Ultra-rare genetic variation in the epilepsies: a whole-exome sequencing study of 17,606 individuals

Epi25 Collaborative[†]

[†]A full list of contributing authors appears in the Consortia section.

Correspondence:

Samuel F. Berkovic (s.berkovic@unimelb.edu.au)

Benjamin M. Neale (bneale@broadinstitute.org)

Abstract

Sequencing-based studies have identified novel risk genes for severe epilepsies and revealed an excess of rare deleterious variation in less severe forms of epilepsy. To identify the shared and distinct ultra-rare genetic risk factors for different types of epilepsies, we performed a whole-exome sequencing (WES) analysis of 9,170 epilepsy-affected individuals and 8,364 controls of European ancestry. We focused on three phenotypic groups; severe developmental and epileptic encephalopathies (DEE), genetic generalized epilepsy (GGE), and non-acquired focal epilepsy (NAFE). We observed that compared to controls, individuals with any type of epilepsy carried an excess of ultra-rare, deleterious variants in constrained genes and in genes previously associated with epilepsy, with the strongest enrichment seen in DEE and the least in NAFE. Moreover, we found that inhibitory $GABA_A$ receptor genes were enriched for missense variants across all three classes of epilepsy, while no enrichment was seen in excitatory receptor genes. The larger gene groups for the GABAergic pathway or cation channels also showed a significant mutational burden in DEE and GGE. Although no single gene surpassed exome-wide significance among individuals with GGE or NAFE, highly constrained genes and genes encoding ion channels were among the top associations, including *CACNA1G*, *EEF1A2*, and *GABRG2* for GGE and *LGI1*, *TRIM3*, and *GABRG2* for NAFE. Our study confirms a convergence in the genetics of severe and less severe epilepsies associated with ultra-rare coding variation and highlights a ubiquitous role for GABAergic inhibition in epilepsy etiology in the largest epilepsy WES study to date.

Introduction

Epilepsy is a group of disorders characterized by repeated seizures due to excessive electrical activity in the brain and is one of the most common neurological conditions affecting every 5-7 of 1000 individuals worldwide[1; 2]. Human genetics research has established that a genetic basis underlies the susceptibility to epilepsy for a majority of the cases[3-6]. However, the multifactorial condition of epilepsy that subsumes a variety of epilepsy types, seizures, levels of severity, and comorbidity has made it a core challenge to disentangle the genetic architecture for different types of epilepsy and to determine the specific genetic risks for individual patients.

In recent years, our understanding of the genetic risk factors of epilepsy has substantially expanded thanks to the rapid advancement in sequencing technology. Currently, gene identification from sequencing-based studies has been primarily limited to rare, monogenic forms of epilepsy, particularly for a group of severe epilepsy syndromes, known as the developmental and epileptic encephalopathies (DEE)[7-11]. DEE typically begin early in life and are characterized by intractable seizures and profound to mild developmental impairment. It was found that 1 in 2,000 infants develop severe epilepsies with onset under 18 months[12]. For these severe epilepsies, dozens of genes with *de novo* pathogenic variants have been identified and the number continues to grow. The other major epilepsy types broadly encompass genetic generalized epilepsy (GGE) and non-acquired focal epilepsy (NAFE), the former characterized by seizures involving both hemispheres of the brain, the latter a localized cortical region. The incidence of these groups is not well-established, but they are recognized as the more common, less severe forms of epilepsy, and epidemiological studies have estimated generalized and focal epilepsies each account for 20-40% of incident epilepsies[13-16]. Similar to DEE, there are several specific electroclinical syndromes within the class of GGE and NAFE, but the genetic etiology is more complex. Genetic investigations into GGE or NAFE thus far support both a role for a oligogenic or polygenic component[17-20] as well as some evidence for monogenic causes for a minority of patients[5]. Despite a significant heritability consistently demonstrated from twin, family,

and genome-wide association studies (GWAS)[4; 19-22], single gene discovery has remained scarce for GGE and NAFE. Most genes identified to date come from monogenic families of focal epilepsies, while attempts to identify risk genes for GGE have been largely unsuccessful[23-25]. For most of the GGE and NAFE patients with a non-familial onset, the specific pathogenic variants are not yet known, and gene findings from small-scale studies have often not been reproducible[26-28].

Two recent whole-exome sequencing (WES) case-control studies leveraged hundreds of familial cases and provided clear evidence of specific gene groups linked to the risk of GGE and NAFE[24; 25]. Specifically, the authors showed that ultra-rare genetic variation in genes associated with DEE was enriched in GGE and NAFE, and that enrichment of missense variants in all genes encoding $GABA_A$ receptors was observed for the first time in GGE. These findings highlight that genes commonly implicated in epilepsy can span a wider range of epilepsy phenotypes than previously postulated. Studying rare genetic variation involving severe to milder electroclinical syndromes of epilepsy can help to better understand the extent of phenotypic pleiotropy and variable expressivity that may inform treatment strategies. On the other hand, the extensive phenotypic and genetic heterogeneity of epilepsy, especially for GGE and NAFE, underscores the need to enlarge the scale of such studies and beyond familial cases.

Here, we evaluate a WES case-control study of epilepsy from the Epi25 collaborative—an ongoing global effort to collect an unprecedented number of patient cohorts for primarily the three major classes of non-lesional epilepsies: DEE, GGE, and NAFE[29]. We aimed to pinpoint the distinct and overlapping genetic risk of ultra-rare coding variants across these different phenotypic groups by evaluating the burden at the individual gene level and in candidate gene sets to understand the role of rare genetic variation in epilepsy and identify specific epilepsy risk genes.

Subjects and Methods

Study design and participants

We collected DNA and detailed phenotyping data on individuals with epilepsy from 37 sites in Europe, North America, Australasia and Asia (**Supplemental Subjects and Methods; Table S1**). Here we analyzed subjects with genetic generalized epilepsy (GGE, also known as idiopathic generalized epilepsy; N=4,453), non-acquired focal epilepsy (NAFE; N=5,331) and developmental and epileptic encephalopathies (DEE; N=1,476); and a small number of other epilepsies were also included in the initiative (**Table S1**).

Control samples were aggregated from local collections at the Broad Institute (Cambridge, MA, USA) or obtained from dbGaP, consisting of 17,669 individuals of primarily European ancestry who were not ascertained for neurological or neuropsychiatric conditions (**Table S2; Supplemental Subjects and Methods**).

Phenotyping procedures

Epilepsies were diagnosed on clinical grounds based on criteria given in the next paragraph (see below for GGE, NAFE and DEE, respectively) by experienced epileptologists and consistent with International League Against Epilepsy (ILAE) classification at the time of diagnosis and recruitment. De-identified (non-PHI [protected health information]) phenotyping data were entered into the Epi25 Data repository hosted at the Luxembourg Centre for Systems Biomedicine via detailed on-line case record forms based on the RedCAP platform. Where subjects were part of previous coordinated efforts with phenotyping on databases (e.g., the Epilepsy Phenome/Genome Project[30] and the EpiPGX project (www.epipgx.eu)), deidentified data were accessed and transferred to the new platform. Phenotyping data underwent review for uniformity among sites and quality control by automated data checking, followed by manual review if required. Where doubt remained about eligibility, cases were reviewed by the phenotyping

committee and sometimes further data was requested from the source site before a decision was made.

Case Definitions

GGE required a convincing history of generalized seizure types (generalized tonic-clonic seizures, absence, or myoclonus) and generalized epileptiform discharges on EEG. We excluded cases with evidence of focal seizures, or with moderate to severe intellectual disability and those with an epileptogenic lesion on neuroimaging (although neuroimaging was not obligatory). If a diagnostic source EEG was not available, then only cases with an archetypal clinical history as judged by the phenotyping committee (e.g., morning myoclonus and generalized tonic-clonic seizures for a diagnosis of Juvenile Myoclonic Epilepsy) were accepted.

Diagnosis of NAFE required a convincing history of focal seizures, an EEG with focal epileptiform or normal findings (since routine EEGs are often normal in focal epilepsy), and neuroimaging showing no epileptogenic lesion except hippocampal sclerosis (MRI was preferred but CT was accepted). Exclusion criteria were a history of generalized onset seizures or moderate to severe intellectual disability.

The DEE group comprised subjects with severe refractory epilepsy of unknown etiology with developmental plateau or regression, no epileptogenic lesion on MRI, and with epileptiform features on EEG. As this is the group with the largest number of gene discoveries to date, we encouraged inclusion of those with non-explanatory epilepsy gene panel results, but we did not exclude those without prior testing (**Table S7**).

Informed Consent

Adult subjects, or in the case of children, their legal guardians, provided signed informed consent at the participating centers according to local national ethical requirements. Samples had been collected over a 20-year period in some centers, so the consent forms reflected standards at the time of collection. Samples were only accepted if the consent did not exclude data sharing. For

samples collected after January 25, 2015, consent forms required specific language according to the NIH Genomic Data Sharing policy (http://gds.nih.gov/03policy2.html).

Whole exome sequencing data generation

All samples were sequenced at the Broad Institute of Harvard and MIT on the Illumina HiSeq X platform, with the use of 151 bp paired-end reads. Exome capture was performed with Illumina Nextera® Rapid Capture Exomes or TruSeq Rapid Exome enrichment kit (target size 38 Mb), except for three control cohorts (MIGen ATVB, MIGen Ottawa, and Swedish SCZ controls) for which the Agilent SureSelect Human All Exon Kit was used (target size 28.6 Mb – 33 Mb). Sequence data in the form of BAM files were generated using the Picard data-processing pipeline and contained well-calibrated reads aligned to the GRCh37 human genome reference. Samples across projects were then jointly called via the Genome Analysis Toolkit (GATK) best practice pipeline[31] for data harmonization and variant discovery. This pipeline detected single nucleotide (SNV) and small insertion/deletion (indel) variants from exome sequence data.

Quality control

Variants were pre-filtered to keep only those passing the GATK VQSR (Variant Quality Score Recalibration) metric and those lying outside of low complexity regions[32]. Genotypes with GQ < 20 and heterozygous genotype calls with allele balance > 0.8 or < 0.2 were set to missing. To control for capture platform difference, we retained variants that resided in GENCODE coding regions where 80% of Agilent and Illumina-sequenced samples show at least 10x coverage. This resulted in the removal of ~50% of the called sites (23% of the total coding variants and 97% of the total non-coding variants) but effectively reduced the call rate difference between cases and controls (**Figure S1**). To further identify potential false positive sites due to technical variation, we performed single variant association tests (for variants with a minor allele frequency MAF > 0.001) among the controls, treating one platform as the pseudo-case group with adjustment for sex and

the first ten principal components (PCs), and removed variants significantly associated with capture labels (p-value < 0.05). We also excluded variants with a call rate < 0.98, case-control call rate difference > 0.005, or Hardy-Weinberg Equilibrium (HWE) test p-value < 1×10^{-6} based on the combined case and control cohort.

Samples were excluded if they had a low average call rate (< 0.98), low mean sequence depth (< 30; **Figure S2**), low mean genotype quality (< 85), high freemix contamination estimate (> 0.04), or high percent chimeric reads (> 1.4%). We performed a series of principal component analyses (PCAs) to identify ancestral backgrounds and control for population stratification, keeping only individuals of European (EUR) ancestry classified by Random Forest with 1000 Genomes data (**Figure S3**). Within the EUR population, we removed controls not well-matched with cases based on the top two PCs, and individuals with an excessive or a low count of synonymous singletons—a number that increases with the North-to-South axis (**Figure S4**). We also removed one sample from each pair of related individuals (proportion identity-by-descent > 0.2) and those whose genetically imputed sex was ambiguous or did not match with self-reported sex. Outliers (>4SD from the mean) of transition/transversion ratio, heterozygous/homozygous ratio, or insertion/deletion ratio within each cohort were further discarded (**Figures S5-7**). At the phenotype level, we removed individuals with epilepsy phenotype to-be-determined or marked as "excluded" from further review.

The number of variant and sample dropouts at each step are detailed in **Tables S3** and **S4.**

Variant annotation

Annotation of variants was performed with Ensembl's Variant Effect Predictor (VEP)[33] for human genome assemble GRCh37. Based on the most severe consequence, we defined four mutually exclusive functional classes of variants using relevant terms and SnpEff[34] impact (**Table S5**): protein-truncating variant (PTV), damaging missense (predicted by PolyPhen-2 and SIFT),

other/benign missense (predicted by PolyPhen-2 and SIFT), and synonymous. To further discriminate likely deleterious missense variants from benign missense variants, we applied an *in silico* missense deleteriousness predictor ("Missense badness, PolyPhen-2, and regional Constraint", or MPC score)[35] that leverages regional constraint information to annotate a subset of missense variants that are highly deleterious (MPC ≥ 2). The MPC ≥ 2 group accounts for a small proportion of the total damaging and benign missense variants annotated by PolyPhen-2 and SIFT. Because many of our control samples were obtained from external datasets used in the Exome Aggregation Consortium (ExAC)[36] (**Table S2)**, we used the DiscovEHR cohort—an external population allele frequency reference cohort that contains 50,726 whole-exome sequences from a largely European and non-diseased adult population[37]—to annotate if a variant is absent in the general population (**Figure S8**).

Gene-set burden analysis

To estimate the excess of rare, deleterious protein-coding variants in individuals with epilepsy, we conducted burden tests across the entire exome, for biologically relevant gene sets and at the individual gene level. We focused on two definitions of "ultra-rare" genetic variation (URV) for the primary analyses—variants not seen in the DiscovEHR database and observed only once among the combined case and control test cohort (allele count AC=1) or absent in DiscovEHR and observed no more than three times in the test cohort (allele count AC≤3)—where the strongest burden of deleterious pathogenic variants have been observed previously[24; 38] and in our study compared to less stringent allele frequency thresholds (**Figure S9 & S10**). We performed these case-control comparisons separately for each of the three primary epilepsy disorders (DEE, GGE, NAFE) and again for all epilepsy-affected individuals combined.

Gene-set burden tests were implemented using logistic regression to examine the enrichment of URVs in individuals with epilepsy versus controls. We performed the test by regressing case-control status on certain classes of URVs aggregated across a target gene set

in an individual, adjusting for sex, the top ten PCs, and exome-wide variant count. This analysis tested the burden of URVs separately for five functional coding annotations: synonymous, benign missense predicted by PolyPhen-2 and SIFT, damaging missense predicted by PolyPhen-2 and SIFT, protein-truncating variants, and missense with MPC≥2 (**Table S5**). To help determine whether our burden model was well calibrated, we used synonymous substitutions as a negative control, where significant burden effects would more likely indicate insufficient control of population stratification or exome capture differences. The inclusion of overall variant count as a covariate—which tracks with ancestry—made our test conservative but allows for better control of residual population stratification not captured by PCs, and effectively reduces inflation of signals in synonymous variants (**Figure S11**). We collected and tested eleven different gene sets, including constrained genes that are intolerant to loss-of-function mutations (pLI > 0.9 and pLI > 0.995[39]) or missense variation (mis-Z > 3.09[39]), brain-enriched genes that express more than 2-fold in brain tissues compared to other tissues based on Genotype-Tissue Expression Consortia (GTEx) data[40], and genes reported to be associated with epilepsy in a dominant fashion[10; 24] or epilepsy-related mechanisms[25] (**Table S6**). Unlike the gene-based burden tests, because most of the gene-set tests were not independent, we used a false discovery rate (FDR) correction for multiple testing that accounted for the number of functional categories (5), gene sets (11) and epilepsy phenotypes (4), totaling 220 tests, and defined a significant enrichment at FDR < 0.05.

Gene-based collapsing analysis

For gene-based tests, we restricted to deleterious URVs annotated as either PTV, missense with MPC≥2, or in-frame insertion/deletion. For each gene, individuals who had at least one copy of these deleterious variants were counted as a carrier, and we used a two-tailed Fisher's Exact test (FET) to assess if the proportion of carriers among epilepsy subgroup cases was significantly higher than controls. Instead of assuming a uniform distribution for p-values under the null, we generated empirical p-values by permuting case-control labels 500 times,

ordering the FET p-values of all genes for each permutation, and taking the average across all permutations to form a rank-ordered estimate of the expected p-value distribution. This was done by modifying functions in the "QQperm" R package[41]. To avoid potential false discoveries, we defined a stringent exome-wide significance as p-value < 6.8e-07, using Bonferroni correction to account for 18,509 consensus coding sequence genes tested and the four individual case-control comparisons.

Considering that recessive pathogenic variants were implicated in a number of epilepsy-associated genes, mostly identified from individuals with a DEE phenotype[8], we conducted a secondary gene-based Fisher's exact test using a recessive model, comparing the proportion of carriers that are homozygous for the minor allele between cases and controls. The recessive model was assessed for PTVs, missense (MPC≥2) variants, and in-frame indels separately. For this analysis, we did not restrict to non-DiscovEHR variants and relaxed the allele frequency up to MAF < 0.01 to account for the sparse occurrences.

Additionally, to evaluate the contribution of low frequency deleterious variants to epilepsy risk, we explored the gene burden of all protein-truncating and damaging missense variants for those with a MAF < 0.01 using SKAT[42], including sex and the top ten PCs as covariates in the analysis. We performed the tests with the default weighting scheme (dbeta(1,25)).

Single variant association

Associations of common and low-frequency variants (MAF > 0.001) with epilepsy were estimated using logistic regression by Firth's method, correcting for sex and the first ten PCs.

Quality control, annotation, and analysis were largely performed using Hail[43], an open-source software for scalable genomic data analysis, in conjunction with R (version 3.4.2).

Results

Whole exome sequencing, quality control, and sample overview

We performed WES on an initial dataset of over 30,000 epilepsy affected and control individuals. After stringent quality control (QC), we identified a total of 9,170 individuals with epilepsy and 8,436 controls without reported neurological or neuropsychiatric-related conditions, all of whom were unrelated individuals of European descent. Among the individuals with epilepsy, 1,021 were diagnosed with DEE, 3,108 with GGE, 3,597 with NAFE, and 1,444 with other epilepsy syndromes (lesional focal epilepsy, febrile seizures, and others). Cases and controls were carefully matched on genetic ancestry to eliminate the possibility of false positive findings induced by population stratification or effects of variable minor allele frequency resolution that occur when studying individuals from differing ancestries. Due to the lack of cosmopolitan controls from non-European populations, cases identified from PCA with a non-European ancestry were removed. Furthermore, to ensure the distribution of rare variants was balanced between cases and controls[44], we removed a subset of case and control-only cohorts (from Sweden, Finland, Cyprus, and Turkey) where the mean synonymous singleton count that significantly deviated from the overall average being the consequence of incomplete ancestry matching (**Figure S4**). We called a total of 1,844,644 sites in 18,509 genes in the final dataset, comprising 1,811,325 SNVs and 33,319 indels, 48.5% of which were absent in the DiscovEHR database[37]. Among the non-DiscovEHR sites, 85% were singletons (defined as only one instance of that variant), and 99% had a minor allele count (AC) not more than three (equivalent to MAF ≤0.01%; **Figure S8**); the missense with MPC≥2 annotation accounted for 2.0% of the total missense variants (5.5% of the damaging and 1.0% of the benign missense variants predicted by PolyPhen-2 and SIFT). In our primary burden analyses, we focused on the “ultra-rare” non-DiscovEHR variants (URVs) that are unique to the 17,606 individuals under study and are seen either only once (AC=1) or no more than three times (AC≤3) in our dataset. These URVs were shown to confer the largest risk of epilepsy compared to singletons observed in DiscovEHR, doubletons, or beyond (**Figure S9 &**

S10). As previously described, epilepsy enrichment signals diminished with an increase in allele frequency[24].

Enrichment of ultra-rare deleterious variants in constrained genes in DEE and GGE

We first tested the burden of singleton URVs for each epilepsy subgroup, as well as for all epilepsy-affected individuals combined, versus controls among gene sets collected based on current understanding and hypothesis of epilepsy causation. These included genes under evolutionary constraint, genes highly expressed in the brain, genes previously associated with epilepsy, $GABA_A$ receptor subunit-encoding genes, genes delineating GABAergic pathways, genes encoding excitatory neuronal receptors, and cation channel-encoding genes. (**Table S6**). To evaluate the burden in constrained genes, we defined "loss-of-function (LoF) intolerant" genes with either a pLI score[36] > 0.9 (3,488 genes) or separately a pLI score > 0.995 (1,583 genes) and those as "missense-constrained" for genes with a missense Z-score > 3.09 (1,730 genes)[39]. Genes marked by these specific cut-offs have been shown to be extremely intolerant to loss-of-function or missense variation and thus help to identify specific classes of variants with a higher burden in diseased individuals[36; 45; 46]. We used a version of the scores derived from the non-neuropsychiatric subset of the Exome Aggregation Consortium (ExAC) samples. Because some of our control cohorts are also in ExAC (**Table S2**), we restricted our constrained gene burden tests to controls outside of the ExAC cohort (N=4,042).

Consistent with a recent study that evaluated *de novo* burden in autism[46], burden signals of PTVs were mostly contained in genes with a pLI > 0.995 compared to pLI > 0.9 (**Figures S12** & **S13**). Focusing on pLI > 0.995 in the all-epilepsy case-control analysis, both protein-truncating and damaging missense (MPC[35]≥2) URVs in LoF-intolerant genes showed a mutational burden with an odds ratio of 1.3 ($adjP = 1.6\times10^{-4}$) and 1.1 ($adjP = 0.039$), respectively. Breaking this down by epilepsy types, there was a significant excess of these deleterious URVs among individuals with DEE ($OR_{PTV} = 1.4$, $adjP_{PTV} = 0.013$; $OR_{MPC} = 1.2$, $adjP_{MPC} = 0.019$), as expected.

This enrichment was also seen in individuals with GGE with a magnitude comparable to that in DEE (OR_{PTV} = 1.4, $adjP_{PTV}$ = 9.1×10^{-5}; OR_{MPC} = 1.2, $adjP_{MPC}$ = 5.5×10^{-3}), but was not significant in individuals with NAFE (OR_{PTV} = 1.2, $adjP_{PTV}$ = 0.062; OR_{MPC} =1.0, $adjP_{MPC}$ = 0.37; **Figure 1**). There was no evidence of excess burden in synonymous URVs, suggesting that enrichment of deleterious pathogenic variants was unlikely to be the result of un-modeled population stratification or technical artifact. Among *in-silico* missense predictors, MPC≥2 annotations consistently showed a higher burden than those predicted by PolyPhen-2 and SIFT. The burden among missense-constrained genes exhibited a similar pattern, with PTVs showing a higher burden in DEE than in GGE and NAFE (**Figure S14**). In addition, both large gene sets were more enriched for PTVs than for damaging missense variants.

Burden in candidate genetic etiologies associated with epilepsy

Among URVs in previously reported epilepsy genes, we found an expected and pronounced difference in the number of singleton protein-truncating URVs in individuals with DEE relative to controls. PTVs were associated with an increased DEE risk in 43 known dominant epilepsy genes[24] (*OR* = 6.3, *adjP* = 2.1×10^{-08}), 50 known dominant DEE genes[10] (*OR* = 9.1, *adjP* = 7.8×10^{-11}), and 33 genes with *de novo* burden in neurodevelopmental disorders with epilepsy[10] (*OR* = 14.8, *adjP* =1.7×10^{-12}). Evidence for an excess of ultra-rare PTVs was also observed in individuals with GGE, with an odds ratio ranging from 2 to 4. No enrichment of PTVs was observed among people with NAFE (**Figure 2A; Table S9**). In contrast, the burden of singleton missense (MPC≥2) URVs was more pervasive across epilepsy types. Compared to controls, there was a 3.6-fold higher rate of these missense URVs in established epilepsy genes in individuals with DEE (*adjP* = 1.6×10^{-10}), a 2.3-fold elevation in individuals with GGE (*adjP* = 6.4×10^{-07}), and a 1.9-fold elevation in individuals with NAFE (*adjP* = 2.8×10^{-4}).

Burden in genes encoding for cation channels and neurotransmitter receptors

Among brain-enriched genes—those defined as genes with at least a 2-fold increase in expression in brain tissues relative to their average expression across tissues based on GTEx data[40]—both protein-truncating and damaging missense (MPC≥2) URVs were significantly enriched in epilepsy cases versus controls, and the missense burden was much higher than the PTV burden (**Figure S15**). We then investigated the burden in four smaller gene sets previously implicated as mechanisms driving the etiology of epilepsy; these included 19 genes encoding $GABA_A$ receptor subunits, 113 genes involved in GABAergic pathways, 34 genes encoding excitatory receptors (ionotropic glutamate receptor subunits and nicotinic acetylcholine receptor subunits), and 86 voltage-gated cation channel genes (e.g., sodium, potassium, calcium—full list in **Table S6**)[25]. We discovered that, relative to damaging missense variants, the distribution of PTVs in most of these gene sets did not differ significantly between epilepsy cases and controls (**Figure 2A**; **Table 1**). The PTV signals that remained significant after FDR correction included, for individuals with DEE, an increased burden in GABAergic pathway genes and voltage-gated cation channels, and noticeably, for individuals with GGE, an increased burden in the inhibitory $GABA_A$ receptors (OR = 4.8, $adjP$ = 0.021). No PTV burden was detected for individuals with NAFE. In contrast, the enrichment of missense (MPC≥2) URVs was more extensive in these gene sets across all epilepsy-control comparisons (**Figure 2A**; **Table 1**). The burden of these damaging missense pathogenic variants was seen in $GABA_A$ receptor genes (OR_{DEE} = 3.7, $adjP_{DEE}$ = 0.028; OR_{GGE} = 3.8, $adjP_{GGE}$ = 1.4×10^{-3}; OR_{NAFE} = 2.7, $adjP_{NAFE}$ = 0.039), GABAergic pathway genes (OR_{DEE} = 2.6, $adjP_{DEE}$ = 4.7×10^{-5}; OR_{GGE} = 1.9, $adjP_{GGE}$ = 9.9×10^{-04}; OR_{NAFE} = 1.4, $adjP_{NAFE}$ = 0.11), and voltage-gated cation channel genes (OR_{DEE} = 2.1, $adjP_{DEE}$ = 1.7×10^{-03}; OR_{GGE} = 1.5, $adjP_{GGE}$ = 0.023; OR_{NAFE} = 1.4, $adjP_{NAFE}$ = 0.081). However, no enrichment was detected in genes encoding excitatory receptors. For individuals with NAFE, the burden signals were consistently the weakest across gene sets compared to the other epilepsy phenotypes. None of the gene sets was enriched for putatively neutral variation, except for a slightly elevated synonymous burden in $GABA_A$ receptor genes (**Table S9**). These results support a recent finding where rare missense

variation in $GABA_A$ receptor genes conferred a significant risk to GGE[25], and together implicate the relative importance and involvement of damaging missense variants in abnormal inhibitory neurotransmission in both severe and less severe forms of epilepsy.

For gene sets other than the three lists of previously associated genes (**Table S6**; 74 non-overlapping genes in total), we evaluated the residual burden of URVs after correcting for events in the 74 known genes. For the gene sets of cation channel and neurotransmitter receptor genes, the adjusted burden signals of singleton deleterious URVs was largely reduced, with some weak associations remaining in $GABA_A$ receptor-encoding or GABAergic genes among individuals with DEE or GGE. For the larger gene groups of constrained genes and brain-enriched genes, burden signals were attenuated but many remained significant, especially the strong enrichment of missense MPC≥2 variants in brain-enriched genes across all three classes of epilepsy (**Figure S16**). These findings suggest that although most gene burden is driven by previously identified genes, more associations could be uncovered with larger sample sizes.

Gene-based collapsing analysis recapture known genes for DEE

For gene discovery, because both protein-truncating and damaging missense (MPC≥2) URVs showed an elevated burden in epilepsy cases, we aggregated both together as deleterious pathogenic variants along with in-frame insertions and deletions in our gene collapsing analysis. This amassed to a total of 46,917 singleton URVs and 52,416 URVs with AC≤3. Surprisingly, for individuals diagnosed with DEE, we re-identified several of the established candidate DEE genes as top associations (**Figure 3A**). Although screening was not performed systematically, many DEE patients were screened-negative using clinical gene panels prior to enrollment (**Table S7**). Based on the results of singleton URVs, *SCN1A* was the only gene that reached exome-wide significance (*OR* = 18.4, $P = 5.8\times10^{-8}$); other top-ranking known genes included *NEXMIF* (previously known as *KIAA2022*; *OR* > 99, $P = 1.6\times10^{-6}$), *KCNB1* (*OR* = 20.8, $P = 2.5\times10^{-4}$), *SCN8A* (*OR* = 13.8, $P = 6.1\times10^{-4}$), and *SLC6A1* (*OR* = 11.1, $P = 3.6\times10^{-3}$) (**Table S11**). Some

carriers of deleterious URVs in lead genes were affected individuals with a normal result for gene panel testing, such as 2 out of the 3 carriers of qualifying URVs for *PURA* and 2 out of 5 for *KCNB1* (**Table S7**). This is primarily because gene panels ordered for a particular diagnosis usually do not screen all of the commonly implicated DEE genes (e.g., one of the carriers of qualifying URVs in *KCNB1* was diagnosed with West syndrome and screened with a customized panel that did not include *KCNB1*). Overall, more than 50 different gene panels were used across sample-contributing sites, which adds to the heterogeneity in screening procedures and interpretation. The gene burden results held up when considering URVs with AC≤3, often showing even stronger associations; two other well-studied genes, *STXBP1* (*OR* = 13.3, $P = 1.4\times10^{-5}$) and *WDR45* (*OR* > 49, $P = 1.2\times10^{-3}$), emerged on top, both of which have been implicated in DEE and developmental disorders (**Table S12**).

Channel and transporter genes implicated in GGE and NAFE

When evaluating gene burden in the GGE and NAFE epilepsy subgroups, we did not identify any exome-wide significant genes. However, several candidate epilepsy genes made up the lead associations, including ion channel and transporter genes, mutations of which are known to cause rare forms of epilepsy. For the GGE case-control analysis in singleton deleterious URVs, the lead associations included four previously-associated genes (*EEF1A2, OR* = 32, $P = 3.8\times10^{-4}$; *GABRG2*, *OR* = 19.0, $P = 6.2\times10^{-4}$; *SLC6A1*, *OR* = 7.3, $P = 2.0\times10^{-3}$; and *GABRA1*, *OR* = 9.5, $P = 2.2\times10^{-3}$), and two genes (*CACNA1G*, *OR* = 9.1, $P = 2.5\times10^{-4}$; *UNC79, OR* = 19.0, $P = 6.2\times10^{-4}$) that were not previously linked to epilepsy but are both highly expressed in the brain and under evolutionary constraint (**Figures 3B**; **Table S13**). Although evidence has been mixed, *CACNA1G* was previously implicated as a potential susceptibility gene for GGE in mutational analysis[47] and reported to modify mutated sodium channel (*SCN2A*) activity in epilepsy[48]. *UNC79* is an essential part of the UNC79-UNC80-NALCN channel complex that influences neuronal excitability by interacting with extracellular calcium ions[49], and this channel complex has been previously

associated with infantile encephalopathy[50]. Notably, all these lead genes were more enriched for damaging missense (MPC≥2) than for protein-truncating URVs despite the lower rate of MPC≥2 variants relative to PTVs (**Table S13**).

For individuals with NAFE, the analysis of singleton deleterious URVs identified *LGI1* and *TRIM3* as the top two genes carrying a disproportionate number of deleterious URVs, however neither reached exome-wide significance (OR > 32, $P = 2.1\times10^{-4}$). *GABRG2*, a lead association in individuals with GGE, was among the top ten most enriched genes, along with two brain-enriched, constrained genes (*PPFIA3*, OR = 8.2, $P = 4.2\times10^{-3}$; and *KCNJ3,* OR = 16.4, $P = 1.2\times10^{-3}$). *GABRG2* has previously been reported to show an enrichment of variants compared to controls in a cohort of individuals with Rolandic epilepsy (childhood epilepsy with centrotemporal spikes) or related phenotypes, the most common group of focal epilepsies of childhood[51]. Two other genes previously associated with epilepsy, *DEPDC5* and *SCN8A* (both OR = 5.5, P = 0.01), were among the top twenty associations (**Figures 3C**; **Table S14**). *LGI1* and *DEPDC5* are established genes for focal epilepsy, and *DEPDC5* was the only exome-wide significant hit in the Epi4K WES study for familial NAFE cases[24]. *TRIM3* has not been previously implicated in epilepsy, but evidence from a mouse model study implicates it in regulation of $GABA_A$ receptor signaling and thus modulation of seizure susceptibility[52]. Single gene burden for both GGE and NAFE remained similar when considering URVs with an allele count up to AC≤3 (**Tables S14 & S16**). Gene burden tests collapsing all epilepsy phenotypes recapitulated the lead genes in each of the subgroup-specific analyses, but none of the genes achieved exome-wide significance (**Tables S17 & S18**). It is worth noting that some of the genes were enriched for deleterious URVs among the “controls”, which is clearly driven by non-neuropsychiatric disease ascertainment for many of the available controls (e.g., *LDLR* in **Table S17**; most control carriers were individuals with cardiovascular diseases from the MIGen cohorts in **Table S2**). Thus, these should not be interpreted as potential protective signals for epilepsy.

Recessive model, SKAT gene test, and single variant association

The secondary gene-based test of a recessive model did not identify genes that differed significantly in the carrier rate of homozygous deleterious variants between epilepsy-affected individuals and controls (**Table S19**). Even if we considered variants up to MAF < 0.01, for most of the lead genes, only one case carrier was identified. For the DEE cohort, these genes included recessive genes previously implicated, such as *ARV1*, *BRAT1*, *CHRDL*[53] with a homozygous PTV and *OPHN1*[53] with a recessive missense (MPC≥2) variant (**Table S19A**). For the GGE and NAFE cohorts, a few studied recessive epilepsy genes were also observed in the lead gene associations, such as *SLC6A8*[53] (a homozygous PTV) for GGE (**Table S19B**), and *SLC6A8* (a homozygous missense-MPC) and *SYN1*[53] (a homozygous PTV) for NAFE (**Table S19C**). One GGE-affected individual was found homozygous for an in-frame deletion on *CHD2*, a dominant DEE gene[53] (**Table S19B**). These findings suggest an even larger cohort will be needed to identify with clarity recessive risk variants for different groups of epilepsy.

Beyond URVs, we studied the contribution of low frequency deleterious variants to epilepsy risk using SKAT (MAF < 0.01). Top associations for individuals with DEE included known genes, such as missense-enriched *STXBP1* ($P = 9.3\times10^{-9}$), *KCNA2* ($P = 1.0\times10^{-5}$; **Figure S18**), and PTV-enriched *NEXMIF* ($P = 7.1\times10^{-8}$), and *SCN1A* ($P = 3.9\times10^{-4}$; **Figure S19**). However, no significant gene enrichment was observed for GGE and NAFE or when combining all epilepsy cases. The tests for PTVs and missense variants with MPC≥2 were mostly underpowered due to sparse observations (**Figure S18 & S19**). No individual low-frequency variant (MAF > 0.001) was significantly associated with overall epilepsy or with any of the studied epilepsy phenotypes (**Figure S20**). The primary gene-based test results and single variant associations are available on our Epi25 WES browser (**Web Resources**).

