



## Research paper

## Distinct gene-set burden patterns underlie common generalized and focal epilepsies

Mahmoud Koko<sup>a</sup>, Roland Krause<sup>b</sup>, Thomas Sander<sup>c</sup>, Dheeraj Reddy Bobbili<sup>b</sup>, Michael Nothnagel<sup>d</sup>, Patrick May<sup>b,\*</sup>, Holger Lerche<sup>a,\*</sup>, Epi25 Collaborative<sup>2</sup>

<sup>a</sup> Department of Neurology and Epileptology, Hertie Institute for Clinical Brain Research, University of Tübingen, Otfried-Müller Str. 27, Tübingen 72076, Germany

<sup>b</sup> Luxembourg Centre for Systems Biomedicine, University of Luxembourg, 6 Avenue du Swing, Belvaux 4367, Luxembourg

<sup>c</sup> Cologne Center for Genomics, University of Cologne, Faculty of Medicine and Cologne University Hospital, Weyertal 115b, Cologne 50931, Germany

<sup>d</sup> Cologne Center for Genomics, University of Cologne, Faculty of Mathematics and Natural Sciences and Cologne University Hospital, Weyertal 115b, Cologne 50931, Germany

## ARTICLE INFO

## Article History:

Received 24 May 2021

Revised 4 September 2021

Accepted 6 September 2021

Available online 24 September 2021

## Keywords:

Burden analysis  
Ultra-rare variants  
Gene-sets  
Epilepsy  
Exome sequencing

## ABSTRACT

**Background:** Analyses of few gene-sets in epilepsy showed a potential to unravel key disease associations. We set out to investigate the burden of ultra-rare variants (URVs) in a comprehensive range of biologically informed gene-sets presumed to be implicated in epileptogenesis.

**Methods:** The burden of 12 URV types in 92 gene-sets was compared between cases and controls using whole exome sequencing data from individuals of European descent with developmental and epileptic encephalopathies (DEE,  $n = 1,003$ ), genetic generalized epilepsy (GGE,  $n = 3,064$ ), or non-acquired focal epilepsy (NAFE,  $n = 3,522$ ), collected by the Epi25 Collaborative, compared to 3,962 ancestry-matched controls.

**Findings:** Missense URVs in highly constrained regions were enriched in neuron-specific and developmental genes, whereas genes not expressed in brain were not affected. GGE featured a higher burden in gene-sets derived from inhibitory vs. excitatory neurons or associated receptors, whereas the opposite was found for NAFE, and DEE featured a burden in both. Top-ranked susceptibility genes from recent genome-wide association studies (GWAS) and gene-sets derived from generalized vs. focal epilepsies revealed specific enrichment patterns of URVs in GGE vs. NAFE.

**Interpretation:** Missense URVs affecting highly constrained sites differentially impact genes expressed in inhibitory vs. excitatory pathways in generalized vs. focal epilepsies. The excess of URVs in top-ranked GWAS risk-genes suggests a convergence of rare deleterious and common risk-variants in the pathogenesis of generalized and focal epilepsies.

**Funding:** DFG Research Unit FOR-2715 (Germany), FNR (Luxembourg), NHGRI (US), NHLBI (US), DAAD (Germany).

© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

## 1. Introduction

Dismantling the genetic architecture behind epilepsy is yet to be within reach in many individuals. The role of genetic causality is apparent in the developmental and epileptic encephalopathies (DEEs) [1–3], sometimes with consequences on precision treatments [4–7]. In contrast, only few individuals with familial or sporadic genetic generalized epilepsies (GGEs) or non-acquired focal

epilepsies (NAFEs) harbour monogenic causative variations [8–11]. Therefore, methodologies investigating the mutational burden of neurobiologically meaningful gene-sets improve the prospects to dissect the joint effects of multiple genetic factors underlying the complex genetic architecture of these common epilepsy syndromes. Such ‘gene-set analysis’ approaches are likely to provide valuable insights into the role of certain gene-sets and pathways in epilepsy. Recent gene-set burden analyses have shown an enrichment in ultra-rare deleterious and intolerant variants both in common and rare epilepsies in genes associated with dominant epilepsy syndromes, DEE genes, and neuro-developmental disorders (NDDs) with epilepsy genes, emphasizing a shared genetic component [8,11]. Evidence for the enrichment of rare missense variants in genes encoding GABA<sub>A</sub> receptors and GABAergic pathway genes in GGE pointed to the

\* Corresponding authors.

E-mail addresses: [patrick.may@uni.lu](mailto:patrick.may@uni.lu) (P. May), [holger.lerche@uni-tuebingen.de](mailto:holger.lerche@uni-tuebingen.de) (H. Lerche).

<sup>1</sup> These authors contributed equally to this work.

<sup>2</sup> Members of the Epi25 Collaborative are listed at the end of this article. See the Appendix for the affiliations.

## Research in context

### Evidence before this study

A systematic analysis of specific neuronal gene-sets underlying common generalized vs. focal epilepsies and developmental and epileptic encephalopathies (DEE) has not been performed to date. To evaluate the available evidence, we searched PubMed for articles published before 31.12.2019 using the English search terms “epilepsy AND gene-sets” or “epilepsy AND burden analysis”, identifying three studies with relevant positive findings. The first study examined generalized and focal epilepsies showing a preferential excess of ultra-rare variants in known disease genes in genetic generalized epilepsy (GGE) and non-acquired focal epilepsy (NAFE), and the second investigated generalized epilepsies highlighting the importance of GABA<sub>A</sub> (inhibitory) receptors in GGE. The third described multiple similarities between rare and common epilepsies in the pattern of rare genetic factors in known disease genes and genes that are most intolerant to genetic variation.

### Added value of this study

We performed an extensive analysis of biologically informed gene-sets in GGE and NAFE. Leveraging an array of metrics to enrich our analysis for pathogenic variants, we detected a substantial difference in the genetic burden between cases and controls in key neuronal gene-sets, including synaptic and developmental genes. We observed a relatively higher burden in inhibitory vs. excitatory neuronal gene-sets in generalized epilepsy but an increased burden in excitatory vs. inhibitory sets in focal epilepsies. Also, we found an excess of ultra-rare variants in generalized and focal epilepsy cases vs. controls in genes otherwise implicated by genome-wide association studies targeting common variants in the same epilepsy types, suggesting that rare and common variants work in concert to cause these common diseases. These novel results add a wider biological context to the previous findings, improving our understanding of the neuronal processes underlying seizure disorders.

### Implications of all the available evidence

Ultra-rare genetic variants in individuals with common epilepsies affect shared gene-sets, highlighting the importance of variation-intolerant sites, but also show specific enrichment patterns that suggest a central role for inhibitory and excitatory pathway defects in generalized and focal epilepsies, respectively, and a convergence of genetic risk caused by common and rare variants in both.

epilepsy [10]. We analyzed subjects from recruitment years 1 and 2 ( $n = 13,197$  before filtering) targeting individuals diagnosed with DEE ( $n = 1,474$ ), GGE ( $n = 4,510$ ), NAFE ( $n = 5,321$ ). The epilepsy classification, phenotyping and consent procedures have been previously described [10,11]. Five control cohorts [10,12] were available for this analysis ( $n = 13,299$ ), including Italian controls from the Epi25 Collaborative ( $n = 300$ ), the Swedish Schizophrenia Study controls ( $n = 6,242$ ), and three Myocardial Infarction Genetics (MIGen) Consortium cohorts: Leicester UK Heart Study ( $n = 1,165$ ), Ottawa Heart Study ( $n = 1,915$ ) and the Italian Atherosclerosis, Thrombosis, and Vascular Biology (ATVB) Study ( $n = 3,677$ ). The ethical approval and consents procedures for the individual cohorts were reported by the Epi25 Collaborative [10]. Subjects investigated by the Epi25 Collaborative provided signed informed consent at the participating centres according to local national ethical requirements and their standards at the time of collection. Approval for data reuse and analysis was obtained from the Epi25 Collaborative (cases) and dbGAP (controls). The data generation process has been previously described [10] (see the Appendix).

### 2.2. Quality control

We considered Non-Finnish European (NFE) individuals diagnosed with DEE, GGE, or NAFE. The ancestry was predicted based on 1000 Genomes data [13] using a Support Vector Machine, removing 1,911 individuals with epilepsy and 146 controls. The quality control procedures [14–22] aimed to ensure adequate case-control matching and minimize the coverage and call rate differences between cohorts. The final analysis set included 7,589 cases (DEE = 1,003, GGE = 3,064, NAFE = 3,522) and 3,962 matched controls (ATVB = 1,673, Leicester = 1,082, Ottawa = 924, Epi25 Italian = 283). The details are outlined in the supplemental methods (see the Appendix). The use of predominantly male or male-only control cohorts from ATVB and Leicester studies resulted in a misbalanced sample sex ratio (53.6% female cases vs. 19.4% female controls). The effect of this imbalance was addressed in a secondary analysis as will be detailed.

### 2.3. Qualifying variants

The variants were annotated using snpEff [23] v4.3 and Annovar [24] v20191024. We focused on URVs as these have shown a strong burden of deleterious pathogenic variants in multiple studies of epilepsy and other neurological disorders [8,10,11,25–29]. URVs were defined based on their Minor Allele Counts (MACs) in the study dataset (internal allele count/frequency) and their estimated frequency in the general population (external Minor Allele Frequencies, MAFs). Specifically, we examined variants that are: (i) Seen in less than three cases and controls ( $MAC \leq 3$ ); (ii) Not seen in DiscovEHR [30] (MAF in DiscovEHR = 0); (iii) Seen at a very low allele frequency in gnomAD [31] r2.1 database ( $MAF \text{ in gnomAD} \leq 2 \times 10^{-5}$ ). We performed three separate analyses for the three epilepsy phenotypes; Therefore, MACs were calculated independently in each analysis. This was intended to provide a better control for inflation compared to calculating MACs from all cases and controls. Accordingly, the reported variant counts in the control sets may differ slightly between the three analyses. Since our controls overlapped partially with gnomAD r2.1, we did not require complete absence of variants in gnomAD. URVs were categorised further into multiple classes based on their functional consequences and collapsed by gene as qualifying variants (QVs). We considered twelve non-synonymous variant classes including protein-truncating variants (presumed loss-of-function) and multiple groups of missense variants (mix of neutral, loss-, and gain-of-function mechanisms) as well as a (thirteenth) synonymous control classes of variants (presumed neutral). The grouping of missense QVs in multiple (partially overlapping) classes focused on three perspectives: conventional in-silico deleteriousness, constraint and paralog conservation. It was based on multiple

importance of the inhibitory pathway [9,10]. We used the large-scale dataset collected by the Epi25 Collaborative [10] for a comprehensive, exome-based case-control study to examine the burden of ultra-rare variants (URVs) in a large number of candidate gene-sets for three different epilepsy forms (DEE, GGE, NAFE), aiming to understand the specific roles of deleterious URVs in key pathways implicated in epileptogenesis. Focusing on regional constraint and paralog conservation, we identified relevant and specific gene-set associations in these three epilepsy forms.

