RESEARCH ARTICLE OPEN ACCESS

Distinguishing Loss-of-Function and Gain-of-Function *SCN8A* Variants Using a Random Forest Classification Model Trained on Clinical Features

Joshua B. Hack, BS, Kyle Horning, PhD, Denise M. Juroske Short, PhD, John M. Schreiber, MD, Joseph C. Watkins, PhD, and Michael F. Hammer, PhD

Correspondence
Dr. Hammer
mfh@email.arizona.edu

Neurol Genet 2023;9:e200060. doi:10.1212/NXG.0000000000200060

Abstract

Background and Objectives

Pathogenic variants at the voltage-gated sodium channel gene, *SCN8A*, are associated with a wide spectrum of clinical disease outcomes. A critical challenge for neurologists is to determine whether patients carry gain-of-function (GOF) or loss-of-function (LOF) variants to guide treatment decisions, yet in vitro studies to infer channel function are often not feasible in the clinic. In this study, we develop a predictive modeling approach to classify variants based on clinical features present at initial diagnosis.

Methods

We performed an exhaustive search for individuals deemed to carry SCN8A GOF and LOF variants by means of in vitro studies in heterologous cell systems, or because the variant was classified as truncating, and recorded clinical features. This resulted in a total of 69 LOF variants: 34 missense and 35 truncating variants, including 9 nonsense, 13 frameshift, 6 splice site, 6 indels, and 1 large deletion. We then assembled a truth set of variants with known functional effects, excluding individuals carrying variants at other loci associated with epilepsy. We then trained a predictive model based on random forest using this truth set of 45 LOF variants and 45 GOF variants randomly selected from a set of variants tested by in vitro methods.

Results

Phenotypic categories assigned to individuals correlated strongly with GOF or LOF variants. All patients with GOF variants experienced early-onset seizures (mean age at onset = 4.5 ± 3.1 months) while only 64.4% patients with LOF variants had seizures, most of which were late-onset absence seizures (mean age at onset = 40.0 ± 38.1 months). With high accuracy (95.4%), our model including 5 key clinical features classified individuals with GOF and LOF variants into 2 distinct cohorts differing in age at seizure onset, development of seizures, seizure type, intellectual disability, and developmental and epileptic encephalopathy.

Discussion

The results support the hypothesis that patients with *SCN8A* GOF and LOF variants represent distinct clinical phenotypes. The clinical model developed in this study has great utility because it provides a rapid and highly accurate platform for predicting the functional class of patient variants during *SCN8A* diagnosis, which can aid in initial treatment decisions and improve prognosis.

From the BIO5 Institute (J.B.H., M.F.H.), University of Arizona, Tucson; CombinedBrain (K.H.), Brentwood; DenGen Consulting (D.M.J.S.), Rocky Top, TN; Children's National Medical Center (J.M.S.), Department of Neurology, Washington, DC; Department of Mathematics (J.C.W.), University of Arizona; and Neurology Department (M.F.H.), University of Arizona, Tucson.

Funding information and disclosures are provided at the end of the article. Full disclosure form information provided by the authors is available with the full text of this article at Neurology.org/NG.

The Article Processing Charge was funded by the authors.

Glossary

BFIE = benign familial infantile epilepsy; **DEE** = developmental and epileptic encephalopathy; **GE** = generalized epilepsy; **GOF** = gain-of-function; **GTC** = generalized tonic-clonic; **ICD10** = *International Classification of Diseases, 10th Revision*; **IDD** = intellectual and developmental disability; **IE** = intermediate epilepsy; **LOF** = loss-of-function; **NDD** = neurodevelopmental delay; **PCA** = principal component analysis; **SCB** = sodium channel blockers; **UE** = unclassified epilepsy; **VUS** = variant of uncertain significance.

Genetic variants at voltage-gated sodium channels (Na_V) are associated with a wide phenotypic spectrum of disease. The Na_V α-subunit gene family comprises 10 genes, each of which exhibits differential gene expression in heart (SCN5A, $Na_V1.5$), muscle (SCN4A, Nav1.4), peripheral nervous system (SCN9A, $Na_V1.7$), and CNS (SCN1A, SCN2A, and SCN8A encoding $Na_V1.1$, $Na_V1.2$, and $Na_V1.6$, respectively). Given the pivotal roles played by these ion channels in governing action potential initiation and propagation in excitable cells, it is not surprising that variants in these genes cause specific inherited ion channelopathies that range from relatively common disorders to very rare diseases.[1] Indeed, decreased or increased Na_V activity caused by loss-of-function (LOF) and gain-of-function (GOF) variants in the corresponding genes, respectively, underlies a broad spectrum of human disorders affecting the function of the heart, kidney, muscles, and peripheral and central nervous systems.[2]

In the case of SCN2A and SCN8A genes, primarily expressed on excitatory neurons in the brain, efficient methods are needed for determining the GOF or LOF character of variants. Biophysical studies have shown that GOF variants cause early-infantile epilepsies of variable severity, with seizure onset typically occurring in the first year of life,[3,4] while LOF variants result in later-onset epilepsies or neurodevelopmental delays and behavioral features without epilepsy.[3,5,6] The distinction between GOF and LOF variants has important implications for the treatment of these disorders. Patients with GOF variants often benefit from sodium channel blockers (SCBs), while SCBs tend to aggravate symptoms of patients with LOF variants.[4]

A critical challenge for neurologists and epileptologists is whether a patient carries a GOF or LOF variant at initial presentation.[7-9] While experimental studies of channel function are important, such studies are expensive, time-consuming, and often not feasible.[8] An alternative approach is to assess whether phenotypic differences exist between patients with LOF and GOF variants and determine whether such differences are present during diagnosis. Toward this goal, we examine a set of *SCN8A* variants that were previously classified as GOF and LOF by biophysical means and compare phenotypic features among patients who carry these variants. We then build predictive models to classify patients and test the accuracy of these models. Classification accuracy is assessed via confusion matrices to test the hypothesis that GOF and LOF represent distinct clinical phenotypes.

