

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:<https://orca.cardiff.ac.uk/id/eprint/98346/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Thomas, Rhys 2017. Ultra-rare genetic variation in common epilepsies: a case-control sequencing study. *Lancet Neurology* 16 (2) , pp. 135-143. 10.1016/S1474-4422(16)30359-3

Publishers page: [http://dx.doi.org/10.1016/S1474-4422\(16\)30359-3](http://dx.doi.org/10.1016/S1474-4422(16)30359-3)

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.





25 We separately compared the sequence data from 640 individuals with familial GGE and 525  
26 individuals with familial NAFE to the same group of 3,877 controls, and found significant  
27 excess of ultra-rare deleterious variation in genes established as causative for dominant  
28 epilepsy disorders (GGE: OR 2.3 [95% CI 1.7–3.2];  $p=9.1\times 10^{-8}$ ) (NAFE: OR 3.6 [95% CI  
29 2.7–4.9];  $p=1.1\times 10^{-17}$ ). Comparing an additional collection of 662 individuals with sporadic  
30 NAFE to controls did not identify study-wide significant signals. For the familial NAFE  
31 cases, we found that five previously known epilepsy genes ranked as the top five genes  
32 enriched for ultra-rare deleterious variation. After accounting for the control carrier rate we  
33 estimate that these five genes contribute to the risk of epilepsy in approximately 8% of  
34 familial NAFE cases. While no individual gene showed study-wide significance in the  
35 familial GGE analyses, known epilepsy genes showed a significant excess ( $p=5.8\times 10^{-8}$ ) of p-  
36 values that were lower than expected from a random sampling of genes.

## 37 **INTERPRETATION**

38 We identified excess ultra-rare variation in known epilepsy genes, which establishes a clear  
39 connection between the genetics of common and rare severe epilepsies, and shows that the  
40 variants responsible for the observed epilepsy risk signal are exceptionally rare in the general  
41 population. Our results suggest that the emerging paradigm of targeting treatments to the  
42 genetic cause in rare devastating epilepsies may also extend to a proportion of common  
43 epilepsies. These findings might allow clinicians to broadly explain the aetiology of these  
44 syndromes to patients, and lay the foundation for possible precision treatments in the future.

## 45 **FUNDING**

46 National Institute of Neurological Disorders and Stroke (NINDS), Epilepsy-Research UK.

47

## INTRODUCTION

48 Next generation sequencing has proven successful in identifying genetic contributions to rare  
49 Mendelian disorders and cancers,<sup>1,2</sup> creating widespread optimism that treatments can be  
50 targeted to underlying causes of disease.<sup>3</sup> Although epilepsy is a common complex disease, it  
51 is emerging as a group of disorders with precision medicine opportunities similar to those in  
52 rare Mendelian disorders and cancers.<sup>4</sup> Unlike many common diseases, epilepsy genetics  
53 research is identifying not only the genes responsible, but also the genetic variants  
54 contributing to disease in individual patients. This is most apparent in the role of *de novo*  
55 mutations in the epileptic encephalopathies.<sup>5,6</sup>

56 Traditional heritability studies of the common epilepsies consistently show strong genetic  
57 effects in non-acquired focal epilepsy (NAFE) and in genetic generalized epilepsy (GGE),  
58 with both shared and distinct genetic contributions to these broadly defined epilepsies.<sup>7,8</sup>  
59 Two important unresolved questions are the extent to which the genes responsible for rare  
60 severe epilepsies contribute to common epilepsies, and whether, as in the rare epilepsies,  
61 genetic risk arises primarily from ultra-rare variants of large effect including *de novo*  
62 mutations,<sup>5,6</sup> or from a constellation of common variants each conferring small or modest  
63 effect.<sup>9-13</sup>

64 Exome sequencing of large case and control cohorts followed by genome-wide collapsing  
65 analyses provide a hypothesis-free approach to discovering novel disease genes and better  
66 understanding the overall contribution of ultra-rare genetic variation to disease.<sup>14</sup> Here, we  
67 assess the contribution of ultra-rare genetic variation to common epilepsies while controlling  
68 for background variation in the general population.

## METHODS

69

### 70 **Participants**

71 For this case-control study, participants with familial or sporadic NAFE or familial GGE  
72 were recruited between November 26, 2007 and August 2, 2013 through the international  
73 Epilepsy Phenome/Genome Project (EPGP) and Epi4K collaborations (appendix), as  
74 previously described.<sup>15</sup> The case samples were sequenced between February 6, 2013 and  
75 August 18, 2015 by the Institute for Genomic Medicine, Columbia University (New York  
76 City, NY, USA). To be clinically classified as having NAFE, patients were required to have  
77 focal seizures and no evidence of an epileptogenic lesion on clinical imaging; however,  
78 hippocampal sclerosis was not considered an exclusion criterion. To be clinically classified as  
79 having GGE, patients were required to have a diagnosis of generalized epilepsy with absence,  
80 myoclonic or tonic-clonic seizures and generalized spike-and-wave on an EEG, and no or  
81 mild intellectual disability. All patients were clinically evaluated by their local clinician or  
82 the clinical team at recruiting centres. Individuals with unclassifiable epilepsy or classified as  
83 having both GGE and NAFE were excluded from the analyses.

84 To be classified as a familial case, at least one reported relative (up to third degree) who had  
85 been diagnosed with epilepsy was required. The sporadic NAFE cohort included participants  
86 who self-reported no known epilepsy family history and were recruited from international  
87 hospital, outpatient, and epilepsy clinics (appendix).<sup>15, 16</sup> Written informed consent was  
88 collected at the time of recruitment at each of the clinical sites. Patient collection and sharing  
89 of anonymised specimens for research was approved by site-specific Institutional Review  
90 Boards and ethic committees.

91 The control cohort comprised of unrelated individuals of European ancestry that had been  
92 selected for control purposes and sequenced through unrelated studies not focused on  
93 neurodevelopmental, neuropsychiatric or severe paediatric disease (appendix).

94 **Procedures**

95 Sequencing was performed at the Institute for Genomic Medicine, Columbia University (New  
96 York City, NY, USA). Samples were exome sequenced using the Agilent All Exon (50MB or  
97 65MB; Agilent Technologies, Santa Clara, CA, USA) or the Nimblegen SeqCap EZ V2.0 or  
98 3.0 Exome Enrichment kit (Roche NimbleGen, Madison, WI, USA) or whole genome  
99 sequenced using HiSeq 2000 or 2500 (Illumina, San Diego, CA, USA) sequencers according  
100 to standard protocols.

101 The sequence data from patients with epilepsy and controls were processed using the same  
102 Institute for Genomic Medicine bioinformatics pipeline (appendix). We focused on 18,668  
103 consensus coding sequence (CCDS; release 14) protein-coding genes. On average, at least  
104 10-fold coverage was achieved for 95.8% (familial GGE), 96.8% (familial NAFE), 97.1%  
105 (sporadic NAFE) and 95.6% (controls) of the 33.27 Mbps of the CCDS. For each protein-  
106 coding site in the CCDS—inclusive of two base intronic extensions to accommodate  
107 canonical splice variants—we determined the percentage of cases and controls that had  $\geq 10$ -  
108 fold coverage at the site. To alleviate confounding due to differential coverage we used a site-  
109 based pruning strategy similar to our previously described exon-pruning strategy.<sup>17</sup> Individual  
110 CCDS sites were excluded from analysis if the absolute difference in the percentage of the  
111 cases compared to controls with adequate coverage of the site differed by greater than 5.19%  
112 (familial GGE vs. controls), 5.14% (familial NAFE vs. controls) and 6.39% (sporadic NAFE  
113 vs. controls) (appendix). Site-based pruning resulted in 8.9% (GGE), 8.3% (familial NAFE)  
114 and 8.3% (sporadic NAFE) of the CCDS bases excluded from the respective analyses to  
115 alleviate issues from differential coverage. Thus, all gene tests were performed on the pruned  
116 CCDS where cases and controls had similar opportunity to call gene variants (appendix).

117

## 118 STATISTICAL ANALYSIS

119 To search for genes that confer risk for common epilepsy syndromes, we implemented a  
120 genic collapsing analysis,<sup>17</sup> in which only a single affected individual (the index case) from  
121 each family was included. We applied standard procedures to address potential bias due to  
122 relatedness and population stratification (appendix). The analyses focused on CCDS protein-  
123 coding sites with minimal variability in coverage between the case and control populations.

124 As in our earlier work,<sup>17</sup> the term “qualifying variants” has been adopted to refer to the subset  
125 of variation within the sequence data that meets specific criteria designed to enrich for  
126 pathogenic variants. We defined qualifying variants in four ways (Table 1). Our primary  
127 analysis focused on ultra-rare variants where a combination of internal (the test samples) and  
128 external data (the Exome Variant Server [EVS]<sup>18</sup> and Exome Aggregate Consortium [ExAC;  
129 release 0.3]<sup>19</sup>). The test cohort was used to identify variants with a minor allele frequency  
130 (MAF) <0.05% among our combined case and control population being tested. The EVS and  
131 ExAC external databases were used to identify variants found among the test samples and  
132 absent (i.e., MAF=0%) among the two external reference control cohorts. The MAF was set  
133 to <0.05% in the combined case and control test collection to accommodate the possibility of  
134 multiple instances of a risk variant among cases. The two freely available EVS and ExAC  
135 external databases were solely used to support the rarity of the identified variants and did not  
136 contribute as control samples to the tests themselves.

137 For the primary analysis, functional annotation focused on single nucleotide substitution and  
138 insertion or deletion variants annotated as having a loss-of-function, inframe insertion or  
139 deletion, or a “probably damaging” missense effect by PolyPhen-2 (HumDiv).<sup>20</sup> Three  
140 secondary analyses were performed to evaluate the contribution to epilepsy risk from: rare  
141 loss-of-function variants with an internal and external population MAF up to 0.1%; rare non-  
142 synonymous variation in the general population with an internal and external MAF up to

143 0.1%; and a presumed neutral model that imposed similar MAF thresholds as our primary  
144 analysis, but focused specifically on protein-coding variants predicted to have a synonymous  
145 effect. The purpose of the presumed neutral model was to further confirm that no cryptic  
146 factors might be increasing qualifying variant calling among one of the groups.

