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 rare, severe epilepsies and revealed a role of rare deleterious variation in common epilepsies. To identify the shared and distinct ultra-rare genetic risk factors for rare and common 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; the rare but severe developmental and epileptic encephalopathies (DEE), and the commoner phenotypes of 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 common and rare 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, one of the most common and burdensome neurological conditions worldwide[1; 2]. A core challenge for epilepsy genetics is identifying and disentangling the genetic architecture and biological mechanisms underlying the variety of epilepsy types (e.g., focal vs. generalized) and electroclinical syndromes. While the occurrence of epilepsy for many affected individuals carries an underlying genetic component[3-5], the highly heterogeneous nature of epileptic seizures, epilepsy types, severity, and comorbidity makes it difficult to determine the specific genetic risks for each patient. For individuals with common, complex types of epilepsy, where inheritance may be due to strongly acting mutations, oligogenic or polygenic, the discovery of genetic risk factors is particularly challenging.

Considerable progress in our understanding of the genetic risk factors for epilepsy has been made in recent years thanks to the rapid growth and advancement in sequencing technology. Dozens of epilepsy-causing genes have been identified in individuals diagnosed with severe epilepsy syndromes[6-10], known as the developmental and epileptic encephalopathies (DEE). DEE are rare in the population and typically begin early in life. Incidence of the entire group is not well established, but a recent epidemiological study of severe epilepsies limited to onset under 18 months found an incidence of 1 in 2000 births[11]. The incidence of Dravet syndrome, one of the important specific forms of DEE, has been shown in several studies as 1 in 22,000[10]. Individuals with DEE usually have developmental impairment, ranging from profound to mild. With such severity, sequencing-based studies continue to discover *de novo* pathogenic variants for DEE and have implicated genes encoding neuronal ion channels and receptors and genes involved in cellular signaling[6-10; 12]. The common subgroups of epilepsy, broadly comprising genetic generalized epilepsy (GGE) and non-acquired focal epilepsy (NAFE), account for a major proportion of all incident epilepsies[2; 13] and have been shown robust heritability in twin, family, and genome-wide association studies (GWAS)[4; 14-16]. Disappointingly, only a limited number of genes

had been discovered to date for the common epilepsies, mostly from rare monogenic families with focal epilepsies, and attempts to identify clear risk genes for GGE have been least successful[12; 17; 18]. In most cases, especially for non-familial onset, the specific pathogenic variants are not yet known, and gene findings from small-scale studies have often not been reproducible[19-21].

Current evidence of the genetic etiology of epilepsies has revealed both extensive phenotypic and genetic heterogeneity. Many of the identified genes are associated with a spectrum of mild to severe epilepsies, showing phenotypic pleiotropy or variable expressivity, and most of the electroclinical syndromes have diverse genetic causes[10; 22]. Two recent studies using whole-exome sequencing (WES) of hundreds of individuals with common familial epilepsies found an enrichment in ultra-rare genetic variation in genes associated with rare epilepsy syndromes[17] and in missense variants in a group of genes encoding all $GABA_A$ receptors, the most important neurotransmitter receptors for neuronal inhibition in the mammalian brain[18]. Given the complex genetic architecture of the epilepsies, it is therefore critical to pinpoint the distinct and overlapping genetic risk factors underlying different groups of epilepsy on a scale much larger than previous sequencing studies and beyond familial cases.

Here, we evaluate a WES case-control study of epilepsy from the Epi25 collaborative—an ongoing global effort that collected an unprecedented number of patient cohorts for primarily the three major classes of non-lesional epilepsies: DEE, GGE, and NAFE[22]. We aimed to characterize the genetic risk of ultra-rare coding variants across these common and rare epilepsy subgroups 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[23] 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

Patients or their legal guardians provided signed informed consent 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

129 forms required specific language according to the NIH Genomic Data Sharing policy
130 (http://gds.nih.gov/03policy2.html).
131
132 ***Whole exome sequencing data generation***
133 All samples were sequenced at the Broad Institute of Harvard and MIT on the Illumina HiSeq X
134 platform, with the use of 151 bp paired-end reads. Exome capture was performed with Illumina
135 Nextera® Rapid Capture Exomes or TruSeq Rapid Exome enrichment kit (target size 38 Mb),
136 except for three control cohorts (MIGen ATVB, MIGen Ottawa, and Swedish SCZ controls) for
137 which the Agilent SureSelect Human All Exon Kit was used (target size 28.6 Mb – 33 Mb).
138 Sequence data in the form of BAM files were generated using the Picard data-processing pipeline
139 and contained well-calibrated reads aligned to the GRCh37 human genome reference. Samples
140 across projects were then jointly called via the Genome Analysis Toolkit (GATK) best practice
141 pipeline[24] for data harmonization and variant discovery. This pipeline detected single nucleotide
142 (SNV) and small insertion/deletion (indel) variants from exome sequence data.
143
144 ***Quality control***
145 Variants were pre-filtered to keep only those passing the GATK VQSR (Variant Quality Score
146 Recalibration) metric and those lying outside of low complexity regions[25]. Genotypes with GQ <
147 20 and heterozygous genotype calls with allele balance > 0.8 or < 0.2 were set to missing. To
148 control for capture platform difference, we retained variants that resided in GENCODE coding
149 regions where 80% of Agilent and Illumina-sequenced samples show at least 10x coverage. This
150 resulted in the removal of ~50% of the called sites (23% of the total coding variants and 97% of
151 the total non-coding variants) but effectively reduced the call rate difference between cases and
152 controls (**Figure S1**). To further identify potential false positive sites due to technical variation, we
153 performed single variant association tests (for variants with a minor allele frequency MAF > 0.001)
154 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)[26] 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[27] 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)[28] 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)[29] (**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[30]—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[17; 31] 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, brain-enriched genes, and genes reported to be associated with epilepsy or epilepsy-related mechanisms[9; 17; 18; 32; 33] (**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[34]. 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[7], 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[35], 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[36], 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. 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[37], 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[30]. 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[17].

