De Novo Truncating Variants in the Last Exon of *SEMA6B* Cause Progressive Myoclonic Epilepsy

Kohei Hamanaka,[1,12] Eri Imagawa,[1,12] Eriko Koshimizu,[1,12] Satoko Miyatake,[1] Jun Tohyama,[2] Takanori Yamagata,[3] Akihiko Miyauchi,[3] Nina Ekhilevitch,[4] Fumio Nakamura,[5] Takeshi Kawashima,[6] Yoshio Goshima,[6] Ahmad Rithauddin Mohamed,[7] Gaik-Siew Ch'ng,[8] Atsushi Fujita,[1] Yoshiteru Azuma,[1] Ken Yasuda,[9] Shintaro Imamura,[10] Mitsuko Nakashima,[11] Hirotomo Saitsu,[11] Satomi Mitsuhashi,[1] Takeshi Mizuguchi,[1] Atsushi Takata,[1] Noriko Miyake,[1] and Naomichi Matsumoto[1,*]

De novo variants (DNVs) cause many genetic diseases. When DNVs are examined in the whole coding regions of genes in next-generation sequencing analyses, pathogenic DNVs often cluster in a specific region. One such region is the last exon and the last 50 bp of the penultimate exon, where truncating DNVs cause escape from nonsense-mediated mRNA decay [NMD(−) region]. Such variants can have dominant-negative or gain-of-function effects. Here, we first developed a resource of rates of truncating DNVs in NMD(−) regions under the null model of DNVs. Utilizing this resource, we performed enrichment analysis of truncating DNVs in NMD(−) regions in 346 developmental and epileptic encephalopathy (DEE) trios. We observed statistically significant enrichment of truncating DNVs in semaphorin 6B (*SEMA6B*) (p value: 2.8×10^{-8}; exome-wide threshold: 2.5×10^{-6}). The initial analysis of the 346 individuals and additional screening of 1,406 and 4,293 independent individuals affected by DEE and developmental disorders collectively identified four truncating DNVs in the SEMA6B NMD(−) region in five individuals who came from unrelated families (p value: 1.9×10^{-13}) and consistently showed progressive myoclonic epilepsy. RNA analysis of lymphoblastoid cells established from an affected individual showed that the mutant allele escaped NMD, indicating stable production of the truncated protein. Importantly, heterozygous truncating variants in the NMD(+) region of *SEMA6B* are observed in general populations, and *SEMA6B* is most likely loss-of-function tolerant. Zebrafish expressing truncating variants in the NMD(−) region of *SEMA6B* orthologs displayed defective development of brain neurons and enhanced pentylenetetrazole-induced seizure behavior. In summary, we show that truncating DNVs in the final exon of *SEMA6B* cause progressive myoclonic epilepsy.

Whole-exome sequencing (WES) has allowed for the comprehensive identification of *de novo* variants (DNVs) in gene coding regions as pathogenic causes for many genetic diseases. The ongoing reduction in sequencing costs has enabled recent WES studies to incorporate increasingly large numbers of individuals, sometimes more than 10,000.[1] However, pathogenic causes remain unidentified in the majority of individuals.[1,2] The simplest way to identify such rare undiscovered causes would be to increase the sample sizes of future studies; however, improvement in the sensitivity of the analytical method is always beneficial.

WES analyzes many genes at once; therefore, a sophisticated statistical framework is required for the identification of *bona fide* disease-causing variants and DNVs. One popular approach is to calculate the probability of the observed number of DNVs in a gene of interest by referring to the theoretical DNV rate. Samocha et al.[3] developed a null model that takes tri-nucleotide context into consideration to provide gene-specific DNV probabilities of each variant type (e.g. nonsense, missense, etc.). This valuable resource for enrichment analysis of DNVs has been widely used in large-scale analyses of DNVs.[2,4–6]

In this model, all coding regions are weighted equally in a calculation of a per-gene DNV probability. However, in the real world, a specific type of DNV in a specific region can cause a rare disease. For example, truncating DNVs in the last exon and in the last 50 bp of the penultimate exon of a gene [hereafter termed the NMD(−) region] can cause disease. Such DNVs likely cause the transcript to escape nonsense-mediated mRNA decay (NMD).[7,8] Diseases can also result from missense DNVs at constrained coding regions.[9] Focusing on specific combinations of variant type and areas of a coding region should increase the sensitivity of enrichment analysis. Such analysis is predicted to identify disease-causative variants that conventional analyses do not.

Here, we focused on truncating DNVs in the NMD(−) region of each gene and performed enrichment analysis in our cohort of developmental and epileptic encephalopathy (DEE). Consequently, we identified truncating DNVs at

[1]Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama 236-0004, Japan; [2]Department of Pediatrics, National Hospital Organization Nishi-Niigata Chuo National Hospital, Niigata 950-2085, Japan; [3]Department of Pediatrics, Jichi Medical University, Tochigi 329-0498, Japan; [4]The Genetics Institute, Rambam Health Care Campus, Haifa 3109601, Israel; [5]Department of Biochemistry, Tokyo Women's Medical University, Tokyo 162-8666, Japan; [6]Department of Molecular Pharmacology and Neurobiology, Yokohama City University Graduate School of Medicine, Yokohama 236-0004, Japan; [7]Department of Pediatric Neurology Unit, Pediatric Institute, Hospital Kuala Lumpur, Kuala Lumpur, Malaysia; [8]Department of Genetics, Hospital Kuala Lumpur, Kuala Lumpur, Malaysia; [9]Department of Neurology, Kyoto University Graduate School of Medicine, Kyoto 606-8507, Japan; [10]Reserch Center for Biochemistry and Food Technology, National Research Institute of Fisheries Science, Yokohama 236-8648, Japan; [11]Department of Biochemistry, Hamamatsu University School of Medicine, Hamamatsu, Shizuoka 431-3192, Japan
[12]These authors contributed equally to this work
*Correspondence: naomat@yokohama-cu.ac.jp
https://doi.org/10.1016/j.ajhg.2020.02.011.
© 2020 American Society of Human Genetics.