Methods

Identifying *SCN8A* LOF Variants

An exhaustive literature review identified all published cases of *SCN8A* LOF variants, including functional testing of such variants in heterologous cells and clinical features of patients carrying LOF variants. A search of the PubMed database between March and June 2022 using the search terms "*SCN8A* loss of function," "*SCN8A* clinical," "*SCN8A* epilepsy," and "*SCN8A* encephalopathy" yielded a total of 56 references. We collected information on any individual described to harbor pathogenic SCN8A variants that were confirmed to be complete or partial LOF, or published as LOF. We also collected information from any individual with a truncating variant (i.e., via nonsense, frameshift, splice site, or insertion/deletion). Only 1 randomly selected individual from any pair of related individuals was included. This resulted in a total of 69 cases with LOF variants, 50 that were in the literature only, 12 that were in the International SCN8A Registry only,[10] and 7 that were in both. A list of all 69 SCN8A individuals with LOF variants is summarized in eTable 1, links.lww.com/NXG/A598.

Constructing a Truth Set for Predictive Modeling

To construct a truth set to train predictive models, we selected a subset of the individuals listed in eTable 1, links.lww.com/NXG/A598 according to 3 criteria: (1) the individual carried a single pathogenic variant at *SCN8A*, (2) the individual did not harbor any known or likely pathogenic variants at loci other than *SCN8A*, and (3) the individual's variant had electrophysiologic data supporting an alteration of channel function. The final truth set contained 45 patients with LOF variants. This was matched with 45 patients with GOF variants randomly selected from 136 individuals with GOF variants in a recently published study.[5]

Phenotypes of the individuals listed in Table 1 included benign familial infantile epilepsy (BFIE), developmental and epileptic encephalopathy (DEE), generalized epilepsy (GE), intermediate epilepsy (IE), neurodevelopmental delay without epilepsy (NDD), and unclassified epilepsy (UE).[5] UE is defined as "both focal and generalized seizure types" or the patient's seizure types "were not adequately described" or "data were missing" (Table 1).[5]

For the clinical model, we reduced the number of features with the goal of maintaining those that are readily accessible

Table 1 Clinical Features of Individuals Included in the Truth Set

	GOF (n = 45)	LOF (n = 45)	Fisher test, *p* Value
Seizure onset			
Experienced seizure	45	29	6.02×10^{-6}
Age at onset (mo)	4.5 (*3.1*)	40.0 (38.1)	7.30×10^{-5}
Seizure type			
Tonic-clonic	31	10	1.65×10^{-5}
Focal	18	0	9.06×10^{-7}
Tonic	17	4	2.25×10^{-3}
Myoclonic	9	4	
Atonic	1	2	
Clonic	2	2	
Hemiclonic	2	1	
Epileptic spasms	3	0	
Absence	0	15	1.51×10^{-5}
Atypical absence	0	5	
Focal nonmotor nonaware	2	0	
Febrile	2	6	
Status epilepticus	7	1	5.82×10^{-2}
Intellectual disability			
Mild	3	13	1.13×10^{-2}
Moderate	9	6	
Severe	20	7	5.24×10^{-3}
Neurotypical	10	6	
Unknown	3	13	
Mutation type			
Missense	45	16	3.91×10^{-12}
Nonsense	0	8	5.56×10^{-3}
Frameshift	0	9	2.51×10^{-3}
Indel	0	6	2.62×10^{-2}
Splice	0	6	2.62×10^{-2}
Deletion	0	0	
Duplication	0	0	
Phenotype			
DEE	31	3	6.90×10^{-10}
BFIE	8	0	5.56×10^{-3}
IE	6	0	2.62×10^{-2}
GE	0	13	8.88×10^{-5}
NDD w/o epilepsy	0	11	4.88×10^{-4}
UE	0	6	2.62×10^{-2}
Unknown	0	12	

Abbreviations: BFIE = benign familial infantile epilepsy; DEE = developmental and epileptic encephalopathy; GE = generalized epilepsy; IE = intermediate epilepsy; NDD = neurodevelopmental delay; UE = unclassified epilepsy.

by a clinician at diagnosis. This was achieved by removing the phenotype[5] and the variant type. The seizure typed featured in the model was the first seizure type and did not include seizures developed later in life. In addition, we combined all seizure types into either "motor/focal" (i.e., generalized motor and focal motor and nonmotor seizures) or "absence" and only retained severe intellectual and developmental disability (IDD). We note that while mild or moderate IDD can sometimes be more difficult to detect in infancy, the clinical model uses severe IDD, typically apparent from an early age.

Individuals with *SCN8A* LOF variants that did not satisfy the criterion for inclusion in the truth set were classified into 3 subsets: those carrying an SCN8A missense variant and another variant at an epilepsy-associated locus other than *SCN8A* that was assessed to be pathogenic or likely pathogenic (variant of uncertain significance [VUS]) (subset 1, n = 6), those carrying an *SCN8A* truncating variant and a VUS at an epilepsy-associated locus other than *SCN8A* (subset 2, n = 6), and those published as missense LOF *SCN8A* without supporting biophysical data (subset 3, n = 12) (eTable 1, links.lww.com/NXG/A598).

Because of uncertainty of the functional effects of splice site variants, we also separately considered a fourth subset. To perform this, we removed the 6 individuals with splice site variants (individuals #40–45 in eTable 1, links.lww.com/NXG/A598) and constructed a restricted truth set of 39 individuals that was matched with 39 randomly chosen individuals with GOF variants. In the end, we had 4 scenarios for predictive modeling–the final truth set and the restricted truth set analyzed using both the full model consisting of all features and the clinical model having features selected based on the availability of relevant data in clinical settings.

Statistical Analyses and Data Visualization

Seizure types, IDD, mutation type, and phenotype were separated into unique binary variables expanding the data set from 6 features to 34 features (Table 1). Comparisons of feature frequencies between individuals with LOF and GOF variants were placed in a 2 × 2 table and analyzed using the Fisher exact test. Several dimension reduction analyses were performed on the truth set and the restricted truth set. Both

2-dimensional and 3-dimensional principal component analysis (PCA) provide a visualization of both GOF and LOF clustering. In addition, *t*-distributed stochastic neighbor embedding and uniform manifold approximation and projection provide additional dimension reduction visualization. All statistical analyses and visualizations were performed using R version 4.2.1.

