

# 1 *De novo* Variants in Neurodevelopmental Disorders with Epilepsy

2  
3  
4 \*Henrike O. Heyne<sup>1,2,5,26</sup>, Tarjinder Singh<sup>2,26</sup>, Hannah Stamberger<sup>21,22,23</sup>, Rami Abou  
5 Jamra<sup>1</sup>, Hande Caglayan<sup>12</sup>, Dana Craiu<sup>13</sup>, Peter De Jonghe<sup>21,22,23</sup>, Renzo Guerrini<sup>14</sup>,  
6 Katherine L. Helbig<sup>7</sup>, Bobby P. C. Koeleman<sup>15</sup>, Jack A. Kosmicki<sup>2,26</sup>, Tarja Linnankivi<sup>16</sup>,  
7 Patrick May<sup>11</sup>, Hiltrud Muhle<sup>19</sup>, Rikke S. Møller<sup>24,25</sup>, Bernd A. Neubauer<sup>18</sup>, Aarno Palotie<sup>2</sup>,  
8 Manuela Pendziwiat<sup>19</sup>, Pasquale Striano<sup>20</sup>, Sha Tang<sup>4</sup>, Sitao Wu<sup>4</sup>, EuroEPINOMICS RES  
9 Consortium\*\*, Annapurna Poduri<sup>6</sup>, Yvonne G. Weber<sup>8</sup>, Sarah Weckhuysen<sup>21,22,23</sup>, Sanjay  
10 M. Sisodiya<sup>9,10</sup>, Mark Daly<sup>2,26</sup>, Ingo Helbig<sup>7,19</sup>, Dennis Lal<sup>2,3,26</sup>, \*Johannes R. Lemke<sup>1</sup>

11  
12 \*\*EuroEPINOMICS RES Consortium

13 Zaid Afawi, Carolien de Kovel, Petia Dimova, Tania Djémié, Milda Endziniene, Dorota  
14 Hoffman-Zacharska, Johanna Jähn, Christian Korff, Anna-Elina Lehesjoki, Carla Marini,  
15 Stefanie H. Müller, Deb Pal, Niklas Schwarz, Kaja Selmer, Jose Serratosa, Ulrich Stephani,  
16 Katalin Štěrbová, Arvid Suls, Steffen Syrbe, Inga Talvik, Shan Tang, Sarah von Spiczak,  
17 Federico Zara

18  
19 *\*corresponding author*

20  
21 1 University of Leipzig Hospitals and Clinics, Leipzig, Germany

22 2 Broad Institute, Stanley Center for Psychiatric Research, Cambridge, MA, USA

23 3 Cologne Center for Genomics (CCG), Cologne, Germany

24 4 Division of Clinical Genomics, Ambry Genetics, Aliso Viejo, California, USA

25 5 IFB AdiposityDiseases, University of Leipzig Hospitals and Clinics, Leipzig, Germany

26 6 Epilepsy Genetics Program, Department of Neurology, Division of Epilepsy and Clinical  
27 Neurophysiology, Boston Children's Hospital, Boston, MA, USA

28 7 Division of Neurology Children's Hospital of Philadelphia, Philadelphia, Pennsylvania,  
29 USA

30 8 Department of Neurology and Epileptology, Hertie Institute for Clinical Brain  
31 Research, University of Tübingen, Tübingen, Germany

32 9 Department of Clinical and Experimental Epilepsy NIHR University College London  
33 Hospitals Biomedical Research Centre UCL Institute of Neurology, London, United  
34 Kingdom

35 10 The Epilepsy Society Chalfont-St-Peter Bucks, United Kingdom

36 11 Luxembourg Centre for Systems Biomedicine, Esch-sur-Alzette, University of  
37 Luxembourg, Luxembourg

38 12 Department of Molecular Biology and Genetics Bogaziçi University Istanbul, Turkey

39 13 "Carol Davila" University of Medicine Bucharest, Department of Clinical  
40 Neurosciences (No.6), Pediatric Neurology Clinic, Alexandru Obregia Hospital, Bucharest,  
41 Romania

42 14 Pediatric Neurology and Neurogenetics Unit and Laboratories A. Meyer Children's  
43 Hospital-University of Florence, Florence, Italy

44 15 Department of Medical Genetics University Medical Center Utrecht, Utrecht, The  
45 Netherlands

46 16 Department of Pediatric Neurology, Children's Hospital, University of Helsinki and  
47 Helsinki University Hospital, Helsinki, Finland

48 17

49 18 Department of Pediatric Neurology, University Hospital Giessen, Giessen, Germany

50 19 Department of Neuropediatrics, University Medical Center Schleswig-Holstein,  
51 Christian-Albrechts University, Kiel, Germany

52 20 Pediatric Neurology and Muscular Diseases Unit, Department of Neurosciences,  
53 Rehabilitation, Ophthalmology, Genetics, and Maternal and Child Health, University of  
54 Genoa 'G. Gaslini' Institute, Genova, Italy  
55 21 Neurogenetics Group, Center for Molecular Neurology, VIB, Antwerp, Belgium  
56 22 Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, Antwerp,  
57 Belgium  
58 23 Division of Neurology, University Hospital Antwerp, Antwerp, Belgium  
59 24 The Danish Epilepsy Centre, Dianalund, Denmark  
60 25 Institute for Regional Health Services, University of Southern Denmark, Odense,  
61 Denmark  
62 26 Analytical and Translational Genetics Unit  
63  
64  
65

## 66 Acknowledgements

67  
68 We like to thank all patients and their families who participated in this study, as well as  
69 the teams who were involved in recruiting patients, samples and data at the respective  
70 study sites. We thank Lisenka Vissers and Christian Gilissen for epilepsy and age  
71 phenotypes from the cohort of Lelieveld *et al*, 2016 and Jeremy McRae for useful  
72 discussions on the DDD cohort (McRae *et al*, 2016). We thank Johannes Krause for  
73 support in figure design and helpful discussions. We are grateful to members of ATGU  
74 and the Institute for Human Genetics in Leipzig for helpful contributions. Supported by  
75 the Eurocores program EuroEPINOMICS, the Fund for Scientific Research Flanders  
76 (FWO), the International Coordination Action (ICA) grant G0E8614N, and the University  
77 of Antwerp (research fund). HOH was supported by stipends from the Federal Ministry  
78 of Education and Research (BMBF), Germany, FKZ: 01EO1501 and the German Research  
79 Foundation (DFG): HE7987/1-1. HS is PhD fellow of the Fund for Scientific Research  
80 Flanders (1125416N). IH and YGW were supported by DFG grants WE4896/3-1 and  
81 HE5415/6-1. RG received funding through the EU 7<sup>th</sup> Framework Program (FP7) under  
82 the project DESIRE grant N602531.

