De novo mutations in epileptic encephalopathies

Epileptic encephalopathies are a devastating group of severe childhood epilepsy disorders for which the cause is often unknown. Here we report a screen for de novo mutations in patients with two classical epileptic encephalopathies: infantile spasms (n = 149) and Lennox–Gastaut syndrome (n = 115). We sequenced the exomes of 264 probands, and their parents, and confirmed 329 de novo mutations. A likelihood analysis showed a significant excess of de novo mutations in the ~4,000 genes that are the most intolerant to functional genetic variation in the human population (P = 2.9 × 10^-8). Among these are GABRB3, with de novo mutations in four patients, and ALG13, with the same de novo mutation in two patients; both genes show clear statistical evidence of association with epileptic encephalopathy. Given the relevant site-specific mutation rates, the probabilities of these outcomes occurring by chance are P = 4.1 × 10^-10 and P = 7.8 × 10^-12, respectively. Other genes with de novo mutations in this cohort include CACNA1A, CHD2, FLNA, GABRA1, GRIN1, GRIN2B, HNRNPU, IQSEC2, MTOR and NEDD4L. Finally, we show that the de novo mutations observed are enriched in specific gene sets including genes regulated by the fragile X protein (P < 10^-8), as has been reported previously for autism spectrum disorders.

Genetics is believed to have an important role in many epilepsy syndromes; however, specific genes have been discovered in only a small proportion of cases. Genome-wide association studies for both local and generalized epilepsies have revealed few significant associations, and rare copy number variants explain only a few per cent of cases of epilepsy. An emerging paradigm in neuropsychiatric disorders is the major impact de novo mutations have on disease risk. We searched for de novo mutations associated with epileptic encephalopathies, a heterogeneous group of severe epilepsy disorders characterized by early onset of seizures with cognitive and behavioural features associated with ongoing epileptic activity. We focused on two ‘classical’ forms of epileptic encephalopathies: infantile spasms and Lennox–Gastaut syndrome, recognizing that some patients with infantile spasms progress to Lennox–Gastaut syndrome.

Exome sequencing of 264 trios (Methods) identified 439 putative de novo mutations. Sanger sequencing confirmed 329 de novo mutations (Supplementary Table 2), and the remainder were either false positives, a result of B-cell immortalization, or in regions where the Sanger assays did not work (Supplementary Table 3).

Across our 264 trios, we found nine genes with de novo single nucleotide variant (SNV) mutations in two or more probands (SCN1A, n = 7; STXBP1, n = 5; GABRB3, n = 4; CDLK5, n = 3; SCN8A, n = 2; SCN2A, n = 2; ALG13, n = 2; DNMT1, n = 2; and HDAC4, n = 2). Of these, SCN1A, STXBP1, SCN8A, SCN2A and CDLK5 are genes that have a previously established association with epileptic encephalopathy. To assess whether the observations in the other genes implicate them as risk factors for epileptic encephalopathies, we determined the probability of seeing multiple mutations in the same gene given the sequence-specific mutation rate, size of the gene, and the number and gender of patients evaluated in this study (Methods). The number of observed de novo mutations in HDAC4 and DNMT1 are not yet significantly greater than the null expectation. However, observing four unique de novo mutations in GABRB3 and two identical de novo mutations in ALG13 were found to be highly improbable (Table 1 and Fig. 1). We performed the same calculations on all of the genes with multiple de novo mutations observed in 610 control trios and found no genes with a significant excess of de novo mutations (Supplementary Table 4). Although mutations in GABRB3 have previously been reported in association with another type of epilepsy, and in vivo mouse studies suggest that GABRB3 haplosufficiency is one of the causes of epilepsy in Angelman’s syndrome, our observations implicate it, for the first time, as a single-gene cause of epileptic encephalopathies and provide the strongest evidence to date for its association with any epilepsy. Likewise, ALG13, an X-linked gene encoding a subunit of the uridine diphosphate-N-acetylgalactosamine transferase, was previously shown to carry a novel de novo mutation in a male patient with a severe congenital glycosylation disorder with microcephaly, seizures and early lethality. Furthermore, the same ALG13 de novo mutation identified in this study was observed as a de novo mutation in an additional female patient with severe intellectual disability and seizures.