Predictive Models

All predictive models and analyses used the following procedure: the truth set and the restricted truth set were split randomly into 70% training set and 30% testing set. Cross-validation was performed to generate the mean misclassification error (mmce) across 50 iterations of each learner. Random forest was determined to have the lowest mmce at 1%, a value that was used for all predictions. The number of independently constructed decision trees was set at 1,000, and 5-fold cross-validation was conducted a total of 10 times. Each model output was evaluated using mmce and by constructing confusion matrices and computing both the accuracy and the Matthew correlation coefficient ϕ.

Standard Protocol Approvals, Registrations, and Patient Consents

The Institutional Review Board at the University of Arizona approved the International SCN8A registry. The human subjects committee approved an online informed consenting process. Informed consent for minors was obtained from parents.

Data Availability

All relevant data are listed in eTable 1, links.lww.com/NXG/A598.

Results

Distinguishing Features of Individuals With LOF and GOF Variants

Table 1 lists the features of the individuals included in the truth set for LOF and GOF, which is composed of the following: (1) the development of seizures, (2) age at seizure onset, (3) seizure types, (4) IDD, (5) mutation type, and (6) phenotype. All individuals with GOF variants experienced seizures while only 29 of the 45 individuals with LOF variants experienced seizures ($p = 6.02 \times 10^{-6}$). The average age at seizure onset for the individuals who had a history of seizures is 4.5 ± 3.1 months and 40.0 ± 38.1 months for GOF and LOF, respectively (*t* test *p* value = 7.30×10^{-5}). Fisher tests on 2 × 2 tables (Table 1) indicate that generalized tonic-clonic (GTC) ($p = 1.65 \times 10^{-5}$), focal ($p = 9.06 \times 10^{-7}$), and tonic seizures ($p = 2.25 \times 10^{-3}$) are more prevalent in the GOF group, while absence and atypical seizures are present exclusively in the LOF group (18 vs 0, $p = 1.24 \times 10^{-7}$). Intellectual disability is more frequently severe in individuals with GOF variants ($p = 5.24 \times 10^{-3}$), while it was more typically mild in individuals with LOF variants ($p = 1.13 \times 10^{-2}$).

The remaining 29 individuals with LOF variants had frameshift (n = 9), indel (n = 6), nonsense (n = 8), and splice site (n = 6) variants. Every individual with a GOF variant had a missense variant, whereas 16 of the 45 individuals with LOF variants in the truth set had a missense variant ($p = 3.91 \times 10^{-12}$). Phenotypes also show distinct differences. For the truth set, all cases of GE (13 vs 0, $p = 8.88 \times 10^{-5}$) and NDD without epilepsy (11 vs 0, $p = 4.88 \times 10^{-4}$) occur in individuals with LOF variants. Similarly, all 6 cases of UE occur in individuals with LOF variants (6 vs 0, $p = 2.62 \times 10^{-2}$). BFIE (8 vs 0, $p = 5.56 \times 10^{-3}$) and IE (6 vs 0, $p = 2.62 \times 10^{-3}$) are features exclusively of individuals with GOF variants. DEE is primarily a feature of individuals with GOF variants (31 vs 3, $p = 6.90 \times 10^{-10}$). For the reduced truth set, 2 of the 3 LOF DEE individuals had splice site variants. In this case, the *p* value for the Fisher test decreases from 4.58×10^{-9} to 2.53×10^{-10}. Other *p* values for the reduced truth set were similar with generally slightly larger *p* values (Table 1).