83 The DDD study presents independent research commissioned by the Health Innovation  
84 Challenge Fund [grant number HICF-1009-003], a parallel funding partnership between  
85 the Wellcome Trust and the Department of Health, and the Wellcome Trust Sanger  
86 Institute [grant number WT098051]. The views expressed in this publication are those  
87 of the author(s) and not necessarily those of the Wellcome Trust or the Department of  
88 Health. The study has UK Research Ethics Committee approval (10/H0305/83, granted  
89 by the Cambridge South REC, and GEN/284/12 granted by the Republic of Ireland REC).  
90 The research team acknowledges the support of the National Institute for Health  
91 Research, through the Comprehensive Clinical Research Network.  
92

93 Abstract

94

95 Epilepsy is a frequent feature of neurodevelopmental disorders (NDD) but little is  
96 known about genetic differences between NDD with and without epilepsy. We analyzed  
97 *de novo* variants (DNV) in 6753 parent-offspring trios ascertained for different NDD. In  
98 the subset of 1942 individuals with NDD with epilepsy, we identified 33 genes with a  
99 significant excess of DNV, of which *SNAP25* and *GABRB2* had previously only limited  
100 evidence for disease association. Joint analysis of all individuals with NDD also  
101 implicated *CACNA1E* as a novel disease gene. Comparing NDD with and without epilepsy,  
102 we found missense DNV, DNV in specific genes, age of recruitment and severity of  
103 intellectual disability to be associated with epilepsy. We further demonstrate to what  
104 extent our results impact current genetic testing as well as treatment, emphasizing the  
105 benefit of accurate genetic diagnosis in NDD with epilepsy.

106

107

108

## 109 Introduction

110

111 Epilepsies, defined as recurrent, unprovoked seizures, affect about 50 million people  
112 worldwide ([www.who.int](http://www.who.int), 03/2017). A significant subset of severe and intractable  
113 epilepsies starts in infancy and childhood and poses a major clinical burden to patients,  
114 families, and society<sup>1</sup>. Early onset epilepsies are often comorbid with  
115 neurodevelopmental disorders (NDD), such as developmental delay, intellectual  
116 disability and autism spectrum disorders (DD, ID, ASD)<sup>2-4</sup>, while up to 26% of  
117 individuals with NDD have epilepsy, depending on the severity of intellectual  
118 impairment<sup>4-6</sup>. Several genes have been implicated in both NDD and epilepsy  
119 disorders<sup>7,8</sup>. The epileptic encephalopathies (EE) comprise a heterogeneous group of  
120 epilepsy syndromes characterized by frequent and intractable seizures that are thought  
121 to contribute to developmental regression<sup>3,9</sup>. Phenotypic categorisation of clinically-  
122 recognizable EE syndromes enabled identification of several associated genes<sup>1,2,10</sup>.  
123 However, the phenotypic spectrum of these disease genes was broader than  
124 expected<sup>11,12</sup>, ranging from EE (e.g. *SCN1A*<sup>13</sup>, *KCNQ2*<sup>14</sup>) to unspecific NDD with or without  
125 epilepsy (e.g. *SCN2A*<sup>15</sup>, *STXBP1*<sup>16</sup>). While clinically distinguishable entities exist, many  
126 patients with NDD and epilepsy are not easily classified into EE syndromes<sup>1,12</sup>.  
127 Consequently, EE is often used synonymously with NDD with epilepsy<sup>17</sup>. Targeted  
128 sequencing of disease-specific gene panels is commonly used in diagnostics of  
129 epilepsies<sup>12,18,19</sup>. However, epilepsy gene panel designs of diagnostic laboratories differ  
130 substantially in gene content<sup>19</sup>.

131 Application of a mutational model<sup>18</sup> to detect enrichment for *de novo* variants  
132 (DNV) has proven to be a powerful approach for identification of disease-associated  
133 genes in neurodevelopmental disorders including ID, congenital heart disease,  
134 schizophrenia and ASD<sup>20-23</sup>. For EE, the currently largest exome-wide DNV burden study  
135 comprised 356 parent-offspring trios of two classic EE syndromes (infantile/epileptic  
136 spasms, IS and Lennox-Gastaut syndrome, LGS) and revealed seven genes at exome-  
137 wide significance<sup>24</sup>. To identify genes that are significantly associated with NDD with  
138 epilepsy, we analysed 6753 parent-offspring trios of NDD, focusing on 1942 cases with  
139 epilepsy including 529 individuals with epileptic encephalopathy. We compared rates of  
140 DNV between EE, NDD with unspecified epilepsies and NDD without epilepsy to identify  
141 genetic differences between these phenotypic groups. We further investigated the  
142 potential impact of our findings on the design of genetic testing approaches and  
143 assessed the extent of therapeutically relevant diagnoses.

## 144 Results

145

### 146 *Description of dataset*

147 We analysed DNV in parent-offspring trios of eight published<sup>7,20,23-27</sup>, one partly  
148 published<sup>28</sup> and three unpublished cohorts of in total 6753 individuals with NDD  
149 stratifying for the 1942 cases with epilepsy (Supplementary Table 1, Figure 1, Online  
150 Methods). These 1942 patients were ascertained for either EE or NDD with unspecified  
151 epilepsy (DD<sup>21</sup>, ASD<sup>11</sup> with ID and ID<sup>20</sup>). We define those two phenotype groups as  
152 NDD<sub>EE</sub> (n = 529) and NDD<sub>uE</sub> (n = 1413), respectively. We later compared DNV in NDD  
153 with epilepsy (NDD<sub>EE+uE</sub>) to DNV in NDD without epilepsy (NDD<sub>woE</sub>, n = 4811). For  
154 genotype-phenotype comparisons, we restricted our analysis to regions that were  
155 adequately captured across different capture solutions (see Online Methods). For ASD  
156 data from the Simon Simplex Consortium<sup>29</sup>, we included only individuals with IQ < 70  
157 (defined as ID) as different studies have found DNV only associated with low-IQ ASD<sup>6,30</sup>.  
158 Individuals with NDD<sub>EE</sub> were diagnosed with following specific syndromes: IS (n = 243),  
159 LGS (n = 145), electrical status epilepticus in sleep (ESES, n = 42), myoclonic-atonic  
160 epilepsy (MAE, n = 39), Dravet syndrome (DS, n = 16), unspecified EE (n = 44). Six of  
161 eight NDD cohorts (n = 6037) included individuals with as well as without  
162 epilepsy<sup>20,23,25-27,31</sup>. Of these, 20.3% of patients had epilepsy. In cohorts with more severe  
163 ID, a higher rate of patients had epilepsy (Spearman-Rank correlation, p-value = 0.012,  
164 rho = 0.89, Supplementary Figure S2), in line with previous literature<sup>4,6</sup>. We considered  
165 DNV of 1911 healthy siblings of patients with ASD as a control group.

166

### 167 *DNV in known EE genes in patients with different NDD diagnoses*

168 We first compared DNV in known EE genes between NDD<sub>EE</sub>, NDD<sub>uE</sub>, NDD<sub>woE</sub> and control  
169 cohorts. We investigated missense and truncating DNV (DNV<sub>mis+trunc</sub>) in 50 known  
170 autosomal dominant or X-linked EE genes (updated list from<sup>19</sup>, Supplementary Table 3).  
171 We excluded DNV present in ExAC<sup>32</sup> to improve power, as these have been shown to  
172 confer no risk to childhood-onset NDD on a group level<sup>33</sup>. The frequency of DNV<sub>mis+trunc</sub> in  
173 EE genes was not significantly different between NDD<sub>EE</sub> (13.0%±3.1, mean, 95%-CI) and  
174 NDD<sub>uE</sub> (11.5%±1.8, mean, 95%-CI, p-value = 0.4, Fisher's Exact Test, Figure 1A, see  
175 Supplementary Figure S2 for individual cohorts), but was significantly greater than in  
176 NDD<sub>woE</sub> (2.7%±0.5, mean, 95%-CI, p-value = 4.4x10<sup>-46</sup>) and in healthy controls  
177 (0.3%±0.2, mean, 95%-CI)<sup>20</sup>. Within three different NDD diagnoses (ID, ASD [with and  
178 without ID], DD), we detected more DNV in EE genes in individuals with epilepsy than  
179 without epilepsy (Cochran–Mantel–Haenszel test, p-value 3.5x10<sup>-43</sup>, common OR 4.6,  
180 95%-CI: 3.7 to 5.9, Figure 2B). This suggests a markedly overlapping genetic spectrum  
181 of NDD<sub>EE</sub> and NDD<sub>uE</sub>. We subsequently performed DNV enrichment analyses on the  
182 combined cohort of NDD<sub>EE+uE</sub>.