Each trio harboured on average 1.25 confirmed de novo mutations, with 181 probands harbouring at least one. Considering only de novo SNVs, each trio harboured on average 1.17 de novo mutations (Supplementary Fig. 1). Seventy-two per cent of the confirmed de novo SNV mutations were missense and 7.5% were putative loss-of-function (splice donor, splice acceptor, or stop-gain mutations). Compared to rates of these classes of mutations previously reported in controls (69.4% missense and 4.2% putative loss-of-function mutations), we observed a significant excess of loss-of-function mutations in patients with infantile spasms and Lennox–Gastaut syndrome (exact binomial P = 0.01), consistent with data previously reported in autism spectrum disorder.

A framework was recently established for testing whether the distribution of de novo mutations in affected individuals differs from the general population. Here, we extend the simulation-based approach of ref. 8 by developing a likelihood model that characterizes this effect and estimates the distribution of de novo mutations among affected individuals in terms of the distribution in the general population, and a set of parameters describing the genetic architecture of the disease. These parameters include the proportion of the exome sequence that can carry disease-influencing mutations (η) and the relative risk (γ) of the mutations (Supplementary Methods). Consistent with what was reported in autism spectrum disorder, we found no significant deviation in the overall distribution of mutations from expected (γ = 1 and/or η = 0). It is, however, now well established that some genes tolerate protein-disrupting mutations without apparent adverse phenotypic consequences, whereas others do not. To take this into account, we used a simple scoring system that uses polymorphism data in the human population to assign a tolerance score to every considered gene (Methods). We then found that genes with a known association with epileptic encephalopathy rank among the most intolerant genes using this scheme (Supplementary Table 8). We therefore evaluated the distribution of de novo mutations within these 4,264 genes that are within the 25th percentile for intolerance and found a significant shift from the null distribution (P = 2.9 × 10^-3). The maximum likelihood estimates of η (percentage of intolerant genes involved in epileptic encephalopathies) was 0.021 and γ (relative risk) was 81, indicating that there are 90 genes among the intolerant genes...
that can confer risk of epileptic encephalopathies and that each mutation carries substantial risk. We also found that putatively damaging de novo variants in our cohort are significantly enriched in intolerant genes compared with control cohorts (Supplementary Methods).

We next evaluated whether the de novo mutations were drawn preferentially from six gene sets (Methods and Supplementary Table 10), including ion channels, genes known to cause monogenic disorders with seizures as a phenotypic feature, genes carrying confirmed de novo mutations in patients with autism spectrum disorder and in patients with intellectual disability and FMRI-regulated genes. Taking into account the size of regions with adequate sequencing coverage to detect a de novo mutation (Methods), we found significant over-representation for all gene lists in our data (Supplementary Table 10), and no over-representation in control genes.

To determine possible interconnectivity among the genes carrying a de novo mutation, we performed a protein–protein interaction analysis and identified a single network of 71 connected proteins (Fig. 2 and Supplementary Fig. 7). These 71 proteins include six encoded by OMIM reported epileptic encephalopathy genes (http://www.omim.org/) where we identified one or more de novo mutations among the epileptic encephalopathy patients in this study. Genes in this protein–protein network were also found to have a much greater probability of overlap with the autism spectrum disorder and severe intellectual disability disorder exome sequencing study genes, and with FMRI-associated genes, than genes not in this network (Supplementary Table 11).

In support of a hypothesis that individual rare mutations in different genes may converge on biological pathways, we draw attention to six mutations that all affect subunits of the GABA (Y-aminobutyric acid) ionotropic receptor (four in GABRB3, and one each in GABRA1 and GABRB1), and highlight two interactions: HNRNPU interacting with HNRNP1 and NDE4L (identified here) binding to TNK2, a gene previously implicated in epileptic encephalopathies (Fig. 2). Although the HNRNPU mutation observed here is a small insertion/deletion variant (indel) in a splice acceptor site, and therefore probably results in a modified protein, the HNRNP1 de novo mutation is synonymous and thus of unknown functional significance (Supplementary Table 2). Notably, a minigene experiment indicates that this synonymous mutation induces skipping of exon 12 (Supplementary Methods).

