Ultra-rare constrained missense variants in the epilepsies: Shared and specific enrichment patterns in neuronal gene-sets

Mahmoud Koko[1], Roland Krause[2], Thomas Sander[3], Dheeraj Reddy Bobbili[2], Michael Nothnagel[4], Patrick May[2*†], Holger Lerche[1*†], Epi25 Collaborative[5]

[1] Department of Neurology and Epileptology, Hertie Institute for Clinical Brain Research, University of Tübingen, Otfried-Müller Str. 25, 72076 Tübingen, Germany.

[2] Luxembourg Centre for Systems Biomedicine, University of Luxembourg, 6 Avenue du Swing, 4367 Belvaux, Luxembourg.

[3] University of Cologne, Faculty of Medicine and Cologne University Hospital, Cologne Center for Genomics, Weyertal 115b, 50931 Cologne, Germany.

[4] University of Cologne, Faculty of Mathematics and Natural Sciences and Cologne University Hospital, Cologne Center for Genomics, Weyertal 115b, 50931 Cologne, Germany.

[5] See contributing authors in the appendix.

[*] Contributed equally as PIs.

[†] Correspondence to:

Prof. Holger Lerche. Address: Department of Neurology and Epileptology, Hertie Institute for Clinical Brain Research, University and University Clinic Tübingen, Hoppe-Seyler-Str. 3, 72076 Tübingen, Germany. E-Mail: holger.lerche@uni-tuebingen.de

or

Dr. Patrick May. Address: Luxembourg Centre for Systems Biomedicine, University of Luxembourg, 6 Avenue du Swing, 4367 Belvaux, Luxembourg. E-Mail: patrick.may@uni.lu

Abstract:

Background: Burden analysis in epilepsy has shown an excess of deleterious ultra-rare variants (URVs) in few gene-sets, such as known epilepsy genes, constrained genes, ion channel or $GABA_A$ receptor genes. We set out to investigate the burden of URVs in a comprehensive range of gene-sets presumed to be implicated in epileptogenesis.

Methods: We investigated several constraint and conservation-based strategies to study whole exome sequencing data from European individuals with developmental and epileptic encephalopathies (DEE, n = 1,003), genetic generalized epilepsy (GGE, n = 3,064), and non-acquired focal epilepsy (NAFE, n = 3,522), collected by the Epi25 Collaborative, compared to 3,962 ancestry-matched controls. The burden of 12 URVs types in 92 gene-sets was compared between epilepsy cases (DDE, GGE, NAFE) and controls using logistic regression analysis.

Results: Burden analysis of brain-expressed genes revealed an excess of different URVs types in all three epilepsy categories which was largest for constrained missense variants. The URVs burden was prominent in neuron-specific, synaptic and developmental genes as well as genes encoding ion channels and receptors, and it was generally higher for DEE and GGE compared to NAFE. The patterns of URVs burden in gene-sets expressed in inhibitory vs. excitatory neurons or receptors suggested a high burden in both in DEE but a differential involvement of inhibitory genes in GGE, while excitatory genes were predominantly affected in NAFE. Top ranking susceptibility genes from a recent genome-wide association study (GWAS) of generalized and focal epilepsies displayed a higher URVs burden in constrained coding regions in GGE and NAFE, respectively.

Conclusions: Using exome-based gene-set burden analysis, we demonstrate that missense URVs affecting mainly constrained sites are enriched in neuronal genes in both common and rare severe epilepsy syndromes. Our results indicate a differential impact of these URVs in genes expressed in inhibitory vs. excitatory neurons and receptors in generalized vs. focal epilepsies. The excess of URVs in top-ranking GWAS risk-genes suggests a convergence of rare deleterious and common risk-variants in the pathogenesis of generalized and focal epilepsies.

Keywords: burden analysis, ultra-rare variants, gene-sets, epilepsy, exome sequencing.

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 (DEE) (1–3), sometimes with consequences on precision treatments (4–7). In contrast, only few individuals with familial or sporadic genetic generalized epilepsies (GGE) or non-acquired focal epilepsies (NAFE) harbor monogenic causative variations (8–10). Therefore, statistical methods 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 variants both in common and rare epilepsies in genes associated with dominant epilepsy syndromes, developmental and epileptic encephalopathy genes, and neuro-developmental disorders (NDD) with epilepsy genes, emphasizing a shared genetic component (8,10). Evidence for the enrichment of rare missense variants in genes encoding $GABA_A$ receptors and GABAergic pathway genes in genetic generalized epilepsies pointed to the 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.

Methods:

Study Samples: The Epi25 Collaborative collected and generated phenotyping and exome sequencing data from individuals with different subtypes of epilepsy (10). We analyzed subjects from recruitment years 1 and 2 (n=13,197). The epilepsy classification, phenotyping and consent procedures have been previously described (10). Five control cohorts, from the database of Genotypes and Phenotypes (11)

77 (dbGAP) and the Epi25 Collaborative, were available for this analysis (n=13,299), including Italian
78 controls from the Epi25 Collaborative, the Swedish Schizophrenia Study (dbGAP: phs000473), and
79 three Myocardial Infarction Genetics (MIGen) Consortium cohorts: Leicester UK Heart Study (dbGAP:
80 phs001000), Ottawa Heart Study (dbGAP: phs000806) and the Italian Atherosclerosis, Thrombosis,
81 and Vascular Biology "ATVB" Study (dbGAP: phs001592). Sequencing was performed on an Illumina
82 HiSeq 2000 or 2500 platform at the Broad Institute (Cambridge, MA, USA). The data generation
83 process has been previously described (10).

84 **Baseline sample QC:** For this analysis, we considered samples from three epilepsy categories (DEE,
85 GGE, and NAFE) as classified by the Epi25 phenotype review (10). Controls with Coronary Artery
86 Disease (CAD) diagnosis from the MIGen cohorts were not considered. Outliers on key sample calling
87 metrics (total SNVs/indels counts, TiTv ratio, Ins-Del ratio, Hom-Het ratio, autosomal heterozygosity),
88 samples with genotyping rate less than 90%, duplicates and related samples up to the 3rd degree (one
89 from each pair), and samples with ambiguous/discordant sequencing gender were removed (Fig. S1,
90 Fig. S2). Using multi-dimensional scaling, the genotypes of the remaining samples were projected on
91 the 1000 Genomes space (12). The major continental ancestry was then predicted using a Support
92 Vector Machine (Fig. S3). Samples labeled as European were further subclassified (non-Finish and
93 Finnish) following visualization of the first two principal components. Those samples with a predicted
94 ancestry other than non-Finnish European were filtered. Following this baseline filtering, 7,836 cases
95 and 8,822 controls (out of 13,197 cases and 13,299 controls) remained for subsequent analysis. These
96 filtering steps were performed using the Genome Analysis Toolkit (GATK) v4.1.4.1 (13), PLINK v1.9
97 (14) and KING v2.2.4 (15). For additional details, see "Baseline sample quality control" in the
98 supplemental methods (Additional file 1).

99 **Baseline variant QC:** Variants located outside Gencode v.33 (16) coding sequencies (CDS) boundaries
100 or in low complexity regions (17) were not considered. Multi-allelic calls were split, and the variants
101 were normalized. Variants with low variant quality score log-odds, covered at mean depth < 10x in the
102 baseline filtered sample set, at minimum depth of 10x in less than 95% of cases and controls, with large
103 difference in mean depth or call rate between the cases and controls, or with allele count equal to 0 were

104 removed. Low depth/quality genotype calls were set to missing. Variant filtering was performed using
105 bcftools/htslib v1.10.2 (18), vt v0.57721 (19) and GATK (13). See the supplemental methods section
106 "Baseline variant quality controls" for further details (Additional file 1).

107 **Case control matching and call rate harmonization:** We performed multiple iterations of principal
108 component analysis (PCA) on our baseline filtered dataset. Following the first round of PCA using
109 PLINK (14), the Swedish Schizophrenia Study control cohort showed poor clustering (PC1/2) with the
110 rest of study samples and was removed. Subsequently, outliers on top principal components were
111 filtered using EIGENSTRAT v6.1.4 (20,21). For details, see "Residual stratification" in supplemental
112 methods (Additional file 1). To handle the residual stratification caused by platform differences, we
113 removed all variants with call rate < 95% in any of the remaining sequencing cohorts (Epi25, Leicester,
114 Ottawa, ATVB) or with difference in call rate > 0.5% between any pair of cohorts. Also, variants with
115 Hardy-Weinberg Equilibrium p-value less than $1x10^{-6}$ (on the combined case-control cohort) were
116 excluded. These variant calling metrics were obtained using bcftools/htslib (18) and PLINK (14). The
117 final analysis set included 7,589 cases (DEE=1,003, GGE=3,064, NAFE=3,522) and 3,962 matched
118 controls (ATVB = 1,673, Leicester=1,082, Ottawa=924, Epi25 Italian=283) of non-Finnish European
119 ancestry (Table S1 and Table S2). The use of predominantly male or male-only control cohorts from
120 ATVB and Leicester studies resulted in a misbalanced sample sex ratio (53.6% female cases vs. 19,4%
121 female controls). We observed a total of 1,267,392 variants in the final dataset, comprising 1,247,342
122 SNVs and 20,050 indels. The QC-ed case-control cohort (Fig. S5) showed a balanced distribution of
123 variants and comparable variant calling metrics (Fig. S6).

124 **Qualifying variants (QVs):** The variants were annotated using snpEff v4.3 (22) and Annovar
125 v20191024 (23). We focused on URVs as these have shown a strong burden of deleterious pathogenic
126 variants in multiple studies of epilepsy and other neurological disorders (8,10,24–28). Here, URVs were
127 defined based on their population Minor Allele Frequencies (MAFs) in DiscovEHR ($MAF_{DiscovEHR}$) and
128 gnomAD r2.1 (MAF_{gnomAD}) population databases (29,30) and their Minor Allele Counts (MACs)
129 calculated separately for each analysis ($MAC_{DEE+Controls}$, $MAC_{GGE+Controls}$, $MAC_{NAFE+Controls}$), as follows
130 (i) $MAF_{DiscovEHR} = 0$ (ii) $MAF_{gnomAD} \leq 2x10^{-5}$ (iii) $MAC_{Epilepsy+Controls} \leq 3$ in the respective analysis group.

131 Performing three separate analyses for the three epilepsy phenotypes, with independent calculation of
132 MAC in each analysis, was intended to provide stringent control for inflation. Accordingly, the reported
133 variant counts in the control sets may differ slightly between the three analyses (see Additional file 3).
134 URVs were categories further into multiple conditions of qualifying variants (QVs) based on their
135 functional consequences. We considered thirteen variant conditions (QVs matching the specified effect
136 classes) including synonymous (presumed neutral), protein-truncating (presumed loss-of-function) and
137 multiple groups of missense variants (mix of loss- and gain-of-function mechanisms). The grouping of
138 missense variants in multiple (partially overlapping) conditions of QVs focused on three perspectives:
139 conventional in-silico deleteriousness, constraint and paralog conservation. It was based on multiple
140 predictions, namely, PolyPhen2 (PPh2) (31), Sorting Intolerant From Tolerant (SIFT) (32), Missense
141 Badness Polyphen and Constraint (MPC) (33), Missense Tolerance Ratio (MTR) (34), Constrained
142 Coding Regions (CCR) (35) and para-Z-score for paralog conservation (36). PPh2 (31) and SIFT (32)
143 are two conventional, commonly used, in-silico missense deleteriousness scores that are widely used in
144 genetic studies to identify likely benign and likely deleterious variants based on a number of features
145 including the sequence, phylogenetic and structural information. MPC score (33) aims to identify
146 regions within genes that are specifically depleted of missense variation and combines this information
147 with variant-level metrics that measures the increased deleteriousness of amino acid substitutions when
148 they occur in missense-constrained regions. MTR score (34) estimates the intolerance of genic regions
149 by comparing the observed proportion of missense variation to the expected proportion in the sequence
150 context of the protein-coding region understudy. While MPC and MTR scores are scaled down to
151 individual missense alterations, CCR score (35) aims to identify coding regions that are completely
152 devoid of variation in population databases. Functionally critical protein regions are usually encoded
153 by bases in regions with high CCR scores. Paralog conservation-based missense variant analysis was
154 recently shown to aid variant prioritization in neurodevelopmental disorders (36). It has been proposed
155 that most disease genes in humans have paralogs (37). The analyzed variant conditions (Table S6) were
156 (i) "Synonymous" variants that served as a control condition for inflation. (ii) Benign missense variants:
157 as predicted by PPh2 and SIFT. (iii) Damaging missense variants as predicted by PPh2 and SIFT. (iv)
158 Protein Truncating Variants (PTVs) that included stop-gained, start-lost, frameshift, splice-donor and

159 splice-acceptor variants. (v) All functional variants combined PTVs, in-frame indels and deleterious
160 missense variants (PPh2 and SIFT) (vi) "MPC 1" missense variants: constrained missense with MPC
161 score ≥ 1. (vii) "MPC 2" missense variants: highly constrained missense with MPC score ≥ 2 (enriched
162 for *de novo* variants). (viii) "MTR ClinVar" missense variants: constrained missense with MTR score
163 ≤ 0.825 which is the median for ClinVar variants not denoted as *de novo*. (ix) "MTR DeNovo" missense
164 variants: highly constrained missense with MTR score ≤ 0.565 which is the median for ClinVar *de novo*
165 variants. (x) "CCR 80" missense variants: highly constrained missense variants in regions with CCR
166 score ≥ 80, with MPC score ≥ 1, and MTR score ≤ 0.825. (xi) "paralog-non-conserved": missense
167 variants located in sites not conserved across paralog genes as indicated by a para-Z-score ≤ 0. (xii)
168 "paralog-conserved": missense variants located in sites conserved across paralog genes as indicated by
169 a para-Z-score > 0. (xiii) "paralog highly conserved": missense variants in highly conserved sites
170 between paralog genes with para-Z-score ≥ 1.

