Cohorts included

Of 220 patients from European centres, mostly recruited to a European project investigating photosensitive epilepsy from Italy, Greece, Serbia and the Netherlands, 215 had epilepsy and PPR, and five had PPR without seizures. All had at least one EEG with PPR to standardised intermittent photic stimulation (IPS) (Kasteleijn-Nolst Trenité *et al.*, 2012) with any type of idiopathic epilepsy (generalised or focal). Of 140 patients from Kiel, Germany, 90 had epilepsy and PPR on standardised IPS; 50 had PPR without epilepsy (Tauer *et al.*, 2005; Lu *et al.*, 2008). 275 patients were recruited from Australia and New Zealand, with photosensitive epilepsy or epilepsy and PPR, but PPR was not a requirement for inclusion (Taylor *et al.*, 2013).

CoGIE (www.esf.org/euroepinomics) Cohort description

The cohort consists of 238 independent individuals with mainly idiopathic/genetic generalized epilepsy (IGE/GGE) including the clinical subsyndromes childhood absence epilepsy (CAE), juvenile absence epilepsy (JAE) and juvenile myoclonic epilepsy (JME) in combination with generalized spike-wave epileptic discharges (GSW) on electroencephalograms (EEG). The patients were recruited in Finland, France, Germany, Italy, The Netherlands and Turkey. All individuals were whole exome sequenced within the Euroepinomics CoGIE-IGE consortium. The index cases with CAE, JAE or JME were chosen from families with at least one additional case of IGE/GGE or generalized spike-wave discharges on EEG. However, ten sporadic cases and 27 patients with generalized-tonic clonic seizures (GTCS) only and GSW on EEG were also included, when they were part of families with mainly CAE, JAE or JME phenotypes and not enough DNA was available from those other family members. Combinations of more than one subsyndrome within one individual occurred as well. The cohort thus consists of the following groups: 118 cases with CAE, 29 JAE, 50 JME, 7 CAE/JME, 7 JAE/JME, and 27 cases with GTCS only.

***CHD2* sequencing**

For the European cohorts, coding regions and splice sites (assay designed for 20bp either side of each exon, variant calling was restricted to the five flanking base pairs for each coding exon) in *CHD2* were sequenced using the Illumina MiSeq™ sequencing system (Illumina Inc, San Diego, CA). Library preparation was carried out using the Illumina® TruSeq Custom Amplicon™ (TSCA) according to the manufacturer's instructions. Briefly, oligonucleotides targeting the regions of interest (RefSeq, hg 19 build) were designed using DesignStudio™, Illumina's web-based tool. The customized probes were hybridized to 250ng of genomic DNA to create a template of the regions of interest. Indexed primers were added to the DNA templates, which were then amplified by PCR, pooled and sequenced on the MiSeq system (Illumina Inc, San Diego, CA). BAM files were created by the MiSeq Reporter Software (Li *et al.*, 2009) and the data processed using VarScan (Koboldt *et al.*, 2012) to detect variants. We imposed the following VarScan filters for base calls and variant calls: a) a minimum read depth of 30x; b) a minimum of five supporting reads with the alternative allele at any given position in order to call a variant; c) a minimum base quality using a PHRED-like score threshold of 20 at any given position to count a read (probability of incorrect base call 1 in 100) (Cock *et al.*, 2010); d) a minimum frequency threshold for a heterozygous variant allele of 20% or 10 reads; e) a minimum frequency threshold for a homozygous variant allele of 75%; and f) a p-value threshold for calling variants of 0·01. Reads were aligned to the GenBank reference sequence from Ensembl Genome browser (GRCh37; www.ensemble.org) and viewed using the Integrative Genomics Viewer (Thorvaldsdóttir *et al.*, 2013). For the Australian cohort, patients had targeted *CHD2* sequencing using molecular inversion probes (O'Roak *et al.*, 2012) to capture all *CHD2* exons and intron/exon boundaries (±5bp flanking). Briefly, pooled molecular inversion probes were used to capture targets from