Evaluation of the clinical phenotypes among patients revealed significant genetic heterogeneity underlying infantile spasms and Lennox–Gastaut syndrome, and begins to provide information about the range of phenotypes associated with mutations in specific genes (Supplementary Table 13). We identified four genes—SCN8A, STXBPI, DNM1 and GABRB3—with de novo mutations in both patients with infantile spasms and patients with Lennox–Gastaut syndrome. Although infantile spasms may progress to Lennox–Gastaut syndrome, in three of these cases the patients with Lennox–Gastaut syndrome did not initially present with infantile spasms, indicating phenotypic heterogeneity associated with mutations in these genes yet supporting the notion of shared genetic susceptibility. Notably, in multiple patients we identified de novo mutations in genes previously implicated in other neurodevelopmental conditions, and in some cases with very distinctive clinical presentations (Supplementary Table 12). Most notably, we found a de novo mutation in MTOR, a gene recently found to harbour a causal variant in mosaic form in a case with hemimegalencephaly. Our patient however showed no detectable structural brain malformation. Similarly, we found one patient with a de novo mutation in DCX and another with a de novo mutation in FLNA, previously associated with lissencephaly and periventricular nodular heterotopia, respectively, neither patient had cortical malformations detected on magnetic resonance imaging.

In addition to de novo variants, we also screened for highly penetrant genotypes by identifying variants that create newly homozygous, compound heterozygous, or hemizygous genotypes in the pedigrees that are not seen in parents or controls (Supplementary Methods). No inherited variants showed significant evidence of association. Additional studies evaluating a larger number of epileptic encephalopathy patients will be required to establish the role of inherited variants in the disease risk associated with infantile spasms and Lennox–Gastaut syndrome.

We have identified novel de novo mutations implicating at least two genes for epileptic encephalopathies, and also describe a genetic architecture that strongly suggests that we have identified additional causal mutations in genes intolerant to functional variation. Given that our sample size already shows many genes with recurrent mutations, it is clear that even modest increases in sample sizes will confirm many new genes now seen in only one of our trios. Our results also emphasize that it may be difficult to predict with confidence the responsible gene, even among known genes, based upon clinical presentation. This makes it clear that the future of genetic diagnostics in epileptic encephalopathies will need to focus on the genome as a whole as opposed to single genes or even gene panels. In particular, several of the genes with de novo mutations in our cohort have also been identified in patients with

---

**Table 1** Probability of observing the reported number of de novo mutations by chance in genes recurrently mutated in this cohort.

| Gene   | Chromosome | Average effectively captured length (bp) | Weighted mutation rate | De novo mutation number | P value† |
|--------|------------|------------------------------------------|------------------------|-------------------------|----------|
| SCN1A  | 2          | 6,063.70                                 | 1.61 × 10⁻⁴            | 5                       | 1.12 × 10⁻⁴  *** |
| STXB1  | 9          | 1,917.51                                 | 6.44 × 10⁻⁵            | 5                       | 1.16 × 10⁻¹¹  *** |
| GABRB3 | 15         | 1,206.86                                 | 3.78 × 10⁻⁵            | 4                       | 4.11 × 10⁻¹⁰  *** |
| CDKL5  | X          | 2,798.38                                 | 5.44 × 10⁻⁵            | 3                       | 4.90 × 10⁻⁸  *** |
| ALG13A | X          | 475.05                                   | 1.03 × 10⁻⁵            | 2                       | 7.77 × 10⁻¹²  *** |
| DNM1   | 9          | 2,323.37                                 | 9.10 × 10⁻⁵            | 2                       | 2.84 × 10⁻⁴  *** |
| HDAC4  | 2          | 2,649.82                                 | 1.16 × 10⁻⁴            | 2                       | 4.57 × 10⁻⁴  *** |
| SCN2A  | 2          | 5,831.21                                 | 1.52 × 10⁻⁴            | 2                       | 1.14 × 10⁻⁸  *** |
| SCN8A  | 12         | 5,814.48                                 | 1.64 × 10⁻⁴            | 2                       | 9.14 × 10⁻⁴  *** |

† Adjusted α is equivalent to 0.05/18,091 = 2.76 × 10⁻⁴ (***), 0.01/18,091 = 5.53 × 10⁻⁵ (**), and 0.001/18,091 = 5.53 × 10⁻⁶ (†).†† Counts exclude three additional patients with an indel or splice site mutation as these are not accounted for in the mutability calculation.