171 **Gene-sets:** 92 gene-sets were tested. In addition to exome-wide burden testing (one gene-set of all
172 protein coding genes), we defined additional 91 specific gene-sets as follows: (a) 34 sets based on gene
173 expression patterns in the brain: brain-expressed genes grouped by their intolerance profiles including
174 loss-of-function intolerant and missense intolerant genes (30,38), genes grouped by their regional brain
175 expression in the cortex and hippocampus from Genotype-Tissue Expression project v8 (39),
176 developmentally-relevant genes (40–42) , brain-enriched genes from the Human Protein Atlas v20.1
177 (43,44), genes enriched in specific cell types (45,46) (neurons, excitatory and inhibitory neurons, glia,
178 astrocytes, microglia, oligodendrocytes, and endothelial cells) and localization in certain neuronal
179 compartments (47,48); (b) 28 functional groups including ion channels (8), $GABA_A$ receptors (9),
180 excitatory receptors (9), GABAergic pathway (9), PSD-95 interactors (8), Gene Ontology (GO) gene-
181 sets of GABAergic and glutamatergic synapses (40,41,49), neuronal pathways from Kyoto
182 Encyclopedia of Genes and Genomes (KEGG) (50) and neuronal gene-sets from Reactome database
183 (51); (c) 14 gene-sets of known disease-related genes including monogenic epilepsy-causing genes (8–
184 10), epilepsy genome-wide association study (GWAS) top-ranking genes (positional mapping within a
185 window of 250 kb of significant loci and mapping based on chromatin interaction between gene

186 promoters and the significant locus) (52), co-regulated genes in the brain (53,54); and (d) 15 non-
187 neuronal gene-sets including genes not expressed in the brain (39,43,44), cancer and metabolic
188 pathways (50). The gene-sets are outlined in Table 1, Table S7 and Table S8 (see Additional file 1 and
189 Additional file 2).

190 **Table (1): Gene-sets investigated in this study.** (see end of text).

191 **Gene-set burden analysis:** We examined the burden of qualifying ultra-rare variants (QVs) in thirteen
192 variant conditions (Table S6) for 92 gene-sets in three epilepsy phenotypes (DEE, GGE, and NAFE)
193 against a set of matched controls. Gene-set burden testing was done using logistic regression by
194 regressing case-control status on the individual QVs counts. In each sample, QVs were collapsed by
195 gene and aggregated (summed) across a target gene-set to get a burden score (assuming equal weights
196 and direction of effects) which was used as a predictor in a binomial model while adjusting for
197 additional covariates (sex, top ten principal components, exome-wide variant count, and exome wide
198 singletons count) using *glm*() function from *stats* package (55). Likelihood ratio test (LRT) from *lmtest*
199 package (56) was used to compare a model with QVs burden and covariates against a null model
200 (covariates only). Log-odds from LRT and their respective 95% confidence intervals and *p* values are
201 presented here as a measure of enrichment in tested gene-sets. We employed a Benjamini-Hochberg
202 false discovery rate (FDR) multiple testing adjustment for *p* values that accounted for 3312 tests (92
203 gene-sets x 3 epilepsy phenotypes x 12 test variant conditions, excluding the synonymous variants) as
204 implemented in *p.adjust*() function from *stats* package (55). The cut-off for substantial enrichment was
205 defined as FDR-corrected p value < 0.05. For simplicity, p values (FDR corrected except for
206 synonymous variants) are indicated throughout the presented plots using stars as follows: no star > 0.05,
207 * < 0.05, ** < 0.005, *** < 0.0005, **** < 0.00005. To estimate the extent of bias that might have been
208 introduced by the imbalance in male-to-female ratios between cases and controls, we performed a
209 secondary analysis excluding chromosome X genes. Also, to ensure adequate control for any bias
210 introduced by differences in capture kits, we performed another supplementary analysis between two
211 groups of control samples (Leicester study controls vs. Ottawa and ATVB controls) representing two
212 main enrichment kits (Illumina ICE vs. Agilent SureSelect kits). The statistical analysis was performed

213 in R 3.3.3 (55). The analysis approach is outlined in Fig. 1 and the methods are detailed in Additional
214 file 1. A list of tested gene-sets and genes in each set is provided in Additional file 2.

215 **Results:**

216 **URV excess in brain-expressed genes**

217 First, we investigated the burden of ultra-rare variants across all protein coding genes following the
218 analysis approach outlined in Fig. 1. This revealed a clear enrichment in constrained missense variants
219 that was maximum in consensus constrained coding regions predicted by Missense-badness Polyphen
220 and Constraint (MPC), Missense Tolerance Ratio (MTR) and Consensus Coding Regions (CCR) scores
221 (Fig. 2). The combination of the three metrics (see methods) identifies highly deleterious variants in
222 functionally critical genic regions. In this particular analysis in all three phenotypes, about half of the
223 cases, in contrast to roughly one-fourth of controls, harbored one or more ultra-rare highly constrained
224 variants (Fig. S10). The primary analysis (10) examined loss-of-function intolerant genes and
225 demonstrated an increased burden in ultra-rare constrained as well as protein truncating variants
226 (PTVs). Here, the examination of brain-expressed intolerant genes showed, similarly, a marked
227 enrichment in PTVs in addition to a burden in constrained missense variants comparable to what is seen
228 exome-wide (Fig. 2 and Fig. S11).

229 When we examined protein coding genes grouped by their relative brain expression, damaging missense
230 variants were only substantially enriched in genes highly expressed in the cortex or hippocampus,
231 whereas those expressed at medium or low levels only showed an enrichment for the most constrained
232 missense variants (Fig. 3). Genes not expressed in brain did not show a substantial enrichment for any
233 variant type (Fig. S15). Genes showing a higher expression in the adult brain compared to other tissues
234 (brain-enriched & brain-enhanced) were also preferentially enriched, as well as genes associated with
235 brain development. Genes related to late rather than early development showed a slightly higher
236 enrichment in all three phenotypic groups (Fig. 3).

237 Focusing further on cell-type specific expression, neuron-specific genes were preferentially affected
238 compared to those enriched in glial cells, particularly in GGE (Fig. 4). To obtain further insight into the

239 nature of this neuronal enrichment, we used sets of genes representing paralogs of mouse genes found
240 to be enriched in excitatory or inhibitory neurons (see Additional file 1). Interestingly, genes
241 preferentially expressed in inhibitory neurons showed a an increased burden only in GGE, whereas
242 those preferentially expressed in excitatory neurons showed a more prominent signal in NAFE. Since
243 well-established epilepsy genes, like ion channels and receptors, show differential distributions in
244 different neuronal compartments (57,58), we examined further sets of genes based on subcellular
245 localization. We found that pre- and postsynaptic genes were enriched with variants in cases vs.
246 controls, as well as a very small set of 17 genes located in axon initial segments (most prominent in
247 DEE) (Fig. S13).

248 **Burden of URVs in ion channel, neurotransmitter receptor encoding and related genes**

249 Next, we examined functional gene-sets that could, more specifically, underlie the observed enrichment
250 in neuronal and synaptic genes. Ion channels, neurotransmitter receptors and transporters are widely
251 implicated in epilepsy, especially in monogenic and familial forms, displaying considerable phenotypic
252 heterogeneity and presenting as mild or severe epilepsies (59–61). Variants in $GABA_A$ receptors were
253 enriched in GGE but not in DEE and NAFE while those in gene-sets representing genes encoding N-
254 Methyl-D-Aspartate receptor and Activity-Regulated Cytoskeleton protein (NMDAR-ARC) were
255 enriched in NAFE and DEE. A comprehensive gene-set for the GABAergic pathway genes (9) showed
256 a prominent signal in GGE and DEE, and less in NAFE. In contrast, a gene-set representing PSD-95
257 interactors showed comparable enrichment in NAFE and GGE (Fig. 4). Brain-expressed ion channels
258 were found to be enriched for highly constrained missense variants (CCR 80 condition) in common as
259 well as rare epilepsies (Fig. 4).

260 **Patterns of burden in gene-sets representing inhibitory vs. excitatory signaling**

261 We then compared the patterns of URVs enrichment in genes involved in the GABAergic (main
262 inhibitory) pathway and synapse against those in the glutamatergic (main excitatory) pathway and
263 synapse in the brain, by examining their unique and overlapping genes based on KEGG pathways (50)
264 and GO synaptic gene-sets (41) and sets of specific receptors (Fig. 5). GGE showed a higher burden in

265 GABAergic vs. glutamatergic synapse (GO) and pathway (KEGG) genes, in $GABA_A$ receptors vs.
266 excitatory receptors/NMDAR-ARC genes, and in GABAergic pathway genes (comprehensive gene-
267 set) vs. PSD-95 interactors, thus matching the higher burden in genes representing inhibitory vs.
268 excitatory neuronal signaling. The CCR 80 analysis of GO gene-sets in NAFE showed a higher burden
269 in glutamatergic vs. GABAergic synapse genes, akin to the pattern seen in genes enriched in excitatory
270 vs. inhibitory neurons. The analysis of KEGG glutamatergic vs. GABAergic pathway genes did not
271 confirm this finding (Fig. 5). It is notable that the overlap between GO synapse and KEGG pathway
272 gene-sets is minimal (Fig. S18), and the size of GO and KEGG gene-sets was comparable in
273 GABAergic but discordant in glutamatergic genes.

274 Altogether, these comparisons of the burden in missense variants in highly constrained sites between
275 GGE and NAFE (Fig. 4 and Fig. 5) suggest the following patterns: (i) brain-expressed ion channels,
276 genes enriched in excitatory neurons, enriched in astrocytes, PSD-95 interactors, GABAergic and
277 glutamatergic synapse/pathway genes show an increased burden in cases vs. controls both in GGE &
278 NAFE; (ii) in GGE, this enrichment is coupled with a stronger enrichment in inhibitory neuronal genes,
279 in $GABA_A$ receptors and in GABAergic synapse-specific genes. (iii) in NAFE, this is accompanied by
280 an absence of enrichment in the later gene-sets and increased burden in the NMDAR-ARC gene-set.

281 **Burden in gene-sets of known epilepsy-related genes**

282 The primary Epi25 Collaborative analysis (10) demonstrated a high burden of missense variants in
283 constrained sites with MPC score ≥ 2 in DEE, GGE, and NAFE, seen in dominant epilepsy genes, DEE
284 genes, and NDD-Epilepsy genes. We observed similar enrichment patterns (Fig. 6) in MPC 2 and MTR
285 DeNovo conditions (enriched for de *novo* mutations). In addition, we saw a substantial enrichment for
286 other analysis conditions (MPC score ≥ 1 and MTR ClinVar) with lower odds ratios. Limiting the
287 analysis to highly constrained genic regions (CCR 80 condition) resulted in a marked increase in
288 missense burden, as was the trend in all the tested gene-sets so far. Testing these sets also unraveled
289 strong enrichment in PTVs and missense variants in paralog-conserved sites. PTVs and missense
290 variants in paralog-conserved sites did not show substantial enrichment in exome-wide analysis and
291 most of other expression-based, localization-based or pathway-based gene-sets. However, we saw a

292 modest increase in PTV burden in highly intolerant genes with probability of Loss-of-function
293 Intolerance (pLI) > 0.995 in all epilepsies (Fig. 2 and Fig. S11). In the gene-set of known DEE genes,
294 where highly intolerant genes are rather prevalent, we saw a prominent enrichment in PTVs burden in
295 DEE. Also, there was an increased burden in missense variants in paralog-conserved sites in sets of
296 epilepsy-related disease genes (DEE genes, dominant Epilepsy genes, NDD-Epilepsy genes). This
297 burden was very strong in DEE but not as remarkable in GGE and NAFE (Fig. 6).

298 **Enrichment in top GWAS hits captures divergence between common epilepsies**

299 Recent efforts from the ILAE consortium on complex epilepsies identified multiple associations in a
300 large GWAS of common epilepsies (52). To examine the hypothesis that genes located near to top
301 GWAS hits are also affected by rare variants, we tested the enrichment in sets of the 100 top-ranking
302 genes derived from the GWAS in generalized, focal, and all epilepsies. Interestingly, when limiting the
303 analysis to Consensus Coding Regions (CCR80 condition), top-ranking genes derived from the GWAS
304 of either GGE or focal epilepsies were preferentially enriched for rare variants in the respective
305 phenotypic groups of GGE and NAFE (Fig. 7). Although the observed enrichment was rather subtle,
306 this result was corroborated by a similar pattern for two rather small sets of known epilepsy genes that
307 are associated with either generalized or focal epilepsy (9).

308 **Brain- and epilepsy-related co-expression modules**

309 We also aimed to touch upon the role of brain co-expression modules identified in post-mortem brain
310 tissues from healthy individuals (54) and contrast these to the networks and modules identified in brain
311 tissue derived from epilepsy patients (53). A brain expression module was found to be substantially
312 enriched for rare deleterious variants in an independent cohort of DEE (54). A link to common epilepsy
313 phenotypes was also inferred, but a burden in ultra-rare variants was not examined so far. This module
314 showed a non-specific enrichment in all three epilepsy subtypes with highest odds in DEE. It is
315 noteworthy that this module overlaps largely with known epilepsy genes (Fig. S19). In resected
316 hippocampi of individuals with temporal lobe epilepsy (TLE), Johnson and colleagues identified two

317 co-expression modules (M1 and M2) within a gene-regulatory transcriptional network (53). A subtle
318 enrichment was seen in these modules in DEE and GGE, but not NAFE (Fig. 7).

