 7 patients, the de novo CNV was of uncertain clinical significance (see Table 1).

Because whole genome array comparative genome hybridization was used to validate de novo CNVs, we also confirmed a large number of inherited CNVs across the cohort. We confirmed 69 inherited CNVs in 54 individuals ranging from 5.2kb to 8.8Mb (mean = 305kb; Supplementary Table 4). Eight individuals (2.3%) each had an inherited CNV >500kb; 4 (1.2%) of these were >1Mb (Table 2). We also looked specifically for inherited CNVs within 3 recurrent deletion regions that have been previously associated with risk for epilepsy[10,11]:

15q11.2, 15q13.3, and 16p13.11. Two patients had inherited deletions of 15q11.2 that may have contributed to their phenotype; another 2 patients each had a small, inherited duplication within the 16p13.11 region of uncertain significance. Aside from the large 15q11-q13 duplications described above, there were no additional CNVs within the 15q13.3 region. Although de novo CNVs are more likely to be pathogenic,[12] it is possible that 1 or more of the inherited CNVs in our cohort was contributory. Three individuals with rare inherited CNVs had a pathogenic SNV, and 1 had a de novo 15q11 duplication, making it less likely that the inherited CNV was causative (see Table 2).

There were 540 unique genes within the 18 de novo CNV regions in our cohort (Supplementary Table 5), 3 of which were known EE genes: SCN1A, SCN2A, and GABRB3. All 5 individuals with CNVs involving these genes had phenotypes consistent with those described for the CNVs they carried (Supplementary Table 7). Eight additional genes (GLIS3, KIAA1324L, NIPA1, PLCG2, RCL1, RFX3, SPG7, YWHAG) within de novo CNV regions were also found to have a de novo sequence variant by trio exome sequencing in the same cohort (see Supplementary Table 5, Allen et al,[2] and unpublished data); these cannot be regarded as confirmed EE genes, but finding both a de novo SNV and a de novo CNV involving each of them suggests that followup in a larger cohort is warranted. In addition, 3 and 30 genes within de novo CNVs were found to have de novo mutations by trio exome sequencing in intellectual disability[13,14] and autism,[15–18] respectively; these genes may warrant follow-up given the overlapping genetic susceptibility of these disorders.

Discussion

We detected CNVs from exome sequencing data in 349 trios from patients with IS or LGS. We confirmed 18 de novo CNVs in 17 of 349 probands (4.8%), providing a definitive diagnosis in 2.9% of patients and a possible explanation for another 2.0%. Notably, 38% of the current cohort had already undergone karyotype and/or chromosome microarray testing prior to enrollment in the study and had not arrived at a diagnosis through clinical testing. Evaluation of patients without prior screening may result in a higher yield; we observed a de novo CNV in 5.6% of the 218 participants in our cohort without previous clinical testing. These results are similar to our prior studies in a broader spectrum of EE, where 4.1% had a definitely pathogenic CNV,[3] and to our recently reported findings in a large clinically ascertained cohort with a broad range of epilepsy diagnoses, where 5% of cases had a causative CNV.[19]

TABLE 2. Selected Inherited CNVs

Trio	CNV (inheritance)	Size, kb	No. of Genes; Possible EE Candidates	Causative d.n. SNV?	Validation Platform
Large [>500kb] inherited CNVs					
jp	2p22 dup (paternal)	620	3 genes; BIRC6, TTC27, LTBP1	No	SNP
ip	17q dup (paternal)	737	13 genes	No	CGH
ad	10q21 del (maternal)	858	1 gene; PCDH15	No	SNP
jg	4p16 dup (maternal)	885	5 genes	WDR45 frameshift	SNP
ki	7q11 dup (paternal)	1,000	9 genes	DNM1 missense	SNP
dg	Xp22 del (paternal)	1,900	8 genes	ALG13 missense	h.c.
bj	Xp22 dup (maternal)	2,000	9 genes	No	h.c.
gq	1q31 dup (paternal)	8,800	23 genes	No; de novo 15q11 dup	CGH, SNP
Recurrent CNV regions previously associated with epilepsy					
j	16p13 dup (paternal)	30	NTAN1, PDXDC1 (16p13.11)	No	h.c.
r	16p13 dup (maternal)	58	ABCC1, ABCC6 (16p13.11)	No	h.c.
d	15q11.2 del (maternal)	213	NIPA2, CYFIP1	No	h.c.
in	15q11.2 del (paternal)	213	NIPA2, CYFIP1	No	SNP

CGH = comparative genome hybridization; CNV = copy number variant; d.n. = de novo; EE = epileptic encephalopathy; h.c. = high-confidence CNV call by CoNIFER; SNP = single nucleotide polymorphism; SNV = single nucleotide variant.