Discussion

In the largest exome study of epilepsies to date, we show that ultra-rare deleterious coding variation—variation absent in a large population-based exome database—is enriched across the severity spectrum for epilepsy syndromes when compared to ancestrally matched controls. When all genes were considered in the tested gene sets, PTVs showed a more significant signal than missense variants with an MPC≥2, and enrichment in deleterious URVs was more pronounced in individuals diagnosed with DEE and GGE relative to NAFE. While no single gene surpassed exome-wide statistical significance for GGE or NAFE, specific gene sets that had previously been associated with epilepsy or encoding biologically interesting entities showed a clear enrichment of deleterious URVs. Specifically, we observed a significant excess of deleterious URVs in constrained genes, established epilepsy genes, and $GABA_A$ receptor subunit genes, a larger group of genes delineating the GABAergic pathway, and also all cation channel-encoding genes. Our results thus support the concept that defects in GABAergic inhibition underlie various forms of epilepsy. The epilepsy-associated excess of deleterious URVs in our study likely comprises signals from both inherited and *de novo* variants, the latter enriched by restricting variant inclusion to a combination of study-specific singletons and absence in a population reference cohort (DiscovEHR)[38; 45]. These findings, based on a more than 5-fold increase in sample size over previous exome-sequencing studies[24-26; 54], clearly support observations that have been hypothesized for GGE and NAFE from studies of rare, large monogenic families, and confirm that the same genes are relevant in both settings. Thus, a further increase in sample size will continue to unravel the complex genetic architecture of GGE and NAFE. The evidence that URVs contribute, in part, to GGE and NAFE is clear, but what remains unclear is the extent to which the excess rate of URVs observed in cases is a consequence of a small subset of patients carrying highly penetrant mutations versus URVs that are conferring risk, but do not rise to the level of Mendelian acting mutations but rather simply contribute to an overall polygenic risk for these syndromes. Interestingly, no enrichment was seen in genes encoding the excitatory glutamate

and acetylcholine receptors. For GGE, this difference between variants in inhibitory versus excitatory receptor genes may be real, as excitatory receptor variants have not been shown so far in single subjects or families. In NAFE, however, we suspect it is probably due to a lack of power and/or genetic heterogeneity, since genetic variants in specific subunits of nicotinic acetylcholine and NMDA receptors have been described extensively in different types of non-acquired familial focal epilepsies[55].

Notably, our overall finding of a mild to moderate burden of deleterious coding URVs in NAFE (**Figure 1** & **2**) contrasts with results reported in the Epi4K WES study, where the familial NAFE cohort showed a strong enrichment signal of ultra-rare functional variation in known epilepsy genes and ion channel genes[24]. In addition, our findings for GGE showed a genetic risk comparable or even stronger than the Epi4K familial GGE cohort. The strong signal in our GGE cohort likely reflects the larger sample size, whereas the weaker signal in our NAFE cohort is most likely due to differences in patient ascertainment. In Epi4K the cohort was deliberately enriched with familial cases, most of whom had an affected first-degree relative and were ascertained in sibling or parent-child pairs ors multiplex families, and familial NAFE is relatively uncommon. In the Epi25 collaboration a positive family history of epilepsy was not a requirement and only 9% of DEE, 12% of GGE, and 5% of NAFE patients had a known affected first-degree relative. Removing these familial cases led to no change in gene set burden (**Figure S17**) and a slightly attenuated association for some of the lead genes in the GGE and NAFE cohorts (**Table S20**). Indeed, our results were consistent with the Epi4K sporadic NAFE cohort, where no signals of enrichment were observed[24; 56]. This difference may reflect the substantial etiological and genetic heterogeneity of epilepsy even within subgroups especially in NAFE. In particular, the dramatically weaker genetic signals, per sample, observed in individuals with NAFE studied here compared with those in the previous Epi4K study illustrate a pronounced difference in the genetic signals associated with familial and non-familial NAFE. The reasons for this striking difference remain to be elucidated. Comparing GGE and NAFE, our findings showed a larger genetic burden

from URVs for GGE relative to NAFE, which could be due to heterogeneity in electroclinical syndromes within each class and should not be viewed as conclusive. On the other hand, in the latest GWAS of common epilepsies of 15,212 cases and 29,677 controls from the ILAE Consortium[20], fewer GWAS hits were discovered and less heritability was explained by common genetic variation for the focal epilepsy cohort (9.2%) compared to the GGE cohort (32.1%), suggesting that current evidence from both common and rare variant studies are converging on a larger genetic component underlying the etiology of non-familial cases of GGE relative to NAFE, as originally postulated.

We found that ultra-rare missense variants with an MPC score[35] ≥ 2 (2.0% of missense variants) were enriched in individuals with epilepsy at an effect size approaching PTVs in the investigated gene groups. For GGE and NAFE, the burden of these missense variants (MPC≥2) was even more prominent than PTVs in known epilepsy genes and GABAergic genes (**Figure 2**). At the gene level, some of the most commonly implicated channel genes (e.g. *GABRG2*, *CACNA1G*) carried a higher number of missense variants (MPC≥2) than PTVs in people with epilepsy. For instance, in the gene-based collapsing analysis considering all epilepsies, 15 *GABRG2* pathogenic variants were found in epilepsy-affected individuals (including 7 GGE and 7 NAFE; **Tables S13, 15 & 17**) versus only 1 pathogenic variant in controls; among the case-specific pathogenic variants, one was a splice site mutation, while the other 14 were all missense variants (MPC≥2) (**Figure S21**), linking to an impaired channel function. This is in line with findings from a recent exome-wide study of 6,753 individuals with neurodevelopmental disorder with and without epilepsy[10] that detected an association of missense *de novo* variants with the presence of epilepsy, particularly when considering only ion channel genes. A disease-association of missense variants rather than PTVs points to a pathophysiological mechanism of protein-alteration (e.g., gain-of-function or dominant-negative effects) rather than haploinsufficiency, but ultimately only functional tests can elucidate these mechanisms. A recent study on the molecular basis of 6 *de novo* missense variants in *GABRG2* identified in DEE reported an overall reduced

inhibitory function of *GABRG2* due to decreased cell surface expression or GABA-evoked current amplitudes, suggesting GABAergic disinhibition as the underlying mechanism[57]. Surprisingly, 2 of those recurrent *de novo* missense variants were seen in two GGE-affected individuals in our study (A106T and R323Q), and another recently reported variant in *GABRB2* (V316I) also occurred both *de novo* in DEE[58] and as an inherited variant in a GGE family showing a loss of receptor function[25]. This suggests that changes in protein function from the same missense pathogenic variant may contribute to not only severe epilepsy syndromes but also epilepsy phenotypes with milder presentations, similar to what is known about variable expressivity in large families carrying *GABRG2* variants[55; 59-61]. Reduced receptor function due to *GABRG2* variants has been also shown for childhood epilepsy with centrotemporal spikes previously[51; 61], which belong to the NAFE group in this study. Moving forward, discovering how variant-specific perturbations of the neurotransmission and signaling system in a gene can link to a spectrum of epilepsy syndromes will require in-depth functional investigation.

Although we have increased the sample size from the Epi4K and EuroEPINOMICS WES studies for both GGE and NAFE subgroups by more than 5-fold, the phenotypic and genetic heterogeneity of these less severe forms of epilepsy—on par with other complex neurological and neuropsychiatric conditions—will require many more samples to achieve statistical power for identifying exome-wide significant genes. We estimated that at least 8,000 cases and 20,000 controls would be required to convert some of the lead genes from the GGE and NAFE cohorts to exome-wide significance (**Table S8**). Furthermore, while we implemented stringent QC to effectively control for the exome capture differences between cases and controls, this concomitantly resulted in a loss of a substantial amount of the called sites and reduced our detection power to identify associated variants. As sample sizes grow, the technical variation across projects and sample collections will remain a challenge in large-scale sequencing studies relying on a global collaborative effort.

With this largest epilepsy WES study to date, we demonstrated a strong replicability of existing gene findings in an independent cohort. $GABA_A$ receptor genes affected by predicted-pathogenic missense pathogenic variants were enriched across the three subgroups of epilepsy. An ongoing debate in epilepsy genetics is the degree to which generalized and focal epilepsies segregate separately, and whether their genetic determinants are largely distinct or sometimes shared[4; 22]. Whilst clinical evidence for general separation of pathophysiological mechanisms in these two forms is strong, and most monogenic epilepsy families segregate either generalized or focal syndromes, the distinction is not absolute. Here, the finding of rare variants in $GABA_A$ receptor genes in both forms adds weight to the case for shared genetic determinants.

Our results suggest that clinical presentations of GGE and NAFE with complex inheritance patterns have a combination of both common and rare genetic risk variants. The latest ILAE epilepsy GWAS of over 15,000 patients and 25,000 controls identified 16 genome-wide significant loci for common epilepsies[20], mapped these loci to ion channel genes, transcriptional factors, and pyridoxine metabolism, and implicated a role in epigenetic regulation of gene expression in the brain. A combination of rare and common genetic association studies with large sample sizes, along with the growing evidence from studies of copy number variation and tandem repeat expansions in epilepsy[23; 62; 63], will further decipher the genetic landscape of GGE and NAFE. The ongoing effort of the Epi25 collaborative is expected to double the patient cohorts in upcoming years with the goal of elucidating shared and distinct gene discoveries for severe and less severe forms of epilepsy, ultimately facilitating precision medicine strategies in the treatment of epilepsy.

Supplemental Data

Supplemental data includes affiliations of the contributing authors, descriptions of patient recruitment and phenotyping from individual participating cohorts, supplemental acknowledgment, 21 figures and 20 tables.

Consortia

A full list of contributing authors:

Yen-Chen Anne Feng, Daniel P. Howrigan, Liam E. Abbott, Katherine Tashman, Felecia Cerrato, Tarjinder Singh, Henrike Heyne, Andrea Byrnes, Claire Churchhouse, Nick Watts, Matthew Solomonson, Dennis Lal, Erin L. Heinzen, Ryan S. Dhindsa, Kate E. Stanley, Gianpiero L. Cavalleri, Hakon Hakonarson, Ingo Helbig, Roland Krause, Patrick May, Sarah Weckhuysen, Slavé Petrovski, Sitharthan Kamalakaran, Sanjay M. Sisodiya, Patrick Cossette, Chris Cotsapas, Peter De Jonghe, Tracy Dixon-Salazar, Renzo Guerrini, Patrick Kwan, Anthony G. Marson, Randy Stewart, Chantal Depondt, Dennis J. Dlugos, Ingrid E. Scheffer, Pasquale Striano, Catharine Freyer, Kevin McKenna, Brigid M. Regan, Susannah T. Bellows, Costin Leu, Caitlin A. Bennett, Esther M.C. Johns, Alexandra Macdonald, Hannah Shilling, Rosemary Burgess, Dorien Weckhuysen, Melanie Bahlo, Terence J. O'Brien, Marian Todaro, Hannah Stamberger, Danielle M. Andrade, Tara R. Sadoway, Kelly Mo, Heinz Krestel, Sabina Gallati, Savvas S. Papacostas, Ioanna Kousiappa, George A. Tanteles, Katalin Štěrbová, Markéta Vlčková, Lucie Sedláčková, Petra Laššuthová, Karl Martin Klein, Felix Rosenow, Philipp S. Reif, Susanne Knake, Wolfram S. Kunz, Gábor Zsurka, Christian E. Elger, Jürgen Bauer, Michael Rademacher, Manuela Pendziwiat, Hiltrud Muhle, Annika Rademacher, Andreas van Baalen, Sarah von Spiczak, Ulrich Stephani, Zaid Afawi, Amos D. Korczyn, Moien Kanaan, Christina Canavati, Gerhard Kurlemann, Karen Müller-Schlüter, Gerhard Kluger, Martin Häusler, Ilan Blatt, Johannes R. Lemke, Ilona Krey, Yvonne G. Weber, Stefan Wolking, Felicitas Becker, Christian Hengsbach, Sarah Rau, Ana F. Maisch, Bernhard J. Steinhoff, Andreas Schulze-Bonhage, Susanne Schubert-Bast, Herbert

612 Schreiber, Ingo Borggräfe, Christoph J. Schankin, Thomas Mayer, Rudolf Korinthenberg, Knut
613 Brockmann, Gerhard Kurlemann, Dieter Dennig, Rene Madeleyn, Reetta Kälviäinen, Pia Auvinen,
614 Anni Saarela, Tarja Linnankivi, Anna-Elina Lehesjoki, Mark I. Rees, Seo-Kyung Chung, William
615 O. Pickrell, Robert Powell, Natascha Schneider, Simona Balestrini, Sara Zagaglia, Vera Braatz,
616 Michael R. Johnson, Pauls Auce, Graeme J. Sills, Larry W. Baum, Pak C. Sham, Stacey S.
617 Cherny, Colin H.T. Lui, Nina Barišić, Norman Delanty, Colin P. Doherty, Arif Shukralla, Mark
618 McCormack, Hany El-Naggar, Laura Canafoglia, Silvana Franceschetti, Barbara Castellotti,
619 Tiziana Granata, Federico Zara, Michele Iacomino, Francesca Madia, Maria Stella Vari, Maria
620 Margherita Mancardi, Vincenzo Salpietro, Francesca Bisulli, Paolo Tinuper, Laura Licchetta,
621 Tommaso Pippucci, Carlotta Stipa, Raffaella Minardi, Antonio Gambardella, Angelo Labate,
622 Grazia Annesi, Lorella Manna, Monica Gagliardi, Elena Parrini, Davide Mei, Annalisa Vetro,
623 Claudia Bianchini, Martino Montomoli, Viola Doccini, Carla Marini, Toshimitsu Suzuki, Yushi Inoue,
624 Kazuhiro Yamakawa, Birute Tumiene, Lynette G. Sadleir, Chontelle King, Emily Mountier, S.
625 Hande Caglayan, Mutluay Arslan, Zuhal Yapıcı, Uluc Yis, Pınar Topaloglu, Bulent Kara, Dilsad
626 Turkdogan, Aslı Gundogdu-Eken, Nerses Bebek, Sibel Uğur-İşeri, Betül Baykan, Barış Salman,
627 Garen Haryanyan, Emrah Yücesan, Yeşim Kesim, Çiğdem Özkara, Annapurna Poduri, Beth R.
628 Shiedley, Catherine Shain, Russell J. Buono, Thomas N. Ferraro, Michael R. Sperling, Warren
629 Lo, Michael Privitera, Jacqueline A. French, Steven Schachter, Ruben I. Kuzniecky, Orrin
630 Devinsky, Manu Hegde, Pouya Khankhanian, Katherine L. Helbig, Colin A. Ellis,
631 Gianfranco Spalletta, Fabrizio Piras, Federica Piras, Tommaso Gili, Valentina Ciullo, Andreas
632 Reif, Andrew McQuillin, Nick Bass, Andrew McIntosh, Douglas Blackwood, Mandy Johnstone,
633 Aarno Palotie, Michele T. Pato, Carlos N. Pato, Evelyn J. Bromet, Celia Barreto Carvalho, Eric D.
634 Achtyes, Maria Helena Azevedo, Roman Kotov, Douglas S. Lehrer, Dolores Malaspina, Stephen
635 R. Marder, Helena Medeiros, Christopher P. Morley, Diana O. Perkins, Janet L. Sobell, Peter F.
636 Buckley, Fabio Macciardi, Mark H. Rapaport, James A. Knowles, Genomic Psychiatry Cohort
637 (GPC) Consortium, Ayman H. Fanous, Steven A. McCarroll, Namrata Gupta, Stacey B. Gabriel,

Mark J. Daly, Eric S. Lander, Daniel H. Lowenstein, David B. Goldstein, Holger Lerche, Samuel F. Berkovic, and Benjamin M. Neale

Author contributions:

S.F.B., H.L., D.B.G., and D.H.L conceived the consortium and contributed to the study design. S.F.B., C.D., D.J.D., D.H.L., A.G.M., I.E.S., and P.S. designed patient recruitment and phenotyping criteria. D.H.L., D.B.G., H.L., S.F.B, C.D., D.J.D., A.G.M., I.E.S., P.S, C.F., R.K., K.M., B.M.R., S.T.B., G.L.C., P.C., C.C., P.D.G., T.D-S., R.G., H.H., E.L.H., I.H., P.K., S.P., S.M.S., R.S., and S.W. advised and approved the consortium activities. Authors from individual patient and control cohorts (Supplemental Data) contributed to site-specific recruitment and phenotyping. B.M.N., M.J.D., D.P.H., and F.C. directed and managed sample aggregation and sequencing effort at the Broad Institute. Sequence data generation was directed by S.B.G and managed by F.C. and N.G., with E.S.L, M.J.D., and B.M.N. overseeing the work. B.M.N, D.B.G., C.C., D.L, L.E.A., M.J.D., E.L.H., G.L.C., H.H., H.L., I.H., R.K., S.W., S.K., and S.P. supervised and reviewed genetic data analyses. Lead analyst Y.A.F. conducted the analyses with D.P.H, L.E.A., and K.T. Y.A.F. wrote the manuscript and interpreted the results with B.M.N., D.P.H., E.S.L., S.F.B., H.L., D.B.G., E.L.H., and D.H.L.

Acknowledgments

We gratefully thank the Epi25 principal investigators, local staff from individual cohorts, and all of the patients with epilepsy who participated in the study for making possible this global collaboration and resource to advance epilepsy genetics research. This work is part of the Centers for Common Disease Genomics (CCDG) program, funded by the National Human Genome Research Institute (NHGRI) and the National Heart, Lung, and Blood Institute (NHLBI). CCDG research activities at the Broad Institute was supported by NHGRI grant UM1 HG008895. The Genome Sequencing Program efforts were also supported by NHGRI grant 5U01HG009088-02.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Supplemental grant for Epi25 phenotyping was supported by “Epi25 Clinical Phenotyping R03”, National Institutes of Health (1R03NS108145-01), with D.H.L and S.F.B as the principal investigators. Additional funding sources and acknowledgment of individual patient and control cohorts were listed in **Supplemental Data**. We thank the Stanley Center for Psychiatric Research at the Broad Institute for supporting sequencing effort and control sample aggregation. The authors would like to thank the Discov-EHR collaboration of Geisinger Health System and Regeneron for providing exome variant data for comparison. We also thank Sali Farhan, Kyle Satterstrom, and Chai-Yen Chen for helpful discussions, Nick Watts and Matthew Solomonson for browser development, and the Hail team for analysis support.

Web Resources

The URLs for the consortium, data, and results presented herein are as follows:

Epi25 Collaborative, http://epi-25.org/

Exome Aggregation Consortium (ExAC), http://exac.broadinstitute.org

The DiscovEHR cohort, http://www.discovehrshare.com

Epi25 Year1 whole-exome sequence data on dbGaP, http://www.ncbi.nlm.nih.gov/gap through accession number phs001489 (the current study includes Year1-2 samples, and the Year2 data will later be made available)

Epi25 WES results browser, http://epi25.broadinstitute.org/

References

1. Fiest, K.M., Sauro, K.M., Wiebe, S., Patten, S.B., Kwon, C.S., Dykeman, J., Pringsheim, T., Lorenzetti, D.L., and Jette, N. (2017). Prevalence and incidence of epilepsy: A systematic review and meta-analysis of international studies. Neurology 88, 296-303.
2. Fisher, R.S., Acevedo, C., Arzimanoglou, A., Bogacz, A., Cross, J.H., Elger, C.E., Engel, J., Jr., Forsgren, L., French, J.A., Glynn, M., et al. (2014). ILAE official report: a practical clinical definition of epilepsy. Epilepsia 55, 475-482.
3. Annegers, J.F., Hauser, W.A., Anderson, V.E., and Kurland, L.T. (1982). The risks of seizure disorders among relatives of patients with childhood onset epilepsy. Neurology 32, 174-179.
4. Berkovic, S.F., Howell, R.A., Hay, D.A., and Hopper, J.L. (1998). Epilepsies in twins: genetics of the major epilepsy syndromes. Ann Neurol 43, 435-445.
5. Helbig, I., Scheffer, I.E., Mulley, J.C., and Berkovic, S.F. (2008). Navigating the channels and beyond: unravelling the genetics of the epilepsies. Lancet Neurol 7, 231-245.
6. Hildebrand, M.S., Dahl, H.H., Damiano, J.A., Smith, R.J., Scheffer, I.E., and Berkovic, S.F. (2013). Recent advances in the molecular genetics of epilepsy. J Med Genet 50, 271-279.
7. Epi4K Consortium. (2016). De Novo Mutations in SLC1A2 and CACNA1A Are Important Causes of Epileptic Encephalopathies. Am J Hum Genet 99, 287-298.
8. Epi4K Consortium, Epilepsy Phenome/Genome Project, Allen, A.S., Berkovic, S.F., Cossette, P., Delanty, N., Dlugos, D., Eichler, E.E., Epstein, M.P., Glauser, T., et al. (2013). De novo mutations in epileptic encephalopathies. Nature 501, 217-221.
9. EuroEPINOMICS-RES Consortium, Epilepsy Phenome/Genome Project, and Epi4K Consortium. (2014). De novo mutations in synaptic transmission genes including DNM1 cause epileptic encephalopathies. Am J Hum Genet 95, 360-370.
10. Heyne, H.O., Singh, T., Stamberger, H., Abou Jamra, R., Caglayan, H., Craiu, D., De Jonghe, P., Guerrini, R., Helbig, K.L., Koeleman, B.P.C., et al. (2018). De novo variants in neurodevelopmental disorders with epilepsy. Nat Genet 50, 1048-1053.
11. McTague, A., Howell, K.B., Cross, J.H., Kurian, M.A., and Scheffer, I.E. (2016). The genetic landscape of the epileptic encephalopathies of infancy and childhood. Lancet Neurol 15, 304-316.
12. Howell, K.B., Eggers, S., Dalziel, K., Riseley, J., Mandelstam, S., Myers, C.T., McMahon, J.M., Schneider, A., Carvill, G.L., Mefford, H.C., et al. (2018). A population-based cost-effectiveness study of early genetic testing in severe epilepsies of infancy. Epilepsia 59, 1177-1187.
13. Banerjee, P.N., Filippi, D., and Allen Hauser, W. (2009). The descriptive epidemiology of epilepsy-a review. Epilepsy Res 85, 31-45.
14. Camfield, P., and Camfield, C. (2015). Incidence, prevalence and aetiology of seizures and epilepsy in children. Epileptic Disord 17, 117-123.
15. Jallon, P., Loiseau, P., and Loiseau, J. (2001). Newly diagnosed unprovoked epileptic seizures: presentation at diagnosis in CAROLE study. Coordination Active du Reseau Observatoire Longitudinal de l' Epilepsie. Epilepsia 42, 464-475.
16. Jallon, P., and Latour, P. (2005). Epidemiology of idiopathic generalized epilepsies. Epilepsia 46 Suppl 9, 10-14.
17. EPICURE Consortium, Leu, C., de Kovel, C.G., Zara, F., Striano, P., Pezzella, M., Robbiano, A., Bianchi, A., Bisulli, F., Coppola, A., et al. (2012). Genome-wide linkage meta-analysis identifies susceptibility loci at 2q34 and 13q31.3 for genetic generalized epilepsies. Epilepsia 53, 308-318.
18. Marini, C., Scheffer, I.E., Crossland, K.M., Grinton, B.E., Phillips, F.L., McMahon, J.M., Turner, S.J., Dean, J.T., Kivity, S., Mazarib, A., et al. (2004). Genetic architecture of idiopathic generalized epilepsy: clinical genetic analysis of 55 multiplex families. Epilepsia 45, 467-478.
19. The International League Against Epilepsy Consortium on Complex Epilepsies. (2014). Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. Lancet Neurol 13, 893-903.

736 20. The International League Against Epilepsy Consortium on Complex Epilepsies. (2018). Genome-
737 wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the
738 common epilepsies. Nat Commun 9, 5269.
739 21. Kjeldsen, M.J., Corey, L.A., Christensen, K., and Friis, M.L. (2003). Epileptic seizures and
740 syndromes in twins: the importance of genetic factors. Epilepsy Res 55, 137-146.
741 22. Peljto, A.L., Barker-Cummings, C., Vasoli, V.M., Leibson, C.L., Hauser, W.A., Buchhalter, J.R.,
742 and Ottman, R. (2014). Familial risk of epilepsy: a population-based study. Brain 137, 795-
743 805.
744 23. Myers, C.T., and Mefford, H.C. (2015). Advancing epilepsy genetics in the genomic era. Genome
745 Med 7, 91.
746 24. Epi4K Consortium, and Epilepsy Phenome/Genome Project. (2017). Ultra-rare genetic variation
747 in common epilepsies: a case-control sequencing study. Lancet Neurol 16, 135-143.
748 25. May, P., Girard, S., Harrer, M., Bobbili, D.R., Schubert, J., Wolking, S., Becker, F., Lachance-
749 Touchette, P., Meloche, C., Gravel, M., et al. (2018). Rare coding variants in genes encoding
750 GABAA receptors in genetic generalised epilepsies: an exome-based case-control study.
751 Lancet Neurol 17, 699-708.
752 26. Heinzen, E.L., Depondt, C., Cavalleri, G.L., Ruzzo, E.K., Walley, N.M., Need, A.C., Ge, D., He,
753 M., Cirulli, E.T., Zhao, Q., et al. (2012). Exome sequencing followed by large-scale
754 genotyping fails to identify single rare variants of large effect in idiopathic generalized
755 epilepsy. Am J Hum Genet 91, 293-302.
756 27. Tan, N.C., and Berkovic, S.F. (2010). The Epilepsy Genetic Association Database (epiGAD):
757 analysis of 165 genetic association studies, 1996-2008. Epilepsia 51, 686-689.
758 28. Cavalleri, G.L., Lynch, J.M., Depondt, C., Burley, M.W., Wood, N.W., Sisodiya, S.M., and
759 Goldstein, D.B. (2005). Failure to replicate previously reported genetic associations with
760 sporadic temporal lobe epilepsy: where to from here? Brain 128, 1832-1840.
761 29. Scheffer, I.E., Berkovic, S., Capovilla, G., Connolly, M.B., French, J., Guilhoto, L., Hirsch, E.,
762 Jain, S., Mathern, G.W., Moshe, S.L., et al. (2017). ILAE classification of the epilepsies:
763 Position paper of the ILAE Commission for Classification and Terminology. Epilepsia 58,
764 512-521.
765 30. EPGP Collaborative, Abou-Khalil, B., Alldredge, B., Bautista, J., Berkovic, S., Bluvstein, J., Boro,
766 A., Cascino, G., Consalvo, D., Cristofaro, S., et al. (2013). The epilepsy phenome/genome
767 project. Clin Trials 10, 568-586.
768 31. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A.,
769 Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high
770 confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc
771 Bioinformatics 43, 11 10 11-33.
772 32. Li, H. (2014). Toward better understanding of artifacts in variant calling from high-coverage
773 samples. Bioinformatics 30, 2843-2851.
774 33. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., and
775 Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol 17, 122.
776 34. Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and
777 Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide
778 polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-
779 2; iso-3. Fly (Austin) 6, 80-92.
780 35. Samocha, K.E., Kosmicki, J.A., Karczewski, K.J., O'Donnell-Luria, A.H., Pierce-Hoffman, E.,
781 MacArthur, D.G., Neale, B.M., and Daly, M.J. (2017). Regional missense constraint improves
782 variant deleteriousness prediction. bioRxiv.
783 36. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria,
784 A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic
785 variation in 60,706 humans. Nature 536, 285-291.
786 37. Dewey, F.E., Murray, M.F., Overton, J.D., Habegger, L., Leader, J.B., Fetterolf, S.N.,
787 O'Dushlaine, C., Van Hout, C.V., Staples, J., Gonzaga-Jauregui, C., et al. (2016).

Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science 354.
38. Genovese, G., Fromer, M., Stahl, E.A., Ruderfer, D.M., Chambert, K., Landen, M., Moran, J.L., Purcell, S.M., Sklar, P., Sullivan, P.F., et al. (2016). Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. Nat Neurosci 19, 1433-1441.
39. Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., McGrath, L.M., Kosmicki, J.A., Rehnstrom, 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.
40. Ganna, A., Genovese, G., Howrigan, D.P., Byrnes, A., Kurki, M., Zekavat, S.M., Whelan, C.W., Kals, M., Nivard, M.G., Bloemendal, A., et al. (2016). Ultra-rare disruptive and damaging mutations influence educational attainment in the general population. Nat Neurosci 19, 1563-1565.
41. Petrovski, S., and Wang, Q. (2016). QQperm: Permutation Based QQ Plot and Inflation Factor Estimation. R package version 1.0.1. https://CRANR-projectorg/package=QQperm.
42. Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet 89, 82-93.
43. Hail, https://github.com/hail-is/hail.
44. Neale, B.M., Rivas, M.A., Voight, B.F., Altshuler, D., Devlin, B., Orho-Melander, M., Kathiresan, S., Purcell, S.M., Roeder, K., and Daly, M.J. (2011). Testing for an unusual distribution of rare variants. PLoS Genet 7, e1001322.
45. Kosmicki, J.A., Samocha, K.E., Howrigan, D.P., Sanders, S.J., Slowikowski, K., Lek, M., Karczewski, K.J., Cutler, D.J., Devlin, B., Roeder, K., et al. (2017). Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. Nat Genet 49, 504-510.
46. Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.-Y., Peng, M., Collins, R., Grove, J., Klei, L., et al. Novel genes for autism implicate both excitatory and inhibitory cell lineages in risk. bioRxiv.
47. Singh, B., Monteil, A., Bidaud, I., Sugimoto, Y., Suzuki, T., Hamano, S., Oguni, H., Osawa, M., Alonso, M.E., Delgado-Escueta, A.V., et al. (2007). Mutational analysis of CACNA1G in idiopathic generalized epilepsy. Mutation in brief #962. Online. Hum Mutat 28, 524-525.
48. Calhoun, J.D., Hawkins, N.A., Zachwieja, N.J., and Kearney, J.A. (2016). Cacna1g is a genetic modifier of epilepsy caused by mutation of voltage-gated sodium channel Scn2a. Epilepsia 57, e103-107.
49. Lu, B., Zhang, Q., Wang, H., Wang, Y., Nakayama, M., and Ren, D. (2010). Extracellular calcium controls background current and neuronal excitability via an UNC79-UNC80-NALCN cation channel complex. Neuron 68, 488-499.
50. Shamseldin, H.E., Faqeih, E., Alasmari, A., Zaki, M.S., Gleeson, J.G., and Alkuraya, F.S. (2016). Mutations in UNC80, Encoding Part of the UNC79-UNC80-NALCN Channel Complex, Cause Autosomal-Recessive Severe Infantile Encephalopathy. Am J Hum Genet 98, 210-215.
51. Reinthaler, E.M., Dejanovic, B., Lal, D., Semtner, M., Merkler, Y., Reinhold, A., Pittrich, D.A., Hotzy, C., Feucht, M., Steinbock, H., et al. (2015). Rare variants in gamma-aminobutyric acid type A receptor genes in rolandic epilepsy and related syndromes. Ann Neurol 77, 972-986.
52. Cheung, C.C., Yang, C., Berger, T., Zaugg, K., Reilly, P., Elia, A.J., Wakeham, A., You-Ten, A., Chang, N., Li, L., et al. (2010). Identification of BERP (brain-expressed RING finger protein) as a p53 target gene that modulates seizure susceptibility through interacting with GABA(A) receptors. Proc Natl Acad Sci U S A 107, 11883-11888.
53. Wang, J., Lin, Z.J., Liu, L., Xu, H.Q., Shi, Y.W., Yi, Y.H., He, N., and Liao, W.P. (2017). Epilepsy-associated genes. Seizure 44, 11-20.
54. Klassen, T., Davis, C., Goldman, A., Burgess, D., Chen, T., Wheeler, D., McPherson, J., Bourquin, T., Lewis, L., Villasana, D., et al. (2011). Exome sequencing of ion channel genes reveals complex profiles confounding personal risk assessment in epilepsy. Cell 145, 1036-1048.

55. Oyrer, J., Maljevic, S., Scheffer, I.E., Berkovic, S.F., Petrou, S., and Reid, C.A. (2018). Ion Channels in Genetic Epilepsy: From Genes and Mechanisms to Disease-Targeted Therapies. Pharmacol Rev 70, 142-173.
56. Epi4K Consortium. (2012). Epi4K: gene discovery in 4,000 genomes. Epilepsia 53, 1457-1467.
57. Shen, D., Hernandez, C.C., Shen, W., Hu, N., Poduri, A., Shiedley, B., Rotenberg, A., Datta, A.N., Leiz, S., Patzer, S., et al. (2017). De novo GABRG2 mutations associated with epileptic encephalopathies. Brain 140, 49-67.
58. Hamdan, F.F., Myers, C.T., Cossette, P., Lemay, P., Spiegelman, D., Laporte, A.D., Nassif, C., Diallo, O., Monlong, J., Cadieux-Dion, M., et al. (2017). High Rate of Recurrent De Novo Mutations in Developmental and Epileptic Encephalopathies. Am J Hum Genet 101, 664-685.
59. Baulac, S., Huberfeld, G., Gourfinkel-An, I., Mitropoulou, G., Beranger, A., Prud'homme, J.F., Baulac, M., Brice, A., Bruzzone, R., and LeGuern, E. (2001). First genetic evidence of GABA(A) receptor dysfunction in epilepsy: a mutation in the gamma2-subunit gene. Nat Genet 28, 46-48.
60. Wallace, R.H., Marini, C., Petrou, S., Harkin, L.A., Bowser, D.N., Panchal, R.G., Williams, D.A., Sutherland, G.R., Mulley, J.C., Scheffer, I.E., et al. (2001). Mutant GABA(A) receptor gamma2-subunit in childhood absence epilepsy and febrile seizures. Nat Genet 28, 49-52.
61. Maljevic, S., Moller, R.S., Reid, C.A., Pérez-Palma, E., Lal, D., May, P., and Lerche, H. (2018). Spectrum of GABAA receptor variants in epilepsy. Curr Opin Neurol In press.
62. Lal, D., Ruppert, A.K., Trucks, H., Schulz, H., de Kovel, C.G., Kasteleijn-Nolst Trenite, D., Sonsma, A.C., Koeleman, B.P., Lindhout, D., Weber, Y.G., et al. (2015). Burden analysis of rare microdeletions suggests a strong impact of neurodevelopmental genes in genetic generalised epilepsies. PLoS Genet 11, e1005226.
63. Tankard, R.M., Bennett, M.F., Degorski, P., Delatycki, M.B., Lockhart, P.J., and Bahlo, M. (2018). Detecting Expansions of Tandem Repeats in Cohorts Sequenced with Short-Read Sequencing Data. Am J Hum Genet 103, 858-873.

Figure titles and legends

Figure 1. Burden of ultra-rare singletons in LoF-intolerant genes (pLI > 0.995)

This analysis was restricted to 4,042 non-ExAC controls for comparison with epilepsy cases. We focused on "ultra-rare" variants not observed in the DiscovEHR database. Significance of association was displayed in FDR-adjusted p-values; odds ratios and 95% CIs were not multiplicity adjusted. The five functional coding annotations were defined as described in **Table S5**. PTV denotes protein-truncating variants; the "damaging missense" and "benign missense" categories were predicted by PolyPhen-2 and SIFT, while "damaging missense-MPC" was a group of missense variants with a missense badness score (MPC) ≥ 2. From top to bottom are the results based on all-epilepsy, DEE, GGE, and NAFE. Epilepsy cases, except for individuals with NAFE, carried a significant excess of ultra-rare PTV and damaging missense (MPC≥2) variants compared to controls (FDR < 0.05). PTV burden was higher than missense (MPC≥2) burden across epilepsy types.