319 **Additional neuronal and non-neuronal pathways**

320 Other neuronal gene-sets were enriched in our analysis (Fig. S14). Genes encoding neurexins and
321 neuroligins, important elements of pre- and post-synaptic interaction promoting adhesion between
322 dendrites and axons (62), were enriched in DEE (Fig. S14). Also, the synaptic vesicle cycle pathway
323 (KEGG) showed a prominent signal in both DEE and GGE. We also examined the burden in the mTOR
324 pathway (KEGG), hypothesizing that it could have potential relevance to focal epilepsies, but did not
325 detect a substantial enrichment (Fig. S14). Interestingly, NAFE analysis displayed a burden in
326 endothelial and astrocyte-specific genes in constrained genic regions (Fig. 4). Detailed results from all
327 tested conditions including the counts of genes with observed QVs, variant counts in cases and controls,
328 logistic regression odds of the individual QVs burden in cases vs. controls and related *p* values are
329 provided as supplemental material (see Additional file 3).

330 **Specificity of the observed enrichment patterns**

331 All four sets of genes not expressed in the brain that were tested (high confidence genes with depleted
332 RNA and protein expression in the brain, genes with no RNA detected in the cortex, the hippocampus
333 or any brain tissue) were not substantially enriched in almost all the tested variant conditions (Fig. S15).
334 In these sets, only one test across all thirteen conditions and three epilepsy subtypes showed an adjusted
335 p value < 0.05 (genes with no expression in brain tissues in GTEx portal; MTR DeNovo condition in
336 GGE). Additionally, we examined eleven metabolic and cancer pathways (KEGG) to have some
337 insights into the specificity of the observed signals to neuronal processes and genes. In tests targeting
338 functional variants (3 epilepsy subtypes, 11 KEGG metabolic and cancer pathways, 12 conditions/types
339 of variants excluding synonymous), 16/396 tests revealed corrected p values < 0.05. At least for some
340 of those, the significance could be explained by an overlap with genes known to play a role in epilepsy.
341 For instance, genes forming the Type II Diabetes KEGG pathway are substantially enriched in DEE
342 (corrected p values of 0.007 for MTR DeNovo and 0.01 for CCR 80 conditions). This pathway contains

343 two genes that are known to cause DEE, namely, *CACNA1A* (63) and *CACNA1E* (64). The enrichment
344 is no longer seen after removal of these two genes (p values > 0.05).

345 **Bias and inflation in gene-set burden testing**

346 The analysis for synonymous and benign missense variants did not show more substantial enrichment
347 than expected by chance, indicating sufficient control for inflation, particularly in exome-wide models
348 and gene-sets with large number of genes. Few gene-sets showed p values < 0.05 in the synonymous
349 variants' analysis (Additional file 3). The proportion of these tests (5%; 14 out of 276 tests of 92 gene-
350 sets and 3 phenotypes) was within the limit expected by chance under a true null hypothesis. Possible
351 alternative explanations for such subtle signals include residual stratification and differences in exome
352 capture not adjusted by covariates (exome-wide variant counts and principal components) and the
353 presence of synonymous variants with functional consequences (65). Another potential source of bias
354 in our burden testing was the imbalance in male-to-female ratios between cases and controls (Table S4).
355 We provide results from a secondary analysis that excluded all genes located on chromosome X, which
356 shows that any bias not captured by the inclusion of sample sex as a covariate is likely marginal
357 (Additional file 5). To exclude any major residual stratification resulting from the use of different
358 enrichment kits, we additionally performed a controls-only analysis (Additional file 5) in which we
359 compared control samples enriched with Illumina ICE capture kits (from Leicester study) to controls
360 enriched using Agilent SureSelect kits (ATVB study and Ottawa study). This analysis reflected a good
361 control for any potential bias introduced by different exome capture systems and also demonstrated that
362 the mixing of controls included (Leicester and Ottawa) or not included (ATVB) in gnomAD is unlikely
363 to have affected our main outcomes.

364 **Discussion:**

365 By analyzing the sequencing data of 11,551 unrelated European individuals, we show an exome-wide
366 burden in ultra-rare missense variants in epilepsy cases compared to controls. These variants are mainly
367 found in constrained sites across three different subtypes of common generalized and focal epilepsies
368 as well as rare and mostly severe developmental and epileptic encephalopathies. This ultra-rare variant

369 burden was prominent for missense variants enriched for *de novo* alterations, yet also substantial in
370 analyses examining a likely combination of inherited and *de novo* variants. While this burden was
371 consistently higher with increased constraint in various gene-sets, PTVs were specifically relevant in
372 known disease genes (Fig. 6) and brain-expressed loss-of-function intolerant genes in comparison to all
373 protein coding genes (Fig. 2). This replicates the findings of the primary analysis (10) which indicated
374 an increased PTV burden in known disease-causing epilepsy related genes and genes with pLI > 0.995
375 but not those with pLI 0.9-0.995 (Fig. S11). In general, PTVs are known to be associated with several
376 phenotypes with their effect being most prominent in LOF-intolerant genes (30,38,66). Similar to PTVs,
377 we did not observe a substantial increase in the burden of missense variants in paralog-conserved sites
378 in a group of all protein coding genes, but it was remarkably high in gene-sets of known disease genes.
379 The increased burden was very prominent in DEE but not in GGE and NAFE. Missense variants in
380 paralog-conserved sites showed prominent enrichment in neurodevelopmental disorders (36), a
381 phenotypic category that overlaps largely with DEE. Although this may reflect a true disparity based
382 on the importance of these highly conserved sites (with their disturbance possibly resulting in severe
383 rather than mild epilepsy phenotypes), it is possible that our analysis lacked power to detect small effect
384 sizes in common epilepsies.

385 In addition to their utility in estimating the contribution of certain variant conditions/types in different
386 epilepsy phenotypes, gene-sets with known relation to epilepsy, especially to monogenic forms,
387 constitute high-effect-size gene-sets that can also serve as technical validation sets. The presented
388 results are consistent with previous analysis of missense variants in a small number of gene-sets
389 examined in similar cohorts (8–10). The systematic analysis of gene-sets and more different
390 conditions/types of variants revealed interesting findings about the neurobiology of distinct types of
391 epilepsy and clearly revealed that the pathological burden is markedly higher in constrained coding sites
392 and regions. Although associated with higher odds ratios of an epilepsy phenotype, these variants are
393 not deterministic on their own, since about one-fourth of the controls also carry a qualifying variant in
394 the CCR 80 analysis (Fig. S10). As such, the phenotype is determined by a constellation of other factors,

395 possibly including patterns of multiple variations, oligogenic contribution from rare variants (67), and
396 polygenic risk from common variants (68).

397 The excess of ultra-rare missense variants in the epilepsies clearly stemmed from genes with enriched
398 expression in the brain compared to other tissues, including developmentally relevant genes. A strong
399 link exists between brain development and rare DEE and NDD, since both present with marked
400 developmental deficits accompanied with different seizure types (69,70). The relation of developmental
401 genes to common epilepsies is not as clear. At the cellular level, the enrichment was primarily seen in
402 neuron-specific genes. Synaptic genes with their various pre- and postsynaptic localizations were main
403 drivers in all epilepsies, with prominent signals in postsynaptic genes (Fig. S13). A similar enrichment
404 in synaptic genes has been observed in neurodevelopmental disorders with epilepsy (70,71),
405 schizophrenia (24) and autism (72), highlighting a shared genetic architecture not only between epilepsy
406 subtypes but also with other related neurological disorders, as has been shown previously for common
407 variants (73). Ion channel genes were enriched for ultra-rare, constrained variants in all epilepsy
408 phenotypes (Fig. 4), as has been shown in previous work focusing on ultra-rare variants (8,10).

409 Despite the common genetic and phenotypic features, DEE, GGE and NAFE represent well-recognized
410 phenotypic clusters with defined electro-encephalographic and clinical characteristics. Given the
411 phenotypic severity of DEE, the prevalence of *de novo* variants and 'monogenic' cases in DEE, and the
412 description of phenotypic spectra for genes involved in DEE that also span the milder GGE or NAFE,
413 the distinction between severe and mild epilepsies could be attributed, at least to some extent, to the
414 severity of the genetic defects, their functional effects or their localization within certain channel regions
415 (61,74–77). The distinction between GGE and NAFE, however, is probably functional, at least in part,
416 as suggested by previous work demonstrating the centrality of GABAergic genes in generalized
417 epilepsies (9,10). Also, it is well recognized that few genes present with focal, but not generalized,
418 epilepsy syndromes (78). Here, phenotype-specific patterns were seen in comparisons of GGE and
419 NAFE. $GABA_A$ receptor genes, GABAergic pathway genes and genes enriched in inhibitory neurons
420 were preferentially affected in GGE in comparison to the glutamatergic pathway genes. In contrast,
421 NMDA receptor and ARC genes, and genes expressed in excitatory neurons were enriched in NAFE.

422 Corroborating this finding for GGE, we found that the enrichment in the GABAergic pathway is
423 stronger in core genes that are not shared with the glutamatergic pathway. An opposing pattern (albeit,
424 not as prominent) was seen in NAFE, where the burden in genes enriched in excitatory neurons and
425 glutamatergic synapse was more prominent than the burden in their inhibitory and GABAergic
426 counterparts, providing evidence for the importance of the former in focal epilepsies.

427 Additional disparities in key gene-sets point to a possible genetic-functional divergence, so that a
428 common background of shared risk seems to be overlaid by specific risk entities. Groups of known
429 genes implicated in focal vs. generalized epilepsy were enriched in NAFE vs. GGE. Interestingly, the
430 same pattern was found for the 100 top ranking genes associated with GWAS hits, which were
431 preferentially enriched in respective phenotypic groups. The enrichment of rare variants in GWAS
432 genes also supports the convergence of ultra-rare and common variants in conferring epilepsy risk.
433 According to our findings, a link between common and rare variants is likely to be also relevant for the
434 phenotypic heterogeneity observed in seizure disorders. Notably, polygenic risk scores also pointed out
435 the specificity of the risk profiles in common epilepsies (68). We also found an enrichment of ultra-rare
436 *de novo* variants in DEE in a previously identified brain co-expression module (54), and the same
437 module was also enriched for constrained variants in GGE and NAFE. Although co-expressed genes
438 are not necessarily part of a single pathway, they represent closely orchestrated networks with possible
439 functional correlations. It is therefore conceivable that differentially expressed genes in individuals with
440 epilepsy would highlight modules in which altered transcription, ultra-rare variants, or both contribute
441 to cause both rare and common epilepsies.

442 **Study limitations**

443 Despite their robustness, the associations presented in this work should be interpreted with the caveats
444 of gene group testing in mind (79). Given that pathways and molecular processes are not consistently
445 defined in different resources, it is not always easy to define genes that represent a certain pathway. For
446 instance, GABAergic and glutamatergic pathway definitions based on GO terms and KEGG databases
447 are widely discordant (Fig. S18). These differences may explain the discrepancies we observed in
448 enrichment patterns in the same pathway. One gene with abundance of qualifying variants in the GO-

449 based analysis of the glutamatergic pathway was *PPFIA3* (see Additional file 4), a highly intolerant
450 gene that encodes a synaptic receptor tyrosine phosphatase highly expressed in the brain. However,
451 *PPFIA3* is not part of the KEGG-based glutamatergic pathway. Attempting to overcome such
452 discrepancies, we examined multiple overlapping gene-sets from different resources. Associations that
453 are seen regardless of the gene-set source (e.g., the consistent pattern of enrichment in inhibitory vs.
454 excitatory neuronal/pathway genes in GGE vs. NAFE) are, therefore, likely to underscore a genuine
455 biological relevance.

456 The analysis presented here has additional limitations which we aimed to overcome using stringent
457 analysis and quality control strategies. The limited use of about half of the controls from the primary
458 analysis to maximize case-control matching affected the overall power. Nevertheless, we were able to
459 reproduce most of the major signals from gene-sets with large effect sizes, the latter thereby active as
460 positive controls. The male-to-female ratios were not well-balanced in our case and control sets. Based
461 on a secondary analysis excluding X chromosomal genes, this does not seem to introduce a substantial
462 bias (see Additional file 5). The overlap between the controls used in this study and gnomAD controls
463 resulting from MIGen Leicester and Ottawa controls (Table S2) created some challenges in defining
464 ultra-rare variants, usually defined as those variants not observed in population databases. For
465 population frequency filtering, we allowed around five alleles in gnomAD (allele frequency of $2x10^{-5}$).
466 This count that exceeds our internal filtering cut-off for ultra-rare variants (three alleles) would allow
467 for the retention of ultra-rare variants from our control that are also seen in gnomAD while still filtering
468 common variants and prevalent sequencing artifacts. Also, multiple in-silico algorithms for predicting
469 missense deleteriousness and estimating constraint (including MPC, MTR, and CCR scores) were
470 derived from or validated in the Exome Aggregation Consortium (80) and gnomAD databases.
471 Examination of control conditions and control gene-sets that are not expected to show an enrichment
472 (e.g., genes not expressed in the brain) did not indicate any prominent inflation and supported the
473 validity of the overall analysis. Interestingly, some signals in 'non-neuronal' gene-sets could be
474 explained by the inclusion of well-established epilepsy genes (Fig. S20), like *CACNA1A* (63) and
475 *CACNA1E* (64), which are in fact key neuronal genes of synaptic transmission.