## 2. Methods

### 2.1. Study samples

The Epi25 Collaborative collected and generated phenotyping and exome sequencing data from individuals with different subtypes of

predictions, namely, PolyPhen2 [32] (PPh2), Sorting Intolerant From Tolerant [33] (SIFT), Missense Badness Polyphen and Constraint [34] (MPC), Missense Tolerance Ratio [35] (MTR), Constrained Coding Regions [36] (CCR) and para-Z-score for paralog conservation [37,38]. The rationale behind the use of these scores is detailed in the supplemental methods (see the Appendix). The analyzed functional classes of variants (Table S6) were: (i) Benign missense variants: as predicted by PPh2 and SIFT. (ii) Damaging missense variants: as predicted by PPh2 and SIFT. (iii) Protein Truncating Variants (PTVs) that included stop-gained, start-lost, frameshift, splice-donor, and splice-acceptor variants. (iv) All functional variants combined: PTVs, in-frame indels, and damaging missense variants. (v) "MPC 1" missense variants: constrained missense with MPC score  $\geq 1$ . (vi) "MPC 2" missense variants: highly constrained missense with MPC score  $\geq 2$  (enriched for *de novo* variants). (vii) "MTR ClinVar" missense variants: constrained missense with MTR score  $\leq 0.825$  which is the median for ClinVar variants not denoted as *de novo*. (viii) "MTR De Novo" missense variants: highly constrained missense with MTR score  $\leq 0.565$  which is the median for ClinVar *de novo* variants. (ix) "CCR 80" missense variants: highly constrained missense variants in regions with CCR score  $\geq 80$ , with MPC score  $\geq 1$ , and MTR score  $\leq 0.825$ . (x) "paralog-non-conserved": missense variants located in sites not conserved across paralog genes as indicated by a para-Z-score  $\leq 0$ . (xi) "paralog-conserved": missense variants located in sites conserved across paralog genes as indicated

by a para-Z-score  $> 0$ . (xii) "paralog highly conserved": missense variants in highly conserved sites between paralog genes with para-Z-score  $\geq 1$ . (xiii) "Synonymous" variants that served as a control class for inflation.

#### 2.4. Gene-sets

92 gene-sets were tested. In addition to exome-wide burden testing (one gene-set of all protein coding genes), we defined additional 91 specific gene-sets as follows: (a) 34 sets based on gene expression patterns in the brain and genic intolerance [39–49]; (b) 28 functional groups including ion channels [8], GABA<sub>A</sub> receptors [9], excitatory receptors [9], GABAergic pathway [9], PSD-95 interactors [8], Gene Ontology (GO) gene-sets of GABAergic and glutamatergic synapses [41,42,50], neuronal pathways from Kyoto Encyclopedia of Genes and Genomes [51] (KEGG) and neuronal gene-sets from Reactome [52] database; (c) 14 gene-sets of known disease-related genes including monogenic epilepsy-causing genes [8–10], epilepsy genome-wide association study (GWAS) top-ranked genes from positional mapping (within a window of 250 kb around significant loci) plus mapping based on chromatin interaction (between gene promoters and the significant locus) [53], co-regulated genes in the brain [54,55]; and (d) 15 non-neuronal gene-sets [51]. The gene-sets are outlined in Table 1 (also see the Appendix).

**Table 1**  
**Gene-sets investigated in this study.** The number of gene-sets in each category is given in parenthesis.

#### Group of all protein coding genes (1):

-all genes annotated by snpEff as protein coding.

**Groups based on brain expression (34):** Expression in the brain, regional, cellular and sub-cellular expression patterns.

#### Brain-expressed LOF-intolerant genes:

excluding genes with no expression in the cortex/hippocampus  
-pLI  $> 0.995$ .  
-pLI 0.9–0.995.  
-pLI 0.8–0.9.

#### Brain-expressed missense-intolerant genes:

excluding genes with no expression in the cortex/hippocampus  
-Z-score  $> 3.09$ .  
-Z-score 2.5–3.09.  
-Z-score 2–2.5.

#### Cortical and hippocampal expression level:

-High, Moderate, Low in the cortex.  
-High, Moderate, Low in the hippocampus.

#### Brain development:

-Brain development genes (Gene-Ontology group).  
-Brain developmental genes (extended group).  
-Early developmental genes.  
-Late developmental genes.

#### Enrichment in the brain:

-Brain-enriched -Brain-enhanced.

#### Cell-type-specific enrichment:

-Neurons -glial cells  
-Excitatory neurons -Inhibitory neurons  
-Astrocytes -Microglia  
-Oligodendrocytes -Endothelium.  
**Neuronal Localization:**  
-Axon Initial Segment.  
-Synaptic (curated group).  
-Synaptic (extended group).  
-Synaptic vesicle and active zone.  
-Pre-synaptic.  
-Post-synaptic.  
-Pre-synaptic only.  
-Post-synaptic only.

**Functional gene-sets (28):** Ion channels, transporters, synaptic cycles, pathways and neurotransmitter cycles.

#### Ion channels, neurotransmitter receptors and related genes:

-Voltage-gated ion channels.  
-Voltage-gated cation channels.  
-Brain-specific voltage-gated ion channels.  
-GABA<sub>A</sub> receptors.  
-GABAergic pathway.  
-Excitatory receptors.  
-NMDAR & ARC interactors.  
-PSD-95 interactors.

#### GABAergic/Glutamatergic synapses (GO groups):

-GABAergic synapse -Glutamatergic synapse  
-only in GABAergic -only in glutamatergic  
-shared genes.

**Disease-associated and intolerant genes (14):** Genes and gene-sets with known associations with epilepsy and related neurological diseases

#### Monogenic disease-causing genes:

-Generalized epilepsy genes.  
-Focal epilepsy genes.  
-Dominant epilepsy genes  
-DEE genes.  
-NDD with epilepsy genes.  
-FMRP targets.  
-MGI seizure genes.

#### Control groups (15):

#### Genes not expressed in the brain:

-RNA not detected in cortex, in hippocampus, or all GTEx regions.  
-Protein is depleted in the brain.

#### GABAergic/Glutamatergic pathways (KEGG database):

-GABAergic pathway -Glutamatergic pathway  
-only in GABAergic -only in glutamatergic  
-shared genes.

#### Additional neuronal pathways (KEGG):

-Cholinergic pathway.  
-Dopaminergic pathway.  
-mTOR pathway.  
-Synaptic vesicle cycle.  
Glutamate release, uptake and clearance cycle.

#### GABA/glutamate cycles (Reactome database;

**pooled from multiple groups):**  
-GABA release, receptor activation, and clearance -  
**Additional neuronal groups (Reactome database):**  
-Presynaptic depolarization.  
-Neurexins and Neuroligins.  
-Synaptic Adhesion molecules.  
-Receptor-type Protein Tyrosine Phosphatases.

#### Top-ranking 100 genes in ILAE2 GWAS:

-Generalized epilepsy GWAS.  
-Focal epilepsy GWAS.  
-All epilepsies GWAS.

#### Brain co-expression module:

-Co-expressed module identified in non-diseased post-mortem brain tissues.  
(enriched for *de novo* variants in DEE).

#### Regulatory and co-expression modules in epilepsy:

-Co-expression network identified in brain tissues of Temporal Lobe Epilepsy patients  
-Two modules within this network.

#### KEGG metabolic pathways:

-Type II Diabetes.  
-Carbohydrate Absorption & Digestion.  
-Protein Absorption & Digestion.  
-Fat Absorption & Digestion.

#### KEGG cancer pathways:

-CA Breast, CA Lung, CA Colon, CA Prostate, Renal Cell Ca, CA Pancreas, Hepatocellular Ca.

## 2.5. Gene-set burden analysis

We examined the burden of QVs in thirteen variant classes (Table S6) for 92 gene-sets in three epilepsy phenotypes (DEE, GGE, and NAFE) against a set of matched controls. Gene-set burden testing was done using logistic regression by regressing the case-control status on the individual QVs counts. In each sample, URVs that matched the specific analysis criteria were collapsed by gene into QVs (each sample was assigned a status indicator: 1 for the presence of a QV or 0 for its absence) and these QVs were aggregated (summed per sample) across a target gene-set to get a burden score (assuming equal weights and direction of effects) which was used as a predictor in a binomial model while adjusting for additional covariates (sex, top ten principal components, exome-wide variant count, and exome wide singletons count) using *glm()* function from *stats* package [56]. Likelihood ratio test (LRT) from *lmtest* package [57] was used to compare a model with QVs burden and covariates as predictors against a null model (covariates only). Log-odds from LRT and their respective 95% confidence intervals and *p* values are presented here as a measure of enrichment in tested gene-sets. We employed a Benjamini-Hochberg false discovery rate (FDR) multiple testing adjustment for *p* values that accounted for 3,312 tests (92 gene-sets × 3 epilepsy phenotypes × 12 non-synonymous variant classes, excluding synonymous variants used as a control class) as implemented in *p.adjust()* function from *stats* package [56]. The cut-off for substantial enrichment was defined as FDR-adjusted *p* value < 0.05. For simplicity, *p* values (FDR-adjusted except for synonymous variants) are indicated throughout the presented plots using stars as follows: no star > 0.05, \* < 0.05, \*\* < 0.005, \*\*\* < 0.0005, \*\*\*\* < 0.00005. To estimate the extent of bias that might have been introduced by the imbalance in male-to-female ratios between cases and controls, we performed a secondary analysis excluding chromosome X genes (Table S11). Also, to ensure adequate control for any bias introduced by differences in capture kits, we performed another secondary analysis between two groups of control samples (Leicester study controls vs. Ottawa and ATVB controls) representing two main enrichment kits (1,100 samples enriched using Illumina ICE kits vs. 2,789 samples enriched using Agilent SureSelect kits). To ensure that the latter analysis would have sufficient power, it was coupled with an analysis of randomly selected individuals with GGE (*n* = 1,100) vs. controls (*n* = 2,789) using the CCR 80 class of variants. We did 500 permutations, taking the average odds, 2.5<sup>th</sup>/97.5<sup>th</sup> centiles of odds and average *p* values as an outcome. Lastly, to explore the extent of the observed differences between GGEs and NAFEs, we performed another limited secondary analysis comparing the CCR80 class of variants directly between individuals with GGE and NAFE. The statistical analysis was performed in R [56] v3.3.3. The analysis approach is outlined in Fig. S1 (see the Appendix). Details of the analysis methods, tables of tested gene-sets (Table S7) and genes in each set (Table S8) are provided in the Appendix.

## 2.6. Role of the funding source

The funding agencies had no role in study design; in the collection, analysis, and interpretation of data; and in the writing and the decision to submit the paper for publication. The authors did not receive payments from companies or other agencies to write this article, had full access to the data in the study and accept final responsibility to submit for publication.

## 3. Results

### 3.1. URVs excess in brain-expressed genes

First, we investigated the burden of URVs across all protein coding genes following the analysis approach outlined in Fig. S1. This revealed a clear enrichment in constrained missense variants that was maximum in consensus constrained coding regions predicted by Missense-badness

Polyphen and Constraint (MPC), Missense Tolerance Ratio (MTR) and Consensus Coding Regions (CCR) scores (Fig. 1). The combination of the three metrics identifies highly deleterious variants in functionally critical genic regions (see methods). In this particular analysis in all three phenotypes, about half of the cases, in contrast to roughly one-fourth of controls, harboured one or more QVs in highly constrained regions (Fig. S11). A previous similar analysis of this [10] and related [11] datasets examined loss-of-function intolerant genes and demonstrated an increased burden in ultra-rare constrained as well as protein truncating variants (PTVs). Here, the examination of brain-expressed intolerant genes showed, similarly, a marked enrichment in PTVs in addition to a burden in highly constrained missense variants that is comparable to what is seen exome-wide (Figs. 1 and S12).

When we examined protein coding genes grouped by their relative brain expression, damaging missense variants were only substantially enriched in genes highly expressed in the cortex or hippocampus, whereas those expressed at medium or low levels only showed an enrichment for the most constrained missense variants (Fig. 2). Genes with depleted expression in the brain did not show a substantial enrichment for any variant type (Fig. S18). Genes showing a higher expression in the adult brain compared to other tissues (brain-enriched & brain-enhanced) were also preferentially enriched, as well as genes associated with brain development. Genes related to late rather than early development showed a slightly higher enrichment in all three phenotypic groups (Fig. 2).