183

### 184 *Discovery of genes with exome-wide DNV burden in NDD with epilepsy*

185 We compared the numbers of DNV in the combined cohort of NDD with epilepsy  
186 (NDD<sub>EE+uE</sub>), to the number of DNV expected by a mutational model<sup>30</sup> revealing global  
187 enrichment of truncating (2.3-fold, p<sub>trunc</sub>= 1 x 10<sup>-47</sup>, Poisson Exact test, see Online  
188 Methods) and missense (1.6-fold, p<sub>mis</sub>=2 x 10<sup>-33</sup>) but not synonymous DNV (0.6 fold,  
189 p<sub>syn</sub>=1.0). We identified 33 genes with an exome-wide significant burden of DNV<sub>mis+trunc</sub>  
190 (Table 1), of which *KCNQ2* (n=21), *SCN2A* (n=20) and *SCN1A* (n=19) were most  
191 frequently mutated. *GABRB2* and *SNAP25* had previously no statistical evidence for  
192 disease association (see Supplementary Note). Beyond the 33 genes with exome-wide  
193 significant DNV burden, 114 genes had at least two DNV<sub>mis+trunc</sub> in our cohort  
194 (Supplementary Table 6). After DNV enrichment analysis, we again excluded DNV in  
195 ExAC<sup>32</sup> to improve specificity<sup>33</sup>.

196 Collectively analysing all patients with NDD with or without epilepsy (n = 6753),  
 197 we found 101 genes with exome-wide DNV burden (Supplementary Table 7). Among  
 198 these 101 genes five were mutated in at least one individual with EE and at least two  
 199 other individuals with epilepsy with DNV in the same variant class. Of these, *SMARCA2*  
 200 *DYNC1H1* and *SLC35A2* were formerly associated with NDD with epilepsy. *KCNQ3* had  
 201 previously limited association with NDD with epilepsy and *CACNA1E* had previously no  
 202 statistical evidence for disease association (Genes further described in Supplementary  
 203 Notes).

204  
 205  
 206

**Table 1, Genes with exome-wide DNV burden in NDD with epilepsy**

Gene	DNV <sub>trunc</sub>		DNV <sub>mis</sub>		DNV <sub>mis+trunc</sub>
	yes	no	yes	no	yes+no
KCNQ2	0	1	21	3	25
SCN2A	2	12	18	5	37
SCN1A	8	0	11	0	19
CHD2	9	1	3	2	15
SYNGAP1	10	7	1	2	20
STXBP1	4	3	7	5	19
SCN8A	0	1	10	3	14
MEF2C	4	1	5	0	10
SLC6A1	2	1	7	3	13
DNM1	0	0	9	2	11
EEF1A2	0	0	8	3	11
CDKL5	2	0	6	0	8
DYRK1A	7	9	0	5	21
SMC1A	7	0	0	2	9
GABRB3	0	0	7	1	8
KIAA2022	6	0	0	0	6
ASXL3	6	12	0	0	18
WDR45	5	5	1	0	11
ARID1B	6	28	0	2	36
GNAO1	0	1	6	2	9
ALG13	0	0	6	0	6
KCNH1	0	0	6	2	8
GRIN2B	0	3	6	9	18
HNRNPU	5	2	0	1	8
PURA	3	4	2	4	13
GABRB2	0	0	5	1	6
COL4A3BP	0	0	5	4	9
MECP2	2	5	3	5	15
FOXG1	2	3	3	3	11
ANKRD11	4	28	0	2	34
SNAP25	1	0	3	0	4
DDX3X	3	19	1	11	34
IQSEC2	3	2	1	3	9

207 \*Genes in order of decreasing numbers of  $DNV_{mis+trunc}$  in NDD with epilepsy

208  
209

210 *Phenotypic, biological and therapeutic properties of genes with DNV burden in NDD with*  
211 *epilepsy*

212 We aimed to explore whether the 33 genes with DNV burden in NDD with epilepsy  
213 ( $NDD_{EE+uE}$ ) were associated with specific phenotypes. Analyses of human phenotype  
214 ontology<sup>34</sup> (HPO) terms revealed most significant enrichment in genes associated with  
215 “epileptic encephalopathy” (see Online Methods, Supplementary Table 8). After  
216 excluding the 529 patients diagnosed with EE from the DNV enrichment analysis, the  
217 most significantly enriched HPO term was still “epileptic encephalopathy” (Bonferroni  
218 p-value  $3.6 \times 10^{-14}$ ), confirming our previous findings (Figure 1). Per DNV-enriched gene,  
219 we plotted distribution of EE phenotypes, sex and seizure phenotypes of generalized,  
220 focal, febrile or spasms (Supplementary Figure S6 - 8).

221 Since the disease onset of NDD with epilepsy is typically in infancy and early  
222 childhood, we evaluated expression levels of the 33 genes with DNV burden in the  
223 developing infant brain (expression data: [brainspan.org](http://brainspan.org), see Online Methods). At a  
224 group level, these genes showed high levels of brain expression (Supplementary Figure  
225 S9A). The DNV-enriched genes were also substantially depleted for truncating and  
226 missense variants in the ExAC control data (Supplementary Figure S9B, S9C). Genes  
227 with at least two DNV in  $NDD_{EE+uE}$ , but no significant DNV burden showed similar  
228 patterns.

229 We finally evaluated if genes with  $DNV_{mis+trunc}$  in NDD with epilepsy were  
230 associated with therapy. For each gene, we used criteria from the Centre for Evidence-  
231 Based Medicine (CEBM)<sup>35</sup> to evaluate the evidence for targeted treatments. Five of the  
232 33 DNV-enriched genes (*SCN1A*, *SCN2A*, *SCN8A*, *KCNQ2*, *MECP2*) had evidence for  
233 therapeutic relevance (CEBM Grade of Recommendation A and B, see Online Methods,  
234 Supplementary Table 9). These five genes accounted for 28% of all  $DNV_{mis+trunc}$  in the  
235 significantly implicated genes. Three additional genes (*PTEN*, *CACNA1A*, *SLC2A1*) with at  
236 least two  $DNV_{mis+trunc}$ , which were also known disease genes, also had therapeutic  
237 relevance according to CEBM criteria. In total 5% (84/1587) of  $DNV_{mis+trunc}$  in NDD with  
238 epilepsy were in genes with therapeutic consequences. According to the guidelines of  
239 the American College of Medical Genetics (ACMG),<sup>36</sup> all DNV that are not in ExAC and  
240 that are in known disease genes or genes with DNV burden in our dataset are  
241 categorized as “likely pathogenic”, while we did not apply all ACMG criteria to individual  
242 DNV (see online methods).