†† Two de novo mutations occur at the same position. The probability of these special cases obtain P = 7.77 × 10⁻¹² and P = 1.14 × 10⁻⁴ for ALG13 and SCN2A, respectively (Methods).
intellectual disability or autism spectrum disorder. Finally, and perhaps most importantly, this work suggests a clear direction for both drug development and treatment personalization in the epileptic encephalopathies, as many of these mutations seem to converge on specific biological pathways.

METHODS SUMMARY

All probands and family members were collected as part of the Epilepsy Phenome/Genome Project (EPGP) cohort (Supplementary Table 1). Detailed inclusion and exclusion criteria are provided in Methods. Patient collection and sharing of specimens for research were approved by site-specific Institutional Review Boards.

We sequenced the exome of each trio, from DNA derived from primary cells (n = 224 trios) or from lymphoblastoid cell lines (LCLs) in one or more family members (n = 40 trios), using the TruSeq Exome Enrichment kit (Illumina). We aligned samples and called variants using established algorithms (Methods) and identified candidate de novo variants at sites included in the exons or splice sites of the consensus coding sequence (CCDS) as those called in the affected child and absent in both parents, despite each parent having at least tenfold coverage at the site. Variants created by the de novo mutation also had to be absent in our internal controls (n = 436), as well as the approximately 6,500 samples represented in the Exome Variant Server (http://evs.gs.washington.edu/EVS), and had to pass visual inspection of alignment quality. Candidate de novo mutations were confirmed to be de novo mutations using Sanger sequencing. In all cases, primary DNA from the proband was used for the Sanger confirmation so that mutations appearing in the transformation process for the 40 trios sequenced from LCLs would be eliminated.

To determine whether our list of de novo mutations was preferentially located in genes contained in the six gene lists we calculated the proportion of CCDS de novo mutation opportunity space for each list (Additional Methods). A binomial probability calculation was used to determine whether the de novo mutations in CCDS transcripts identified in this cohort of epileptic encephalopathy patients were selectively enriched within the coding sequence of genes within a particular gene list (Supplementary Table 10).

Ingenuity Pathway Analysis (Ingenuity Systems) was used to assess the connectivity of proteins harbouring a de novo mutation.

Full Methods and any associated references are available in the online version of the paper.

Received 3 March; accepted 9 July 2013.
Published online 11 August 2013.

1. Berg, A. T. et al. Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005–2009. Epilepsia 51, 676–685 (2010).
De novo mutations in the sodium-channel gene SCN1A cause severe pharmacoresistant epilepsy. Am. J. Hum. Genet. 88, 1327–1332 (2011).

Saito, H. et al. De novo mutations in the gene encoding STXBP1 (MUNC18-1) cause early infantile epileptic encephalopathy. Nature Genet. 40, 782–788 (2008).

Otsuka, M. et al. STXBP1 mutations cause not only Ohtahara syndrome but also West syndrome—result of Japanese cohort study. Epilepsia 51, 2449–2452 (2010).

Veeramah, K. R. et al. De novo pathogenic SCN2A mutation identified by genome sequencing of a family quartet affected by infantile epilepsy. Eur. J. Hum. Genet. 20, 502–510 (2012).

Kamiya, K. et al. A nonsense mutation of the sodium channel gene SCN2A in a patient with intractable epilepsy and mental delay. J. Neurol. Sci. 260, 2690–2694 (2007).