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 (**Table S6**). To evaluate the burden in constrained genes, we defined "loss-of-function (LoF) intolerant" genes with either a pLI score[29] > 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)[33]. 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).

We found that, consistent with a recent study that evaluated *de novo* burden in autism[38], burden signals of PTVs were mostly contained in genes with a pLI > 0.995 compared to pLI > 0.9 (**Figure S12** & **S13**). Focusing on pLI > 0.995 in the all-epilepsy case-control analysis, both protein-truncating and damaging missense (MPC[28]≥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 ($OR_{PTV} = 1.4$, $adjP_{PTV} = 9.1\times10^{-5}$; $OR_{MPC} = 1.2$, $adjP_{MPC} = 5.5\times10^{-3}$), 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 the common epilepsy types (**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[17] (*OR* = 6.3, *adjP* = 2.1×10^{-08}), 50 known dominant DEE genes[9] (*OR* = 9.1, *adjP* = 7.8×10^{-11}), and 33 genes with *de novo* burden in neurodevelopmental disorders with epilepsy[9] (*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 S8**). 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[32]—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**)[18]. 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, $adjP_{GGE} = 1.4\times10^{-3}$; OR_{NAFE} = 2.7, $adjP_{NAFE}$ = 0.039), GABAergic pathway genes (OR_{DEE} = 2.6, $adjP_{DEE} = 4.7\times10^{-5}$; OR_{GGE} = 1.9, $adjP_{GGE} = 9.9\times10^{-04}$; OR_{NAFE} = 1.4, $adjP_{NAFE}$ = 0.11), and voltage-gated cation channel genes (OR_{DEE} = 2.1, $adjP_{DEE} = 1.7\times10^{-03}$; 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 S8**). These results support a recent finding where rare missense variation in $GABA_A$ receptor genes conferred a significant risk to GGE[18], and together implicate the relative importance and involvement of damaging missense variants in abnormal inhibitory neurotransmission in both rare and common epilepsy types.

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 for these genes using clinical gene panels prior to enrollment. 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 S10**). Some carriers of deleterious URVs in lead genes were affected individuals with a normal result for gene panel testing; for example, 2 out of the 3 carriers of qualified URVs for *PURA* and 2 out of 5 for *KCNB1* had undergone previous genetic screening. (**Table S7**). This could be because different sample-contributing sites adopted different gene panels and not all of them included the lead genes found to carry variants qualifying from this study, or that during screening patients were found to carry a variant of uncertain significance that did not satisfy the ACMG guidelines[39]. 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*

385 > 49, $P = 1.2 \times 10^{-3}$), emerged on top, both of which have been implicated in DEE and
386 developmental disorders (**Table S11**).
387
388 ***Channel and transporter genes implicated in common epilepsies***
389 When evaluating gene burden in the GGE and NAFE epilepsy subgroups, we did not identify any
390 exome-wide significant genes. However, several candidate epilepsy genes made up the lead
391 associations, including ion channel and transporter genes known to cause rare forms of epilepsy.
392 For the GGE case-control analysis in singleton deleterious URVs, the lead associations included
393 four previously-associated genes (*EEF1A2, OR* = 32, $P = 3.8\times10^{-4}$; *GABRG2*, *OR* = 19.0, $P =$
394 6.2×10^{-4}; *SLC6A1*, *OR* = 7.3, $P = 2.0\times10^{-3}$; and *GABRA1*, *OR* = 9.5, $P = 2.2\times10^{-3}$), and two genes
395 (*CACNA1G*, *OR* = 9.1, $P = 2.5\times10^{-4}$; *UNC79, OR* = 19.0, $P = 6.2\times10^{-4}$) that were not previously
396 linked to epilepsy but are both highly expressed in the brain and under evolutionary constraint
397 (**Figures 3B**; **Table S12**). Although evidence has been mixed, *CACNA1G* was previously
398 implicated as a potential susceptibility gene for GGE in mutational analysis[40] and reported to
399 modify mutated sodium channel (*SCN2A*) activity in epilepsy[41]. *UNC79* is an essential part of the
400 UNC79-UNC80-NALCN channel complex that influences neuronal excitability by interacting with
401 extracellular calcium ions[42], and this channel complex has been previously associated with
402 infantile encephalopathy[43]. Notably, all these lead genes were more enriched for damaging
403 missense (MPC≥2) than for protein-truncating URVs despite the lower rate of MPC≥2 variants
404 relative to PTVs (**Table S12**).
405 For individuals with NAFE, the analysis of singleton deleterious URVs identified *LGI1* and
406 *TRIM3* as the top two genes carrying a disproportionate number of deleterious URVs, however
407 neither reached exome-wide significance (*OR* > 32, $P = 2.1\times10^{-4}$). *GABRG2*, a lead association
408 in individuals with GGE, was among the top ten most enriched genes, along with two brain-
409 enriched, constrained genes (*PPFIA3*, *OR* = 8.2, $P = 4.2\times10^{-3}$; and *KCNJ3, OR* = 16.4, $P =$