476 **Conclusions:**

477 Missense URVs affecting constrained sites in brain-expressed genes are key genetic drivers in epilepsy.
478 Patients with both common and rare epilepsy subtypes show an increased exome-wide burden of such
479 variants that is primarily derived from neuronal genes, where key gene-sets including ion channels,
480 developmental and synaptic genes are enriched across the phenotypic spectrum of epilepsy. Genes
481 implicated by common GWAS variants may also be disrupted by URVs in various epilepsy phenotypes,
482 suggesting a convergence of rare disruptive variants, and common variants in the pathogenesis of
483 epilepsy. Enrichment patterns of URV-affected genes suggest a preferential involvement of inhibitory
484 genes in GGE and excitatory genes in focal epilepsies.

485 **Additional files:**

486 - Additional file 1: PDF file. Supplemental methods, tables and figures; affiliations of the Epi25
487 Collaborative members.
488 - Additional file 2: Excel xlsx file. List of genes and gene-sets.
489 - Additional file 3: Excel xlsx file. Gene-set burden analysis results.
490 - Additional file 4: Excel xlsx file. Top-ranking genes per gene-set in the CCR 80 analysis.
491 - Additional file 5: Excel xlsx file. Secondary analysis results.

492 **List of abbreviations:**

493 ARC: Activity-Regulated Cytoskeleton protein.

494 ATVB: Atherosclerosis, Thrombosis, and Vascular Biology Study.

495 CAD: Coronary Artery Disease.

496 CCR: Constrained Coding Regions.

497 CDS: Coding sequences.

498 dbGAP: Database of Genotypes and Phenotypes.

499 DEE: Developmental and Epileptic Encephalopathies.

500 FDR: False Discovery Rate.

501 FMRP: Fragile-X Mental Retardation Protein.

502 GATK: Genome Analysis Toolkit.

503 GGE: Genetic Generalized Epilepsy.

504 GO: Gene Ontology.

505 GWAS: Genome-Wide Association Study.

506 KEGG: Kyoto Encyclopedia of Genes and Genomes.

507 ILAE: International League Against Epilepsy.

508 MAC: Minor Allele Counts.

509 MAF: Minor Allele Frequencies.

510 MGI: Mouse Genome Informatics.

511 MIGen: Myocardial Infarction Genetics Consortium.

512 MPC: Missense Badness Polyphen and Constraint.

513 MTR: Missense Tolerance Ratio.

514 NAFE: Non-Acquired Focal Epilepsy.

515 NDD: Neuro-Developmental Disorders.

516 NMDA: N-methyl D-Aspartate.

517 PCA: Principal Component Analysis.

518 PPh2: PolyPhen2

519 PSD-95: Post-synaptic density protein 95.

520 PTVs: Protein Truncating Variants.

521 QVs: Qualifying Variants.

522 SIFT: Sorting Intolerant From Tolerant

523 URV: Ultra-Rare Variants.

524 **Declarations:**

525 Ethics approval and consent to participate: This analysis utilized previously analyzed and published
526 data from the Epi25 Collaborative and the MIGen Consortium. Subjects investigated by the
527 Collaborative provided signed informed consent at the participating centers according to local national
528 ethical requirements and their standards at the time of collection. Approval for data reuse and analysis
529 was obtained from the Epi25 Collaborative (cases) and dbGAP (controls). The ethical approval and
530 consents procedures for the individual cohorts were reported by the Epi25 Collaborative (see methods
531 section).

532 Consent for publication: not applicable.

533 Availability of data and materials: The data/analyses presented in the current publication are based on
534 the use of study data from the Epi25 Collaborative (http://epi-25.org/) and the dbGaP
535 (https://ncbi.nlm.nih.gov/gap/) under accessions phs000473, phs001000, phs000806, phs001592 and
536 phs001489. Data supporting the results reported in the article can be found in this published article and
537 its supplementary information files.

538 Competing interests: None declared.

539 Funding: RK, TS, MN, PM and HL (and IH and YW as part of the Epi25 collaborative) were supported
540 by the Research Unit FOR-2715 of the German Research Foundation and the Fond Nationale de la
541 Recherche of Luxembourg (DFG/FNR, grants INTER/DFG/17/11583046, Le1030/16-1, Sa434/6-1,
542 No755/6-1, He5415/7-1, We4896/4-1). MK was supported by the German Academic Exchange Service
543 (DAAD personal funding program number 57214224). PM obtained additional FNR funding as part of
544 the National Centre of Excellence in Research on Parkinson's disease (NCER-PD, FNR11264123). The
545 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. Data used for the analyses described in this manuscript were obtained from the GTEx Portal (dbGaP accession number phs000424.v8.p2) on 30/11/2020.

Authors' contributions: HL, PM, MK conceived the study. HL, PM, RK, and MK contributed to the study design and/or data acquisition. MK and PM performed the analysis with input from HL, RK, TS, DRB and MN. MK, PM, and HL interpreted the results and wrote the manuscript. Authors from the Epi25 Collaborative recruited and phenotyped patients and/or contributed to data generation as indicated in the appendix. All authors reviewed the manuscript for intellectual content and approved the final version of the manuscript.

Acknowledgements: We are cordially thankful to the Epi25 Collaborative staff and all patients participating in the Epi25 project, who all made this study possible.

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 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 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, S. Hande Caglayan, Mutluay Arslan, Zuhal Yapıcı, Uluc Yis, Pınar 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.

601 Sobell, Peter F. Buckley, Fabio Macciardi, Mark H. Rapaport, James A. Knowles, Genomic Psychiatry
602 Cohort (GPC) Consortium, Ayman H. Fanous, Steven A. McCarroll, Namrata Gupta, Stacey B. Gabriel,
603 Mark J. Daly, Eric S. Lander, Daniel H. Lowenstein, David B. Goldstein, Holger Lerche, Samuel F.
604 Berkovic, and Benjamin M. Neale. See Additional file 1 for the affiliations.

605 **References:**

606 1. McTague A, Howell KB, Cross JH, Kurian MA, Scheffer IE. The genetic landscape of the
607 epileptic encephalopathies of infancy and childhood. Lancet Neurol. 2016 Mar;15(3):304–16.
608 Available from: https://linkinghub.elsevier.com/retrieve/pii/S1474442215002501

609 2. Happ HC, Carvill GL. A 2020 View on the Genetics of Developmental and Epileptic
610 Encephalopathies. Epilepsy Curr. 2020 Mar;20(2):90–6. Available from:
611 http://journals.sagepub.com/doi/10.1177/1535759720906118

612 3. Hebbar M, Mefford HC. Recent advances in epilepsy genomics and genetic testing.
613 F1000Research. 2020 Mar 12;9:185. Available from: https://f1000research.com/articles/9-185/v1

614 4. EpiPM Consortium. A roadmap for precision medicine in the epilepsies. Lancet Neurol. 2015
615 Dec;14(12):1219–28. Available from:
616 https://linkinghub.elsevier.com/retrieve/pii/S1474442215001994

617 5. Striano P, Minassian BA. From Genetic Testing to Precision Medicine in Epilepsy.
618 Neurotherapeutics. 2020 Apr;17(2):609–15. Available from:
619 http://link.springer.com/10.1007/s13311-020-00835-4

620 6. Sisodiya SM. Precision medicine and therapies of the future. Epilepsia. 2021 Mar;62(S2).
621 Available from: https://onlinelibrary.wiley.com/doi/10.1111/epi.16539

622 7. Lerche H. Drug-resistant epilepsy — time to target mechanisms. Nat Rev Neurol. 2020
623 Nov;16(11):595–6. Available from: http://www.nature.com/articles/s41582-020-00419-y

624 8. Epi4K Consortium, Epilepsy Phenome/Genome Project. Ultra-rare genetic variation in common
625 epilepsies: a case-control sequencing study. Lancet Neurol. 2017 Feb;16(2):135–43. Available
626 from: https://linkinghub.elsevier.com/retrieve/pii/S1474442216303593

627 9. May P, Girard S, Harrer M, Bobbili DR, Schubert J, Wolking S, et al. Rare coding variants in
628 genes encoding GABAA receptors in genetic generalised epilepsies: an exome-based case-control

study. Lancet Neurol. 2018 Aug;17(8):699–708. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1474442218302151

10. Epi25 Collaborative. Ultra-Rare Genetic Variation in the Epilepsies: A Whole-Exome Sequencing Study of 17,606 Individuals. Am J Hum Genet. 2019 Aug;105(2):267–82. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0002929719302071

11. 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 Jan;42(D1):D975–9. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkt1211

12. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015 Oct;526(7571):68–74. Available from: http://www.nature.com/articles/nature15393

13. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010 Sep 1;20(9):1297–303. Available from: http://genome.cshlp.org/cgi/doi/10.1101/gr.107524.110

14. 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 Dec;4(1):7. Available from: https://academic.oup.com/gigascience/article-lookup/doi/10.1186/s13742-015-0047-8

15. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010 Nov 15;26(22):2867–73. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btq559

16. Frankish A, Diekhans M, Ferreira A-M, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. 2019 Jan 8;47(D1):D766–73. Available from: https://academic.oup.com/nar/article/47/D1/D766/5144133

17. Krusche P, Trigg L, Boutros PC, Mason CE, De La Vega FM, Moore BL, et al. Best practices for benchmarking germline small-variant calls in human genomes. Nat Biotechnol. 2019 May;37(5):555–60. Available from: http://www.nature.com/articles/s41587-019-0054-x

18. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078–9. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btp352

659 19. Tan A, Abecasis GR, Kang HM. Unified representation of genetic variants. Bioinformatics. 2015
660 Jul 1;31(13):2202–4. Available from: https://academic.oup.com/bioinformatics/article-
661 lookup/doi/10.1093/bioinformatics/btv112

662 20. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components
663 analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006
664 Aug;38(8):904–9. Available from: http://www.nature.com/articles/ng1847

665 21. Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. PLoS Genet.
666 2006;2(12):e190. Available from: https://dx.plos.org/10.1371/journal.pgen.0020190

667 22. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating
668 and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of
669 Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012 Jun;6(2):80–92.

670 23. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-
671 throughput sequencing data. Nucleic Acids Res. 2010 Sep 1;38(16):e164–e164. Available from:
672 https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkq603

673 24. Genovese G, Fromer M, Stahl EA, Ruderfer DM, Chambert K, Landén M, et al. Increased burden
674 of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. Nat Neurosci.
675 2016 Nov;19(11):1433–41. Available from: http://www.nature.com/articles/nn.4402

676 25. Demontis D, Satterstrom K, Duan J, Lescai F, Dinesen Østergaard S, Lesch K-P, et al. The Role
677 of Ultra-Rare Coding Variants In ADHD. Eur Neuropsychopharmacol. 2019;29:S724–5. Available
678 from: https://linkinghub.elsevier.com/retrieve/pii/S0924977X1730319X

679 26. Wilfert AB, Turner TN, Murali SC, Hsieh P, Sulovari A, Wang T, et al. Recent ultra-rare
680 inherited mutations identify novel autism candidate risk genes. Genetics; 2020 Feb . Available
681 from: http://biorxiv.org/lookup/doi/10.1101/2020.02.10.932327

682 27. Singh T, Walters JTR, Johnstone M, Curtis D, Torniainen M, Rees E, et al. The contribution of
683 rare variants to risk of schizophrenia in individuals with and without intellectual disability. Nat
684 Genet. 2017 Aug;49(8):1167–73. Available from: http://www.nature.com/articles/ng.3903

685 28. Bennett CA, Petrovski S, Oliver KL, Berkovic SF. ExACtly zero or once: A clinically helpful
686 guide to assessing genetic variants in mild epilepsies. Neurol Genet. 2017 Aug;3(4):e163. Available
687 from: http://ng.neurology.org/lookup/doi/10.1212/NXG.0000000000000163

688 29. Dewey FE, Murray MF, Overton JD, Habegger L, Leader JB, Fetterolf SN, et al. Distribution and
689 clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study.
690 Science. 2016 Dec 23;354(6319):aaf6814. Available from:
691 https://www.sciencemag.org/lookup/doi/10.1126/science.aaf6814

692 30. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational
693 constraint spectrum quantified from variation in 141,456 humans. Nature. 2020
694 May;581(7809):434–43. Available from: http://www.nature.com/articles/s41586-020-2308-7

695 31. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and
696 server for predicting damaging missense mutations. Nat Methods. 2010 Apr;7(4):248–9. Available
697 from: http://www.nature.com/articles/nmeth0410-248

698 32. Sim N-L, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of
699 amino acid substitutions on proteins. Nucleic Acids Res. 2012 Jul 1;40(W1):W452–7. Available
700 from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gks539

701 33. Samocha KE, Kosmicki JA, Karczewski KJ, O'Donnell-Luria AH, Pierce-Hoffman E, MacArthur
702 DG, et al. Regional missense constraint improves variant deleteriousness prediction. bioRxiv. 2017
703 Jun; Genomics. Available from: http://biorxiv.org/lookup/doi/10.1101/148353

704 34. Traynelis J, Silk M, Wang Q, Berkovic SF, Liu L, Ascher DB, et al. Optimizing genomic
705 medicine in epilepsy through a gene-customized approach to missense variant interpretation.
706 Genome Res. 2017 Oct;27(10):1715–29. Available from:
707 http://genome.cshlp.org/lookup/doi/10.1101/gr.226589.117

708 35. Havrilla JM, Pedersen BS, Layer RM, Quinlan AR. A map of constrained coding regions in the
709 human genome. Nat Genet. 2019 Jan;51(1):88–95. Available from:
710 http://www.nature.com/articles/s41588-018-0294-6

711 36. Lal D, May P, Perez-Palma E, Samocha KE, Kosmicki JA, Robinson EB, et al. Gene family
712 information facilitates variant interpretation and identification of disease-associated genes in
713 neurodevelopmental disorders. Genome Med. 2020 Dec;12(1):28. Available from:
714 https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-020-00725-6

715 37. Dickerson JE, Robertson DL. On the Origins of Mendelian Disease Genes in Man: The Impact of
716 Gene Duplication. Mol Biol Evol. 2012 Jan 1;29(1):61–9. Available from:
717 https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msr111

718 38. Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, et al. A framework for
719 the interpretation of de novo mutation in human disease. Nat Genet. 2014 Sep;46(9):944–50.
720 Available from: http://www.nature.com/articles/ng.3050

721 39. Genotype Tissue Expression Project portal. Available from: https://www.gtexportal.org. Accessed
722 2020 Dec 1.