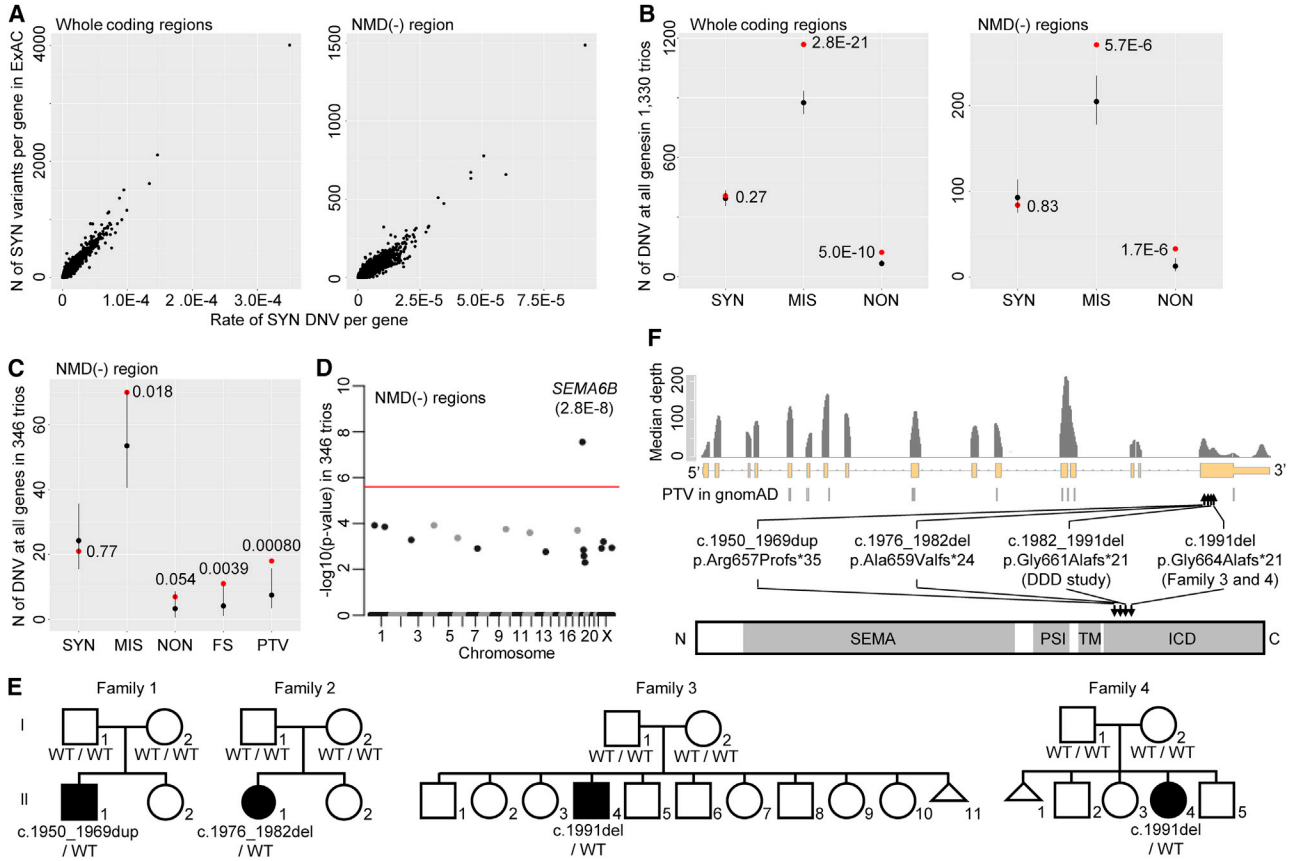

Figure 1. Enrichment of Truncating DNVs in the NMD(−) Region of *SEMA6B* in 346 DEE Trios

(A) Correlation between the observed number of synonymous variants in ExAC and the expected rate of synonymous DNVs by the trinucleotide-context-based model at each of 20,042 genes. Whole coding regions (left) and NMD(−) regions (right) were analyzed.

(B) Expected and observed numbers of synonymous, missense, and nonsense DNVs in 20,042 genes in 1,330 trios with variable rare diseases. p values are shown above dots and were calculated as the probability of the observed number or more in a Poisson distribution of λ (mean) of expected number. Black and red dots: expected and observed numbers of DNVs, respectively.Eerror bar: 95% CI of the Poisson distribution.

(C) Expected and observed numbers of synonymous, nonsense, frameshift, and truncating (nonsense+frameshift) DNVs in NMD(−) regions in 346 DEE trios. Statistical testing and figure presentation are as in Figure 1B.

(D) Manhattan plot showing enrichment of truncating DNVs in NMD(−) regions of each gene in 346 DEE trios. Statistical testing is performed as in Figure 1C but is applied to each gene. The red line shows a threshold for Bonferroni-corrected exome-wide statistical significance (0.05/20,042).

(E) Pedigrees of the four families with truncating DNVs in the NMD(−) region of *SEMA6B* from 346 DEE trios and 1,406 independent individuals. Information for the other family from the DDD study was unavailable. The *SEMA6B* genotype is described below each member. WT: wild-type allele.

(F) Locations of truncating DNVs in *SEMA6B* identified in this study or in the "non-neuro" subset of gnomAD in the gene (middle) and encoded protein (lower). Median per-base depth at exons and flanking 20 bp of 100 randomly selected samples sequenced with V5 kitare shown (upper). Domains are shown as gray bars. Abbreviations are as follows: SEMA, Sema domain; PSI, plexin-semaphorin-integrin domain; TM, transmembrane domain; and ICD, intracellular domain.

(A, B, C, and F) Abbreviations are as follows: SYN, synonymous variant; MIS, missense variant; NON, nonsense variant; FS, frameshift indel; and PTV, protein-truncating variant.

the last exon of semaphorin 6B (*SEMA6B*) (MIM: 608873) as a genetic cause of progressive myoclonic epilepsy (PME). Notably, the gene would be missed by conventional enrichment analysis. DNV rates in the NMD(−) region of each gene are available (Tables S1 and S2). This study was approved by the institutional review board of Yokohama City University School of Medicine (Yokohama, Japan). Written informed consent was obtained from all children's parents.

First, we confirmed that the previous trinucleotide-context-based model of DNV rate could predict the DNV rate in NMD(−) regions.[3] Because the DNV rate for a genomic interval under no natural selection is correlated with the number of variant sites in the interval (Figure S1, see Supplemental Note), we intended to show a correlation between the DNV rate expected by the model and the number of variant sites in the NMD(−) region of each gene. After we adjusted the DNV rate predicted by the model for depth,[10] which affects variant detection (Figure S2A, see Supplemental Methods), we observed a high Pearson's correlation coefficient (r = 0.97) between the observed number of synonymous variants and the expected rate of synonymous DNVs in the whole coding

region of each gene (Figure 1A, left), as previously described.[3] Similarly, we observed a high correlation ($r = 0.94$) in NMD(−) regions of each gene (Figure 1A, right). This result indicates that the trinucleotide-context-based model provided accurate relative rates of DNVs among the NMD(−) regions.

Next, we confirmed that the model could predict a total absolute rate of DNVs in the NMD(−) regions of all genes. We performed WES, as previously described,[11] on 1,330 trios in which the parents were healthy and their child showed variable rare diseases. We expected that number of synonymous DNVs observed in the 1,330 trios would be comparable to that expected from the model, whereas numbers of missense or nonsense DNVs would be higher than expected. After we optimized our flow for DNV detection (see Supplemental Note) and adjusted the DNV rate expected from the model for depth (Figure S2B, see Supplemental Methods), we confirmed that the expected and observed numbers of synonymous DNVs were comparable in the 1,330 trios, whereas the observed numbers of missense and nonsense DNVs were higher than expected (p: 0.27 for synonymous, 2.8×10^{-21} for missense, and 5.0×10^{-10} for nonsense, Figure 1B). This was also the case for NMD(−) regions (p: 0.83 for synonymous, 5.7×10^{-6} for missense, and 1.7×10^{-6} for nonsense, Figure 1B). These results indicate that the model predicted a total absolute rate of DNVs in the NMD(−) regions of all genes. Together with the above finding (Figure 1A), these results indicate that the trinucleotide-context-based model predicts an accurate absolute rate of DNVs in the NMD(−) region of each gene.