410 1.2×10^{-3}). *GABRG2* has previously been reported to show an enrichment of variants compared
411 to controls in a cohort of individuals with Rolandic epilepsy (childhood epilepsy with
412 centrotemporal spikes) or related phenotypes, the most common group of focal epilepsies of
413 childhood[44]. Two other genes previously associated with epilepsy, *DEPDC5* and *SCN8A* (both
414 *OR* = 5.5, *P* = 0.01), were among the top twenty associations (**Figures 3C**; **Table S14**). *LGI1* and
415 *DEPDC5* are established genes for focal epilepsy, and *DEPDC5* was the only exome-wide
416 significant hit in the Epi4K WES study for familial NAFE cases[17]. *TRIM3* has not been previously
417 implicated in epilepsy, but evidence from a mouse model study implicates it in regulation of
418 $GABA_A$ receptor signaling and thus modulation of seizure susceptibility[45]. Single gene burden for
419 both GGE and NAFE remained similar when considering URVs with an allele count up to AC≤3
420 (**Tables S13 & S15**). Gene burden tests collapsing all epilepsy phenotypes recapitulated the lead
421 genes in each of the subgroup-specific analyses, but none of the genes achieved exome-wide
422 significance (**Tables S16 & S17**). It is worth noting that some of the genes were enriched for
423 deleterious URVs among the "controls", which is clearly driven by non-neuropsychiatric disease
424 ascertainment for many of the available controls (e.g., *LDLR* in **Table S16**; most control carriers
425 were individuals with cardiovascular diseases from the MIGen cohorts in **Table S2**). Thus, these
426 should not be interpreted as potential protective signals for epilepsy.
427
428 ***Recessive model, SKAT gene test, and single variant association***
429 The secondary gene-based test of a recessive model did not identify genes that differed
430 significantly in the carrier rate of homozygous deleterious variants between epilepsy-affected
431 individuals and controls (**Table S18**). Even if we considered variants up to MAF < 0.01, for most
432 of the lead genes, only one case carrier was identified. For the DEE cohort, these genes included
433 recessive genes previously implicated, such as *ARV1*, *BRAT1*, *CHRDL*[46] with a homozygous PTV
434 and *OPHN1*[46] with a recessive missense (MPC≥2) variant (**Table S18A**). For the two common
435 forms of epilepsy, a few studied recessive epilepsy genes were also observed in the lead gene

436 associations, such as *SLC6A8*[46] (a homozygous PTV) for GGE (**Table S18B**), and *SLC6A8* (a
437 homozygous missense-MPC) and *SYN1*[46] (a homozygous PTV) for NAFE (**Table S18C**). One
438 GGE-affected individual was found homozygous for an in-frame deletion on *CHD2*, a dominant
439 DEE gene[46] (**Table S18B**). These findings suggest an even larger cohort will be needed to identify
440 with clarity risk genes that act in a recessive manner for different groups of epilepsy.

441 Beyond URVs, we studied the contribution of low frequency deleterious variants to
442 epilepsy risk using SKAT (MAF < 0.01). Top associations for individuals with DEE included known
443 genes, such as missense-enriched *STXBP1* ($P = 9.3\times10^{-9}$), *KCNA2* ($P = 1.0\times10^{-5}$; **Figure S17**),
444 and PTV-enriched *NEXMIF* ($P = 7.1\times10^{-8}$), and *SCN1A* ($P = 3.9\times10^{-4}$; **Figure S18**). However, no
445 significant gene enrichment was observed for the two common types of epilepsy or when
446 combining all epilepsy cases. The tests for PTVs and missense variants with MPC≥2 were mostly
447 underpowered due to sparse observations (**Figure S17 & S18**). No individual low-frequency
448 variant (MAF > 0.001) was significantly associated with overall epilepsy or with any of the studied
449 epilepsy phenotypes (**Figure S19**). The primary gene-based test results and single variant
450 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 in both rare and common 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 significance in the non-hypothesis-driven analysis for GGE or NAFE, specific gene groups previously 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, $GABA_A$ receptor subunit genes, a larger group of genes delineating the GABAergic pathway, and all cation channel-encoding genes. Our results thus support the concept that defects in GABAergic inhibition underlie various forms of epilepsy. These findings, based on a more than 5-fold increase in sample size over previous exome-sequencing studies[17-19; 47], clearly reveal observations that have been hypothesized for common epilepsies 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 common epilepsies. 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 that 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[48].

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[17]. 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 or 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. Indeed, our results were consistent with the Epi4K sporadic NAFE cohort, where no signals of enrichment were observed[17; 49]. 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 the two common classes of epilepsy, 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[15], 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[28] ≥ 2 (2.0% of missense variants) were enriched in individuals with epilepsy at an effect size approaching PTVs in the