Figure 2. Burden of ultra-rare singletons annotated as (A) protein-truncating variants or (B) damaging missense (MPC≥2) variants

"Ultra-rare" variants (URVs) were defined as not observed in the DiscovEHR database. Gene sets were defined in **Table S6**, with the number of genes specified in the parenthesis. DEE stands for individuals with developmental and epileptic encephalopathies, GGE for genetic generalized epilepsy, NAFE for non-acquired focal epilepsy, and EPI for all epilepsy; NDD-EPI genes are genes with *de novo* burden in neurodevelopmental disorders with epilepsy. Star signs indicate significance after FDR control ("*": FDR-adjusted p-value < 0.05; "**": adjusted p-value < 1×10^{-3}; "***": adjusted p-value < 1×10^{-5}). PTVs were enriched in candidate epilepsy genes for individuals with DEE relative to other epilepsy subgroups, but did not show a strong signal in inhibitory, excitatory receptors or voltage-gated cation channel genes. The burden of damaging missense (MPC≥2) variants, on the other hand, was stronger across these gene sets compared to PTVs,

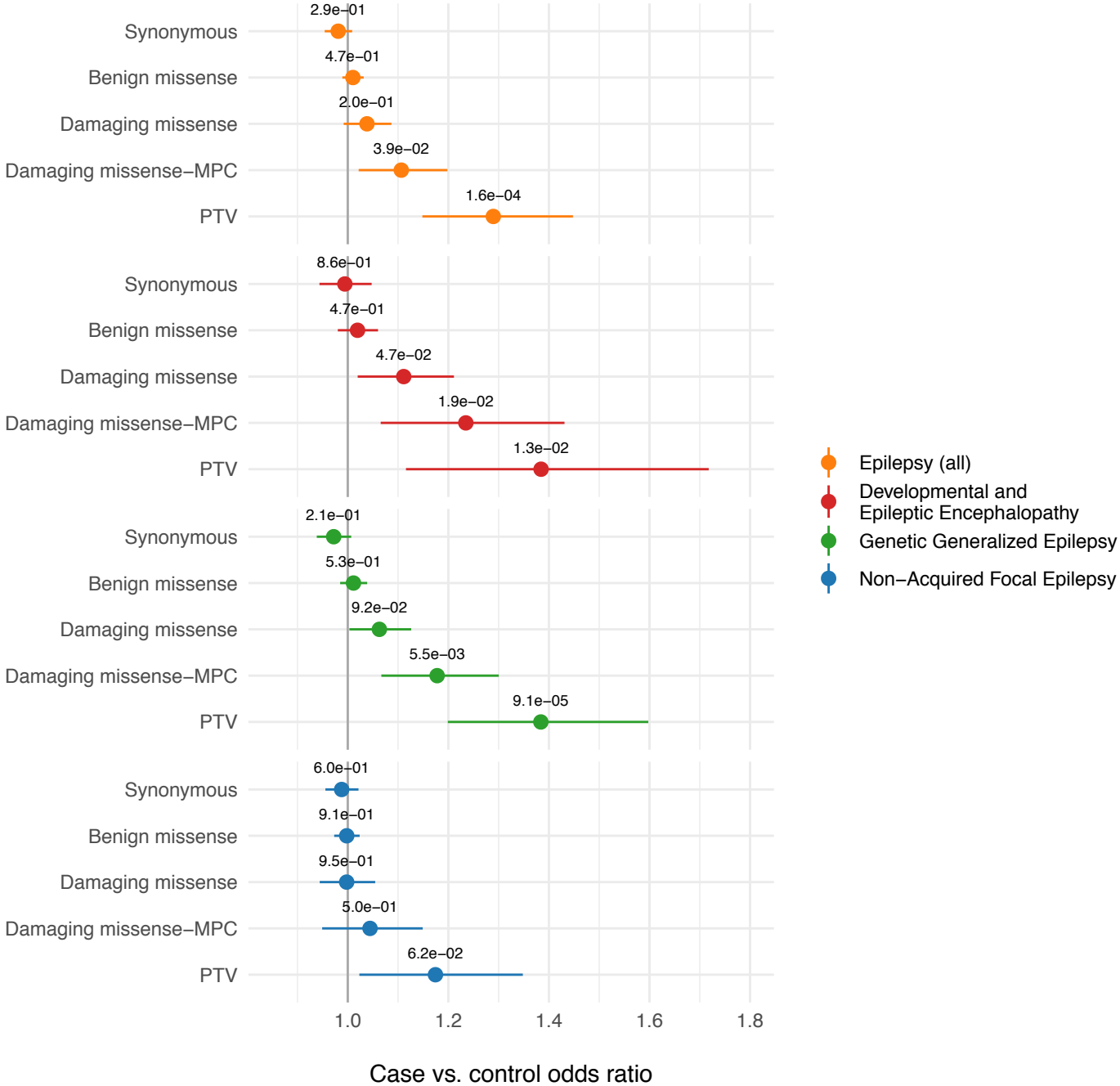
especially for $GABA_A$ receptor genes and genes involved in GABAergic pathways. Relative to other epilepsy types, individuals with NAFE consistently showed the least burden of deleterious URVs. No enrichment was observed from excitatory receptors.

Figure 3. Gene burden for individuals diagnosed with (A) developmental and epileptic encephalopathies, (B) genetic generalized epilepsy, or (C) non-acquired focal epilepsy

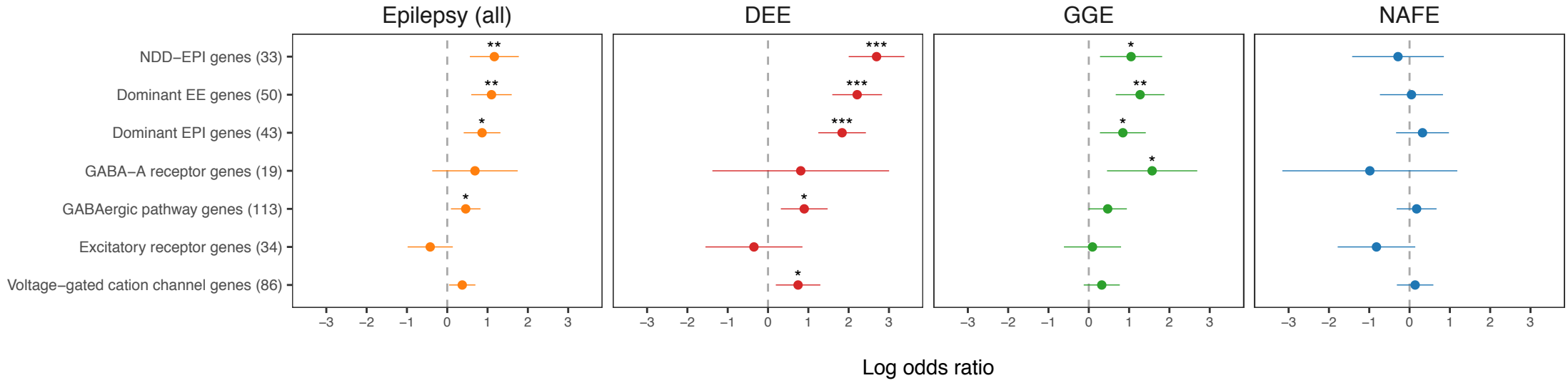
This analysis focused on ultra-rare (non-DiscovEHR) singleton variants annotated as PTV, damaging missense (MPC≥2), or in-frame insertion/deletion and used Fisher's exact test to identify genes with a differential carrier rate of these ultra-rare deleterious variants in individuals with epilepsy compared to controls. Exome-wide significance was defined as p-value < 6.8e-07 after Bonferroni correction (Methods). Only *SCN1A* achieved exome-wide significance for individuals with DEE.

Table 1. Enrichment of ultra-rare protein-truncating or damaging missense (MPC≥2) singletons in epilepsy
This analysis compared the burden of deleterious pathogenic variants between cases and controls using logistic regression, adjusting for sex, the first ten principal components, and overall variant count. FDR correction was based on a full list of burden tests in **Table S9**. Tested epilepsy types included all epilepsies (**EPI**; N=9,170), developmental and epileptic encephalopathies (**DEE**; N=1,021), genetic generalized epilepsy (**GGE**; N=3,108), and non-acquired focal epilepsy (**NAFE**; N=3,597). All were compared against 8,436 control samples. **Figure 2** shows the enrichment pattern of PTVs and MPC≥2 variants across the seven gene sets listed here.

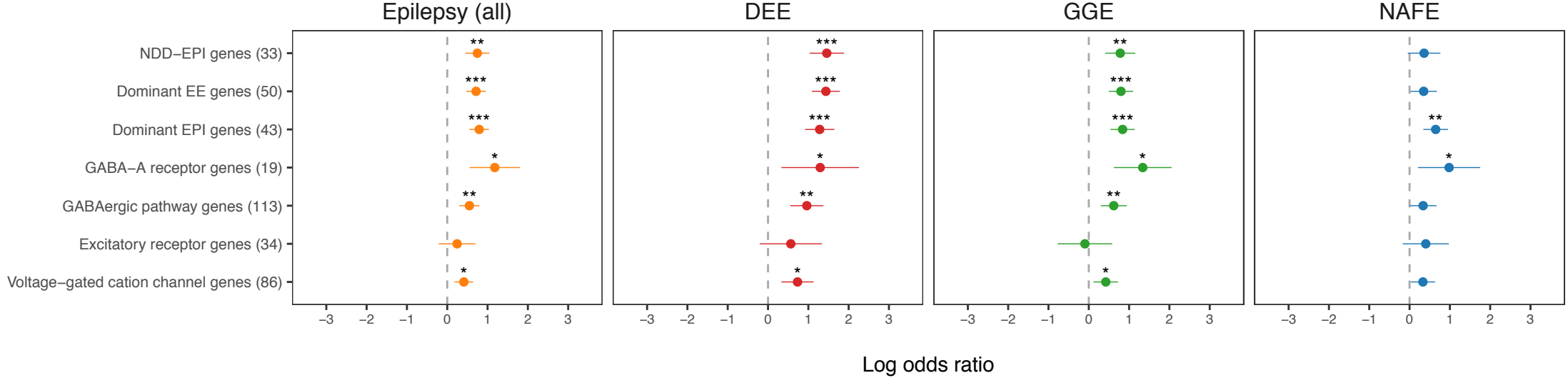
Gene set (# genes)	Mutation (# variants)	Epilepsy type	Carriers (N)		OR	95%CI	P-value	FDR adj. P
			cases	controls				
Known epilepsy genes (43)	PTV (95)	EPI	67	27	2.37	(1.50-3.74)	2.0e-04	1.2e-03
		DEE	24	27	6.28	(3.48-11.3)	1.0e-09	2.1e-08
		GGE	22	27	2.33	(1.32-4.11)	3.6e-03	1.4e-02
		NAFE	15	27	1.38	(0.72-2.66)	3.4e-01	4.7e-01
	MPC≥2 (335)	EPI	235	98	2.21	(1.74-2.81)	1.1e-10	2.8e-09
		DEE	47	98	3.60	(2.50-5.19)	5.0e-12	1.6e-10
		GGE	85	98	2.31	(1.71-3.12)	4.4e-08	6.4e-07
		NAFE	80	98	1.91	(1.41-2.60)	3.3e-05	2.8e-04
Known DEE genes (50)	PTV (89)	EPI	68	21	3.00	(1.82-4.95)	1.8e-05	1.6e-04
		DEE	27	21	9.13	(4.93-16.9)	2.1e-12	7.8e-11
		GGE	25	21	3.57	(1.95-6.54)	3.7e-05	3.0e-04
		NAFE	10	21	1.05	(0.48-2.29)	9.1e-01	9.3e-01
	MPC≥2 (327)	EPI	224	101	2.05	(1.61-2.60)	6.5e-09	1.2e-07
		DEE	54	101	4.20	(2.97-5.95)	6.0e-16	1.3e-13
		GGE	85	101	2.22	(1.64-3.00)	2.0e-07	2.6e-06
		NAFE	63	101	1.42	(1.02-1.97)	3.7e-02	8.8e-02
Neuro-developmental disorders with epilepsy (33)	PTV (63)	EPI	49	14	3.22	(1.75-5.90)	1.6e-04	9.9e-04
		DEE	29	14	14.77	(7.4-29.49)	2.3e-14	1.7e-12
		GGE	14	14	2.86	(1.32-6.17)	7.7e-03	2.7e-02
		NAFE	4	14	0.75	(0.24-2.34)	6.2e-01	7.2e-01
	MPC≥2 (215)	EPI	149	65	2.11	(1.57-2.84)	9.4e-07	1.1e-05
		DEE	36	65	4.30	(2.81-6.57)	1.8e-11	5.1e-10
		GGE	54	65	2.18	(1.50-3.17)	4.2e-05	3.2e-04
		NAFE	41	65	1.43	(0.96-2.15)	8.0e-02	1.6e-01
GABA-A receptors (19)	PTV (17)	EPI	12	5	1.99	(0.69-5.74)	2.0e-01	3.2e-01
		DEE	1	5	2.25	(0.25-20.2)	4.7e-01	6.0e-01
		GGE	9	5	4.81	(1.57-14.7)	5.9e-03	2.1e-02
		NAFE	1	5	0.37	(0.04-3.27)	3.7e-01	5.0e-01
	MPC≥2 (62)	EPI	49	13	3.25	(1.74-6.07)	2.1e-04	1.2e-03
		DEE	7	13	3.65	(1.39-9.54)	8.3e-03	2.8e-02
		GGE	21	13	3.81	(1.86-7.81)	2.5e-04	1.4e-03
		NAFE	15	13	2.67	(1.23-5.77)	1.3e-02	3.9e-02
GABAergic pathway (113)	PTV (127)	EPI	81	44	1.58	(1.10-2.28)	1.4e-02	4.4e-02
		DEE	16	44	2.46	(1.37-4.39)	2.4e-03	1.0e-02
		GGE	28	44	1.60	(0.99-2.57)	5.3e-02	1.1e-01
		NAFE	24	44	1.19	(0.73-1.95)	4.9e-01	6.1e-01
	MPC≥2 (287)	EPI	185	101	1.73	(1.35-2.22)	1.6e-05	1.6e-04
		DEE	34	101	2.62	(1.74-3.95)	4.5e-06	4.7e-05
		GGE	68	101	1.86	(1.35-2.56)	1.6e-04	9.9e-04
		NAFE	58	101	1.40	(1.00-1.95)	4.7e-02	1.1e-01
Excitatory receptors (34)	PTV (54)	EPI	22	32	0.66	(0.37-1.15)	1.4e-01	2.5e-01
		DEE	3	32	0.71	(0.21-2.35)	5.7e-01	6.7e-01
		GGE	11	32	1.10	(0.54-2.23)	8.0e-01	8.4e-01
		NAFE	5	32	0.44	(0.17-1.15)	9.5e-02	1.8e-01
	MPC≥2 (80)	EPI	47	33	1.28	(0.81-2.02)	2.9e-01	4.3e-01
		DEE	9	33	1.76	(0.81-3.81)	1.5e-01	2.6e-01
		GGE	12	33	0.91	(0.46-1.79)	7.8e-01	8.3e-01
		NAFE	20	33	1.50	(0.84-2.65)	1.7e-01	2.8e-01
Voltage-gated cation channels (86)	PTV (163)	EPI	100	63	1.45	(1.05-2.01)	2.5e-02	7.0e-02
		DEE	18	63	2.11	(1.21-3.66)	8.2e-03	2.8e-02
		GGE	31	63	1.38	(0.88-2.16)	1.6e-01	2.7e-01
		NAFE	30	63	1.15	(0.73-1.81)	5.5e-01	6.7e-01
	MPC≥2 (329)	EPI	206	121	1.51	(1.20-1.90)	4.7e-04	2.4e-03
		DEE	34	121	2.08	(1.40-3.10)	3.1e-04	1.7e-03
		GGE	73	121	1.52	(1.12-2.07)	6.6e-03	2.3e-02
		NAFE	74	121	1.39	(1.03-1.88)	3.1e-02	8.1e-02

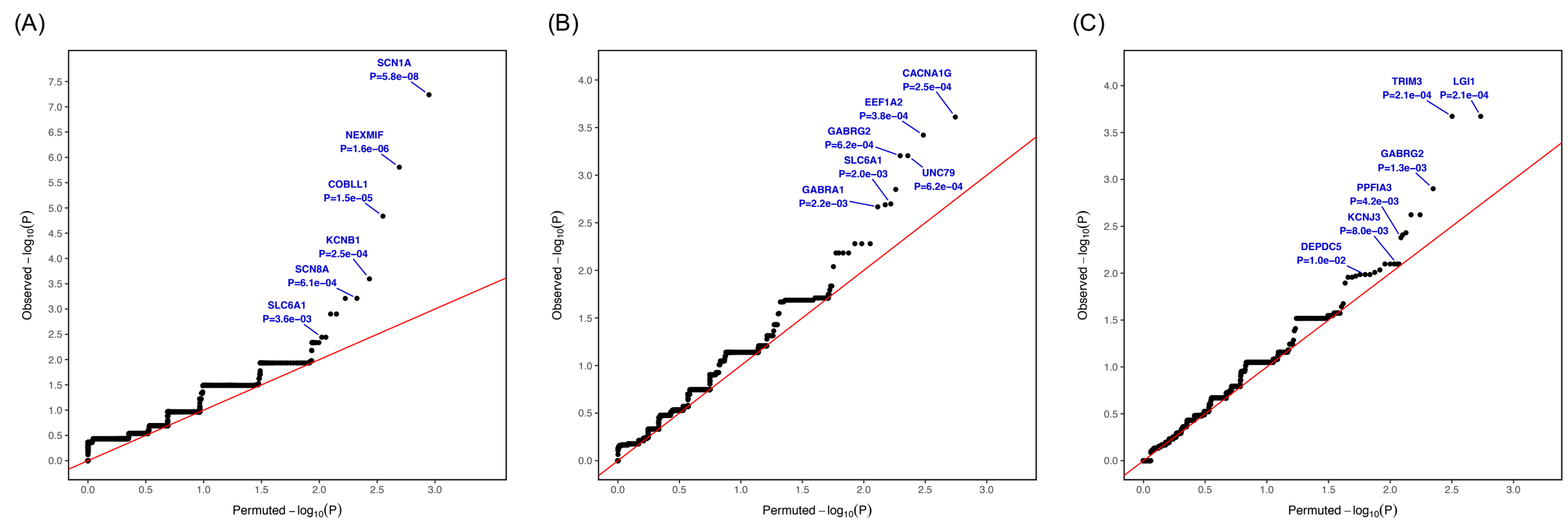

2.9e−01
Synonymous
4.7e−01
Benign missense
2.0e−01
Damaging missense
3.9e−02
Damaging missense−MPC
1.6e−04
PTV
8.6e−01
Synonymous
4.7e−01
Benign missense
4.7e−02
Damaging missense
1.9e−02
Damaging missense−MPC
1.3e−02
PTV
2.1e−01
Synonymous
5.3e−01
Benign missense
9.2e−02
Damaging missense
5.5e−03
Damaging missense−MPC
9.1e−05
PTV
6.0e−01
Synonymous
9.1e−01
Benign missense
9.5e−01
Damaging missense
5.0e−01
Damaging missense−MPC
6.2e−02
PTV
1.0
1.2
1.4
1.6
1.8
Case vs. control odds ratio
Epilepsy (all)
Developmental and Epileptic Encephalopathy
Genetic Generalized Epilepsy
Non−Acquired Focal Epilepsy

(A) Burden of ultra-rare PTVs



(B) Burden of ultra-rare damaging missense (MPC≥2) variants




(A)
SCN1A
P=5.8e−08
NEXMIF
P=1.6e−06
COBLL1
P=1.5e−05
KCNB1
P=2.5e−04
SCN8A
P=6.1e−04
SLC6A1
P=3.6e−03
Observed − log10(P)
Permuted − log10(P)
(B)
CACNA1G
P=2.5e−04
EEF1A2
P=3.8e−04
GABRG2
P=6.2e−04
SLC6A1
P=2.0e−03
UNC79
P=6.2e−04
GABRA1
P=2.2e−03
Observed − log10(P)
Permuted − log10(P)
(C)
TRIM3
P=2.1e−04
LGI1
P=2.1e−04
GABRG2
P=1.3e−03
PPFIA3
P=4.2e−03
KCNJ3
P=8.0e−03
DEPDC5
P=1.0e−02
Observed − log10(P)
Permuted − log10(P)

Supplemental Data

Table of contents

Supplemental Subjects and Methods

A full list of authors and affiliations

Epi25 sequencing, analysis, project management, and browser development at the Broad Institute
Yen-Chen Anne Feng[1-4], Daniel P. Howrigan[1,3,4], Liam E. Abbott[1,3,4], Katherine Tashman[1,3,4], Felecia Cerrato[3], Tarjinder Singh[1,3,4], Henrike Heyne[1,3,4], Andrea E. Byrnes[1,3,4], Claire Churchhouse[1,3,4], Nick Watts[1,3], Matthew Solomonson[1,3], Dennis Lal[4,5], Namrata Gupta[3], Stacey B. Gabriel[146], Mark J. Daly[1,3,4], Eric S. Lander[146,147,148], Benjamin M. Neale[1,3,4]

Epi25 executive committee
Samuel F. Berkovic[9], Holger Lerche[8], David B. Goldstein[6], Daniel H. Lowenstein[7]

Epi25 strategy, phenotyping, analysis, informatics, and project management comittees
Samuel F. Berkovic[9], Holger Lerche[8], David B. Goldstein[6], Daniel H. Lowenstein[7], Gianpiero L. Cavalleri[67,70], Patrick Cossette[106], Chris Cotsapas[111], Peter De Jonghe[12-14], Tracy Dixon-Salazar[112], Renzo Guerrini[82], Hakon Hakonarson[101], Erin L. Heinzen[6], Ryan S. Dhindsa[6], Kate E. Stanley[6], Ingo Helbig[28,29,102], Patrick Kwan[10,11], Anthony G. Marson[58], Slavé Petrovski[11,116], Sitharthan Kamalakaran[6], Sanjay M. Sisodiya[57], Randy Stewart[113], Sarah Weckhuysen[12-14], Chantal Depondt[15], Dennis J. Dlugos[101], Ingrid E. Scheffer[9], Pasquale Striano[74], Catharine Freyer[7], Roland Krause[114], Patrick May[114], Kevin McKenna[7], Brigid M. Regan[9], Susannah T. Bellows[9], Costin Leu[4,5,57]

Authors from individual Epi25 cohorts:

Australia: Melbourne (AUSAUS)
Samuel F. Berkovic[9], Ingrid E. Scheffer[9], Brigid M. Regan[9], Caitlin A. Bennett[9], Susannah T. Bellows[9], Esther M.C. Johns[9], Alexandra Macdonald[9], Hannah Shilling[9], Rosemary Burgess[9], Dorien Weckhuysen[9], Melanie Bahlo[119,120]

Australia: Royal Melbourne (AUSRMB)
Terence J. O'Brien[10,11], Patrick Kwan[10,11], Slavé Petrovski[11,116], Marian Todaro[10,11]

Belgium: Antwerp (BELATW)
Sarah Weckhuysen[12-14], Hannah Stamberger[12-14], Peter De Jonghe[12-14]

Belgium: Brussels (BELULB)
Chantal Depondt[15]

Canada: Andrade (CANUTN)
Danielle M. Andrade[16,17], Tara R. Sadoway[17], Kelly Mo[17]

Switzerland: Bern (CHEUBB)
Heinz Krestel[18], Sabina Gallati[19]

Cyprus (CYPCYP)
Savvas S. Papacostas[20], Ioanna Kousiappa[20], George A. Tanteles[21]

Czech Republic: Prague (CZEMTH)
Katalin Štěrbová[22], Markéta Vlčková[23], Lucie Sedláčková[22], Petra Laššuthová[22]

Germany: Frankfurt/Marburg (DEUPUM)
Karl Martin Klein[24,25], Felix Rosenow[24,25], Philipp S. Reif[24,25], Susanne Knake[25]

Germany: Bonn (DEUUKB)
Wolfram S. Kunz[26,27], Gábor Zsurka[26,27], Christian E. Elger[27], Jürgen Bauer[27], Michael Rademacher[27]

Germany: Kiel (DEUUKL)
Ingo Helbig[28,29,102], Karl Martin Klein[24,25], Manuela Pendziwiat[29], Hiltrud Muhle[29], Annika Rademacher[29], Andreas van Baalen[29], Sarah von Spiczak[29], Ulrich Stephani[29], Zaid Afawi[30], Amos D. Korczyn[31], Moien Kanaan[32], Christina Canavati[32], Gerhard Kurlemann[33], Karen Müller-Schlüter[34], Gerhard Kluger[35,36], Martin Häusler[37], Ilan Blatt[31,115]

Germany: Leipzig (DEUULG)
Johannes R. Lemke[38], Ilona Krey[38]

Germany: Tuebingen (DEUUTB)
Holger Lerche[8], Yvonne G. Weber[8,151], Stefan Wolking[8], Felicitas Becker[8,39], Christian Hengsbach[8], Sarah Rau[8], Ana F. Maisch[8], Bernhard J. Steinhoff[40], Andreas Schulze-Bonhage[41], Susanne Schubert-Bast[42], Herbert Schreiber[43], Ingo Borggräfe[44], Christoph J. Schankin[45], Thomas Mayer[46], Rudolf Korinthenberg[47], Knut Brockmann[48], Gerhard Kurlemann[33], Dieter Dennig[49], Rene Madeleyn[50]

Finland: Kuopio (FINKPH)
Reetta Kälviäinen[51], Pia Auvinen[51], Anni Saarela[51]

Finland: Helsinki (FINUVH)
Tarja Linnankivi[52], Anna-Elina Lehesjoki[53]

Wales: Swansea (GBRSWU)
Mark I. Rees[54,55], Seo-Kyung Chung[54,55], William O. Pickrell[54], Robert Powell[54,56]

UK: UCL (GBRUCL)
Sanjay M. Sisodiya[57], Natascha Schneider[57], Simona Balestrini[57], Sara Zagaglia[57], Vera Braatz[57]

UK: Imperial/Liverpool (GBRUNL)
Anthony G. Marson[58], Michael R. Johnson[59], Pauls Auce[60], Graeme J. Sills[61]

Hong Kong (HKGHKK)
Patrick Kwan[10,11,62], Larry W. Baum[117,118,63], Pak C. Sham[117,118,63], Stacey S. Cherny[64], Colin H.T. Lui[65]

Croatia (HRVUZG)
Nina Barišić[66]

Ireland: Dublin (IRLRCI)
Gianpiero L. Cavalleri[67,70], Norman Delanty[67,70], Colin P. Doherty[68,70], Arif Shukralla[69], Mark McCormack[67], Hany El-Naggar[69,70]

Italy: Milan (ITAICB)

Laura Canafoglia[71], Silvana Franceschetti[71], Barbara Castellotti[72], Tiziana Granata[73]

Italy: Genova (ITAIGI)
Pasquale Striano[74], Federico Zara[75], Michele Iacomino[75], Francesca Madia[75], Maria Stella Vari[74], Maria Margherita Mancardi[75], Vincenzo Salpietro[74]

Italy: Bologna (ITAUBG)
Francesca Bisulli[76,77], Paolo Tinuper[76,77], Laura Licchetta[76,77], Tommaso Pippucci[78], Carlotta Stipa[79], Raffaella Minardi[76]

Italy: Catanzaro (ITAUMC)
Antonio Gambardella[80], Angelo Labate[80], Grazia Annesi[81], Lorella Manna[81], Monica Gagliardi[81]

Italy: Florence (ITAUMR)
Renzo Guerrini[82], Elena Parrini[82], Davide Mei[82], Annalisa Vetro[82], Claudia Bianchini[82], Martino Montomoli[82], Viola Doccini[82], Carla Marini[82]

Japan: RIKEN Institute (JPNRKI)
Toshimitsu Suzuki[83], Yushi Inoue[84], Kazuhiro Yamakawa[83]

Lithuania (LTUUHK)
Birute Tumiene[85,86]

New Zealand: Otago (NZLUTO)
Lynette G. Sadleir[87], Chontelle King[87], Emily Mountier[87]

Turkey: Bogazici (TURBZU)
S. Hande Caglayan[88], Mutluay Arslan[89], Zuhal Yapıcı[90], Uluc Yis[91], Pınar Topaloglu[90], Bulent Kara[92], Dilsad Turkdogan[93], Aslı Gundogdu-Eken[88]

Turkey: Istanbul (TURIBU)
Nerses Bebek[94,95], Sibel Uğur-İşeri[95], Betül Baykan[94], Barış Salman[95], Garen Haryanyan[94], Emrah Yücesan[149], Yeşim Kesim[94], Çiğdem Özkara[96]

USA: BCH (USABCH)
Annapurna Poduri[97,98], Beth R. Shiedley[97,98], Catherine Shain[97,98]

USA: Philadelphia/CHOP (USACHP) and Philadelphia/Rowan (USACRW)
Russell J. Buono[99,100,101], Thomas N. Ferraro[99,102], Michael R. Sperling[100], Dennis J. Dlugos[101,102], Warren Lo[103], Michael Privitera[104], Jacqueline A. French[105], Patrick Cossette[106], Steven Schachter[107], Hakon Hakonarson[101]

USA: EPGP (USAEGP)
Daniel H. Lowenstein[7], Ruben I. Kuzniecky[108], Dennis J. Dlugos[101,102], Orrin Devinsky[105]

USA: NYU HEP (USAHEP)
Daniel H. Lowenstein[7], Ruben I. Kuzniecky[108], Jacqueline A. French[105], Manu Hegde[7]

USA: Penn/CHOP (USAUPN)
Ingo Helbig[28,102], Pouya Khankhanian[109,110], Katherine L. Helbig[28], Colin A. Ellis[110]

Authors from individual control cohorts (not including samples obtained from dbGaP):

Italian controls
Gianfranco Spalletta[121,122], Fabrizio Piras[121], Federica Piras[121], Tommaso Gili[123,121], Valentina Ciullo[121,124]

German controls
Andreas Reif[125,126]

UK/IRL controls 1
Andrew McQuillin[127], Nick Bass[127]

UK/IRL controls 2
Andrew McIntosh[128], Douglas Blackwood[128], Mandy Johnstone[128]

FINRISK controls
Aarno Palotie[1,2,4,129,130]

Genomic Psychiatry Cohort (GPC) controls
Michele T. Pato[131], Carlos N. Pato[131], Evelyn J. Bromet[132], Celia Barreto Carvalho[133], Eric D. Achtyes[134], Maria Helena Azevedo[135], Roman Kotov[132], Douglas S. Lehrer[136], Dolores Malaspina[137], Stephen R. Marder[138], Helena Medeiros[131], Christopher P. Morley[139], Diana O. Perkins[140], Janet L Sobell[141], Peter F. Buckley[142], Fabio Macciardi[143], Mark H. Rapaport[144], James A. Knowles[131], Genomic Psychiatry Cohort (GPC) Consortium, Ayman H. Fanous[131,145], Steven A. McCarroll[3,4,150]

Affiliations:

[1] Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA
[2] Psychiatric & Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA
[3] Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, 7 Cambridge Center, Cambridge, MA 02142, USA
[4] Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA
[5] Genomic Medicine Institute, Cleveland Clinic, Cleveland, OH 44195, USA
[6] Institute for Genomic Medicine, Columbia University, New York, NY 10032, USA
[7] Department of Neurology, University of California, San Francisco, CA 94110, USA
[8] Department of Neurology and Epileptology, Hertie Institute for Clinical Brain Research, University of Tübingen, 72076 Tübingen, Germany
[9] Epilepsy Research Centre, Department of Medicine, University of Melbourne, Victoria, Australia
[10] Department of Neuroscience, Central Clinical School, Monash University, Alfred Hospital, Melbourne, Australia
[11] Departments of Medicine and Neurology, University of Melbourne, Royal Melbourne Hospital, Parkville, Australia
[12] Neurogenetics Group, Center for Molecular Neurology, VIB, Antwerp, Belgium
[13] Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, Belgium
[14] Division of Neurology, Antwerp University Hospital, Antwerp, Belgium
[15] Department of Neurology, Université Libre de Bruxelles, Brussels, Belgium
[16] Department of Neurology, Toronto Western Hospital, Toronto, ON M5T 2S8, Canada

[17] University Health Network, University of Toronto, Toronto, ON, Canada
[18] Departments of Neurology and BioMedical Research, Bern University Hospital and University of Bern, Bern, Switzerland
[19] Institute of Human Genetics, Bern University Hospital, Bern, Switzerland
[20] Neurology Clinic B, The Cyprus Institute of Neurology and Genetics, 2370 Nicosia, Cyprus
[21] Department of Clinical Genetics, The Cyprus Institute of Neurology and Genetics, 2370 Nicosia, Cyprus
[22] Department of Paediatric Neurology, 2nd Faculty of Medicine, Charles University and Motol Hospital, Prague, Czech Republic
[23] Department of Biology and Medical Genetics, 2nd Faculty of Medicine, Charles University and Motol Hospital, Prague, Czech Republic
[24] Epilepsy Center Frankfurt Rhine-Main, Center of Neurology and Neurosurgery, Goethe University Frankfurt, Frankfurt, Germany
[25] Epilepsy Center Hessen-Marburg, Department of Neurology, Philipps University Marburg, Marburg, Germany
[26] Institute of Experimental Epileptology and Cognition Research, University Bonn, 53127 Bonn, Germany
[27] Department of Epileptology, University Bonn, 53127 Bonn, Germany
[28] Division of Neurology, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA
[29] Department of Neuropediatrics, Christian-Albrechts-University of Kiel, 24105 Kiel, Germany
[30] Sackler School of Medicine, Tel-Aviv University, Ramat Aviv, Israel
[31] Tel-Aviv University Sackler Faculty of Medicine, Ramat Aviv 69978, Israel
[32] Hereditary Research Lab, Bethlehem University, Bethlehem, Palestine
[33] Department of Neuropediatrics, Westfälische Wilhelms-University, Münster, Germany
[34] Epilepsy Center for Children, University Hospital Neuruppin, Brandenburg Medical School, Neuruppin, Germany
[35] Neuropediatric Clinic and Clinic for Neurorehabilitation, Epilepsy Center for Children and Adolescents, Vogtareuth, Germany
[36] Research Institute Rehabilitation / Transition / Palliation, PMU Salzburg, Austria
[37] Division of Neuropediatrics and Social Pediatrics, Department of Pediatrics, University Hospital, RWTH Aachen, Aachen, Germany
[38] Institute of Human Genetics, Leipzig, Germany
[39] RKU-University Neurology Clinic of Ulm, Ulm, Germany
[40] Kork Epilepsy Center, Kehl-Kork, Germany
[41] Epilepsy Center, University of Freiburg, Freiburg im Breisgau, Germany
[42] Section Neuropediatrics and Inborn Errors of Metabolism, University Children's Hospital, Heidelberg, Germany
[43] Neurological Practice Center & Neuropoint Patient Academy, Ulm, Germany.
[44] Department of Pediatric Neurology and Developmental Medicine, LMU Munich, Munich, Germany
[45] Department of Neurology, University of Munich Hospital-Großhadern, Munich, Germany
[46] Saxonian Epilepsy Center Radeberg, Radeberg, Germany
[47] Division of Neuropediatrics and Muscular Disorders, University Hospital Freiburg, Freiburg, Germany
[48] University Children's Hospital, Göttingen, Germany
[49] Private Neurological Practice, Stuttgart, Germany
[50] Department of Pediatrics, Filderklinik, Filderstadt, Germany
[51] Neurocenter, Kuopio University Hospital, Kuopio Finland and Institute of Clinical Medicine, University of Eastern Finland, Finland
[52] Child Neurology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
[53] Medicum, University of Helsinki, Helsinki, Finland and Folkhälsan Research Center, Helsinki, Finland
[54] Neurology Research Group, Swansea University Medical School, Swansea University SA2 8PP, UK
[55] Faculty of Medicine and Health, University of Sydney, Sydney, Australia
[56] Department of Neurology, Morriston Hospital, Abertawe BroMorgannwg HealthBoard, Swansea, UK
[57] Department of Clinical and Experimental Epilepsy, UCL Queen Square Institute of Neurology, London, UK and Chalfont Centre for Epilepsy, Chalfont St Peter, UK
[58] Department of Molecular and Clinical Pharmacology, University of Liverpool, Liverpool, UK
[59] Division of Brain Sciences, Imperial College London, London, UK
[60] Department of Neurology, Walton Centre NHS Foundation Trust, Liverpool, UK