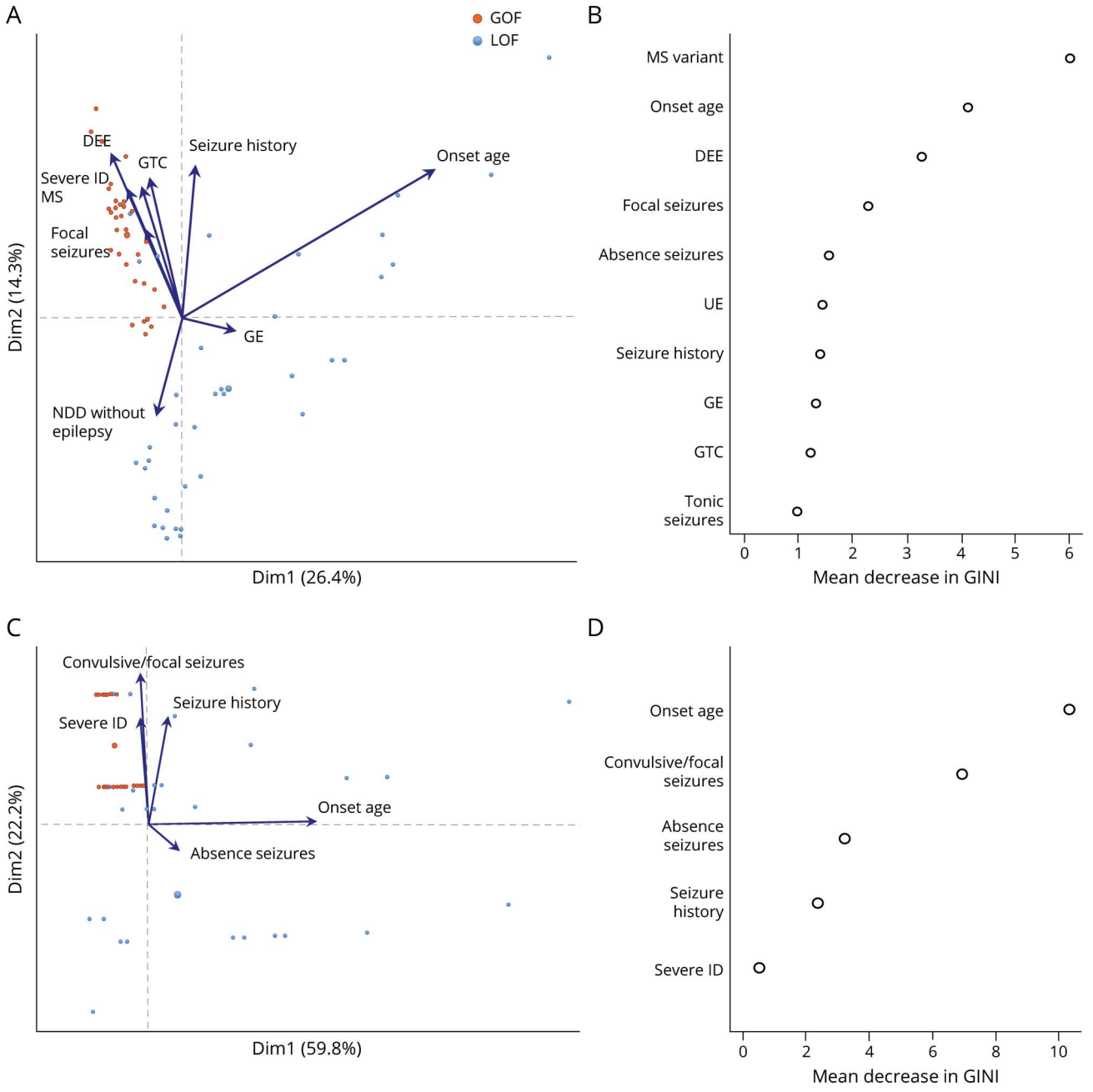
Data Visualization

For the full model as applied to the truth set, the PCA shows that 72.3% of the variance is explained with 10 features (Figure 1A). The vertical and horizontal axes in the bi-plot (the top 2 eigenvectors) explained 26.4% and 14.3% of the total variance, respectively, among individuals with GOF and LOF variants. The former strongly follow a vertical line that extends mainly in quadrant II, while the latter more diffusely follow a diagonal line in quadrants I and III. In the 2-dimensional full model PCA, the features DEE, GTC and focal seizures, severe IDD, and missense variants make moderate and similar contributions. This can be seen because the associated arrows point in nearly the same or the opposite direction, whereas age at seizure onset plays a distinct and nearly orthogonal role in distinguishing individuals with GOF and LOF variants. Note that arrows pointing in nearly opposite directions indicate that an increase in one feature and a decrease in the other makes the same type of contribution to the variance among individual phenotypes. To clarify the overlap in vectors in the 2-dimensional PCA, we provide a 3-dimensional visualization of the PCA plot in eFigure 1, links.lww.com/NXG/A596. For the clinical model as applied to the truth set, the PCA gave 5 features with variance contributions as follows: (1) age at seizure onset (73.7%), (2) absence seizures (12.1%), (3) development of seizures (7.7%), (4) severe IDD (3.8%), and (5) motor/focal seizures (2.7%). Figure 1C displays individuals in a 2-dimensional PCA plot while eFigure 2, links.lww.com/NXG/A597 shows a 3-dimensional version of the PCA. Individuals with GOF variants form 2 distinct tight clusters in quadrant II, while those with LOF variants are more diffusely distributed in quadrants I, III, and IV. Thus, for the clinical model, for example, severe IDD, motor/focal seizures, and a seizure history have a similar interpretation.

Random Forest Analysis

For the full model, the top 10 contributing features were calculated using mean decrease Gini coefficient (Figure 1B). Seven of these top 10 features (age at seizure onset, missense mutation, GTC seizures, GE, absence seizures, DEE, focal

Figure 1 Visualization of Truth Set and Feature Importance



(A) Principal component analysis bi-plot with top 11 contributing features from Full Model. GOF (red) and LOF (blue) are shown on first 2 principal components. Feature contribution is represented by the length of the vector: age at *onset* (29.4%), DEE (7.7%), GE (6.2%), absence seizures (6.0%), missense mutation (5.1%), UE (4.4%), focal seizures (3.9%), severe IDD (3.7%), mild IDD (3.2%), and unknown IDD (2.8%). (B) Feature importance in random forest determined by mean decrease Gini in the full model: missense mutation (6.00), age at onset (4.12), DEE (3.25), focal seizures (2.28), absence seizures (1.54), UE (1.43), development of seizures (1.40), GE (1.32), GTC seizures (1.21), and tonic seizures (0.98). (C) Principal component analysis bi-plot of clinical model. (D) Feature importance in random forest model for the clinical model. Gini coefficient importance rankings were as follows: age at seizure onset (10.29), motor/focal seizures (6.92), absence seizures (3.23), development of seizures (2.38), and severe IDD (0.53). DEE = developmental and epileptic encephalopathy; GOF = gain-of-function; GTC = generalized tonic-clonic; IDD = intellectual and developmental disability; LOF = loss-of-function; UE = unclassified epilepsy.

seizures, and development of seizures) were shared with the top 10 features contributing to variation in the PCA. Probabilities for LOF classification are listed in eTable 1, links.lww.com/NXG/A598. The confusion matrix based on predictions from the random trees is summarized in Table 2A (top) with accuracy 0.999 and Matthew correlation coefficient $\phi = 0.998$. In the full model, individuals with LOF variants were falsely classified as having GOF variants in approximately 1.1% of the runs. Figure 1D shows the Gini coefficient importance rankings for the clinical model, the accuracy of which was approximately 95.4% and the Matthew correlation coefficient was $\phi = 0.910$ (Table 2A, bottom).

Table 2 Relative Confusion Matrix for GOF/LOF Predictions Based on Full and Clinical Models Across 50 Iterations

		Predicted	
	True	**Gain**	**Loss**
Full model (φ = 0.998)	Gain	50.0%	0.0%
	Loss	0.1%	49.9%
Clinical model (φ = 0.910)	Gain	50.0%	0.0%
	Loss	4.7%	45.3%

Abbreviations: GOF = gain-of-function; LOF = loss-of-function; ϕ, Matthew correlation coefficient.

Figure 2A shows the probability distribution from random forest prediction for the full model. There is a clear distinction between GOF and LOF with no overlap among the 90 individuals included in the analysis. Figure 2B shows the probability distribution using the clinical model. Similarly, there is a clear distinction in the distribution among individuals in the truth set carrying GOF and LOF variants, except in the following cases. Two individuals predicted as those with GOF variants in the clinical model were predicted as those with LOF variants in the full model: individual #29 carries a frameshift mutation in the inactivation gate, and individual #32 carries a nonsense mutation at the 3′ end of the gene (probLOF = 0.08 in both cases). Both individuals had early-onset seizures (GTC and tonic or myoclonic), one of which was classified as DEE and the other remained as UE (eTable 1, links.lww.com/NXG/A598).