503 investigated gene groups. For common epilepsy types, the burden of these missense variants
504 (MPC≥2) was even more prominent than PTVs in known epilepsy genes and GABAergic genes
505 (**Figure 2**). At the gene level, some of the top channel genes (e.g. *GABRG2*, *CACNA1G*) carried
506 a higher number of missense variants (MPC≥2) than PTVs in people with epilepsy. For instance,
507 in the gene-based collapsing analysis considering all epilepsies, 15 *GABRG2* pathogenic variants
508 were found in epilepsy-affected individuals (including 7 GGE and 7 NAFE; **Tables S12, 14 & 16**)
509 versus only 1 pathogenic variant in controls; among the case-specific pathogenic variants, one
510 was a splice site mutation, while the other 14 were all missense variants (MPC≥2) (**Figure S20**),
511 linking to an impaired channel function. This is in line with findings from a recent exome-wide
512 study of 6,753 individuals with neurodevelopmental disorder with and without epilepsy[9] that
513 detected an association of missense *de novo* variants with the presence of epilepsy, particularly
514 when considering only ion channel genes. A disease-association of missense variants rather than
515 PTVs points to a pathophysiological mechanism of protein-alteration (e.g., gain-of-function or
516 dominant-negative effects) rather than haploinsufficiency, but ultimately only functional tests can
517 elucidate these mechanisms. A recent study on the molecular basis of 6 *de novo* missense
518 variants in *GABRG2* identified in DEE reported an overall reduced inhibitory function of *GABRG2*
519 due to decreased cell surface expression or GABA-evoked current amplitudes, suggesting
520 GABAergic disinhibition as the underlying mechanism[50]. Surprisingly, 2 of those recurrent *de novo*
521 missense variants were seen in two GGE-affected individuals in our study (A106T and R323Q),
522 and another recently reported variant in *GABRB2* (V316I) also occurred both *de novo* in DEE[51]
523 and as an inherited variant in a GGE family showing a loss of receptor function[18]. This suggests
524 that changes in protein function from the same missense pathogenic variant may cause not only
525 severe epilepsy syndromes, but also contribute to common epilepsies with milder presentations,
526 similar to what is known about variable expressivity in large families carrying *GABRG2* variants[48;
527 52-54]. Reduced receptor function due to *GABRG2* variants has been also shown for childhood
528 epilepsy with centrotemporal spikes previously[44; 54], which belong to the NAFE group in this study.

529 Moving forward, discovering how variant-specific perturbations of the neurotransmission and
530 signaling system in a gene can link to a spectrum of epilepsy syndromes will require in-depth
531 functional investigation.

532 Although we have increased the sample size from the Epi4K and EuroEPINOMICS WES
533 studies for both GGE and NAFE subgroups by more than 5-fold, the phenotypic and genetic
534 heterogeneity of common epilepsies—on par with other complex neurological and
535 neuropsychiatric conditions—will require many more samples to achieve statistical power for
536 identifying exome-wide significant genes. Furthermore, while we implemented stringent QC to
537 effectively control for the exome capture differences between cases and controls, this
538 concomitantly resulted in a loss of a substantial amount of the called sites and reduced our
539 detection power to identify associated variants. As sample sizes grow, the technical variation
540 across projects and sample collections will remain a challenge in large-scale sequencing studies
541 relying on a global collaborative effort.

542 With this largest epilepsy WES study to date, we demonstrated a strong replicability of
543 existing gene findings in an independent cohort. $GABA_A$ receptor genes affected by predicted-
544 pathogenic missense variants were enriched across the three subgroups of epilepsy. An ongoing
545 debate in epilepsy genetics is the degree to which generalized and focal epilepsies segregate
546 separately, and whether their genetic determinants are largely distinct or sometimes shared[4; 55].
547 Whilst clinical evidence for general separation of pathophysiological mechanisms in these two
548 forms is strong, and most monogenic epilepsy families segregate either generalized or focal
549 syndromes, the distinction is not absolute. Here, the finding of rare variants in $GABA_A$ receptor
550 genes in both forms adds weight to the case for shared genetic determinants.

551 Our results suggest that clinical presentations of common epilepsy types with complex
552 inheritance patterns have a combination of both common and rare genetic risk variants. The latest
553 ILAE epilepsy GWAS of over 15,000 patients and 25,000 controls identified 16 genome-wide
554 significant loci for common epilepsies[15], mapped these loci to ion channel genes, transcriptional

555 factors, and pyridoxine metabolism, and implicated a role in epigenetic regulation of gene
556 expression in the brain. A combination of rare and common genetic association studies with large
557 sample sizes, along with the growing evidence from studies of copy number variation and tandem
558 repeat expansions in epilepsy[12; 56; 57], will further decipher the genetic landscape of common
559 epilepsy subgroups. The ongoing effort of the Epi25 collaborative is expected to double the patient
560 cohorts in upcoming years with the goal of elucidating shared and distinct gene discoveries for
561 common and rare forms of epilepsy, ultimately facilitating precision medicine strategies in the
562 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, 20 figures and 18 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, Dennis Lal, Erin L. Heinzen, 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 Schreiber, Ingo Borggräfe, Christoph J. Schankin, Thomas