Tanna, M. A., De Lorey, T. M., Delgado-Escueta, A. & Olsen, R. W. in Jasper’s Basic Mechanisms of the Epilepsies (eds Noels, J. L. et al.) (2012).

De Lorey, T. M. et al. Mice lacking the p3 subunit of the GABAA receptor have the epilepsy phenotype and many of the behavioral characteristics of Angelman syndrome. J. Neurosci. 18, 8505–8514 (1998).

Timal, S. et al. Genetic characterization in the cohort of genetic dysgenesis of gyration type I by whole-exome sequencing. Hum. Mol. Genet. 21, 4151–4161 (2012).

de Ligt, J. et al. Diagnostic exome sequencing in persons with severe intellectual disability. N. Engl. J. Med. 367, 1921–1929 (2012).

O’Roak, B. J. et al. Exome sequencing reveals that autism is strongly associated with autism. Science 334, 524–528 (2011).

Sandberg, S. J. et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 485, 246–250 (2012).

Petrovski, S. et al. Genomic inheritance to functional variation and the interpretation of personal genomes. PLoS Gen. (in the press) (2013).

Klassen, T. et al. Exome sequencing of ion channel genes reveals complex profiles confounding personal risk assessment in epilepsy. Cell 145, 1036–1048 (2011).

Lemon, J. R. et al. Targeted next generation sequencing as a diagnostic tool in epileptic disorders. Epilepsia 53, 1387–1396 (2012).

Rauch, A. et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet 380, 1673–1682 (2012).

Hitomi, Y. et al. Mutations in TNK2 in severe autosomal recessive infantile-onset epilepsy. Ann. Neurol. http://dx.doi.org/10.1002/ana.23934 (2013).

Lee, J. H. et al. De novo somatic mutations in components of the PKB-AXT3- mTOR pathway cause hemimegalencephaly. Nature Genet. 44, 941–945 (2012).

Lemoine, E. et al. Epilepsy exome studies: a meta-analysis of diverse sequencing strategies in epilepsy. Epilepsy Research, Epilepsy Therapy Project, Finding a Cure for Epilepsy and Seizures, IDEÁ Lead, InfantileSpasms.com, Lennox-Gastaut Syndrome Foundation, PatientsLikeMe, People Against Childhood Epilepsy, PVHN Support & Awareness, and Seizures & Epilepsy Education. We thank the EPGen Consortium (epi4k@duke.edu).

The collection of control samples was funded in part by ARRA 1RC2NS070342, NIAID R56AI098588, the Ellison Foundation New Scholar award AG-N5-0441-08, a award from SAID-Frederick, Inc. (M11-074), and with federal funds by the Center for HIV/AIDS Vaccine Immunology (“CHAVI”) under a grant from the National Institute of Allergy and Infectious Diseases, National Institutes of Health (U01AI067854).

**Author contributions** Initial design of EPGP: B.K.A., D.D., D.M., P.Kuz., D.H.L., R.O., E.H.S. and M.R.W. EPGen patient recruitment and phenotyping: B.A.-K., J.F.B., S.F.B., G.C., D.C., P.C., O.D., D.M., M.F., N.B.F., D.B., E.B.G., T.G., G.S., S.R.H., J.K., S.H., L.E.K., R.C.K., E.H.K., R.Kuz., K.Ruz., D.H.L., S.M.M., P.V.M., E.J.N., J.M.Pao., J.M.Par., K.P., A.P., I.E.S., J.J.S., R.S., J.S.I., M.C.S., L.L.T., A.V., E.P.G.V., G.K.V.A., J.L.W. and P.W.-W. Phenotype data analysis: B.A.-K., B.K.A., A.B.C., G.O., D.D., F.F., T.G., S.J., A.J.K., R.C.K., R.Kuz., D.H.L., R.O., J.M.Pao., A.P., I.E.S., R.A.S., E.H.S., J.J.S., J.Su., P.W.-W. and M.R.W. Initial design of Epik: S.F.B., P.Cao., N.D., D.D., E.E.F., M.P.E., T.G., D.B.G., E.L.H., M.R.J., R.Kuz., D.H.L., A.B.M., H.C.M., T.J.O., R.O., A.P., I.E.S. and E.H.S. Epilepsy encephalopathy phenotyping strategy: S.F.B., P.Cao., D.D., R.Kuz., D.H.L., R.O., I.E.S. and E.H.S. Encephalopathy phenotyping: D.D., K.B.H., M.R.Z.M., H.C.M., A.P., I.E.S. and C.H.Y. Sequence data analysis and statistical interpretation: A.S.A., D.B.G., E.L.H., R.A.S., E.H.S. and C.H.Y. Functional evaluation of identified mutations: D.B.G., E.L.H., Y.Hi. and Y.-F.L. Writing of manuscript: A.S.A., S.F.B., D.D., D.B.G., Y.Ha., E.L.H., M.R.J., D.H.L., R.O., A.P., S.P., E.K.R., I.E.S. and E.H.S.