147 For each of the four models, we tested the complete list of 18,668 CCDS genes. For each  
148 gene, an indicator variable (1/0 states) was assigned to each individual based on the presence  
149 of at least one qualifying variant in the gene (state 1) or no qualifying variants in that gene  
150 (state 0). We used a two-tailed Fisher's exact test to identify genes where there was a  
151 significant enrichment of qualifying variants in the case or control group. To control for the  
152 type-I error rate within each epilepsy phenotype, we defined study-wide significance as  
153  $p=8.9 \times 10^{-7}$ , correcting for 18,668 CCDS genes studied across three models ( $0.05/[3 \times 18668]$ ).  
154 We did not correct for the neutral control model.

155 All collapsing analyses were performed using an in-house package, Analysis Tool for  
156 Annotated Variants ([ATAV](#)). Binomial tests were used to evaluate whether there was an  
157 enrichment of previously reported pathogenic variants among the case collection of  
158 qualifying variants. Hypergeometric tests were performed to assess whether among the  
159 collapsing analysis results the known epilepsy genes preferentially achieved lower p-values  
160 relative to the rest of the genome. Cochran-Mantel-Haenszel tests were adopted to combine  
161 the results of the gender stratified sex chromosome collapsing analyses.

162 We also used the primary analysis results from each of the patient groups to assess  
163 enrichment among six biologically informed gene-sets that were chosen and described in our  
164 earlier studies of the epileptic encephalopathies,<sup>5,21</sup> including a list of 43 established  
165 dominant epilepsy genes (appendix).<sup>3</sup> To account for background variation in gene-set tests  
166 we applied a logistic regression model (appendix).

167 To assess the contribution to epilepsy risk coming from variants with increasing minor allele  
168 frequencies (MAF), we developed a multivariable logistic regression model that focuses on  
169 the known epilepsy genes and relates disease risk to the presence of variants among  
170 increasing MAF bins (appendix).

171 These additional binomial, hypergeometric, Cochran-Mantel-Haenszel, and logistic  
172 regression tests were completed using R package ‘stats’ version 3.2.2.

### 173 **ROLE OF THE FUNDING SOURCE**

174 The funders of the study had no role in study design, data collection, data analysis, data  
175 interpretation or writing of the report. The corresponding author had full access to the data in  
176 the study and had final responsibility for the decision to submit for publication.

177

## 178 **RESULTS**

179 We sequenced the exomes of 1,827 patients with epilepsy—640 unrelated individuals with a  
180 diagnosis of familial GGE and 525 unrelated individuals with a diagnosis of familial NAFE  
181 of European ancestry. We also sequenced an additional 662 individuals with sporadic NAFE.  
182 We compared these three groups of patients with epilepsy to 3,877 controls, who were  
183 unrelated individuals of European ancestry with no known epilepsy diagnosis.

184

185 Among our familial GGE cohort, no individual gene achieved study-wide significant  
186 enrichment for qualifying variants (Figure 1, appendix). Of the total 76,313 qualifying  
187 variants in the GGE primary analysis, 15.0% were found among cases in the familial GGE  
188 cohort. We then found that among the 76,313 qualifying variants, four unique variants  
189 overlapped a codon previously reported to have a pathogenic-classified epilepsy variant

190 based on the disease-associated variant catalogues of ClinVar, the Online Mendelian  
191 Inheritance in Man (OMIM), or the Human Gene Mutation Database (HGMD). All four  
192 variants (two *SCN1A*, one *GABRG2*, and one *SCN1B*; appendix) were found among the  
193 familial GGE cohort, an improbable enrichment given the expected proportion of 15.0%  
194 ( $p=5.1 \times 10^{-4}$ , two-tailed exact binomial test). Through an evaluation of the scientific literature,  
195 these four cases were confirmed as unrelated to those families reported in the literature.

196

197 While no single gene attained study-wide significance in the familial GGE analysis, three  
198 known epilepsy genes (*KCNQ2*, *GABRG2*, and *SCN1A*), were among the top ten case-  
199 enriched genes in the primary analysis (Figure 1). A hypergeometric test was run at each of  
200 the gene ranks occupied by one of the 43 established epilepsy genes (appendix), and we  
201 found that the enrichment was greatest at rank 151 whereby seven of the 43 known epilepsy  
202 genes had been accounted for (hypergeometric  $p=5.8 \times 10^{-8}$ ; appendix).

203

204 When we assessed enrichment among six biologically informed gene-sets, we found that the  
205 familial GGE cohort had a significant enrichment of ultra-rare functional variation among 43  
206 known dominant epilepsy genes ( $p=9.1 \times 10^{-8}$ , OR=2.3 [95% C.I. 1.7–3.2]; Table 2) and a  
207 subset of 33 genes known to contribute to epileptic encephalopathy ( $p=2.6 \times 10^{-7}$ , OR=2.6  
208 [95% C.I. 1.8–3.6]).<sup>3</sup> We confirmed that the signal of enrichment for qualifying variants  
209 among known epilepsy genes was consistently greater than the control rate across groupings  
210 of the familial GGE cohort, reflecting the number of affected relatives (appendix). While they  
211 did not achieve study-wide significance (defined as  $p < 8.9 \times 10^{-7}$ ), we also investigated  
212 qualifying variant enrichment among the fragile X mental retardation protein associated  
213 genes,<sup>22</sup> the genes encoding the NMDA receptor (NMDAR), and neuronal activity-regulated

214 cytoskeleton-associated protein, postsynaptic signalling complexes,<sup>23</sup> mouse seizure-  
215 associated orthologs,<sup>24</sup> and ion channel protein-coding genes<sup>25</sup> (Table 2). None of these gene-  
216 set tests reported enrichment of neutral variation.

217

218 Among the primary analysis of our familial NAFE cohort (figure 2A), *DEPDC5* achieved  
219 study-wide significance (OR 8.1 [95% C.I. 3.6–18.3],  $p=1.8 \times 10^{-7}$ ). *LGII* did not achieve  
220 study-wide significance (OR 29.9 [95% C.I. 6.0–288.0],  $p=1.4 \times 10^{-6}$ ). Established epilepsy  
221 genes *PCDH19* (OR 22.4 [95% C.I. 4.0–226.4],  $p=6.4 \times 10^{-5}$ ), *SCN1A* (OR 5.5 [95% C.I. 2.3–  
222 12.9],  $p=9.0 \times 10^{-5}$ ) and *GRIN2A* (OR 7.5 [95% C.I. 2.2–25.1],  $p=5.3 \times 10^{-4}$ ) occupied the 3<sup>rd</sup> –  
223 5<sup>th</sup> genome-wide ranks (appendix), but were not study-wide significant after correcting for  
224 the 56,004 tests (Bonferroni corrected  $p = 1$ ). A hypergeometric test indicated that it was  
225 highly improbable for five of the 43 known dominant epilepsy genes to occupy the top five  
226 positions of the primary analysis by chance ( $p=5.7 \times 10^{-14}$ ) (appendix).

227

228 Of 74,272 qualifying variants identified in the primary analysis of 525 individuals with  
229 familial NAFE and 3,877 controls, 9,092 (12.2%) of these were found among the familial  
230 NAFE cases. Among the 74,272 qualifying variants, nine variants overlapped a codon of a  
231 ClinVar, OMIM, or HGMD literature-reported pathogenic variant in a confirmed unrelated  
232 family. All nine unique variants (three *DEPDC5*, three *PCDH19*, one *CHRN2*, one *GRIN2A*  
233 and one *LGII* variant; appendix) were found among nine distinct NAFE cases of the  
234 combined 4,402 unrelated samples used in the familial NAFE collapsing analysis, despite the  
235 expected proportion being 12.2% (exact binomial test  $p=6.2 \times 10^{-9}$ ).

236 The known dominant epilepsy gene-set (OR=3.6 [95% CI 2.7–4.9],  $p=1.1 \times 10^{-17}$ ) and the  
237 epileptic encephalopathy gene-set (OR=3.3 [95% CI 2.3–4.7],  $p=5.0 \times 10^{-11}$ ) were study-wide  
238 significantly enriched for qualifying variants among the primary analysis of familial NAFE  
239 cases (Table 2). As observed in the familial GGE cases, the signal of enrichment for

240 qualifying variants among known epilepsy genes remained consistently greater than the  
241 control rate across groupings of the familial NAFE cohort stratified by the number of affected  
242 relatives (appendix). Presumably neutral variation was not significantly enriched among any  
243 gene-set. Under the loss-of-function model, *DEPDC5* achieved study-wide significance  
244 (OR=53.07, [95% C.I. 12.1–481.3],  $p=9.6 \times 10^{-12}$ ), with 14 (2.7%) of familial NAFE cases  
245 having a *DEPDC5* loss-of-function variant compared to only two (0.05%) controls. Focusing  
246 solely on PolyPhen-2 ‘probably damaging’ missense *DEPDC5* qualifying variants showed  
247 that they were non-significant for enrichment (3 [0.6%] of 525 cases vs. 12 [0.3%] of  
248 3877 controls; OR=1.9 [95% C.I. 0.3–6.9],  $p=0.41$ ; Figure 2B and appendix). Results from  
249 the list of 43 known dominant epilepsy genes that achieved an uncorrected  $p < 0.05$  in the  
250 primary or loss-of-function models are listed in the appendix.