589 Mayer, Rudolf Korinthenberg, Knut Brockmann, Gerhard Kurlemann, Dieter Dennig, Rene
590 Madeleyn, Reetta Kälviäinen, Pia Auvinen, Anni Saarela, Tarja Linnankivi, Anna-Elina Lehesjoki,
591 Mark I. Rees, Seo-Kyung Chung, William O. Pickrell, Robert Powell, Natascha Schneider, Simona
592 Balestrini, Sara Zagaglia, Vera Braatz, Michael R. Johnson, Pauls Auce, Graeme J. Sills, Larry
593 W. Baum, Pak C. Sham, Stacey S. Cherny, Colin H.T. Lui, Nina Barišić, Norman Delanty, Colin
594 P. Doherty, Arif Shukralla, Mark McCormack, Hany El-Naggar, Laura Canafoglia, Silvana
595 Franceschetti, Barbara Castellotti, Tiziana Granata, Federico Zara, Michele Iacomino, Francesca
596 Madia, Maria Stella Vari, Maria Margherita Mancardi, Vincenzo Salpietro, Francesca Bisulli, Paolo
597 Tinuper, Laura Licchetta, Tommaso Pippucci, Carlotta Stipa, Raffaella Minardi, Antonio
598 Gambardella, Angelo Labate, Grazia Annesi, Lorella Manna, Monica Gagliardi, Elena Parrini,
599 Davide Mei, Annalisa Vetro, Claudia Bianchini, Martino Montomoli, Viola Doccini, Carla Marini,
600 Toshimitsu Suzuki, Yushi Inoue, Kazuhiro Yamakawa, Birute Tumiene, Lynette G. Sadleir,
601 Chontelle King, Emily Mountier, S. Hande Caglayan, Mutluay Arslan, Zuhal Yapıcı, Uluc Yis, Pınar
602 Topaloglu, Bulent Kara, Dilsad Turkdogan, Aslı Gundogdu-Eken, Nerses Bebek, Sibel Uğur-İşeri,
603 Betül Baykan, Barış Salman, Garen Haryanyan, Emrah Yücesan, Yeşim Kesim, Çiğdem Özkara,
604 Annapurna Poduri, Russell J. Buono, Thomas N. Ferraro, Michael R. Sperling, Warren Lo,
605 Michael Privitera, Jacqueline A. French, Steven Schachter, Ruben I. Kuzniecky, Orrin Devinsky,
606 Manu Hegde, Pouya Khankhanian, Katherine L. Helbig, Colin A. Ellis, Gianfranco Spalletta,
607 Fabrizio Piras, Federica Piras, Tommaso Gili, Valentina Ciullo, Andreas Reif, Andrew McQuillin,
608 Nick Bass, Andrew McIntosh, Douglas Blackwood, Mandy Johnstone, Aarno Palotie, Michele T.
609 Pato, Carlos N. Pato, Evelyn J. Bromet, Celia Barreto Carvalho, Eric D. Achtyes, Maria Helena
610 Azevedo, Roman Kotov, Douglas S. Lehrer, Dolores Malaspina, Stephen R. Marder, Helena
611 Medeiros, Christopher P. Morley, Diana O. Perkins, Janet L. Sobell, Peter F. Buckley, Fabio
612 Macciardi, Mark H. Rapaport, James A. Knowles, Genomic Psychiatry Cohort (GPC) Consortium,
613 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. 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.
2. Banerjee, P.N., Filippi, D., and Allen Hauser, W. (2009). The descriptive epidemiology of epilepsy-a review. Epilepsy Res 85, 31-45.
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. Epi4K Consortium. (2016). De Novo Mutations in SLC1A2 and CACNA1A Are Important Causes of Epileptic Encephalopathies. Am J Hum Genet 99, 287-298.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. Myers, C.T., and Mefford, H.C. (2015). Advancing epilepsy genetics in the genomic era. Genome Med 7, 91.
13. Jallon, P., and Latour, P. (2005). Epidemiology of idiopathic generalized epilepsies. Epilepsia 46 Suppl 9, 10-14.
14. 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.
15. The International League Against Epilepsy Consortium on Complex Epilepsies. (2018). Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. Nat Commun 9, 5269.

16. Helbig, I., and Lowenstein, D.H. (2013). Genetics of the epilepsies: where are we and where are we going? Curr Opin Neurol 26, 179-185.
17. 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.
18. 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.
19. Heinzen, E.L., Depondt, C., Cavalleri, G.L., Ruzzo, E.K., Walley, N.M., Need, A.C., Ge, D., He, M., Cirulli, E.T., Zhao, Q., et al. (2012). Exome sequencing followed by large-scale genotyping fails to identify single rare variants of large effect in idiopathic generalized epilepsy. Am J Hum Genet 91, 293-302.
20. Tan, N.C., and Berkovic, S.F. (2010). The Epilepsy Genetic Association Database (epiGAD): analysis of 165 genetic association studies, 1996-2008. Epilepsia 51, 686-689.
21. Cavalleri, G.L., Lynch, J.M., Depondt, C., Burley, M.W., Wood, N.W., Sisodiya, S.M., and Goldstein, D.B. (2005). Failure to replicate previously reported genetic associations with sporadic temporal lobe epilepsy: where to from here? Brain 128, 1832-1840.
22. Scheffer, I.E., Berkovic, S., Capovilla, G., Connolly, M.B., French, J., Guilhoto, L., Hirsch, E., Jain, S., Mathern, G.W., Moshe, S.L., et al. (2017). ILAE classification of the epilepsies: Position paper of the ILAE Commission for Classification and Terminology. Epilepsia 58, 512-521.
23. 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.
24. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 43, 11 10 11-33.
25. Li, H. (2014). Toward better understanding of artifacts in variant calling from high-coverage samples. Bioinformatics 30, 2843-2851.
26. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol 17, 122.
27. Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6, 80-92.
28. Samocha, K.E., Kosmicki, J.A., Karczewski, K.J., O'Donnell-Luria, A.H., Pierce-Hoffman, E., MacArthur, D.G., Neale, B.M., and Daly, M.J. (2017). Regional missense constraint improves variant deleteriousness prediction. bioRxiv.

29. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285-291.
30. Dewey, F.E., Murray, M.F., Overton, J.D., Habegger, L., Leader, J.B., Fetterolf, S.N., 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.
31. 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.
32. 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.
33. 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.
34. 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.
35. 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.
36. Hail, https://github.com/hail-is/hail.
37. 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.
38. 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.
39. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17, 405-424.
40. 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.
41. 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.