Using the DNV rate in NMD(−) regions, we performed enrichment analysis of DNVs in NMD(−) regions in 346 trios in which attending physicians diagnosed the child as having DEE and the parents were healthy.[12] After optimizing our workflow of DNV detection (see Supplemental Materials and Methods "Variant filtering and *de novo* calling," Figure S3), we observed statistically significant enrichment for truncating (nonsense and frameshift) DNVs ($p = 8.0 \times 10^{-4}$) in all genes but no enrichment for synonymous DNVs ($p = 0.77$) (Figure 1C). We then performed enrichment analysis for truncating DNVs in the NMD(−) region of each gene (Table S3). One gene, *SEMA6B*, exceeded the Bonferroni-corrected exome-wide statistical significance threshold ($p = 2.8 \times 10^{-8}$; observed number: 2, expected number: 0.000237, Figure 1C). A causal association between *SEMA6B* and DEE has not previously been reported. The observed DNVs were as follows: c.1976_1982del (p.Ala659Valfs*24) and c.1991del (p.Gly664Alafs*21). We then screened for truncating DNVs in the NMD(−) region of *SEMA6B* in WES data of an additional 1,406 independent individuals with DEE. We identified two frameshift DNVs, c.1950_1969dup (p.Arg657Profs*35) and c.1991del, the latter being identical to one of the variants found through the initial analysis of the 346 trios. We also screened for truncating DNVs in the NMD(−) region of *SEMA6B* in the Deciphering Developmental Disorders (DDD) cohort, which comprises 4,293 trios with developmental disorders[2] that often involve epilepsy. This analysis identified another truncating DNV, c.1982_1991del (p.Gly661Alafs*21). Thus, we identified four truncating DNVs in the NMD(−) region of *SEMA6B* in five unrelated families from among 1,752 DEE- (346 + 1,406) and 4,293 DDD-affected individuals. Collectively, the statistical significance was 1.9×10^{-13} when we used a non-depth-adjusted DNV rate (observed number: 5; expected number: 0.0075, see the "Per-gene enrichment analysis of DNVs" section of Supplemental Materials and Methods) because information about the depth of the 4,293 trios in the DDD cohort was unavailable. This statistical analysis using non-depth-adjusted rate is more conservative than that using depth-adjusted rate because the analysis assumes sufficient depth in all the regions analyzed and expects a higher number of DNVs. On the other hand, no DNVs at the NMD(+) region of *SEMA6B* were observed in the 346 DEE and 4,293 DDD trios after filtering DNVs with scoring systems used in each dataset. We confirmed all variants from our four families by Sanger sequencing (Figure 1E); the DNA samples from the one family from the DDD study was unavailable for confirmatory sequencing.

All four variants occurred in the last exon of *SEMA6B* and lead to truncation of the intracellular domain (ICD), the region closer to the C terminus than the transmembrane domain (InterPro database, Figure 1F).[13] Truncating variants in the NMD(−) region of *SEMA6B* were not observed in general populations (i.e., the "non-neuro" subset of gnomAD), except for one at the extreme C terminus, whreas truncating variants were observed in the NMD(+) region (Figure 1F). *SEMA6B* has a low probability of being loss-of-function intolerant (pLI = 0.06). We investigated whether the variants induce NMD in a lymphoblastoid cell line of individual 2 (individual II-1 of family 2, Figure 1E), which was the only cell line available to us among all the individuals. Sanger sequencing of real-time PCR-amplified *SEMA6B* transcripts confirmed a mutant-allele peak in the electropherogram, whose intensity was comparable to that of the wild-type allele (Figure S4), indicating that the DNV was unlikely to induce NMD. Collectively, these results indicate that the pathomechanism of truncating DNVs in the NMD(−) region of *SEMA6B* was most likely dominant-negative or gain-of-function, but not haploinsufficiency.

We retrospectively investigated the phenotypes of the affected individuals (Supplemental Note and Table 1; note that the clinical information available for individual 3, individual 4, and the individual from the DDD study was limited). Their summarized clinical courses were as follows. Developmental milestones were mildly delayed but apparently normal until 2 years. Epilepsy started from the age of 1 year 5 months until about 6 years. Regression started from 2–4 years in individual 3 and individual 4, whereas its precise commencing time in individual 1 and individual 2 was unknown. The affected individuals were

Table 1. Clinical Findings in Four Affected Individuals with Truncating Variants in the NMD(−) Region of *SEMA6B*

	Individual 1	Individual 2	Individual 3	Individual 4
Current age	22 years	28 years	14 years	11 years
Sex	male	female	male	female
Nationality	Japanese	Japanese	Israeli	Malaysian
Mutation	c.1950_1969dup	c.1976_1982del	c.1991del	c.1991del
Protein change	p.Arg657Profs*35	p.Ala659Valfs*24	p.Gly664Alafs*21	p.Gly664Alafs*21
Inheritance	*de novo*	*de novo*	*de novo*	*de novo*
Age at onset	6 years	11 months	2 years	4 years
Development milestones	rolling over: 12 months; meaningful words: 24–36 months	walking without support: 28 months	walking without support: 24 months	eye pursuit, 5 months; walking without support, 24 months; meaningful words, 30 months;
Initial symptom	seizure, developmental delay	seizure	seizure	seizure and developmental delay,
Initial walking	1 year 5 months	2 years 4 months	2 years	2 years,
Intellectual disability	severe (IQ = 25 at 17 years)	severe (IQ = 25 at 12 years)	severe	severe
Language	few words	few words	no words	few words
Microcephaly	−	+ (−2.0 SD)	+ (−2.5 SD)	+ (2^{nd} percentile)
Seizure type at onset	GTCS since 6 years, absence seizure since 9 years, atonic seizure since 11 years	GTCS since 11 months; loss of consciousness with abnormal eye movement since 5 years; complex partial seizure since 10 years; atonic seizure since 10 years,	absence seizure since 2 years	atonic seizure since 4 years
Response to therapy (seizure)	intractable	intractable	responsive	intractable, but improved by clobazam and sulthiame (responsive)
Regression	+ (motor skill and dysarthria)	+ (motor skill)	+ (motor and verbal skills)	+
Ataxia	+	+	+	+
Intention tremor	+	+	+	+
Rigidity	+	+	ND	ND
Myoclonus	+	+	ND	+
Spasticity	+	+	ND	ND
Increased deep tendon reflex	+ (upper and lower limbs)	+ (upper and lower limbs)	ND	−
Pathogenic reflex	+ (Rossolimo sign; positive, Mendel-Bechterew sign; positive)	−	ND	−
Dysmorphic features	−	−	−	−
Motor disturbance	wheelchair	wheelchair	wheelchair	walking with support
Brain MRI	normal	mild cerebellar atrophy;	small vermis	normal
EEG	abnormal discharge in right hemisphere (6 years); burst of diffuse irregular spikes and slow waves (9 years); diffuse spike and slow waves in frontal, parietal and temporal regions (14 years)	diffuse slow wave with 2–3 Hz and spike-and-wave in bilateral frontal region (3 years and 4 years); diffuse theta waves with 4–5 Hz and spike-and-wave burst with 2–3 Hz (9 years); multifocal spikes in left parietal region and bilateral frontal regions (12 years); multispikes in left occipital region (13 years); slow waves at baselines (23 years)	abnormal background activity (1 year); slow abnormal sleep features with paucity of sleep spindles (13 years)	focal bifrontal epileptiform discharges accentuated during sleep (4 years); frequent frontocentral discharges during awake state (5 years); frequent intermittent slow spikes in right posterior region (11 years)