251

252 Sanger sequencing was used to validate a subset of qualifying variants found among 19  
253 established and 13 candidate epilepsy genes (appendix). Our rate of Sanger validation was  
254 97.0% (128/132) of the qualifying variants identified through the collapsing tests (appendix).  
255 When available, we also Sanger sequenced qualifying variants among affected first-degree  
256 relatives of index cases used in the collapsing analyses. We looked at six genes where we had  
257 enough affected first-degree relatives to be sufficiently powered to achieve an uncorrected  
258  $p < 0.05$  from a test of preferential segregation (appendix). Comparing to the expected rate of  
259 50%, *SCN1A* (88.2% co-occurrence;  $p=1.2 \times 10^{-3}$ ), *DEPDC5* (100% co-occurrence;  $p=4.9 \times 10^{-4}$ )  
260 and *GRIN2A* (100% co-occurrence;  $p=7.8 \times 10^{-3}$ ) had significant co-occurrence among  
261 affected first-degree family members, after correcting for the six studied genes (adjusted  
262  $\alpha=8.3 \times 10^{-3}$ ; appendix).

263

264 To explore which variants, as a function of MAF, are most important to the observed risk  
265 signal we performed conditional analyses (appendix). These analyses show that among the  
266 observed epilepsy risk signal, beyond the ultra-rare qualifying variants (i.e., absent in EVS  
267 and ExAC) there is no significant contribution from variants with minor-allele frequencies up  
268 to 0.1% population MAF. This was true for both the familial GGE and familial NAFE  
269 populations (Figure 3; appendix).

270

271 Comparing 662 sporadic NAFE cases to controls did not identify study-wide significant  
272 genes across any of the three models (appendix). Of the five previously described familial  
273 NAFE top ranked genes, we found that only *LGII* achieved an uncorrected p-value of less  
274 than 0.05, (OR 8.8 [95% C.I. 1.0–105.7],  $p=0.025$ ). None of the tested gene-sets were  
275 significantly enriched with qualifying variants among sporadic NAFE cases (Table 2, Figure  
276 3).

277

278

## DISCUSSION

279 In this study, we demonstrate the presence of clear genetic risk signal for common epilepsies  
280 across genes established as responsible for familial and rare severe epilepsies. In our analysis  
281 of a cohort of individuals with familial NAFE, we found that five established epilepsy genes  
282 (*DEPDC5*, *LGII*, *PCDH19*, *SCN1A* and *GRIN2A*) occupy the top five positions genome-  
283 wide, and after correcting for background variation, the collection of these five genes  
284 contribute to approximately 8% of patients with familial NAFE. Sampling from a similarly  
285 sized familial GGE collection identified three established epilepsy genes (*KCNQ2*, *SCN1A*,  
286 and *GABRG2*) ranking among the top ten genes. Power estimates highlight the potential for  
287 new epilepsy gene discovery using this framework on larger sample sizes (appendix). Using

288 the example from *LGII*, while we found only two qualifying variants among 3,877 controls  
289 (0.05%), identifying eight familial NAFE case carriers in the primary analysis (1.5% of the  
290 familial NAFE cohort) was still inadequate to achieve study-wide significance ( $p < 8.9 \times 10^{-7}$ )  
291 for this known familial NAFE gene. Assuming the sampled rates for *LGII* case and control  
292 carriers remain the same, we estimate that *LGII* would achieve study-wide significance with  
293 the inclusion of approximately twice as many controls and 70 more unrelated familial NAFE  
294 cases.

295 As in earlier studies, our data show that *SCN1A* contributes to risk in both the familial GGE  
296 and familial NAFE epilepsy cohorts<sup>11</sup> and this enrichment is not explained by diagnoses of  
297 generalized epilepsy with febrile seizures plus (GEFS+). *SLC9A2* was also among the top 20  
298 genes in both the familial NAFE and familial GGE cohort analyses; however, it did not reach  
299 study-wide significance. No clear risk signal for epilepsy was found among the sporadic  
300 NAFE cohort. This might be explained by the possibility that non-genetic (acquired) causes  
301 play a more important role among individuals with sporadic NAFE, leading to substantially  
302 reduced power but otherwise similar genetics. Other unexplored genetic contributions to the  
303 sporadic NAFE cohort include somatic mutations arising later in development, limited to the  
304 brain or at undetectable levels in blood-extracted DNA using conventional whole-exome  
305 sequencing.

306 Among the most important findings in this work is our ability to identify clear risk signal in  
307 these data and subsequently show that the observed risk signal is concentrated among the  
308 rarest variants in the human population. In fact, among the 43 established dominant epilepsy  
309 genes we have shown that there is no evidence of risk contribution from variants observed at  
310 greater than 0.005% allelic frequency. This, however, does not preclude any other  
311 contributions to risk being present among currently unrecognized epilepsy risk genes. This  
312 work not only illustrates the value of large reference control variant databases,<sup>19</sup> but provides

313 clinically relevant information concerning the frequency spectrum of risk variants for a  
314 common complex disease.

315 A new paradigm is emerging for the treatment of rare devastating epilepsies, where  
316 treatments are being targeted to the precise genetic cause of disease.<sup>3, 26-28</sup> For example,  
317 children with *KCNT1* gain-of-function mutations have been treated with quinidine<sup>27, 29</sup> while  
318 patients with *GRIN2A* gain-of-function mutations have been treated with memantine, a  
319 specific NMDA receptor blocker.<sup>28, 30</sup> As this paradigm becomes more established, a critical  
320 question for the field is whether the approach will also apply to common epilepsies. If so, the  
321 field, which is currently accustomed to undertaking large randomised controlled trials in  
322 broad phenotypes, needs to rapidly develop a framework for classification based on ultra-rare  
323 variants in what is effectively a collection of rare genetic diseases. The work presented here  
324 demonstrates that many genes responsible for devastating rare and familial epilepsies also  
325 contribute to more common epilepsies, and it is still the ultra-rare variants that are relevant in  
326 those genes. This suggests that the emerging precision medicine paradigm of targeting  
327 treatments to the underlying causes of disease in the rarest epilepsies may also find  
328 application among the common epilepsies.

## 329 **RESEARCH IN CONTEXT**

### 330 **Evidence before this study**

331 The genetic underpinnings of the common epilepsies are largely unknown, especially the  
332 relative contributions of common variants of small effect size versus rare variants of large  
333 effect, where opportunities for novel therapeutic strategies may be greater. We searched  
334 PubMed for the terms “exome sequencing” and “common epilepsy” for reports published  
335 before June 28, 2016, with no language restrictions. There were no reports of exome  
336 sequencing of large case collections of common complex epilepsies. Although exome  
337 sequencing studies have been successful in implicating numerous genes and finding the  
338 relevant mutations for individuals with rare severe paediatric epilepsies, including epileptic  
339 encephalopathies, estimating the risk contribution from the ultra-rare protein-coding variants  
340 has been less clear for many of the common epilepsy syndromes.

### 341 **Added value of this study**

342 We used whole-exome sequencing on a large collection of two common epilepsy syndromes,  
343 genetic generalized epilepsy (GGE) and non-acquired focal epilepsy (NAFE), to search for an  
344 excess of ultra-rare deleterious qualifying variants, and compared the qualifying variant rates  
345 found among cases to background rates estimated from sequenced controls. Among familial  
346 index cases sampled from the common epilepsies, we found a significant excess of ultra-rare  
347 deleterious variation within known epileptic encephalopathy genes. We also demonstrate that  
348 the epilepsy risk signal observed in the known epilepsy genes is accounted for by the ultra-  
349 rare class of variants that are absent among large reference control cohorts, such as ExAC and  
350 EVS. Variants in known epilepsy genes that were predicted to be deleterious, but found at  
351 very low frequencies among the population reference cohorts, showed no evidence of  
352 contribution to the observed epilepsy risk signal.

353 **Implications of all the available evidence**

354 The present findings provide three key conclusions important to our understanding of the  
355 common epilepsies. First, identifying significant enrichment of ultra-rare deleterious variants  
356 among established epilepsy genes illustrates that there are genuine signals to be found using  
357 the analysis framework presented here. Secondly, we showed that the precision medicine  
358 framework that is emerging for rare epilepsies can be expected to find applications among  
359 more common epilepsies. Finally, we showed that the risk signals among the common  
360 complex forms of epilepsy come from the rarest variants in the human population, providing  
361 the clearest insight currently available into the genetic variants underlying this common  
362 complex disorder. Further research is warranted to understand to what extent these findings  
363 can be applied to clinical practice.

364

365

366 **CONTRIBUTORS**

367 EPGP project design: B.K.A., O.D., D.D., M.P.E., R.Kuz., D.H.L., R.O., E.H.S. and M.R.W.  
368 EPGP patient recruitment and phenotyping: B.A.-K., D.A., E.A., F.A., J.F.B., S.F.B., J.Bl.,  
369 G.D.C., D.C., P.Cr., O.D., D.D., M.F., N.B.F., D.F., E.B.G., T.G., S.G., K.H., S.R.H., J.H.,  
370 S.L.H., H.E.K., R.C.K., E.H.K., R.Kup., R.Kuz., D.H.L., P.V.M., E.J.N., J.M.Pao., J.M.Par.,  
371 K.P., A.P., L.G.S., I.E.S., J.J.S., S.S., R.K.S., J.Si., M.C.S., L.L.T., A.V., E.P.G.V.,  
372 G.K.V.A., J.L.W. and P.W.-W. Phenotype data analysis: B.A.-K., B.K.A., A.B., G.D.C.,  
373 O.D., D.D., M.P.E., J.F., C.F., T.G., S.J., A.K., R.C.K., R.Kuz., D.H.L., R.O., J.M.Pao., A.P.,  
374 I.E.S., R.A.S., E.H.S., J.J.S., S.S., J.Su., P.W.-W. and M.R.W.

375 Epi4K project design: A.S.A., S.F.B., P.Co., N.D., D.D., M.P.E., D.B.G., E.L.H., M.R.J.,  
376 R.Kuz., D.H.L., A.G.M., H.C.M., T.J.O., R.O., S.Petrou, S.Petrov, A.P., and I.E.S. Patient  
377 recruitment and phenotyping: S.T.B., S.F.B., R.B., G.C., S-K.C., P.Co., N.D., D.D., M.R.J.,  
378 R.Kuz., D.H.L., A.G.M., T.J.O., R.O., W.O.P., R.A.R., M.I.R., B.M.R., I.E.S., G.J.S., and  
379 R.H.T. Bioinformatics processing: J.Br., S.Petrov., Z.R. and Q.W. Sequencing and  
380 segregation of variants: M.S.H. and C.M. Sequence data analysis and Statistical  
381 interpretation: lead analyst S.Petrov with A.S.A., D.B.G. and E.L.H. Writing of manuscript:  
382 lead writer S.Petrov with A.S.A., S.F.B., D.B.G., E.L.H., and D.H.L.