**Author Information** Exome sequence data will be available in dbGaP (Epik4: Gene Discovery in 4,000 Epilepsy Genomes). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Epik4 (epik4@duke.edu).
METHODS

Study subjects. Infantile spasms and Lennox–Gastaut syndrome patients evaluated in this study were collected through the Epilepsy Phenome/Genome Project (EPGP, http://www.epgp.org)20. Patients were enrolled across 23 clinical sites. Informed consent was obtained for all patients in accordance with the site-specific Institutional Review Boards. Phenotypic information has been centrally databased and DNA specimens stored at the Corell Institute–NINDS Genetics Repository (Supplementary Table 1). Infantile spasms patients were required to have hypsarhythmia or a hypsarrhythmia variant on EEG. Lennox–Gastaut syndrome patients were required to have EEG background slowing or disorganization for age and generalized spike and wave activity of any frequency or generalized paroxysmal fast activity (GPFA). Background slowing was defined as <8 Hz posterior dominant rhythm in patients over 3 years of age, and <5 Hz in patients over 2 years of age. EEGs with normal backgrounds were accepted if the generalized spike and wave activity was 2.5 Hz or less and/or if GFAV was present.

All patients were required to have no evidence of moderate-to-severe developmental impairment or diagnosis of autistic disorder or pervasive developmental disorder before the onset of seizures. Severe developmental delay was defined by 50% or more delay in any area: motor, social, language, cognition, or activities of living or global delay. Mild delay was defined as delay of less than 50% of expected milestones in one area, or less than 30% of milestones across more than one area. All patients had no known confirmed genetic or metabolic diagnosis, and no history of congenital TORCH infection, premature birth (before 32 weeks gestation), neonatal hypoxic-ischaemic encephalopathy or neonatal seizures, meningitis/encephalitis, stroke, intracranial haemorrhage, significant head trauma, or evidence of acquired epilepsy. All infantile spasms and Lennox–Gastaut syndrome patients had an MRI or CT scan interpreted as normal, mild diffuse atrophy or focal cortical dysplasia. (Our case with the mutation in HNRNPU had had a reportedly normal MRI but on review of past records, a second more detailed MRI was found showing small regions of pteround nodular heterotopia.) To participate, both biological parents had to have no past medical history of seizures (except febrile or metabolic/toxic seizures).

A final diagnosis form was completed by local patients and Lennox–Gastaut syndrome patients had an MRI or CT scan interpreted as normal, mild diffuse atrophy or focal cortical dysplasia. (Our case with the mutation in HNRNPU had had a reportedly normal MRI but on review of past records, a second more detailed MRI was found showing small regions of pteround nodular heterotopia.) To participate, both biological parents had to have no past medical history of seizures (except febrile or metabolic/toxic seizures).

Exome sequencing, alignment and variant calling. Exome sequencing was carried out within the Genomic Analysis Facility in the Center for Human Genome Variation (Duke University). Sequencing libraries were prepared from primary DNA extracted from leukocytes of parents and probands using the Illumina TruSeq library preparation kit following the manufacturer’s protocol. Illumina TruSeq Exome Enrichment kit was used to selectively amplify the coding regions of the genome according to the manufacturer’s protocol. Six individual barcoded samples (two complete trios) were sequenced in parallel across two lanes of an Illumina HiSeq 2000 sequencer.