(*Continued on next page*)

Table 1. ***Continued***

	Individual 1	Individual 2	Individual 3	Individual 4
SEP	prolonged N20 latency and high amplitude of P24-N33	giant SEP	ND	ND
Other findings	ND	SLE	ND	ND

SEMA6B variants are based on GenBank: NM_032108.4. Abbreviations are as follows: EEG, electroencephalogram; GTCS, generalized tonic-clonic seizures; ND, not described; MRI, magnetic resonance imaging; SEP, somatosensory evoked potential; SLE, systemic lupus erythematosus; +, present; and −, not present.

wheelchair bound or only walked with support at 10–14 years. They could communicate with a few words—perhapstwo-word sentences—or could not communicate at 14–28 years. The types of epilepsy were generalized tonic-clonic seizure (GTCS) in individual 1 and individual 2, absence seizure in individual 1 and individual 3, and atonic seizure in individual 1, individual 2, and individual 4. Microcephaly was recognized in individual 2, individual 3, and individual 4. Additional neurological findings included extrapyramidal symptoms (rigidity and/or myoclonus) in individual 1, individual 2, and individual 4; pyramidal-tract signs (spasticity, increased deep tendon reflex, and/or pathological reflex) in individual 1 and individual 2; and cerebellar findings (ataxia, intention tremor, and/or mild cerebellar atrophy on MRI) in all four individuals. Somatosensory evoked potential (SEP) tests showed high amplitudes of P24 and N33 electrical potentials in individual 1 and giant SEP in individual 2. The findings such as myoclonus, epileptic seizures, and progressive neurologic decline indicated that the affected individuals had PME with phenotypic consistency, indicating a shared etiology among the individuals.[14] Clinically, pyramidal-tract signs indicate cerebral cortex layer 5 (L5) excitatory neurons, abnormal SEP[15,16] indicates the primary somatosensory cortex, and cerebellar findings indicate variable types of cerebellar neurons.[17]

Next, we investigated relationships between the observed phenotypes and the expression patterns of *SEMA6B* across different brain regions. We analyzed *Sema6b in situ* hybridization (ISH) data in the post-natal day (P)56, P14, and P4 and embryonic day (E)18.5 mouse brain in the Allen Brain Atlas.[18] *Sema6b* was expressed broadly in the brain at P56 and P14 (Figure S5A, left). All layers of the cerebral cortex and the Purkinje cell layer of the cerebellum (Purkinje cells and/or Bergmann glia) were stained (Figure S5A, middle and right). At P4, *Sema6b* was expressed in L5 and L6 of the frontal region (Figure S5A, P4, left) and in the cortical plate that will form L2 to L4 in later developmental stages (Figure S5A, P4). In the cerebellum, the Purkinje cell layer (migrating Purkinje cells, Bergmann glia, interneurons, and/or granule cells) was stained. At E18.5, *Sema6b* was expressed in the subplate and cortical plate, which forms L2 to L6 in later stages (Figure S5A, E18.5). In the E18.5 cerebellum, the outer part of Purkinje cell clusters (migrating Purkinje cells, Bergmann glia, interneurons, and/or granule cells) was stained (within the black dotted outline, Figure S5A, E18.5). We also analyzed two single-cell RNA sequencing (scRNA-seq) datasets: one from P2 and P11 mouse brains[19] and the other from P12–P60 mouse brains[20] (Figure S5B, Tables S4 and S5). The cell type was determined in the original analyses,[19,20] and additionally, we confirmed this by using Allen Brain Atlas ISH data (Table S4 and S5).[18] In P12–P60 mouse brains, *Sema6b* was expressed in excitatory neurons in the cerebral cortex and in cerebellar Purkinje cells (Figure S5B, left). P2 and P11 mouse brains showed a similar pattern (Figure S5B, right). We also analyzed *SEMA6B* expression in human brain tissue: bulk RNA-seq data of variable brain regions at variable developmental stages in the BrainSpan Atlas of the Developing Human Brain and single-nucleus RNA-seq (snRNA-seq) of variable regions of the adult brain in the Allen Brain Map. The bulk RNA-seq showed broad expression of *SEMA6B* at variable brain regions across variable developmental stages, especially at infantile periods (Figure S6A). The snRNA-seq showed stronger expression in neuronal cells, especially excitatory neurons, than non-neuronal cells (Figure S6B). These results collectively indicate that *SEMA6B* is expressed in excitatory neurons of various layers, including L5 of the cerebral cortex and possibly Purkinje cells, consistent with the phenotypes described above.

To determine whether the truncation of *SEMA6B* in the NMD(−) region affects neurological function *in vivo*, we created a zebrafish model to investigate the pathogenicity of the truncating human variants. The zebrafish genome contains two orthologs of human *SEMA6B*, *sema6ba* and *sema6bb*, which are both expressed in neuronal cells at 24 to 72 h post-fertilization (hpf), when neurons are migrating and neuronal connections are forming.[21] To introduce truncation of *sema6ba* and *sema6bb* by nonhomologous end joining (NHEJ)-based genome editing in each of the two paralogs, we employed the CRISPR-Cas9 system and analyzed the first generation of F0 crispants. We generated CRISPR RNAs (crRNAs) targeting the last coding exon of *sema6ba* (*sema6ba*-ex.17) or *sema6bb* (*sema6bb*-ex.17). We also generated crRNAs targeting the sequences encoding the first methionine in exon 2 of *sema6ba* (*sema6ba*-ex.2) or *sema6bb* (*sema6bb*-ex.2) to create loss-of-function mutations (Figure S7A, Table S6). T7 endonuclease I (T7E1) assays and subcloning of PCR products with a genomic DNA template derived from five mixed embryos showed >63% mosaicism (Figure S7B and S7D). We also confirmed the editing efficiency in single F0 crispants by heteroduplex mobility assay (Figure S7C). Sanger sequencing confirmed deletion at the ICD. The expression