To further investigate the variability in predicting GOF or LOF effects in individuals carrying splice site variants, we repeated the random forest analyses for the clinical model on the restricted truth set. The results for the confusion matrix, the Matthew correlation coefficient, and the accuracy are summarized in Table 2B. Subset 4 contains 6 individuals who possess splice site variants (predicted to cause exon skipping). The results confirm that only 2 of these individuals were classified as those with GOF variants (probLOF = 0.08 and 0.25) (Table 3, Figure 2B).

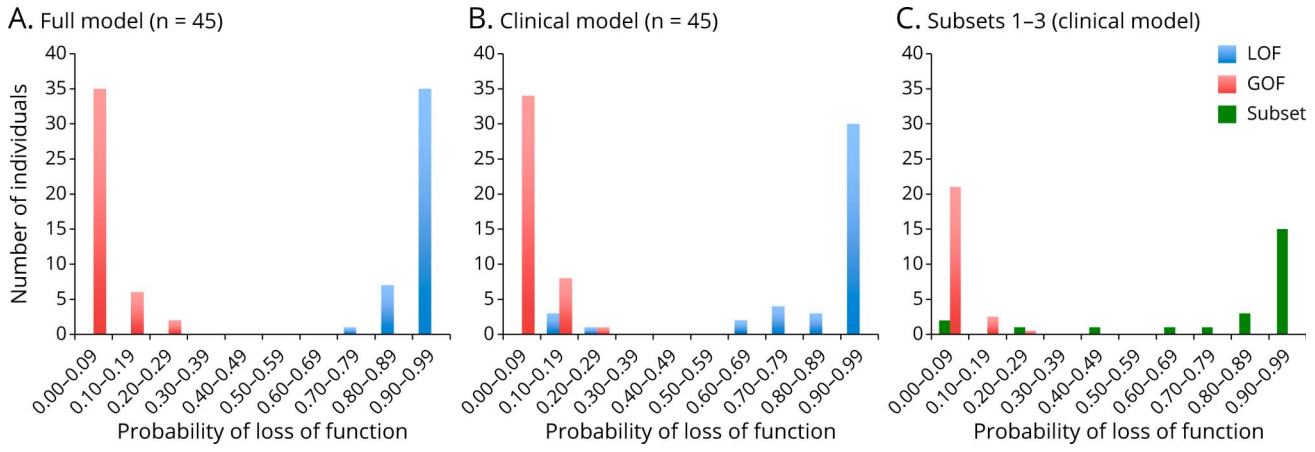
Subset Analysis

The clinical model was used to classify individuals in each subset category in eTable 1, links.lww.com/NXG/A598 (Table 3). Four individuals in subset 1 were clearly classified as those with LOF variants (#48–51, eTable 1, links.lww.com/NXG/A598), while individuals #46 and #47 had intermediate probabilities: probLOF = 0.64 and 0.47, respectively (Figure 2C). Both of these individuals carried 2 missense variants at SCN8A (eTable 1, links.lww.com/NXG/A598). Individual #47 experienced tonic seizures beginning at 7 months and severe IDD. For subset 2, 2 individuals were classified as those with GOF variants (#51 and #53 in eTable 2, links.lww.com/NXG/A598) (Table 3, Figure 2C). Individual #51 has several variants at other loci, while individual #53 carries a second variant at SCN8A leading to an amino acid substitution. All 12 individuals in subset 3 were classified as those with LOF variants (Table 3, Figure 2C).

Discussion

Pathogenic variants at the voltage-gated sodium channel gene, *SCN8A*, are associated with a wide spectrum of clinical disease outcomes. Perhaps one of the most important determinants of clinical manifestation, as well as for informing treatment decisions in the clinic, is the effect a particular variant has on $Na_v1.6$ channel activity. For instance, *SCN8A* GOF and LOF variants exhibit discrete biophysical properties for enhanced

Figure 2 Distribution of Probability of LOF for Each Individual in (A) the Full Model Truth Set, (B) the Clinical Model Truth Set, and (C) Subsets 1–3 (Green)



Individual's true classification is indicated by color: GOF (red) and LOF (blue). Those with probability LOF less than 0.5 are classified as GOF and those greater than 0.5 are classified as LOF. GOF = gain-of-function; LOF = loss-of-function.

Table 3 Prediction Results for Subsets of Individuals Excluded From the Truth Set Using the Clinical Model

	n	GOF	LOF
Subset 1: missense with VUS	6	1	5
Subset 2: truncating with VUS	6	2	4
Subset 3: published LOF without in vitro testing	12	0	12
Subset 4[a]: splice site mutations	6	2	4

Abbreviations: GOF = gain-of-function; LOF = loss-of-function; VUS = variant of uncertain significance.
[a] Also included in truth set.

and reduced Na_v activities, respectively. Patients with GOF variants generally present with early-infantile epilepsies of variable severity, while those with LOF variants typically present with neurodevelopmental delay with or without epilepsy. Patients with GOF variants often benefit from SCBs, while symptoms of those with LOF variants can be exacerbated by further reduction in sodium channel activity.[4] When the functional consequence of a particular variant is known, this information informs treatment decisions in the clinic. However, this is not often the case because in vitro experiments to determine the biophysical properties of variants are expensive and time-consuming. Clinicians therefore must use their best judgment and a trial and error approach when mapping out a treatment plan for an individual patient with a novel or unclassified *SCN8A* variant.

We compared clinically determined features of individuals deemed to carry *SCN8A* GOF and LOF variants by means of in vitro studies or because the variant was truncating, resulting in a total of 69 variants: 34 MS and 35 truncating variants, including 9 NS, 13 FS, 6 splice site, 6 indels, and 1 large deletion. We then assembled a truth set of variants with known functional effects (Table 1), excluding individuals carrying variants at other loci associated with epilepsy. We then trained a predictive model using a truth set on a full model of 45 LOF variants (Table 1) and 45 GOF variants randomly selected from a set of variants tested by in vitro methods.[5]

Using the full model including 34 clinical features (Table 1), individuals could be classified into 2 groups with high precision as measured by positive predictive value (98.5%) and with an accuracy of 99.8% (Table 2). Individuals with GOF and LOF variants were separated into 2 distinct clinical cohorts differing in age at seizure onset, development of seizures, seizure type, IDD, and DEE. Given that several patient phenotypes were tightly linked to GOF or LOF status, we wanted to assess the extent to which clinical indicators that were likely to be present during diagnosis could be used to predict the functional category of the variant. To perform this, we reduced the complexity of the full model in a stepwise manner while testing the predictive value to ensure that little or no loss of precision occurred at each step. The final set of clinical features resulted in a clinical model that separately classified patients with GOF and LOF variants with a positive predictive value of 91.0% and an accuracy = 95.3 for the truth set (Table 2).