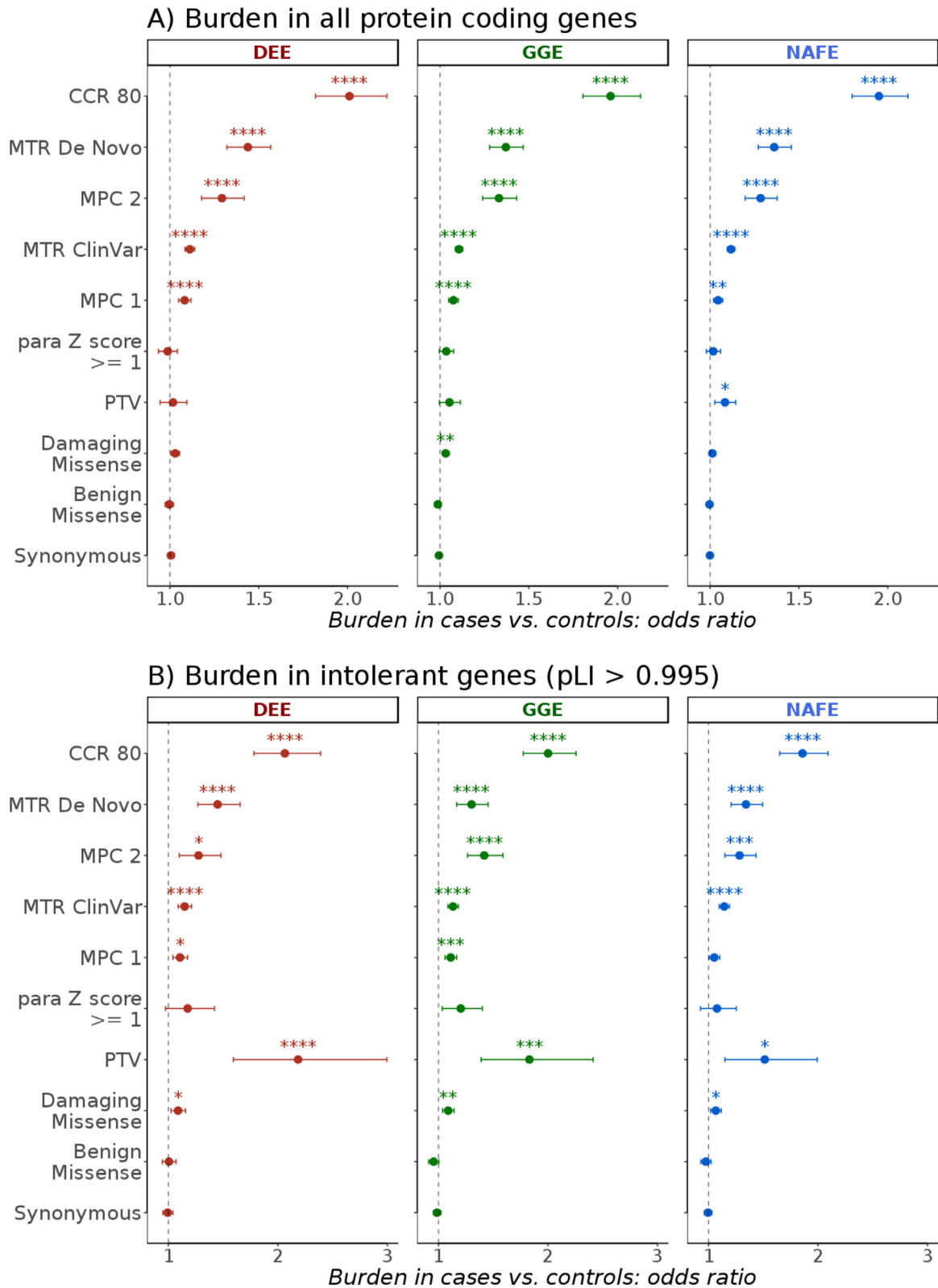
Focusing further on cell-type specific expression, neuron-specific genes were preferentially affected compared to those enriched in glial cells, particularly in GGE (Fig. 3). To obtain further insights into the nature of this neuronal enrichment, we used sets of genes representing paralogs of mouse genes found to be enriched in excitatory or inhibitory neurons (see the Appendix). Interestingly, genes preferentially expressed in inhibitory neurons showed an increased burden only in GGE, whereas those preferentially expressed in excitatory neurons showed a more prominent signal in NAFE. Since well-established epilepsy genes, like ion channels and receptors, show differential distributions in different neuronal compartments [58,59], we examined further sets of genes based on subcellular localization. We found that pre- and postsynaptic genes were enriched with variants in cases vs. controls, as well as a very small set of 17 genes located in axon initial segments (most prominent in DEE) (Fig. S16).

### 3.2. Burden of URVs in ion channel, neurotransmitter receptor encoding and related genes

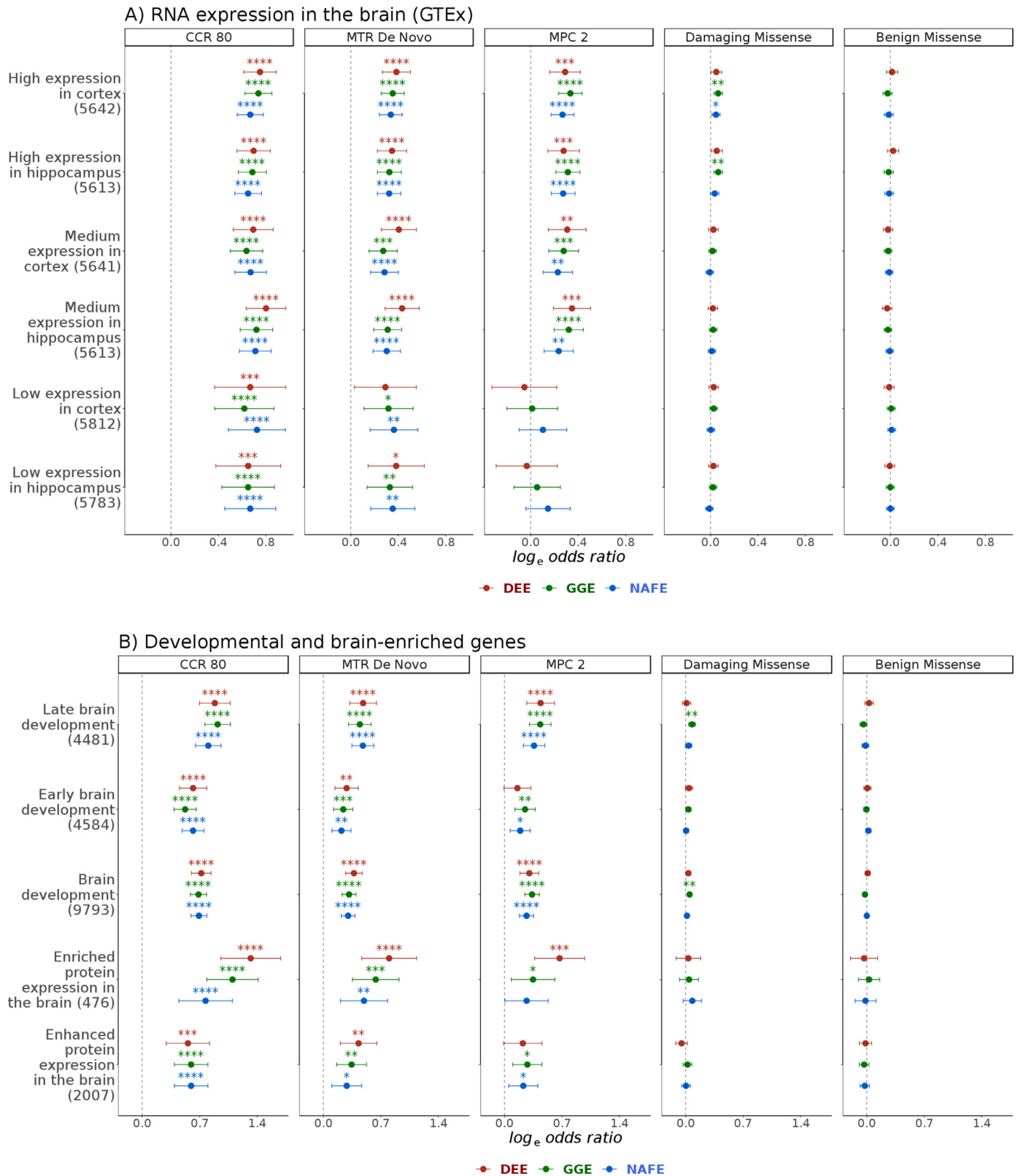
Next, we examined functional gene-sets that could, more specifically, underlie the observed enrichment in neuronal and synaptic genes. Ion channels, neurotransmitter receptors and transporters are widely implicated in epilepsy, especially in monogenic and familial forms, displaying considerable phenotypic heterogeneity and presenting as mild or severe epilepsies [60–62]. Variants in GABA<sub>A</sub> receptors were enriched in GGE but not in DEE or NAFE while those in gene-sets representing genes encoding N-Methyl-D-Aspartate receptor and Activity-Regulated Cytoskeleton protein [8] (NMDAR-ARC) interactors were enriched in NAFE and DEE. A comprehensive gene-set for the GABAergic pathway genes [9] showed a prominent signal in GGE and DEE, and less in NAFE. In contrast, a gene-set representing PSD-95 interactors showed comparable enrichment in NAFE and GGE (Fig. 3). Brain-expressed ion channels were found to be enriched for highly constrained missense variants (CCR 80 class of variants) in common as well as rare epilepsies (Fig. 3).

### 3.3. Patterns of burden in gene-sets representing inhibitory vs. excitatory signalling

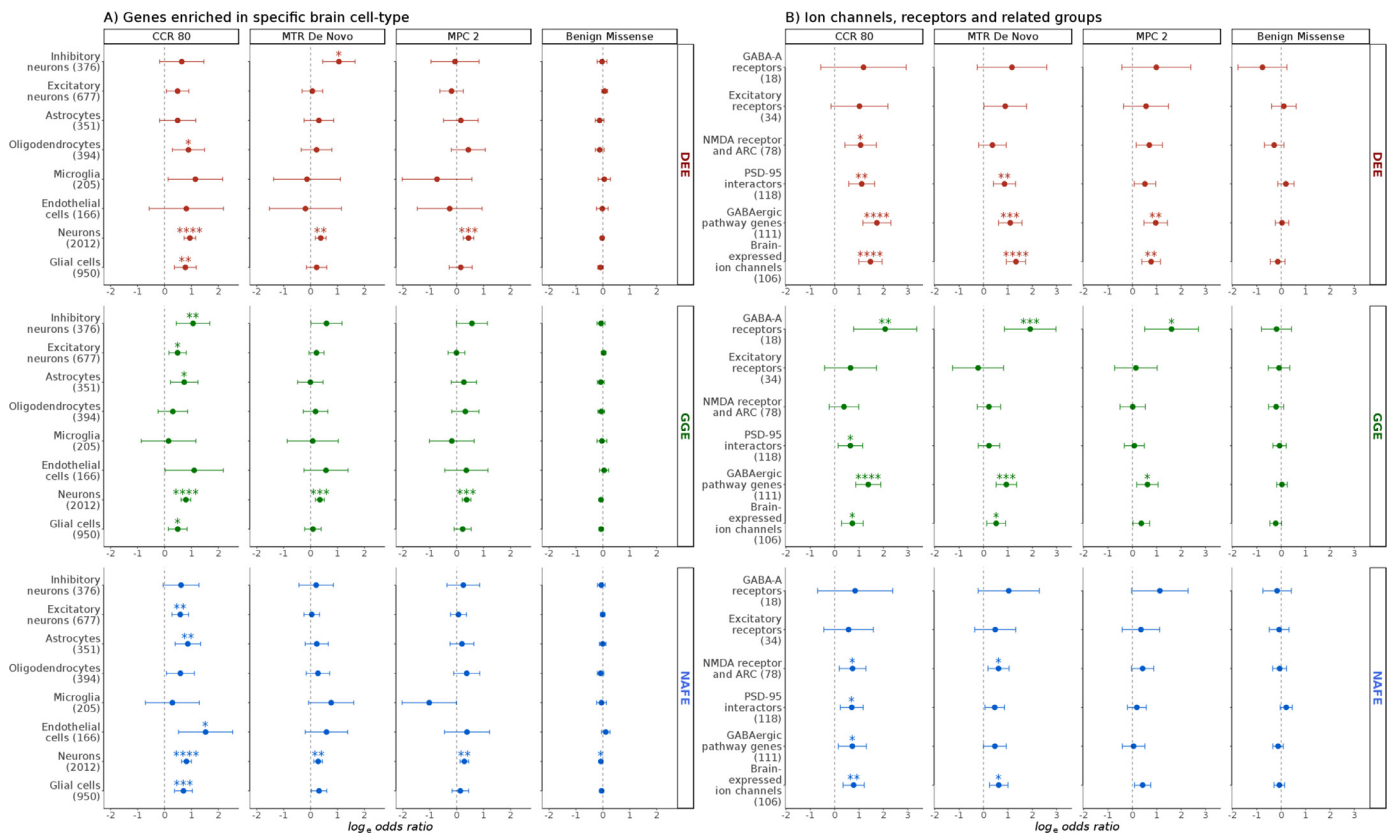
We then compared the patterns of URVs burden in genes involved in the GABAergic (main inhibitory) pathway and synapse against