243

244 *Comparing DNV between NDD with and without epilepsy*

245 We compared frequencies of  $DNV_{mis+trunc}$  in NDD with epilepsy ( $NDD_{EE+uE}$ ) to  $NDD_{woE}$   
246 across all 107 DNV-enriched genes (logistic regression, see Online Methods). Increasing  
247 age at time of recruitment increased likelihood of epilepsy (three-year OR 1.11, 95%-CI  
248 1.04 to 1.18, p-value =  $3 \times 10^{-3}$ , individual genes in Supplementary Figure S5). Sex was not  
249 associated with epilepsy status (p-value = 0.5). Individuals with  $DNV_{mis}$  were more likely  
250 to have epilepsy than individuals with  $DNV_{trunc}$  (Figure 2,  $OR_{mis}$  2.1, 95%-CI 1.6 to 2.8, p-  
251 value  $2 \times 10^{-7}$ ). In line with previous reports<sup>15</sup>, we observed this pattern on a single gene  
252 level for *SCN2A* (Firth regression,  $OR_{mis}$  23.5, 95%-CI 3.8 to 277, p-value 0.0003, Table 1).  
253 Confirming previous findings<sup>24,37</sup>, DNV in ion channel genes were associated with  
254 epilepsy (OR 6.0, 95%-CI 3.9 to 9.2, p-value  $1 \times 10^{-16}$ ). 83% (110/133) of DNV in ion  
255 channel genes were  $DNV_{mis}$ . However, in the subset of 910 DNV not in ion channel genes,  
256  $DNV_{mis}$  were still associated with epilepsy (OR 1.5, p-value 0.005, 95%-CI 1.1 to 2.1),  
257 implying that the effect of  $DNV_{mis}$  on epilepsy was not entirely driven by ion channel  
258 genes. We observed a higher rate of  $DNV_{mis}$  in  $NDD_{EE}$  than in  $NDD_{uE}$ , though only with  
259 nominal significance (Fisher’s exact test, OR 1.8, 95%-CI 1.04 to 3.4, p-value 0.03,  
260 Supplementary Figure S10B). Four genes were more frequently mutated in NDD with

261 epilepsy (NDD<sub>EE+uE</sub>) than NDD<sub>woE</sub> (Fisher's Exact Test, Figure 2A/2B, Table 1,  
262 Supplementary Table 10). With the exception of *SCN1A*, frequencies of DNV were not  
263 significantly different per gene between NDD<sub>EE</sub> and NDD<sub>uE</sub> for DNV<sub>mis</sub> or DNV<sub>trunc</sub>  
264 (Supplementary Figure S10, Supplementary Table 11).

265

#### 266 *Evaluation of diagnostic gene panels for epilepsy disorders*

267 Targeted sequencing of disease-specific gene panels is widely employed in diagnostics  
268 of epilepsies<sup>18,19</sup>. We compared our results to 24 diagnostic panels for epilepsy or EE  
269 (see Online Methods, full list in Supplementary Table 12). In total, the 24 different  
270 panels covered 358 unique genes ( $81.5 \pm 8.8$  genes per panel, mean  $\pm$  sd). Applying  
271 these 24 diagnostic panels on our data set would only have detected on average 59% of  
272 DNV<sub>mis+trunc</sub> in the 33 DNV-enriched genes (Supplementary Figure S11). However, similar  
273 to most other research studies involving clinical WES<sup>7</sup>, we cannot fully assess the extent  
274 of potential pre-screening. We investigated whether genes in the 24 panels had some  
275 evidence for disease association given the following features that we (and others<sup>23,33</sup>)  
276 observed in genes with DNV burden in NDD: depletion for truncating and missense  
277 variants in ExAC<sup>32</sup> controls as well as brain expression (Online Methods, Supplementary  
278 Figure S9). We restricted this analysis to autosomal dominant and X-linked acting panel  
279 genes ( $n_{\text{dominant+X-linked}} = 191$ , Supplementary Table 13). 95% (52/55) of panel genes that  
280 had two or more DNV<sub>mis+trunc</sub> in our study were both constraint and brain-expressed.  
281 However, only 63% (86/136) of panel genes with one or less DNV<sub>mis+trunc</sub> in our study  
282 were constraint and brain-expressed (Fisher's exact test, OR 10.2, 95%-CI 3.0 to 53.0, p-  
283 value  $2.3 \times 10^{-6}$ ). We applied evidence of disease association as defined by the ClinGen  
284 Gene Curation Workgroup<sup>38</sup>, to those 50 panel genes lacking two of the criteria  
285 DNV/brain expression/constraint. We found that ten of the 50 genes had no, eight had  
286 limited and seven had conflicting published evidence for disease association  
287 (Supplementary Table 14). Thirteen genes showed moderate, strong or definitive  
288 evidence for association to entities where neither NDD nor epilepsy were major features  
289 which may partly be explained by a panel design containing genes associated with  
290 diseases beyond the spectrum of NDD (for further details see Online Methods and  
291 Supplementary Figure S11).



## 292 Discussion

293

294 In this study, we systematically investigated DNV in NDD with and without epilepsy. In  
295 NDD with epilepsy, we could hardly distinguish individuals ascertained for epileptic  
296 encephalopathy and NDD with unspecified epilepsy on a genetic level. Thus, we  
297 conclude that these phenotype groups share a spectrum of disease genes predominantly  
298 including genes initially reported as EE genes. We identified 33 genes with DNV burden  
299 in NDD with epilepsy, of which the majority was expressed in the infant brain and  
300 depleted for functional variation in ExAC<sup>32</sup>, as previously described for NDD genes<sup>23,33</sup>.  
301 We report statistically robust disease association for *SNAP25*, *GABRB2* and *CACNA1E*,  
302 which was previously lacking (Supplementary Notes).

303

304 We found, that individuals with DNV<sub>mis</sub> were generally more likely to have epilepsy than  
305 individuals with DNV<sub>trunc</sub>. This association was largely driven by ion channel genes,  
306 which confirms longstanding statements that many epilepsy disorders act as  
307 channelopathies<sup>2,37,24</sup>. Heterozygous DNV<sub>mis</sub> have been shown to cause epilepsy via  
308 dominant negative (e.g. *KCNQ2*<sup>39</sup>) or gain-of-function (e.g. *SCN8A*<sup>40</sup>) effects on ion  
309 channels. On the individual gene level, missense variants in *SCN2A*<sup>15</sup> and *SCN8A*<sup>41</sup> were  
310 more strongly implicated in epilepsy than protein truncating variants, which we  
311 statistically confirm for *SCN2A*. Yet, we found that DNV<sub>mis</sub> were also associated with  
312 epilepsy independent of ion channel genes. This may imply that alteration of protein  
313 function quantitatively plays a larger role than haploinsufficiency<sup>42</sup> in the  
314 pathophysiology of NDD with epilepsy compared to NDD without epilepsy. We found  
315 multiple gene sets enriched for DNV<sub>mis</sub> in epilepsy compared to no epilepsy (see  
316 Supplementary Note). The majority was related to ion channels, while others related to  
317 neuronal cells (e.g. axon part, synaptic transmission). However, biological interpretation  
318 should be done with caution given that previous studies have found that many of these  
319 gene sets share a large number of underlying genes<sup>22</sup> and gene annotations are biased<sup>43</sup>.  
320 We further replicate a previous finding that the rate of epilepsy was correlated with  
321 severity of intellectual disability<sup>4-6</sup>, implying that brain function could contribute to  
322 epileptogenesis or genetic variants cause both epilepsy and NDD. Alternatively, severe  
323 epileptic activity may also damage brain function and thereby contribute to NDD, which  
324 constitutes the original definition of EE<sup>9,17</sup>. This is supported by many cases of clinical  
325 regression after onset of epilepsy and improvement of NDD through seizure control.