383

384 **DECLARATION OF INTERESTS**

385 DBG reports personal fees from Pairnomix and EpiPM /Clarus outside the submitted work.  
386 DA reports grants from NINDS during the conduct of the study. EA reports grants from  
387 NINDS/NIH during the conduct of the study, and grants from Sunovion Pharmaceuticals Inc.  
388 outside the submitted work. FA reports grants from NINDS/NIH during the conduct of the  
389 study, and grants from Sunovion Pharmaceuticals Inc. outside the submitted work. SFB

390 reports grants from National Health and Medical Research Council and NINDS during the  
391 conduct of the study; grants from UCB Pharma, Novartis Pharmaceuticals, Sanofi-Aventis,  
392 Jansen Cilag, outside the submitted work; In addition, SFB has a patent for SCN1A testing  
393 held by Bionomics Inc and licensed to various diagnostic companies. No financial return  
394 although was a consultant to Bionomics and Athena diagnostics over 4 years ago; Patent  
395 pending for DEPDC5 testing applied for by University of South Australia and University of  
396 Melbourne. JLC reports personal fees from Shire, and grants from UCB Pharma, outside the  
397 submitted work. PC reports grants from NINDS (NS053998) during the conduct of the study.  
398 OD served on the scientific advisory board of PairoMix during the conduct of the study. DD  
399 reports grants from NIH/NINDS during the conduct of the study. MPE reports grants from  
400 NIH/NINDS during the conduct of the study. EBG reports grants and personal fees from  
401 Neuropace Inc outside the submitted work. TG reports grants from NIH (NS053998) during  
402 the conduct of the study, and personal fees from AssureX Health, outside the submitted work;  
403 also has a patent (PCT/US2006/045631) with royalties paid by AssureX Health. EHK served  
404 on the Data Safety Board of GW Pharma, served on the Scientific Advisory Board of Atkins  
405 Nutritionals, and reports grants from Nutricia outside the reported work. RK reports grants  
406 from NINDS (NS053998) during the conduct of the study. HCM reports grants from  
407 NINDS/NIH during the conduct of the study. RO reports other from Trigeminal Solutions Inc  
408 outside the submitted work. KP reports grants from NIH during the conduct of the study.  
409 SPetrou reports equity from PairoMix outside the submitted work; a patent for the diagnosis  
410 and treatment of epilepsy (WO2004/085674) issued to Bionomics Ltd, and a patent for  
411 treatment and diagnosis of epilepsy by detecting mutations in the SCN1A  
412 gene(WO2006/133508) issued to Bionomics Ltd. SPetrov reports interests in PairoMix  
413 outside the submitted work. IES reports grants from NHMRC and NIH during the conduct of  
414 the study; personal fees from UCB, Transgenomics, GlaxoSmithKline, Sanofi, and Eisai  
415 outside the submitted work; patent for Epilepsy and Mental Retardation Limited to Females

416 with royalties paid (WO/2009/086591), and a patent for treatment, and diagnosis of epilepsy  
417 by detecting mutations in the SCN1A gene licensed (WO/2006/133508). RAS reports grants  
418 from NIH, Pediatric Epilepsy Research Foundation, and American Sleep Medicine  
419 Foundation outside the submitted work; reports personal fees from UpToDate outside the  
420 submitted work. EHS reports personal fees from Invitae outside the submitted work. JJS  
421 reports grants from NIH during the conduct of the study. SS reports personal fees from  
422 Accordia, AstraZenica, Neurelis, Questcor, Upsher Smith, and Xeris outside the submitted  
423 work. LLT reports grants from NIH, Epilepsy Study Consortium, and Eisai outside the  
424 submitted work. RT has received personal fees for speaking from UCB Pharma and Eisai  
425 outside the submitted work; personal fees for consulting from Sanofi outside the submitted  
426 work.

427 All other authors declare no competing interests.

## 428 **ACKNOWLEDGEMENTS**

429 We are grateful to the patients, their families, clinical research coordinators and referring  
430 physicians for participating in the various recruitment sites that provided the phenotype data  
431 and DNA samples used in this study. We thank the following professional and lay  
432 organizations for substantial assistance in publicizing EPGP and therefore enabling us to  
433 recruit participants effectively: AED Pregnancy Registry, American Epilepsy Society,  
434 Association of Child Neurology Nurses, California School Nurses Organization, Child  
435 Neurology Society, Citizens United for Research in Epilepsy, Epilepsy Alliance of Orange  
436 County, Epilepsy Foundation, Epilepsy Therapy Project, and Finding a Cure for Epilepsy and  
437 Seizures. We thank the EPGP Administrative Core (K.F. Schardein, R. Fahlstrom, S.  
438 Cristofaro and K. McGovern), EPGP Bioinformatics Core (G. Nesbitt, K. McKenna, V.  
439 Mays), staff at the Coriell Institute – NINDS Genetics Repository, and members of the  
440 Institute for Genomic Medicine (C. Malone, B. Krueger, S. Kamalakaran, C. Davidson, B.

441 Copeland, S. Kisselev, D. Collado, D. Fernandez, J. Charoensri and P. Cansler) for their  
442 dedication and commitment to this work. We also thank R. Stewart, K. Gwinn, R. Corriveau,  
443 B. Fureman and V. Whittemore from the National Institute of Neurological Disorders and  
444 Stroke for their careful oversight and guidance of both EPGP and Epi4K. This work was  
445 supported by grants from the National Institute of Neurological Disorders and Stroke (The  
446 Epilepsy Phenome/Genome Project NS053998; Epi4K—Administrative Core NS077274;  
447 Epi4K—Sequencing, Biostatistics and Bioinformatics Core NS077303; Epi4K—Multiplex  
448 Families & Pairs Project NS077367 and Epi4K—Phenotyping and Clinical Informatics Core  
449 NS077276) and project grant (P1104) to R.H.T. from Epilepsy-Research UK.

450

451 The authors would like to thank the NHLBI GO Exome Sequencing Project and its ongoing  
452 studies which produced and provided exome variant calls for comparison: the Lung GO  
453 Sequencing Project (HL-102923), the WHI Sequencing Project (HL-102924), the Broad GO  
454 Sequencing Project (HL-102925), the Seattle GO Sequencing Project (HL-102926) and the  
455 Heart GO Sequencing Project (HL-103010). The authors would like to thank the Exome  
456 Aggregation Consortium and the groups that provided exome variant data for comparison. A  
457 full list of contributing groups can be found at <http://exac.broadinstitute.org/about>.

458 **REFERENCES**

459 1. Gilissen C, Hoischen A, Brunner HG, Veltman JA. Unlocking Mendelian disease using exome  
460 sequencing. *Genome biology* 2011; **12**(9): 228.

461 2. Collisson EA, Cho RJ, Gray JW. What are we learning from the cancer genome? *Nature*  
462 *reviews Clinical oncology* 2012; **9**(11): 621-30.

463 3. EpiPM Consortium. A roadmap for precision medicine in the epilepsies. *The Lancet*  
464 *Neurology* 2015; **14**(12): 1219-28.

465 4. Berg AT, Berkovic SF, Brodie MJ, et al. Revised terminology and concepts for organization of  
466 seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005-  
467 2009. *Epilepsia* 2010; **51**(4): 676-85.

468 5. Epi4K Consortium, Epilepsy Phenome/Genome Project. De novo mutations in epileptic  
469 encephalopathies. *Nature* 2013; **501**(7466): 217-21.

470 6. Euro E-RESC, Epilepsy Phenome/Genome P, Epi KC. De novo mutations in synaptic  
471 transmission genes including DNMT1 cause epileptic encephalopathies. *Am J Hum Genet* 2014; **95**(4):  
472 360-70.

473 7. Vadlamudi L, Milne RL, Lawrence K, et al. Genetics of epilepsy: The testimony of twins in the  
474 molecular era. *Neurology* 2014; **83**(12): 1042-8.

475 8. Peljto AL, Barker-Cummings C, Vasoli VM, et al. Familial risk of epilepsy: a population-based  
476 study. *Brain : a journal of neurology* 2014; **137**(Pt 3): 795-805.

477 9. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through  
478 whole-genome sequencing. *Nature reviews Genetics* 2010; **11**(6): 415-25.

479 10. Goldstein DB. Common genetic variation and human traits. *The New England journal of*  
480 *medicine* 2009; **360**(17): 1696-8.

481 11. International League Against Epilepsy Consortium on Complex Epilepsies. Electronic address  
482 e-aeua. Genetic determinants of common epilepsies: a meta-analysis of genome-wide association  
483 studies. *The Lancet Neurology* 2014; **13**(9): 893-903.

484 12. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases.  
485 *Nature* 2009; **461**(7265): 747-53.

486 13. Gibson G. Rare and common variants: twenty arguments. *Nature reviews Genetics* 2011;  
487 **13**(2): 135-45.

488 14. Goldstein DB, Allen A, Keebler J, et al. Sequencing studies in human genetics: design and  
489 interpretation. *Nature reviews Genetics* 2013; **14**(7): 460-70.

490 15. Epi4K Consortium. Epi4K: gene discovery in 4,000 genomes. *Epilepsia* 2012; **53**(8): 1457-67.

491 16. Speed D, Hoggart C, Petrovski S, et al. A genome-wide association study and biological  
492 pathway analysis of epilepsy prognosis in a prospective cohort of newly treated epilepsy. *Human*  
493 *molecular genetics* 2014; **23**(1): 247-58.

494 17. Cirulli ET, Lasseigne BN, Petrovski S, et al. Exome sequencing in amyotrophic lateral sclerosis  
495 identifies risk genes and pathways. *Science* 2015; **347**(6229): 1436-41.

496 18. EVS. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP). Seattle, WA.