**Fig. 1. Exome-wide burden of ultra-rare variants in the epilepsies.** The burden in developmental and epileptic encephalopathies (DEE), genetic generalized epilepsies (GGE) and non-acquired focal epilepsies (NAFE) in (A) 19,402 protein coding genes and (B) 1,743 genes with probability of loss-of-function intolerance (pLI) score > 0.995 is shown in multiple classes of variants (y-axis; see methods) as odds ratio (x-axis) from Likelihood Ratio Test (bars indicate 95% confidence intervals). False-Discovery-Rate-adjusted *p* values (synonymous variants analysis *p* values were not adjusted) are indicated with stars as follows: no star > 0.05, \* < 0.05, \*\* < 0.005, \*\*\* < 0.0005, \*\*\*\* < 0.00005.



**Fig. 2.** Burden of ultra-rare missense variants in brain expressed and developmental genes. The burden of benign or damaging missense variants and missense variants in highly constrained sites in developmental and epileptic encephalopathies (DEE), genetic generalized epilepsies (GGE) and non-acquired focal epilepsies (NAFE) is shown in gene-sets based on levels of RNA/protein expression in the cortex and hippocampus (A) or enrichment in adult or developing brain (B). Gene-sets are shown on the y-axis (number of genes in parenthesis). Log odds ratio (Likelihood Ratio Test) are shown on the x-axis (error bars indicate 95% confidence intervals). The variant classes are shown in vertical panels. False-Discovery-Rate-adjusted *p* values are indicated with stars as follows: no star > 0.05, \* < 0.05, \*\* < 0.005, \*\*\* < 0.0005, \*\*\*\* < 0.00005. High, medium and low expression categorization was based on expression levels in Gene Tissue Expression Project portal (GTEx). Brain-enriched (with more than four-fold expression in the brain compared to other tissues) and brain-enhanced genes (higher but less than four-fold expression) were obtained from the Human Protein Atlas.



**Fig. 3. Burden in neuronal and glial cells, ion channels, receptors and related interactors.** The burden in developmental and epileptic encephalopathies (DEE), genetic generalized epilepsies (GGE) and non-acquired focal epilepsies (NAFE) is shown on the x-axis (log-odds from Likelihood Ratio Test; error bars indicate 95% confidence intervals). Gene-sets are shown on the y-axis (number of genes in parenthesis). The variant classes are shown in vertical panels. False-Discovery-Rate-adjusted *p* values are indicated with stars as follows: no star > 0.05, \* < 0.05, \*\* < 0.005, \*\*\* < 0.0005, \*\*\*\* < 0.00005. (A) Burden in genes enriched in specific brain cells including neuron- or glia-enriched genes and their subtypes. (B) Burden in key biologically informed neuronal gene-sets with known or suspected relation to epilepsy. NMDA: N-Methyl-Dextro-Aspartate. ARC: neuronal activity-regulated cytoskeleton-associated protein (interactors). PSD-95: Post-Synaptic-Density protein 95 (interactors).

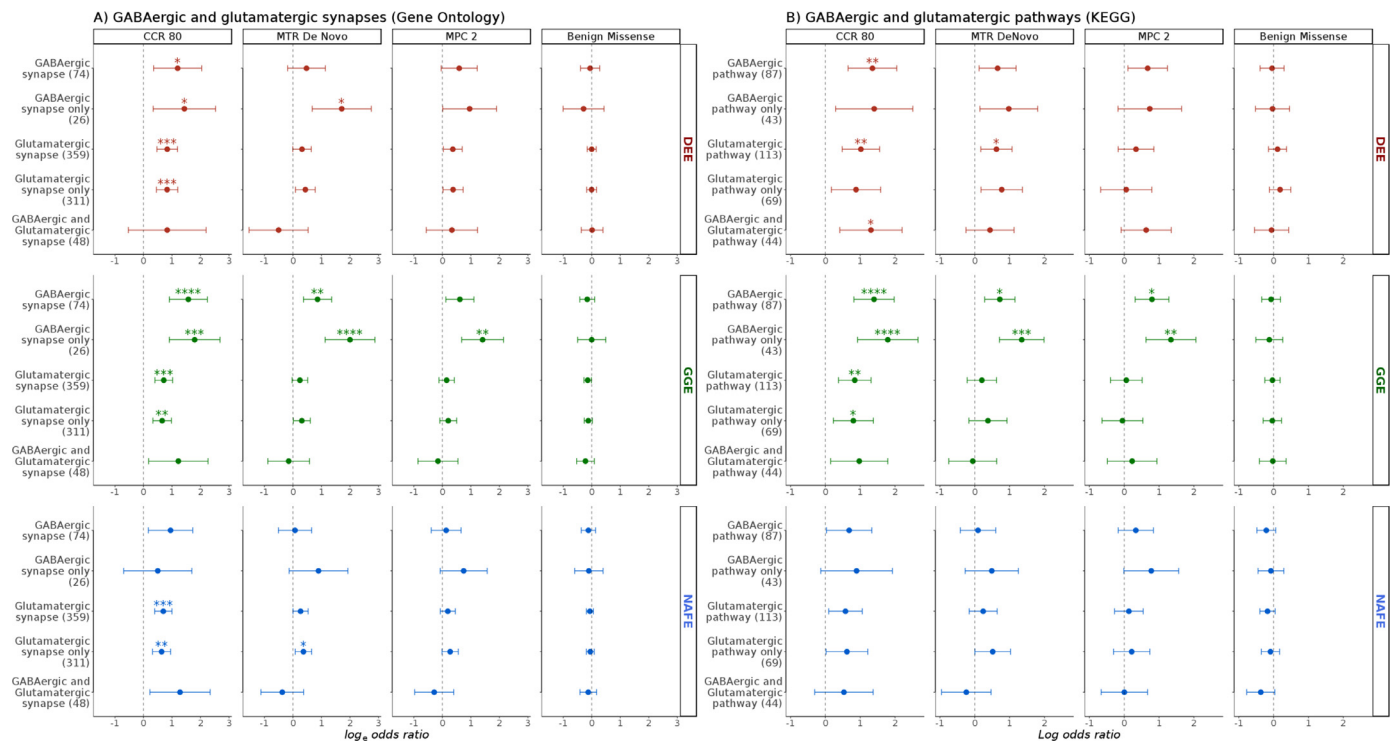
those in the glutamatergic (main excitatory) pathway and synapse in the brain, by examining their unique and overlapping genes based on KEGG pathways [51] or GO synaptic gene-sets [41] and sets of specific receptors (Fig. 4). GGE showed a higher burden in GABAergic vs. glutamatergic synapse (GO) and pathway (KEGG) genes, in genes encoding GABA<sub>A</sub> receptors vs. excitatory receptors/NMDAR-ARC interactors, and in GABAergic pathway genes (comprehensive gene-set) vs. genes encoding PSD-95 interactors, thus matching the higher burden in genes representing inhibitory vs. excitatory neuronal signalling. The CCR 80 analysis of GO gene-sets in NAFE showed a higher burden in glutamatergic vs. GABAergic synapse genes, akin to the pattern seen in genes enriched in excitatory vs. inhibitory neurons. The analysis of KEGG glutamatergic vs. GABAergic pathway genes did not confirm this finding (Fig. 4). It is notable that the overlap between GO synapse and KEGG pathway gene-sets is minimal, and the size of GO and KEGG gene-sets was comparable in GABAergic but discordant in glutamatergic genes (Fig. S21).

Altogether, these comparisons of the burden in missense variants in highly constrained sites between GGE and NAFE (Figs. 3 and 4) suggest the following patterns: (i) brain-expressed ion channels, genes enriched in excitatory neurons, enriched in astrocytes, PSD-95 interactors, GABAergic and glutamatergic synapse/pathway genes show an increased burden in cases vs. controls both in GGE & NAFE; (ii) in GGE, this enrichment is coupled with a stronger enrichment in inhibitory neuronal genes, in genes coding for GABA<sub>A</sub> receptors and in GABAergic synapse-specific genes (higher burden in inhibitory vs. excitatory gene-sets); and (iii) in NAFE, this is accompanied by an absence of enrichment in the later gene-sets and increased burden in the NMDAR-ARC interactors gene-set (higher burden in excitatory vs.

inhibitory gene-sets). A direct comparison of GGEs vs. NAFEs supported the observation of a substantially higher burden of highly constrained variants (CCR 80 class of missense variants) in GABAergic pathway genes in GGEs (Fig. S15).

### 3.4. Burden in gene-sets of known epilepsy-related genes

The previous Epi25 Collaborative analyses [10,11] demonstrated a high burden of missense variants in constrained (intolerant) sites in DEE, GGE, and NAFE, seen in dominant epilepsy genes, DEE genes, and NDD-Epilepsy genes. We observed similar enrichment patterns (Fig. 5) in MPC 2 and MTR De Novo classes of variants (enriched for *de novo* mutations). Limiting the analysis to highly constrained genic regions (CCR 80 class of variants) resulted in a marked increase in URVs burden, as was the trend in all the tested gene-sets so far. Testing these sets also unravelled strong enrichment in PTVs and missense variants in paralog-conserved sites. PTVs and missense variants in paralog-conserved sites did not show substantial enrichment in exome-wide analysis and most of other expression-based, localization-based or pathway-based gene-sets. However, we saw a modest increase in PTV burden in highly intolerant genes with probability of Loss-of-function Intolerance (pLI) > 0.995 in all epilepsies (Fig. 1). The choice of the pLI score cut-off was based on the outcomes of a previous analysis [10] which demonstrated that the burden in PTVs in genes with pLI > 0.9 is driven primarily by genes with pLI > 0.995 rather 0.9–0.995, a pattern that we were able to reproduce (Fig. S12). In a gene-set of known DEE genes, where highly intolerant genes are rather prevalent, we saw a prominent enrichment in PTVs burden in DEE. Also, there was an increased burden in missense



**Fig. 4. Enrichment in major neuronal synapses and pathways.** Panels show comparison of enrichment patterns in developmental and epileptic encephalopathies (DEE), genetic generalized epilepsies (GGE) and non-acquired focal epilepsies (NAFE) in GABAergic and glutamatergic synapses and pathway genes based on (A) Gene-Ontology (GO) and (B) Kyoto Encyclopedia of Genes and Genomes (KEGG). The burden is shown on the x-axis (log-odds from Likelihood Ratio Test; error bars indicate 95% confidence intervals). Gene-sets are shown on the y-axis (number of genes in parenthesis). The variant classes are shown in vertical panels. False-Discovery-Rate-adjusted  $p$  values are indicated with stars as follows: no star  $> 0.05$ , \*  $< 0.05$ , \*\*  $< 0.005$ , \*\*\*  $< 0.0005$ , \*\*\*\*  $< 0.00005$ . Complete groups, genes specific to one of the two synapses/pathways as well as their intersection were tested.

variants in paralogue-conserved sites in sets of epilepsy-related disease genes (DEE genes, dominant Epilepsy genes, NDD-Epilepsy genes). This burden was very strong in DEE but not as remarkable in GGE and NAFE (Fig. 5).