326

327 In NDD with epilepsy we found no genetic differences between unspecified epilepsy and  
328 EE, with the exception of *SCN1A* (Supplementary Note). Phenotypic heterogeneity has  
329 been described for the majority of EE genes<sup>1,11</sup>, i.e. variants in the same gene could lead  
330 to a spectrum of different phenotypes. Due to pleiotropy, individuals that carry a  
331 pathogenic DNV in an EE gene and fulfil diagnostic criteria of EE may also be eligible for  
332 another NDD diagnosis and thus by chance be assigned to an ASD, DD or ID and not an  
333 EE screening cohort. In line with this hypothesis, we found typically EE-associated  
334 seizure types (e.g. epileptic spasms) in cohorts with unspecified epilepsy. Some of the  
335 diagnostic criteria for EE<sup>1,10</sup> may present ambiguously, leading to uncertainty in  
336 terminology<sup>17</sup>. Thus, 43% (21/49) of individuals diagnosed with EE in the Epi4K-E2<sup>24</sup>  
337 study initially presented with DD prior to seizure onset conflicting with the original  
338 definition of EE<sup>3,17</sup>. Clear phenotypic distinction between encephalopathic versus non-  
339 encephalopathic epilepsies may therefore be difficult. Accordingly, mechanisms that  
340 result in an encephalopathic course of a genetic NDD remain elusive.

341

342 Restricting DNA sequencing or DNA sequence analysis to panels of known disease genes  
343 is widely used in diagnosis of genetic diseases including epilepsy (<sup>19</sup>, 100,000 genomes  
344 project [[www.genomicsengland.co.uk](http://www.genomicsengland.co.uk)]). We confirmed that epilepsy gene panels from

345 diagnostic laboratories differ substantially in gene content<sup>18</sup> with at least 25 genes with  
346 low evidence for disease association (ClinGen criteria<sup>38</sup>). Statistically not robust gene-  
347 disease associations occasionally resulted in false-positive reports of causality posing  
348 challenges for correct diagnosis in research and clinical settings<sup>11,44</sup>. Our data provide  
349 grounds for replacing genes with limited evidence by genes with higher evidence in the  
350 design of gene panels for NDD with epilepsy.

351  
352 Therapeutic approaches, tailored to the patient's underlying genetic variant, have  
353 successfully been applied for several EE<sup>2</sup> including treatment with ezogabine in *KCNQ2*  
354 encephalopathy<sup>45</sup> or ketogenic diet in *SLC2A1*-related disorders<sup>46</sup>. 5% of DNV<sub>mis+trunc</sub> in  
355 our study were in eight genes (Supplementary Table 9) for which we could confirm  
356 therapeutic consequences with established evidence-based medicine criteria<sup>35</sup>. This  
357 finding reinforces the urgency of making a genetic diagnosis in NDD with epilepsy. We  
358 expect that with increasing understanding of the underlying pathomechanisms, the  
359 group of genetic epilepsies with relevant therapeutic consequences will continue to  
360 grow.

## 361 Online Methods

362

### 363 *Patient cohorts*

364 For this study, we ascertained 8,529 patients with the following neurodevelopmental  
365 disorders (NDD): developmental delay (DD<sup>31</sup>, n=4293), autism spectrum disorder  
366 (ASD<sup>20</sup>, n=2508), epileptic encephalopathy<sup>24,28</sup> (NDD<sub>EE</sub>, n=529), intellectual  
367 disability<sup>23,25-27</sup> (ID, n=1035), and epilepsy with NDD<sup>28</sup> (n=164). From this cohort, we  
368 selected 6753 individuals, for which the presence or absence of epilepsy was  
369 ascertained and of whom ca. 88% had ID (based on assumption of 81.7% ID in the DDD  
370 study<sup>47</sup>, 89.8% ID in a diagnostic cohort from AmbryGenetics<sup>28</sup> and 100% ID in all other  
371 cohorts.) Among individuals with ASD who were phenotyped within the Simon Simplex  
372 Consortium<sup>29</sup>, we restricted our analysis to patients with ID (IQ < 70) as it has been  
373 shown that DNV play only a minor role, in normal IQ ASD<sup>6,30</sup>. Previously sequenced trios  
374 (n = 1911), from unaffected siblings of a child with ASD<sup>20,29</sup>, served as control trios. For  
375 our main analyses, we stratified this combined cohort of patients with NDD for patients  
376 comorbid or primarily diagnosed with epilepsy (NDD<sub>EE+uE</sub>, n=1942)<sup>20,29</sup>. Two EE cohorts  
377 and one ID cohort comprising a combined 144 patients were not previously published;  
378 one cohort was only partly published (see Supplementary Table 1). Medical doctors,  
379 mostly clinical geneticists, but also neurologists, paediatricians and for ASD<sup>29</sup> some  
380 primary care physicians reported out phenotypes, including presence of epilepsy, in all  
381 patients. Our analysis is based on the assumption that medical professionals are  
382 sufficiently qualified to diagnose the presence or absence of epilepsy correctly.

383

### 384 *Subphenotypes*

385 We obtained information on specific EE syndromes on 98% of 518/529 individuals with  
386 NDD<sub>EE</sub> (see main text). We obtained specific seizure types (febrile, focal, spasms,  
387 generalized) for 55% (140/256) and age of seizure onset for 30% (77/256) of  
388 individuals with DNV<sub>mis+trunc</sub> in genes with DNV burden in NDD<sub>EE+uE</sub>. (See Supplementary  
389 Figure S5 and S6). We did not obtain EEG data per patient. Some patients may have  
390 developed epilepsy after inclusion in the study, so we ascertained age at recruitment,  
391 that we obtained for 94% (1087/1157) of all individuals with NDD with DNV<sub>mis+trunc</sub> in  
392 DNV-enriched genes (median age at recruitment: 74.8 months). We obtained age of  
393 seizure onset for 30% (77/256) of individuals with epilepsy and DNV<sub>mis+trunc</sub> in DNV-  
394 enriched genes (Supplementary Figure S5). We identified 30 individuals with potentially  
395 epilepsy-relevant brain malformations (abnormalities of neuronal migration, structural  
396 abnormalities of corpus callosum, midbrain, brainstem as schiz-, megal-,  
397 holoprosencephaly) in individuals with DNV<sub>mis+trunc</sub> in DNV-enriched genes (29 from  
398 DDD<sup>29</sup>, 1 from Hamdan *et al.*<sup>7</sup>). 11 of them (37%) also had seizures.

399

### 400 *Whole exome sequencing of parent-patient trios*

401 In all cohorts, both patients and their unaffected parents underwent whole exome  
402 sequencing (WES). Variants that were not present in either parent were considered *de*  
403 *novo* variants (DNV). 1942 individuals with NDD with epilepsy (NDD<sub>EE+uE</sub>) had 1687  
404 DNV<sub>mis</sub> and 396 DNV<sub>trunc</sub> (i.e. stopgain, frameshift, essential splice site). 4811 individuals  
405 with NDD<sub>woE</sub> had 4227 DNV<sub>mis</sub> and 1120 DNV<sub>trunc</sub> (Supplementary Table 2, for individual  
406 cohorts see Supplementary Figure S3). The study was approved by the ethics committee  
407 of the University of Leipzig (224/16-ek, 402/16-ek) and additional local ethics  
408 committees. A list of all published and unpublished cohorts used in this paper can be  
409 found in Supplementary Table 1.