100ng of each proband's DNA. PCR was performed using universal primers with the introduction of unique eight-base barcodes on the tagged reverse primer. Pooled libraries were subject to massively parallel sequencing using a 100bp paired-end protocol on a MiSeq (Illumina). Raw read data was mapped with Burrows-Wheeler Aligner (BWA, version 0.5.9, http://bio-bwa.sourceforge.net/). Single nucleotide variant and indel calling and filtering was performed using the Genome Analysis Tool Kit (GATK, version 2.2, http://www.broadinstitute.org/gatk/). Only samples that had >85% of *CHD2* exonic intervals covered at a minimum of 50x were included for further analysis. Variants were annotated with SeattleSeq (SeattleSeq, v134, http://snp.gs.washington.edu/SeattleSeqAnnotation134/). Where family members were available, segregation analysis was carried out for all rare *CHD2* variants.

Whole exome sequencing

Library construction was undertaken using Nextera chemistry (Illumina Inc., San Diego, CA). Dual indexed, paired-end libraries were prepared from 50ng of genomic DNA. Libraries were enriched for the exome (62Mb) according to standard protocols supplied by Illumina (Illumina Inc., San Diego, CA), and then sequenced on a HiSeq2500 sequencing system (Illumina Inc., San Diego, CA). PCR duplicates were removed using Picard, as were reads without a unique mapping location. WES data was then mapped to the human reference assembly (GRCh37; www.genome.ucsc.edu), using Novoalign Software. Variants were extracted using the Maq model in SAMtools and filtered).

***CHD2* coverage and read depth**

ExAC: Coverage information for *CHD2* was extracted from the coverage assessment provided by the ExAC consortium (www.exac.broadinstitute.org), based on a sampling of

10% of the total number of samples. The mean average read depth across all *CHD2* target intervals in ExAC was 56·2x, while the mean average across all exons of the *CHD2* reference transcript (RefSeq: NM_001271.3) was 66·2x, excluding the 5' UTR exon (range 29·9-93·6x). On average, all exons were called at 20x in 94·4% and at 30x in 86·5% of the samples. For the 11 unique variants in the cases with epilepsy and PPR, the mean average read depth in ExAC was 67·4x (range 41-93x). For the variant in case (i) with PPR without epilepsy, the average read depth was 47x. All 12 nucleotide positions of the unique variants in the cases were called in 99·9% of the ExAC samples with at least 10x.

MiSeq, UK: A minimum read depth of 30x was required for an allele to be called with at least 10x coverage of the alternate allele present. Variants of interest and alleles with less than 30x read depth were re-sequenced using the Sanger method. Primer details are available on request.

Molecular inversion probes, USA: A minimum read depth of 30x was required for an allele to be called with at least 10x coverage of the alternate allele. All samples included in this study had a minimum of 85% of bases covered at 50x.

Whole exome sequencing, UK (5 EMA cases): Variants were filtered by the following criteria: a) consensus quality >30; b) SNP quality >30; and c) root mean square mapping quality >30. These variants were further filtered against dbSNP build 135 (UCSC genome browser) and 1000 Genomes dataset using Annovar. The mean average coverage for the *CHD2* reference sequence (RefSeq: NM_001271.3) was 38·6x (range 28·3-67·9x). In average, 46·2% of all *CHD2* target bases achieved 20x or greater coverage (range 40·8-56·6%). Two of the five whole exome-sequenced cases were also studied using molecular inversion probes, with exactly the same results.

CoGIE: All samples included in this study had a minimum of 74% (average of 94·5% +/- 4·81%) of bases covered at least 30x for CHD2 target sequence (CCDS10374.2, including +5bp intronic sequence at the exon/intron boundaries).

Heinzen *et al.*(2012): In order for a variant to be called present or absent in a study sample, the coverage of the site was required to be at least 10x and any variant had to have a quality score of at least 30.