### 3.5. Enrichment in top GWAS hits captures divergence between common epilepsies

Recent efforts from the International League Against Epilepsy (ILAE) consortium on complex epilepsies identified multiple associations in a large GWAS of common epilepsies [53]. To examine the hypothesis that genes located near the top GWAS hits are also affected by rare variants, we tested the enrichment in sets of the 100 top-ranked genes derived from the ILAE GWAS in generalized, focal, and all epilepsies. Interestingly, when limiting the analysis to Consensus Coding Regions (CCR 80 class of variants), top-ranked genes derived from the GWAS of either generalized or focal epilepsies were preferentially enriched for rare variants in the respective phenotypic groups of GGE and NAFE (Fig. 6). Although the observed enrichment was rather subtle, this result was corroborated by a similar pattern in two, rather small, sets of known epilepsy genes that are predominantly associated with either generalized or focal epilepsy [9].

### 3.6. Brain- and epilepsy-related co-expression modules

We also aimed to touch upon the role of brain co-expression modules identified in post-mortem brain tissues from healthy individuals [55] and to contrast these to the networks and modules identified in brain tissue derived from epilepsy patients [54]. A brain expression module was found to be substantially enriched for rare deleterious variants in an independent cohort of DEE [55]. A link to common epilepsy phenotypes was also inferred, but a burden in URVs was not examined so far. This module showed a non-specific enrichment in all three epilepsy subtypes with highest odds in DEE. It is noteworthy

that this module overlaps largely with known epilepsy genes (Fig. S22). In resected hippocampi of individuals with temporal lobe epilepsy (TLE), Johnson and colleagues identified two co-expression modules within a gene-regulatory transcriptional network [54]. A subtle enrichment was seen in these modules in DEE and GGE, but not NAFE (Fig. 6).

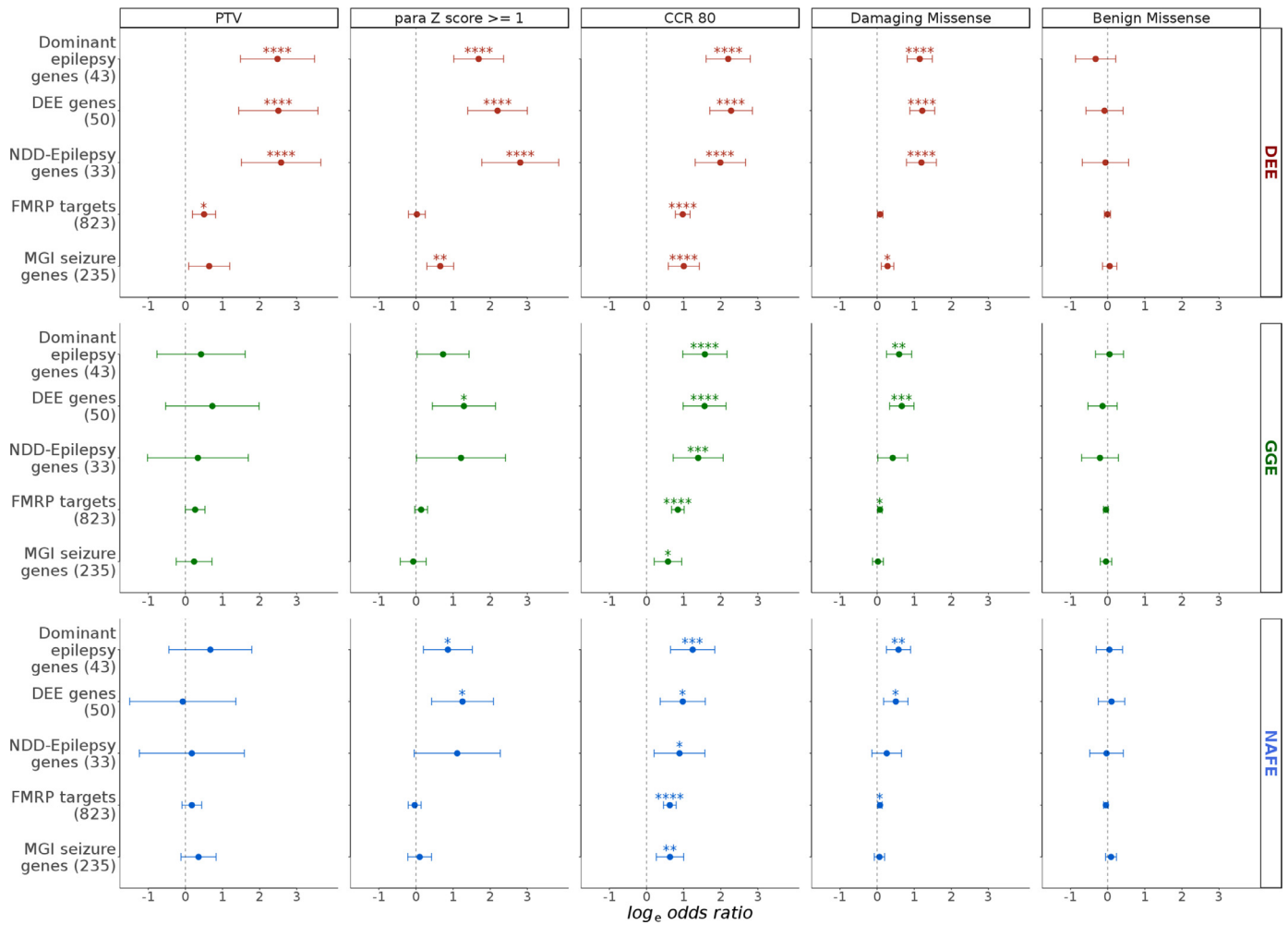
### 3.7. Additional neuronal and non-neuronal pathways

Other neuronal gene-sets were enriched in our analysis. Genes encoding neurexins and neuroligins, important elements of pre- and post-synaptic interaction promoting adhesion between dendrites and axons [63], were enriched in DEE (Fig. S17). Also, the synaptic vesicle cycle pathway (KEGG) showed a prominent signal in both DEE and GGE. We also examined the burden in the mTOR pathway (KEGG), hypothesizing that it could have potential relevance to focal epilepsies, but did not detect a substantial enrichment (Fig. S17). Interestingly, NAFE analysis displayed a burden in endothelial and astrocyte-specific genes in highly constrained genic regions (Fig. 3). Detailed results from all tested classes including the counts of genes with observed QVs, variant counts in cases and controls, logistic regression odds of the individual QVs burden in cases vs. controls and related  $p$  values are provided in Table S9.

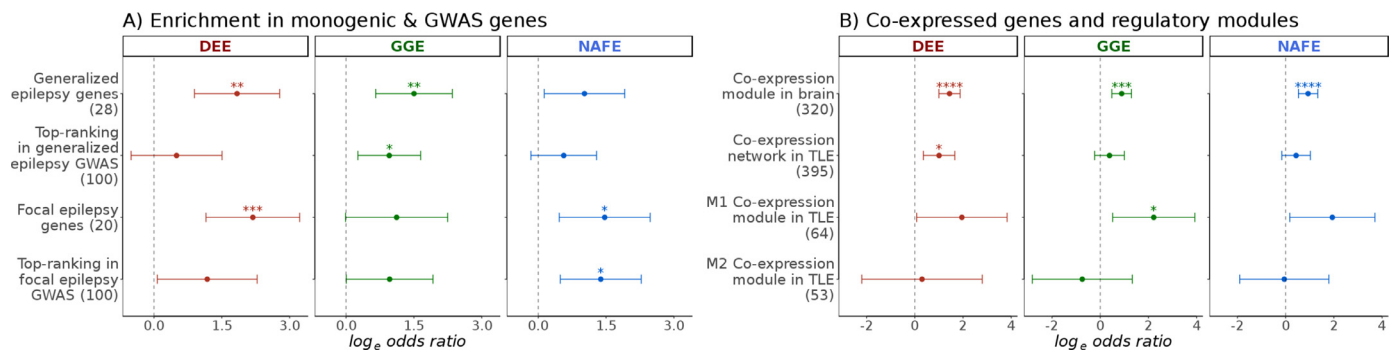
### 3.8. Specificity of the observed enrichment patterns

Examination of control classes and control gene-sets that are not expected to show an enrichment supported the validity of our analysis. Four sets of genes not expressed in the brain that were tested (high confidence genes with depleted RNA and protein expression in the brain, genes with no RNA detected in the cortex, the hippocampus or any brain tissue) were not substantially enriched in most of the tested variant classes (Fig. S18). Also, we examined eleven metabolic and cancer pathways (KEGG) to have additional insights into





**Fig. 5. Burden of ultra-rare variants in groups of epilepsy-related known disease genes.** The burden in five gene-sets (y-axis; number of genes in parenthesis) in developmental and epileptic encephalopathies (DEE), genetic generalized epilepsies (GGE) and non-acquired focal epilepsies (NAFE) (horizontal panel) in selected variant classes (vertical panels) is shown on the x-axis (log odds ratios from Likelihood Ratio Test; error bars indicate 95% confidence intervals). False-Discovery-Rate-adjusted *p* values are indicated with stars as follows: no star > 0.05, \* < 0.05, \*\* < 0.005, \*\*\* < 0.0005, \*\*\*\* < 0.00005. NDD-Epilepsy: neurodevelopmental disorders with epilepsy. FMRP: Fragile-X Mental Retardation Protein. MGI: Mouse Genome Informatics database.



**Fig. 6. Risk elements in GWAS top-ranked genes and co-expression modules.** The burden of missense variants in highly constrained sites (log-odds on the x-axis; error bars indicate 95% confidence intervals) in developmental and epileptic encephalopathies (DEE), genetic generalized epilepsies (GGE) and non-acquired focal epilepsies (NAFE) is shown in gene-sets (y-axis; number of genes in parenthesis) representing (A) Generalized or Focal epilepsy (presumed monogenic) genes as well as top-ranked 100 genes from GWAS of generalized and focal epilepsies, and (B) co-expressed genes identified in post-mortem brain tissues of healthy individuals (module of 320 genes) or in brain tissues from TLE patients (network of 395 genes) as well as two sub-modules of this network (M1 and M2). False-Discovery-Rate-adjusted *p* values are indicated with stars as follows: no star > 0.05, \* < 0.05, \*\* < 0.005, \*\*\* < 0.0005, \*\*\*\* < 0.00005.

the specificity of the observed signals to neuronal processes and genes. Among 540 tests targeting functional variants in these non-neuronal gene-sets (3 epilepsy subtypes, 15 sets representing genes not expressed in the brain, KEGG metabolic and cancer pathways, 12 non-synonymous functional classes of variants), 18 tests (3.3%) had

an FDR-adjusted *p* values < 0.05. At least for some of those, the enrichment could be explained by an overlap with genes known to play a role in epilepsy. For instance, genes forming the Type II Diabetes KEGG pathway are substantially enriched in DEE (FDR-adjusted *p* values of 0.007 for MTR DeNovo and 0.01 for CCR 80 class of

variants). This pathway contains two genes that are known to cause DEE, namely, *CACNA1A* [64] and *CACNA1E* [65]. The enrichment was no longer prominent ( $p$  values  $> 0.05$ ) after the removal of these two genes (Fig. S23).