[61] School of Life Sciences, University of Glasgow, Glasgow, UK
[62] Department of Medicine and Therapeutics, Chinese University of Hong Kong, Hong Kong, China
[63] Department of Psychiatry, University of Hong Kong, Hong Kong, China
[64] Department of Epidemiology and Preventive Medicine and Department of Anatomy and Anthropology, Sackler Faculty of Medicine, Tel Aviv University, Israel
[65] Department of Medicine, Tseung Kwan O Hospital, Hong Kong, China
[66] Department of Pediatric University Hospital centre Zagreb, Croatia
[67] The Department of Molecular and Cellular Therapeutics, The Royal College of Surgeons in Ireland, Dublin, Ireland
[68] Neurology Department, St. James Hospital, Dublin, Ireland
[69] The Department of Neurology, Beaumont Hospital, Dublin, Ireland
[70] The FutureNeuro Research Centre, Ireland
[71] Neurophysiopathology, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy
[72] Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy
[73] Department of Pediatric Neuroscience, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy
[74] Pediatric Neurology and Muscular Diseases Unit, Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genoa, "G. Gaslini" Institute, Genova, Italy
[75] Laboratory of Neurogenetics, "G. Gaslini" Institute, Genova, Italy
[76] IRCCS, Institute of Neurological Sciences of Bologna, Bologna, Italy
[77] Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy
[78] Medical Genetics Unit, Polyclinic Sant'Orsola-Malpighi University Hospital, Bologna, Italy
[79] Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy
[80] Institute of Neurology, Department of Medical and Surgical Sciences, University “Magna Graecia”, Catanzaro, Italy
[81] Institute of Molecular Bioimaging and Physiology, CNR, Section of Germaneto, Catanzaro, Italy
[82] Pediatric Neurology, Neurogenetics and Neurobiology Unit and Laboratories, Children's Hospital A. Meyer, University of Florence, Italy
[83] Laboratory for Neurogenetics, RIKEN Center for Brain Science, Saitama, Japan
[84] National Epilepsy Center, Shizuoka Institute of Epilepsy and Neurological Disorder, Shizuoka, Japan
[85] Institute of Biomedical Sciences, Faculty of Medicine, Vilnius University, Vilnius, Lithuania
[86] Centre for Medical Genetics, Vilnius University Hospital Santaros Klinikos, Vilnius, Lithuania
[87] Department of Paediatrics and Child Health, University of Otago, Wellington
[88] Department of Molecular Biology and Genetics, Bogaziçi University, Istanbul, Turkey
[89] Department of Child Neurology, Gulhane Education and Research Hospital, Health Sciences University, Ankara, Turkey
[90] Department of Child Neurology, Istanbul Faculty of Medicine, Istanbul University, Istanbul, Turkey
[91] Department of Child Neurology, Medical School, Dokuz Eylul University, Izmir, Turkey
[92] Department of Child Neurology, Medical School, Kocaeli University, Kocaeli, Turkey
[93] Department of Child Neurology, Medical School, Marmara University, Istanbul, Turkey
[94] Department of Neurology, Istanbul Faculty of Medicine, Istanbul University, Istanbul, Turkey
[95] Department of Genetics, Aziz Sancar Institute of Experimental Medicine, Istanbul University, Istanbul, Turkey
[96] Department of Neurology, Faculty of Medicine, Cerrahpaşa University Istanbul, Istanbul, Turkey
[97] Epilepsy Genetics Program, Department of Neurology, Boston Children's Hospital, Boston, MA 02115, USA
[98] Department of Neurology, Harvard Medical School, Boston, MA 02115, USA
[99] Cooper Medical School of Rowan University, Camden, NJ 08103, USA
[100] Thomas Jefferson University, Philadelphia, PA 19107, USA
[101] The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA
[102] Perelman School of Medicine, University of Pennsylvania, PA 19104, USA
[103] Nationwide Children's Hospital, Columbus, OH 43205, USA
[104] University of Cincinnati, Cincinnati, OH 45220, USA
[105] Department of Neurology, New York University/Langone Health, New York, NY 10016, USA
[106] University of Montreal, Montreal, QC H3T 1J4, Canada

[107] Beth Israel Deaconess/Harvard, Boston, MA 02115, USA
[108] Department of Neurology, Hofstra-Northwell Medical School, New York, NY 11549, USA
[109] Center for Neuro-engineering and Therapeutics, University of Pennsylvania, Philadelphia, PA 19104 , USA
[110] Department of Neurology, Hospital of University of Pennsylvania, Philadelphia, PA 19104, USA
[111] School of Medicine, Yale University, New Haven, CT 06510, USA
[112] LGS Foundation, NY 11716, USA
[113] National Institute of Neurological Disorders and Stroke, MD 20852, USA
[114] Luxembourg Centre for Systems Biomedicine, University Luxembourg, Esch-sur-Alzette, Luxembourg
[115] Department of Neurology, Sheba Medical Center, Ramat Gan, Israel
[116] Centre for Genomics Research, Precision Medicine and Genomics, IMED Biotech Unit, AstraZeneca, Cambridge, UK
[117] The State Key Laboratory of Brain and Cognitive Sciences, University of Hong Kong, Hong Kong, China
[118] Centre for Genomic Sciences, University of Hong Kong, Hong Kong, China
[119] Population Health and Immunity Division, the Walter and Eliza Hall Institute of Medical Research, Parkville 3052, VIC, Australia
[120] Department of Medical Biology, The University of Melbourne, Melbourne 3010, VIC, Australia
[121] Neuropsychiatry Laboratory, IRCCS Santa Lucia Foundation, Rome, Italy
[122] Division of Neuropsychiatry, Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, TX, USA
[123] IMT School for Advanced Studies Lucca, Lucca, Italy
[124] Department of Neurosciences, Psychology, Drug Research and Child Health, University of Florence, Florence, Italy
[125] Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital Frankfurt
[126] Department of Psychiatry, Psychotherapy and Psychosomatics, University Hospital Würzburg
[127] Division of Psychiatry, University College London, London, UK
[128] Division of Psychiatry, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK
[129] Institute for Molecular Medicine Finland, University of Helsinki, 00014, Finland
[130] Department of Neurology, Massachusetts General Hospital, Boston, MA, USA
[131] Department of Psychiatry and Behavioral Sciences, SUNY Downstate Medical Center, Brooklyn, NY, USA
[132] Department of Psychiatry, Stony Brook University, Stony Brook, NY, USA
[133] Faculty of Social and Human Sciences, University of Azores, PT
[134] Cherry Health and Michigan State University College of Human Medicine, Grand Rapids, MI, USA
[135] Institute of Medical Psychology, Faculty of Medicine, University of Coimbra, Coimbra, PT
[136] Department of Psychiatry, Wright State University, Dayton, OH, USA
[137] Departments of Psychiatry, Genetics & Genomics, Icahn School of Medicine at Mount Sinai, NY, USA
[138] Semel Institute for Neuroscience at UCLA, Los Angeles, CA, USA
[139] Departments of Public Health and Preventive Medicine, Family Medicine, and Psychiatry and Behavioral Sciences, State University of New York, Upstate Medical University, Syracuse, NY, USA
[140] Department of Psychiatry, University of North Carolina, Chapel Hill, NC, USA
[141] Department of Psychiatry & Behavioral Sciences, University of Southern California, Los Angeles, CA, USA
[142] Department of Psychiatry, Virginia Commonwealth University School of Medicine, Richmond, VA, USA
[143] Department of Psychiatry and Human Behavior, University of California, Irvine, CA, USA
[144] Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, GA, USA
[145] Department of Psychiatry, Veterans Administration New York Harbor Healthcare System, Brooklyn, NY, USA
[146] Broad Institute of MIT and Harvard, Cambridge, MA, USA
[147] Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA
[148] Department of Systems Biology, Harvard Medical School, Boston, MA, USA
[149] Bezmialem Vakif University, Institute of Life Sciences and Biotechnology, Istanbul, Turkey
[150] Department of Genetics, Harvard Medical School, Boston, MA, USA
[151] Department of Neurosurgery, University of Tübingen, Tübingen, Germany

Details of individual participating Epi25 cohorts

Australia: Austin Hospital, Melbourne (AUSAUS)
The Epilepsy Research Centre at the Austin Hospital in Melbourne, Australia, has been investigating the genetic basis of the epilepsies for over 20 years. The cohort in the Epi25 Collaborative were recruited to the epilepsy genetics research program over this period from the Austin Hospital, epilepsy clinics around Melbourne, and referrals from neurologists Australia-wide. Informed consent was obtained from patients or their parent/guardian as appropriate. DNA was extracted from blood or saliva samples. A skilled team of researchers and clinicians conducted detailed clinical phenotyping which involved a systematic review of medical records, including EEG and MRI reports, and a validated epilepsy questionnaire. Information on family history of seizures and other neurological disorders has also been collected via interviews with the patients and their families. Patients with GGE, non-acquired focal epilepsy, or a DEE were included in Epi25. A very heterogenous collection of epilepsy syndromes are represented in the cohort, including patients with EOAE, CAE, JME and late-onset GGE in the GGE cohort and patients with TLE, FLE, and benign childhood focal epilepsies in the non-acquired focal cohort. The DEE cohort is particularly heterogenous and includes patients with a range of DEE syndromes such as Ohtahara syndrome, Lennox-Gastaut syndrome, epilepsy with myoclonic-atonic seizures, and non-syndromic DEE. An additional subset of patients with lesional focal epilepsy, such as malformations of cortical development or acquired epilepsy, were also included, as well as a selection of patients with familial febrile seizures or FS+.

Most patients in the cohort are of European descent ('Anglo-Australian') although there is a diverse range of ethnic backgrounds including Asian, Middle Eastern, Indigenous Australian and mixed ethnicities. There is a known family history of seizures in 46% of the cohort (58% in the subset with GGE). The majority of the patients have had some previous genetic testing, including CNV testing and single gene testing. The DEE cohort have been extensively investigated with multiple iterations of a research panel of known, novel and putative genes for epilepsy. In addition, many patients with focal epilepsy have had a panel of known genes.

Australia: Royal Melbourne Hospital, Melbourne (AUSRMB)
The cohort includes patients that have been referred to a seizure clinic. These clinics include the First Seizure Clinic, Video EEG Monitoring Unit and the Epilepsy Clinic based at the Royal Melbourne Hospital. Patients were also referred from the private rooms of neurologists around Victoria.

Belgium: Antwerp (BELATW)
Patients were recruited by the Neurogenetics Group of the University of Antwerp through epilepsy clinics at the different university hospitals in Belgium. All patients were diagnosed with a (so far) unexplained presumed genetic epilepsy, and should have had at least 1 MRI of the brain excluding acquired causal lesions.The study was approved by the ethics committee of the University of Antwerp, and parents or the legal guardian of each proband signed an informed consent form for participation in the study. Genomic DNA of individuals was extracted from peripheral blood according to standard procedures. Clinical information was extracted from clinical files, as reported by their treating (paediatric) neurologists, and a subset was reviewed independently by two research team clinicians to ensure data quality and consistency.

Belgium: Brussels (BELULB)
Adult patients with epilepsy were recruited consecutively through outpatient clinics and hospitalizations at Hôpital Erasme, Brussels, Belgium (between October 2004 and June 2017)

and UZ Gasthuisberg, Leuven, Belgium (between October 2004 and June 2009). The study was approved by the Institutions' Review Boards. All patients provided written informed consent for data collection; patients with learning disability were included after consent from a parent or guardian. DNA was extracted from peripheral blood lymphocytes[1].

Canada: Andrade (CANUTN)
There are 96 patients (49 DEE, 3 NAFE, 44IGE) in the Andrade cohort, typically from the Greater Toronto region of Ontario, Canada. There are 44 males and 52 females, mostly of Caucasian ancestry, but also African, South Asian, East Asian, Latino, Middle Eastern, Jewish and Indigenous. Patients were recruited to each group through an REB protocol allowing for the collection of blood or saliva and data collaboration. Patients that were previously consented were reconsented to allow for whole exome sequencing and data sharing with the EPI25 group. After collection, the sample was de-identified, and then extracted and stored at the Hospital for Sick Children, Toronto, Canada.

Switzerland: Bern University Hospital and University of Bern, Bern (CHEUBB)
In the recruitment of our cohort, the Departments of Neurology and BioMedical Research, Bern University Hospital and University of Bern, Bern, Switzerland, and the Institute of Human Genetics, Bern University Hospital, Bern, Switzerland, were involved. The Swiss study population encompasses 28 patients with epilepsy between 2 and 63 years of age. All patients have been codified for the Epi25 Study. The patient ascertainment protocol was according to Epi25 phenotyping requirements. Phenotyping information was taken from medical records, stored in the hospital's database, and entered in codified form into a RedCap database provided by Ep25. DNA source was patients' venous blood. DNA was extracted with standard kits at the Institutes of Human Genetics or Clinical Chemistry of Bern University Hospital and stored there at -80 degrees C. Informed consent declarations are available from all patients and have been approved by the American Institutional Review Board involved in the Epi25 Study. The Cantonal Ethics Committee Bern, Switzerland, granted permission for participation of Bern University Hospital and University of Bern in the Epi25 Study including all steps described above. The patient cohort has not been described yet elsewhere.

Cyprus: The Cyprus Institute of Neurology and Genetics (CYPCYP)
Epilepsy-affected subjects of the Cyprus cohort were largely recruited and enrolled in the Epi25 Consortium by physicians during routine clinical visits in the Cyprus Institute of Neurology and Genetics. Phenotypic data were collected at the time of enrollment and submitted into the Epi25 RedCap database in a de-identified manner. There are 123 unrelated individuals of Southern European ancestry in the Cyprus cohort, 59 GGE subjects, 53 NAFE and 11 DEE. All subjects selected for this study had clinical, neuroimaging and EEG or video-EEG characteristics meeting the International League against Epilepsy (ILAE) 2017 Seizure Classification. The controls cohort consisted of a group of 32 individuals. of Southern European ancestry and were not diagnosed with epilepsy or other neuropsychiatric phenotypes. Genomic DNA samples were extracted from whole blood with the Gentra Puregene Blood Kit (Qiagen, Hilden, Germany) according to the manufacturer's guidelines. This study was carried out in compliance with the Cyprus National Bioethics Committee (ΕΕΒΚ/ΕΠ/2015/22). Written informed consent was obtained from all study participants or their legal guardians at the Cyprus Institute of Neurology and Genetics.

Czech Republic: University Hospital Motol, Prague (CZEMTH)
Patients in our cohort have been diagnosed with West syndrome, myoclonic-astatic epilepsy or developmental and epileptic encephalopathy of unknown aetiology. Brain magnetic resonance imaging and metabolic screening excluded any underlying pathology. Patients were collected at the Department of Child Neurology of the 2nd Medical Faculty and University Hospital Motol.

Legal guardians of patients signed an informed consent. The study was approved by the local ethics committee.

Germany: Epilepsy Center Frankfurt Rhine-Main, Goethe University, Frankfurt, and Epilepsy Center Hessen, Philipps University, Marburg (DEUPUM)
The principal investigators (Karl Martin Klein, Felix Rosenow, Philipp S. Reif, and Susanne Knake) contributed 259 samples. The patients were recruited from the outpatient clinics and the video EEG monitoring units at the Epilepsy Centers Frankfurt Rhine-Main and Hessen-Marburg and contained 64 samples with genetic generalized epilepsy, 130 samples with non-acquired focal epilepsy, 60 samples with lesional focal epilepsy and 4 samples with DEE. All patients were phenotyped in detail by epilepsy specialists (KMK, FR, SK, PSR) within the EpimiRNA project (European Union's 'Seventh Framework' Programme (FP7) under Grant Agreement no. 602130). EEGs and MRIs were performed as part of the clinical workup. Phenotypic classification and data entry for the biobank for paroxysmal neurological disorders was performed by KMK. Patient selection and data export from the biobank for the study was performed by KMK. DNA was extracted from peripheral blood or saliva. All patients provided written informed consent.

Germany: University of Bonn, Bonn (DEUUKB)
The sample recruitment site is the Department of Epileptology at the University of Bonn. The collection of blood DNA samples from patients with epilepsy was conducted from 2007 till 2015 within the studies Epicure (Functional Genomics in Neurobiology of Epilepsy: A Basis for New Therapeutic Strategies) and NGEN-Plus (Genetic basis of Levetiracetam pharmacoresistance and side effects in human epilepsy) and has been approved by the Ethics committee of University Bonn Medical Center (040/07). Genomic DNA was isolated from 10 ml aliquots of EDTA-anticoagulated blood by a salting-out technique[2]. From selected samples of this cohort GWAS data have been published in several studies[3-6].

Germany: University Hospital Schleswig-Holstein, Kiel (DEUUKL)
Patients were recruited by the Neuropedriatics Group of the University Hospital of Schleswig-Holstein and through the Israeli-Palestinian Family Consortium. The recruitment and analysis of these samples is covered by the Kiel IRB. Patients, their parents or the legal guardian of each proband signed an informed consent form for participation in the study. Clinical data was collected from clinical files and a subset of patients from Israel or Palestine was interviewed by a research team of clinicians to provide their clinical data. Genomic DNA of patients was extracted from peripheral blood according to standard procedures.

Germany: TLE Leipzig (DEUULG)
Patients were recruited by the Swiss Epilepsy Center in Zurich, Switzerland and samples were transferred for research and storage to the Institute of Human Genetics at the University of Leipzig, Germany. All patients were diagnosed with temporal lobe epilepsy due to an indicative EEG. Most patients had at least 1 MRI of the brain with focus on focal abnormalities, especially of the temporal lobe / hippocampal structures. The study was approved by the "Kantonale Ethikkommission Zürich". Parents or the legal guardian of each proband signed an informed consent form for participation in research studies including whole genome analyses. Genomic DNA of individuals was extracted from peripheral blood according to standard procedures. Clinical information was extracted from clinical files, as reported by their treating neurologists.

Germany: University of Tübingen, Tübingen (DEUUTB)
Our study cohort consists of over 1000 samples with mainly Caucasian origin. These samples were recruited at Tubingen and 38 other cooperating departments of neurology of university clinics and outpatient clinics in Germany. The Ethics / informed consent was approved by the

ethics committee of the Medical Faculty of the University and the University Clinic of Tübingen. Patients with the diagnosis of genetic generalized epilepsies, epileptic encephalopathies and non-acquired focal epilepsies were systematically recruited by a letter of invitation. After approving the informed consent by the participants, the data were collected retrospectively from medical reports and in exceptional cases, by personal interviews of patients or their relatives. The DNA source was blood.

Finland: Kuopio University Hospital, Kuopio (FINKPH)
Patients diagnosed with epilepsy and visiting Epilepsy Center, Kuopio University Hospital, Finland have given their written informed consent to record their clinical data to a research registry and collect a blood sample for DNA analysis. The ethics committee of Kuopio University Hospital has approved the study.

Finland: University of Helsinki, Helsinki (FINUVH)
Patients were recruited at the University of Helsinki through pediatric neurology clinics in the Helsinki and Tampere University Hospitals in Finland. All patients were diagnosed with a presumed genetic epilepsy, the etiology remaining unknown. All patients had an MRI done to exclude acquired causal lesions. The study was approved by an ethics committee of The Hospital District of Helsinki and Uusimaa, Finland. The parents or the legal guardian of each proband has signed an informed consent form. Genomic DNA of the patients was extracted from peripheral blood or saliva according to standard procedures. Clinical information was extracted from clinical files, as reported by the treating pediatric neurologists.

Wales: Swansea (GBRSWU)
The samples from Wales: Swansea are recruited from regional NHS HealthBoard Clinics and consented into the IRAS-ethical permissions framework of the Swansea Neurology Biobank (17/WA/0290). Informed consent is given for samples to be used for research purposes and for third-party consortia with appropriate MTA agreements. Patents attend epilepsy or general neurology clinics and are prioritised if they reach the clinical evidence for submission to studies. Blood samples are sent to the UK Porton Down EcACC facility and returned to SNB in batches where they are checked and validated through gender testing. The samples submitted to Epi25 must pass the DNA QC standards for NGS pipelines and have the level of certainty for clinical diagnosis.

UK: University College London, London (GBRUCL)
Participant recruitment took place at the National Hospital for Neurology and Neurosurgery (United Kingdom). Written informed consent or assent was obtained between 10/01/2000 and 01/25/2015 from all participants according to local and national requirements and blood samples were collected for DNA extraction. 709 epilepsy cases were submitted for analysis. Allocation to the following groups was based on the clinical diagnosis and the specific inclusion and exclusion criteria of the Epi25 consortium: generalized genetic epilepsy (n=393, 145 male), non-acquired focal epilepsy (n=313, 146 male), developmental and epileptic encephalopathy (n=3, 2 male). Additionally, relatives were included, where samples were available (n=3, 2 male). Phenotypic information was obtained from local medical records by clinical or trained non-clinical researchers.

UK: University of Liverpool, Liverpool and Imperial College London, London (GBRUNL)
GBRUNL samples are derived from four separate, UK-wide, ethically approved studies coordinated by the University of Liverpool (UK) and Imperial College London (UK). The SANAD and MESS linked DNA Bank and Relational Database study recruited individuals with newly-diagnosed focal, generalised or unclassified epilepsy from out-patient neurology clinics between 2003-2006[7; 8]. The Pharmacogenetics of GABAergic Mechanisms of Benefit and Harm in Epilepsy

study recruited patients with refractory focal epilepsy, previously or prospectively exposed to adjunctive treatment with clobazam or vigabatrin, from out-patient neurology clinics between 2005-2009. The Refractory Juvenile Myoclonic Epilepsy Cohort (ReJuMEC) study recruited individuals with valproic acid resistant juvenile myoclonic epilepsy from out-patient neurology clinics between 2009 and 2010. The ongoing Standard and New Antiepileptic Drugs (SANAD-II) study, which is recruiting individuals with newly-diagnosed focal, generalised or unclassified epilepsy from out-patient neurology clinics between 2013-2019. In all cases, study participants provided written informed consent to the collection (via blood or saliva sampling) and analysis of their DNA for use in genetic and pharmacogenetic research related to epilepsy and its treatment. All studies were approved by research ethics committees in operation at the relevant time (SANAD DNA bank, North West MREC ref 02/8/45; GABAergic mechanisms, UCLH REC ref 04/Q0505/95; ReJuMEC, Cheshire REC ref 09/H1017/55; SANAD-II, North West REC ref 12/NW/0361). Assembly of the GBRUNL cohort was supported by generous funding from The Wellcome Trust, the Imperial College NIHR Biomedical Research Centre, the Department of Health (UK), the Medical Research Council (UK), and the National Institute of Health Research (UK).

Hong Kong: Chinese University of Hong Kong (HKGHKK)
Epilepsy patients of Han Chinese ethnicity aged between 2 and 91 years were recruited from neurology clinics of five regional hospitals in Hong Kong covering a combined catchment population of approximately 3 million. Syndromic classification was adapted from the revised international organization of phenotypes in epilepsy. DNA was extracted from venous blood. The study was approved by ethics committees of the participating hospitals, and all patients or their legal guardians gave written informed consent. The sample collection methodology has been described previously[9].

Croatia: University Clinical Centre Zagreb, Zagreb (HRVUZG)
Pediatric patients were recruited from University Medical Centre Zagreb and 2 patients from 2 other epilepsy clinics in Croatia. All patients were diagnosed as possible genetic epilepsy not yet explained. All patients underwent MR brain imaging at least once, the acquired epilepsy causes were excluded. DNA was extracted from peripheral blood according to the accepted protocol. The study was approved by Hospital ethical committee and all parents or legal guardian of probands signed informed consent for participation in the study. Clinical information was extracted from clinical files. The cohort was also reviewed by reviewed by collaborative research team clinicians from University of Antwerp to ensure data quality and consistency.

Ireland: Dublin (IRLRCI)
Patients were recruited from a specialized epilepsy clinic at Beaumont Hospital and St. James' Hospital, Dublin, Ireland. Patients were mostly of Irish ethnicity. DNA was extracted from a combination of lymphocytes and saliva. All participants provided written informed consent.

Italy: Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan (ITAICB)
Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy (ITAICB): Our cohort included the DNA samples of 207 patients and 62 controls (healthy subjects not related with the patients, and without epilepsy history). The patient population included 106 patients with generalized epilepsy (GGE), 51 patients with developmental and epileptic encephalopathy (DEE), 33 patients with focal epilepsy due to cerebral malformations (nodular heterotopia), and 15 patients with non-acquired focal epilepsy (NAFE). All of the patients were diagnosed and followed at our Institute. Diagnosis of epilepsy was based on clinical, EEG and neurophysiological data, neuroimaging (MRI). Metabolic screening, karyotype, CGH array, analyses of single genes and customized panels were performed in some cases, when appropriate. The patients did not undergo to exome

sequencing analysis (before the Epi25 collection). The DNA of the patients was extracted from peripheral blood, according with standard procedures, after signature of an Informed Consent form. The genetic study was approved by The Ethic Committee of our Institute. No publications have described genetic findings pertaining to the collected patients until now. Clinical information was extracted from clinical files, as reported by their treating (paediatric and adult) neurologists.

Italy: Gaslini Institute, Genova (ITAIGI)
Patients with generalized and focal epilepsy or developmental and epileptic encephalopathy followed- up or referred for genetic analysis at 'Gaslini Institute'. The study was approved by the IRB and written informed consent was signed by the patients/parents. Clinical information, including data on EEG and antiepileptic therapy, were recorded on data collection forms. Genomic DNA isolation and genetic analysis was carried out with the Nimblegen-SeqCapEZ-V244M enrichment kit on the Illumina HiSeq2000 system.

Italy: IRCCS Institute of Neurological Science of Bologna, Bologna (ITAUBG)
Patients were recruited by the Adult and Pediatric Neurologists of the IRCCS Institute of Neurological Sciences, Bellaria Hospital, Bologna. Patients with generalized epilepsy, focal epilepsy (with or without brain lesions) and developmental and epileptic encephalopathy were referred by epilepsy clinics. All patients were diagnosed with a (so far) epilepsy of uncertain aetiology and underwent neuro-radiological imaging (CT or MRI) and EEG. The study was approved by the local ethics committee (cod. CE:16057). Patients themselves or parents or the legal guardian of each proband signed an informed consent form for the study participation. Genomic DNA of individuals was extracted from peripheral blood according to standard procedures. Clinical information was collected from medical records, as reported by their treating neurologists[10; 11].

Italy: University Magna Graecia, Catanzaro (ITAUMC)
Patients were recruited by the Epilepsy Group of the University Magna Graecia of Catanzaro (Italy) that includes a Pediatric and Adult Neurologic Unit with a specific focus on genetic epilepsy. In each patient, the diagnosis of epilepsy syndrome is based on comprehensive clinical, neuropsychological, electroencephalographic, and MR evaluations. Clinical data are stored into a database. The study was approved by the ethics committee of the University of Catanzaro Italy, and parents or the legal guardian of each proband signed an informed consent form for participation in the study. Genomic DNA of individuals was extracted from peripheral blood according to standard procedures

Italy: Meyer Hospital, Florence (ITAUMR)
Patients were recruited by the Neurology and Neurogenetics Group of the Meyer Hospital of Florence. All patients were diagnosed with a unexplained presumed genetic epilepsy. The study was approved by the Pediatric Ethics Committee of the Regione Toscana, and parents or the legal guardian of each proband signed an informed consent form for participation in the study. Genomic DNA of individuals was extracted from peripheral blood according to standard procedures. Clinical information was extracted from clinical files, as reported by their treating (paediatric) neurologists.

Japan: RIKEN Institute, Tokyo (JPNRKI)
Japanese patients with epilepsies were recruited by National Epilepsy Center, Shizuoka Institute of Epilepsy and Neurological Disorder. Epileptic seizures and epilepsy syndrome diagnoses were performed according to the International League Against Epilepsy classification of epileptic syndromes. Genomic DNA was extracted from peripheral venous blood samples using QIAamp DNA Blood Midi Kit according to manufacturer's protocol (Qiagen). The experimental protocols

were approved by the Ethical Committee of RIKEN Institution and National Epilepsy Center, Shizuoka Institute of Epilepsy and Neurological Disorder. Written informed consent was obtained from all individuals and/or their families in compliance with the relevant Japanese regulations.

Lithuania: Vilnius University Hospital Santaros Klinikos, Vilnius (LTUUHK)
Patients were recruited in Vilnius University Hospital Santaros Klinikos by a clinical geneticist through a referral of a neurologist or a child neurologist and according to inclusion/ exclusion criteria. All patients were diagnosed with one of the three forms of epilepsy - genetic generalized, focal non-lesional or developmental and epileptic encephalopathy - after exclusion of acquired causes and should have had at least 1,5T brain MRI and a diagnostic EEG. The study was approved by Vilnius Regional Biomedical Research Ethics Committee, and each proband or parents/ legal guardians of a proband signed an informed consent form for participation in the study. Samples of genomic DNA were obtained during the routine procedure for blood sampling for genetic testing done in a clinical testing and the majority of patients had chromosomal microarray, metabolic testing and/or gene/gene panel testing prior to the inclusion into the study. Clinical information was extracted from clinical files and obtained during the clinical genetic consultation.

New Zealand: University of Otago, Wellington (NZLUTO)
Cases were recruited as part of a larger study from neurology, paediatric and genetic outpatient services throughout New Zealand. Participants were between 1 month and 63 years of age from the following ethnic groups: New Zealand European, Māori, Pasifika, Asian, Hispanic, Ethiopian. Using a structured interview and review of medical records diagnosis was based on the International League of Epilepsy (ILAE) classification and made by a paediatric neurologist. The study protocol was approved by the New Zealand Health and Disability Ethics Committee. Participants gave written informed consent for clinical and genetic analysis. DNA was extracted from blood or saliva.

Turkey: Bogazici University, Istanbul (TURBZU)
Patients were recruited by the Child Neurology and Neurology clinics at the different university hospitals in Turkey. The study was approved by the Institutional Review Board for Research with Human Subjects (INAREK) of Boğaziçi University, and parents or the legal guardian of each proband signed an informed consent form for participation in the study. Genomic DNA of individuals was extracted from peripheral blood according to standard procedures. Clinical information was reported by their treating (pediatric) neurologists. All developmental and epileptic encephalopathy patients had severe epilepsy, with developmental delay and regression, normal neuroimaging and epileptiform activity on EEG. Turkish population control group included individuals with no symptoms of any neurological disorder.

Turkey: Istanbul University, Istanbul (TURIBU)
Epilepsy patients were recruited from Epilepsy Clinic, Department of Neurology, Istanbul Faculty of Medicine, Istanbul University. The study population consisted of patients with idiopathic/genetic generalized epilepsies, lesional or non-lesional focal epilepsies and epileptic encephalopathies, including sporadic and familial cases. All patients were long-term follow-up. Seizure types, age of onset, neurological examinations, past and family history, prognosis and response to treatment, features of electroencephalography and neuroimaging were evaluated. Ethics committee approval was obtained. Peripheral blood samples were collected from all patients following written informed consent. DNA isolation was performed in Department of Genetics, Aziz Sancar Institute of Experimental Medicine, Istanbul University.

USA: Boston Children's Hospital, Boston (USABCH)

Cases from Boston Children's Hospital (BCH) were ascertained from 3 local repositories. All repository protocols are approved by the BCH Institutional Review Board and participants were consented under one (or more) of the following protocols. The Genetics of Epilepsy and Related Disorders protocol, led by Dr. Annapurna Poduri, enrolls patients with a clinical epilepsy diagnosis for genotype/phenotype correlation. Samples are obtained from BCH and non-BCH patients and biological samples collected for genetic sequencing. Patient medical records (BCH and outside records) are reviewed for phenotyping purposes[12]. The Phenotyping and Banking Repository of Neurological Disorders is a local repository led by Dr. Mustafa Sahin. BCH patients with any neurological phenotype, including epilepsy, are enrolled and biological samples collected. Boston Children's Biobank for Health Discovery is a local repository led by Dr. Kenneth Mandl that enrolls any patient of BCH, regardless of diagnosis or phenotype. Samples from these two broader repositories are available to BCH researchers through an application process, including a supporting IRB-approved protocol. Patients with a clinical diagnosis of epilepsy were reviewed for Epi25 eligibility using their BCH medical records.

USA: Philadelphia/CHOP (USACHP) and Philadelphia/Rowan (USACRW)

The Philadelphia Cohort began in 1997 and collected blood, saliva and brain tissues from patients with common forms of idiopathic human epilepsy, mostly genetic generalized epilepsy (GGE) and non-acquired focal epilepsy (NAFE). The collection began at Thomas Jefferson University Hospital in Philadelphia and expanded to include six other sites: The Children's Hospital of Philadelphia, The University of Pennsylvania, The University of Cincinnati, Nationwide Children's Hospital, Beth Israel Deaconess and The University of Montreal. The cohort consists of 2615 samples from epilepsy patients collected and supported during two periods of NIH funding (R01NS493060, 2001-2007 RJ Buono PI and R01NS06415401, 2009-2012 RJ Buono and H Hakonarson Co- PI). Over 1000 additional samples from first degree relatives of the patients were also collected. Many of these samples are available to the research community via the NINDS sample repository at the Coriell Institute in Camden NJ. All studies were approved by Institutional Review Boards at each participating site. All patients were identified and recruited by trained epileptologists at tertiary care centers using inclusion and exclusion criteria previously published. Diagnostic methods applied included EEG, MRI, and collection of deep phenotypic information on family history, medications, risk factors, age of onset, and other information. Inclusion and exclusion criteria for the cohort were published previously[6; 13]. For the Epi25K project, blood and saliva were used as the source of DNA.

USA: Epilepsy Phenome/Genome Project (USAEGP)

Infantile spasms (IS), Lennox–Gastaut syndrome (LGS), genetic generalized epilepsy (GGE), and non-acquired focal epilepsy (NAFE) patients were collected through the Epilepsy Phenome/Genome Project[14] (EPGP, http://www.epgp.org). More than 4,000 participants in EPGP were enrolled across 27 clinical sites from around the world. The subset of samples included in Epi25 were enrolled from 20 sites across the USA and in Australia. IS patients were required to have hypsarrhythmia or a hypsarrhythmia variant on EEG. LGS 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). IS and LGS cases were enrolled as trios with both biological parents. Participants with NAFE and GGE were required to have a first degree relative who also had NAFE or GGE (did not have to be concordant). All patients had no 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. Enrollment required detailed confirmation of detailed phenotypic data including medical record review and abstraction, patient

interviews, EEG and MRI, and comprehensive review by expert scientific cores for EEG, MRI, and clinical final diagnosis.

USA: NYU Human Epilepsy Project (USAHEP)
Participants were recruited for the Human Epilepsy Project at 33 different medical centers located in US, Canada, Australia, Austria, Finland, and Ireland. All participants were between 12 and 60 at the age of enrollment and had a clinical history consistent with a diagnosis of focal epilepsy, as determined by an eligibility panel of epilepsy specialists. Participants were required to have two or more spontaneous seizures with clinically observable features in the past 12 months, and 4 or fewer months of anticonvulsant treatment. Those with major medical comorbidities, intellectual disability, or significant psychiatric disease were excluded, as were those with progressive neurological lesions on imaging or known neurodegenerative disease. Participants completed daily electronic diaries tracking seizures, medication adherence, and mood. Mood and cognition were assessed periodically via standardized instruments, and brain MRIs and EEGs were obtained for all participants. Blood was collected and banked annually, allowing for study of DNA, RNA and protein. HEP was approved by the IRBs at all participating sites, and all participants or their parent/legal guardian gave written informed consent. Minors also gave written assents. HEP was funded by the Epilepsy Study Consortium.

USA: Penn/CHOP, Philadelphia (USAUPN)
The USAUPN cohort was recruited at the Children's Hospital of Philadelphia (CHOP) and Hospital of the University of Pennsylvania (UPenn), including pediatric and adult patients with epilepsies through dedicated IRB protocols at CHOP and UPenn. Patients were recruited in an inpatient and outpatient setting and samples were also contributed through the Penn Biobank.

Details of individual control cohorts (not including samples obtained from dbGaP)

Italian controls (PI: Spalletta)
Right-handed healthy individuals were recruited and had whole blood drawn by local advertisements at Santa Lucia Foundation in Rome, Italy. All of the individuals were born and educated in Italy and had Italian-Caucasian ancestry, to reduce the possibility of artifactual association caused by ethnic stratification. Exclusion criteria were: (i) major medical illnesses and/or known or suspected history of alcoholism or drug dependency and abuse; (ii) mental disorders (i.e. schizophrenia, mood, anxiety, personality and/or any other significant mental disorders) according to the DSM-IV-TR criteria assessed by the Structured Clinical Interviews for DSM-IV-TR [SCID-I and SCID-II] and/or neurological disorders diagnosed by an accurate clinical neurological examination; (iii) presence of vascular brain lesions, brain tumour and/or marked cortical and/or subcortical atrophy on magnetic resonance imaging (MRI) scan; and (iv) suspicion of cognitive impairment or dementia based on Mini Mental State Examination (MMSE) scores ≤24 (a cut-off point for dementia screening in the Italian population)[15] and confirmed by a clinical neuropsychological evaluation using the Mental Deterioration Battery[16] and the NINCDS-ADRDA criteria for dementia[17]. The presence of anxiety symptoms was assessed using the Hamilton Rating Scale for Anxiety (HAM-A)[18]. Written, informed consent was obtained from all subjects participating in the study, which was approved by the local ethics committee at the Santa Lucia Foundation of Rome (protocol number CE/11.9)[19].