Alignment of the sequenced DNA fragments to Human Reference Genome (NCBI Build 37) was performed using the Burrows–Wheeler Alignment Tool (BWA) (version 0.5.10). The reference sequence we use is identical to the 1000 Genomes Phase II reference and it consists of chromosomes 1–22, X, Y, MT, unplaced and unlocalized contigs, the human herpesvirus 4 type 1 (NC_007605), and decoy sequences (hs37d5) as the reference and it consists of chromosomes 1–22, X, Y, MT, unplaced and unlocalized contigs, the human herpesvirus 4 type 1 (NC_007605), and decoy sequences (hs37d5)

Exome Variant Server, NHLBI Exome Sequencing Project (http://evs.gs.washington.edu/EVS/; accessed August 2012). We first filtered within the EVS database and eliminated from further consideration genes where the number of tenfold covered bases was less than 70% of its total extent. In calculating a score, we focused on departures from the average common functional variant frequency spectrum, corrected for the total mutation burden in a gene. We construct this score as follows. Let Y be the total number of common, minor allele frequency >0.1%, missense and nonsense (including splice) variants and let X be total number of variants (including synonymous) observed within a gene. We regress Y on X and take the standardized residual as the score (S). Thus, the raw residual is divided by an estimate of its standard deviation and thus account for differences in variability that comes with differing mutational burdens. S measures the departure from the average number of common functional mutations found in genes with a similar amount of mutational burden. When S ≈ 0 the gene is in agreement with the average number of functional variants given its total mutational burden. Genes where S < 0 have less common functionals than average for their mutational burden and thus, would seem to be less tolerant of functional mutation, indicating the presence of weak purifying selection.

We further investigated how different ‘intolerance’ thresholds of S captured known epileptic encephalopathy genes (Supplementary Table 8). Supplementary Fig. 6 illustrates how different percentages of S lead to the classification of different proportions of the known epileptic encephalopathy genes as ‘intolerant’. Note that ARX is not in these analyses as this gene did not meet a 70% of gene coverage threshold. The dashed vertical line in Supplementary Fig. 6 illustrates the 25th percentile of S and shows that using S = 0.05 we could classify all known epileptic encephalopathy genes as ‘intolerant’. On the basis of this analysis, we used this 25th percentile threshold in classifying genes as intolerant in all subsequent analyses. Supplementary Table 9 lists the 25th percentile of most intolerant genes that had Sanger confirmed de novo mutations among the infantile spasms/Lennox–Gastaut syndrome probands.

Defining the CCDS opportunity space for detecting de novo mutations. For each trio, we defined callable exonic bases that had the opportunity for identification of a coding de novo mutation, by restricting to bases where each of the three family members had at least tenfold coverage, obtained a multi-sampling (GATK) raw phred-scaled confidence score of ≥20 in the presence or absence of a variant, and were within the two base pairs at each end of exons to allow for splice acceptor and donor variants. Using these three criteria, the average CCDS-defined de novo mutation opportunity space across 264 trios was found to be 28.84 ± 0.92 Mb (range of 25.46–30.25 Mb).
To explore at the gene level, we similarly assessed the de novo calling opportunity within any given trio for every gene with a CCDS transcript. For genes with instances of non-overlapping CCDS transcripts, we merged the corresponding regions into a consensus summary of all CCDS-defined bases for that gene. Using these criteria, over 85% of the CCDS-defined exonic regions were sequenced to at least tenfold coverage across the three family members in over 90% of trios. All 264 trios covered at least 79% of the CCDS-defined regions under the CCDS opportunity space criteria.

Calculations of CCDS opportunity space for calling a de novo mutation, aside from the Y chromosome, were used in both the gene-list enrichment and architecture calculations.

30. Kryukov, G. V., Pennacchio, L. A. & Sunyaev, S. R. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. Am. J. Hum. Genet. 80, 727–739 (2007).
31. Kong, A. et al. Rate of de novo mutations and the importance of father’s age to disease risk. Nature 488, 471–475 (2012).