### 3.9. Bias and inflation in gene-set burden testing

The analysis for synonymous variants did not show more substantial enrichment than expected by chance, indicating sufficient control for inflation, particularly in exome-wide models and gene-sets with large number of genes. In this control analysis (synonymous variants), few tests showed  $p$  values  $< 0.05$  (15 out of 276 tests of 92 gene-sets and 3 phenotypes: 5.4%). The analysis for benign missense variants, another class that is not expected to show an increased burden in cases vs. controls [10], did not show substantial enrichment as well. Nine out of 276 tests for benign missense variants (3.2%) showed  $p$  values  $< 0.05$  (only 2 with FDR-adjusted  $p$  values  $< 0.05$ ). Possible alternative explanations for such subtle signals include residual population stratification, differences in exome capture not adjusted by covariates and the presence of synonymous variants with functional consequences [66]. However, these proportions are close to the limit expected by chance under a true null hypothesis (5% with  $\alpha = 0.05$ ). A potential source of bias in our burden testing was the imbalance in male-to-female ratios between cases and controls (Table S4). We provide results from a secondary analysis that excluded all genes located on chromosome X, which shows that any bias not captured by the inclusion of sample sex as a covariate is likely marginal (Table S11). To exclude any major residual stratification resulting from the use of different enrichment kits, we additionally performed a controls-only analysis in which we compared control samples enriched with Illumina ICE capture kits (from Leicester study) to controls enriched using Agilent SureSelect kits (ATVB study and Ottawa study). This analysis reflected a good control for any potential bias introduced by different exome capture systems and also demonstrated that the mixing of controls included (Leicester and Ottawa) or not included (ATVB) in gnomAD is unlikely to have affected our main outcomes (Table S11).

## 4. Discussion

By analyzing the sequencing data of 11,551 unrelated European individuals (1,003 individuals with DEE, 3,064 individuals with GGE, and 3,522 individuals with NAFE vs. 3,962 controls), we show an increased burden in ultra-rare missense variants in highly constrained sites in epilepsy cases compared to controls, not only in intolerant and known epilepsy-related genes, as previously shown [10,11], but also exome-wide in all protein coding genes. Similar to the observations made in several other phenotypes, the burden in PTVs was most prominent in known disease genes and brain-expressed loss-of-function intolerant genes [31,39,67]. Consistent with their enrichment in neurodevelopmental disorders [37], the burden in missense variants in paralog-conserved sites was prominent in DEEs. The lower burden of these variants in GGEs and NAFEs may reflect a true disparity between rare and common epilepsies. The presented results are also consistent with previous analyses of missense variants in a small number of gene-sets examined in similar cohorts [8–11].

The systematic analysis of additional gene-sets and a wider variety of classes of variants revealed interesting findings about the neurobiology of distinct types of epilepsy. Although associated with higher odds ratios of an epilepsy phenotype, enriched variants are not deterministic on their own, since about one-fourth of the controls also carry qualifying variants in the CCR 80 analysis (Fig. S11). As such, the phenotype is determined by a constellation of other factors, possibly including the severity of variants [11], patterns of multiple variations, oligogenic contribution from rare variants [68], and

polygenic risk from common variants [69]. Developmental genes were key drivers in all epilepsies suggesting that the impairment of developmental processes is not limited to DEEs with marked developmental deficits [70,71]. The enrichment in synaptic genes is another shared feature between the epilepsies that has also been observed in neurodevelopmental disorders with epilepsy [72,73], schizophrenia [25], and autism [74]. This highlights a shared genetic architecture not only between epilepsy subtypes but also with other related neurological disorders, as has been shown previously for common variants [75].

Despite the common genetic and phenotypic features, DEEs, GGEs and NAFEs represent well-recognized phenotypic clusters with defined electro-encephalographic and clinical characteristics. Given the phenotypic severity of DEEs, the prevalence of *de novo* variants and monogenic cases in DEE (those with pathogenic and likely pathogenic variants in known monogenic genes), and the description of phenotypic spectra for genes involved in DEE that also span the milder GGE or NAFE, the distinction between severe and mild epilepsies could be attributed, at least to some extent, to the severity of the genetic defects, their functional effects or their localization within certain channel regions [11,62,76–79]. The distinction between GGE and NAFE, however, is probably functional, at least in part, as suggested by previous work demonstrating the centrality of GABAergic genes in generalized epilepsies [9,10]. Also, it is well recognized that few genes present with focal epilepsy and are not linked generalized epilepsy syndromes [80]. Here, phenotype-specific patterns in gene-sets representing neuronal inhibitory vs. excitatory signalling were observed in comparisons of GGE and NAFE.

Additional disparities in key gene-sets (genes implicated in monogenic generalized & focal epilepsy, the 100 top-ranked genes associated with GWAS hits in generalized & focal epilepsy) point to a possible genetic-functional divergence, so that a common background of shared risk seems to be overlaid by specific risk entities. The enrichment of rare variants in GWAS genes also supports the convergence of ultra-rare and common variants in conferring epilepsy risk, in concordance with the observed enrichment of epilepsy GWAS hits for monogenic epilepsy genes [53]. According to our findings, a link between common and rare variants is likely to be also relevant for the phenotypic heterogeneity observed in seizure disorders. Notably, polygenic risk scores also pointed out the specificity of the risk profiles in common epilepsies [69]. Based on previous findings of an increased URV burden in DEEs [55] and the current findings in GGEs and NAFEs, it is also conceivable that differentially expressed genes in individuals with epilepsy, representing closely orchestrated networks with possible functional correlations, would highlight modules in which altered transcription, URVs, or both contribute to cause both rare and common epilepsies.

The associations presented in this work should be interpreted with the caveats of gene-set testing in mind [81]. Pathways and molecular processes are not consistently defined in different resources (Fig. S21). These differences may explain the discrepancies in enrichment patterns in the same pathway. We examined multiple overlapping gene-sets from different sources to corroborate the findings that underscore a genuine biological relevance. Our analysis has additional limitations which we aimed to overcome using stringent analysis and quality control strategies. The limited use of about half of the controls from the primary analysis affected the overall power. Nevertheless, we were able to reproduce most of the major signals from gene-sets with large effect sizes, the latter thereby acting as positive controls. Multiple secondary analyses suggested that the impalpable of male-to-female ratios in our case and control sets and the use of sequencing data from ExAC [82], gnomAD [31] or DiscovEHR [30] to develop, train or validate *in-silico* algorithms used for estimating constraint [34–36] do not seem to have introduced a substantial bias (Table S11). The overlap between the controls used in this study and gnomAD controls (Table S2) created some challenges

in defining URVs. For population frequency filtering, we allowed around five alleles in gnomAD (allele frequency of  $2 \times 10^{-5}$ ) to retain URVs from our control that are also seen in gnomAD while still filtering common variants and prevalent sequencing artifacts.

In conclusion, missense URVs affecting constrained sites in brain-expressed genes show distinct signatures in epilepsy. Enrichment patterns of URVs-affected genes suggest a preferential involvement of inhibitory genes in GGE and excitatory genes in focal epilepsies. Genes implicated by common GWAS variants may also be disrupted by URVs in various epilepsy phenotypes, suggesting a convergence of rare disruptive variants, and common variants in the pathogenesis of epilepsy.

**Online appendix:** Supplementary materials including supplemental methods, Tables S1–S6, Figs. S1–S20, and affiliations of the Epi25 Collaborative members accompany the online version of this article. Additional large supplemental Tables S7–S12 are available on Mendeley Data.

### Contributors

HL, PM, MK conceived the study. HL, PM, RK, and MK contributed to the study design and/or data acquisition. Authors from the Epi25 Collaborative recruited and phenotyped patients and/or contributed to data generation as indicated in the supplemental author information (see the [Appendix](#)). MK and PM performed the analysis and verified the underlying data with input from HL, RK, TS, DRB and MN. MK, PM, and HL interpreted the results and wrote the manuscript with input from TS. All authors reviewed the manuscript for intellectual content and approved the final version of the manuscript. PM and HL contributed equally as PIs.

### Epi25 collaborative

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, Wolkong, 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, Mayer, Rudolf, Korinthenberg, Knut, Brockmann, Gerhard, Kurlemann, Dieter, Dennig, Rene, Madeleyn, Reetta, Kälviäinen, Pia, Auvinen, Anni, Saarela, Tarja, Linnankivi, Anna-Elina, Lehesjoki, Mark, I, Rees, Seo-Kyung, Chung, William, O, Pickrell, Robert, Powell, Natascha, Schneider, Simona, Balestrini, Sara, Zagaglia, Vera, Braatz, Michael, R, Johnson, Pauls,

Auce, Graeme, J, Sills, Larry, W, Baum, Pak, C, Sham, Stacey, S, Cherny, Colin, H T, Lui, Nina, Barišić, Norman, Delanty, Colin, P, Doherty, Arif, Shukralla, Mark, McCormack, Hany, El-Naggar, Laura, Canafoglia, Silvana, Franceschetti, Barbara, Castellotti, Tiziana, Granata, Federico, Zara, Michele, Iacomino, Francesca, Madia, Maria, Stella, Vari, Maria, Margherita, Mancardi, Vincenzo, Salpietro, Francesca, Bisulli, Paolo, Tinuper, Laura, Licchetta, Tommaso, Pippucci, Carlotta, Stipa, Raffaella, Minardi, Antonio, Gambardella, Angelo, Labate, Grazia, Annesi, Lorella, Manna, Monica, Gagliardi, Elena, Parrini, Davide, Mei, Annalisa, Vetro, Claudia, Bianchini, Martino, Montomoli, Viola, Doccini, Carla, Marini, Toshimitsu, Suzuki, Yushi, Inoue, Kazuhiro, Yamakawa, Birute, Tumiene, Lynette, G, Sadleir, Chontelle, King, Emily, Mountier, Hande, S, Caglayan, Mutluay, Arslan, Zuhaf, Yapıcı, Uluc, Yis, Pinar, Topaloglu, Bulent, Kara, Dilsad, Turkdogan, Aslı, Gundogdu-Eken, Nerses, Bebek, Sibel, Uğur-İşeri, Betül, Baykan, Barış, Salman, Garen, Haryanyan, Emrah, Yücesan, Yeşim, Kesim, Çiğdem, Özkara, Annapurna, Poduri, Beth, R, Shiedley, Catherine, Shain, Russell, J, Buono, Thomas, N, Ferraro, Michael, R, Sperling, Warren, Lo, Michael, Privitera, Jacqueline, A, French, Steven, Schachter, Ruben, I, Kuzniecky, Orrin, Devinsky, Manu, Hegde, Pouya, Khankhanian, Katherine, L, Helbig, Colin, A, Ellis, Gianfranco, Spalletta, Fabrizio, Piras, Federica, Piras, Tommaso, Gili, Valentina, Ciullo, Andreas, Reif, Andrew, McQuillin, Nick, Bass, Andrew, McIntosh, Douglas, Blackwood, Mandy, Johnstone, Aarno, Palotie, Michele, T, Pato, Carlos, N, Pato, Evelyn, J, Bromet, Celia, Barreto, Carvalho, Eric, D, Achtyes, Maria, Helena, Azevedo, Roman, Kotov, Douglas, S, Lehrer, Dolores, Malaspina, Stephen, R, Marder, Helena, Medeiros, Christopher, P, Morley, Diana, O, Perkins, Janet, L, Sobell, Peter, F, Buckley, Fabio, Macciardi, Mark, H, Rapaport, James, A, Knowles, Genomic Psychiatry Cohort (GPC), 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, Benjamin, M, Neale. See the [Appendix](#) for the affiliations of the Epi25 Collaborative members.

### Data availability

The data/analyses presented in the current publication are based on the use of study data from the Epi25 Collaborative (<http://epi-25.org/>), available with controlled access through dbGaP (<https://ncbi.nlm.nih.gov/gap/>) and AnVIL project (<https://anvilproject.org/data>) with the accession number [phs001489](#), and control exome data from dbGAP that are accessible with appropriate permissions under accession numbers [phs000473](#) (Swedish Schizophrenia Study), [phs001000](#) (Leicester UK Heart Study), [phs000806](#) (Ottawa Hear Study), and [phs000814](#) (Italian Atherosclerosis, Thrombosis, and Vascular Study). Supplemental data supporting the analyses presented are available on Mendeley Data (<https://doi.org/10.17632/nmmz4wjvxxk.1>).