497 19. ExAC. Exome Aggregation Consortium (ExAC). Cambridge, MA.

498 20. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging  
499 missense mutations. *Nature methods* 2010; **7**(4): 248-9.

500 21. EuroEpinomics- R. E. S. Consortium, Epilepsy Phenome/Genome Project, Epi4K Consortium.  
501 De novo mutations in synaptic transmission genes including DNMT1 cause epileptic  
502 encephalopathies. *American journal of human genetics* 2014; **95**(4): 360-70.

503 22. Darnell JC, Van Driesche SJ, Zhang C, et al. FMRP stalls ribosomal translocation on mRNAs  
504 linked to synaptic function and autism. *Cell* 2011; **146**(2): 247-61.

505 23. Kirov G, Pocklington AJ, Holmans P, et al. De novo CNV analysis implicates specific  
506 abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Molecular*  
507 *psychiatry* 2012; **17**(2): 142-53.

- 508 24. Eppig JT, Blake JA, Bult CJ, Kadin JA, Richardson JE, Mouse Genome Database G. The Mouse  
509 Genome Database (MGD): facilitating mouse as a model for human biology and disease. *Nucleic*  
510 *acids research* 2015; **43**(Database issue): D726-36.
- 511 25. Pawson AJ, Sharman JL, Benson HE, et al. The IUPHAR/BPS Guide to PHARMACOLOGY: an  
512 expert-driven knowledgebase of drug targets and their ligands. *Nucleic acids research* 2014;  
513 **42**(Database issue): D1098-106.
- 514 26. Milligan CJ, Li M, Gazina EV, et al. KCNT1 gain of function in 2 epilepsy phenotypes is  
515 reversed by quinidine. *Annals of neurology* 2014; **75**(4): 581-90.
- 516 27. Mikati MA, Jiang YH, Carboni M, et al. Quinidine in the treatment of KCNT1 positive  
517 epilepsies. *Annals of neurology* 2015.
- 518 28. Pierson TM, Yuan H, Marsh ED, et al. mutation and early-onset epileptic encephalopathy:  
519 personalized therapy with memantine. *Annals of clinical and translational neurology* 2014; **1**(3): 190-  
520 8.
- 521 29. Bearden D, Strong A, Ehnot J, DiGiovine M, Dlugos D, Goldberg EM. Targeted treatment of  
522 migrating partial seizures of infancy with quinidine. *Annals of neurology* 2014; **76**(3): 457-61.
- 523 30. Yuan H, Hansen KB, Zhang J, et al. Functional analysis of a de novo GRIN2A missense  
524 mutation associated with early-onset epileptic encephalopathy. *Nature communications* 2014; **5**:  
525 3251.

526

527

| Cohort           | Model                | Internal MAF(%) | External MAF(%) | Variant Effects   | # Genes with >0 qualifying variant(s) | CCDS represented in the tests (%) |
|------------------|----------------------|-----------------|-----------------|---|---------------------------------------|-----------------------------------|
| Familial<br>GGE  | Primary^             | 0.05%           | 0%              | LoF<br>inframe insertions or deletions<br>PolyPhen-2 (HumDiv) “probably” damaging | 15,515                                | 30.3Mbp (91.1%)                   |
|                  | LoF                  | 0.1%            | 0.1%            | LoF   | 10,712                                |                                   |
|                  | Common<br>(0.1% MAF) | 0.1%            | 0.1%            | LoF<br>inframe insertions or deletions<br>PolyPhen-2 (HumDiv) “probably” damaging | 17,118                                |                                   |
|                  | Presumed<br>Neutral  | 0.05%           | 0%              | Synonymous substitution   | 14,959                                |                                   |
| Familial<br>NAFE | Primary^             | 0.05%           | 0%              | LoF<br>inframe insertions or deletions<br>PolyPhen-2 (HumDiv) “probably” damaging | 15,438                                | 30.5Mbp (91.7%)                   |
|                  | LoF                  | 0.1%            | 0.1%            | LoF   | 10,601                                |                                   |
|                  | Common<br>(0.1% MAF) | 0.1%            | 0.1%            | LoF<br>inframe insertions or deletions<br>PolyPhen-2 (HumDiv) “probably” damaging | 17,089                                |                                   |
|                  | Presumed             | 0.05%           | 0%              | Synonymous substitution   | 14,871                                |                                   |

|                  |                      |       |      |   |        |                 |
|------------------|----------------------|-------|------|---|--------|-----------------|
|                  | Neutral              |       |      |   |        |                 |
| Sporadic<br>NAFE | Primary <sup>^</sup> | 0.05% | 0%   | LoF<br>inframe insertions or deletions<br>PolyPhen-2 (HumDiv) “probably” damaging | 15,507 | 30.5Mbp (91.7%) |
|                  | LoF                  | 0.1%  | 0.1% | LoF   | 10,729 |                 |
|                  | Common<br>(0.1% MAF) | 0.1%  | 0.1% | LoF<br>inframe insertions or deletions<br>PolyPhen-2 (HumDiv) “probably” damaging | 17,108 |                 |
|                  | Presumed<br>Neutral  | 0.05% | 0%   | Synonymous substitution   | 14,956 |                 |

528 **Table 1.** Qualifying variant criteria in the four models.

529 <sup>^</sup>Primary analysis permits minor allele frequency (MAF) to be up to 0.05% (i.e., up to four alleles in the combined case and control test population) to accommodate for possible  
530 recurrent pathogenic variants that might be relevant to multiple cases. GGE = genetic generalized epilepsy. NAFE = non-acquired focal epilepsy. LoF = loss-of-function. MAF  
531 = minor allele frequency. CCDS = consensus coding sequence