German controls
Subjects have been recruited at the Department of Psychiatry and Psychotherapy, University of Würzburg, Germany, with the exception of the TK samples (n=63, they are anonymous blood donors). All subjects have been screened for the absence of mental disorders (by MINI) as well

as severe medical and neurological (including epilepsy) disorders (by self-report). Ethnicity is Caucasian by self-report in all cases, and DNA source is blood. Studies were approved by the IRB, University Hospital Würzburg; all participants gave written informed consent.

UK/IRL controls 1 (PI: McQuillin)
The UCL control sample consisted of 480 genomic DNA samples that were extracted from EBV transformed peripheral blood lymphocytes from unscreened healthy British blood donors (https://www.phe-culturecollections.org.uk/products/dna/hrcdna/hrcdna). The remaining DNA samples were extracted from whole blood samples from healthy volunteers of UK or Irish ancestry who were interviewed with the initial clinical screening questions of the SADS-L and selected on the basis of not having a past or present personal history of any RDC-defined mental disorder. Heavy drinking and a family history of schizophrenia, alcohol dependence or bipolar disorder, were also used as exclusion criteria for controls. UK National Health Service multi-centre and local research ethics approvals were obtained and all subjects gave signed informed consent[20].

UK/IRL controls 2 (PIs: McIntosh, Blackwood, and Johnstone)
Participants were recruited from clinical service around Edinburgh and Scotland and screened using the SADS-L. DNA samples were extracted from whole blood for genotyping and sequencing. Research was conducted after research ethics and NHS management approvals[21].

FINRISK controls
The controls from FINRISK that contributed to the Epi25 WES study were part of the FINRISK inflammatory bowel disease (IBD) cohort. The population-based FINRISK study have been followed up for IBD and other disease end-points using annual record linkage with the Finnish National Hospital Discharge Register, the National Causes-of-Death Register and the National Drug Reimbursement Register. Controls were chosen to have a high polygenic risk score for IBD without an IBD diagnosis. A detailed description of the FINRISK cohort can be found at Borodulin *et al*[22].

Genomic Psychiatry Cohort (GPC) controls
The controls from GPC that contributed to the Epi25 WES study were a subset of the overall control participants of European ancestry with no personal or family history of schizophrenia or bipolar disorder. All the samples were exome-sequenced at the Broad Institute. A detailed description of the GPC cohort can be found at Pato *et al*[23].

Supplemental acknowledgments

Additional funding sources and acknowledgment of individual patient cohorts are as follows: The **AUSAUS** cohort was supported by NHMRC Program Grant (ID: 1091593). Effort at **AUSRMB** was supported by NHMRC Project Grant (#1059858) and NHMRC Program Grant (#1091593). H.S. from the **BELATW** cohort is a PhD fellow of the Fund for Scientific Research Flanders (FWO, 1125416N); and S.W. receives funding support from the BOF-University of Antwerp (FFB180053) and FWO (Senior Clinical Investigator Fellowship 1861419N). Funding that supported activities at the **BELULB** site included Fonds National de la Recherche Scientifique, Fonds Erasme, Université Libre de Bruxelles, and European Commission grant 279062. Work conducted at **CANUTN** was supported by funding from Ontario Brain Institute – EpLink. The **CHEUBB** cohort was supported by directorate for Public Health and Social prevention of the Canton of Bern: Medical Innovation “next-generation-sequencing epilepsy” 2015. Work at the **CYPCYP** site was funded by TELETHON/33173158. Effort at **CZEMTH** was supported by Ministry of Health of Czech Republic AZV 15–33041 and DRO 00064203. At **DEUPUM**, all patients were

phenotyped in detail by epilepsy specialists (KMK, FR, SK, PSR) within the EpimiRNA project (European Union's 'Seventh Framework' Programme (FP7) under Grant Agreement no. 602130). KMK, FR and PSR were supported by the Detlev-Wrobel-Fonds for Epilepsy Research Frankfurt. At **DEUUKB**, sample recruitment of the University Bonn cohort was funded by the European Community (FP7 project EpiPGX, grant 279062 to WSK). The work at **USAUPN** and **DEUUKL** was supported by intramural funds of the Children's Hospital of Philadelphia, the German Research Foundation (HE5415/3-1 to I.H.) within the EuroEPINOMICS framework of the European Science Foundation, the German Research Foundation (DFG, HE5415/5-1, HE5415/6-1, HE5415/7-1 to I.H.), and the Genomics Research and Innovation Network (GRIN, grinnetwork.org), a collaboration amongst the Children's Hospital of Philadelphia, Boston Children's Hospital, Cincinnati Children's Hospital Medical Center. I.H. also received support through the International League Against Epilepsy. This work was also partially supported by NIH grant NS108874. Further funding for the **DEUUTB** cohort was obtained from the German Research Foundation and the Fond Nationale de la Recherche in Luxembourg in the frame of the Research Unit FOR-2715 (grants Le1030/16-1, Kr5093/2-1, We4896/4-1, He5415/7-1), by the German Society for Epileptology, and by the foundation no-epilep. The **FINKPH** cohort was supported by funding from the Saastamoinen Foundation, Vaajasalo Foundation and Kuopio University Hospital State Research. Effort at **FINUVH** was supported by Folkhälsan Research Foundation, Helsinki Finland and Foundation for Pediatric Research, Helsinki, Finland. Funding for the **GBRSWU** cohort included Health and Care Research Wales, Wales Gene Park, The BRAIN Unit. The **GBRUCL** cohort thanks the Epilepsy Society, Muir Maxwell Trust and Wellcome Trust Strategic Award (WT104033AIA) for funding support. The work conducted at **GBRUCL** was undertaken at University College London Hospitals, which received a proportion of funding from the NIHR Biomedical Research Centres funding scheme. Assembly of the **GBRUNL** cohort was supported by generous funding from The Wellcome Trust, the Imperial College NIHR Biomedical Research Centre, the Department of Health (UK), the Medical Research Council (UK), and the National Institute of Health Research (UK). The **HKGHKK** cohort was supported by Research Grants Council of the Hong Kong Special Administrative Region, China (project number CUHK4466/06M). The **HRVUZG** cohort acknowledges Sarah Weckhuysen and Hannah Stamberger for help with sample shipping. The work at **IRLRCI** was supported by research grants from Science Foundation Ireland (SFI) (16/RC/3948 and 13/CDA/2223) and co-funded under the European Regional Development Fund and by FutureNeuro industry partners. Effort at **ITAICB** was supported by European project, Grant agreement 602531 (DESIRE) (Silvana Franceschetti only was involved) The **ITAIGI** cohort thanks the Genetic Commission of Italian League Against Epilepsy (LICE) for its contribution to this work. The **ITAUBG** cohort thanks Dr Baldassari S; Dr Caporali L and Prof Carelli V and all the staff of the Laboratory of Neuro-genetics in Bologna; and Dr Margherita Santucci and the Child Neurology Unit, Bellaria Hospital in Bologna, as well as all physicians and nurses of the Epilepsy Centre from the Bellaria Hospital in Bologna. Work at **ITAUMR** was supported by the European Commission Seventh Framework Programme under the project DESIRE (grant agreement no. 602531) (to R.G.). Research at the **JPNRKI** cohort was (partly) supported by AMED under Grant Number JP18dm0107092 and JP18ek0109388 (K.Y.). Effort at **NZLUTO** was supported by Health Research Council of New Zealand: 15/070 – Gene Discover in Epilepsy: the building block of precision medicine; Cure Kids New Zealand: 3576 – Genetics of Epilepsy. The **TURBZU** cohort was supported by Bogazici University Research Fund, Project No: 17B01P3. **USACHP** and **USACRW** were supported by NIH funding (R01NS493060, 2001-2007 RJ Buono PI and R01NS06415401, 2009-2012 RJ Buono and H Hakonarson Co- PI). U01HG006830 (PI Hakonarson) and The Children's Hospital of Philadelphia Endowed Chair in Genomic Research (PI Hakonarson). Work at **USAEGP** was supported by National Institute of Neurological Diseases and Stroke (NINDS) grant U01 NS053998, as well as planning grants from the Finding a Cure for Epilepsy and Seizures (FACES) Foundation and the Richard Thalheimer Philanthropic Fund. At **USAHEP**, H.E.P is sponsored by The Epilepsy Study Consortium (ESCI).

ESCI is a non-profit organization dedicated to accelerating the development of new therapies in epilepsy to improve patient care. The funding provided to ESCI to support HEP comes from industry, philanthropy and foundations (UCB Pharma, Eisai, Pfizer, Lundbeck, Sunovion, The Andrews Foundation, The Vogelstein Foundation, Finding A Cure for Epilepsy and Seizures (FACES), Friends of Faces and others). C.A.E. from the **USAUPN** cohort was supported in part by a Ruth L. Kirschstein National Research Service Award (NRSA) Institutional Research Training Grant, T32 NS091008-01.

Additional funding sources and acknowledgment of individual control cohorts are as follows: Collection of the **Italian controls** was supported by RF-2013-02359074; NET-2011-02346784 (Italian Ministry of Health) and was sequenced as part of the CCDG funding. Sequencing of the **German controls** was funded by the Dalio Foundation DFG grant TRR-CRC 58, Z02. The **UK/IRL controls** were collected and sequenced with support from the Neuroscience Research Charitable Trust, the Central London NHS (National Health Service) Blood Transfusion Service and the National Institute of Health Research (NIHR) funded Mental Health Research Network. The **FINRISK controls** were part of the FINRISK studies supported by THL (formerly KTL: National Public Health Institute) through budgetary funds from the government, with additional funding from institutions such as the Academy of Finland, the European Union, ministries and national and international foundations and societies to support specific research purposes. The **GPC controls** were sequenced with funding from NIH grants MH085548 and MH085542 and the Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard. Public controls used for the analysis were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/gap. Controls from the MIGen studies were obtained through dbGaP accession numbers phs000806.v1.p1, phs001000.v1.p1, and phs000814.v1.p1. We thank the Broad Institute for generating high-quality sequence data for the MIGen studies supported by NHGRI funds (grant # U54 HG003067) with Eric Lander as PI. The Sweden-Schizophrenia (SCZ) controls were obtained through dbGaP accession number phs000473.v2.p2. These samples were provided by the Swedish Cohort Collection supported by the NIMH grant R01MH077139, the Sylvan C. Herman Foundation, the Stanley Medical Research Institute and The Swedish Research Council (grants 2009-4959 and 2011-4659). Support for the Swedish-SCZ exome sequencing was provided by the NIMH Grand Opportunity grant RCMH089905, the Sylvan C. Herman Foundation, a grant from the Stanley Medical Research Institute and multiple gifts to the Stanley Center for Psychiatric Research at the Broad Institute of MIT and Harvard.

Supplemental Figures

(A) Raw data

(B) Pre-filtered

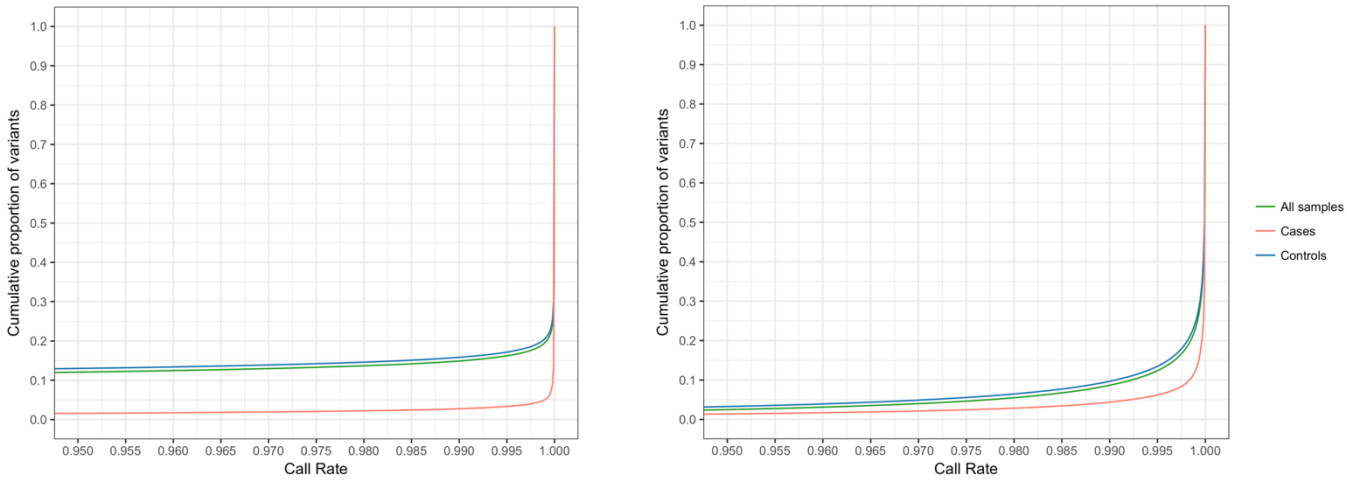
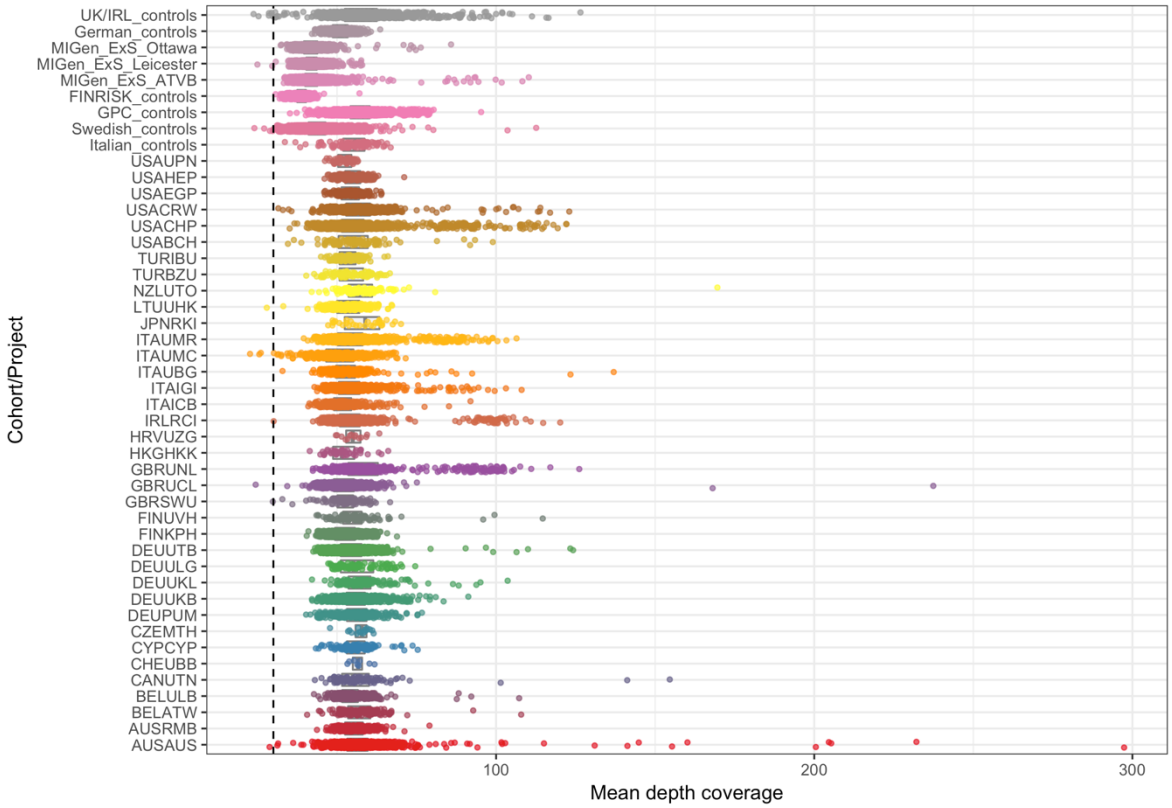


Figure S1. Variant pre-filtering: restricting to well-covered exome intervals across platforms reduced variant calling bias due to capture difference

As the first step of the QC process, we removed sites that fell outside regions with at least 10x coverage in 80% of the samples across sequencing batches and platforms (Illumina and Agilent), regardless of targeted capture intervals. This pre-filtering effectively minimized the call rate difference between cases and controls.

(A)



(B)

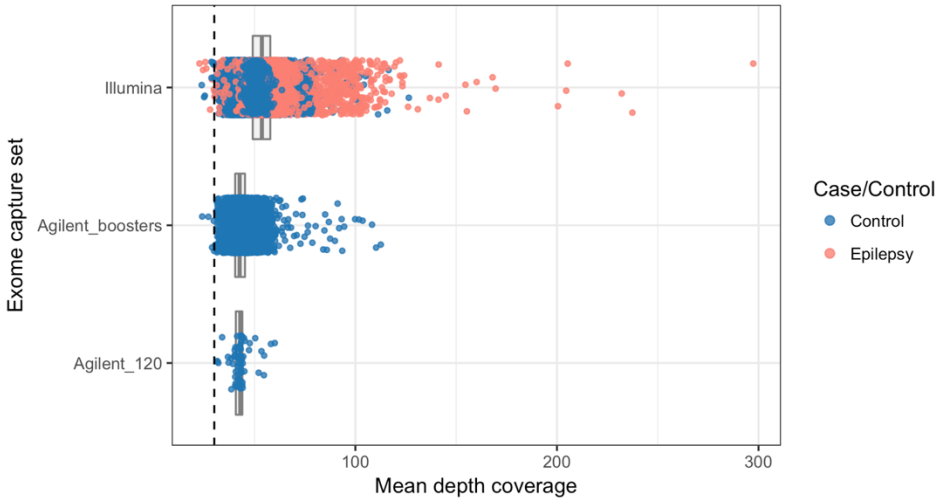
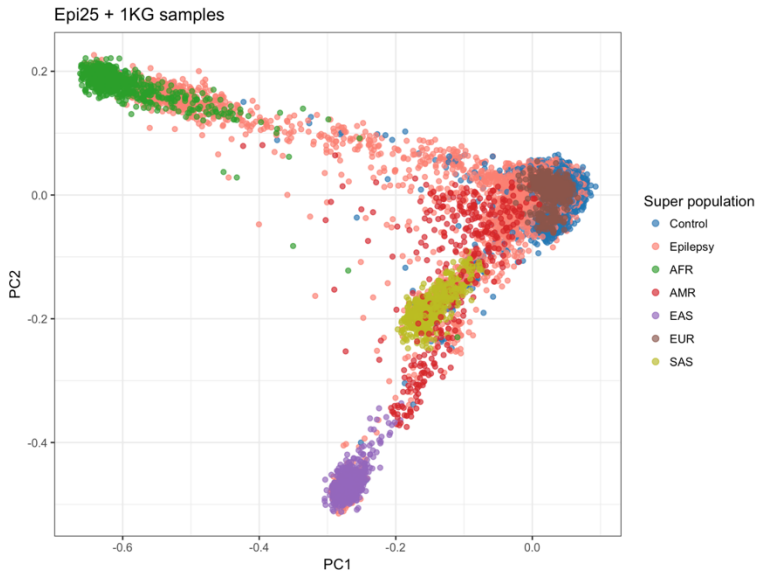


Figure S2. Average sample sequencing depth for pre-filtered data across cohorts (A) and by capture platform (B)

Cohorts with a six-letter name were case cohorts, and the rest were controls cohorts. Epilepsy cases overall had better average coverage than control samples due to capture platform difference (Illumina vs. Agilent exome enrichment kit). Dashed lines indicated the QC threshold, below which samples were removed (mean depth < 30).

(A)



(B)

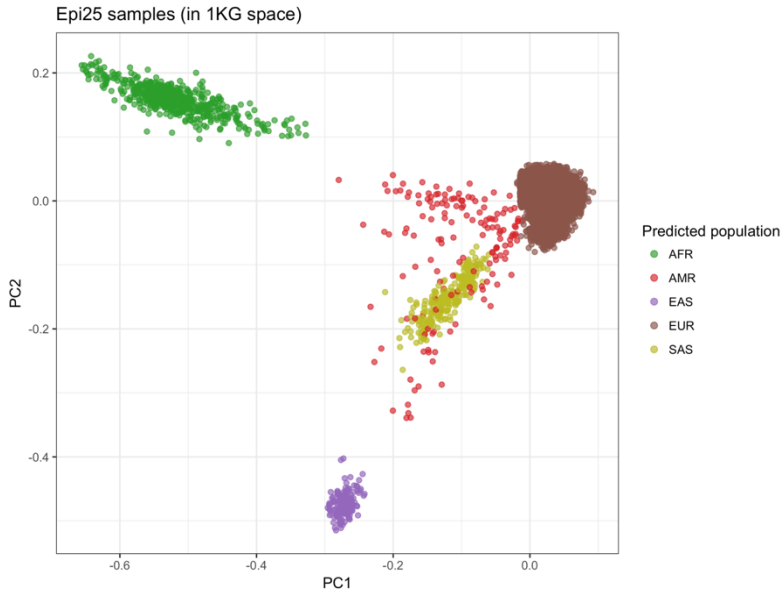
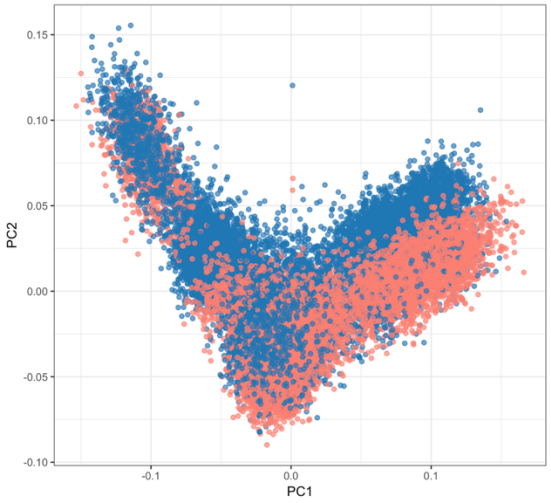
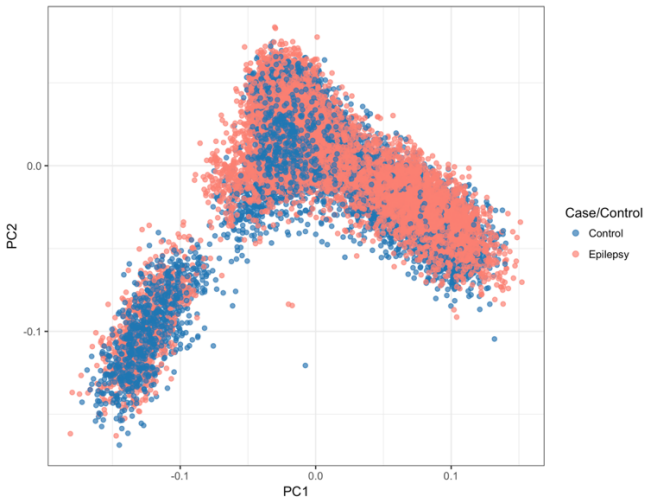


Figure S3. Population structure of the initial cases and controls based on principal component analysis (PCA)
(A) PCA was run on the study samples along with 1000 Genomes (1KG) phase 3 super populations to infer genetic ancestry. Epilepsy cases consisted of individuals from each of the super populations, while control samples were mostly of European descent.
(B) Genetic ancesty of the epi25 cases and controls were predicted by the Random Forest classifier based on top 6 PCs, using 1KG samples as the training data. Individuals were assgined to a particular 1KG-ancestry with a predicted probability >0.9, as depicted in the figure. We removed samples with unknown or unclassified ancestry and focused on European samples—the largest population under study with abundant cases and controls—moving forward.

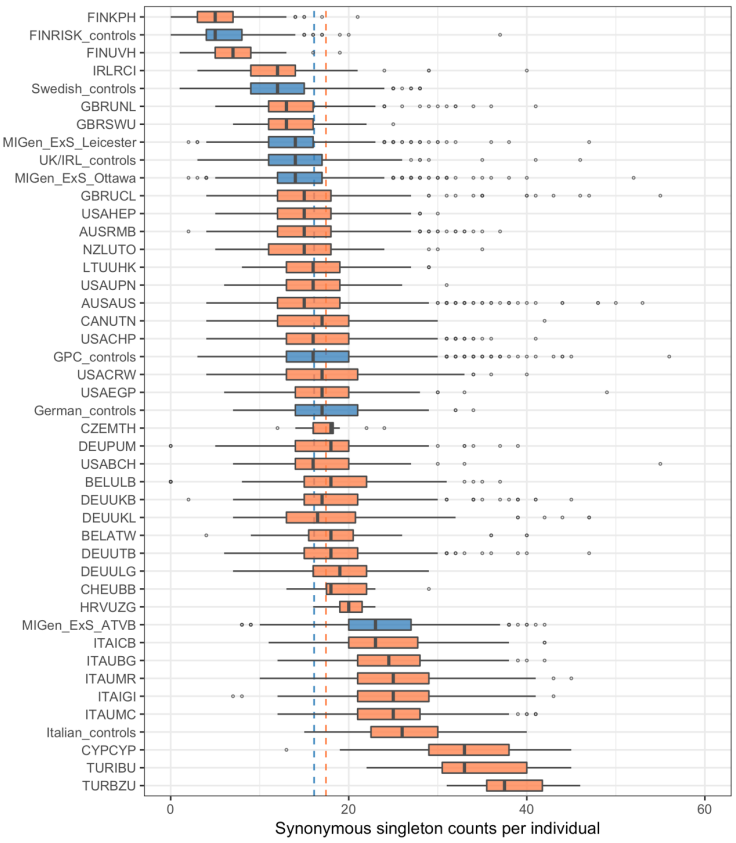
(A)


0.15
0.10
0.05
0.00
-0.05
-0.10
PC2
-0.1
0.0
0.1
PC1

(B)


0.0
-0.1
PC2
-0.1
0.0
0.1
PC1
Case/Control
Control
Epilepsy

(C)


FINKPH
FINRISK_controls
FINUVH
IRLRCI
Swedish_controls
GBRUNL
GBRSWU
MIGen_ExS_Leicester
UK/IRL_controls
MIGen_ExS_Ottawa
GBRUCL
USAHEP
AUSRMB
NZLUTO
LTUUHK
USAUPN
AUSAUS
CANUTN
USACHP
GPC_controls
USACRW
USAEGP
German_controls
CZEMTH
DEUPUM
USABCH
BELULB
DEUUKB
DEUUKL
BELATW
DEUUTB
DEUULG
CHEUBB
HRVUZG
MIGen_ExS_ATVB
ITAICB
ITAUBG
ITAUMR
ITAIGI
ITAUMC
Italian_controls
CYPCYP
TURIBU
TURBZU
0
20
40
60
Synonymous singleton counts per individual

(D)

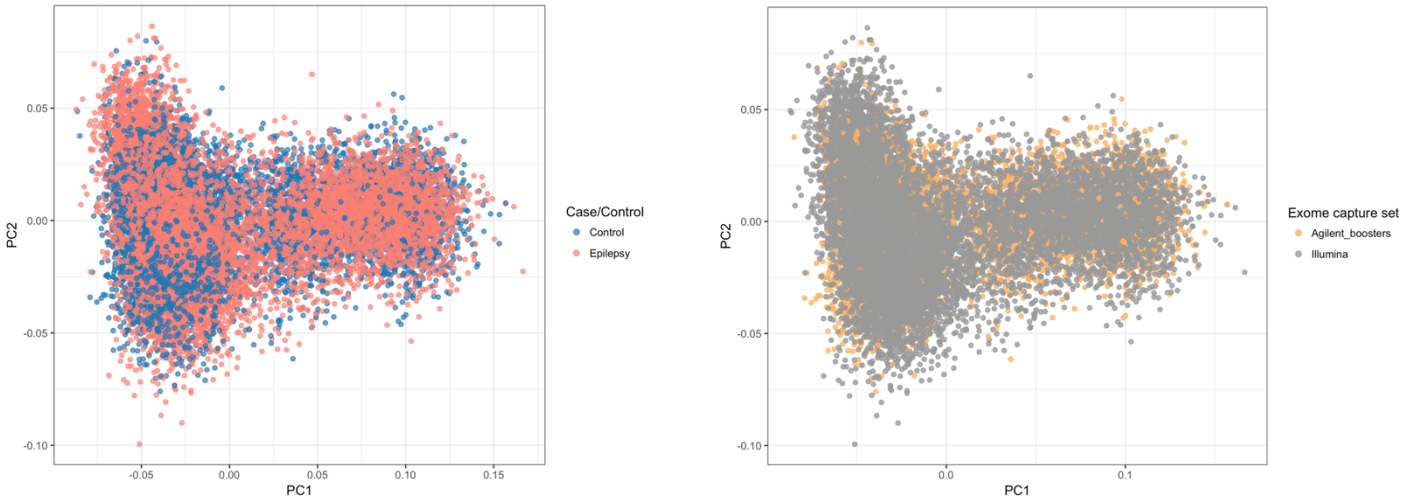


Figure S4. PCA on European samples

(A) Initial PCA on individuals classified as having a European descent. Ancestry was not well-matched between cases and controls.
(B) PCA after removing controls not pair-matched with cases based on top 2 PCs. Discarded controls were mostly Swedish and Finnish due to their unique genetic architecture.
(C) Synonymous singleton count distribution of ancestry-matched samples. This number tracks with intra-continental ancestry, showing a larger count in Southern Europeans relative to Northern Europeans. Case cohorts (orange) and control cohorts (blue) overall had a similar rate of singleton synonymous variants, indicated by the group averages (dashed lines). A few cohorts (Finland, Cyprus, Turkey) showed deviation from the mean and were removed from further analysis.
(D) Post-QC PC plots (PC1 vs. PC2), by case-control status and exome capture platform, showing a genetically homogeneous population for which analyses would be performed

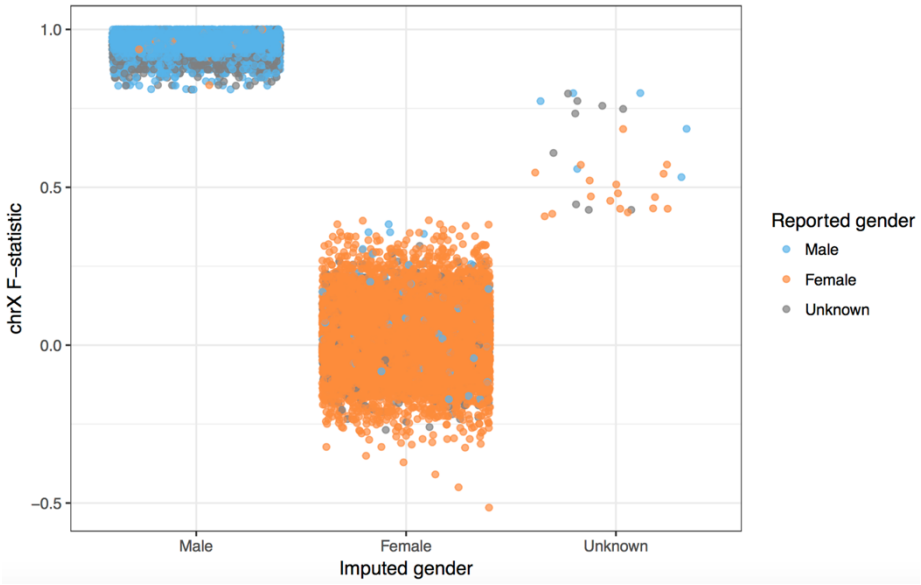


Figure S5. Sex check: concordance between genetically-imputed sex and self-reported sex

Sex was determined using the X-chromosome homozygosity rate, or F-statistics (≤0.4: female; ≥0.8: male; in between: unknown). Samples with either undefined genetic sex or discordant between imputed and reported gender were discarded.

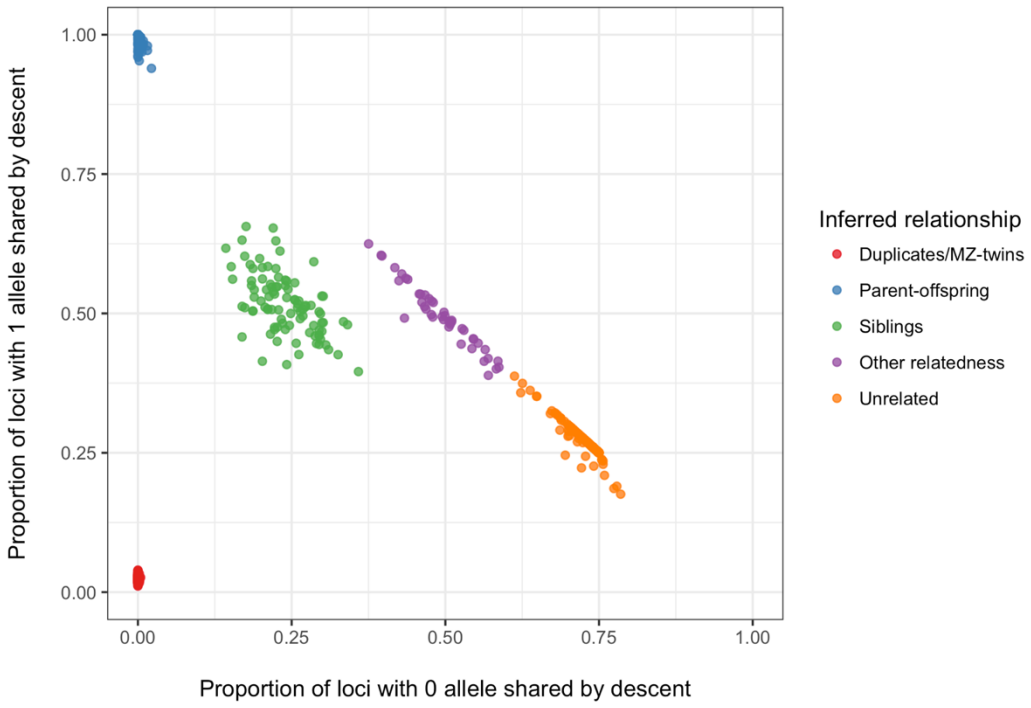


Figure S6. Identify-by-descent (IBD) estimate for ancestry-matched (EUR) samples

This plot shows only pairs of individuals with a relatedness > 0.125 for clarity. One individual in each of the related pairs with IBD > 0.2 was removed. Samples appearing in multiple related pairs were removed first, and cases were preserved over controls.