779 42. Lu, B., Zhang, Q., Wang, H., Wang, Y., Nakayama, M., and Ren, D. (2010). Extracellular calcium
780 controls background current and neuronal excitability via an UNC79-UNC80-NALCN cation
781 channel complex. Neuron 68, 488-499.
782 43. Shamseldin, H.E., Faqeih, E., Alasmari, A., Zaki, M.S., Gleeson, J.G., and Alkuraya, F.S. (2016).
783 Mutations in UNC80, Encoding Part of the UNC79-UNC80-NALCN Channel Complex, Cause
784 Autosomal-Recessive Severe Infantile Encephalopathy. Am J Hum Genet 98, 210-215.
785 44. Reinthaler, E.M., Dejanovic, B., Lal, D., Semtner, M., Merkler, Y., Reinhold, A., Pittrich, D.A.,
786 Hotzy, C., Feucht, M., Steinbock, H., et al. (2015). Rare variants in gamma-aminobutyric acid
787 type A receptor genes in rolandic epilepsy and related syndromes. Ann Neurol 77, 972-986.
788 45. Cheung, C.C., Yang, C., Berger, T., Zaugg, K., Reilly, P., Elia, A.J., Wakeham, A., You-Ten, A.,
789 Chang, N., Li, L., et al. (2010). Identification of BERP (brain-expressed RING finger protein)
790 as a p53 target gene that modulates seizure susceptibility through interacting with GABA(A)
791 receptors. Proc Natl Acad Sci U S A 107, 11883-11888.
792 46. Wang, J., Lin, Z.J., Liu, L., Xu, H.Q., Shi, Y.W., Yi, Y.H., He, N., and Liao, W.P. (2017). Epilepsy-
793 associated genes. Seizure 44, 11-20.
794 47. Klassen, T., Davis, C., Goldman, A., Burgess, D., Chen, T., Wheeler, D., McPherson, J.,
795 Bourquin, T., Lewis, L., Villasana, D., et al. (2011). Exome sequencing of ion channel genes
796 reveals complex profiles confounding personal risk assessment in epilepsy. Cell 145, 1036-
797 1048.
798 48. Oyrer, J., Maljevic, S., Scheffer, I.E., Berkovic, S.F., Petrou, S., and Reid, C.A. (2018). Ion
799 Channels in Genetic Epilepsy: From Genes and Mechanisms to Disease-Targeted
800 Therapies. Pharmacol Rev 70, 142-173.
801 49. Epi4K Consortium. (2012). Epi4K: gene discovery in 4,000 genomes. Epilepsia 53, 1457-1467.
802 50. Shen, D., Hernandez, C.C., Shen, W., Hu, N., Poduri, A., Shiedley, B., Rotenberg, A., Datta,
803 A.N., Leiz, S., Patzer, S., et al. (2017). De novo GABRG2 mutations associated with epileptic
804 encephalopathies. Brain 140, 49-67.
805 51. Hamdan, F.F., Myers, C.T., Cossette, P., Lemay, P., Spiegelman, D., Laporte, A.D., Nassif, C.,
806 Diallo, O., Monlong, J., Cadieux-Dion, M., et al. (2017). High Rate of Recurrent De Novo
807 Mutations in Developmental and Epileptic Encephalopathies. Am J Hum Genet 101, 664-
808 685.
809 52. Baulac, S., Huberfeld, G., Gourfinkel-An, I., Mitropoulou, G., Beranger, A., Prud'homme, J.F.,
810 Baulac, M., Brice, A., Bruzzone, R., and LeGuern, E. (2001). First genetic evidence of
811 GABA(A) receptor dysfunction in epilepsy: a mutation in the gamma2-subunit gene. Nat
812 Genet 28, 46-48.
813 53. Wallace, R.H., Marini, C., Petrou, S., Harkin, L.A., Bowser, D.N., Panchal, R.G., Williams, D.A.,
814 Sutherland, G.R., Mulley, J.C., Scheffer, I.E., et al. (2001). Mutant GABA(A) receptor
815 gamma2-subunit in childhood absence epilepsy and febrile seizures. Nat Genet 28, 49-52.
816 54. Maljevic, S., Moller, R.S., Reid, C.A., Pérez-Palma, E., Lal, D., May, P., and Lerche, H. (2018).
817 Spectrum of GABAA receptor variants in epilepsy. Curr Opin Neurol In press.

55. Peljto, A.L., Barker-Cummings, C., Vasoli, V.M., Leibson, C.L., Hauser, W.A., Buchhalter, J.R., and Ottman, R. (2014). Familial risk of epilepsy: a population-based study. Brain 137, 795-805.

56. 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.