### Declaration of Competing Interest

M. Koko reports grants from DAAD, during the conduct of the study; Dr. R. Krause reports grants from FNR, during the conduct of the study; Dr. med. habil. T. Sander reports grants from DFG, during the conduct of the study; Dr. D. R. Bobbili has nothing to disclose; Prof. Dr. M. Nothnagel reports grants from DFG, during the conduct of the study; Dr. P. May reports grants from FNR, during the conduct of the study; Prof. Dr. med. H. Lerche reports grants from DFG, during the conduct of the study.

### Acknowledgments

We are cordially thankful to the Epi25 Collaborative staff and all patients participating in the Epi25 project, who all made this study possible. RK, TS, MN, PM and HL (and IH and YGW from the Epi25 Collaborative) were supported by the Research Unit FOR-2715 of the

German Research Foundation and the Fond Nationale de la Recherche of Luxembourg (DFG/FNR, grants INTER/DFG/17/11583046, Le1030/16-1, Sa434/6-1, No755/6-1, He5415/7-1, We4896/4-1). The Epi25 project 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-funded Epi25 research activities at the Broad Institute, including genomic data generation in the Broad Genomics Platform, are supported by NHGRI grant UM1 HG008895 (M.J.D.). The Genome Sequencing Program efforts were also supported by NHGRI grant 5U01HG009088-02. MK was supported by the German Academic Exchange Service (DAAD personal funding program Number 57214224). PM obtained additional FNR funding as part of the National Centre of Excellence in Research on Parkinson's disease (NCER-PD, FNR11264123). Data used for the analyses described in this manuscript were obtained in part from the GTEx Portal (dbGaP accession number phs000424.v8.p2) on 30/11/2020. We acknowledge support by the FNR Open Access Fund.

## Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.ebiom.2021.103588](https://doi.org/10.1016/j.ebiom.2021.103588).

## References

- McTague A, Howell KB, Cross JH, Kurian MA, Scheffer IE. The genetic landscape of the epileptic encephalopathies of infancy and childhood. *Lancet Neurol* 2016;15(3):304–16. doi: [10.1016/S1474-4422\(15\)00250-1](https://doi.org/10.1016/S1474-4422(15)00250-1).
- Happ HC, Carvill GL. A 2020 view on the genetics of developmental and epileptic encephalopathies. *Epilepsy Curr* 2020;20(2):90–6. doi: [10.1177/1535759720906118](https://doi.org/10.1177/1535759720906118).
- Hebbar M, Mefford HC. Recent advances in epilepsy genomics and genetic testing. *F1000Res* 2020;9:185. doi: [10.12688/f1000research.21366.1](https://doi.org/10.12688/f1000research.21366.1).
- EpiPM Consortium. A roadmap for precision medicine in the epilepsies. *Lancet Neurol* 2015;14(12):1219–28. doi: [10.1016/S1474-4422\(15\)00199-4](https://doi.org/10.1016/S1474-4422(15)00199-4).
- Striano P, Minassian BA. From genetic testing to precision medicine in epilepsy. *Neurotherapeutics* 2020;17(2):609–15. doi: [10.1007/s13311-020-00835-4](https://doi.org/10.1007/s13311-020-00835-4).
- Sisodiya SM. Precision medicine and therapies of the future. *Epilepsia* 2021;62(S2). doi: [10.1111/epi.16539](https://doi.org/10.1111/epi.16539).
- Lerche H. Drug-resistant epilepsy—time to target mechanisms. *Nat Rev Neurol* 2020;16(11):595–6. doi: [10.1038/s41582-020-00419-y](https://doi.org/10.1038/s41582-020-00419-y).
- Epi4K Consortium. Epilepsy Phenome/Genome Project. Ultra-rare genetic variation in common epilepsies: a case-control sequencing study. *Lancet Neurol* 2017;16(2):135–43. doi: [10.1016/S1474-4422\(16\)30359-3](https://doi.org/10.1016/S1474-4422(16)30359-3).
- May P, Girard S, Harrer M, Bobbili DR, Schubert J, Wolkong S, et al. Rare coding variants in genes encoding GABAA receptors in genetic generalised epilepsies: an exome-based case-control study. *Lancet Neurol* 2018;17(8):699–708. doi: [10.1016/S1474-4422\(18\)30215-1](https://doi.org/10.1016/S1474-4422(18)30215-1).
- Epi25 Collaborative. Ultra-rare genetic variation in the epilepsies: a whole-exome sequencing study of 17,606 individuals. *Am J Hum Genet* 2019;105(2):267–82. doi: [10.1016/j.ajhg.2019.05.020](https://doi.org/10.1016/j.ajhg.2019.05.020).
- Epi25 Collaborative. Sub-genic intolerance, clinvar, and the epilepsies: a whole-exome sequencing study of 29,165 individuals. *Am J Hum Genet* 2021;108(6):965–82. doi: [10.1016/j.ajhg.2021.04.009](https://doi.org/10.1016/j.ajhg.2021.04.009).
- Tryka KA, Hao L, Sturcke A, Jin Y, Wang ZY, Ziyabari L, et al. NCBI's database of genotypes and phenotypes: dbGaP. *Nucleic Acids Res* 2014;42(D1):D975–9. doi: [10.1093/nar/gkt1211](https://doi.org/10.1093/nar/gkt1211).
- The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015;526(7571):68–74. doi: [10.1038/nature15339](https://doi.org/10.1038/nature15339).
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20(9):1297–303. doi: [10.1101/gr.107524.110](https://doi.org/10.1101/gr.107524.110).
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 2015;4(1):7. doi: [10.1186/s13742-015-0047-8](https://doi.org/10.1186/s13742-015-0047-8).
- Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics* 2010;26(22):2867–73. doi: [10.1093/bioinformatics/btq559](https://doi.org/10.1093/bioinformatics/btq559).
- Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* 2019;47(D1):D766–73. doi: [10.1093/nar/gky955](https://doi.org/10.1093/nar/gky955).
- Krusche P, Trigg L, Boutros PC, Mason CE, De La, De La Vega FM, Moore BL, et al. Best practices for benchmarking germline small-variant calls in human genomes. *Nat Biotechnol* 2019;37(5):555–60. doi: [10.1038/s41587-019-0054-x](https://doi.org/10.1038/s41587-019-0054-x).
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25(16):2078–9. doi: [10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352).
- Tan A, Abecasis GR, Kang HM. Unified representation of genetic variants. *Bioinformatics* 2015;31(13):2202–4. doi: [10.1093/bioinformatics/btv112](https://doi.org/10.1093/bioinformatics/btv112).
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38(8):904–9. doi: [10.1038/ng1847](https://doi.org/10.1038/ng1847).
- Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet* 2006;2(12):e190. doi: [10.1371/journal.pgen.0020190](https://doi.org/10.1371/journal.pgen.0020190).
- Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, snpEFF. *Fly* 2012;6(2):80–92. doi: [10.4161/fly.19695](https://doi.org/10.4161/fly.19695).
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38(16):e164. doi: [10.1093/nar/gkq603](https://doi.org/10.1093/nar/gkq603).
- Genovese G, Fromer M, Stahl EA, Ruderfer DM, Chambert K, Landén M, et al. Increased burden of ultra-rare protein-altering variants among 4877 individuals with schizophrenia. *Nat Neurosci* 2016;19(11):1433–41. doi: [10.1038/nn.4402](https://doi.org/10.1038/nn.4402).
- Demontis D, Satterstrom K, Duan J, Lescai F, Dinesen Ostergaard S, Lesch K-P, et al. The role of ultra-rare coding variants in ADHD. *Eur Neuropsychopharmacol* 2019;29:S724–5. doi: [10.1016/j.euroneuro.2017.06.042](https://doi.org/10.1016/j.euroneuro.2017.06.042).
- Wilfert AB, Turner TN, Murali SC, Hsieh P, Sulovari A, Wang T, et al. Recent ultra-rare inherited variants implicate new autism candidate risk genes. *Nat Genet* 2021;53(8):1125–34. doi: [10.1038/s41588-021-00899-8](https://doi.org/10.1038/s41588-021-00899-8).
- Singh T, Walters JTR, Johnstone M, Curtis D, Torniainen M, Rees E, et al. The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. *Nat Genet* 2017;49(8):1167–73. doi: [10.1038/ng.3903](https://doi.org/10.1038/ng.3903).
- Bennett CA, Petrovski S, Oliver KL, Berkovic SF. ExACTly zero or once: a clinically helpful guide to assessing genetic variants in mild epilepsies. *Neurol Genet* 2017;3(4):e163. doi: [10.1212/NXG.000000000000163](https://doi.org/10.1212/NXG.000000000000163).
- Dewey FE, Murray MF, Overton JD, Habegger L, Leader JB, Fetterolf SN, et al. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science* 2016;354(6319):aaf6814. doi: [10.1126/science.aaf6814](https://doi.org/10.1126/science.aaf6814).
- Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;581(7809):434–43. doi: [10.1038/s41586-020-2308-7](https://doi.org/10.1038/s41586-020-2308-7).
- Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7(4):248–9. doi: [10.1038/nmeth0410-248](https://doi.org/10.1038/nmeth0410-248).
- Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* 2012;40(W1):W452–7. doi: [10.1093/nar/gks539](https://doi.org/10.1093/nar/gks539).
- Samocha KE, Kosmicki JA, Karczewski KJ, O'Donnell-Luria AH, Pierce-Hoffman E, MacArthur DG, et al. Regional missense constraint improves variant deleteriousness prediction. *bioRxiv* 2017. doi: [10.1101/148353](https://doi.org/10.1101/148353).
- Traynelis J, Silk M, Wang Q, Berkovic SF, Liu L, Ascher DB, et al. Optimizing genomic medicine in epilepsy through a gene-customized approach to missense variant interpretation. *Genome Res* 2017;27(10):1715–29. doi: [10.1101/gr.226589.117](https://doi.org/10.1101/gr.226589.117).
- Havrilla JM, Pedersen BS, Layer RM, Quinlan AR. A map of constrained coding regions in the human genome. *Nat Genet* 2019;51(1):88–95. doi: [10.1038/s41588-018-0294-6](https://doi.org/10.1038/s41588-018-0294-6).
- Lal D, May P, Perez-Palma E, Samocha KE, Kosmicki JA, Robinson EB, et al. Gene family information facilitates variant interpretation and identification of disease-associated genes in neurodevelopmental disorders. *Genome Med* 2020;12(1):28. doi: [10.1186/s13073-020-00725-6](https://doi.org/10.1186/s13073-020-00725-6).
- Dickerson JE, Robertson DL. On the origins of mendelian disease genes in man: the impact of gene duplication. *Mol Biol Evol* 2012;29(1):61–9. doi: [10.1093/molbev/msr111](https://doi.org/10.1093/molbev/msr111).
- Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, et al. A framework for the interpretation of de novo mutation in human disease. *Nat Genet* 2014;46(9):944–50. doi: [10.1038/ng.3050](https://doi.org/10.1038/ng.3050).
- Genotype Tissue Expression Project portal. Available from: <https://www.gtexportal.org>. Accessed 2020 Dec 1.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. *Nat Genet* 2000;25(1):25–9. doi: [10.1038/75556](https://doi.org/10.1038/75556).
- The Gene Ontology Consortium. The gene ontology resource: enriching a GOLD mine. *Nucleic Acids Res* 2021;49(D1):D325–34. doi: [10.1093/nar/gkaa1113](https://doi.org/10.1093/nar/gkaa1113).
- Cardoso-Moreira M, Halbert J, Valloton D, Velten B, Chen C, Shao Y, et al. Gene expression across mammalian organ development. *Nature* 2019;571(7766):505–9. doi: [10.1038/s41586-019-1338-5](https://doi.org/10.1038/s41586-019-1338-5).
- Human Protein Atlas. Available from: <https://www.proteinatlas.org>. Accessed 2020 Dec 1.
- Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. *Science* 2015;347(6220). doi: [10.1126/science.1260419](https://doi.org/10.1126/science.1260419).
- Dusart P, Hallström BM, Renné T, Odeberg J, Uhlén M, Butler LM. A Systems-based map of human brain cell-type enriched genes and malignancy-associated endothelial changes. *Cell Rep* 2019;29(6):1690–706. doi: [10.1016/j.celrep.2019.09.088](https://doi.org/10.1016/j.celrep.2019.09.088).
- Mo A, Mukamel EA, Davis FP, Luo C, Henry GL, Picard S, et al. Epigenomic signatures of neuronal diversity in the mammalian brain. *Neuron* 2015;86(6):1369–84. doi: [10.1016/j.neuron.2015.05.018](https://doi.org/10.1016/j.neuron.2015.05.018).