410

### 411 *Sequencing pipelines of previously unpublished/partly published cohorts (cohorts 8-11)*

412 Libraries were prepared from parents' and patients' DNA, exome captured and  
413 sequenced on Illumina sequencers. Raw data was processed and technically filtered

414 with established pipelines at the respective academic or diagnostic laboratories. DNV  
415 data from all cohorts was re-annotated for this study (see below). Specific pipelines of  
416 cohorts 10 to 14 are described below.

417

418 Cohort 8 (Ambry Genetics):

419 Diagnostic WES was performed on parent-offspring trios at Ambry Genetics (Aliso Viejo,  
420 CA) in 216 individuals with a history of seizures who have been previously described<sup>28</sup>.  
421 Genomic DNA extraction, exome library preparation, sequencing, bioinformatics  
422 pipeline, and data analyses were performed as previously described<sup>48</sup>. Briefly, samples  
423 were prepared and sequenced using paired-end, 100 cycle chemistry on the Illumina  
424 HiSeq 2500 sequencer. Exome enrichment was performed using either the SureSelect  
425 Target Enrichment System 3.0 (Agilent Technologies) or SeqCap EZ VCRome 2.0 (Roche  
426 NimbleGen). The sequencing reads were aligned to human reference genome (GRCh37)  
427 and variants were called by using CASAVA software (Illumina). The following variants  
428 filters were applied to generate a list of high confident de novo variant calls: 1) mutation  
429 base coverage  $\geq 20\times$  in all members of the trio; 2) heterozygous read ratio in probands  
430  $>30\%$  and  $<80\%$ ; 3) heterozygous read ratio in parents  $<2\%$ ; 4) genotype quality cutoffs  
431 SNV  $> 100$  and indels  $> 300$  and 5) exclusion of known sequencing artefacts (based on  
432 Ambry Genetics' internal databases).

433

434

435 Cohorts 9 (EuroEPINOMICS RES) and 10 (DFG atypical EE):

436 Exonic and adjacent intronic sequences were enriched from genomic DNA using the  
437 NimbleGen SeqCap EZ Human Exome Library v2.0 enrichment kit. WES was performed  
438 using a 100bp paired-end read protocol due to the manufacturer's recommendations on  
439 an Illumina HiSeq2000 sequencer by the Cologne Center for Genomics (CCG), Cologne,  
440 Germany. Reads were mapped on the human hg19 reference genome (bwa-aln software,  
441 [bio-bwa.sourceforge.net/](http://bio-bwa.sourceforge.net/)). The UnifiedGenotyper (GATK,  
442 [www.broadinstitute.org/gatk/](http://www.broadinstitute.org/gatk/)) and Mpileup (Samtools,  
443 <http://samtools.sourceforge.net/>) software were used to call variants. The paired  
444 sample feature from the DeNovoGear software was further used to examine potential de  
445 novo mutations in twin pairs. Data analysis and filtering of mapped target sequences  
446 was performed with the 'Varbank' exome and genome analysis pipeline v.2.1  
447 (unpublished; <https://varbank.ccg.uni-koeln.de>). In particular, we filtered for high-  
448 quality (coverage of more than six reads, fraction of allele carrying reads at least 25%, a  
449 minimum genotype quality score of 10, VQSLOD greater than -8) and rare (Caucasian  
450 population allele frequency  $< 0.5\%$ ) variations on targeted regions + flanking 100bp. In  
451 order to exclude pipeline specific artifacts, we also filtered against an in-house cohort of  
452 variations, which were created with the same analysis pipeline.

453

454 Cohort 11 (University of Leipzig):

455 Exome capture was carried out with Illumina's Nextera Rapid Capture Exome Kit  
456 (Illumina, Inc., San Diego, CA, USA). WES was on an NextSeq500 or HiSeq4000  
457 sequencer (Illumina, Inc.) to  $2 \times 150$ bp reads at the Centogene AG, Rostock, Germany.  
458 Raw sequencing reads were converted to standard fastq format using bcl2fastq software  
459 2.17.1.14 (Illumina, Inc.), and fed to a pipeline at Centogene AG based on the 1000  
460 Genomes Project (1000G) data analysis pipeline and GATK best practice  
461 recommendations. Sequencing reads were aligned to the GRCh37 (hg19) build of the  
462 human reference genome using bwa-mem ([bio-bwa.sourceforge.net/](http://bio-bwa.sourceforge.net/)). In addition to  
463 GATK HaplotypeCaller ([www.broadinstitute.org/gatk/](http://www.broadinstitute.org/gatk/)), variant calling was performed  
464 with freebayes (<https://github.com/ekg/freebayes>) and samtools  
465 (<http://samtools.sourceforge.net/>). Quality filtering of sequencing reads in both parents  
466 and children was done according to the following criteria: read depth  $> 20$ , quality  $> 500$ ,  
467 frequency of alternative allele between 30 and 70% for the child and not present in the

468 parents, frequency < 1% in internal database, variant called by at least two different  
469 genotype callers.

470

471 *False positive rates of DNV*

472 In cohorts 1 to 4, all DNV were validated by Sanger sequencing to eliminate false  
473 positive calls. In cohorts 5 to 7, through random selection of variants for Sanger  
474 validation, the false positive rate was estimated to be approximately 1.4% and < 5 %,  
475 respectively. In the clinical cohorts 8 to 11, variants defined as variants worth reporting  
476 back to patients (variants of unknown significance or [likely] pathogenic) are normally  
477 validated by Sanger sequencing. With this experience, false discovery rates in these  
478 cohorts were estimated to be < 5% (personal communications).

479

480 *Annotation and Filtering*

481 DNV files were generated and quality-filtered by the individual groups. All DNV were re-  
482 annotated with the following pipeline. Variants were annotated with Ensembl's Variant  
483 Effect Predictor ([http://grch37.ensembl.org/Homo\\_sapiens/Tools/VEP](http://grch37.ensembl.org/Homo_sapiens/Tools/VEP)) of version 82  
484 using database 83 of GRCh37 as reference genome. Per variant, the transcript with the  
485 most severe impact, as predicted by VEP, was selected for further analyses. The  
486 decreasing order of variant impacts was HIGH, MODERATE, MODIFIER, LOW. Only  
487 protein – altering DNV (DNV<sub>mis</sub> or DNV<sub>trunc</sub> [premature stop codon, essential splice site,  
488 frameshift]) were included in further analyses. Synonymous DNV (DNV<sub>syn</sub>) were  
489 analysed as a negative control, as most DNV<sub>syn</sub> have no effect on amino acid sequence in  
490 the protein. Variants that were present in ExAC<sup>32</sup>, an aggregation of 60,706 exome  
491 sequences from adult individuals without severe childhood-onset diseases, were  
492 excluded after DNV enrichment, as these have been shown to convey no detectable risk  
493 to NDD on a group level<sup>33</sup>. For DNV rates per cohort see Supplementary Figure S2. We  
494 did not investigate pathogenicity of individual DNV according to the guidelines of the  
495 American College of Medical Genetics (ACMG). However, ACMG criteria PS2 (de novo  
496 occurrence, with maternity and paternity confirmed) and PM2 (absence from controls)  
497 apply to all DNV in our cohort. The combination of PS2 and PM2 classifies a variant as at  
498 least “likely pathogenic”. ACMG criteria are only applicable to variants in disease  
499 associated genes<sup>36</sup>. Therefore, all DNV in known disease genes and genes with genome-  
500 wide DNV burden in our dataset are presumed likely pathogenic DNV.