57. 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,

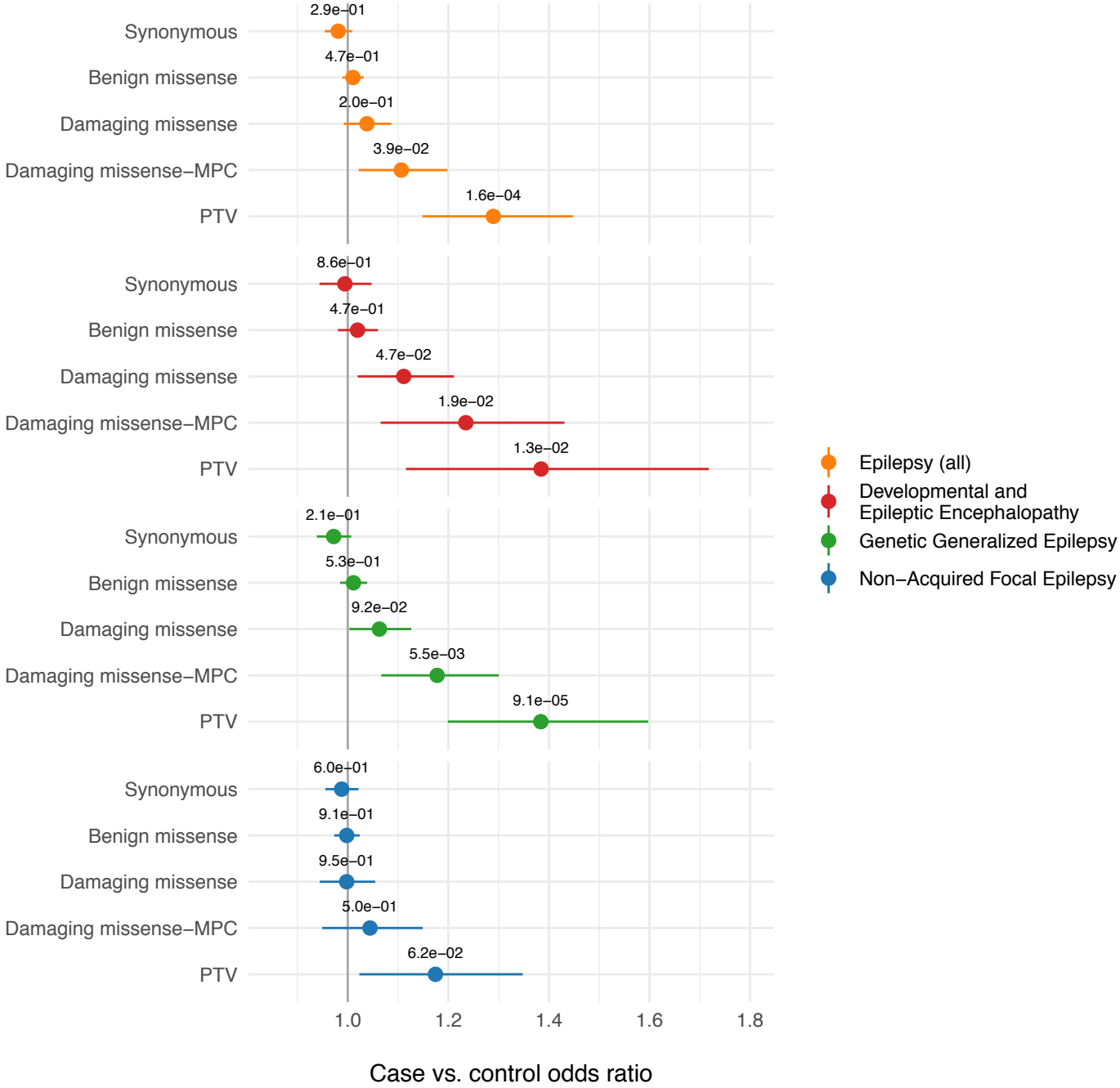
855 especially for $GABA_A$ receptor genes and genes involved in GABAergic pathways. Relative to
856 other epilepsy types, individuals with NAFE consistently showed the least burden of deleterious
857 URVs. No enrichment was observed from excitatory receptors.
858
859 **Figure 3. Gene burden for individuals diagnosed with (A) developmental and epileptic**
860 **encephalopathies, (B) genetic generalized epilepsy, or (C) non-acquired focal epilepsy**
861 This analysis focused on ultra-rare (non-DiscovEHR) singleton variants annotated as PTV,
862 damaging missense (MPC≥2), or in-frame insertion/deletion and used Fisher's exact test to
863 identify genes with a differential carrier rate of these ultra-rare deleterious variants in individuals
864 with epilepsy compared to controls. Exome-wide significance was defined as p-value < 6.8e-07
865 after Bonferroni correction (Methods). Only *SCN1A* achieved exome-wide significance for
866 individuals with DEE.

Table 1. Enrichment of ultra-rare protein-truncating or damaging missense (MPC≥2) singletons in epilepsy

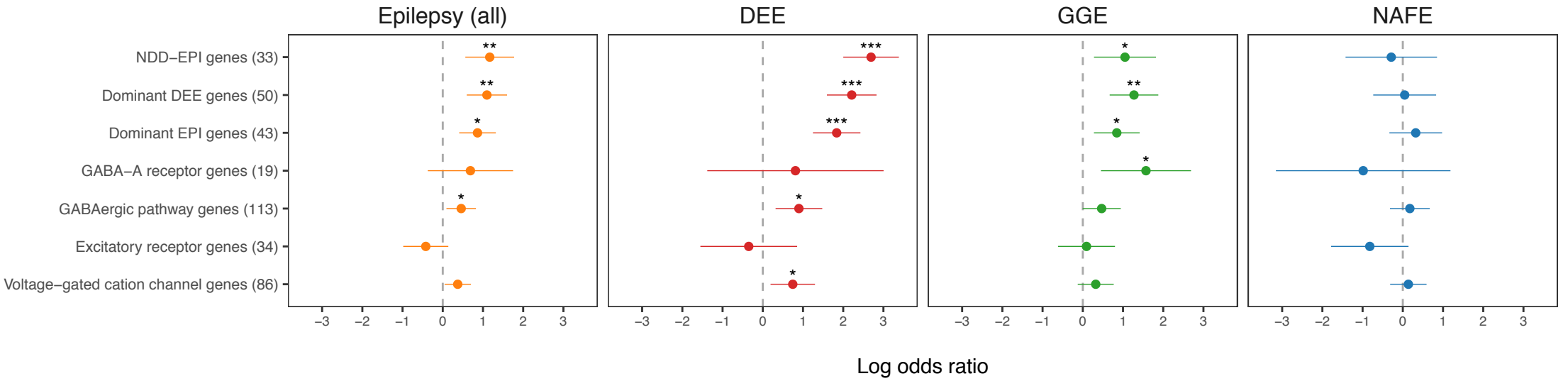
This analysis compared the burden of deleterious pathogenic variants between cases and controls using logistic regression, adjusting for sex, the first ten principal components, and overall variant count. FDR correction was based on a full list of burden tests in **Table S8**. Tested epilepsy types included 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