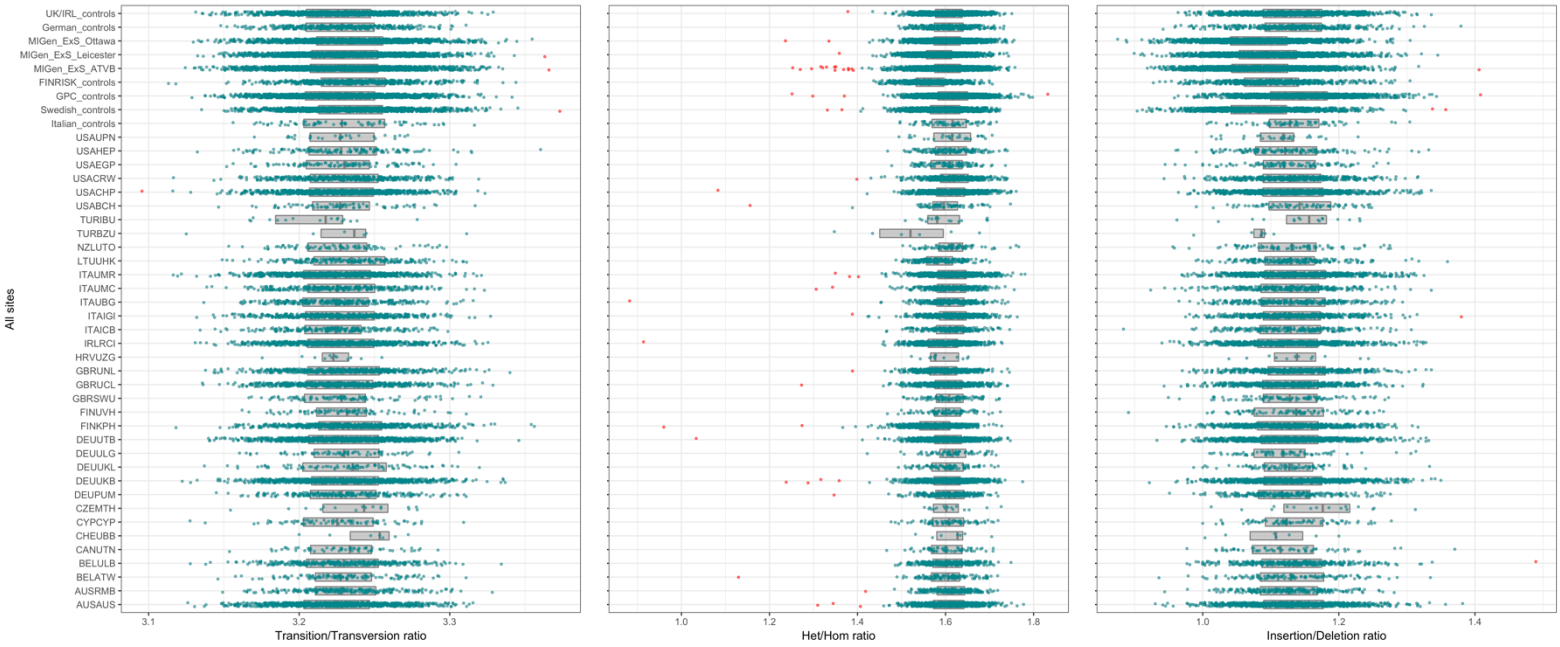
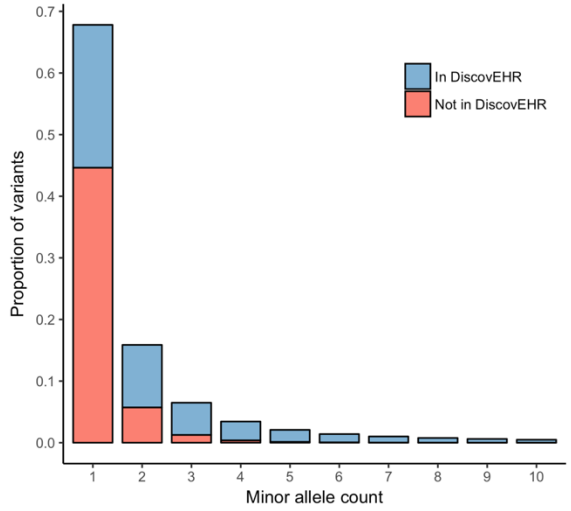


Figure S7. Sample outliers of QC metrics: transition to transversion ratio (Ti/Tv), heterozygous to homozygous ratio (Het/Hom), and insertion to deletion ratio.

Samples in each cohort that had a value outside the range of mean±4SD were removed (red dots).

(A)



(B)

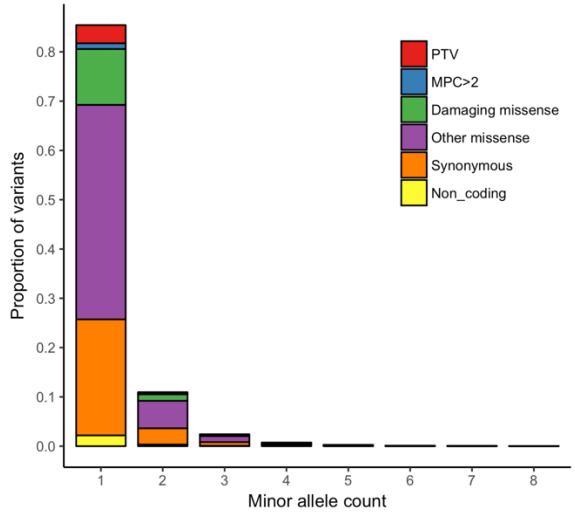


Figure S8. Post-QC allele count distribution

(A) showed the proportion of variants absent in DiscovEHR (48%) for which a majority were singletons
(B) showed the proportion of *non-DiscovEHR* sites by functional annotation. For the purpose of visualization, plotted here were non-overlapping proportions from the six functional classes of variants. Specifically, 4.1% were PTVs, 1.3% were missense variants with MPC≥2, 13% were non-MPC damaging missense, 50.1% were non-MPC benign missense variants (predicted by Polyphen-2 and SIFT), 27.9% were synonymous variants, and 2.8% were non-coding variants. In the analyses, however, the full damaging and benign missense annotations (predicted by PolyPhen-2 and SIFT) contained 59% and 41% of the MPC≥2 variants, respectively.

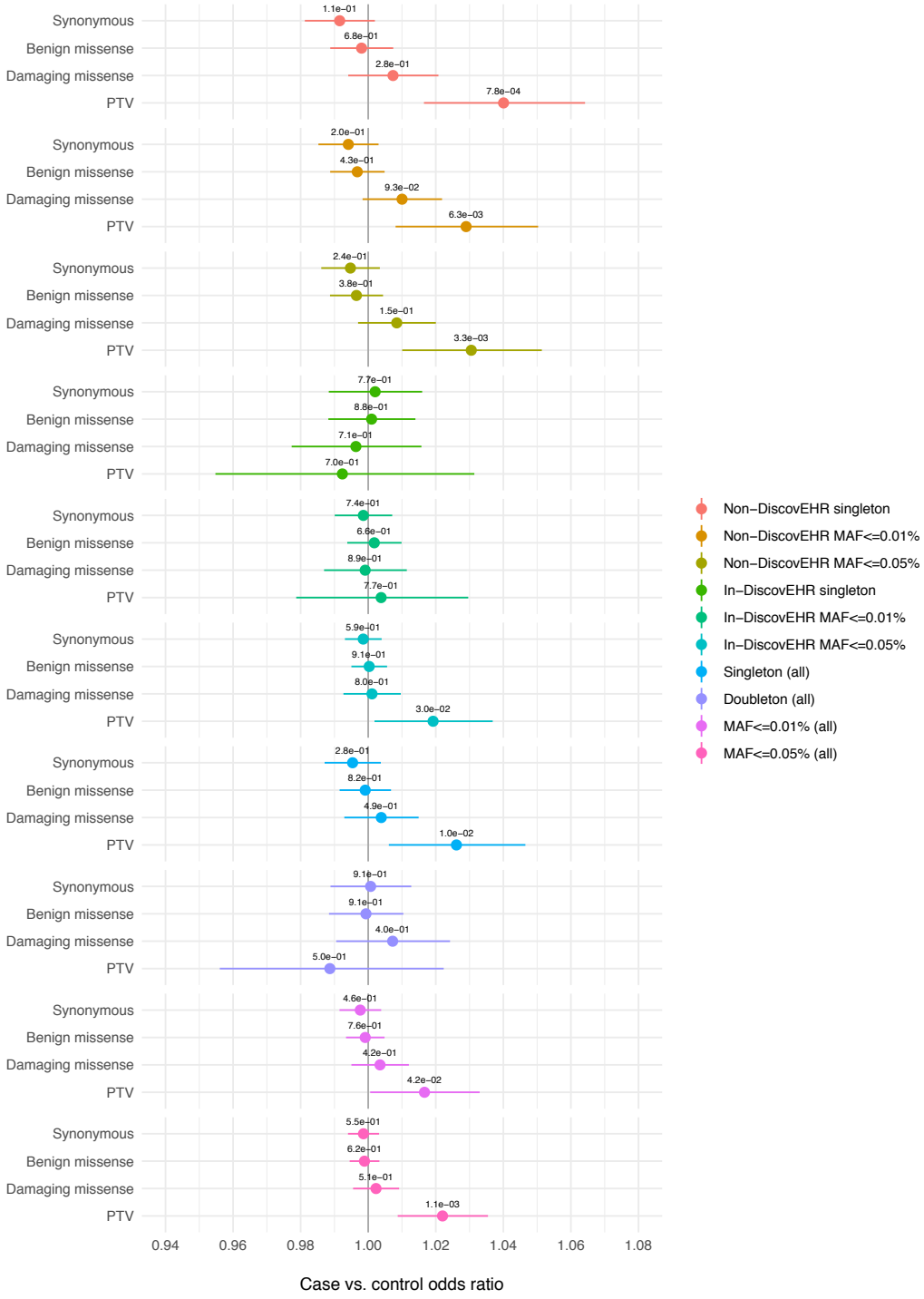


Figure S9. Exome-wide burden of rare variants in epilepsy versus controls across allele frequency categories

The burden of deleterious variants increased with a decrease in allele frequency. Risk of epilepsy was predominantly enriched among the singleton variants absent from the DiscovEHR cohort (a population allele frequency database).

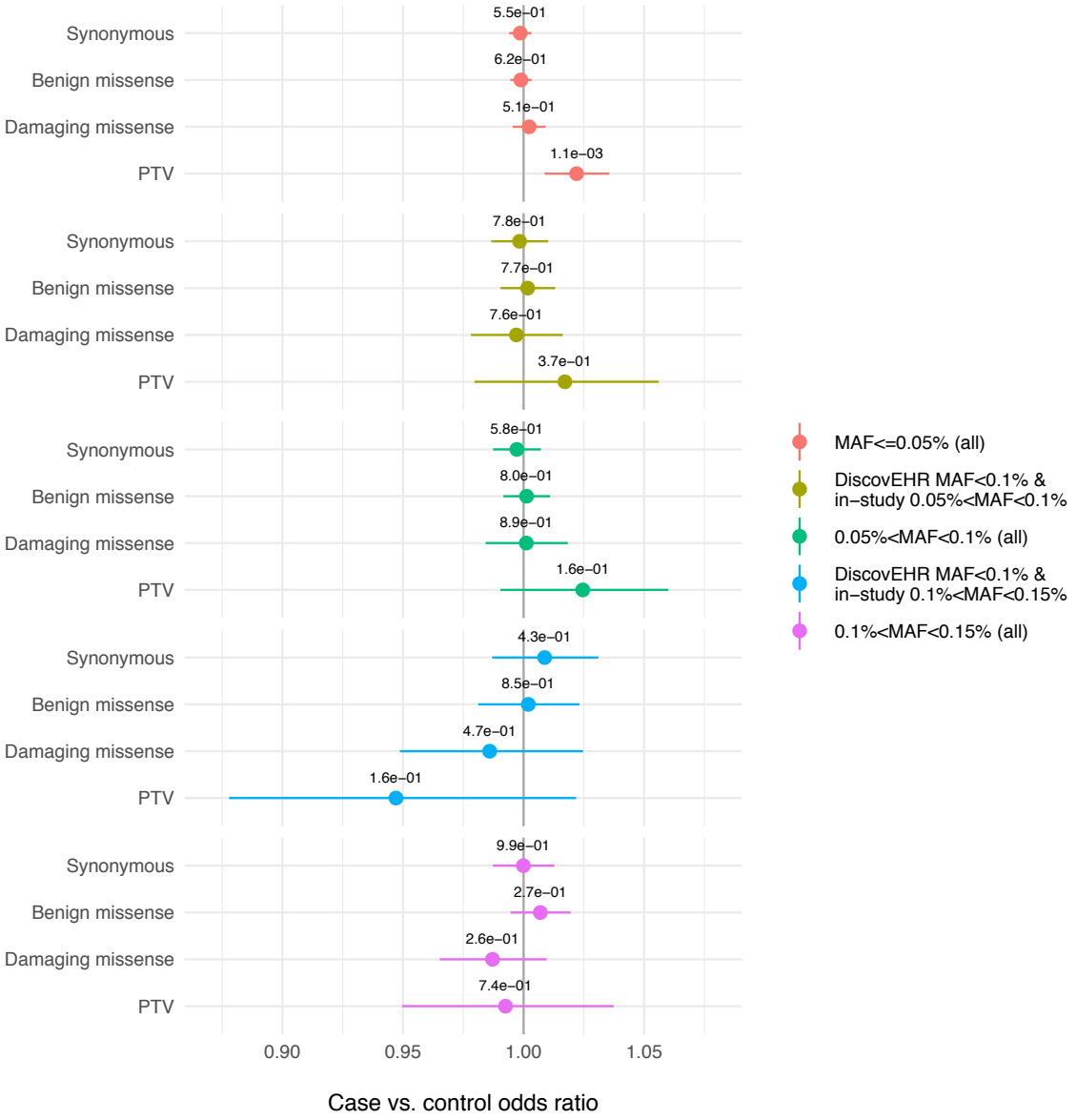


Figure S10. Exome-wide burden of rare and low-frequency variants in epilepsy versus controls

The excess burden of deleterious variants diminished in epilepsy with an increased in allele frequency, and are contained within ultra-rare variants (**Figure S9**).

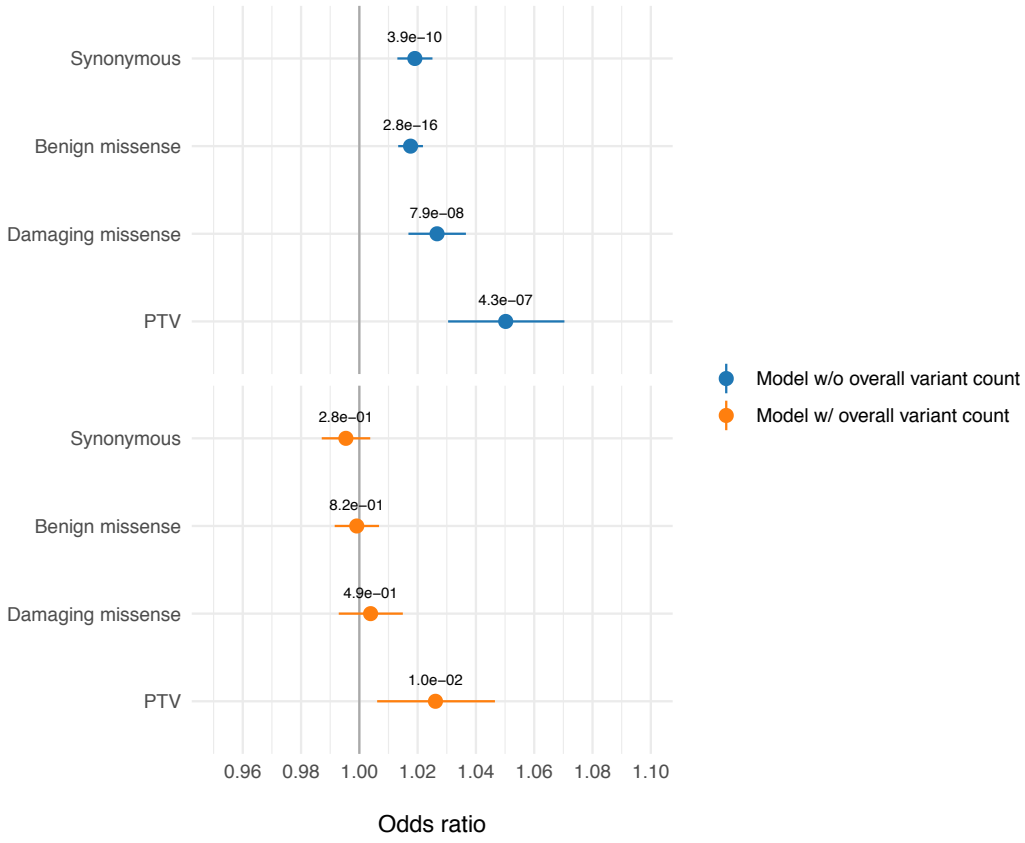


Figure S11. Correcting for exome-wide burden in gene-set burden test: controlling for overall variant count calibrated burden test p-values

This example was a burden analysis that assessed the exome-wide singleton enrichment in all epilepsy cases versus controls. Logistic regression model adjusting for total singleton count regardless of annotation (orange) successfully reduced the inflated burden across all annotation categories due to residual population stratification not captured by PCA.

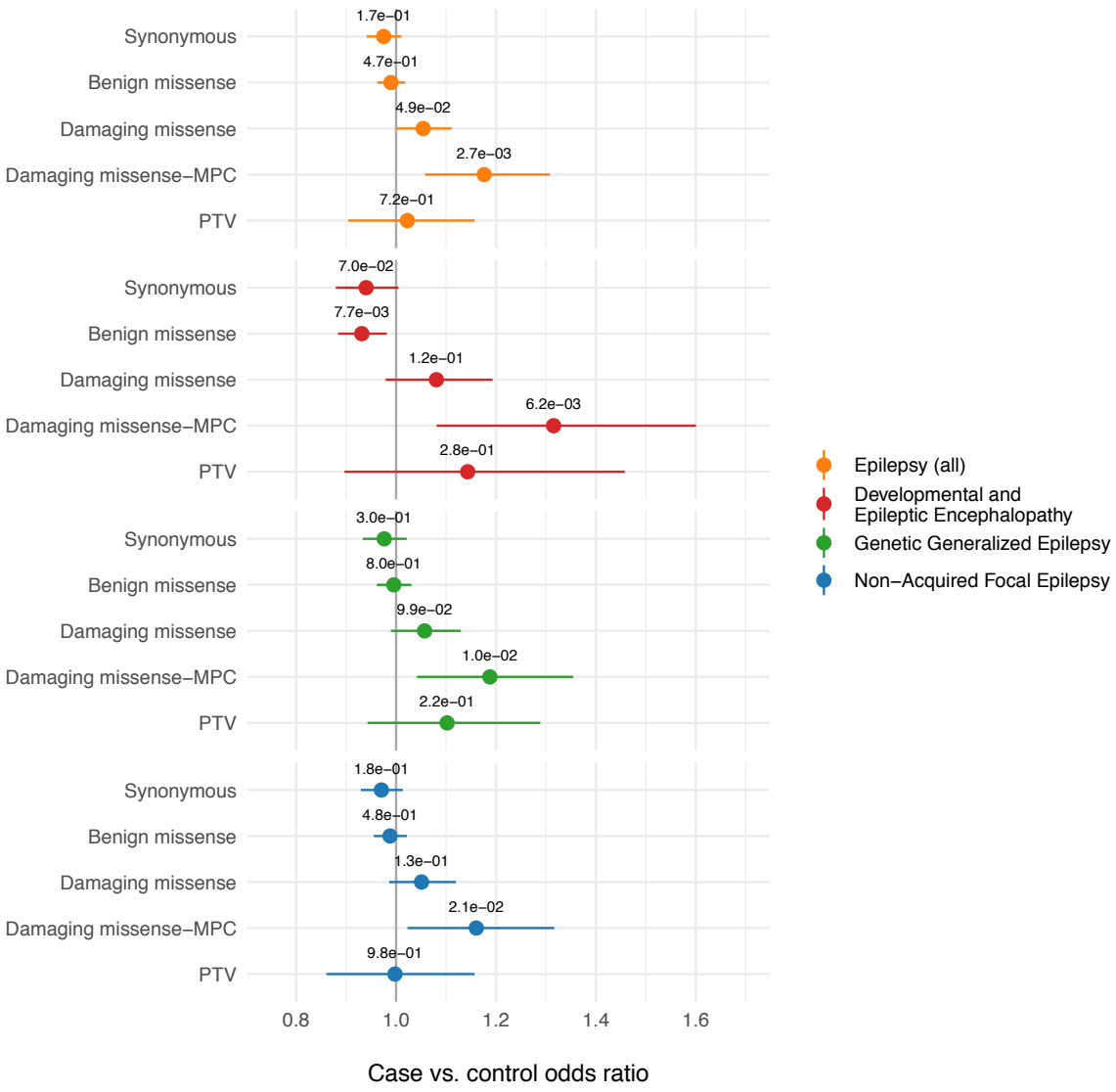


Figure S12. Burden of ultra-rare singletons in LoF-intolerant genes with 0.9 < pLI < 0.995), using only non-ExAC controls (N=4,042)

Burden signals of PTVs disappeared after restricting to genes within the range of 0.9 < pLI < 0.995, suggesting that most enrichment of PTVs in haploinsufficient genes associated with epilepsy was driven by genes with a pLI > 0.995.

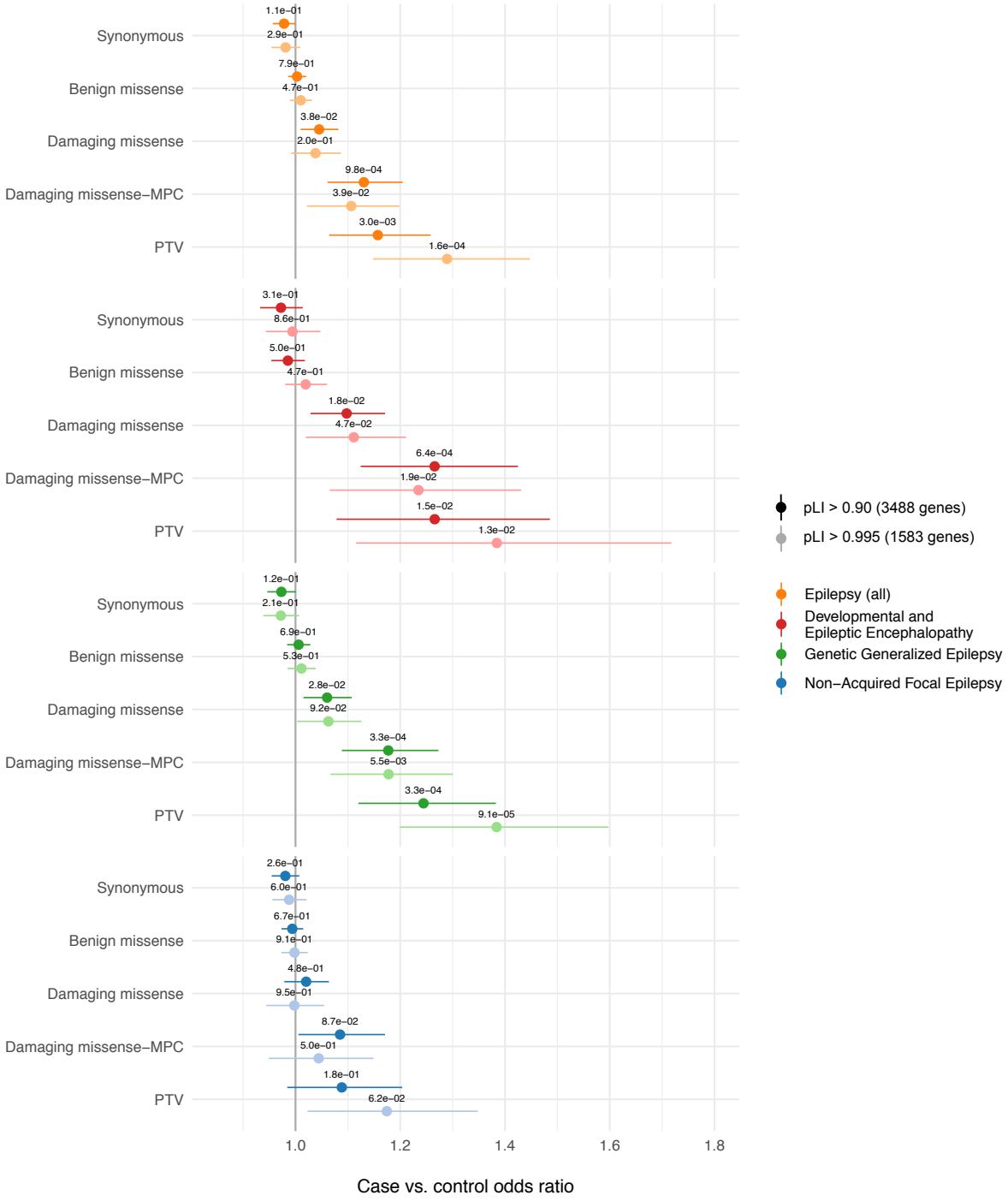


Figure S13. Burden of ultra-rare singletons in LoF-intolerant genes (pLI > 0.9 and > 0.995), using only non-ExAC controls (N=4,042)

Significance of association was displayed in FDR-adjusted p-values. A strongly significant enrichment of PTVs and damaging missense variants (MPC≥2) were seen in patients with DEE and GGE, but not NAFE (FDR < 0.05). PTV burden was even stronger when restricting to genes with pLI>0.995. Only non-ExAC controls were used as pLI scores were derived from ExAC samples.

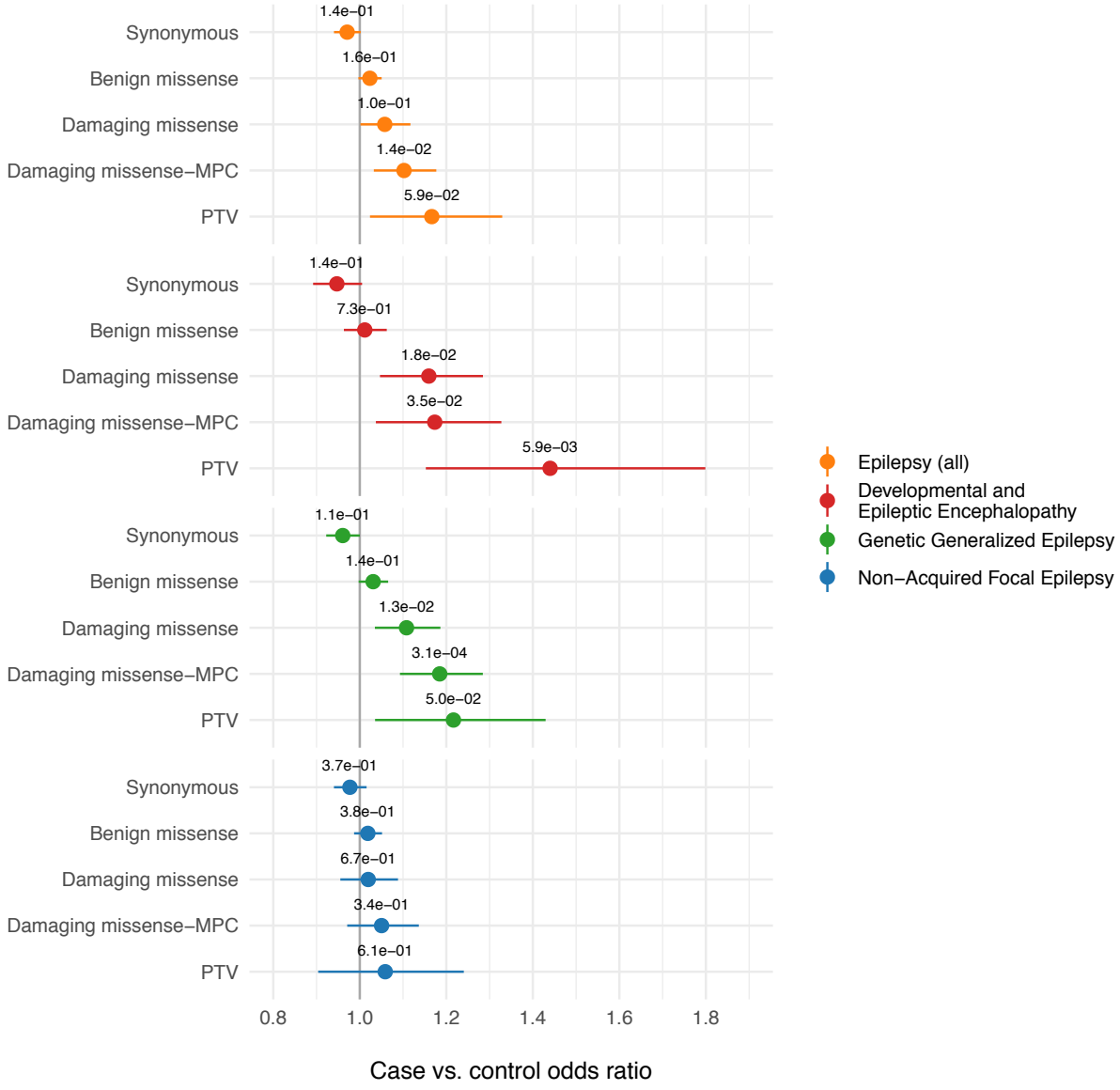


Figure S14. Burden of ultra-rare singletons in missense constrained genes (mi-Z > 3.09), using only non-ExAC controls (N=4,042)

Significance of association was displayed in FDR-adjusted p-values. PTV and damaging missense (MPC≥2) burden were significant for all epilepsy phenotypes at FDR < 0.05 except NAFE.

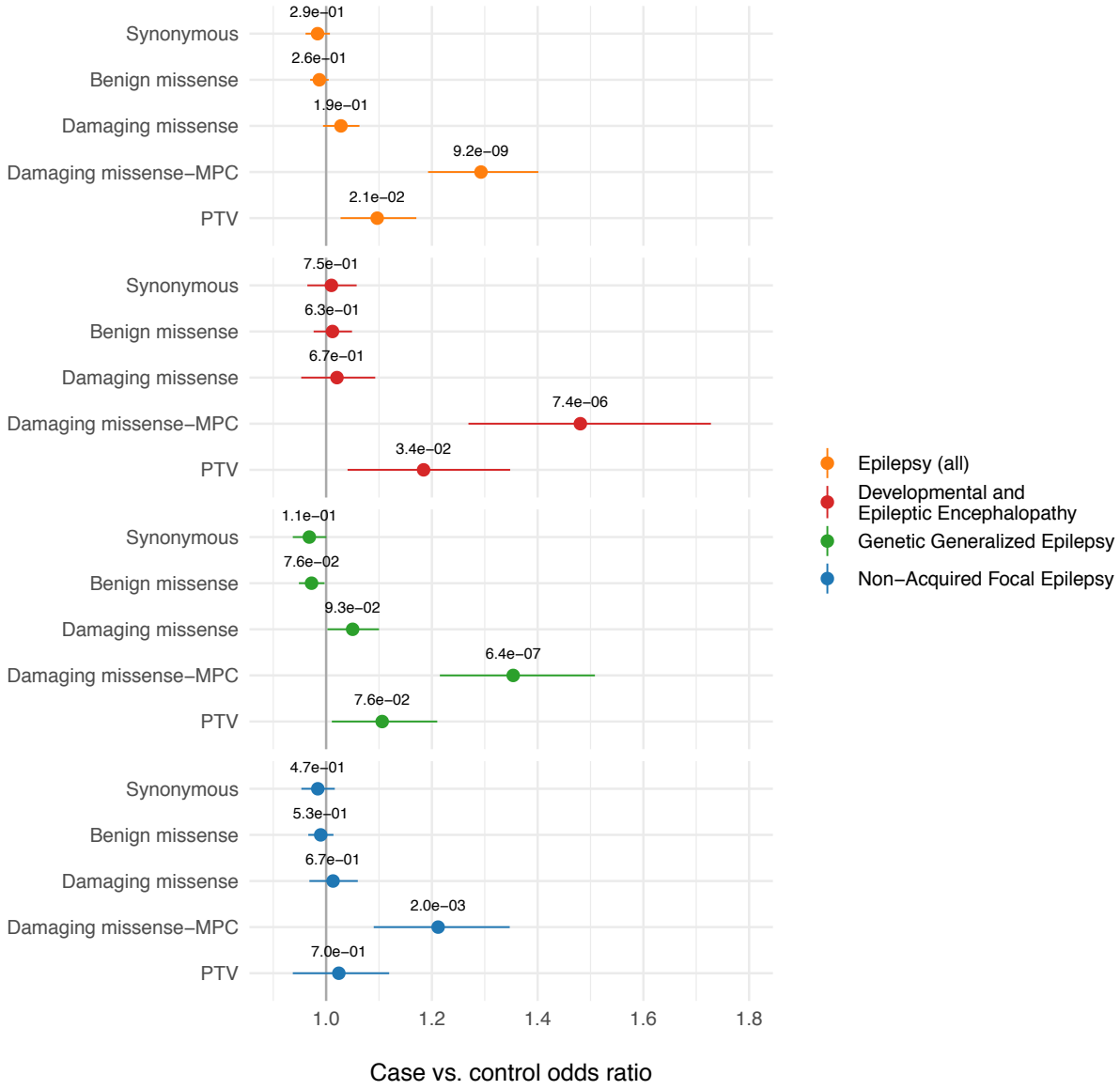
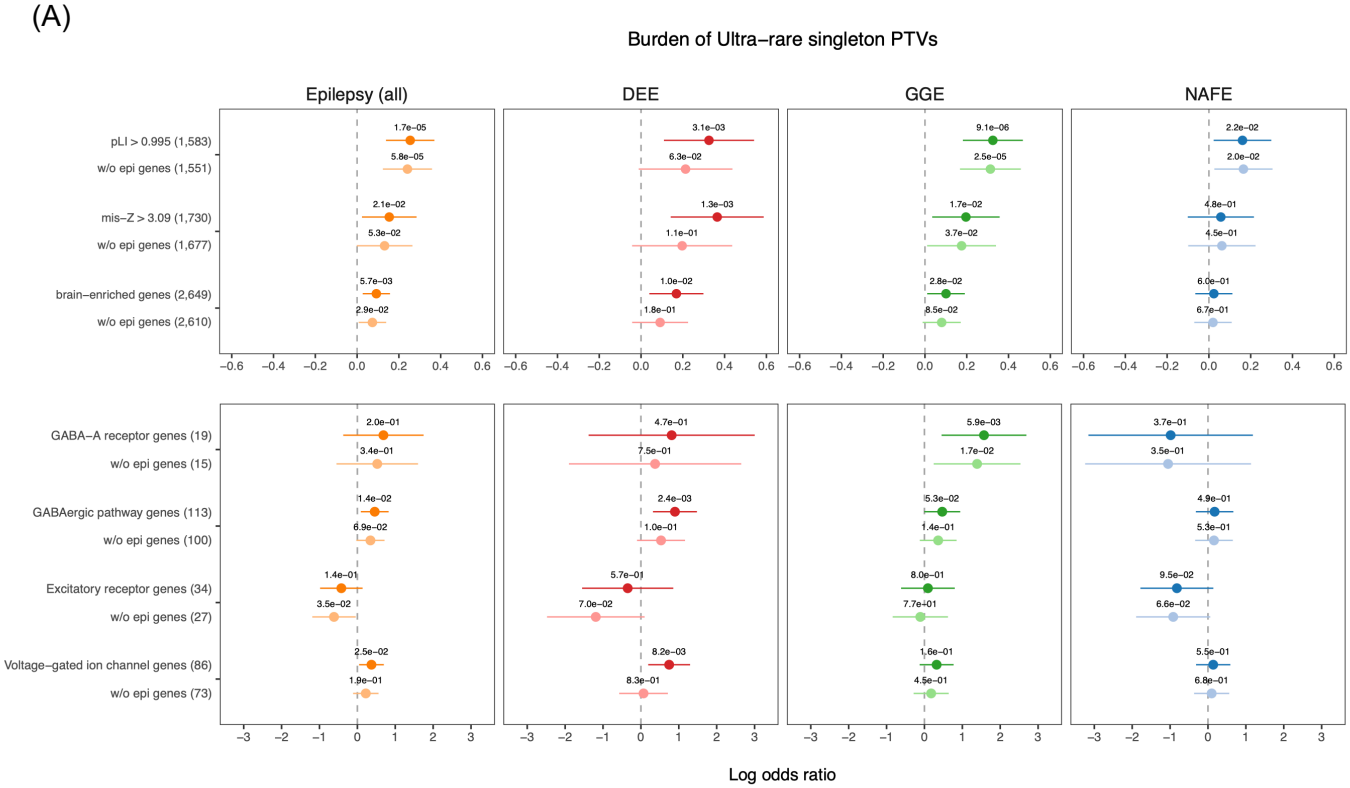


Figure S15. Burden of ultra-rare singletons in 2,649 brain-enriched genes

Significance of association was displayed in FDR-adjusted *p*-values. Damaging missense (MPC≥2) variants in brain-enriched genes were significantly enriched in all epilepsy types and had a much higher burden relative to PTVs. PTV burden was stronger in patients with DEE (FDR < 0.05) than with common epilepsy types.