We found that the clinical model was slightly more consistent in classifying cases when the patient carried a missense vs nonsense or frameshift variant because 2 of the latter individuals were classified as those with GOF variants (eTable 1, links.lww.com/NXG/A598). The functional effects of splice site variants are more difficult to test, given the complexity of the splicing mechanism. Indeed, there is evidence that splice site variants can lead to GOF or LOF effects.[11-13] To investigate whether splice site variants were associated with LOF clinical features in our data set, we constructed a restricted truth set that excluded all splice site variants (Table 1). The precision of the predictive model was slightly reduced (i.e., from 98.8% to 97.2%) and maintained an accuracy of 96.6%. Thus, caution is warranted when interpreting the functional effects of splice site variants. We also note that the effect of small indels is difficult to test in vitro, and little evidence exists to support an LOF effect for such variants in *SCN8A*. However, in-frame indels result in reduced expression and LOF properties of $Na_V1.1$ and $Na_V1.7$ in patients with Dravet syndrome and reduced sensitivity to pain, respectively.[14,15] We included 3 small indels in our truth set that were deemed to have LOF properties[5] and excluded 3 others for which there were no such inferences. Despite uncertainty, all 6 indels had high probabilities of LOF in the clinical model, ranging from 0.73 to 1.00 (eTable 1, links.lww.com/NXG/A598).

In some cases, the range of symptoms associated with specific variants are not easily explained at a molecular and functional level. This is partly due to differences in genetic background[4] and often by variants (of uncertain significance) at other genes implicated in epilepsy and listed on the genetic report.[16] For example, we found that 2 individuals harboring 2 *SCN8A* LOF alleles had intermediate probabilities of LOF (Figure 2C, eTable 1, links.lww.com/NXG/A598). These dominant variants result in partial or complete LOF of the $Na_V1.6$ channel and are associated with mild cognitive impairment in heterozygous carriers and severe DEE in individuals inheriting 2 mutant alleles. It was suggested that the clinical consequences of carrying 2 partial or complete loss-of-function alleles are variable and depend on genetic background.[17]

We also found that 2 individuals carrying a truncating variant and a second VUS were classified as those with GOF variants. The first individual (#53, eTable 1, links.lww.com/NXG/A598) carried a second SCN8A variant (c.3985A > G, N1329D) not found in gnomAD that has been seen in another individual with SCN8A-related epilepsy with clear GOF features (eTable 1, links.lww.com/NXG/A598). The second individual carries a variant (I619L) in the EF-hand domain containing 1 gene that has been found in several individuals with juvenile myoclonic epilepsy or GTC seizures on

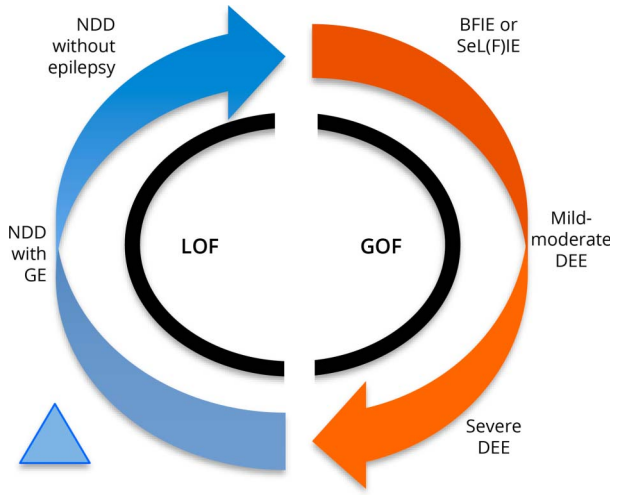
awakening.[18] We also used the clinical model to classify a set of 12 individuals published in the literature (subset 3) inferred to carry *SCN8A* LOF variants despite the lack of in vitro testing. All 12 individuals carried MS variants and were classified as those with LOF variants (Table 3, Figure 2C), suggesting that clinical data alone can be quite accurate in classifying individuals in some cases.

Figure 3 advances a model integrating phenotypic subcategories with LOF and GOF variant functional classes, primarily reflecting the high predictive value of the random forest learner in categorizing patients by their phenotypic features in the full model. Similarly, the classification of individuals using a reduced set of features in the clinical model had high predictive value, primarily depending on whether a patient presents with either early-onset motor/focal seizures, NDD without epilepsy, or NDD accompanied by late-onset absence seizures. Figure 3 depicts 3 and 2 phenotypic subcategories associated with GOF and LOF variants, respectively. The triangle in Figure 3 represents a possible third subcategory of patients with LOF variants that share a subset of features with patients harboring GOF variants (e.g., tonic-clonic, tonic, and myoclonic seizures; Table 1). These patients are phenotypically distinguished from those with GOF variants because they have diagnostic features found to be associated with LOF variants in our model (e.g., late age at seizure onset).[5]

Because GE occurs in association with NDD and that late-onset absence seizures are often the primary seizure type, we prefer the nomenclature NDD with GE over GE alone. Moreover, the term "generalized" is often applied incorrectly when describing seizure types clinically that may be more accurately characterized as focal or unknown onset. Patients with *SCN8A*-DEE may appear to have both generalized and focal seizures clinically,[20-22] while ictal EEG often reveals focal onset[23] (though not always[20]). We also prefer the term moderate DEE over the IE terminology (as shown in Figure 3) because IE occurs in patients with GOF variants (Table 1), and some variants may produce a widely variable phenotype encompassing IE and DEE. Figure 3 also considers DEE as a spectrum from mild to severe because these patients tend to share common clinical features including focal seizures and developmental delay. We note that BFIE has recently been renamed as part of self-limited familial infantile epilepsy or SeL(F)IE.[19] Individuals identified with UE are described with either LOF or GOF.[5] However, every individual analyzed with UE in this study had a LOF variant as indicated by electrophysiologic testing or presence of a protein-truncating variant. Five of 8 had seizure onset greater than or equal to 11 months.