723 40. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for
724 the unification of biology. Nat Genet. 2000 May;25(1):25–9. Available from:
725 http://www.nature.com/articles/ng0500_25

726 41. The Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine. Nucleic
727 Acids Res. 2021 Jan 8;49(D1):D325–34. Available from:
728 https://academic.oup.com/nar/article/49/D1/D325/6027811

729 42. Cardoso-Moreira M, Halbert J, Valloton D, Velten B, Chen C, Shao Y, et al. Gene expression
730 across mammalian organ development. Nature. 2019 Jul;571(7766):505–9. Available from:
731 http://www.nature.com/articles/s41586-019-1338-5

732 43. Human Protein Atlas. Available from: https://www.proteinatlas.org. Accessed 2020 Dec 1.

733 44. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based
734 map of the human proteome. Science. 2015 Jan 23;347(6220):1260419–1260419. Available from:
735 https://www.sciencemag.org/lookup/doi/10.1126/science.1260419

736 45. Dusart P, Hallström BM, Renné T, Odeberg J, Uhlén M, Butler LM. A Systems-Based Map of
737 Human Brain Cell-Type Enriched Genes and Malignancy-Associated Endothelial Changes. Cell
738 Rep. 2019 Nov;29(6):1690-1706.e4. Available from:
739 https://linkinghub.elsevier.com/retrieve/pii/S2211124719312926

740 46. Mo A, Mukamel EA, Davis FP, Luo C, Henry GL, Picard S, et al. Epigenomic Signatures of
741 Neuronal Diversity in the Mammalian Brain. Neuron. 2015 Jun;86(6):1369–84. Available from:
742 https://linkinghub.elsevier.com/retrieve/pii/S0896627315004250

743 47. Pirooznia M, Wang T, Avramopoulos D, Valle D, Thomas G, Huganir RL, et al. SynaptomeDB:
744 an ontology-based knowledgebase for synaptic genes. Bioinformatics. 2012 Mar 15;28(6):897–9.
745 Available from: https://academic.oup.com/bioinformatics/article-
746 lookup/doi/10.1093/bioinformatics/bts040

747 48. Koopmans F, van Nierop P, Andres-Alonso M, Byrnes A, Cijsouw T, Coba MP, et al. SynGO:
748 An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. Neuron. 2019
749 Jul;103(2):217-234.e4. Available from:
750 https://linkinghub.elsevier.com/retrieve/pii/S0896627319304271

751 49. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular
752 Signatures Database Hallmark Gene Set Collection. Cell Syst. 2015 Dec;1(6):417–25. Available
753 from: https://linkinghub.elsevier.com/retrieve/pii/S2405471215002185

754 50. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses
755 and cellular organisms. Nucleic Acids Res. 2021 Jan 8;49(D1):D545–51. Available from:
756 https://academic.oup.com/nar/article/49/D1/D545/5943834

757 51. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The reactome pathway
758 knowledgebase. Nucleic Acids Res. 2019 Nov 6;gkz1031. Available from:
759 https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkz1031/5613674

760 52. The International League Against Epilepsy Consortium on Complex Epilepsies. Genome-wide
761 mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common
762 epilepsies. Nat Commun. 2018 Dec;9(1):5269. Available from:
763 http://www.nature.com/articles/s41467-018-07524-z

764 53. Johnson MR, Behmoaras J, Bottolo L, Krishnan ML, Pernhorst K, Santoscoy PLM, et al. Systems
765 genetics identifies Sestrin 3 as a regulator of a proconvulsant gene network in human epileptic
766 hippocampus. Nat Commun. 2015 May;6(1):6031. Available from:
767 http://www.nature.com/articles/ncomms7031

768 54. Delahaye-Duriez A, Srivastava P, Shkura K, Langley SR, Laaniste L, Moreno-Moral A, et al.
769 Rare and common epilepsies converge on a shared gene regulatory network providing opportunities
770 for novel antiepileptic drug discovery. Genome Biol. 2016 Dec;17(1):245. Available from:
771 http://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-1097-7

772 55. R Core Team. R: A language and environment for statistical computing. Available from:
773 https://www.R-project.org/

774 56. Zeileis, Hothorn. Diagnostic Checking in Regression Relationships. Rnews. 2002. Available
775 from: https://CRAN.R-project.org/doc/Rnews/

776 57. Vacher H, Mohapatra DP, Trimmer JS. Localization and Targeting of Voltage-Dependent Ion
777 Channels in Mammalian Central Neurons. Physiol Rev. 2008 Oct;88(4):1407–47. Available from:
778 https://www.physiology.org/doi/10.1152/physrev.00002.2008

779 58. Martenson JS, Tomita S. Synaptic localization of neurotransmitter receptors: comparing
780 mechanisms for AMPA and GABAA receptors. Curr Opin Pharmacol. 2015 Feb;20:102–8.
781 Available from: https://linkinghub.elsevier.com/retrieve/pii/S1471489214001660

782 59. Lerche H, Shah M, Beck H, Noebels J, Johnston D, Vincent A. Ion channels in genetic and
783 acquired forms of epilepsy: Ion channels in epilepsy. J Physiol. 2013 Feb;591(4):753–64. Available
784 from: http://doi.wiley.com/10.1113/jphysiol.2012.240606

785 60. Oyrer J, Maljevic S, Scheffer IE, Berkovic SF, Petrou S, Reid CA. Ion Channels in Genetic
786 Epilepsy: From Genes and Mechanisms to Disease-Targeted Therapies. Pharmacol Rev. 2018
787 Jan;70(1):142–73. Available from:
788 http://pharmrev.aspetjournals.org/lookup/doi/10.1124/pr.117.014456

789 61. Maljevic S, Møller RS, Reid CA, Pérez-Palma E, Lal D, May P, et al. Spectrum of GABAA
790 receptor variants in epilepsy. Curr Opin Neurol. 2019 Apr;32(2):183–90. Available from:
791 https://journals.lww.com/00019052-201904000-00003

792 62. Craig AM, Kang Y. Neurexin–neuroligin signaling in synapse development. Curr Opin
793 Neurobiol. 2007 Feb;17(1):43–52. Available from:
794 https://linkinghub.elsevier.com/retrieve/pii/S095943880700013X

795 63. Epi4K Consortium. De Novo Mutations in SLC1A2 and CACNA1A Are Important Causes of
796 Epileptic Encephalopathies. Am J Hum Genet. 2016 Aug;99(2):287–98. Available from:
797 https://linkinghub.elsevier.com/retrieve/pii/S000292971630204X

798 64. Helbig KL, Lauerer RJ, Bahr JC, Souza IA, Myers CT, Uysal B, et al. De Novo Pathogenic
799 Variants in CACNA1E Cause Developmental and Epileptic Encephalopathy with Contractures,
800 Macrocephaly, and Dyskinesias. Am J Hum Genet. 2018 Nov;103(5):666–78. Available from:
801 https://linkinghub.elsevier.com/retrieve/pii/S0002929718303173

802 65. Takata A, Ionita-Laza I, Gogos JA, Xu B, Karayiorgou M. De Novo Synonymous Mutations in
803 Regulatory Elements Contribute to the Genetic Etiology of Autism and Schizophrenia. Neuron.
804 2016 Mar 2;89(5):940–7.

805 66. Ganna A, Satterstrom FK, Zekavat SM, Das I, Kurki MI, Churchhouse C, et al. Quantifying the
806 Impact of Rare and Ultra-rare Coding Variation across the Phenotypic Spectrum. Am J Hum Genet.
807 2018 Jun 7;102(6):1204–11.

808 67. Takata A, Nakashima M, Saitsu H, Mizuguchi T, Mitsuhashi S, Takahashi Y, et al.
809 Comprehensive analysis of coding variants highlights genetic complexity in developmental and
810 epileptic encephalopathy. Nat Commun. 2019 Dec;10(1):2506. Available from:
811 http://www.nature.com/articles/s41467-019-10482-9

812 68. Leu C, Stevelink R, Smith AW, Goleva SB, Kanai M, Ferguson L, et al. Polygenic burden in
813 focal and generalized epilepsies. Brain. 2019 Nov 1;142(11):3473–81. Available from:
814 https://academic.oup.com/brain/article/142/11/3473/5585821

815 69. Scheffer IE, Berkovic S, Capovilla G, Connolly MB, French J, Guilhoto L, et al. ILAE
816 classification of the epilepsies: Position paper of the ILAE Commission for Classification and
817 Terminology. Epilepsia. 2017 Apr;58(4):512–21. Available from:
818 http://doi.wiley.com/10.1111/epi.13709

819 70. Heyne HO, Singh T, Stamberger H, Abou Jamra R, Caglayan H, Craiu D, et al. De novo variants
820 in neurodevelopmental disorders with epilepsy. Nat Genet. 2018 Jul;50(7):1048–53.

821 71. EuroEPINOMICS-RES Consortium, Epilepsy Phenome/Genome Project, Epi4K Consortium. De
822 Novo Mutations in Synaptic Transmission Genes Including DNM1 Cause Epileptic
823 Encephalopathies. Am J Hum Genet. 2014 Oct;95(4):360–70. Available from:
824 https://linkinghub.elsevier.com/retrieve/pii/S0002929714003838

825 72. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, et al. Synaptic,
826 transcriptional and chromatin genes disrupted in autism. Nature. 2014 Nov;515(7526):209–15.
827 Available from: http://www.nature.com/articles/nature13772

828 73. The Brainstorm Consortium. Analysis of shared heritability in common disorders of the brain.
829 Science. 2018 Jun 22;360(6395):eaap8757. Available from:
830 https://www.sciencemag.org/lookup/doi/10.1126/science.aap8757

831 74. Johannesen KM, Gardella E, Linnankivi T, Courage C, de Saint Martin A, Lehesjoki A-E, et al.
832 Defining the phenotypic spectrum of *SLC6A1* mutations. Epilepsia. 2018 Feb;59(2):389–402.
833 Available from: http://doi.wiley.com/10.1111/epi.13986

834 75. Liu Y, Schubert J, Sonnenberg L, Helbig KL, Hoei-Hansen CE, Koko M, et al. Neuronal
835 mechanisms of mutations in *SCN8A* causing epilepsy or intellectual disability. Brain. 2019 Feb
836 1;142(2):376–90. Available from: https://academic.oup.com/brain/article/142/2/376/5277044

837 76. Wolking S, May P, Mei D, Møller RS, Balestrini S, Helbig KL, et al. Clinical spectrum of *STX1B*
838 -related epileptic disorders. Neurology. 2019 Feb 8;10.1212/WNL.0000000000007089. Available
839 from: http://www.neurology.org/lookup/doi/10.1212/WNL.0000000000007089

840 77. Vardar G, Gerth F, Schmitt XJ, Rautenstrauch P, Trimbuch T, Schubert J, et al. Epilepsy-causing
841 STX1B mutations translate altered protein functions into distinct phenotypes in mouse neurons.
842 Brain. 2020 Jul 1;143(7):2119–38. Available from:
843 https://academic.oup.com/brain/article/143/7/2119/5861027

844 78. Wang J, Lin Z-J, Liu L, Xu H-Q, Shi Y-W, Yi Y-H, et al. Epilepsy-associated genes. Seizure.
845 2017 Jan;44:11–20. Available from:
846 https://linkinghub.elsevier.com/retrieve/pii/S1059131116302989

847 79. de Leeuw CA, Neale BM, Heskes T, Posthuma D. The statistical properties of gene-set analysis.
848 Nat Rev Genet. 2016 Jun;17(6):353–64. Available from:
849 http://www.nature.com/articles/nrg.2016.29

850 80. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fenell T, et al. Analysis of protein-
851 coding genetic variation in 60,706 humans. Nature. 2016 Aug;536(7616):285–91. Available from:
852 http://www.nature.com/articles/nature19057

853 **Table (1): Gene-sets investigated in this study.**

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:***	***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).	***Cell-type-specific enrichment:*** -Neurons -glial cells -Excitatory neurons -Inhibitory neurons -Astrocytes -Microglia -Oligodendrocytes -Endothelium. ***Neuronal Localization:*** -Axon Initial Segment.

excluding genes with no expression in the cortex/hippocampus -Z-score > 3.09. -Z-score 2.5-3.09. -Z-score 2-2.5.	-Brain developmental genes (extended group). -Early developmental genes. -Late developmental genes. ***Enrichment in the brain:*** -Brain-enriched -Brain-enhanced.	-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. -$GABA_A$ receptors. -GABAergic pathway. -Excitatory receptors. -NMDAR & ARC. -PSD-95 interactors. ***GABAergic/Glutamatergic synapses (GO groups):*** -GABAergic synapse -Glutamatergic synapse -only in GABAergic -only in glutamatergic -shared genes.	***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.
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.	***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.
Control groups (15):		
Genes not expressed in the brain:	***KEGG metabolic pathways:*** -Type II Diabetes. -Carbohydrate Absorption & Digestion.	***KEGG cancer pathways:*** - CA Breast, CA Lung, CA Colon, CA Prostate, Renal

-RNA not detected in cortex, in hippocampus, or all GTEx regions. -Protein is depleted in the brain.	-Protein Absorption & Digestion. -Fat Absorption & Digestion.	Cell Ca, CA Pancreas, Hepatocellular Ca.