(A)



(B)

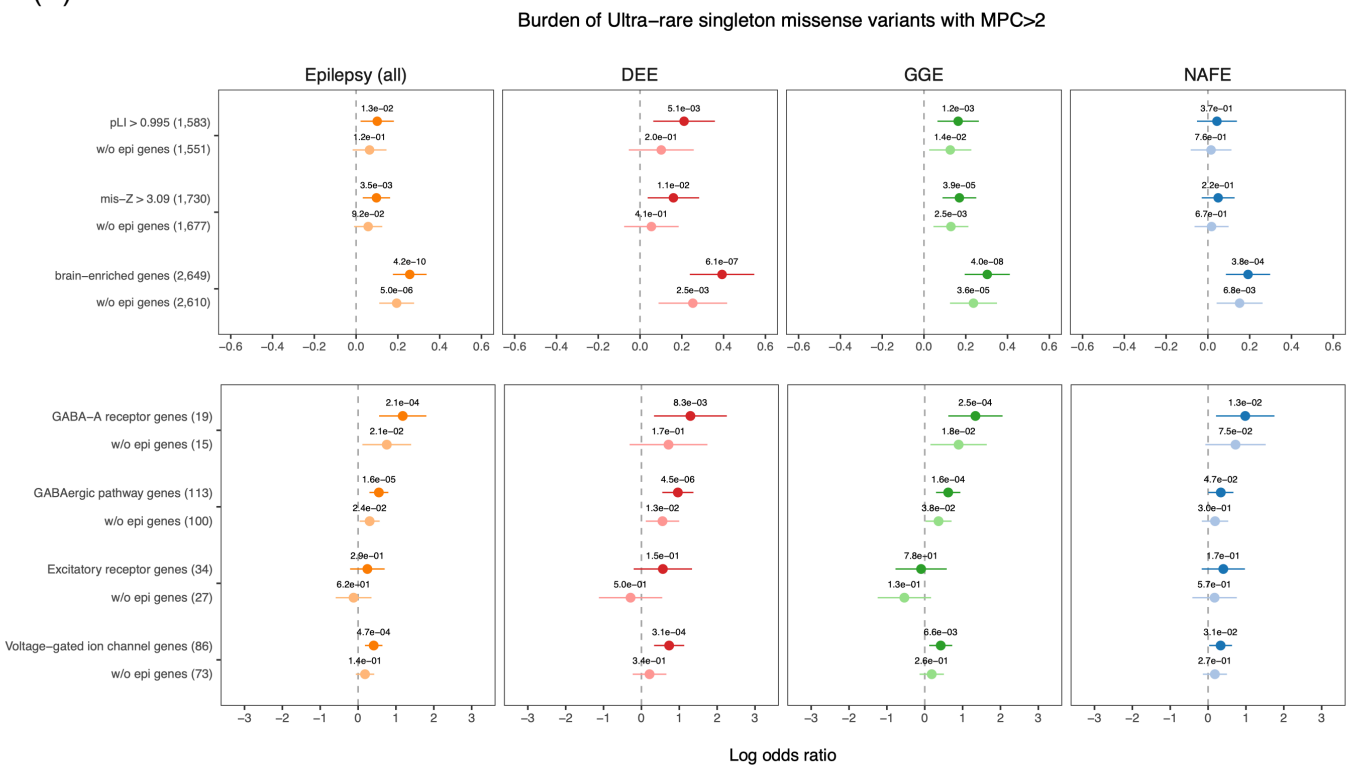


Figure S16. Burden of ultra-rare singleton PTVs (A) and missese-MPC≥2 variants (B) before and after correcting for events in genes previously associated with epilepsy

Genes previously associated with epilepsy included 74 non-overlapping genes from the 43 dominant epilepsy genes, 50 dominant DEE genes, and 33 NDD-EPI genes (**Table S6**). For each gene set, the darker (upper) color showed burden signals before adjusting for counts in the 74 genes and the lighter (lower) color showed the residual signals after correcting for associations driven by the 74 genes. P-values were shown in original values.

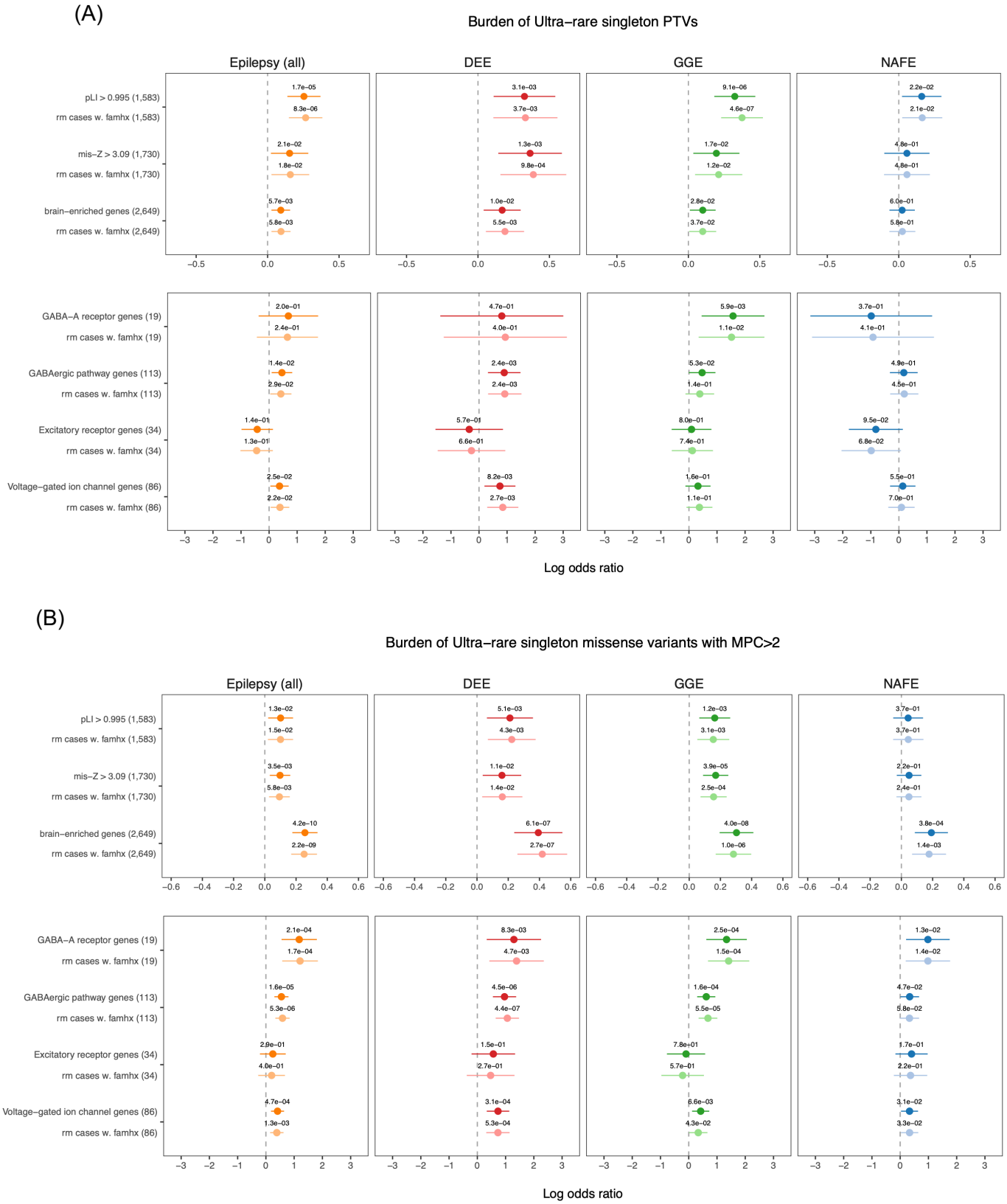


Figure S17. Burden of ultra-rare singleton PTVs (A) and missese-MPC≥2 variants (B) before and after removing patients with a reported family history (of an affected first-degree relative)

The number of individuals reported to have an affected first-degree relative include 96 DEE, 380 GGE, and 197 NAFE cases. Showing on the figure is the estimated burden before (darker color) and after (lighter (lower) removing these individuals for each gene set. P-values were shown in original values.

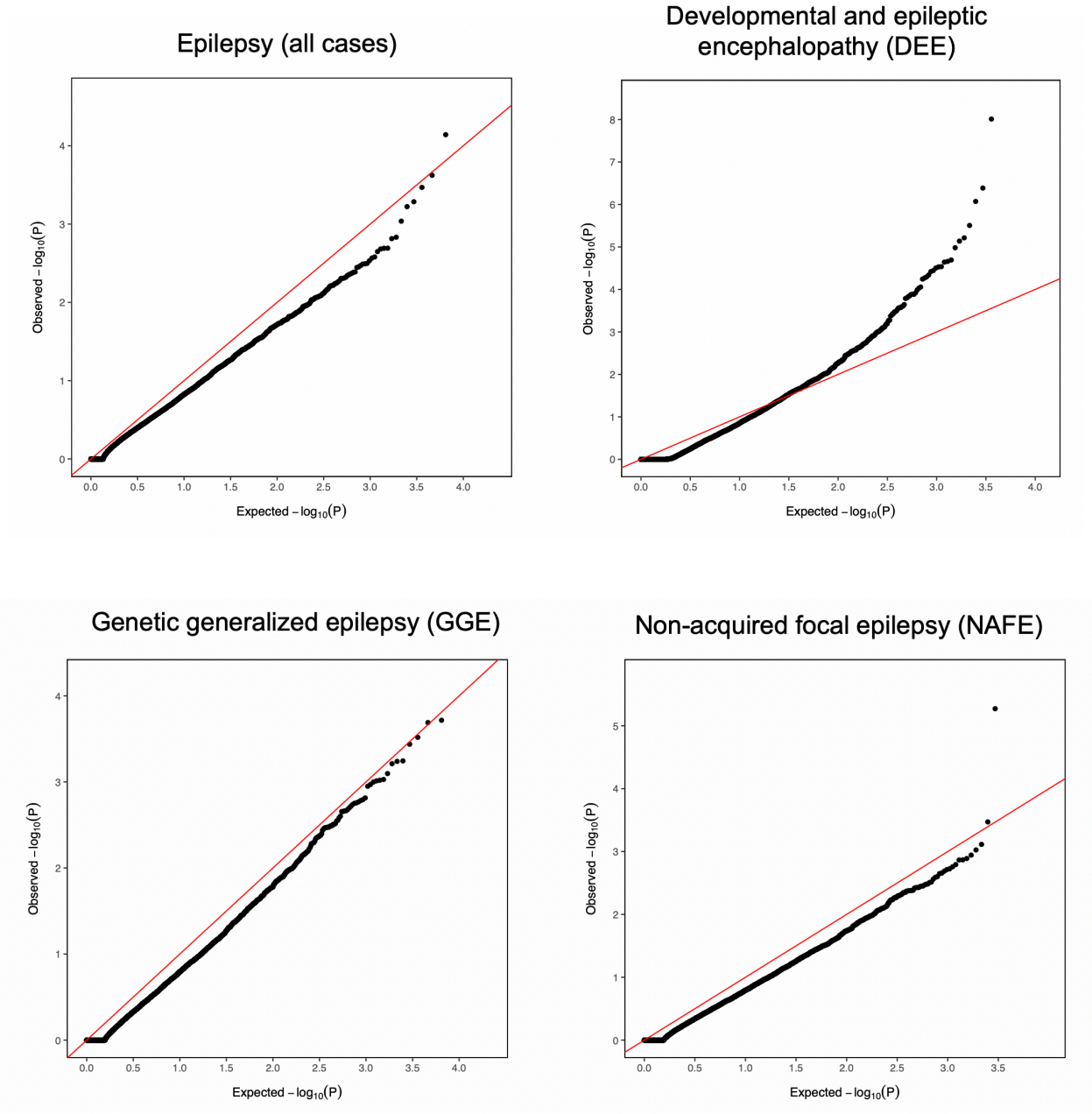


Figure S17. Burden of all low-frequency damaging missense variants assessed using SKAT

Shown were the SKAT gene burden results based on damaging missense variants (predicted by PolyPhen-2 and SIFT) with MAF < 0.01. Exome-wide significant gene associations were observed only among DEE patients compared to controls, where the top genes included previously associated genes (e.g. *STXBP1*, P=9.3e-09; *KCNA2*, P=1.0e-05). Burden of missense variants with MPC≥2 showed a very similar pattern but was in general more underpowered due to low counts (data not shown).

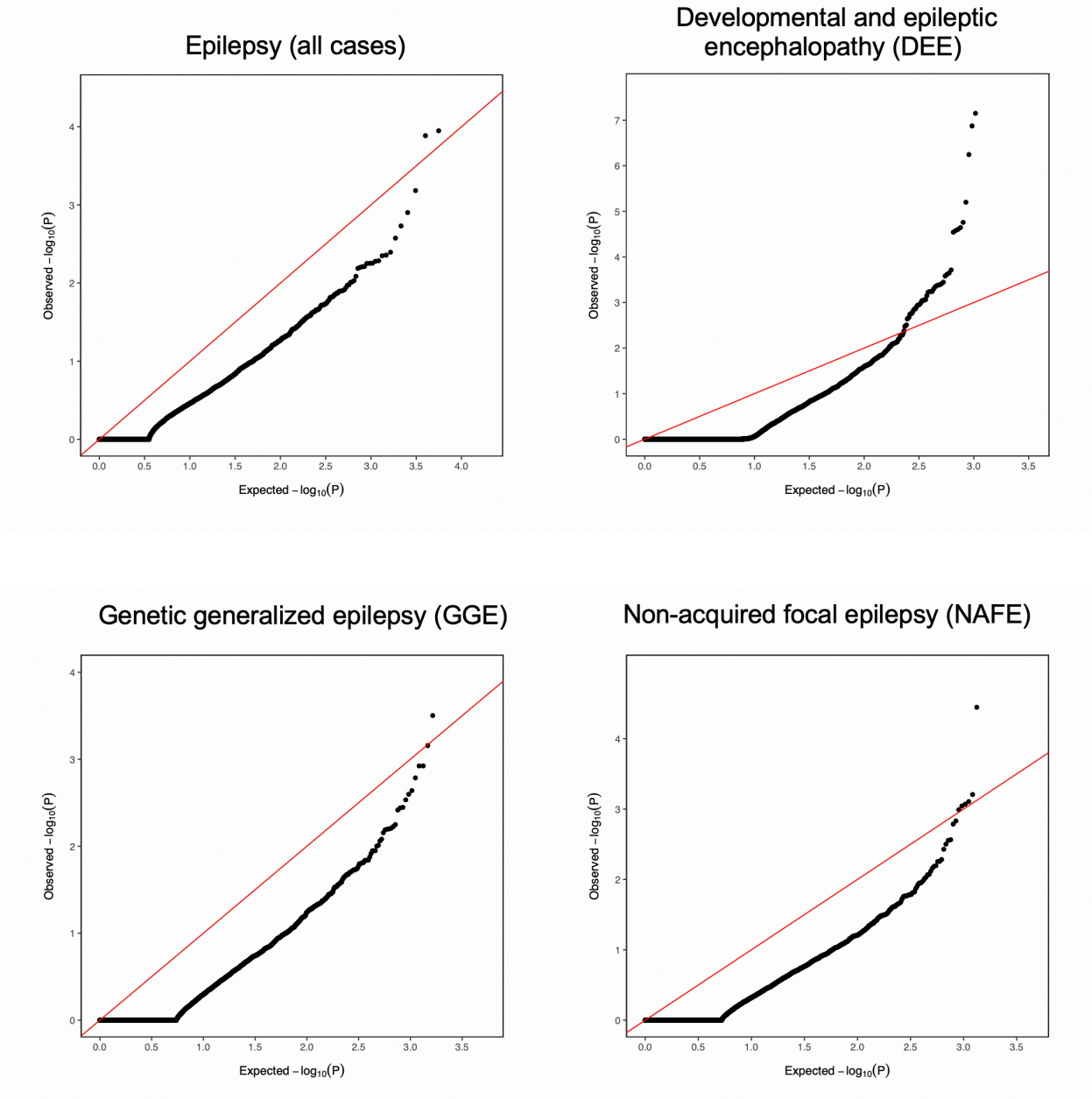


Figure S18. Burden of all low-frequency protein truncating variants assessed using SKAT

Shown were the SKAT gene burden results based on protein truncating variants with MAF < 0.01. Exome-wide significant gene associations were observed only among DEE patients compared to controls, with the top genes including some previously associated genes (e.g. *KIAA2022, P* = 7.1e-08; *SCN1A*, *P* = 3.9e-04). Associations with other epilepsy phenotypes were largely underpowered due to low counts.

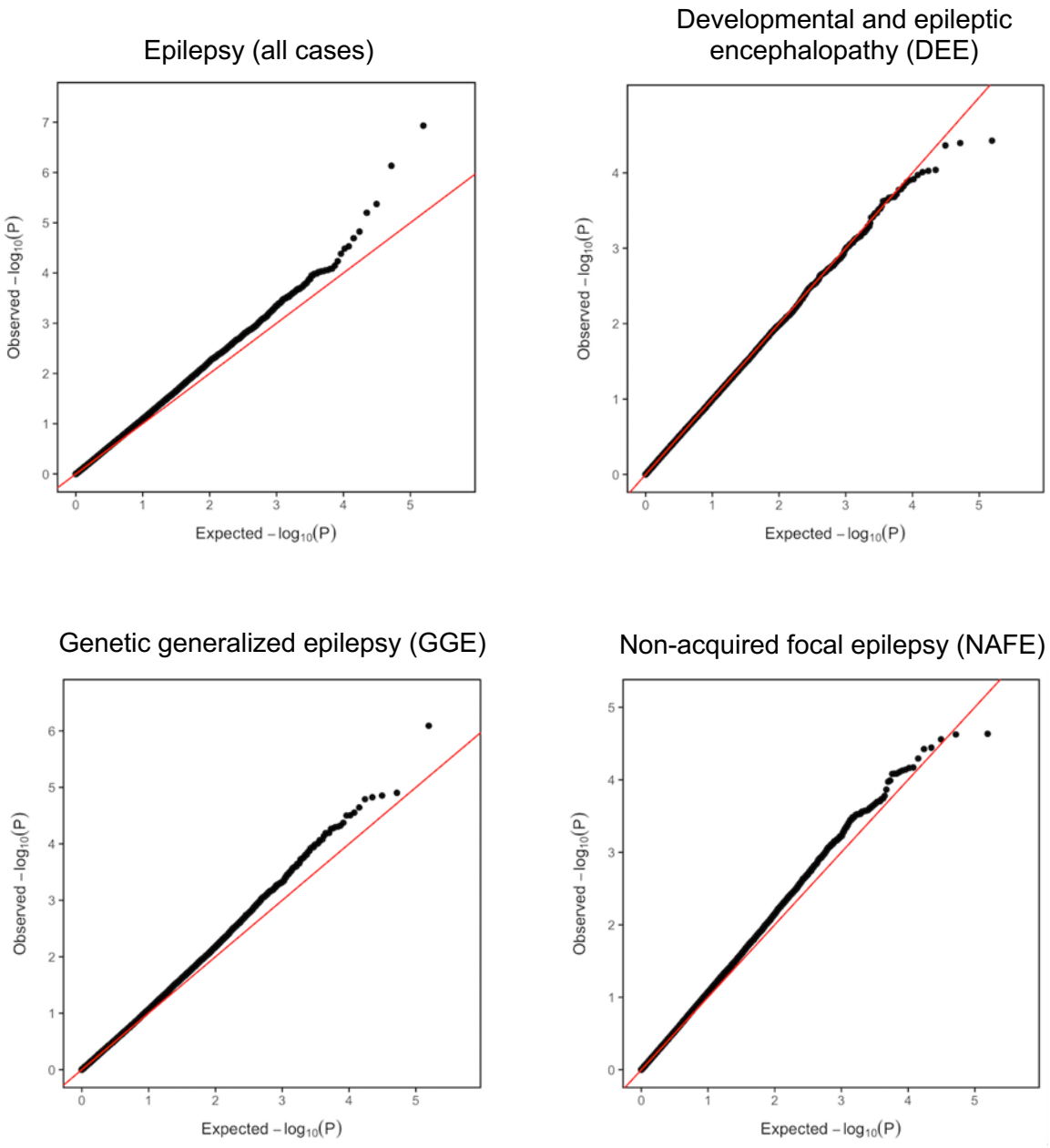
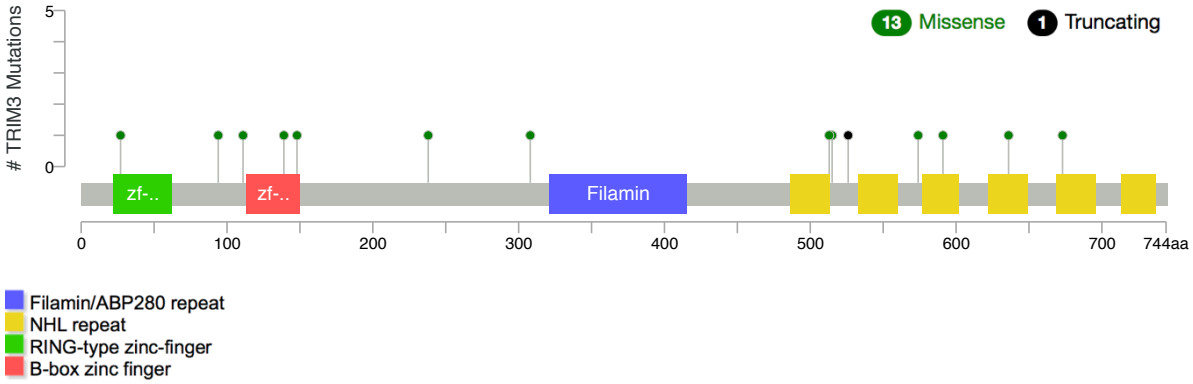


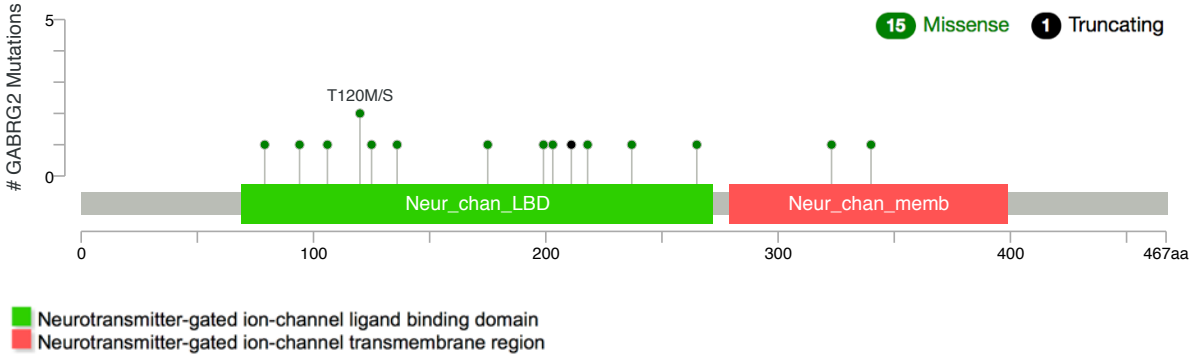
Figure S19. Association with common and low frequency variants (for variants with MAF > 0.001)

After multiple-testing correction, no coding variants were significantly associated with overall epilepsy (all cases) or within any epilepsy type.

(A) ***TRIM3*** (2 DEE, 2 GGE, 7 NAFE, 3 Lesional)



(B) ***GABRG2*** (7 GGE, 7 NAFE, 1 Lesional, 1 Control)



(C) ***CACNA1G*** (2 DEE, 10 GGE, 4 NAFE, 3 Lesional, 3 Control)

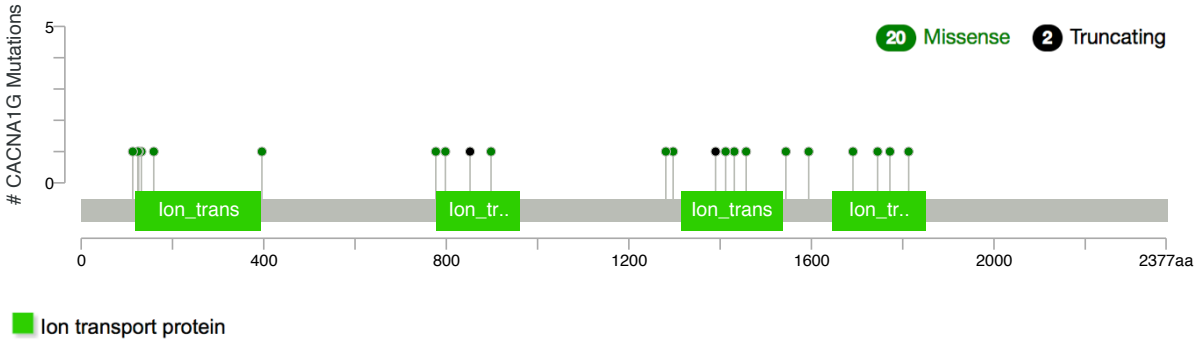


Figure S20. Mutational diagrams of selected top genes in all-epilepsy gene burden analysis

These genes were shown to harbor a higher burden of missense (MPC≥2) than protein-truncating URVs in epilepsy patients.

Table S1. Epi25 epilepsy patients for whole-exome sequencing (WES)

DEE: developmental and epileptic encephalopathy; GGE: genetic generalized epilepsy; NAFE: non-acquired focal epilepsy; LFE: lesional focal epilepsy; FS: febrile seizure; UE: unclassified epilepsy (including those under review or labeled excluded)

Site Name	Site Code	Epilepsy						TOTAL
		DEE	GGE	NAFE	LFE	FS	UE	
Australia: Melbourne	**AUSAUS**	139	412	366	236	57	6	1216
Australia: Royal Melbourne	**AUSRMB**	0	88	128	42	0	12	270
Belgium: Antwerp	**BELATW**	87	36	20	0	0	0	143
Belgium: Brussels	**BELULB**	6	71	190	120	0	0	387
Canada: Andrade	**CANUTN**	37	43	8	3	0	5	96
Switzerland: Bern	**CHEUBB**	0	0	6	2	0	0	8
Cyprus	**CYPCYP**	7	53	52	4	0	2	118
Czech Republic: Prague	**CZEMTH**	16	0	0	0	0	0	16
Germany: Frankfurt/Marburg	**DEUPUM**	4	58	130	60	0	7	259
Germany: Bonn	**DEUUKB**	0	206	557	61	193	36	1053
Germany: Kiel	**DEUUKL**	47	81	20	0	0	1	149
Germany: Leipzig	**DEUULG**	0	0	89	0	0	0	89
Germany: Tuebingen	**DEUUTB**	63	346	256	351	33	6	1055
Finland: Kuopio	**FINKPH**	20	55	636	21	0	2	734
Finland: Helsinki	**FINUVH**	26	51	17	1	0	0	95
Wales: Swansea	**GBRSWU**	0	45	56	0	0	6	107
UK: UCL	**GBRUCL**	5	304	314	2	0	87	712
UK: Imperial/Liverpool	**GBRUNL**	0	190	343	0	0	0	533
Hong Kong	**HKGHKK**	0	21	23	0	0	0	44
Croatia	**HRVUZG**	11	2	0	0	0	1	14
Ireland: Dublin	**IRLRCI**	14	145	464	155	15	18	811
Italy: Milan	**ITAICB**	50	97	14	29	0	6	196
Italy: Genova	**ITAIGI**	100	269	13	26	0	2	410
Italy: Bologna	**ITAUBG**	82	60	117	34	0	6	299
Italy: Catanzaro	**ITAUMC**	5	72	264	28	0	4	373
Italy: Florence	**ITAUMR**	402	186	128	155	0	56	927
Japan: RIKEN Institute	**JPNRKI**	0	31	0	0	0	0	31
Lithuania	**LTUUHK**	33	96	70	14	0	3	216
New Zealand: Otago	**NZLUTO**	27	42	42	6	0	0	117
Turkey: Bogazici	**TURBZU**	89	0	0	0	0	0	89
Turkey: Istanbul	**TURIBU**	5	50	52	6	0	1	114
USA: BCH	**USABCH**	60	12	9	7	1	5	94
USA: Philadelphia/CHOP	**USACHP**	0	988	479	0	0	0	1467
USA: Philadelphia/Rowan	**USACRW**	0	324	236	0	0	0	560
USA: EPGP	**USAEGP**	126	2	1	1	0	0	130
USA: NYU HEP	**USAHEP**	0	0	214	0	0	0	214
USA: Penn/CHOP	**USAUPN**	15	17	17	1	0	0	50
TOTAL		**1476**	**4453**	**5331**	**1365**	**299**	**272**	**13196**

Table S2. Control collection for Epi25 WES analysis

Control cohorts	General control	MI/CAD/CHD[1]	Exome capture platform	Source	dbGaP Accession or PMID	In ExAC
MIGen ATVB	1802	1875	Agilent	dbGaP	phs000814.v1.p1	Yes
MIGen Leicester	1100	0	Illumina	dbGaP	phs001000.v1.p1	No
MIGen Ottawa Heart Study	987	993	Agilent	dbGaP	phs000806.v1.p1	Yes
Genomic Psychiatry Cohort (GPC) controls	1946	0	Illumina	Broad local	23650244[23]	No
German controls	414	0	Illumina	Broad local	N/A	No
FINRISK controls	681	0	Illumina	Broad local	29165699[22]	No
Swedish SCZ controls	6242	0	Agilent	dbGaP	phs000473.v2.p2	Yes
UK/IRL controls	1223	0	Illumina	Broad local	N/A	No
Italian controls	106	0	Illumina	Broad local	N/A	No
Epi25 Italian controls (ITAICB & ITAIGI)	300	0	Illumina	Epi25	N/A	No
Total	**17669**					

[1]Myocardial infarction, coronary artery disease, or coronary heart disease

Table S3. Sample QC and number of sample dropout at each QC step

Sample QC metric	Number of cases (%)	Number of controls (%)
Initial numbers	**13,196** (100%)	**17,669** (100%)
Initial sample QC		
Call rate < 0.98 / %Chimeras > 0.014 Avg. sequence depth < 30 Avg. genotype quality < 85 Freemix contamination > 0.04	84 (0.63%)	372 (2.11%)
PCA		
Non-European	2034 (15.4%)	338 (1.91%)
Case-Control ancestry matching (using top PCs)	0	6838 (38.7%)
Relatedness filtering		
IBD > 0.2	218 (1.65%)	149 (0.84%)
Sex check		
Ambiguous imputed sex	6 (0.05%)	5 (0.03%)
Imputed sex not equal to reported sex	148 (1.12%)	25 (0.14%)
Outlier removal		
Per-cohort outliers (>4 SD) of Ti/Tv, Het/Hom, or indel ratio	34 (0.19%)	34 (0.19%)
Control for residual population stratification (using #synonymous singletons)		
Swedish/Finnish: lower counts	1232 (9.33%)	1472 (8.33%)
Cypriot/Turkish: higher counts	99 (0.75%)	0
TBD-phenotype cases / Excluded by review	171 (1.30%)	0
Final numbers	**9,170** (69.5%)	**8,436** (47.7%)

Table S4. Variant QC and number of variant dropout at each QC step

Variant QC metric	Number of variants removed	%Removed
Pre-filtering:		
Failing VQSR	614,459	8.90%
In low complexity region	69,003	1.00%
Outside well-covered regions	2,996,409	43.40%
Initial QC:		
Call rate < 0.975	220,614	3.20%
Case- & Ctrl-call rate < 0.975		
Case/Ctrl call rate diff > 0.005	224,033	3.20%
pHWE < 1e-06	662	0.01%
AC = 0	732,640	10.60%
Final QC:		
Call rate < 0.98	195,586	2.80%
Case/Ctrl call rate diff > 0.005		
pHWE < 1e-06		
AC = 0		
Associated with capture platform ($p<0.05$)	5,906	0.09%

Table S5. Grouping of functional consequences of the called sites

Consequence	**VEP Terms/annotation**
Protein-truncating variant (PTV)	"transcript_ablation", "splice_acceptor_variant", "splice_donor_variant", "stop_gained", or "frameshift_variant"
Missense	"stop_lost", "start_lost", "transcript_amplification", "inframe_insertion", "inframe_deletion", "missense_variant", "protein_altering_variant", or "splice_region_variant"
Benign/Other missense	Not classified by PolyPhen-2 and SIFT as damaging or deleterious
Damaging missense	PolyPhen-2 "probably_damaging" & SIFT "deleterious"
Damaging missense-MPC	MPC score (**M**issense badness, **P**olyPhen-2, and **C**onstraint) ≥ 2
Synonymous	"incomplete_terminal_codon_variant", "stop_retained_variant", "synonymous_variant"

Table S6. Gene sets collected for burden analysis

Gene set	Number of genes	Reference		**Genes** (display only smaller gene sets)
Constrained genes				
LoF-intolerant (pLI > 0.9)	3,488	Samocha 2014[24]	ExAC non-psych	
LoF-intolerant (pLI > 0.995)	1,583	Samocha 2014[24]	ExAC non-psych	
Missense constrained (misZ > 3.09)	1,730	Samocha 2014[24]	ExAC non-psych	
Tissue-specific				
Brain enriched (GTEx)	2,649	Ganna 2016[25]		
Candidate epilepsy genes				
Known epilepsy genes	43	Epi4K 2017[26]	Table S5	*SPTAN1, SCN8A, CHD2, SCN2A, GRIN2B, GRIN2A, SYNGAP1, KCNMA1, PRICKLE2, SCN1A, CDKL5, KCNT1, GRIN1, PCDH19, CHRNA4, KCNQ2, DEPDC5, KCNQ3, HCN1, CHRNA2, LGI1, SLC6A1, GNAO1, SLC2A1, GABRG2, KCNC1, DNM1, CHRNB2, KCNA2, HNRNPU, EEF1A2, STXBP1, MEF2C, GABRB3, GABRA1, SCN1B, SLC35A2, STX1B, KCNB1, PRRT2, SCN9A, SIK1, ALG13*
Known DEE genes	50	Heyne 2018[27]	Table S3	*ALG13 ,ARHGEF9, ARX, CACNA1A, CASK, CDKL5, CHD2, DNM1, EEF1A2, FOXG1, GABRA1, GABRB3, GNAO1, GNB1, GPHN, GRIN1, GRIN2A, GRIN2B, HCN1, IQSEC2, KCNA2, KCNB1, KCNQ2, KCNT1, MBD5, MECP2, MEF2C, PCDH19, PIGA, PURA, SCN1A, SCN2A, SCN8A, SIK1, SLC2A1, SLC35A2, SLC6A1, SLC6A8, SLC9A6, SPTAN1, STXBP1, SYN1, SYNGAP1, TSC1, TSC2, UBE3A, WDR45, ZEB2, SLC1A2, GRIN2D*
Genes in neurodevelopmental disorders (NDDs) with epilepsy	33	Heyne 2018[27]	Table S6	*KCNQ2, SCN2A, SCN1A, STXBP1, CHD2, CDKL5, DNM1, DYRK1A, MEF2C, SYNGAP1, GABRB3, EEF1A2, SLC6A1, SCN8A, PURA, WDR45, GNAO1, HNRNPU, SMC1A, FOXG1, ARID1B, GRIN2A, GRIN2B, ALG13, ASXL3, KCNH1, GABRB2, NEXMIF, MECP2, SNAP25, COL4A3BP, SLC35A2, ARHGEF9*
Neurotransmission				
GABA-A receptor genes	19	May 2018[28]	Table S5	*GABRA1, GABRA2, GABRA3, GABRA4, GABRA5, GABRA6, GABRB1, GABRB2, GABRB3, GABRD, GABRE, GABRG1, GABRG2, GABRG3, GABRP, GABRQ, GABRR1, GABRR2, GABRR3*

GABAergic pathway genes	113	May 2018[28]	Table S5	*ABAT, ADCY1, ADCY2, ADCY3, ADCY4, ADCY5, ADCY6, ADCY7, ADCY8, ADCY9, ANK2, ANK3, ARHGEF9, DISC1, DLC1, DLC2, DNAI1, FGF13, GABARAP, GABARAPL1, GABARAPL2, GABBR1, GABBR2, GABRA1, GABRA2, GABRA3, GABRA4, GABRA5, GABRA6, GABRB1, GABRB2, GABRB3, GABRD, GABRE, GABRG1, GABRG2, GABRG3, GABRP, GABRQ, GABRR1, GABRR2, GABRR3, GAD1, GAD2, GLS, GLS2, GLUL, GNAI1, GNAI2, GNAI3, GNAO1, GNB1, GNB2, GNB3, GNB4, GNB5, GNG10, GNG11, GNG12, GNG13, GNG2, GNG3, GNG4, GNG5, GNG7, GNG8, GNGT1, GNGT2, GPHN, HAP1, KCNB2, KCNC1, KCNC2, KCNC3, KCNJ6, KIF5A, KIF5B, KIF5C, MAGI, MKLN1, MYO5A, NLGN2, NRXN1, NSF, PFN1, PLCL1, PRKACA, PRKACB, PRKACG, PRKCA, PRKCB, PRKCG, RAFT1, RDX, SCN1A, SCN1B, SCN2B, SCN3A, SCN8A, SEMA4D, SLC12A2, SLC12A5, SLC32A1, SLC38A1, SLC38A2, SLC38A3, SLC38A5, SLC6A1,SLC6A11, SLC6A13, SRC, TRAK1, TRAK2*
Excitatory receptor genes (Glutamate ionotropic receptors & cholinergic receptors)	34	May 2018[28]	Table S5	*CHRNA1, CHRNA10, CHRNA2, CHRNA3, CHRNA4, CHRNA5, CHRNA6, CHRNA7, CHRNA9, CHRNB1, CHRNB2, CHRNB3, CHRNB4, CHRND, CHRNE, CHRNG, GRIA1, GRIA2, GRIA3, GRIA4, GRIK1, GRIK2, GRIK3, GRIK4, GRIK5, GRIN1, GRIN2A, GRIN2B, GRIN2C, GRIN2D, GRIN3A, GRIN3B, GRID1, GRID2*
Voltage-gated cation channel genes	86	May 2018[28]	Table S5	*SCN10A, SCN11A, SCN1A, SCN1B, SCN2A2, SCN2B, SCN3A, SCN3B, SCN4A, SCN4B, SCN5A, SCN7A, SCN8A, SCN9A, CACNA1A, CACNA1B, CACNA1C, CACNA1D, CACNA1E, CACNA1F, CACNA1G, CACNA1H, CACNA1I, CACNA1S, CACNA2D1, CACNA2D2, CACNA2D3, CACNA2D4, CACNB1, CACNB2, CACNB3, CACNB4, KCNA1, KCNA10, KCNA2, KCNA3, KCNA4, KCNA5, KCNA6,KCNA7, KCNAB1, KCNAB2, KCNAB3, KCNB1, KCNB2, KCNC1, KCNC2, KCNC3, KCNC4, KCND1, KCND2, KCND3, KCNE1, KCNE1L, KCNE2, KCNE3, KCNE4, KCNF1, KCNG1, KCNG2, KCNG3, KCNG4, KCNH1, KCNH2, KCNH3, KCNH4, KCNH5, KCNH6, KCNH7, KCNH8, KCNQ1, KCNQ2, KCNQ3, KCNQ5, KCNQ4, KCNRG, KCNS1, KCNS2, KCNS3, KCNT1, KCNV1, KCNV2, HCN1, HCN2, HCN3, HCN4*

Table S7. Prior gene panel screening of the 1,021 DEE-affected individuals
More than 50 different gene panels were used across our patient cohorts. This table summarizes the overall screening results, but heterogeneity exists in the numbers and the types of genes tested, how the testing was performed, etc.