Altered channel activity associated with GOF or LOF variants at other Na_Vs often corresponds to distinct clinical disease manifestations, as well as to differences in drug response.[8] For example, in *SCN5A*, LOF variants can cause Brugada syndrome, whereas GOF variants can lead to long QT syndrome.[25] In the case of $Na_V1.7$, GOF variants cause severe painful disorders, while LOF variants cause congenital insensitivity to pain.[26] GOF and LOF variants at *SCN4A* are associated with myotonic muscle stiffness and muscle weakness, respectively.[27]

Figure 3 Proposed Model Integrating Phenotypic Subcategories of SCN8A Patients With LOF and GOF Variant Functional Classes



Major division between LOF and GOF is mainly governed by the presence of early-onset motor or focal seizures (GOF) and neurodevelopmental delay (NDD) without seizures or NDD with seizures (e.g., late-onset absence seizures) (LOF). Subphenotypic nomenclature is discussed in the text. The black inner circular line represents variant function, while the outer arrows represent the phenotypic spectrum associated with GOF and LOF variants. The triangle represents a possible third subcategory of patients with LOF variants (e.g., individuals in Table 1 with high probLOF scores and DEE and/or tonic-clonic, tonic, or myoclonic seizures). DEE = developmental and epileptic encephalopathy; GOF = gain-of-function; LOF = loss-of-function; NDD = neurodevelopmental delay without epilepsy.

SCN8A-related disorders show a similar pattern. This study supports the hypothesis that patients with *SCN8A* GOF and LOF variants represent distinct clinical phenotypes. The clinical utility of the random forest classifier is robustly illustrated by extending the model from a full model to a model with 5 key phenotypic features typically present during diagnosis. Just as in the case of the other sodium channelopathies, we recommend that the clinical phenotypes of patients harboring GOF and LOF variants be treated as distinct neurologic disorders that require different treatment approaches. One way forward would be to note these disorders with different ICD10 codes to clarify treatment options (i.e., avoid SCBs in individuals with LOF variants) and improve prognosis.[4,8,9,21,28]

This study demonstrates that *SCN8A*-mediated neurodevelopmental disorders essentially encompass 2 distinct groups of patients, those with GOF variants vs those with LOF variants. Our predictive model currently has a low error rate in categorizing these patients based on a limited number of clinical features. While we performed an exhaustive search to identify every possible individual having verified LOF variants, the data set was not gathered under a well-constructed design procedure. This was counterbalanced by the fact that we did not select a subset of individuals with particular features to include in the full data set. Yet, there are

only a relatively small number of cases in the literature that have both in vitro and in vivo data. As the ability to recognize vases with *SCN8A* LOF variants improves, we may find a broadening spectrum of features. This will certainly affect our predictive modeling. Selection biases may still exist because we do not know how cases were selected for in vitro testing.

We must also be aware that in vitro ion channel function may not always translate into in vivo channel behavior. As for the utility of the clinical model for assessing probability of LOF early in the patient's diagnostic journey, we acknowledge that more work is needed to classify phenotypic differences associated with *SCN8A* LOF and GOF variants and to determine whether such differences are present during diagnosis. In the meantime, we suggest referring to patients by phenotype rather than by putative variant function and using the classification of patients by variant type for guidance in treatment options. This approach reinforces decades of epilepsy practice, where patients are treated according to epilepsy phenotype without knowledge of underlying genotype. For instance, patients with focal epilepsy might be expected to respond favorably to sodium channel–blocking medications while those with GE would not. Despite these limitations, we believe that the clinical model has great utility because it provides a rapid and highly accurate platform for predicting the functional class of patient variants during *SCN8A* diagnosis, which can aid in initial treatment decisions and improve prognosis.

Acknowledgment

The authors thank Neurocrine Biosciences and the Shay Emma Hammer Research Foundation for funding. The authors also thank Gabrielle Conecker and JayEtta Hecker of the International SCN8A Alliance for feedback on the design of the study and comments on earlier versions of this manuscript, and Terry Jo Bichell (CombinedBrain) as well for helpful comments. MFH (University of Arizona) conceived the study, wrote the manuscript, J.B. Hack performed analysis and wrote the manuscript, J.C. Watkins performed statistical analysis and wrote the manuscript, K. Horning and D.M. Juroske Short performed literature searches, and J.M. Schreiber wrote the manuscript.

Study Funding

Neurocrine Biosciences, Shay Emma Hammer Research Foundation.

Disclosure

J.B. Hack, K. Horning, D.M. Juroske Short, J.M. Schreiber, J.C. Watkins, M.F. Hammer. Full disclosure form information provided by the authors is available with the full text of this article at Neurology.org/NG.

Publication History

Received by *Neurology: Genetics* October 10, 2022. Accepted in final form January 12, 2023. Submitted and externally peer reviewed. The handling editor was Massimo Pandolfo, MD, FAAN.

Appendix Authors

Name	Location	Contribution
Joshua B. Hack, BS	BIO5 Institute, University of Arizona, Tucson, AZ	Drafting/revision of the article for content, including medical writing for content; study concept or design; analysis or interpretation of data
Kyle Horning, PhD	CombinedBrain, Brentwood, TN	Major role in the acquisition of data
Denise M. Juroske Short, PhD	DenGen Consulting, Rocky Top, TN	Major role in the acquisition of data
John M. Schreiber, MD	Children's National Medical Center, Department of Neurology, Washington, DC	Drafting/revision of the article for content, including medical writing for content; analysis or interpretation of data
Joseph C. Watkins, PhD	Department of Mathematics, University of Arizona, Tucson, AZ	Drafting/revision of the article for content, including medical writing for content; analysis or interpretation of data
Michael F. Hammer, PhD	BIO5 Institute, University of Arizona, Tucson AZ; Neurology Department, University of Arizona, Tucson, AZ	Drafting/revision of the article for content, including medical writing for content; major role in the acquisition of data; study concept or design; and analysis or interpretation of data