854

855

856 **Figure legends:**

857 **Fig. 1: Outlines of the burden analysis method.** Thirteen (synonymous and twelve functional)
858 variants models with focus on missense variants in constrained or paralog-conserved sites were tested
859 in the three epilepsy phenotypes against a shared set of matched controls. The burden was examined in
860 92 gene-sets (detailed in table 1) using a logistic regression model with the count of qualifying variants
861 per sample as a predictor and sample sex, ten principal components, singletons and exome-wide variant
862 counts as covariates. Two secondary analyses were performed: one analysis restricting gene-sets to
863 autosomal genes (to exclude bias introduced by male-to-female ratio imbalances) and another analysis
864 testing controls prepared for exome sequencing using Illumina ICE capture kits against controls
865 prepared with Agilent SureSelect capture kits (to exclude bias caused by differences in enrichment
866 kits).

867 **Fig. 2: Exome-wide burden of ultra-rare variants in the epilepsies.** The burden in developmental
868 and epileptic encephalopathies (DEE), genetic generalized epilepsies (GGE) and non-acquired focal
869 epilepsies (NAFE) in 19,402 protein coding genes **(A)** and 1,743 loss-of-function intolerant genes with
870 pLI score > 0.995 **(B)** is shown in multiple test and control conditions (y-axis) as odds ratio (x-axis)
871 from Likelihood Ratio Test (bars indicate 95% confidence intervals). FDR corrected *p* values (not
872 corrected for synonymous variants) are indicated with stars as follows: no star > 0.05, * < 0.05, ** <
873 0.005, *** < 0.005, **** < 0.0005. There is an incrementing burden with a higher level of missense
874 constraint. Both synonymous and benign missense alterations are not enriched, suggesting sufficient
875 control for inflation.

876 **Fig. 3: Burden of ultra-rare missense variants in brain expressed and developmental genes.** The
877 burden of benign or damaging missense variants and missense variants in highly paralog-conserved or
878 highly constrained sites in developmental and epileptic encephalopathies (DEE), genetic generalized

879 epilepsies (GGE) and non-acquired focal epilepsies (NAFE) is shown in gene-sets based on levels of
880 RNA/protein expression in the cortex and hippocampus **(A)** or enrichment in adult or developing brain
881 **(B)**. Gene-sets are shown on the y-axis (number of genes between brackets). Log odds ratio (Likelihood
882 Ratio Test) are shown on the x-axis (error bars indicate 95% confidence intervals). The variant
883 conditions are shown in vertical panels. FDR corrected *p* values (not corrected for synonymous variants)
884 are indicated with stars as follows: no star > 0.05, * < 0.05, ** < 0.005, *** < 0.005, **** < 0.0005.
885 Brain enriched genes (Human Protein Atlas: more than four-fold expression compared to other tissues)
886 show higher burden in DEE compared to those genes with only enhanced expression (higher but less
887 than four-fold expression in the brain). This difference is less prominent in GGE and NAFE.
888 Developmental genes enriched in late development show higher burden than genes enriched in early
889 development.

890 **Fig. 4: Burden in neuronal and glial cells, ion channels, receptors and related interactors.** The
891 burden in developmental and epileptic encephalopathies (DEE), genetic generalized epilepsies (GGE)
892 and non-acquired focal epilepsies (NAFE) in shown on the x-axis (log-odds from Likelihood Ratio
893 Test; error bars indicate 95% confidence intervals). Gene-sets are shown on the y-axis (number of genes
894 between brackets). The variant conditions are shown in vertical panels. FDR corrected *p* values (not
895 corrected for synonymous variants) are indicated with stars as follows: no star > 0.05, * < 0.05, ** <
896 0.005, *** < 0.005, **** < 0.0005. **(A)** Comparisons based on the cellular expression patterns show
897 that the enrichment is prominent in neuron-enriched compared to glial genes, particularly in GGE. GGE
898 shows relatively higher enrichment in inhibitory neuron genes while NAFE shows a relatively higher
899 burden in excitatory neuronal genes. **(B)** Variants in $GABA_A$ receptors and GABAergic pathway are
900 preferentially enriched in GGE compared to groups of genes coding for NMDA receptor and neuronal
901 activity-regulated cytoskeleton-associated protein (ARC). In NAFE, the variants are enriched in the
902 NMDA receptor & ARC gene-sets, but not $GABA_A$ receptors.

903 **Fig. 5: Enrichment in major neuronal synapses and pathways.** Comparison of enrichment patterns
904 in GABAergic and glutamatergic synapses and pathway genes based on Gene-Ontology (GO) **(A)** or
905 Kyoto Encyclopedia for Genes and Genomes (KEGG) **(B)** The burden in developmental and epileptic

906 encephalopathies (DEE), genetic generalized epilepsies (GGE) and non-acquired focal epilepsies
907 (NAFE) in shown on the x-axis (log-odds from Likelihood Ratio Test; error bars indicate 95%
908 confidence intervals). Gene-sets are shown on the y-axis (number of genes between brackets). The
909 variant conditions are shown in vertical panels. FDR corrected *p* values (not corrected for synonymous
910 variants) are indicated with stars as follows: no star > 0.05, * < 0.05, ** < 0.005, *** < 0.005, **** <
911 0.0005. Complete groups, genes specific to one of the two synapses/pathways as well as their
912 intersection were tested. Both GO and KEGG gene sets show an increased burden in GGE when
913 "GABAergic only" genes are tested. GGE shows a relatively higher burden in GABAergic compared
914 to glutamatergic gene sets.

915 **Fig. 6: Burden of ultra-rare variants in groups of epilepsy-related known disease genes.** The
916 burden in five gene-sets (y-axis; number of genes between brackets) in developmental and epileptic
917 encephalopathies (DEE), genetic generalized epilepsies (GGE) and non-acquired focal epilepsies
918 (NAFE) (horizontal panel) in selected variant conditions (vertical panels) is shown on the x-axis (log
919 odd ratios from Likelihood Ratio Test; error bars indicate 95% confidence intervals). FDR corrected *p*
920 values (not corrected for synonymous variants) are indicated with stars as follows: no star > 0.05, * <
921 0.05, ** < 0.005, *** < 0.005, **** < 0.0005. NDD-Epilepsy: neurodevelopmental disorders with
922 epilepsy. FMPR: Fragile-X Mental Retardation Protein targets. MGI: Mouse Genome Informatics
923 database.

924 Fig. 7: **Risk elements in GWAS top-ranking genes and co-expression modules.** The burden of
925 missense variants in highly constrained sites (log-odds on the x-axis; error bars indicate 95% confidence
926 intervals) in developmental and epileptic encephalopathies (DEE), genetic generalized epilepsies
927 (GGE) and non-acquired focal epilepsies (NAFE) is shown in monogenic epilepsy genes, top-ranking
928 epilepsy GWAS genes, brain co-expression modules and TLE-related co-expression modules (y-axis;
929 number of genes between brackets). FDR corrected p values (not corrected for synonymous variants)
930 are indicated with stars as follows: no star > 0.05, * < 0.05, ** < 0.005, *** < 0.005, **** < 0.0005.
931 (A) Generalized or focal epilepsies (presumed monogenic) as well as top-ranking 100 genes from
932 GWAS of generalized and focal epilepsies are preferentially enriched for constrained missense variants

933 (CCR 80) in respective phenotypic cohorts indicating a possible convergence between common and
934 rare variants in GWAS genes. (B) Enrichment in co-expressed genes identified in post-mortem brain
935 tissues of healthy individuals (module of 320 genes) or in brain tissues from TLE patients (network of
936 395 genes) as well as two sub-modules of this network (M1 and M2).

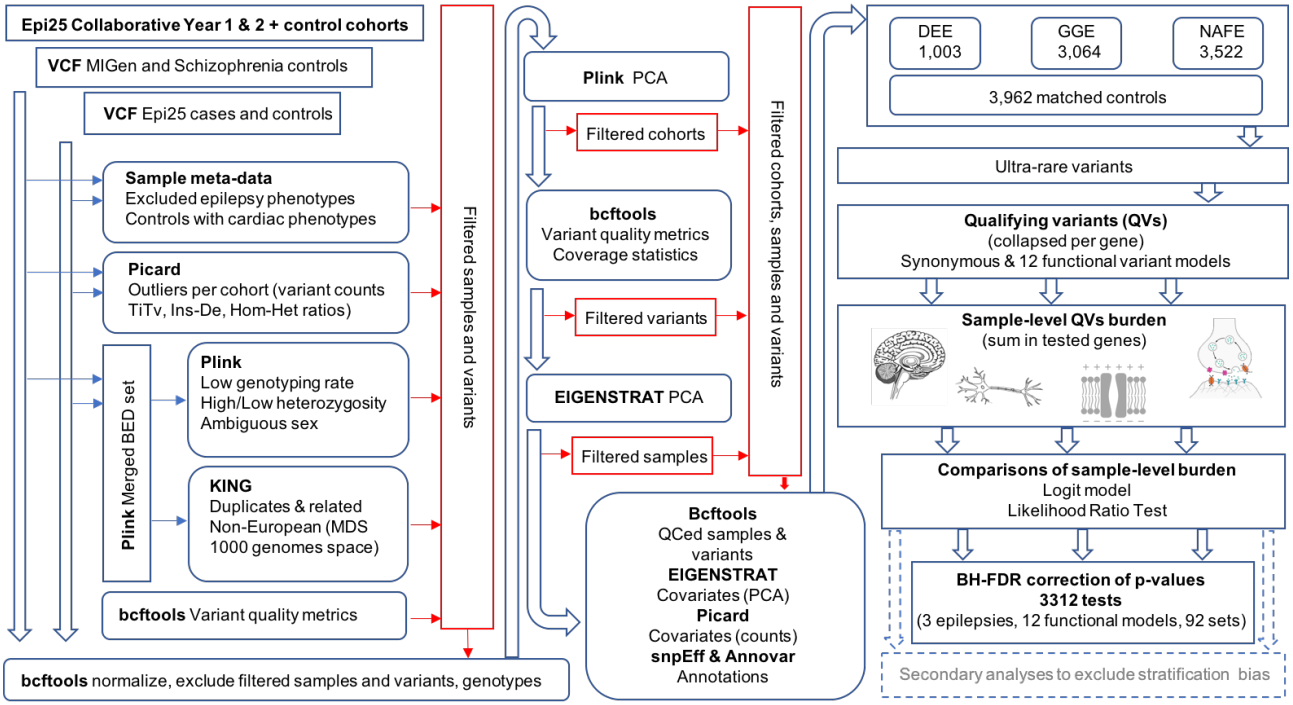

Fig. 1: Outlines of the burden analysis method. Thirteen (synonymous and twelve functional) variants models with focus on missense variants in constrained or paralog-conserved sites were tested in the three epilepsy phenotypes against a shared set of matched controls. The burden was examined in 92 gene groups (detailed in table 1) using a logistic regression model with the count of qualifying variants per sample as a predictor and sample sex, ten principal components, singletons and exome-wide variant counts as covariates. Two secondary analyses were performed: one analysis restricting gene groups to autosomal genes (to exclude bias introduced by male-to-female ratio imbalances) and another analysis testing controls prepared for exome sequencing using Illumina ICE capture kits against controls prepared with Agilent SureSelect capture kits (to exclude bias caused by differences in enrichment kits).