Gene panel screening	Results		
	Abnormal or unknown	Normal	Total
Yes	89	257	346
No	-	-	247
Not entered/answered	-	-	428

Table S8. Estimating sample size required for lead genes of GGE and NAFE to achieve exome-wide significance

Focusing on top genes for GGE (e.g., *CACNA1G*, *UNC79*, *GABRG2*) and NAFE (e.g., *TRIM3*, *LGI1*, *GABRG2*), the aggregated allele frequency (AF) of the deleterious URVs in these genes is approximately 0.001 in cases and 0.0001 in controls (see Tables S13 & S15). If we are willing to hold the case- and control-specific AF constant, as well as assuming we can maintain the case:control ratio at around 1:2.5, the following table shows the corresponding Fisher's Exact p-values with increasing sample size:
(take *GABRG2* in the GGE cohort for example, the first row shows the current sample size and AF)

N_case	N_control	AC_case	AC_control	AF_case	AF_control	P-value
3108	8436	7	1	~0.0011	~0.0001	6.2×10^{-4}
6000	15000	12	3	0.001	0.0001	5.4×10^{-5}
7000	17500	14	3.5	0.001	0.0001	2.2×10^{-5}
8000	20000	16	4	0.001	0.0001	2.8×10^{-6}
8500	21250	17	4.25	0.001	0.0001	9.7×10^{-7}

Consider testing one epilepsy type at a time, Bonferroni correction suggests an exome-wide significance at a p-value < 1/20,000 (genes) = 2.5×10^{-6}. That means, under this particular scenario, we will need about 8000 cases and 20,000 controls to identify exome-wide significant genes for GGE and NAFE, separately.

Supplemental References

1. Conte, F., Legros, B., Van Paesschen, W., Avbersek, A., Muglia, P., and Depondt, C. (2018). Long-term seizure outcomes in patients with drug resistant epilepsy. Seizure 62, 74-78.
2. Miller, S.A., Dykes, D.D., and Polesky, H.F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16, 1215.
3. Brainstorm, C., Anttila, V., Bulik-Sullivan, B., Finucane, H.K., Walters, R.K., Bras, J., Duncan, L., Escott-Price, V., Falcone, G.J., Gormley, P., et al. (2018). Analysis of shared heritability in common disorders of the brain. Science 360.
4. McCormack, M., Gui, H., Ingason, A., Speed, D., Wright, G.E.B., Zhang, E.J., Secolin, R., Yasuda, C., Kwok, M., Wolking, S., et al. (2018). Genetic variation in CFH predicts phenytoin-induced maculopapular exanthema in European-descent patients. Neurology 90, e332-e341.
5. Lal, D., Ruppert, A.K., Trucks, H., Schulz, H., de Kovel, C.G., Kasteleijn-Nolst Trenite, D., Sonsma, A.C., Koeleman, B.P., Lindhout, D., Weber, Y.G., et al. (2015). Burden analysis of rare microdeletions suggests a strong impact of neurodevelopmental genes in genetic generalised epilepsies. PLoS Genet 11, e1005226.
6. The International League Against Epilepsy Consortium on Complex Epilepsies. (2014). Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. Lancet Neurol 13, 893-903.
7. Leschziner, G., Jorgensen, A.L., Andrew, T., Pirmohamed, M., Williamson, P.R., Marson, A.G., Coffey, A.J., Middleditch, C., Rogers, J., Bentley, D.R., et al. (2006). Clinical factors and ABCB1 polymorphisms in prediction of antiepileptic drug response: a prospective cohort study. Lancet Neurol 5, 668-676.
8. Speed, D., Hoggart, C., Petrovski, S., Tachmazidou, I., Coffey, A., Jorgensen, A., Eleftherohorinou, H., De Iorio, M., Todaro, M., De, T., et al. (2014). A genome-wide association study and biological pathway analysis of epilepsy prognosis in a prospective cohort of newly treated epilepsy. Hum Mol Genet 23, 247-258.
9. Guo, Y., Baum, L.W., Sham, P.C., Wong, V., Ng, P.W., Lui, C.H., Sin, N.C., Tsoi, T.H., Tang, C.S., Kwan, J.S., et al. (2012). Two-stage genome-wide association study identifies variants in CAMSAP1L1 as susceptibility loci for epilepsy in Chinese. Hum Mol Genet 21, 1184-1189.
10. Bisulli, F., Menghi, V., Vignatelli, L., Licchetta, L., Zenesini, C., Stipa, C., Morigi, F., Gizzi, M., Avoni, P., Provini, F., et al. (2018). Epilepsy with auditory features: Long-term outcome and predictors of terminal remission. Epilepsia 59, 834-843.
11. Licchetta, L., Bisulli, F., Vignatelli, L., Zenesini, C., Di Vito, L., Mostacci, B., Rinaldi, C., Trippi, I., Naldi, I., Plazzi, G., et al. (2017). Sleep-related hypermotor epilepsy: Long-term outcome in a large cohort. Neurology 88, 70-77.
12. Smith, L.A., Ullmann, J.F., Olson, H.E., Achkar, C.M., Truglio, G., Kelly, M., Rosen-Sheidley, B., and Poduri, A. (2017). A Model Program for Translational Medicine in Epilepsy Genetics. J Child Neurol 32, 429-436.
13. Buono, R.J., Lohoff, F.W., Sander, T., Sperling, M.R., O'Connor, M.J., Dlugos, D.J., Ryan, S.G., Golden, G.T., Zhao, H., Scattergood, T.M., et al. (2004). Association between variation in the human KCNJ10 potassium ion channel gene and seizure susceptibility. Epilepsy Res 58, 175-183.
14. EPGP Collaborative, Abou-Khalil, B., Alldredge, B., Bautista, J., Berkovic, S., Bluvstein, J., Boro, A., Cascino, G., Consalvo, D., Cristofaro, S., et al. (2013). The epilepsy phenome/genome project. Clin Trials 10, 568-586.
15. Measso, G., Cavarzeran, F., Zappalà, G., Lebowitz, B.D., Crook, T.H., Pirozzolo, F.J., Amaducci, L.A., Massari, D., and Grigoletto, F. (1993). The mini-mental state examination: normative study of an Italian random sample. Developmental Neuropsychology 9, 77-85.
16. Carlesimo, G.A., Caltagirone, C., and Gainotti, G. (1996). The Mental Deterioration Battery: normative data, diagnostic reliability and qualitative analyses of cognitive impairment. The Group for the Standardization of the Mental Deterioration Battery. Eur Neurol 36, 378-384.

17. McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., and Stadlan, E.M. (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 34, 939-944.
18. Hamilton, M. (1967). Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 6, 278-296.
19. Cerasa, A., Quattrone, A., Piras, F., Mangone, G., Magariello, A., Fagioli, S., Girardi, P., Muglia, M., Caltagirone, C., and Spalletta, G. (2014). 5-HTTLPR, anxiety and gender interaction moderates right amygdala volume in healthy subjects. Soc Cogn Affect Neurosci 9, 1537-1545.
20. Sharp, S.I., Lange, J., Kandaswamy, R., Daher, M., Anjorin, A., Bass, N.J., and McQuillin, A. (2017). Identification of rare nonsynonymous variants in SYNE1/CPG2 in bipolar affective disorder. Psychiatr Genet 27, 81-88.
21. Schizophrenia Working Group of the Psychiatric Genomics, C. (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421-427.
22. Borodulin, K., Tolonen, H., Jousilahti, P., Jula, A., Juolevi, A., Koskinen, S., Kuulasmaa, K., Laatikainen, T., Mannisto, S., Peltonen, M., et al. (2017). Cohort Profile: The National FINRISK Study. Int J Epidemiol.
23. Pato, M.T., Sobell, J.L., Medeiros, H., Abbott, C., Sklar, B.M., Buckley, P.F., Bromet, E.J., Escamilla, M.A., Fanous, A.H., Lehrer, D.S., et al. (2013). The genomic psychiatry cohort: partners in discovery. Am J Med Genet B Neuropsychiatr Genet 162B, 306-312.
24. Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., McGrath, L.M., Kosmicki, J.A., Rehnstrom, 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.
25. Ganna, A., Genovese, G., Howrigan, D.P., Byrnes, A., Kurki, M., Zekavat, S.M., Whelan, C.W., Kals, M., Nivard, M.G., Bloemendal, A., et al. (2016). Ultra-rare disruptive and damaging mutations influence educational attainment in the general population. Nat Neurosci 19, 1563-1565.
26. Epi4K Consortium, and Epilepsy Phenome/Genome Project. (2017). Ultra-rare genetic variation in common epilepsies: a case-control sequencing study. Lancet Neurol 16, 135-143.
27. Heyne, H.O., Singh, T., Stamberger, H., Abou Jamra, R., Caglayan, H., Craiu, D., De Jonghe, P., Guerrini, R., Helbig, K.L., Koeleman, B.P.C., et al. (2018). De novo variants in neurodevelopmental disorders with epilepsy. Nat Genet 50, 1048-1053.
28. May, P., Girard, S., Harrer, M., Bobbili, D.R., Schubert, J., Wolking, S., Becker, F., Lachance-Touchette, P., Meloche, C., Gravel, M., et al. (2018). Rare coding variants in genes encoding GABAA receptors in genetic generalised epilepsies: an exome-based case-control study. Lancet Neurol 17, 699-708.

Table S19. Top 20 genes with burden of deleterious URVs for the recessive model by epilepsy subgroup

(A-1) DEE: PTV burden

Gene	Ncase_carrie	Ncase_nonca	Ncontrol_car	Ncontrol_noı	OR	Pval.obs
MAGEC1	4	1017	2	8434	16.5860374	0.00169389
EFCAB6	2	1019	1	8435	16.5554465	0.03242686
ADAM8	1	1020	0	8436	Inf	0.10796236
ARV1	1	1020	0	8436	Inf	0.10796236
BRAT1	1	1020	0	8436	Inf	0.10796236
TMEM255A	1	1020	0	8436	Inf	0.10796236
MFRP	1	1020	0	8436	Inf	0.10796236
PTCHD3	1	1020	0	8436	Inf	0.10796236
GPRASP1	1	1020	0	8436	Inf	0.10796236
CNTNAP2	1	1020	0	8436	Inf	0.10796236
METTL7B	1	1020	0	8436	Inf	0.10796236
PCDH11X	1	1020	0	8436	Inf	0.10796236
ABCC12	1	1020	0	8436	Inf	0.10796236
FCN3	1	1020	0	8436	Inf	0.10796236
ULBP3	1	1020	0	8436	Inf	0.10796236
KLHL7	1	1020	0	8436	Inf	0.10796236
PODN	1	1020	0	8436	Inf	0.10796236
ACBD6	1	1020	0	8436	Inf	0.10796236
NEXMIF	1	1020	0	8436	Inf	0.10796236
CHRDL1	1	1020	0	8436	Inf	0.10796236

(A-2) DEE: MPC burden

Gene	Ncase_carrie	Ncase_nonca	Ncontrol_car	Ncontrol_noı	OR	Pval.obs
PRKCD	1	1020	0	8436	Inf	0.10796236
PHF8	1	1020	0	8436	Inf	0.10796236
ARNT	1	1020	0	8436	Inf	0.10796236
ANKRD20A1	1	1020	0	8436	Inf	0.10796236
TBX5	1	1020	0	8436	Inf	0.10796236
LDOC1	1	1020	0	8436	Inf	0.10796236
OPHN1	1	1020	0	8436	Inf	0.10796236
ZNF280C	1	1020	0	8436	Inf	0.10796236
ZFX	1	1020	1	8435	8.26960784	0.20427903
ZIC3	1	1020	1	8435	8.26960784	0.20427903
IDH3G	1	1020	1	8435	8.26960784	0.20427903
GANAB	1	1020	1	8435	8.26960784	0.20427903
GOLGA8S	1	1020	1	8435	8.26960784	0.20427903
RPS6KA3	1	1020	1	8435	8.26960784	0.20427903
FRMPD4	1	1020	3	8433	2.75588235	0.36686055
PHKA2	1	1020	3	8433	2.75588235	0.36686055
TAF1	1	1020	5	8431	1.65313725	0.49624927
G6PD	1	1020	6	8430	1.37745098	0.55066991
HCFC1	2	1019	14	8422	1.18070938	0.68865261
CLCN4	1	1020	9	8427	0.91797386	1

(A-3) DEE: in-frame indel burden

Gene	Ncase_carrie	Ncase_nonca	Ncontrol_car	Ncontrol_noı	OR	Pval.obs
KCNK17	2	1019	0	8436	Inf	0.01164569
PQBP1	2	1019	2	8434	8.2767419	0.06023877
PASD1	7	1014	26	8410	2.2329692	0.08160848
TSNAXIP1	1	1020	0	8436	Inf	0.10796236
TREML1	1	1020	0	8436	Inf	0.10796236
C11orf63	1	1020	0	8436	Inf	0.10796236
CCNB3	1	1020	0	8436	Inf	0.10796236
ARMCX5	1	1020	0	8436	Inf	0.10796236
CPSF6	1	1020	0	8436	Inf	0.10796236
KIAA1551	1	1020	0	8436	Inf	0.10796236
SLC7A6	1	1020	0	8436	Inf	0.10796236
NIPA2	1	1020	0	8436	Inf	0.10796236
PLEKHH3	1	1020	0	8436	Inf	0.10796236
GABRE	1	1020	0	8436	Inf	0.10796236
PGK1	1	1020	0	8436	Inf	0.10796236
KLHL4	1	1020	1	8435	8.26960784	0.20427903
ARG2	1	1020	1	8435	8.26960784	0.20427903
RMDN1	1	1020	1	8435	8.26960784	0.20427903
VSIG1	1	1020	1	8435	8.26960784	0.20427903
AFF2	1	1020	2	8434	4.13431373	0.29020511

(B-1) GGE: PTV burden

Gene	Ncase_carrie	Ncase_nonca	Ncontrol_car	Ncontrol_noı	OR	Pval.obs
ACTRT1	1	3107	18	8418	0.15052033	0.03580342
PDIA2	2	3106	0	8436	Inf	0.07246816
FLG	2	3106	0	8436	Inf	0.07246816
IGSF10	2	3106	0	8436	Inf	0.07246816
OR13H1	2	3106	1	8435	5.43142305	0.17840152
ITIH6	0	3108	7	8429	0	0.2008314
SLC39A2	1	3107	0	8436	Inf	0.26923077
SLC6A8	1	3107	0	8436	Inf	0.26923077
WNT10A	1	3107	0	8436	Inf	0.26923077
HIST1H3A	1	3107	0	8436	Inf	0.26923077
MYH7B	1	3107	0	8436	Inf	0.26923077
PJA1	1	3107	0	8436	Inf	0.26923077
DCAF17	1	3107	0	8436	Inf	0.26923077
CHDC2	1	3107	0	8436	Inf	0.26923077
PCDHA9	1	3107	0	8436	Inf	0.26923077
SPNS3	1	3107	0	8436	Inf	0.26923077
C17orf80	1	3107	0	8436	Inf	0.26923077
TBC1D8B	1	3107	0	8436	Inf	0.26923077
ANKRD30A	1	3107	0	8436	Inf	0.26923077
CST9L	1	3107	0	8436	Inf	0.26923077

(B-2) GGE: MPC burden

Gene	Ncase_carrie	Ncase_nonca	Ncontrol_car	Ncontrol_noı	OR	Pval.obs
WEE1	3	3105	0	8436	Inf	0.01950148
RPS6KA3	3	3105	1	8435	8.14975845	0.06226585
TBC1D25	0	3108	6	8430	0	0.20084626
CNNM4	1	3107	0	8436	Inf	0.26923077
CSTF2	1	3107	0	8436	Inf	0.26923077
PABPC5	1	3107	0	8436	Inf	0.26923077
RAP2C	1	3107	0	8436	Inf	0.26923077
CRHR1	1	3107	0	8436	Inf	0.26923077
CT83	1	3107	0	8436	Inf	0.26923077
SPIN4	1	3107	0	8436	Inf	0.26923077
COL4A3BP	1	3107	0	8436	Inf	0.26923077
CASK	1	3107	0	8436	Inf	0.26923077
JADE3	1	3107	0	8436	Inf	0.26923077
GABRA3	1	3107	0	8436	Inf	0.26923077
NFYC	1	3107	0	8436	Inf	0.26923077
PRAMEF4	1	3107	0	8436	Inf	0.26923077
STAG2	1	3107	0	8436	Inf	0.26923077
APOBEC3B	1	3107	0	8436	Inf	0.26923077
FAM199X	1	3107	0	8436	Inf	0.26923077
RBMX	1	3107	0	8436	Inf	0.26923077

(B-3) GGE: in-frame indel burden

Gene	Ncase_carrie	Ncase_nonca	Ncontrol_car	Ncontrol_noı	OR	Pval.obs
TRO	5	3103	1	8435	13.5916855	0.00656994
ATRX	2	3106	0	8436	Inf	0.07246816
RFX7	1	3107	0	8436	Inf	0.26923077
CHD2	1	3107	0	8436	Inf	0.26923077
TRPM4	1	3107	0	8436	Inf	0.26923077
RAB40AL	1	3107	0	8436	Inf	0.26923077
SLA2	1	3107	0	8436	Inf	0.26923077
TSPYL2	1	3107	0	8436	Inf	0.26923077
F9	1	3107	0	8436	Inf	0.26923077
MAGEC1	1	3107	0	8436	Inf	0.26923077
PRRC2A	1	3107	0	8436	Inf	0.26923077
SECISBP2L	1	3107	0	8436	Inf	0.26923077
NIPA2	1	3107	0	8436	Inf	0.26923077
DDX3X	1	3107	0	8436	Inf	0.26923077
LDOC1	1	3107	0	8436	Inf	0.26923077
SH3BGR	1	3107	0	8436	Inf	0.26923077
TFDP3	1	3107	0	8436	Inf	0.26923077
RGAG1	1	3107	0	8436	Inf	0.26923077
ACRC	3	3105	5	8431	1.62917874	0.45111715
TMC2	1	3107	1	8435	2.71483746	0.46599338

(C-1) NAFE: PTV burden

Gene	Ncase_carrie	Ncase_nonca	Ncontrol_car	Ncontrol_noı	OR	Pval.obs
OR13H1	3	3594	1	8435	7.0409015	0.08285734
AMER1	2	3595	0	8436	Inf	0.0893405
CXorf22	0	3597	8	8428	0	0.11537685
VSIG4	5	3592	4	8432	2.93429844	0.13828031
MAGEC1	3	3594	2	8434	3.52003339	0.16162053
KRT83	2	3595	1	8435	4.69262865	0.21462958
EFCAB6	2	3595	1	8435	4.69262865	0.21462958
TNFRSF10C	2	3595	1	8435	4.69262865	0.21462958
CNGA2	2	3595	1	8435	4.69262865	0.21462958
OR2T11	2	3595	1	8435	4.69262865	0.21462958
SDSL	2	3595	1	8435	4.69262865	0.21462958
KIAA1377	1	3596	0	8436	Inf	0.29892795
TRIM31	1	3596	0	8436	Inf	0.29892795
ZNF135	1	3596	0	8436	Inf	0.29892795
PTGER3	1	3596	0	8436	Inf	0.29892795
CD36	1	3596	0	8436	Inf	0.29892795
ZNF813	1	3596	0	8436	Inf	0.29892795
TMC2	1	3596	0	8436	Inf	0.29892795
SYN1	1	3596	0	8436	Inf	0.29892795
PCDHA9	1	3596	0	8436	Inf	0.29892795

(C-2) NAFE: MPC burden

Gene	Ncase_carrie	Ncase_nonca	Ncontrol_car	Ncontrol_noı	OR	Pval.obs
G6PD	7	3590	6	8430	2.73955432	0.07103174
ZDHHC9	3	3594	1	8435	7.0409015	0.08285734
RPS6KA3	3	3594	1	8435	7.0409015	0.08285734
TRPC5	0	3597	8	8428	0	0.11537685
FMR1	3	3594	2	8434	3.52003339	0.16162053
NDUFA1	12	3585	44	8392	0.63841765	0.18911587
IDH3G	2	3595	1	8435	4.69262865	0.21462958
HDAC6	2	3595	1	8435	4.69262865	0.21462958
SLC6A8	1	3596	0	8436	Inf	0.29892795
ZMYM3	1	3596	0	8436	Inf	0.29892795
BCAP31	1	3596	0	8436	Inf	0.29892795
SLC25A5	1	3596	0	8436	Inf	0.29892795
PHF8	1	3596	0	8436	Inf	0.29892795
NOS1	1	3596	0	8436	Inf	0.29892795
KDM5C	1	3596	0	8436	Inf	0.29892795
NONO	1	3596	0	8436	Inf	0.29892795
RABL2A	1	3596	0	8436	Inf	0.29892795
PRPS1	1	3596	0	8436	Inf	0.29892795
BEST2	1	3596	0	8436	Inf	0.29892795
MED14	1	3596	0	8436	Inf	0.29892795

(C-3) NAFE: in-frame indel burden

Gene	Ncase_carrie	Ncase_nonca	Ncontrol_car	Ncontrol_noı	OR	Pval.obs
BMP15	9	3588	43	8393	0.48959711	0.04859612
MAGEC1	2	3595	0	8436	Inf	0.0893405
KRTAP5-1	2	3595	0	8436	Inf	0.0893405
SHROOM2	2	3595	1	8435	4.69262865	0.21462958
LUZP4	2	3595	1	8435	4.69262865	0.21462958
KLB	1	3596	0	8436	Inf	0.29892795
ZNF577	1	3596	0	8436	Inf	0.29892795
TSNAXIP1	1	3596	0	8436	Inf	0.29892795
REXO4	1	3596	0	8436	Inf	0.29892795
PCDHA9	1	3596	0	8436	Inf	0.29892795
PHKA2	1	3596	0	8436	Inf	0.29892795
HELB	1	3596	0	8436	Inf	0.29892795
NAP1L3	1	3596	0	8436	Inf	0.29892795
NEFM	1	3596	0	8436	Inf	0.29892795
MORC4	1	3596	0	8436	Inf	0.29892795
GDPD2	1	3596	0	8436	Inf	0.29892795
PRAMEF1	1	3596	0	8436	Inf	0.29892795
MPP1	1	3596	0	8436	Inf	0.29892795
NXPE1	1	3596	0	8436	Inf	0.29892795
FLVCR2	1	3596	0	8436	Inf	0.29892795

(D-1) EPI: PTV burden

Gene	Ncase_carrie	Ncase_nonca	Ncontrol_car	Ncontrol_noı	OR	Pval.obs
DGAT2L6	0	9170	5	8431	0	0.02524127
ITIH6	1	9169	7	8429	0.13132761	0.03232174
LILRA2	0	9170	4	8432	0	0.05269174
MAP7D2	0	9170	4	8432	0	0.05269174
ACTRT1	9	9161	18	8418	0.45944766	0.05540856
MAGEC1	9	9161	2	8434	4.14288833	0.06787037
TP53AIP1	0	9170	3	8433	0	0.10998847
CYSLTR1	0	9170	3	8433	0	0.10998847
PIR	0	9170	3	8433	0	0.10998847
CXorf22	3	9167	8	8428	0.34476928	0.13261292
OR13H1	5	9165	1	8435	4.60174577	0.2211133
IFI44L	0	9170	2	8434	0	0.22957518
PLXNA3	0	9170	2	8434	0	0.22957518
USP6	0	9170	2	8434	0	0.22957518
CHP2	0	9170	2	8434	0	0.22957518
ZNF418	0	9170	2	8434	0	0.22957518
XAGE5	0	9170	2	8434	0	0.22957518
ITGB1BP2	0	9170	2	8434	0	0.22957518
ZCCHC16	0	9170	2	8434	0	0.22957518
TLDC2	0	9170	2	8434	0	0.22957518

(D-2) EPI: MPC burden

Gene	Ncase_carrie	Ncase_nonca	Ncontrol_car	Ncontrol_noı	OR	Pval.obs
TBC1D25	1	9169	6	8430	0.15323372	0.06027489
RPS6KA3	7	9163	1	8435	6.44385027	0.07215284
TLR7	0	9170	3	8433	0	0.10998847
WEE1	4	9166	0	8436	Inf	0.12626134
TRPC5	3	9167	8	8428	0.34476928	0.13261292
SLC16A2	0	9170	2	8434	0	0.22957518
KLHL15	0	9170	2	8434	0	0.22957518
CGB1	0	9170	2	8434	0	0.22957518
GDI1	0	9170	2	8434	0	0.22957518
NKRF	0	9170	2	8434	0	0.22957518
HNRNPH2	0	9170	2	8434	0	0.22957518
ATRX	0	9170	2	8434	0	0.22957518
BTK	0	9170	2	8434	0	0.22957518
GRIA3	0	9170	2	8434	0	0.22957518
CDKL5	0	9170	2	8434	0	0.22957518
IL1RAPL1	0	9170	2	8434	0	0.22957518
PMS2	0	9170	2	8434	0	0.22957518
ZMYM3	3	9167	0	8436	Inf	0.25126104
USP9X	2	9168	5	8431	0.36784468	0.27107731
G6PD	11	9159	6	8430	1.68741129	0.33926598

(D-3) EPI: in-frame indel burden

Gene	Ncase_carrie	Ncase_nonca	Ncontrol_car	Ncontrol_noı	OR	Pval.obs
BMP15	30	9140	43	8393	0.64065442	0.06126956
TRO	7	9163	1	8435	6.44385027	0.07215284
C3orf30	0	9170	3	8433	0	0.10998847
FOXR2	0	9170	2	8434	0	0.22957518
CD200R1	0	9170	2	8434	0	0.22957518
GPRASP1	0	9170	2	8434	0	0.22957518
OR2T12	0	9170	2	8434	0	0.22957518
MAP1A	0	9170	2	8434	0	0.22957518
TMEM187	0	9170	2	8434	0	0.22957518
MAGEC1	3	9167	0	8436	Inf	0.25126104
KRTAP5-1	3	9167	0	8436	Inf	0.25126104
MAGEC2	1	9169	3	8433	0.30657651	0.35544825
DMD	1	9169	3	8433	0.30657651	0.35544825
PQBP1	5	9165	2	8434	2.30060011	0.45587951
KIAA1377	0	9170	1	8435	0	0.47915483
EDA	0	9170	1	8435	0	0.47915483
SLC6A8	0	9170	1	8435	0	0.47915483
DICER1	0	9170	1	8435	0	0.47915483
PRSS16	0	9170	1	8435	0	0.47915483
PRM3	0	9170	1	8435	0	0.47915483

Table S20. Top 20 genes with burden of deleterious URVs (AC=1), removing patients with family history (of an affected first-degree relative)

(A) DEE

Ncase = 1021, Ncontrol = 8436			
Gene	**Ncase carrier**	**Ncontrol carrier**	**Pval**
SCN1A	11	5	5.8E-08
NEXMIF	6	0	1.6E-06
COBLL1	5	0	1.5E-05
KCNB1	5	2	2.5E-04
SCN8A	5	3	6.1E-04
PASK	4	1	6.2E-04
CEPT1	3	0	1.3E-03
PHYHIP	3	0	1.3E-03
TAS1R3	4	3	3.6E-03
SLC6A1	4	3	3.6E-03
YWHAG	3	1	4.6E-03
GTSE1	3	1	4.6E-03
PIP4K2B	3	1	4.6E-03
CCDC50	3	1	4.6E-03
KDM5C	3	1	4.6E-03
APOBEC3D	3	1	4.6E-03
ATP9A	3	1	4.6E-03
PAPSS2	3	1	4.6E-03
APC	3	1	4.6E-03
LRRC8A	4	4	6.6E-03

Ncase = 925, Ncontrol = 8436			
Gene	**Ncase carrier (no famhx)**	**Ncontrol carrier**	**Pval**
SCN1A	11	5	2.3E-08
NEXMIF	6	0	9.2E-07
COBLL1	5	0	9.3E-06
SCN8A	5	3	4.0E-04
PHYHIP	3	0	9.6E-04
KCNB1	4	2	1.2E-03
TAS1R3	4	3	2.6E-03
SLC6A1	4	3	2.6E-03
PASK	3	1	3.6E-03
GTSE1	3	1	3.6E-03
PIP4K2B	3	1	3.6E-03
KDM5C	3	1	3.6E-03
APOBEC3D	3	1	3.6E-03
ATP9A	3	1	3.6E-03
PAPSS2	3	1	3.6E-03
APC	3	1	3.6E-03
LRRC8A	4	4	4.8E-03
TET3	4	4	4.8E-03
STXBP1	4	4	4.8E-03
AGO1	4	4	4.8E-03

(B) GGE

Ncase = 3108, Ncontrol = 8436			
Gene	**Ncase carrier**	**Ncontrol carrier**	**Pval**
CACNA1G	10	3	2.5E-04
EEF1A2	6	0	3.8E-04
UNC79	7	1	6.2E-04
GABRG2	7	1	6.2E-04
ALDH4A1	5	0	1.4E-03
SLC6A1	8	3	2.0E-03
RC3H2	6	1	2.0E-03
GABRA1	7	2	2.2E-03
DNAJC13	4	0	5.2E-03
ZBTB2	4	0	5.2E-03
LRRFIP1	4	0	5.2E-03
HDLBP	5	1	6.6E-03
SLC9A1	5	1	6.6E-03
TTC7A	5	1	6.6E-03
FBXO42	5	1	6.6E-03
FRY	8	5	9.1E-03
CYLD	6	3	1.5E-02
ATR	6	3	1.5E-02
DNMT3A	0	14	1.6E-02
PPFIA3	5	2	1.8E-02

Ncase = 2728, Ncontrol = 8436			
Gene	**Ncase carrier (no famhx)**	**Ncontrol carrier**	**Pval**
ALDH4A1	5	0	8.7E-04
CACNA1G	8	3	1.0E-03
SLC6A1	8	3	1.0E-03
GABRG2	6	1	1.2E-03
EEF1A2	4	0	3.6E-03
LRRFIP1	4	0	3.6E-03
GABRA1	6	2	3.7E-03
UNC79	5	1	4.2E-03
FBXO42	5	1	4.2E-03
FRY	8	5	4.8E-03
ATR	6	3	8.8E-03
SPEN	5	2	1.2E-02
FAT2	5	2	1.2E-02
ZBTB17	7	5	1.2E-02
SLC12A5	7	5	1.2E-02
RC3H2	4	1	1.4E-02
ENTPD5	4	1	1.4E-02
CIC	4	1	1.4E-02
HDLBP	4	1	1.4E-02
CADM4	4	1	1.4E-02

(C) NAFE

Ncase = 3597, Ncontrol = 8436			
Gene	**Ncase carrier**	**Ncontrol carrier**	**Pval**
LGI1	7	0	2.1E-04
TRIM3	7	0	2.1E-04
GABRG2	7	1	1.3E-03
ZNF587	5	0	2.4E-03
ACSM2B	5	0	2.4E-03
FXR1	6	1	3.7E-03
CORO1A	9	4	3.9E-03
PPFIA3	7	2	4.2E-03
FBXL6	4	0	8.0E-03
KCNJ3	4	0	8.0E-03
ZDHHC11	4	0	8.0E-03
CSNK2B-LY6G	4	0	8.0E-03
NFATC3	4	0	8.0E-03
ARFGEF1	8	4	9.2E-03
SYNE1	2	24	9.8E-03
SCN8A	7	3	1.0E-02
DEPDC5	7	3	1.0E-02
ABCA4	7	3	1.0E-02
TRPM6	5	1	1.1E-02
EXPH5	6	2	1.1E-02

Ncase = 3400, Ncontrol = 8436			
Gene	**Ncase carrier (no famhx)**	**Ncontrol carrier**	**Pval**
TRIM3	7	0	1.6E-04
LGI1	6	0	5.6E-04
ZNF587	5	0	2.0E-03
ACSM2B	5	0	2.0E-03
FXR1	6	1	3.0E-03
GABRG2	6	1	3.0E-03
PPFIA3	7	2	3.3E-03
FBXL6	4	0	6.8E-03
KCNJ3	4	0	6.8E-03
ZDHHC11	4	0	6.8E-03
NFATC3	4	0	6.8E-03
CORO1A	8	4	7.1E-03
DEPDC5	7	3	8.1E-03
ABCA4	7	3	8.1E-03
TRPM6	5	1	8.9E-03
EXPH5	6	2	8.9E-03
KCNH7	6	2	8.9E-03
COL7A1	8	5	1.4E-02
SYNE1	2	24	1.6E-02
ARFGEF1	7	4	1.7E-02