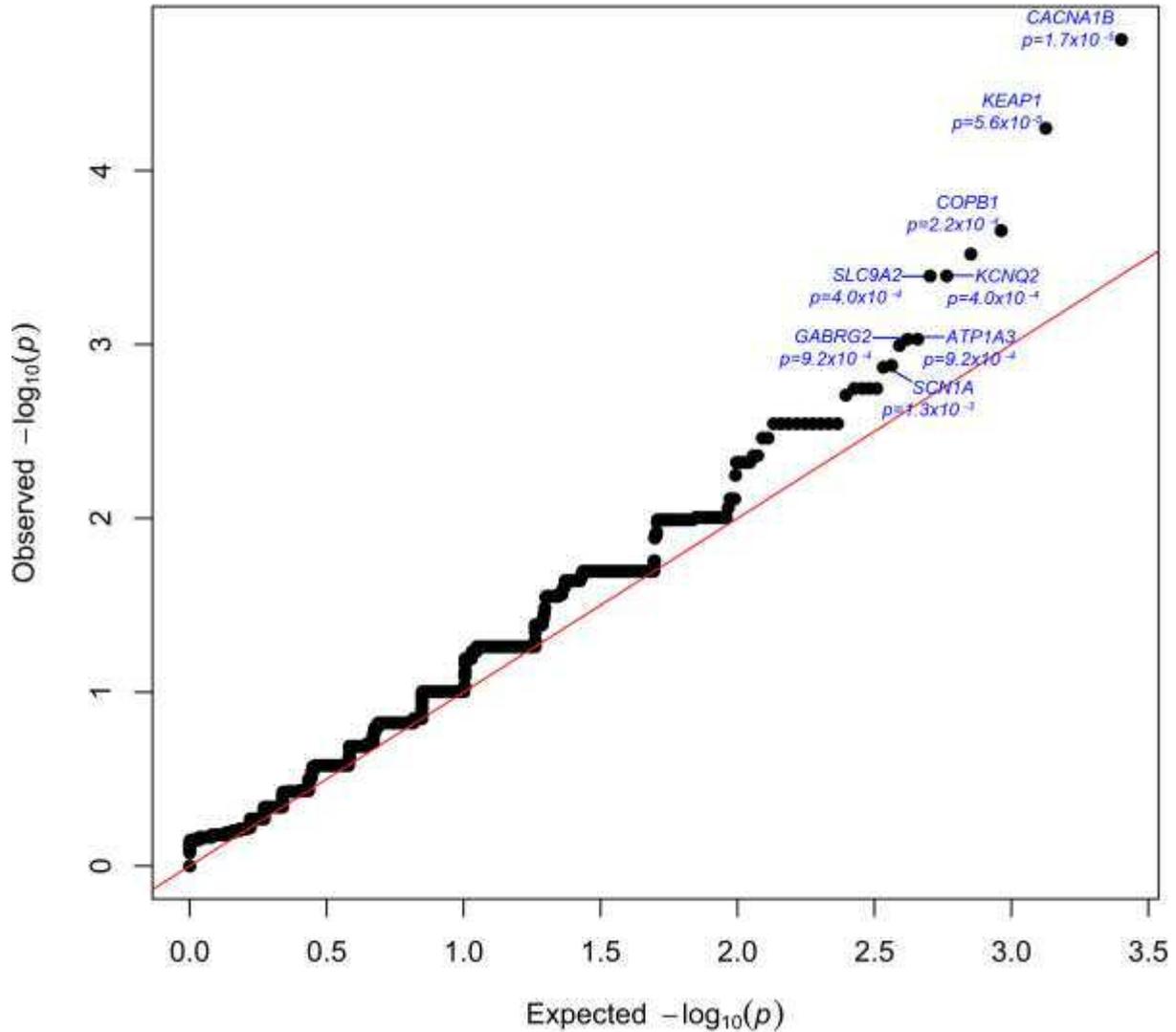
532

| Group         | Gene set    | Number of genes | Average qualifying variants <sup>a</sup> | Qualifying variants enrichment p-value (Odds Ratio [95% CI]) | Neutral variation enrichment p-value | Enrichment after removing the 43 epilepsy genes p-value |
|---------------|-------------|-----------------|--|--|--------------------------------------|---|
| Familial GGE  | Known       | 43              | 0.052                                    | $p = 9.1 \times 10^{-8}$<br>(OR=2.3 [95% CI 1.7 - 3.2])      | $p = 0.86$                           | N/A   |
|               | Known (EE)  | 33              | 0.037                                    | $p = 2.6 \times 10^{-7}$<br>(OR=2.6 [95% CI 1.8 - 3.6])      | $p = 0.34$                           | N/A   |
|               | Ion Channel | 209             | 0.264                                    | $p = 0.028$<br>(OR=1.2 [95% CI 1.0 - 1.5])                   | $p = 0.73$                           | $p = 0.21$  |
|               | FMRP        | 823             | 1.481                                    | $p = 0.034$<br>(OR=1.3 [95% CI 1.0 - 1.6])                   | $p = 0.94$                           | $p = 0.04$  |
|               | NMDAR & ARC | 78              | 0.067                                    | $p = 0.004$<br>(OR=1.6 [95% CI 1.1 - 2.1])                   | $p = 0.80$                           | $p = 0.007$   |
|               | MGI Seizure | 235             | 0.269                                    | $p = 0.003$<br>(OR=1.3 [95% CI 1.1 - 1.6])                   | $p = 0.97$                           | $p = 0.17$  |
| Familial NAFE | Known       | 43              | 0.055                                    | $p = 1.1 \times 10^{-17}$<br>(OR=3.6 [95% CI 2.7 - 4.9])     | $p = 0.87$                           | N/A   |
|               | Known (EE)  | 33              | 0.037                                    | $p = 5.0 \times 10^{-11}$<br>(OR=3.3 [95% CI 2.3 - 4.7])     | $p = 0.65$                           | N/A   |
|               | Ion Channel | 209             | 0.264                                    | $p = 1.9 \times 10^{-4}$<br>(OR=1.5 [95% CI 1.2 - 1.8])      | $p = 0.47$                           | $p = 0.05$  |
|               | FMRP        | 823             | 1.466                                    | $p = 0.77$<br>(OR=1.0 [95% CI 0.8 - 1.2])                    | $p = 0.77$                           | $p = 0.38$  |
|               | NMDAR & ARC | 78              | 0.061                                    | $p = 0.43$<br>(OR=0.8 [95% CI 0.5 - 1.3])                    | $p = 0.62$                           | $p = 0.40$  |
|               | MGI Seizure | 235             | 0.261                                    | $p = 0.05$<br>(OR=1.2 [95% CI 1.0 - 1.5])                    | $p = 0.81$                           | $p = 0.87$  |
| Sporadic NAFE | Known       | 43              | 0.045                                    | $p = 0.27$<br>(OR=1.2 [95% CI 0.8 - 1.8])                    | $p = 0.27$                           | N/A   |
|               | Known (EE)  | 33              | 0.030                                    | $p = 0.79$<br>(OR=0.9 [95% CI 0.5 - 1.5])                    | $p = 0.49$                           | N/A   |
|               | Ion Channel | 209             | 0.251                                    | $p = 0.34$<br>(OR=0.9 [95% CI 0.7 - 1.1])                    | $p = 0.88$                           | $p = 0.25$  |
|               | FMRP        | 823             | 1.461                                    | $p = 0.95$<br>(OR=1.0 [95% CI 0.8 - 1.2])                    | $p = 0.92$                           | $p = 0.94$  |
|               | NMDAR & ARC | 78              | 0.063                                    | $p = 0.65$<br>(OR=1.1 [95% CI 0.8 - 1.5])                    | $p = 0.49$                           | $p = 0.70$  |
|               | MGI Seizure | 235             | 0.254                                    | $p = 0.36$<br>(OR=0.9 [95% CI 0.7 - 1.1])                    | $p = 0.33$                           | $p = 0.33$  |

533 **Table 2. Gene-set enrichment tests.** P-values are from a logistic regression model that regresses the  
534 case/control status of a sample on the presence (1) or absence (0) of at least one qualifying variant among the  
535 corresponding gene set (Primary model). Reported p-values are uncorrected; the study-wide multiplicity-  
536 adjusted significance threshold  $\alpha = 8.9 \times 10^{-7}$ . All tests use the individual's gender, exome-wide tally of  
537 qualifying variants, and the individual's gene-list-specific tally of rare neutral (synonymous) variation as  
538 correction factors (appendix). **Known** = 43 established dominant human epilepsy genes.<sup>3</sup> **Known (EE)** = A  
539 subset of genes securely implicated with epileptic encephalopathies. **Ion Channel** = genes coding for ion

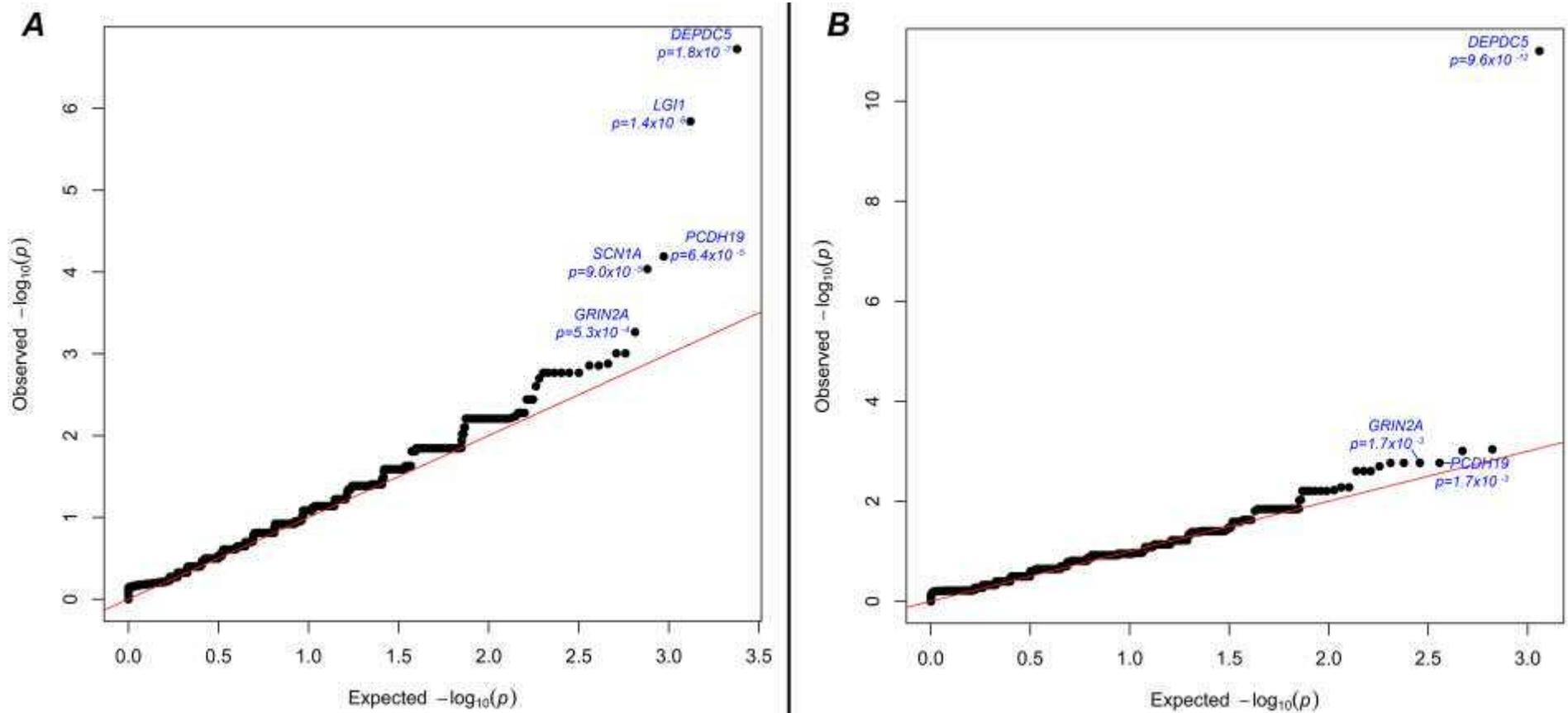
540 channels.<sup>25</sup> **FMRP** = fragile X mental retardation protein associated genes.<sup>22</sup> **NMDAR & ARC** = NMDA  
541 receptor and neuronal activity-regulated cytoskeleton-associated protein synaptic transmission genes.<sup>23</sup> **MGI**  
542 **Seizure** = mouse orthologs linked with seizure phenotypes in the Mouse Genome Database.<sup>24</sup> <sup>a</sup> Average number  
543 of qualifying variants in the corresponding gene set, per sample in the test population.

544 **Figure 1: Familial GGE primary model analysis.** 15,515 genes had at least one case or control  
545 carrier (table 1). Qualifying variants were defined as a minor allele frequency <0.05% in internal case  
546 and control, and absent among external reference cohorts. Variants are annotated as loss-of-function,  
547 inframe insertions or deletions, or missense predicted to be “probably damaging” by PolyPhen-2  
548 (HumDiv). No gene achieved study-wide significance (adjusted  $\alpha < 0.05/[18668 * 3] = 8.9 \times 10^{-7}$ ).