References

1. Imbrici P, Liantonio A, Camerino GM, et al. Therapeutic approaches to genetic ion channelopathies and perspectives in drug discovery. *Front Pharmacol.* 2016;7:121.
2. Gripp KW, Smithson SF, Scurr IJ, et al. Syndromic disorders caused by gain-of-function variants in KCNH1, KCNK4, and KCNN3-a subgroup of K(+) channelopathies. *Eur J Hum Genet.* 2021;29:1384-1395.
3. Berecki G, Howell KB, Heighway J, et al. Functional correlates of clinical phenotype and severity in recurrent SCN2A variants. *Commun Biol.* 2022;5:515.
4. Meisler MH, Hill SF, Yu W. Sodium channelopathies in neurodevelopmental disorders. *Nat Rev Neurosci.* 2021;22:152-166.
5. Johannesen KM, Liu Y, Koko M, et al. Genotype-phenotype correlations in SCN8A-related disorders reveal prognostic and therapeutic implications. *Brain.* 2022;145:2991-3009.
6. Wagnon JL, Barker BS, Ottolini M, et al. Loss-of-function variants of SCN8A in intellectual disability without seizures. *Neurol Genet.* 2017;3:e170.
7. Carbone E, Mori Y. Ion channelopathies to bridge molecular lesions, channel function, and clinical therapies. *Pflugers Arch.* 2020;472:733-738.
8. Heyne HO, Baez-Nieto D, Iqbal S, Palmer DS, Brunklaus A, May P; Epi25 Collaborative, Johannesen KM, Lauxmann S, Lemke JR, Møller RS, Pérez-Palma E, Scholl UI, Syrbe S, Lerche H, Lal D, Campbell AJ, Wang HR, Pan J, Daly MJ. Predicting functional effects of missense variants in voltage-gated sodium and calcium channels. *Sci Transl Med.* 2020;12:eaay6848.
9. Wolff M, Johannesen KM, Hedrich UBS, et al. Genetic and phenotypic heterogeneity suggest therapeutic implications in SCN2A-related disorders. *Brain.* 2017;140:1316-1336.
10. Hammer MF. *The International SCN8A Registry [online]*. Accessed October 2022. scn8a.net.
11. Chen J, Kleyman TR, Sheng S. Deletion of α-subunit exon 11 of the epithelial Na+ channel reveals a regulatory module. *Am J Physiol Ren Physiol.* 2014;306:F561-F567.
12. Heighway J, Sedo A, Garg A, et al. Sodium channel expression and transcript variation in the developing brain of human, Rhesus monkey, and mouse. *Neurobiol Dis.* 2022;164:105622.
13. Wu F, Mi W, Fu Y, Struyk A, Cannon SC. Mice with an NaV1.4 sodium channel null allele have latent myasthenia, without susceptibility to periodic paralysis. *Brain.* 2016;139:1688-1699.
14. Thompson CH, Porter JC, Kahlig KM, Daniels MA, George AL Jr. Nontruncating SCN1A mutations associated with severe myoclonic epilepsy of infancy impair cell surface expression. *J Biol Chem.* 2012;287:42001-42008.
15. Wang JY, Tang B, Sheng WX, et al. Clinical and functional features of epilepsy-associated in-frame deletion variants in SCN1A. *Front Mol Neurosci.* 2022;15:828846.

16. Encinas AC, Watkins JC, Longoria IA, Johnson JP Jr, Hammer MF. Variable patterns of mutation density among NaV1.1, NaV1.2 and NaV1.6 point to channel-specific functional differences associated with childhood epilepsy. *PLoS One.* 2020;15: e0238121.
17. Wengert ER, Tronhjem CE, Wagnon JL, et al. Biallelic inherited SCN8A variants, a rare cause of SCN8A-related developmental and epileptic encephalopathy. *Epilepsia.* 2019;60:2277-2285.
18. Gonsales MC, Ribeiro PAO, Betting LE, et al. Revisiting the clinical impact of variants in EFHC1 in patients with different phenotypes of genetic generalized epilepsy. *Epilepsy Behav.* 2020;112:107469.
19. Zuberi SM, Wirrell E, Yozawitz E, et al. ILAE classification and definition of epilepsy syndromes with onset in neonates and infants: position statement by the ILAE Task Force on Nosology and Definitions. *Epilepsia.* 2022;63:1349-1397.
20. Schreiber JM, Tochen L, Brown M, et al. A multi-disciplinary clinic for SCN8A-related epilepsy. *Epilepsy Res.* 2020;159:106261.
21. Larsen J, Carvill GL, Gardella E, et al, EuroEPINOMICS RES Consortium CRP. The phenotypic spectrum of SCN8A encephalopathy. *Neurology.* 2015;84:480-489.
22. Kong W, Zhang Y, Gao Y, et al. SCN8A mutations in Chinese children with early onset epilepsy and intellectual disability. *Epilepsia.* 2015;56:431-438.
23. Gardella E, Marini C, Trivisano M, et al. The phenotype of SCN8A developmental and epileptic encephalopathy. *Neurology.* 2018;91:e1112-e1124.
24. Johannesen KM, Gardella E, Encinas AC, et al. The spectrum of intermediate SCN8A-related epilepsy. *Epilepsia.* 2019;60:830-844.
25. Cerrone M, Remme CA, Tadros R, Bezzina CR, Delmar M. Beyond the one gene-one disease paradigm: complex genetics and pleiotropy in inheritable cardiac disorders. *Circulation.* 2019;140:595-610.
26. Drenth JP, Waxman SG. Mutations in sodium-channel gene SCN9A cause a spectrum of human genetic pain disorders. *J Clin Invest.* 2007;117:3603-3609.
27. Jurkat-Rott K, Holzherr B, Fauler M, Lehmann-Horn F. Sodium channelopathies of skeletal muscle result from gain or loss of function. *Pflugers Arch.* 2010;460:239-248.
28. Schoonjans A, Paelinck BP, Marchau F, et al. Low-dose fenfluramine significantly reduces seizure frequency in Dravet syndrome: a prospective study of a new cohort of patients. *Eur J Neurol.* 2017;24:309-314.