501

502

503 *Harmonization of different cohorts*

504 The core analysis of our study is the enrichment of DNV<sub>mis+trunc</sub> compared to expectation  
505 by a mutational model in individuals with NDD<sub>EE+nsE</sub>. For this analysis, we were  
506 conservative in assuming that every gene was well captured across all cohorts. However,  
507 when comparing DNV burden across different phenotypes we aimed to separate  
508 technical from biological differences with the following methods. In exome sequencing,  
509 different capture solutions capture specific exonic regions with different efficiencies.  
510 These differences have shown to be quite stable within and across different samples of  
511 the same capture kits<sup>49</sup>. We therefore generated a list of exons that displayed consistent  
512 high coverage across different capture solutions. We collected published and internal  
513 data aiming for the highest possible variety of capture kits using 3,000 samples of 5  
514 different capture kits, including NimbleGen SeqCap v2 and v3, Agilent SureSelect v2, v3,  
515 and v5). We generated a list of exons where at least 80% of all samples had at least 10x  
516 coverage. We excluded the oldest capture kits before calculating the high coverage  
517 exons as well as excluding the two oldest cohorts<sup>26,27</sup> from our list of DNV. Restricting to  
518 high coverage regions resulted in a loss of ca. 11% of DNV in DNV-enriched genes. We  
519 consequently performed all genotype phenotype comparisons across cohorts (Figures  
520 1A, 2, Supplementary Figures S6-10) with this restricted DNV set. Further, we compared  
521 the frequency of DNV<sub>syn</sub> across all cohorts and excluded cohorts of which DNV<sub>syn</sub> were

522 not available. In the subset of DNV in high coverage exons, rates of supposedly neutral  
523  $DNV_{syn}$  were not different between individuals with and without epilepsy (Poisson Exact  
524 test, p-value = 0.48, RR=0.99),  $NDD_{uE}$  and  $NDD_{EE}$  (p-value = 0.65, RR= 0.94) or NDD and  
525 controls (p-value = 0.58, RR=0.99). The frequency of  $DNV_{mis+trunc}$  was also not different  
526 between individuals with and without epilepsy (p-value=0.5, RR=1.02). Our chances to  
527 identify  $DNV_{mis+trunc}$  in EE genes in the epilepsy cohort were therefore not inflated by a  
528 higher baseline rate of  $DNV_{mis+trunc}$  in comparison to  $NDD_{woE}$ . We reannotated all DNV in  
529 the same way as described above.

530

### 531 *Statistical analysis*

532 All statistical analyses were done with the R programming language ([www.r-](http://www.r-project.org)  
533 [project.org](http://www.r-project.org)). Fisher's Exact Test for Count Data, Wilcoxon rank sum test, Poisson Exact  
534 Test, Cochran-Mantel-Haenszel test, logistic regression, Firth regression, Spearman  
535 correlation, Welch two-sided t-test and calculation of empirical p-values were  
536 performed as referenced in the results. For datasets assumed to be normally distributed  
537 after visual inspection, mean and standard deviation (sd) are written as mean  $\pm$  sd.  
538 When performing Poisson Exact Tests, we reported effect size as rate ratio (RR), which  
539 is the quotient of the two rates compared in the test. For Fisher's Exact Test and logistic  
540 regression analyses, we reported odds ratios (OR). 95% confidence intervals were  
541 abbreviated as 95%-CI. The R code used to perform the statistical analyses and figures is  
542 available upon request.

543

### 544 *DNV enrichment analyses*

545 To identify genes with a significant DNV burden, we compared numbers of observed  
546 with numbers of expected missense, truncating and synonymous DNV per gene using an  
547 established framework of gene-specific mutation rates<sup>30</sup>. The analysis was done with the  
548 R package *denovolyzer*<sup>50</sup>, that compares observed versus expected DNV using a Poisson  
549 Exact test. We corrected the obtained p-values with the Bonferroni method for the  
550 number of genes for which gene specific mutation rates<sup>30</sup> were available (n= 18225) and  
551 six tests resulting in a p-value significance threshold of  $5 \times 10^{-7}$ . Genes that passed that  
552 significance threshold for either missense, truncating or both missense plus truncating  
553 DNV were considered genes with an exome-wide DNV burden. To compare DNV  
554 between disease groups, DNV enrichment analyses were carried out in the cohort of all  
555 patients with NDD (n=6753) as well as in patients with epilepsy ( $NDD_{EE+uE}$ , n=1942) and  
556 without epilepsy ( $NDD_{woE}$ , n=4811), but only genes with a  $DNV_{mis+trunc}$  burden in the NDD  
557 with epilepsy cohort and the combined NDD cohort were reported.

558

### 559 *HPO enrichment analyses*

560 Significantly enriched Human phenotype ontology (HPO) terms were computed with the  
561 R package of *g:Profiler*<sup>34</sup>, using ordered enrichment analysis on significance-ranked  
562 proteins (see Supplementary Table 8). Different gene sets were queried using the  
563 background gene set of all 18225 genes for which gene specific mutation rates were  
564 available<sup>30</sup>. Only terms that were statistically significant with a Bonferroni corrected p-  
565 value < 0.01 were reported, as our negative controls (genes with at least two  $DNV_{mis+trunc}$   
566 in healthy control) were not enriched for any functional categories below this p-value.

567

### 568 *Therapeutic relevance*

569 To assess if DNV in our cohort were in genes of therapeutic relevance, we searched the  
570 literature for treatment recommendations for all established disease genes with at least  
571 two  $DNV_{mis+trunc}$  in our NDD with epilepsy cohort. We rated the publications with the  
572 standardized score of the Oxford Centre for Evidence-Based Medicine<sup>35</sup>. We only  
573 reported and considered genes for which at least one treatment recommendation  
574 achieved level of evidence of II or higher. For a list of all genes and levels of evidence see  
575 Supplementary Table 9.

576

577 *Acquisition and processing of brain gene expression data*

578 We downloaded the Developmental Transcriptome dataset of ‘BrainSpan: Atlas of the  
579 Developing Human Brain’ ([www.brainspan.org](http://www.brainspan.org), funded by ARRA Awards  
580 1RC2MH089921-01, 1RC2MH090047-01, and 1RC2MH089929-01, 2011). The atlas  
581 includes RNA sequencing data generated from tissue samples of developing postmortem  
582 brains of neurologically unremarkable donors covering 8 to 16 brain structures. We  
583 extracted brain expression data from the 5 donors that were infants aged 0 to 12  
584 months. Per gene, we obtained the median RPKM value of all infant individuals and  
585 across brain regions. In all calculations and figures gene expression values are displayed  
586 as median ( $\log_2 + 1$ )-transformed RPKM values. We defined infant brain gene  
587 expression as median ( $\log_2 + 1$ )-transformed RPKM value  $> 1$ . More details about tissue  
588 acquisition and sequencing methodology can be found in the BrainSpan website’s  
589 documentation.