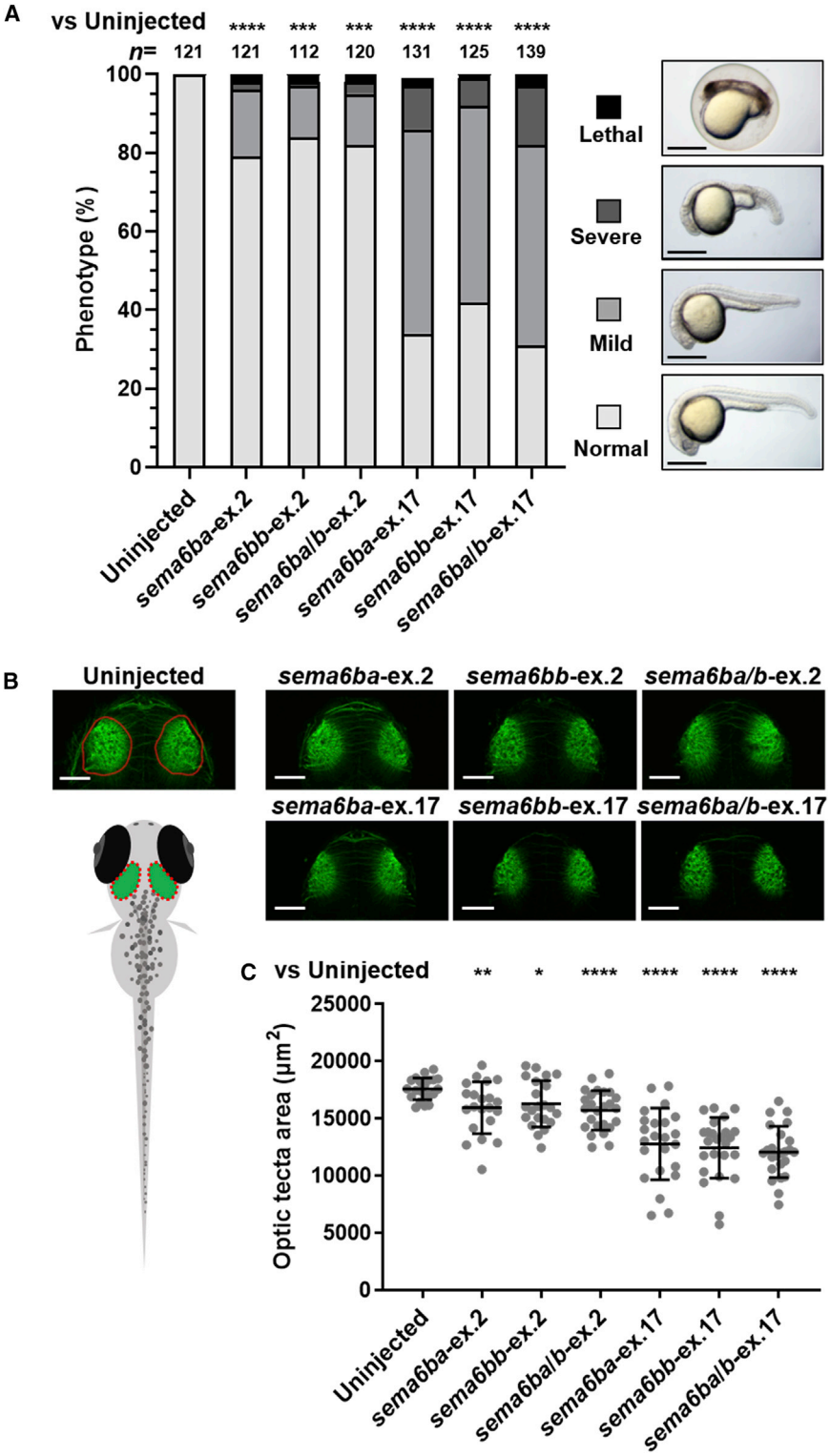

Figure 2. Disruption of Zebrafish *sema6ba* and *sema6bb* 3′ Regions Results in CNS Defects

(A) Frequency of developmental defects and CNS abnormalities in embryos at 24 hpf. Zebrafish embryos were injected with the following crRNAs: *sema6ba*-ex.2, *sema6bb*-ex.2, *sema6ba/b*-ex.2, *sema6ba*-ex.17, *sema6bb*-ex.17, or *sema6ba/b*-ex.17, all of which were duplexed with tracrRNA and Cas9 protein. Uninjected embryos and F0 crispants were scored on the basis of morphological anomalies; lethal (cell death distributed over the entire body), severe (smaller head with shortened body axis), mild (developmental delay with smaller head), and normal. (n = 112–139 in each group). All experiments were performed three times. Scale bar, 500 μm. ***p < 0.0005, ****p < 0.0001 by the chi-square test.

(B) Acetylated tubulin staining of uninjected embryo and F0 crispants shows brain structures at 2.5 dpf (dorsal view). Schematic representation of neuroanatomical structures shows the area of the optic tectum (red dotted oval).

(C) F0 crispants with *sema6ba*-ex.17, *sema6bb*-ex.17, and *sema6ba/b*-ex.17 displayed a smaller optic tectum than uninjected embryos or *sema6ba*-ex.2, *sema6bb*-ex.2, or *sema6ba/b*-ex.2 F0 crispants. Scale bar, 100 μm. Data are represented as the mean ± SD. *p < 0.05, **p < 0.01, ***p < 0.0005, ****p < 0.0001 by Student's t test.

levels of *sema6ba* and *sema6bb* transcripts were relatively decreased in crispants injected with *sema6ba/b*-ex.2 but not in those injected with *sema6ba/b*-ex.17 (Figure S8). Compared with *sema6ba*-ex.2-, *sema6bb*-ex.2-, and *sema6ba/b*-ex.2-injected embryos, the *sema6ba*-ex.17-, *sema6bb*-ex.17-, and *sema6ba/b*-ex.17-injected embryos presented with more severe developmental defects, such as small head and short body (Figure 2A). Immunostaining of the central nervous system (CNS) with an anti-acetylated tubulin antibody (an axonal marker) and measurement of neuroanatomical structures showed a smaller optic tectum, the area between the eyes as a surrogate for brain size, in *sema6ba*-ex.17-, *sema6bb*-ex.17-, and *sema6ba/b*-ex.17-injected embryos compared with either control embryos or embryos injected with *sema6ba*-ex.2, *sema6bb*-ex.2, or *sema6ba/b*-ex.2 (Figure 2B). Behavioral manifestations of pentylenetetrazole (PTZ)-induced electrographic seizures, including episodes of excessive locomotor activity and clonus-like behavior, are well characterized in zebrafish models of epilepsy.[22–24] We tracked and quantified total swimming distance in response to PTZ exposure (Figures 3A and 3B). Uninjected control larvae and larvae injected with *sema6ba/b*-ex.2 or *sema6ba/b*-ex.17 7 days post fertilization (dpf), in which most of the main neuronal clusters and axon tracts had been formed, were exposed to PTZ at concentrations of 2.5, 5, or 15 mM.