- [48] Pirooznia M, Wang T, Avramopoulos D, Valle D, Thomas G, Hugarir RL, et al. SynaptomeDB: an ontology-based knowledgebase for synaptic genes. *Bioinformatics* 2012;28(6):897–9. doi: [10.1093/bioinformatics/bts040](https://doi.org/10.1093/bioinformatics/bts040).
- [49] Koopmans F, van Nierop P, Andres-Alonso M, Byrnes A, Cijssouw T, Coba MP, et al. SynGO: an evidence-based, expert-curated knowledge base for the synapse. *Neuron* 2019;103(2):217–34. doi: [10.1016/j.neuron.2019.05.002](https://doi.org/10.1016/j.neuron.2019.05.002).
- [50] Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database hallmark gene set collection. *Cell Syst* 2015;1(6):417–25. doi: [10.1016/j.cels.2015.12.004](https://doi.org/10.1016/j.cels.2015.12.004).
- [51] Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res* 2021;49(D1):D545–51. doi: [10.1093/nar/gkaa970](https://doi.org/10.1093/nar/gkaa970).
- [52] Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The reactome pathway knowledgebase. *Nucleic Acids Res* 2019;48(D1):D498–503. doi: [10.1093/nar/gkz1031/5613674](https://doi.org/10.1093/nar/gkz1031/5613674).
- [53] The International League Against Epilepsy Consortium on Complex Epilepsies. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. *Nat Commun* 2018;9(1):5269. doi: [10.1038/s41467-018-07524-z](https://doi.org/10.1038/s41467-018-07524-z).
- [54] Johnson MR, Behmoaras J, Bottolo L, Krishnan ML, Pernhorst K, Santoscoy PLM, et al. Systems genetics identifies *Sestrin 3* as a regulator of a proconvulsant gene network in human epileptic hippocampus. *Nat Commun* 2015;6(1):6031. doi: [10.1038/ncomms7031](https://doi.org/10.1038/ncomms7031).
- [55] Delahaye-Duriez A, Srivastava P, Shkura K, Langley SR, Laaniste L, Moreno-Moral A, et al. Rare and common epilepsies converge on a shared gene regulatory network providing opportunities for novel antiepileptic drug discovery. *Genome Biol* 2016;17(1):245. doi: [10.1186/s13059-016-1097-7](https://doi.org/10.1186/s13059-016-1097-7).
- [56] R Core Team. R: a language and environment for statistical computing. The Comprehensive R Archive Network 2017. Available from: <https://www.R-project.org/>
- [57] Zeileis A, Hothorn T. Diagnostic checking in regression relationships. *Rnews* 2002. Available from: <https://cran.r-project.org/web/packages/lmtest/vignettes/lmtest-intro.pdf>
- [58] Vacher H, Mohapatra DP, Trimmer JS. Localization and targeting of voltage-dependent ion channels in mammalian central neurons. *Physiol Rev* 2008;88(4):1407–47. doi: [10.1152/physrev.00002.2008](https://doi.org/10.1152/physrev.00002.2008).
- [59] Martenson JS, Tomita S. Synaptic localization of neurotransmitter receptors: comparing mechanisms for AMPA and GABA<sub>A</sub> receptors. *Curr Opin Pharmacol* 2015;20:102–8. doi: [10.1016/j.coph.2014.11.011](https://doi.org/10.1016/j.coph.2014.11.011).
- [60] Lerche H, Shah M, Beck H, Noebels J, Johnston D, Vincent A. Ion channels in genetic and acquired forms of epilepsy: ion channels in epilepsy. *J Physiol* 2013;591(4):753–64. doi: [10.1113/jphysiol.2012.240606](https://doi.org/10.1113/jphysiol.2012.240606).
- [61] Oyrer J, Maljevic S, Scheffer IE, Berkovic SF, Petrou S, Reid CA. Ion channels in genetic epilepsy: from genes and mechanisms to disease-targeted therapies. *Pharmacol Rev* 2018;70(1):142–73. doi: [10.1124/pr.117.014456](https://doi.org/10.1124/pr.117.014456).
- [62] Maljevic S, Møller RS, Reid CA, Pérez-Palma E, Lal D, May P, et al. Spectrum of GABA<sub>A</sub> receptor variants in epilepsy. *Curr Opin Neurol* 2019;32(2):183–90. doi: [10.1097/WCO.0000000000000657](https://doi.org/10.1097/WCO.0000000000000657).
- [63] Craig AM, Kang Y. Neurexin–neuroligin signaling in synapse development. *Curr Opin Neurobiol* 2007;17(1):43–52. doi: [10.1016/j.conb.2007.01.011](https://doi.org/10.1016/j.conb.2007.01.011).
- [64] Epi4K Consortium. De novo mutations in *SLC1A2* and *CACNA1A* are important causes of epileptic encephalopathies. *Am J Hum Genet* 2016;99(2):287–98. doi: [10.1016/j.ajhg.2016.06.003](https://doi.org/10.1016/j.ajhg.2016.06.003).
- [65] Helbig KL, Lauerer RJ, Bahr JC, Souza IA, Myers CT, Uysal B, et al. De novo pathogenic variants in *CACNA1E* cause developmental and epileptic encephalopathy with contractures, macrocephaly, and dyskinesias. *Am J Hum Genet* 2018;103(5):666–78. doi: [10.1016/j.ajhg.2018.09.006](https://doi.org/10.1016/j.ajhg.2018.09.006).
- [66] Takata A, Ionita-Laza I, Gogos JA, Xu B, Karayiorgou M. De novo synonymous mutations in regulatory elements contribute to the genetic etiology of autism and schizophrenia. *Neuron* 2016;89(5):940–7. doi: [10.1016/j.neuron.2016.02.024](https://doi.org/10.1016/j.neuron.2016.02.024).
- [67] Ganna A, Satterstrom FK, Zekavat SM, Das I, Kurki MI, Churchhouse C, et al. Quantifying the impact of rare and ultra-rare coding variation across the phenotypic spectrum. *Am J Hum Genet* 2018;102(6):1204–11. doi: [10.1016/j.ajhg.2018.05.002](https://doi.org/10.1016/j.ajhg.2018.05.002).
- [68] Takata A, Nakashima M, Saitou H, Mizuguchi T, Mitsuhashi S, Takahashi Y, et al. Comprehensive analysis of coding variants highlights genetic complexity in developmental and epileptic encephalopathy. *Nat Commun* 2019;10(1):2506. doi: [10.1038/s41467-019-10482-9](https://doi.org/10.1038/s41467-019-10482-9).
- [69] Leu C, Stevelink R, Smith AW, Goleva SB, Kanai M, Ferguson L, et al. Polygenic burden in focal and generalized epilepsies. *Brain* 2019;142(11):3473–81. doi: [10.1093/brain/awz292](https://doi.org/10.1093/brain/awz292).
- [70] Scheffer IE, Berkovic S, Capovilla G, Connolly MB, French J, Guilhoto L, et al. ILAE classification of the epilepsies: position paper of the ILAE commission for classification and terminology. *Epilepsia* 2017;58(4):512–21. doi: [10.1111/epi.13709](https://doi.org/10.1111/epi.13709).
- [71] Lal D, Ruppert A, Trucks H, Schulz H, de Kovel C, Trenité DK, et al. Burden analysis of rare microdeletions suggests a strong impact of neurodevelopmental genes in genetic generalised epilepsies. *PLOS Genet* 2015;11(5):e1005226. doi: [10.1371/journal.pgen.1005226](https://doi.org/10.1371/journal.pgen.1005226).
- [72] Heyne HO, Singh T, Stamberger H, Abou Jamra R, Caglayan H, Craiu D, et al. De novo variants in neurodevelopmental disorders with epilepsy. *Nat Genet* 2018;50(7):1048–53. doi: [10.1038/s41588-018-0143-7](https://doi.org/10.1038/s41588-018-0143-7).
- [73] EuroEPINOMICS-RES Consortium. Epilepsy Phenome/Genome Project. Epi4K Consortium. De novo mutations in synaptic transmission genes including *DNN1* cause epileptic encephalopathies. *Am J Hum Genet* 2014;95(4):360–70. doi: [10.1016/j.ajhg.2014.08.013](https://doi.org/10.1016/j.ajhg.2014.08.013).
- [74] De Rubeis S, He X, Goldberg AP, Poultnery CS, Samocha K, Cicek AE, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 2014;515(7526):209–15. doi: [10.1038/nature13772](https://doi.org/10.1038/nature13772).
- [75] The Brainstorm Consortium. Analysis of shared heritability in common disorders of the brain. *Science* 2018;360(6395):eaap8757. doi: [10.1126/science.aap8757](https://doi.org/10.1126/science.aap8757).
- [76] Johannesen KM, Gardella E, Linnankivi T, Courage C, de Saint Martin A, Lehesjoki AE, et al. Defining the phenotypic spectrum of *SLC6A1* mutations. *Epilepsia* 2018;59(2):389–402. doi: [10.1111/epi.13986](https://doi.org/10.1111/epi.13986).
- [77] Liu Y, Schubert J, Sonnenberg L, Helbig KL, Hoei-Hansen CE, Koko M, et al. Neuronal mechanisms of mutations in *SCN8A* causing epilepsy or intellectual disability. *Brain* 2019;142(2):376–90. doi: [10.1093/brain/awy326](https://doi.org/10.1093/brain/awy326).
- [78] Wolking S, May P, Mei D, Møller RS, Balestrini S, Helbig KL, et al. Clinical spectrum of *STX1B*-related epileptic disorders. *Neurology* 2019;92(11):e1238–49. doi: [10.1212/WNL.0000000000007089](https://doi.org/10.1212/WNL.0000000000007089).
- [79] Vardar G, Gerth F, Schmitt XJ, Rautenstrauch P, Trimbuch T, Schubert J, et al. Epilepsy-causing *STX1B* mutations translate altered protein functions into distinct phenotypes in mouse neurons. *Brain* 2020;143(7):2119–38. doi: [10.1093/brain/awaa151](https://doi.org/10.1093/brain/awaa151).
- [80] Wang J, Lin ZJ, Liu L, Xu HQ, Shi YW, Yi YH, et al. Epilepsy-associated genes. *Seizure* 2017;44:11–20. doi: [10.1016/j.seizure.2016.11.030](https://doi.org/10.1016/j.seizure.2016.11.030).
- [81] de Leeuw CA, Neale BM, Heskes T, Posthuma D. The statistical properties of gene-set analysis. *Nat Rev Genet* 2016;17(6):353–64. doi: [10.1038/nrg.2016.29](https://doi.org/10.1038/nrg.2016.29).
- [82] Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fenell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536(7616):285–91. doi: [10.1038/nature19057](https://doi.org/10.1038/nature19057).