Three individuals each had a de novo duplication consistent with 15q11q13 duplication syndrome, characterized by hypotonia, seizures, developmental delay, and behavior problems. A late onset LGS phenotype has been described in some patients. Other de novo CNVs in our cohort that have been previously associated with epilepsy include 7q11 deletion, 9p terminal deletion, 2q24 duplication, and SCN1A deletion. One patient harbored a de novo intragenic deletion of the GPHN gene, which encodes a protein that is responsible for the clustering of glycine and γ-aminobutyric acid receptors at inhibitory synapses. Inherited or de novo deletions involving GPHN were recently described in 6 patients with autism, schizophrenia, or seizures.[20] The deletion in our patient is the largest of those described and also involves the FAM17D and MIPP5 genes.

Comparison of the genes within de novo CNV regions in our cohort to those in which at least 1 other patient in this cohort had a de novo sequence variant identifies several novel candidate genes that deserve follow-up in a larger cohort. Furthermore, several patients harbor de novo CNVs involving only 1 to 4 genes. Although these CNVs are of uncertain significance, identification of de novo SNVs in the same genes encompassed by certain CNVs would support the finding that these CNVs are related to disease.

In the large EPGP cohort of IS and LGS patients, the addition of this CNV data to the de novo SNV findings shows that a definitive genetic diagnosis can be reached in >15% of cases for which there was previously no known cause. As whole exome sequencing is becoming widely used, one might ask whether CNV data can be efficiently and reliably extracted in a clinical setting, thus bypassing the need for array-based CNV assays. Our experience, especially as shown by the false-positive rate, suggests that array-based technologies are currently still required. A logical clinical approach to a patient with IS or LGS of unknown etiology should include a chromosome microarray for patients with epilepsy and additional findings such as abnormal MRI, developmental delays, or dysmorphic features, followed by an epilepsy-focused targeted gene panel and then whole exome sequencing in

cases that remain undiagnosed. As prediction algorithms improve, exome and eventually whole genome sequencing will provide a genetic diagnosis in an even greater proportion of patients in the clinical setting, improving medical management and genetic counseling in this patient population.

Acknowledgment

Authorship

Contributing authors are listed below. A full list of authors with affiliations is listed in Supplementary Table 8. Author Contributions are listed in Supplementary Table 9.

Bassel Abou-Khalil, Brian K. Alldredge, Andrew S. Allen, Eva Andermann, Frederick Andermann, Dina Amrom, Jocelyn F. Bautista, Samuel F. Berkovic, Alex Boro, Gregory Cascino, Bradley P. Coe, Damian Consalvo, Joseph Cook, Patrick Cossette, Patricia Crumrine, Norman Delanty, Orrin Devinsky, Dennis Dlugos, Evan E. Eichler, Michael P. Epstein, Miguel Fiol, Nathan B. Fountain, Jacqueline French, Daniel Friedman, Eric B. Geller, Tracy Glauser, Simon Glynn, David B. Goldstein, Sheryl R. Haut, Jean Hayward, Erin L. Heinzen, Sandra L. Helmers, Michael R. Johnson, Sucheta Joshi, Andres Kanner, Heidi E. Kirsch, Robert C. Knowlton, Eric H. Kossoff, Nik Krumm, Rachel Kuperman, Ruben Kuzniecky, Daniel H. Lowenstein, Anthony G. Marson, Shannon M. McGuire, Heather C. Mefford, Paul V. Motika, Ben Nelson, Sahar Esmaeeli Nieh, Edward J. Novotny, Terence J. O'Brien, Ruth Ottman, Juliann M. Paolicchi, Jack Parent, Kristen Park, Stephen Petrou, Slavé Petrovski, Annapurna Poduri, Archana Raja, Elizabeth K. Ruzzo, Ingrid E. Scheffer, Renée A. Shellhaas, Elliott Sherr, Jerry J. Shih, Rani Singh, Joseph Sirven, Michael C. Smith, Joe Sullivan, Liu Lin Thio, Anu Venkat, Eileen P.G. Vining, Gretchen K. Von Allmen, Judith L. Weisenberg, Peter Widdess-Walsh, Melodie R. Winawer

Potential Conflicts of Interest

Nothing to report.