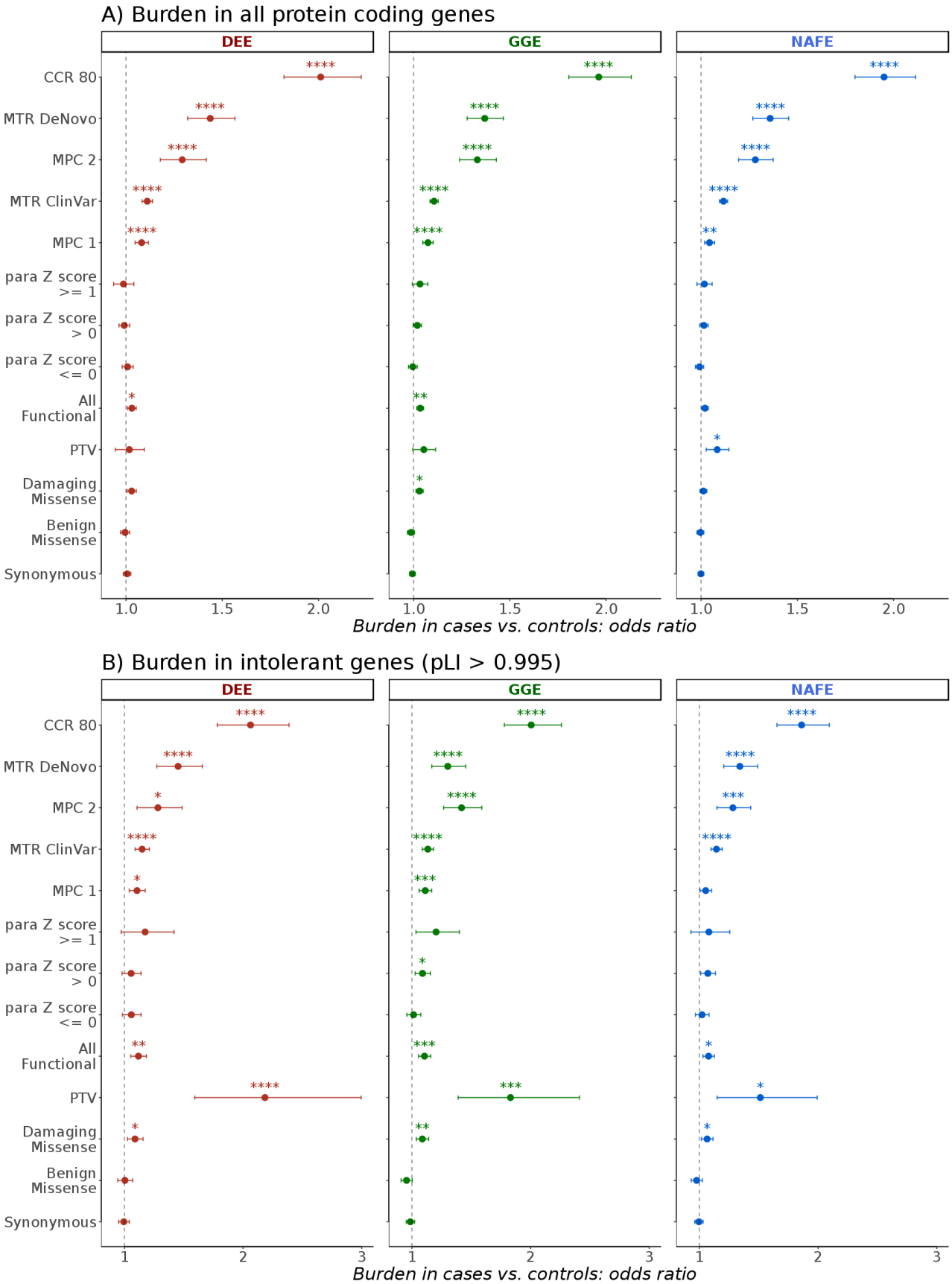

Fig. 2: Exome-wide burden of ultra-rare variants in the epilepsies. The burden in developmental and epileptic encephalopathies (DEE), genetic generalized epilepsies (GGE) and non-acquired focal epilepsies (NAFE) in 19,402 protein coding genes **(A)** and 1,743 loss-of-function intolerant genes with pLI score > 0.995 **(B)** is shown in multiple test and control conditions (y-axis) as odds ratio (x-axis) from Likelihood Ratio Test (bars indicate 95% confidence intervals). FDR corrected *p* values (not corrected for synonymous variants) are indicated with stars as follows: no star > 0.05, * < 0.05, ** < 0.005, *** < 0.005, **** < 0.0005. There is an incrementing burden with higher level of missense constraint. Both synonymous and benign missense alterations are not enriched, suggesting sufficient control for inflation.

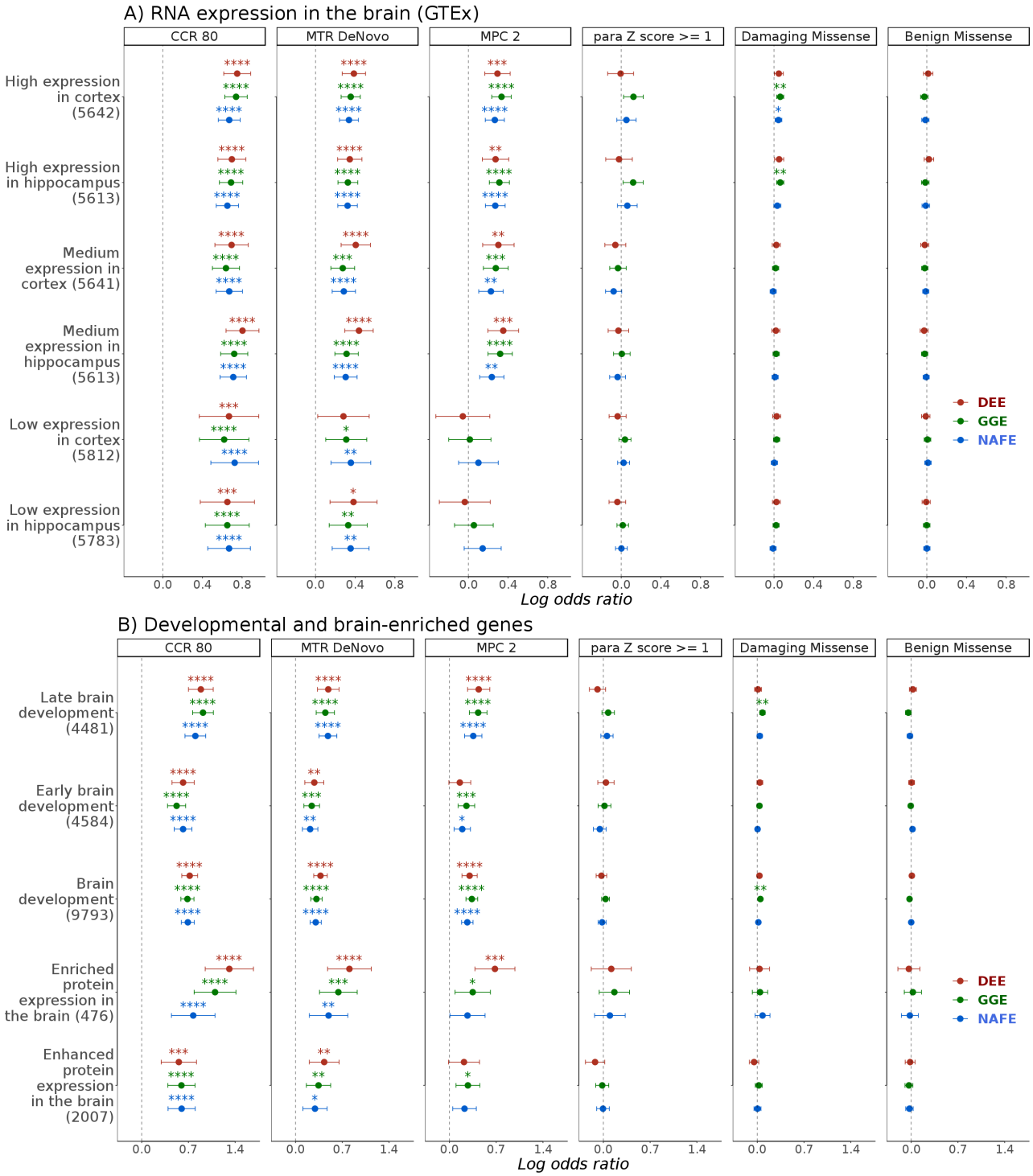

Fig. 3: Burden of ultra-rare missense variants in brain expressed and developmental genes. The burden of benign or damaging missense variants and missense variants in highly paralogue-conserved or highly constrained sites in developmental and epileptic encephalopathies (DEE), genetic generalized epilepsies (GGE) and non-acquired focal epilepsies (NAFE) is shown in gene groups based on levels of RNA/protein expression in the cortex and hippocampus **(A)** or enrichment in adult or developing brain **(B)**. Gene groups are shown on the y-axis (number of genes between brackets). Log odds ratio (Likelihood Ratio Test) are shown on the x-axis (error bars indicate 95% confidence intervals). The variant conditions are shown in vertical panels. FDR corrected *p* values (not corrected for synonymous variants) are indicated with stars as follows: no star > 0.05, * < 0.05, ** < 0.005, *** < 0.005, **** < 0.0005. Brain enriched genes (Human Protein Atlas: more than four-fold expression compared to other tissues) show higher burden in DEE compared to those genes with only enhanced expression (higher but less than four-fold expression in the brain). This difference is less prominent in GGE and NAFE. Developmental genes enriched in late development show higher burden than genes enriched in early development.

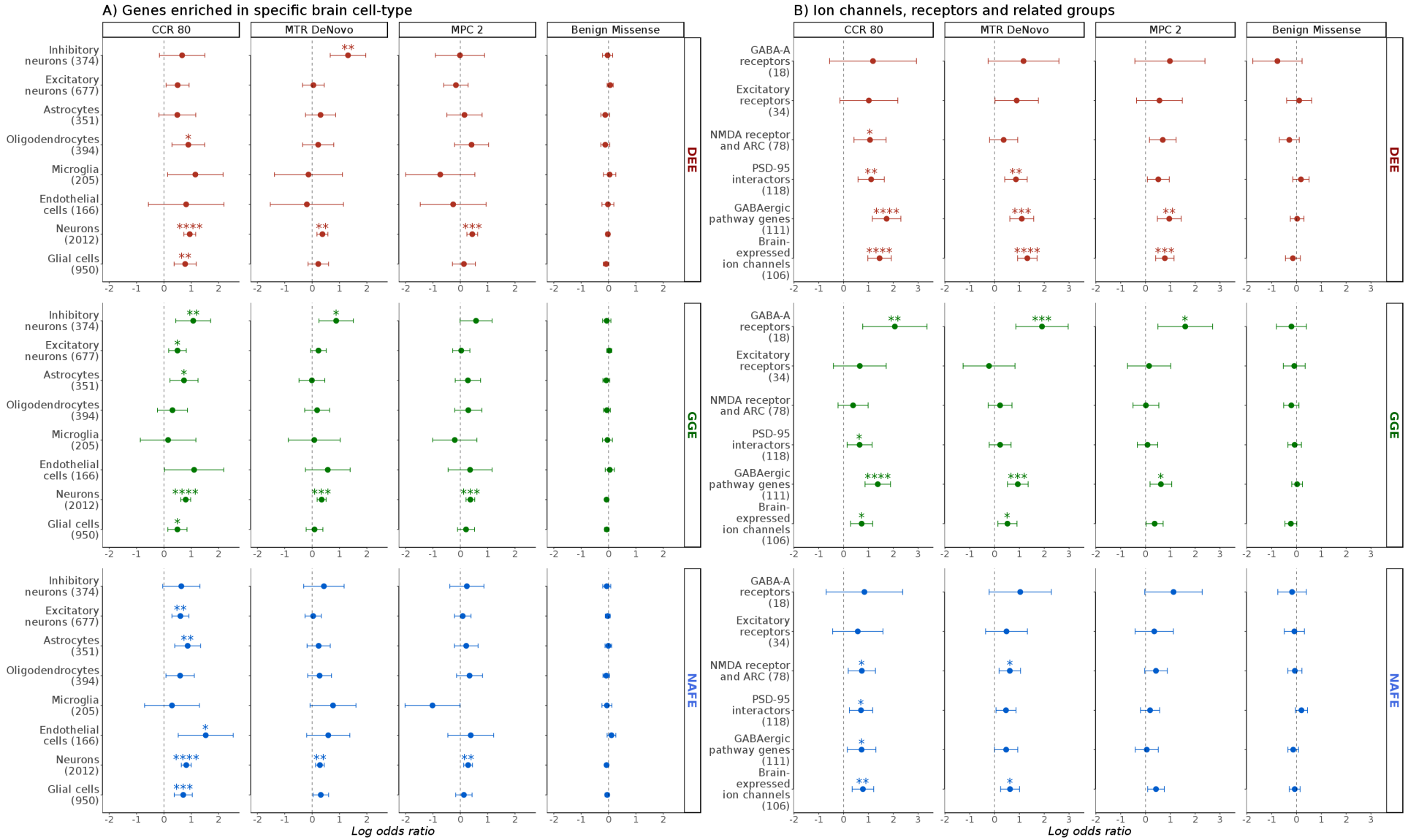

Fig. 4: Burden in neuronal and glial cells, ion channels, receptors and related interactors. The burden in developmental and epileptic encephalopathies (DEE), genetic generalized epilepsies (GGE) and non-acquired focal epilepsies (NAFE) in shown on the x-axis (log-odds from Likelihood Ratio Test; error bars indicate 95% confidence intervals). Gene groups are shown on the y-axis (number of genes between brackets). The variant conditions are shown in vertical panels. FDR corrected *p* values (not corrected for synonymous variants) are indicated with stars as follows: no star > 0.05, * < 0.05, ** < 0.005, *** < 0.005, **** < 0.0005. **(A)** Comparisons based on the cellular expression patterns show that the enrichment is prominent in neuron-enriched compared to glial genes, particularly in GGE. GGE show relatively higher enrichment in inhibitory neuron genes while NAFE show a relatively higher burden in excitatory neuronal genes. **(B)** Variants in $GABA_A$ receptors and GABAergic pathway are preferentially enriched in GGE compared to groups of genes coding for NMDA receptor and neuronal activity-regulated cytoskeleton-associated protein (ARC). In NAFE, the variants are enriched in NMDA receptor & ARC gene group, but not $GABA_A$ receptors.