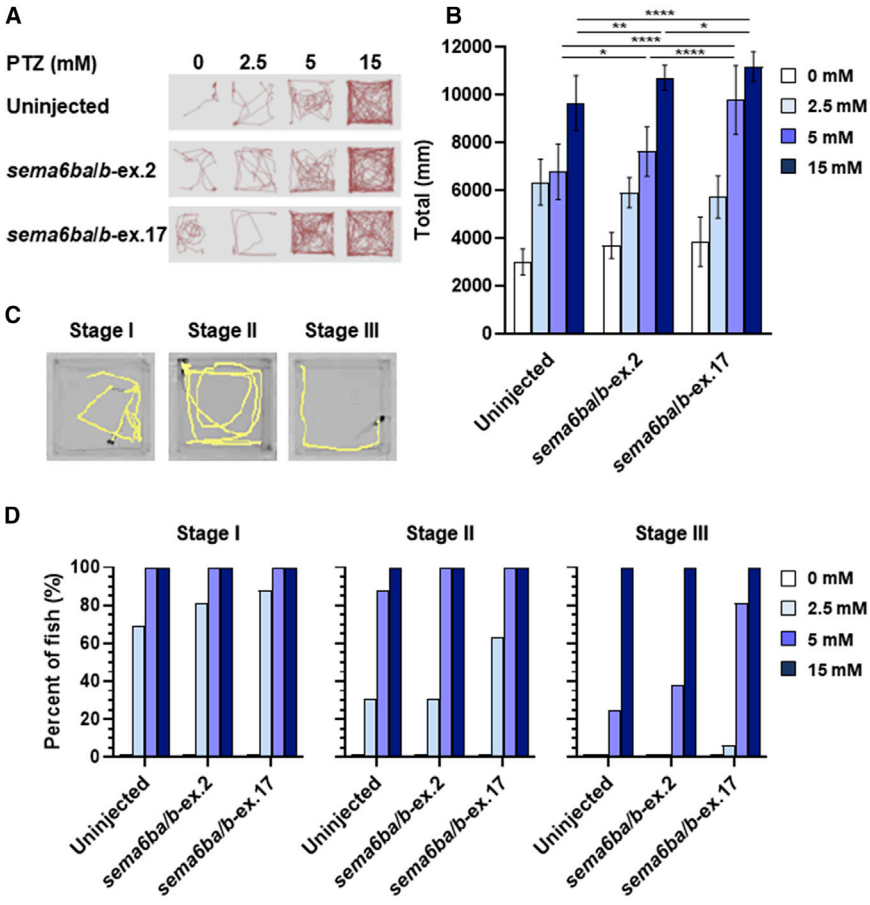

Figure 3. Behavioral Response Following PTZ Exposure in *sema6ba* and *sema6bb* F0 Crispants

(A) Representative trajectory plots are shown, with the tracking from one 7 dpf larva per well, during exposure to PTZ for 3 min (PTZ concentrations of 0, 2.5, 5, and 15 mM).

(B) Quantification of the swimming distance of 7 dpf larvae during 15 min after treatment with different concentrations of PTZ. Data are represented as the mean ± SD. *$p < 0.05$, **$p < 0.01$, **** $p < 0.0001$ by Student's t test.

(C) Sample locomotion tracking plots of individual 7 dpf larva. Plots are shown for typical seizure behaviors. A PTZ-induced seizure is characterized in three stages: stage I (increased swim activity), stage II (increased swim activity with circular motion), and stage III (convulsions).

(D) Percentages of 7 dpf larvae in different behavioral categories (stage I, stage II, stage III) with exposure to 2.5, 5, and 15 mM PTZ and observation for 15 min.

Seizure behavior was not observed in any larvae that were not treated with PTZ (Figures 3A and 3B, Video S1). Total swimming distance increased in a dose-dependent manner in controls and crispants (Figures 3A and 3B, Video S1). It should be noted that incubation with 5 mM PTZ produced hyperactive responses in *sema6ba*/*b*-ex.17-injected larvae compared with *sema6ba*/*b*-ex.2-injected or uninjected larvae (Figures 3A and 3B, Video S1). PTZ exposure increases larvae swimming activity (stage I), causes rapid "whirlpool" motion circular swimming (stage II), and results in loss of posture and activity for 1 to 3 s in a clonus-like seizure (stage III) (Figure 3C).[22] Lower concentrations of PTZ (e.g., 2.5 mM) produced stage I and stage II behavior, but all larvae exhibited clonus-like convulsion (stage III) in the presence of high PTZ concentrations (15 mM) (Figure 3D, Video S1). Exposure to 2.5 mM PTZ produced stage II but not stage III behavior in 31% of uninjected or *sema6ba*/*b*-ex.2-injected larvae. However, 5 mM PTZ exposure produced stage III behavior in 25%, 38%, and 81% of uninjected, *sema6ba*/*b*-ex.2-injected, or *sema6ba*/*b*-ex.17-injected larvae, respectively. These results indicated that larvae with 3′ truncations in *sema6ba* and *sema6bb* had increased sensitivity to PTZ and exhibited a stage III (clonus-like) behavior when exposed to 5 mM PTZ (Figure 3D).

In this study, we performed enrichment analysis for truncating DNVs in NMD(−) regions. The p value of the enrichment analysis for *SEMA6B* was 2.8×10^{-8} (expected number in 346 trios: 0.000237), whereas the p value for the whole coding region of *SEMA6B* was 5.0×10^{-6} (expected number in 346 trios: 0.00316) (exome-wide p value threshold: 2.5×10^{-6}). The large difference was partly due to the low depth of the last two *SEMA6B* exons (Figure 1F). Thus, we demonstrated that enrichment analysis targeting specific regions can identify novel causative genes that are missed by conventional analysis of whole coding regions. This framework can be applied to other coding regions for other types of variants. For example, some coding regions are constrained for missense variants, and missense variants at such regions are likely to cause genetic diseases.[9] Enrichment analysis of missense DNVs specific to constrained coding regions theoretically has higher statistical power and could identify novel causative genes.

The theoretical mutation rates used in this and other studies have some limitation in accuracy for frameshift variants. To calculate the rates of frameshift DNVs per gene, the studies multiplied rates of nonsense variants per gene by the general ratio (1.25 or 1.07) of very rare (or *de novo*) frameshift to nonsense variants in all genes.[2,3,5] The calculation ignores different sequence context in different genes, and the way to calculate the sequence-context-aware rate of frameshift DNVs should be developed in the near future.

To investigate the effects of the truncated protein *in vivo*, we generated a zebrafish model with *sema6ba* and *sama6bb* CRISPR-Cas9-mediated mutations. The *sema6ba* and *sema6bb* ICD-truncated F0 crispants showed developmental defects and CNS anomalies consistent with the

affected individuals' phenotypes. In contrast, F0 *sema6ba* and *sama6bb* loss-of-function crispants showed a very mild phenotype. ICD truncation, therefore, produces more significant CNS anomalies than the loss-of-function in a zebrafish model. To generate a seizure model, we exposed zebrafish larvae to PTZ, a gamma-aminobutyric acid (GABA) type A receptor channel blocker that blocks fast synaptic inhibition. Members of the class 4 semaphorin family are important regulators of both glutamatergic and GABAergic synapse development.[25,26] Furthermore, *Sema6A* contributes to GABAergic interneuron migration during brain development, and its disruption might be an underlying cause of autism spectrum disorder (ASD).[27] In addition, this pathophysiological mechanism might be shared with epilepsy.[27,28] Consistent with this hypothesis, *SEMA6B* was expressed in GABAergeic neurons (Figure S6B), and disruption of *SEMA6B* function in GABAergic neurons might contribute to epilepsy in humans. Our zebrafish *sema6ba*/*b*-ex.17-injected crispants demonstrated increased PTZ-mediated hyperactive responses compared with the response of the *sema6ba*/*b*-ex.2-injected larvae. These results support the pathogenic effects of the *sema6ba* and *sema6bb* ICD-truncated proteins *in vivo*. The results are consistent with those in a chick model, in which a truncated version of the *SEMA6B* homolog lacking the ICD was unable to rescue a loss-of-function phenotype involving abnormal commissural axon guidance.[29] Our analysis using the first generation of F0 crispants can be effective, but its limitations must also be considered. Because F0 crispants are mosaic, to more closely resemble the pathological conditions of the NMD(−) variants in humans, we would need to establish F1 generations and perform the correct crosses of zebrafish. Although further work using *in vivo* or *in vitro SEMA6B* models is needed to elucidate the functional mechanism of ICD truncation in PME, these results indicate an important role for *SEMA6B* in the pathogenesis of neurological disorders.