References

1. Berg AT, Berkovic SF, Brodie MJ, et al. Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005–2009. Epilepsia 2010;51:676–685.

2. Allen AS, Berkovic SF, Cossette P, et al. De novo mutations in epileptic encephalopathies. Nature 2013;501:217–221.

3. Mefford HC, Yendle SC, Hsu C, et al. Rare copy number variants are an important cause of epileptic encephalopathies. Ann Neurol 2011;70:974–985.

4. Abou-Khalil B, Alldredge B, Bautista J, et al. The epilepsy phenome/genome project. Clin Trials 2013;10:568–586.

5. Widdess-Walsh P, Dlugos D, Fahlstrom R, et al. Lennox-Gastaut syndrome of unknown cause: phenotypic characteristics of patients in the Epilepsy Phenome/Genome Project. Epilepsia 2013;54:1898–1904.

6. Krumm N, Sudmant PH, Ko A, et al. Copy number variation detection and genotyping from exome sequence data. Genome Res 2012;22:1525–1532.

7. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics 2007; 23:657–663.

8. Krumm N, O'Roak BJ, Karakoc E, et al. Transmission disequilibrium of small CNVs in simplex autism. Am J Hum Genet 2013;93:595–606.

9. 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.

10. de Kovel CG, Trucks H, Helbig I, et al. Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies. Brain 2010;133(pt 1):23–32.

11. Helbig I, Mefford HC, Sharp AJ, et al. 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. Nat Genet 2009; 41:160–162.

12. Vulto-van Silfhout AT, Hehir-Kwa JY, van Bon BW, et al. Clinical significance of de novo and inherited copy-number variation. Hum Mutat 2013;34:1679–1687.

13. de Ligt J. Diagnostic exome sequencing in persons with severe intellectual disability. N Engl J Med 2012;367:1921–1929.

14. Rauch A, Wieczorek D, Graf E, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet 2012;380: 1674–1682.

15. Neale BM, Kou Y, Liu L, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 2012;485: 242–245.

16. O'Roak BJ, Vives L, Girirajan S, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 2012;485:246–250.

17. Sanders SJ, Murtha MT, Gupta AR, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 2012;485:237–241.

18. Iossifov I, Ronemus M, Levy D, et al. De novo gene disruptions in children on the autistic spectrum. Neuron 2012;74:285–299.

19. Olson H, Shen Y, Avallone J, et al. Copy number variation plays an important role in clinical epilepsy. Ann Neurol 2014;75:943–958.

20. Lionel AC, Vaags AK, Sato D, et al. Rare exonic deletions implicate the synaptic organizer Gephyrin (GPHN) in risk for autism, schizophrenia and seizures. Hum Mol Genet 2013;22: 2055–2066.

21. Cooper GM, Coe BP, Girirajan S, et al. A copy number variation morbidity map of developmental delay. Nat Genet 2011;43:838–846.

University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Author/s:
Abou-Khalil, B; Alldredge, BK; Allen, AS; Andermann, E; Andermann, F; Amrom, D; Bautista, JF; Berkovic, SF; Boro, A; Cascino, G; Coe, BP; Consalvo, D; Cook, J; Cossette, P; Crumrine, P; Delanty, N; Devinsky, O; Dlugos, D; Eichler, EE; Epstein, MP; Fiol, M; Fountain, NB; French, J; Friedman, D; Geller, EB; Glauser, T; Glynn, S; Goldstein, DB; Haut, SR; Hayward, J; Heinzen, EL; Helmers, SL; Johnson, MR; Joshi, S; Kanner, A; Kirsch, HE; Knowlton, RC; Kossoff, EH; Krumm, N; Kuperman, R; Kuzniecky, R; Lowenstein, DH; Marson, AG; McGuire, SM; Mefford, HC; Motika, PV; Nelson, B; Nieh, SE; Novotny, EJ; O'Brien, TJ; Ottman, R; Paolicchi, JM; Parent, J; Park, K; Petrou, S; Petrovski, S; Poduri, A; Raja, A; Ruzzo, EK; Scheffer, IE; Shellhaas, RA; Sherr, E; Shih, JJ; Singh, R; Sirven, J; Smith, MC; Sullivan, J; Liu, LT; Venkat, A; Vining, EPG; Von Allmen, GK; Weisenberg, JL; Widdess-Walsh, P; Winawer, MR

Title:
Copy number variant analysis from exome data in 349 patients with epileptic encephalopathy

Date:
2015-08-01

Persistent Link:
http://hdl.handle.net/11343/257719

File Description:
Published version