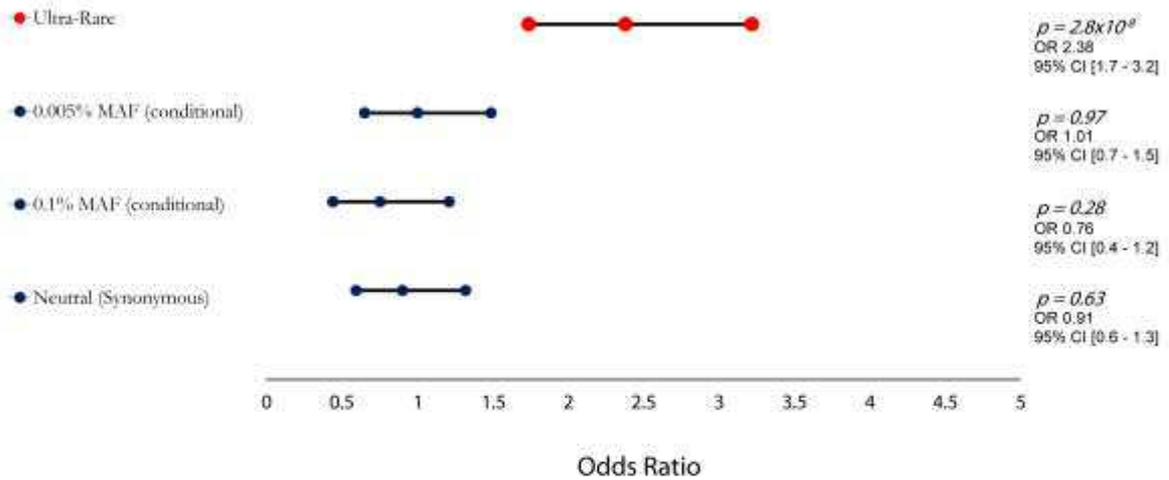
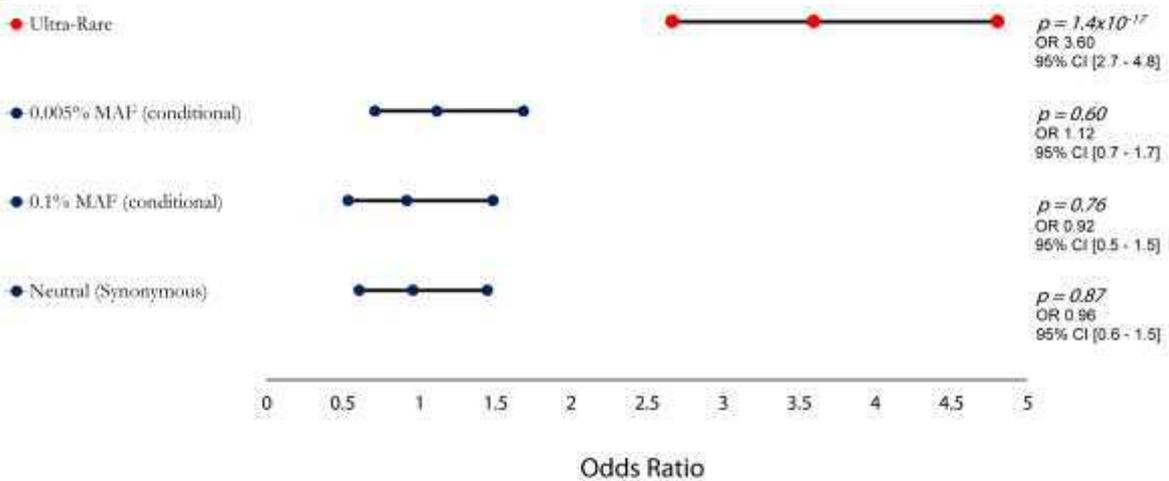
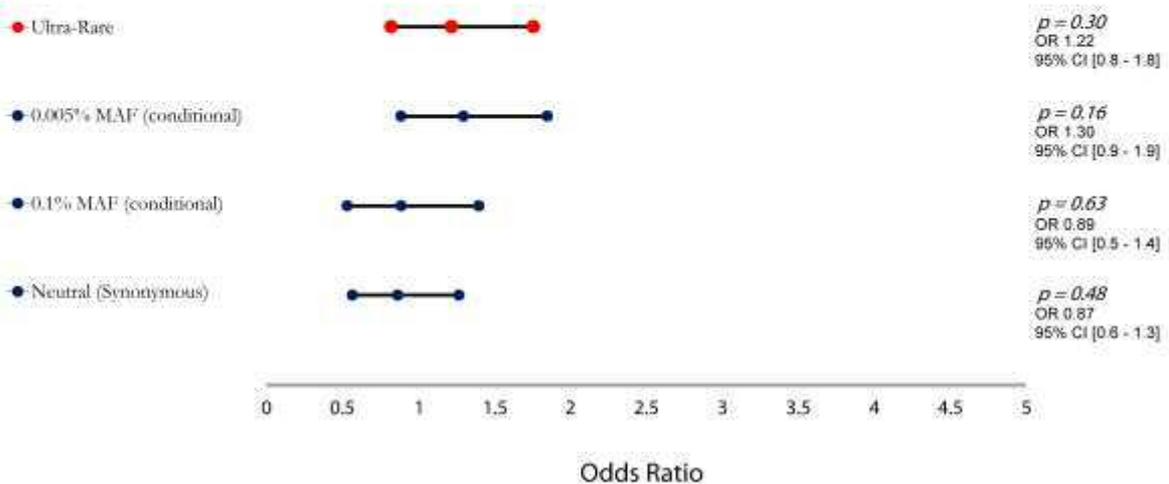
549  
550

551 **Figure 2: Familial NAFE primary model analysis.** (A) 15,438 genes had at least one case or control carrier (table 1). Qualifying variants have a minor allele  
 552 frequency <0.05% in internal case and control, and are absent among external reference cohorts. Variants are annotated as loss-of-function, inframe insertions or  
 553 deletions, or missense predicted to be “probably damaging” by PolyPhen-2 (HumDiv). Only *DEPDC5*, achieved study-wide significance (adjusted  $\alpha <$   
 554  $0.05/[18668 * 3] = 8.9 \times 10^{-7}$ ). (B) 10,601 genes had at least one case or control carrier (table 1). Qualifying variants are variants with a population MAF<0.1% and  
 555 annotated as loss-of-function effects. Only *DEPDC5* achieved study-wide significance.



556

557 **Figure 3: Enrichment of qualifying variants among 43 known epilepsy genes across increasing**  
558 **minor allele frequency bins.** The ultra-rare variation bin reflects qualifying variants from the  
559 primary analyses. The 0.005% MAF (conditional) bin represents qualifying variants with a MAF  
560 greater than 0% but no greater than 0.005% in ExAC. The 0.1% MAF (conditional) bin represents  
561 qualifying variants with a MAF greater than 0.005% but no greater than 0.1% in ExAC. The neutral  
562 (synonymous)bin represents ultra-rare putatively neutral variants across the 43 epilepsy genes.  
563 Multivariate conditional analyses for the **(A)** familial GGE population **(B)** familial NAFE population  
564 **(C)** sporadic NAFE  
565

**A****B****C**

## CONSORTIA

### **Epi4K Consortium**

Andrew S. Allen, Susannah T. Bellows, Samuel F. Berkovic, Joshua Bridgers, Rosemary Burgess, Gianpiero Cavalleri, Seo-Kyung Chung, Patrick Cossette, Norman Delanty, Dennis Dlugos, Michael P. Epstein, Catharine Freyer, David B. Goldstein, Erin L. Heinzen, Michael S. Hildebrand, Michael R. Johnson, Ruben Kuzniecky, Daniel H. Lowenstein, Anthony G. Marson, Richard Mayeux, Caroline Mebane, Heather C. Mefford, Terence J. O'Brien, Ruth Ottman, Steven Petrou, Slavé Petrovski, William O. Pickrell, Annapurna Poduri, Rodney A. Radtke, Mark I. Rees, Brigid M. Regan, Zhong Ren, Ingrid E. Scheffer, Graeme J. Sills, Rhys H. Thomas & Quanli Wang

### **Epilepsy Phenome/Genome Project**

Bassel Abou-Khalil, Brian K. Alldredge, Dina Amrom, Eva Andermann, Frederick Andermann, Jocelyn F. Bautista, Samuel F. Berkovic, Judith Bluvstein, Alex Boro, Gregory D. Cascino, Damian Consalvo, Patricia Crumrine, Orrin Devinsky, Dennis Dlugos, Michael P. Epstein, Miguel Fiol, Nathan B. Fountain, Jacqueline French, Catharine Freyer, Daniel Friedman, Eric B. Geller, Tracy Glauser, Simon Glynn, Kevin Haas, Sheryl R. Haut, Jean Hayward, Sandra L. Helmers, Sucheta Joshi, Andres Kanner, Heidi E. Kirsch, Robert C. Knowlton, Eric H. Kossoff, Rachel Kuperman, Ruben Kuzniecky, Daniel H. Lowenstein, Paul V. Motika, Edward J. Novotny, Ruth Ottman, Juliann M. Paolicchi, Jack M. Parent, Kristen Park, Annapurna Poduri, Lynette G. Sadleir, Ingrid E. Scheffer, Renée A. Shellhaas, Elliott H. Sherr, Jerry J. Shih, Shlomo Shinnar, Rani K Singh, Joseph Sirven, Michael C. Smith, Joseph Sullivan, Liu Lin Thio, Anu Venkat, Eileen P. G. Vining, Gretchen K. Von Allmen, Judith L. Weisenberg, Peter Widdess-Walsh & Melodie R. Winawer

## **Affiliations**

**Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina 27710, USA.**

Andrew S. Allen (PhD)

**Epilepsy Research Centre, Department of Medicine, University of Melbourne (Austin Health), Heidelberg, Victoria 3084, Australia.**

Susannah T. Bellows (MPsych(Ed)), Samuel F. Berkovic (MD), Rosemary Burgess (PhD), Michael S. Hildebrand (PhD), Slavé Petrovski (PhD), Brigid Regan (BSc) & Ingrid E. Scheffer (MD)

**Institute for Genomic Medicine, Columbia University Medical Center, New York, New York 10032, USA.**

Joshua Bridgers (MSc), David B. Goldstein (PhD), Erin L. Heinzen (PhD), Caroline Mebane (BSc), Slavé Petrovski (PhD), Zhong Ren (MSc) & Quanli Wang (MSc)

**Department of Molecular and Cellular Therapeutics, The Royal College of Surgeons in Ireland, Dublin, Ireland.**

Gianpiero Cavalleri (PhD), Norman Delanty (MD)

**Institute of Life Science, Swansea University Medical School, Swansea University, Swansea SA2 8PP, UK**

Seo-Kyung Chung (PhD), William O. Pickrell (MD) & Mark I. Rees (PhD)

**Centre of Excellence in Neuromics and CHUM Research Center, Université de Montréal, CHUM-Hôpital Notre-Dame Montréal, Quebec H2L 4M1, Canada.**

Patrick Cossette (MD)

**Department of Neurology, Beaumont Hospital and Royal College of Surgeons, Dublin, Ireland.**

Norman Delanty (MD)