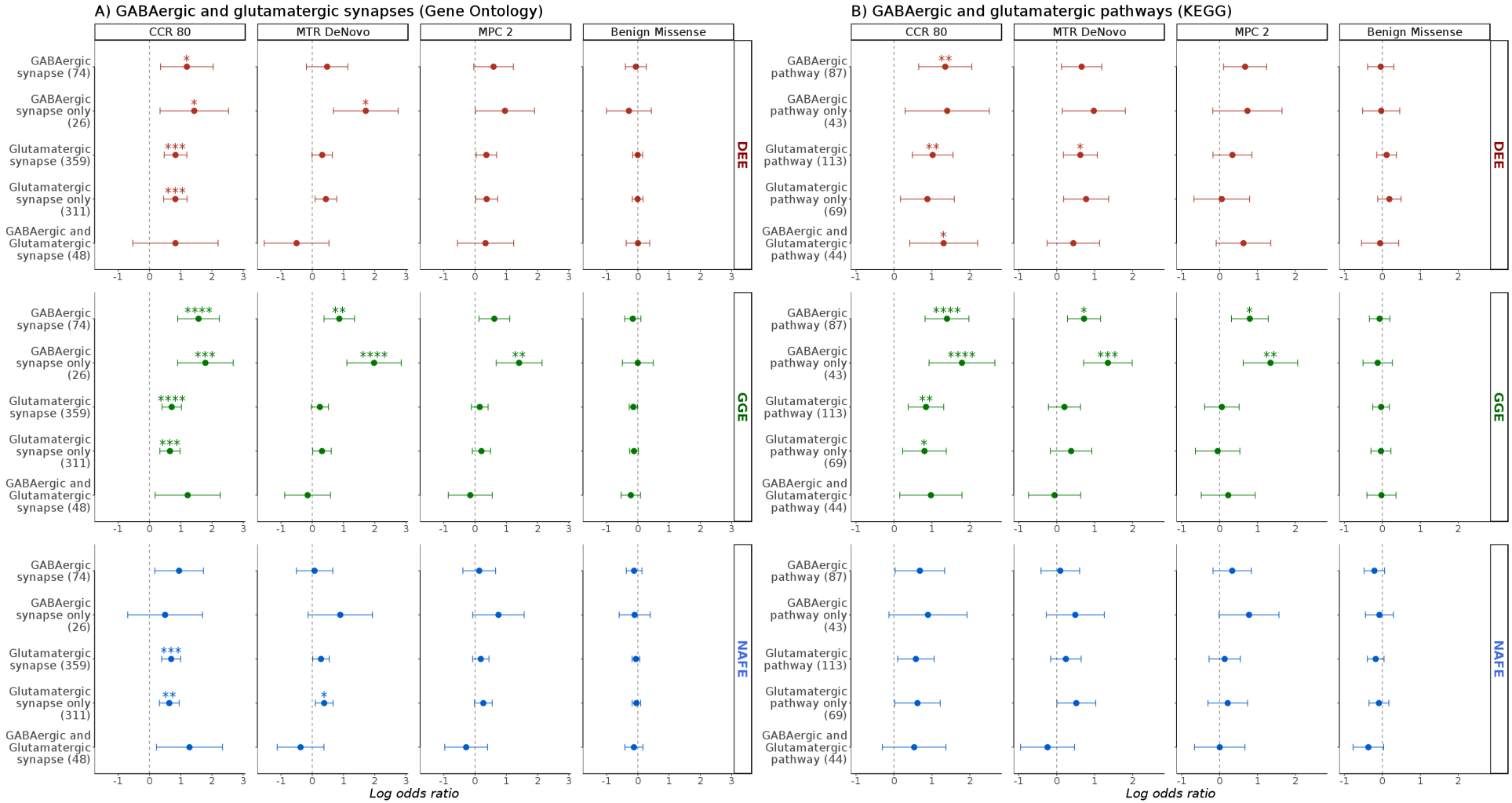

Fig. 5: Enrichment in major neuronal synapses and pathways. Comparison of enrichment patterns in GABAergic and glutamatergic synapses and pathway genes based on Gene-Ontology (GO) **(A)** or Kyoto Encyclopedia for Genes and Genomes (KEGG) **(B)** The burden in developmental and epileptic encephalopathies (DEE), genetic generalized epilepsies (GGE) and non-acquired focal epilepsies (NAFE) in shown on the x-axis (log-odds from Likelihood Ratio Test; error bars indicate 95% confidence intervals). Gene groups are shown on the y-axis (number of genes between brackets). The variant conditions are shown in vertical panels. FDR corrected *p* values (not corrected for synonymous variants) are indicated with stars as follows: no star > 0.05, * < 0.05, ** < 0.005, *** < 0.005, **** < 0.0005. Complete groups, genes specific to one of the two synapses/pathways as well as their intersection were tested. Both GO and KEGG gene sets show an increased burden in in GGE when "GABAergic only" genes are tested. GGE show a relatively higher burden in GABAergic compared to glutamatergic gene sets.

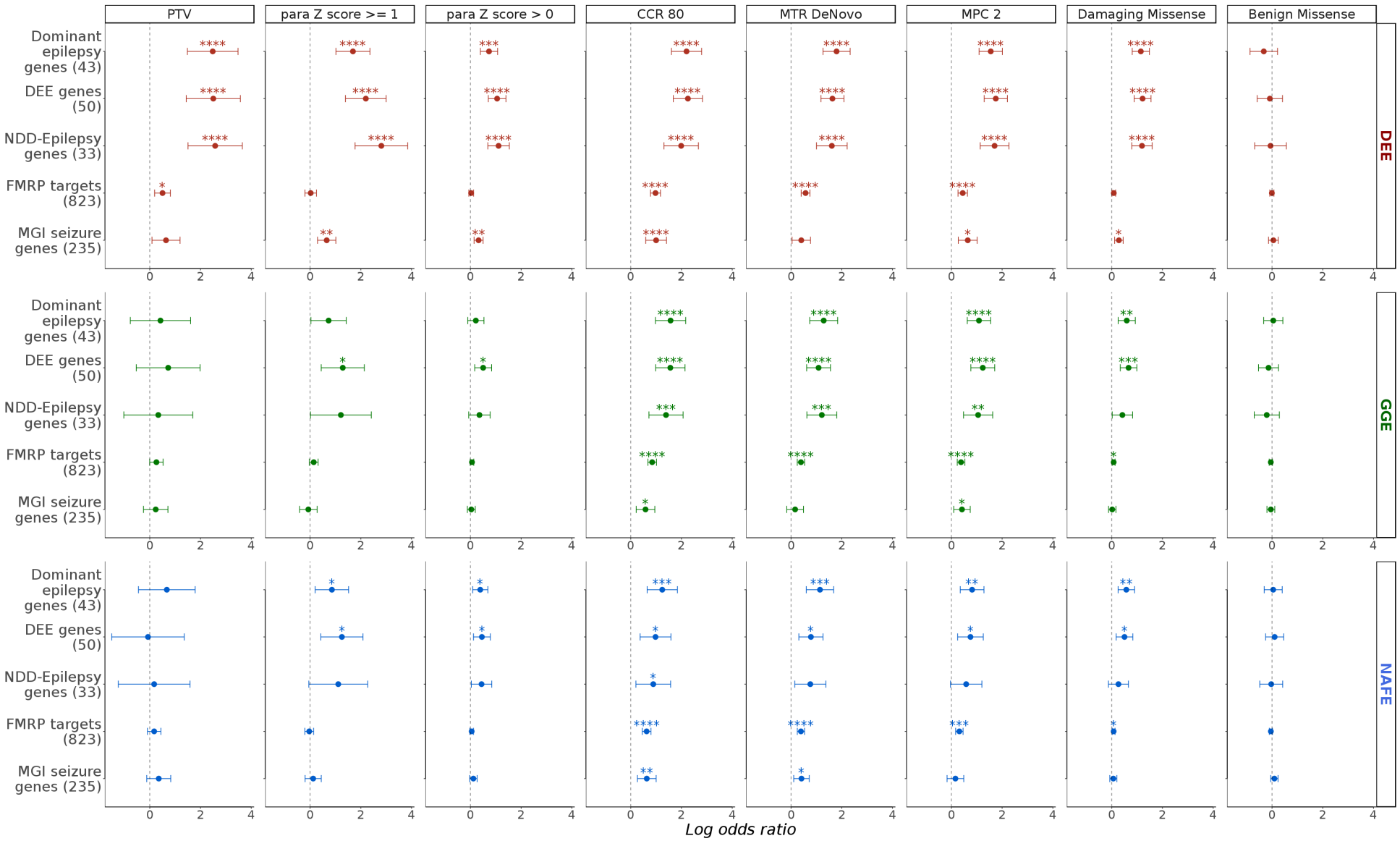

Fig. 6: Burden of ultra-rare variants in groups of epilepsy-related known disease genes. The burden in five gene groups (y-axis; number of genes between brackets) in developmental and epileptic encephalopathies (DEE), genetic generalized epilepsies (GGE) and non-acquired focal epilepsies (NAFE) (horizontal panel) in selected variant conditions (vertical panels) is shown on the x-axis (log odd ratios from Likelihood Ratio Test; error bars indicate 95% confidence intervals). FDR corrected *p* values (not corrected for synonymous variants) are indicated with stars as follows: no star > 0.05, * < 0.05, ** < 0.005, *** < 0.005, **** < 0.0005. NDD-Epilepsy: neurodevelopmental disorders with epilepsy. FMPR: Fragile-X Mental Retardation Protein targets. MGI: Mouse Genome Informatics database.

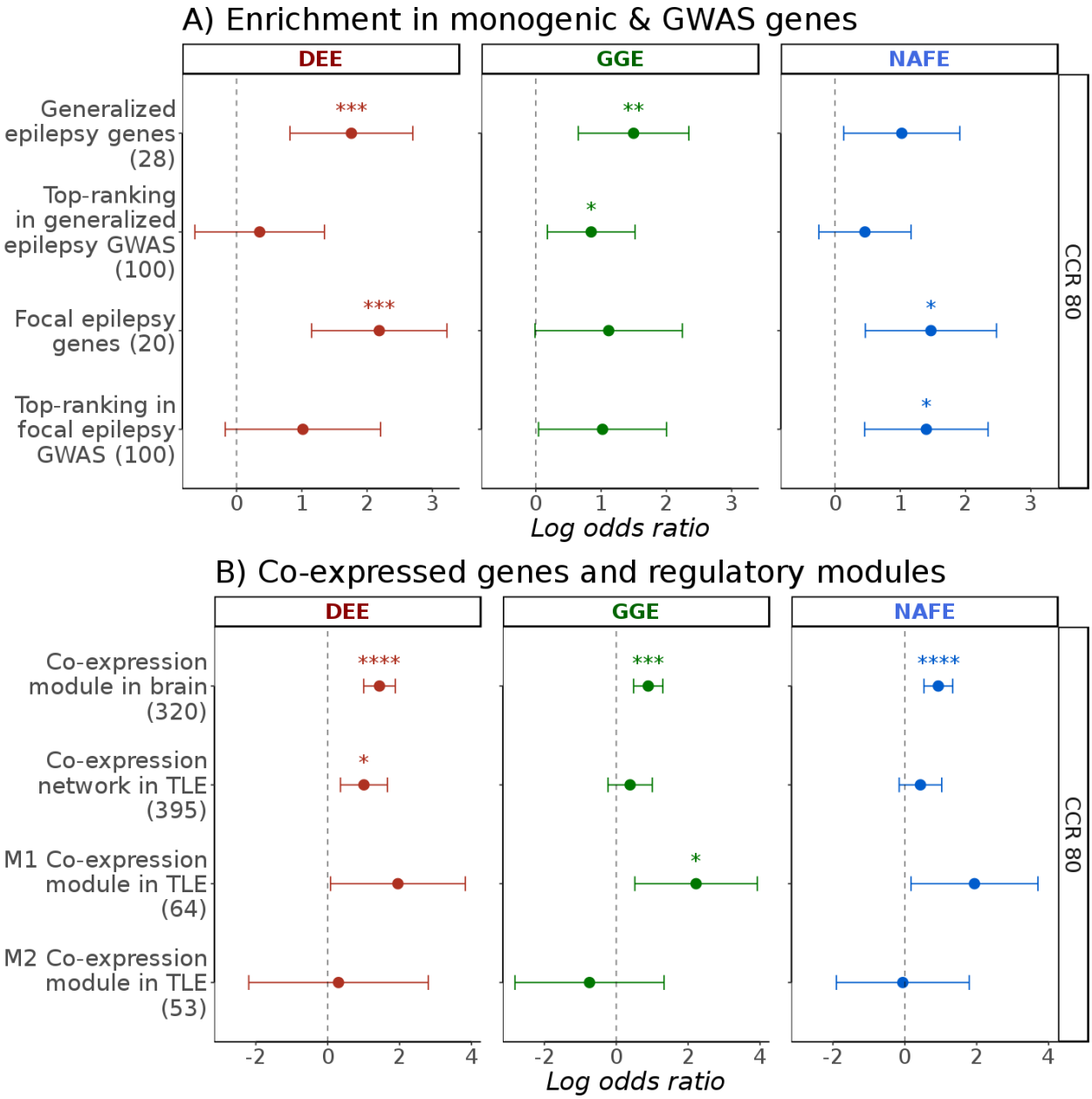

Fig. 7: Risk elements in GWAS top-ranking genes and co-expression modules. The burden of missense variants in highly constrained sites (log-odds on the x-axis; error bars indicate 95% confidence intervals) in developmental and epileptic encephalopathies (DEE), genetic generalized epilepsies (GGE) and non-acquired focal epilepsies (NAFE) is shown in monogenic epilepsy genes, top-ranking epilepsy GWAS genes, brain co-expression modules and TLE-related co-expression modules (y-axis; number of genes between brackets). FDR corrected *p* values (not corrected for synonymous variants) are indicated with stars as follows: no star > 0.05, * < 0.05, ** < 0.005, *** < 0.005, **** < 0.0005. **(A)** Generalized or focal epilepsies (presumed monogenic) as well as top-ranking 100 genes from GWAS of generalized and focal epilepsies are preferentially enriched for constrained missense variants (CCR 80) in respective phenotypic cohorts indicating a possible convergence between common and rare variants in GWAS genes. **(B)** Enrichment in 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).