What is the pathomechanism of truncating variants at NMD(−) regions of *SEMA6B*? Genes depleted for truncating variants at NMD(−) regions in general populations tended to encode proteins with domains for oligomerization (but did so statistically insignificantly) (Supplemental Materials and Methods, Table S7 and S8). Therefore, when such genes have truncating variants at NMD(−) regions in diseases, the truncated proteins might exert their pathogenicity via oligomerization. Similarly, class 6 semaphorin proteins (Sema6) form homodimers via their sema domain at the plasma membrane and interact with other transmembrane proteins in a juxtacrine manner.[30] The juxtacrine interaction leads to inward signaling (called reverse signaling) via interaction of the ICD of semaphorin proteins with other proteins.[31,32] Without its ICD, Sema6A did not exert the reverse signaling in mice, and ectopic expression of Sema6B without its ICD resulted in defective reverse signaling in a dominant-negative manner in chickens.[29,32] The reverse signaling of semaphorin proteins has versatile functions, including targeting of axons and dendrites, synapse formation, and cell migration.[33–36] Thus, SEMA6B without ICD might impair such functions of reverse signaling.

In this study, we demonstrated (1) strong statistical enrichment of DNVs in the NMD(−) region of *SEMA6B* in DEE-affected individuals, (2) multiple individuals in general populations had heterozygous truncating variants in the NMD(+) region of *SEMA6B*, (3) the spatial expression pattern of *SEMA6B* among various neuron types was consistent with the individuals' phenotype, and (4) zebrafish models expressing variants with truncations in the NMD(−) region in *SEMA6B* orthologs showed neurological impairments, whereas (5) zebrafish models expressing variants in the NMD(+) region produced a very mild phenotype. From these results, we demonstrated that truncating DNVs in the NMD(−) region of *SEMA6B* cause PME.

Supplemental Data

Supplemental Data can be found online at https://doi.org/10.1016/j.ajhg.2020.02.011.

Acknowledgments

We thank all of the participants for their cooperation in this research. We also thank Ms. K. Takabe, Mr. T. Miyama, Ms. N. Watanabe, Ms. M. Sato, Mr. S. Nakamura, and Ms. S. Sugimoto at the Department of Human Genetics, Yokohama City University Graduate School of Medicine, for their technical assistance. We are also grateful to Edanz (www.edanzediting.co.jp) for editing a draft of this manuscript. This work was supported by AMED under grant numbers JP19ek0109280, JP19dm0107090, JP19ek0109301, JP19ek0109348, and JP19kk020515 (N.M.); JSPS KAKENHI grant numbers JP17H01539 (N.M.), JP16H05160 (H.S.), JP16H05357 (N.M.), JP16H06254 (A.T.), JP15K10367 (M.N.), JP19H03621 (S.M.), JP17K15630 (T.M.), JP17H06994 (A.F.), and JP19K16921 (E.K.); and grants from the Ministry of Health, Labor and Welfare (N.M.), the Takeda Science Foundation (N.M., H.S., and N.M.), and The Ichiro Kanehara Foundation for the Promotion of Medical Science & Medical Care (S.M.).

Declaration of Interests

The authors declare no competing interests.

Received: November 20, 2019
Accepted: February 19, 2020
Published: March 12, 2020

Web Resources

Allen Brain Atlas, https://portal.brain-map.org/
CRISPRdirect, https://crispr.dbcls.jp/
dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/
DenovoFilter, https://github.com/jeremymcrae/denovoFilter
Denovogear, https://github.com/ultimatesource/denovogear
DNMfilter, https://github.com/yongzhuang/DNMFilter
ExAC browser, http://exac.broadinstitute.org/
Genome Analysis Toolkit, https://software.broadinstitute.org/gatk/

gnomAD, https://gnomad.broadinstitute.org/
InterPro, https://www.ebi.ac.uk/interpro/
NHLBI Exome Sequencing Project, http://exac.broadinstitute.org
Novoalign 3.00, http://www.novocraft.com
OMIM, http://www.omim.orgPicard software, http://broadinstitute.github.io/picard/
Seurat, https://satijalab.org/seurat/install.html
snpEff, http://snpeff.sourceforge.net/
Triodenovo, https://genome.sph.umich.edu/wiki/Triodenovo