Analysis of ExAC data

The ExAC variants were called by the consortium using the Genome Analysis Toolkit (GATK) 3.1 Haplotype Caller with Variant Quality Score Recalibration (VQSR). The VQSR was calibrated by the Consortium at SNP sensitivity 99.6%, and Indel sensitivity 95.0%, representing an attempt at a reasonable trade-off between missing true calls and including artifacts. Low-frequency variants are more likely than other variants to be filtered out if the selected sensitivity is too conservative. Because of the design of our study, we decided to include all SNPs and Indels in the ExAC dataset, with VQSR set to 100% sensitivity (all variants of the training set called), providing the lowest probability of missing true calls for our unique variant design. The trade-off for this sensitivity is only to reduce the power of the study to detect a true association.

Sanger sequencing

All putative variants were verified by Sanger sequencing analysed using Sequencher v5.2 analysis software (Gene Codes Corporation, USA). Primers for Sanger sequencing were designed using Primer3web v4.0 software (http://primer3.ut.ee) and primer specificity checked using UCSC in silico PCR software (http://genome.ucsc.edu/cgi-bin/hgPcr). For some variants, a second independent molecular inversion probe capture was used to confirm

variants.

Parental testing

Microsatellite analysis using the PowerPlex S5 system [Promega] was performed on all parents of probands with a de novo mutation to confirm maternity and paternity. In certain instances we were unable to determine the inheritance of a rare variant due to the unavailability of DNA from one or more parent.

Prediction of variant consequences

To assess the deleteriousness of the unique variants in the patients, we used a combination of several predictive resources (PolyPhen2, SIFTindel, splice-site inference), where appropriate, and the recently published Combined Annotation Dependent Depletion method (CADD) (Kircher *et al.*, 2014).

The CADD method is a framework that integrates multiple annotations into one metric, with the advantage that it allows the ranking of every variant, based on the predicted deleteriousness, among all GRCh37/hg19 reference single nucleotide variants (~8.6 billion).

Tectal field recordings in zebrafish

Zebrafish larvae were raised at 28·5°C in 0·3X Danieau's medium (1·5 mM HEPES, pH 7·6, 17·4 mM NaCl, 0·21 mM KCl, 0·12 mM $MgSO_4$, and 0·18 mM $Ca(NO_3)_2$) in a dark incubator. Each larva was embedded in 2% low-melting-point agarose and a glass electrode filled with artificial cerebrospinal fluid (composed of 124 mM NaCl, 2 mM KCl, 2 mM

$MgSO_4$, 2 mM $CaCl_2$, 1·25 mM KH_2PO_4, 26 mM $NaHCO_3$ and 10 mM glucose (resistance 1–5 MV)). The electrode was placed into the optic tectum and recordings were performed in current clamp mode, low-pass filtered at 1 kHz, high-pass filtered 0·1 Hz, digital gain 10, sampling interval 10 milliseconds (MultiClamp 700B amplifier, Digidata 1440A digitizer, both Axon instruments, USA) for a total of 10 minutes.

References for Supplementary Material

Cock PJA, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res. 2010; 38: 1767–71.

Heinzen EL, Depondt C, Cavalleri GL, Ruzzo EK, Walley NM, Need AC, et al. Exome sequencing followed by large-scale genotyping fails to identify single rare variants of large effect in idiopathic generalized epilepsy. Am J Hum Genet. 2012; 91: 293–302.

Kasteleijn-Nolst Trenité D, Rubboli G, Hirsch E, Martins da Silva A, Seri S, Wilkins A, Parra J, et al. Methodology of photic stimulation revisited: updated European algorithm for visual stimulation in the EEG laboratory. Epilepsia. 2012; 53: 16–24.

Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014; 46: 310–5.

Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012; 22: 568–76.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25: 2078–9.

Lu Y, Waltz S, Stenzel K, Muhle H, Stephani U. Photosensitivity in epileptic syndromes of childhood and adolescence. Epileptic Disord Int Epilepsy J Videotape. 2008; 10: 136–43.

O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science. 2012; 338: 1619–22

Tauer U, Lorenz S, Lenzen KP, Heils A, Muhle H, Gresch M, et al. Genetic dissection of photosensitivity and its relation to idiopathic generalized epilepsy. Ann Neurol. 2005; 57: 866–73.

Taylor I, Berkovic SF, Scheffer IE. Genetics of epilepsy syndromes in families with photosensitivity. Neurology. 2013; 80: 1322–9.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013; 14: 178–92.