**Department of Neurology and Pediatrics, The Children's Hospital of Philadelphia, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.**

Dennis Dlugos (MD)

**Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia 30322, USA.**

Michael P. Epstein (PhD)

**Department of Neurology, University of California, San Francisco, San Francisco, California 94143, USA.**

Catharine Freyer (BA), Daniel H. Lowenstein (MD), Heidi E. Kirsch (MD) & Joseph Sullivan (MD)

**Centre for Clinical Translation Division of Brain Sciences, Imperial College London, London SW7 2AZ, UK.**

Michael R. Johnson (MD)

**Comprehensive Epilepsy Center, Department of Neurology, NYU School of Medicine, New York, New York 10016, USA.**

Ruben Kuzniecky (MD), Judith Bluvstein (MD), Jacqueline French (MD) & Daniel Friedman (MD)

**Department of Molecular and Clinical Pharmacology, University of Liverpool, Clinical Sciences Centre, Lower Lane, Liverpool L9 7LJ, UK.**

Anthony G. Marson (MD) & Graeme J. Sills (PhD)

**Department of Neurology and the Taub Institute on Alzheimer's Disease and the Aging Brain, Gertrude H. Sergievsky Center, Columbia University, New York, NY, USA**

Richard Mayeux (MD)

**Department of Pediatrics, Division of Genetic Medicine, University of Washington, Seattle, Washington 98115, USA.**

Heather C. Mefford (MD)

**Departments of Medicine and Neurology, The Royal Melbourne Hospital, Parkville, Victoria 3146, Australia.**

Terence J. O'Brien (MD), Steven Petrou (PhD) & Slavé Petrovski (PhD)

**Departments of Epidemiology and Neurology, and the G. H. Sergievsky Center, Columbia University; and Division of Epidemiology, New York State Psychiatric Institute, New York, New York 10032, USA.**

Ruth Ottman (PhD)

**The Florey Institute of Neuroscience and Mental Health, The University of Melbourne, VIC 3010, Australia.**

Steven Petrou (PhD) & Ingrid E. Scheffer (MD)

**The Centre for Neural Engineering, The University of Melbourne, VIC 3010, Australia.**

Steven Petrou (PhD)

**Division of Epilepsy and Clinical Neurophysiology, Department of Neurology Boston Children's Hospital, Boston, Massachusetts 02115, USA.**

Annapurna Poduri (MD)

**Duke Hospital Neurodiagnostic Laboratory, Duke Department of Neurology, Duke University School of Medicine, Durham NC 27705, USA.**

Rodney Radtke (MD)

**Department of Pediatrics, University of Melbourne, Royal Children's Hospital, Melbourne, VIC 3010, Australia**

Ingrid E. Scheffer (MD)

**Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Cardiff CF24 4HQ, UK**

Rhys Thomas (MD)

**Department of Neurology, Vanderbilt University Medical Center, Nashville, Tennessee 37232, USA.**

Bassel Abou-Khalil (MD), Kevin Haas (MD), & Juliann M. Paolicchi (MD)

**Department of Clinical Pharmacy, UCSF School of Pharmacy, Department of Neurology, UCSF School of Medicine, San Francisco, California 94143, USA.**

Brian K. Alldredge (PharmD)

**Neurogenetics Unit, Montreal Neurological Hospital and Institute; Departments of Neurology & Neurosurgery, McGill University, Montreal QC H3A 2B4 Canada**

Dina Amrom (MD)

**Neurogenetics Unit and Epilepsy Research Group, Montreal Neurological Hospital and Institute; Departments of Neurology & Neurosurgery and Human Genetics, McGill University, Montreal QC H3A 2B4 Canada**

Eva Andermann (MD)

**Epilepsy Research Group, Montreal Neurological Hospital and Institute; Departments of Neurology & Neurosurgery and Pediatrics, McGill University, Montreal QC H3A 2B4 Canada**

Frederick Andermann (MD)

**Department of Neurology, Cleveland Clinic Lerner College of Medicine & Epilepsy Center of the Cleveland Clinic Neurological Institute, Cleveland, Ohio 44195, USA.**

Jocelyn F. Bautista (MD)

**Department of Neurology, Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, New York 10467, USA.**

Alex Boro (MD)

**Divison of Epilepsy, Mayo Clinic, Rochester, Minnesota 55905, USA.**

Gregory D. Cascino (MD)

**Epilepsy Center, Neurology Division, Ramos Mejía Hospital, Buenos Aires 1221, Argentina.**

Damian Consalvo (MD)

**Medical Epilepsy Program & EEG & Child Neurology, Children's Hospital of Pittsburgh of UPMC, Pediatrics, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15224, USA.**

Patricia Crumrine (MD)

**NYU and Saint Barnabas Epilepsy Centers, NYU School of Medicine, New York, New York 10016, USA.**

Orrin Devinsky (MD)

**Department of Neurology, Epilepsy Care Center, University of Minnesota Medical School, Minneapolis 55414, USA.**

Miguel Fiol (MD)

**FE Dreifuss Comprehensive Epilepsy Program, University of Virginia, Charlottesville, Virginia 22908, USA.**

Nathan B. Fountain (MD)

**Division of Neurology, Saint Barnabas Medical Center, Livingston, New Jersey 07039, USA.**

Eric B. Geller (MD) & Peter Widdess-Walsh (MD)

**Division of Neurology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA.**

Tracy Glauser (MD)

**Department of Neurology, Comprehensive Epilepsy Program, University of Michigan Health System, Ann Arbor, Michigan 48109, USA.**

Simon Glynn (MD)

**Comprehensive Epilepsy Center, Montefiore Medical Center, Bronx, New York 10467, USA.**

Sheryl R. Haut (MD)

**The Kaiser Permanente Group, Oakland, California 94611, USA.**

Jean Hayward (MD)

**Neurology and Pediatrics, Emory University School of Medicine, Atlanta, Georgia 30322, USA.**

Sandra L. Helmers (MD)

**University of Michigan Division of Pediatric Neurology, Department of Pediatrics & Communicable Diseases, Ann Arbor, Michigan, 48109, USA.**

Sucheta Joshi (MD) & Renée A. Shellhaas (MD)

**Department of Neurological Sciences, Rush Epilepsy Center, Rush University Medical Center, Chicago, Illinois 60612, USA.**

Andres Kanner (MD) & Michael C. Smith (MD)

**Department of Radiology, University of California, San Francisco, California 94143, USA.**

Heidi E. Kirsch (MD)

**Neurology, University of Texas Medical School, Houston, Texas 77030, USA.**

Robert C. Knowlton (MD)

**Neurology and Pediatrics, Child Neurology, Pediatric Neurology Residency Program,  
Johns Hopkins Hospital, Baltimore, Maryland 21287, USA.**

Eric H. Kossoff (MD)

**Epilepsy Program, Children's Hospital & Research Center Oakland, Oakland,  
California 94609, USA.**

Rachel Kuperman (MD)

**Comprehensive Epilepsy Center, Oregon Health and Science University, Portland,  
Oregon 97239, USA.**

Paul V. Motika (MD)

**Departments of Neurology and Pediatrics, University of Washington School of  
Medicine, Seattle Children's Hospital, Seattle, Washington 98105, USA.**

Edward J. Novotny (MD)

**Weill Cornell Medical Center, New York, New York 10065, USA.**

Juliann M. Paolicchi (MD)

**Department of Neurology and Neuroscience Graduate Program, University of Michigan  
Medical Center, Ann Arbor, Michigan 49108, USA.**

Jack M. Parent (MD)

**Ann Arbor Veterans Administration Healthcare System, Ann Arbor, Michigan 48105,  
USA.**

Jack M. Parent (MD)

**Departments of Neurology and Pediatrics, University of Colorado School of Medicine,  
Children's Hospital Colorado, Denver, Colorado 80045, USA.**

Kristen Park (MD)

**Department of Paediatrics and Child Health, University of Otago, Wellington,  
Newtown, Wellington, 6021, New Zealand.**

Lynette G. Sadleir (MD)

**University of Michigan, Pediatric Neurology, Ann Arbor, Michigan 48109, USA.**

Renée A. Shellhaas (MD)

**Departments of Neurology, Pediatrics and Institute of Human Genetics, University of  
California, San Francisco, San Francisco, California 94158, USA.**

Elliott H. Sherr (MD)

**Department of Neurology, Mayo Clinic, Jacksonville, Florida 32224, USA.**

Jerry J. Shih (MD)

**Departments of Neurology, Pediatrics, and Epidemiology and Population Health, and  
the Comprehensive Epilepsy Management Center, Montefiore Medical Center, Albert  
Einstein College of Medicine, Bronx, NY**

Shlomo Shinnar (MD)

**Department of Pediatrics/Division of Child Neurology, University of Alabama School Of  
Medicine/Children's Hospital of Alabama, Birmingham, AL 35233, USA.**

Rani K. Singh (MD)

**Department of Neurology, Mayo Clinic, Scottsdale, Arizona 85259, USA.**

Joseph Sirven (MD)

**Department of Neurology, Division of Pediatric and Developmental Neurology,  
Washington University School of Medicine, St. Louis, Missouri 63110, USA.**

Liu Lin Thio (MD)

**Department of Pediatrics, The Children's Hospital at Saint Peters University Hospital  
and Rutgers University, New Brunswick, NJ 08901, USA.**

Anu Venkat (MD)

**Department of Neurology, Johns Hopkins Hospital, Baltimore, Maryland 21287, USA.**

Eileen P. G. Vining (MD)

**Division of Child & Adolescent Neurology, Departments of Pediatrics, University of  
Texas Medical School, Houston, Texas 77030, USA.**

Gretchen K. Von Allmen (MD)

**Department of Neurology, Division of Pediatric Neurology, Washington University  
School of Medicine, St Louis, Missouri 63110, USA.**

Judith L. Weisenberg (MD)

**Department of Neurology and the G.H. Sergievsky Center, Columbia University, New  
York, New York 10032, USA.**

Melodie R. Winawer (MD)