References

1. Coe, B.P., Stessman, H.A.F., Sulovari, A., Geisheker, M.R., Bakken, T.E., Lake, A.M., Dougherty, J.D., Lein, E.S., Hormozdiari, F., Bernier, R.A., and Eichler, E.E. (2019). Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. Nat. Genet. *51*, 106–116.
2. McRae, J., Clayton, S., Fitzgerald, T., Kaplanis, J., Prigmore, E., Rajan, D., Sifrim, A., Aitken, S., Akawi, N., Alvi, M., et al.; Deciphering Developmental Disorders Study (2017). Prevalence and architecture of de novo mutations in developmental disorders. Nature *542*, 433–438.
3. Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., McGrath, L.M., Kosmicki, J.A., Rehnström, K., Mallick, S., Kirby, A., et al. (2014). A framework for the interpretation of de novo mutation in human disease. Nat. Genet. *46*, 944–950.
4. Lord, J., Gallone, G., Short, P.J., McRae, J.F., Ironfield, H., Wynn, E.H., Gerety, S.S., He, L., Kerr, B., Johnson, D.S., et al.; Deciphering Developmental Disorders study (2019). Pathogenicity and selective constraint on variation near splice sites. Genome Res. *29*, 159–170.
5. Fitzgerald, T., Gerety, S., Jones, W., van Kogelenberg, M., King, D., McRae, J., Morley, K., Parthiban, V., Al-Turki, S., Ambridge, K., et al.; Deciphering Developmental Disorders Study (2015). Large-scale discovery of novel genetic causes of developmental disorders. Nature *519*, 223–228.
6. Short, P.J., McRae, J.F., Gallone, G., Sifrim, A., Won, H., Geschwind, D.H., Wright, C.F., Firth, H.V., FitzPatrick, D.R., Barrett, J.C., and Hurles, M.E. (2018). De novo mutations in regulatory elements in neurodevelopmental disorders. Nature *555*, 611–616.
7. Jansen, S., Geuer, S., Pfundt, R., Brough, R., Ghongane, P., Herkert, J.C., Marco, E.J., Willemsen, M.H., Kleefstra, T., Hannibal, M., et al.; Deciphering Developmental Disorders Study (2017). De Novo Truncating Mutations in the Last and Penultimate Exons of PPM1D Cause an Intellectual Disability Syndrome. Am. J. Hum. Genet. *100*, 650–658.
8. White, J., Mazzeu, J.F., Hoischen, A., Jhangiani, S.N., Gambin, T., Alcino, M.C., Penney, S., Saraiva, J.M., Hove, H., Skovby, F., et al.; Baylor-Hopkins Center for Mendelian Genomics (2015). DVL1 frameshift mutations clustering in the penultimate exon cause autosomal-dominant Robinow syndrome. Am. J. Hum. Genet. *96*, 612–622.
9. Havrilla, J.M., Pedersen, B.S., Layer, R.M., and Quinlan, A.R. (2019). A map of constrained coding regions in the human genome. Nat. Genet. *51*, 88–95.
10. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.
11. Hamanaka, K., Miyatake, S., Zerem, A., Lev, D., Blumkin, L., Yokochi, K., Fujita, A., Imagawa, E., Iwama, K., Nakashima, M., et al. (2018). Expanding the phenotype of IBA57 mutations: related leukodystrophy can remain asymptomatic. J. Hum. Genet. *63*, 1223–1229.
12. El Kosseifi, C., Cornet, M.C., and Cilio, M.R. (2019). Neonatal Developmental and Epileptic Encephalopathies. Semin. Pediatr. Neurol. *32*, 100770.
13. Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., et al. (2009). InterPro: the integrative protein signature database. Nucleic Acids Res. *37*, D211–D215.
14. Kälviäinen, R. (2015). Progressive Myoclonus Epilepsies. Semin. Neurol. *35*, 293–299.
15. Shibasaki, H., Yamashita, Y., Neshige, R., Tobimatsu, S., and Fukui, R. (1985). Pathogenesis of giant somatosensory evoked potentials in progressive myoclonic epilepsy. Brain *108*, 225–240.
16. Hitomi, T., Ikeda, A., Kondo, T., Imamura, H., Inouchi, M., Matsumoto, R., Terada, K., Kanda, M., Matsuhashi, M., Nagamine, T., et al. (2011). Increased cortical hyperexcitability and exaggerated myoclonus with aging in benign adult familial myoclonus epilepsy. Mov. Disord. *26*, 1509–1514.
17. Hoxha, E., Balbo, I., Miniaci, M.C., and Tempia, F. (2018). Purkinje Cell Signaling Deficits in Animal Models of Ataxia. Front. Synaptic Neurosci. *10*, 6.
18. Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. Nature *445*, 168–176.
19. Rosenberg, A.B., Roco, C.M., Muscat, R.A., Kuchina, A., Sample, P., Yao, Z., Graybuck, L.T., Peeler, D.J., Mukherjee, S., Chen, W., et al. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science *360*, 176–182.
20. Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Häring, M., Braun, E., Borm, L.E., La Manno, G., et al. (2018). Molecular Architecture of the Mouse Nervous System. Cell *174*, 999–1014.e22.
21. Ebert, A.M., Lamont, R.E., Childs, S.J., and McFarlane, S. (2012). Neuronal expression of class 6 semaphorins in zebrafish. Gene Expr. Patterns *12*, 117–122.
22. Baraban, S.C., Taylor, M.R., Castro, P.A., and Baier, H. (2005). Pentylenetetrazole induced changes in zebrafish behavior, neural activity and c-fos expression. Neuroscience *131*, 759–768.
23. Afrikanova, T., Serruys, A.S., Buenafe, O.E., Clinckers, R., Smolders, I., de Witte, P.A., Crawford, A.D., and Esguerra, C.V. (2013). Validation of the zebrafish pentylenetetrazol seizure model: locomotor versus electrographic responses to antiepileptic drugs. PLoS ONE *8*, e54166.
24. Wang, K., Chen, X., Liu, J., Zou, L.P., Feng, W., Cai, L., Wu, X., and Chen, S.Y. (2018). Embryonic exposure to ethanol increases the susceptibility of larval zebrafish to chemically induced seizures. Sci. Rep. *8*, 1845.
25. Pasterkamp, R.J., and Giger, R.J. (2009). Semaphorin function in neural plasticity and disease. Curr. Opin. Neurobiol. *19*, 263–274.
26. Paradis, S., Harrar, D.B., Lin, Y., Koon, A.C., Hauser, J.L., Griffith, E.C., Zhu, L., Brass, L.F., Chen, C., and Greenberg, M.E. (2007). An RNAi-based approach identifies molecules required

for glutamatergic and GABAergic synapse development. Neuron *53*, 217–232.

27. Karlie Menzel, G.S. Yuchio Yanagawa, Turhan Cocksaygan, Céline Plachez. (2019). GABAergic cell loss in mice lacking autism-associated gene Sema6A. bioRxiv. 10.1101/663419.
28. Håkansson, K., Runker, A.E., O'Sullivan, G.J., Mitchell, K.J., Waddington, J.L., and O'Tuathaigh, C.M. (2017). Semaphorin 6A knockout mice display abnormalities across ethologically-based topographies of exploration and in motor learning. Neurosci. Lett. *641*, 70–76.
29. Andermatt, I., Wilson, N.H., Bergmann, T., Mauti, O., Gesemann, M., Sockanathan, S., and Stoeckli, E.T. (2014). Semaphorin 6B acts as a receptor in post-crossing commissural axon guidance. Development *141*, 3709–3720.
30. Jones, E.Y. (2015). Understanding cell signalling systems: paving the way for new therapies. Philos Trans A Math Phys Eng Sci *373*, 20130155.
31. Battistini, C., and Tamagnone, L. (2016). Transmembrane semaphorins, forward and reverse signaling: have a look both ways. Cell. Mol. Life Sci. *73*, 1609–1622.
32. Perez-Branguli, F., Zagar, Y., Shanley, D.K., Graef, I.A., Chédotal, A., and Mitchell, K.J. (2016). Reverse Signaling by Semaphorin-6A Regulates Cellular Aggregation and Neuronal Morphology. PLoS ONE *11*, e0158686.
33. Godenschwege, T.A., Hu, H., Shan-Crofts, X., Goodman, C.S., and Murphey, R.K. (2002). Bi-directional signaling by Semaphorin 1a during central synapse formation in Drosophila. Nat. Neurosci. *5*, 1294–1301.
34. Jeong, S., Juhaszova, K., and Kolodkin, A.L. (2012). The Control of semaphorin-1a-mediated reverse signaling by opposing pebble and RhoGAPp190 functions in drosophila. Neuron *76*, 721–734.
35. Komiyama, T., Sweeney, L.B., Schuldiner, O., Garcia, K.C., and Luo, L. (2007). Graded expression of semaphorin-1a cell-autonomously directs dendritic targeting of olfactory projection neurons. Cell *128*, 399–410.
36. Toyofuku, T., Zhang, H., Kumanogoh, A., Takegahara, N., Yabuki, M., Harada, K., Hori, M., and Kikutani, H. (2004). Guidance of myocardial patterning in cardiac development by Sema6D reverse signalling. Nat. Cell Biol. *6*, 1204–1211.