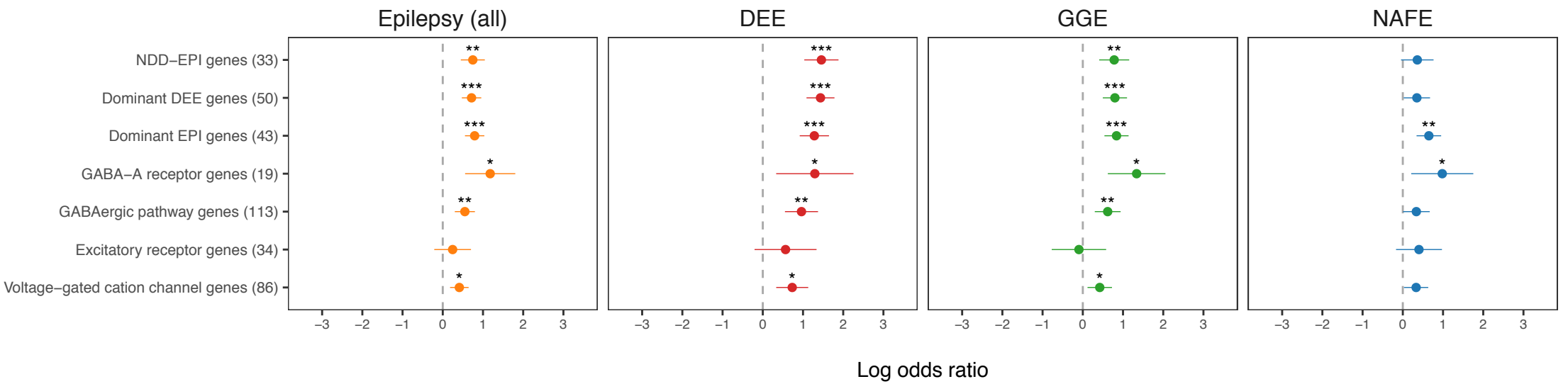
Figure 1.

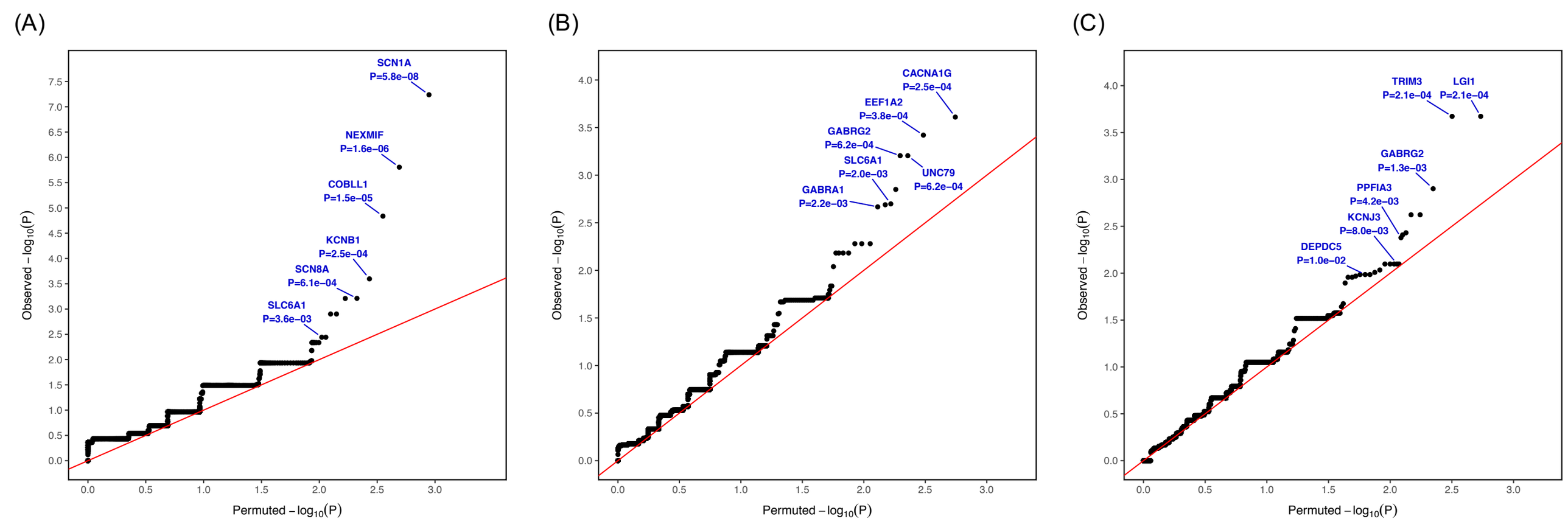

2.9e−01
Synonymous
4.7e−01
Benign missense
2.0e−01
Damaging missense
3.9e−02
Damaging missense−MPC
1.6e−04
PTV
8.6e−01
Synonymous
4.7e−01
Benign missense
4.7e−02
Damaging missense
1.9e−02
Damaging missense−MPC
1.3e−02
PTV
2.1e−01
Synonymous
5.3e−01
Benign missense
9.2e−02
Damaging missense
5.5e−03
Damaging 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 singleton PTVs



(B) Burden of ultra-rare singleton 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)