590

591 *Evaluation of genes’ intolerance to protein altering variants*

592 We assessed individual gene tolerance to truncating or missense variants in the general  
593 population with the pLI score (probability of being loss-of-function intolerant) and  
594 missense z-score. These scores indicate depletion of truncating and missense variants in  
595 ExAC<sup>32</sup> (60,706 individuals without childhood onset diseases), respectively. We used  
596 gene constraint cut-offs  $>0.9$  for pLI and  $>3.09$  for missense-z scores as recommended  
597 by the score developers<sup>32</sup>. We calculated empirical p-values to evaluate if pLI scores of  
598 exome-wide and nominally DNV-enriched genes were significantly higher compared to  
599 pLI scores of random gene sets as described in<sup>23</sup>. Briefly, we computed the expected pLI  
600 for a given gene set with size  $n$  by randomly drawing 1,000,000 gene sets with size  $n$   
601 from the total 18,225 pLI annotated genes. We computed, how many times the median  
602 pLI score of randomly sampled gene sets would exceed the median pLI of the gene set  
603 under investigation. To that number we added 1 and divided by the number of total  
604 samplings +1 to obtain the empirical p-value.

605

606 *Comparing DNV in  $NDD_{EE}$ ,  $NDD_{uE}$  and  $NDD_{woE}$*

607 We investigated  $DNV_{mis+trunc}$  in  $NDD_{EE+uE+woE}$  across all 107 genes that were DNV-  
608 enriched in  $NDD_{EE+uE}$ ,  $NDD_{woE}$  and/or  $NDD_{EE+uE+woE}$ . We restricted our analysis to DNV not  
609 in ExAC<sup>23</sup> and in high coverage regions. To investigate, if age at time of recruitment, sex  
610 or variant class ( $DNV_{mis}/DNV_{trunc}$ ) influenced the presence of epilepsy, we tested them as  
611 covariates in a logistic regression model with epilepsy as response variable. We aimed  
612 to explore, whether DNV in NDD with epilepsy might be associated with ion channels  
613 compared to NDD without epilepsy, as it is a long-established hypothesis, that many  
614 epilepsies are channelopathies<sup>37</sup>. We extracted a comprehensive gene set of 237 known  
615 ion channel genes from 1766 previously described<sup>22</sup> curated gene sets derived from  
616 public pathway databases and publications (see Supplementary Note). To investigate if  
617 ion channel genes were associated with epilepsy we included annotation as ion channel  
618 gene as a categorical predictor in the logistic regression model. We used Firth  
619 regression to assess the effect of variant class on the presence of epilepsy for individual  
620 genes. We used Fisher’s Exact test to compare frequencies of DNV per gene between  
621 phenotype groups. To account for multiple testing, we corrected p-values for the  
622 number of tests performed (Bonferroni method).

623

624 *Diagnostic gene panels for epileptic encephalopathy/ comprehensive epilepsy from 24  
625 academic/ commercial providers*

626 We set out to compare our results to diagnostic gene panels for epileptic  
627 encephalopathy of international commercial and academic providers. We searched the  
628 Genetic Testing Registry (GTR)<sup>51</sup> of NCBI (date: 01/2017) for providers of tests for  
629 “Epileptic encephalopathy, childhood-onset” and identified 16 diagnostic epilepsy

630 panels. We excluded 3 panels with  $< 20$  or  $> 200$  genes and added 11 additional  
631 diagnostic providers not registered at GTR to evaluate 24 diagnostic panels targeting  
632 epilepsy in general ( $n=11$ ) or EE specifically ( $n=13$ ). The gene content covered in each of  
633 the 24 gene panels can be found in Supplementary Table 11. Gene lists were freely  
634 available for download at the respective providers' websites. For each of the 33 genes  
635 with DNV burden in NDD with epilepsy, we calculated to what proportion they were  
636 included in 24 commercial or academic providers of gene panels for epileptic  
637 encephalopathy/comprehensive epilepsy. For each gene, we then multiplied the  
638 percentage of inclusion in any of the 24 panels by the total number of  $DNV_{mis+trunc}$  of that  
639 gene in the cohort of 1942 individuals with  $NDD_{EE+uE}$ .

640 We investigated if there were genes in the 24 diagnostic gene panels without evidence  
641 for implication in NDD with epilepsy. We focused on 191 dominant or X-linked panel  
642 genes (listed in Supplementary Table 14). We tested these genes for three criteria of  
643 association with NDD with epilepsy: Firstly, if genes had at least two  $DNV_{mis+trunc}$  in our  
644 study; secondly, whether genes were expressed in the infant brain defined by a median  
645 RPKM of all samples and brain regions  $> 1$ ; thirdly, whether genes had a pLI  $> 0.9$  or  
646 missense z-score  $> 3.09$  indicating intolerance to truncating or missense variants<sup>32</sup>. We  
647 intersected these lists to nominate genes that did not display features of DNV-enriched  
648 genes in this study. On these genes we applied ClinGen criteria<sup>38</sup> for gene-disease  
649 association.

650

#### 651 *Data availability*

652 The authors declare that all data used for computing results supporting the findings of  
653 this study are available within the paper and its supplementary information files. Raw  
654 sequencing data of published cohorts are referenced at the respective publications. Raw  
655 sequencing data of cohort EuroEPINOMICS RES have been deposited in the European  
656 Genome-phenome Archive (EGA) with the accession code EGAS00001000048  
657 (<https://www.ebi.ac.uk/ega/datasets/EGAD00001000021>). Raw sequencing data of  
658 cohort 10 (DFG atypical EE) will be deposited in a public repository after finalization of  
659 the individual project.

660



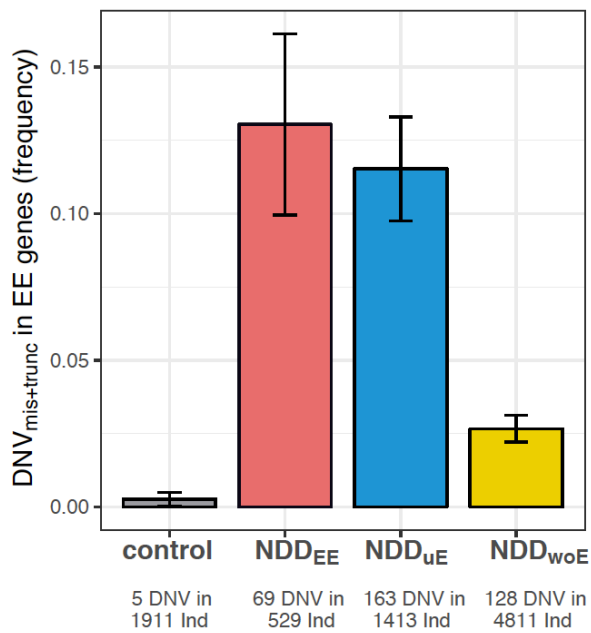
661 **Figures**

662

663 **Figure 1.** DNV<sub>mis+trunc</sub> in EE genes in different cohorts of NDD. **A,** The proportion of  
 664 DNV<sub>mis+trunc</sub> in EE genes is not significantly different between patients with NDD<sub>EE</sub> (red)  
 665 and NDD<sub>uE</sub> (blue), but higher than NDD<sub>woE</sub> (yellow) or healthy controls (grey). Cohort  
 666 size is given as number of individuals (Ind). **B,** Proportion of DNV in EE genes in patients  
 667 with versus without epilepsy across different NDD (DD, ASD, ID). P-values are plotted  
 668 next to respective odds ratios (red dots), while 95%-CI are shown as yellow bars  
 669 (Fisher's exact test for individual cohorts, Cochran-Mantel-Haenszel test for combined  
 670 cohorts).

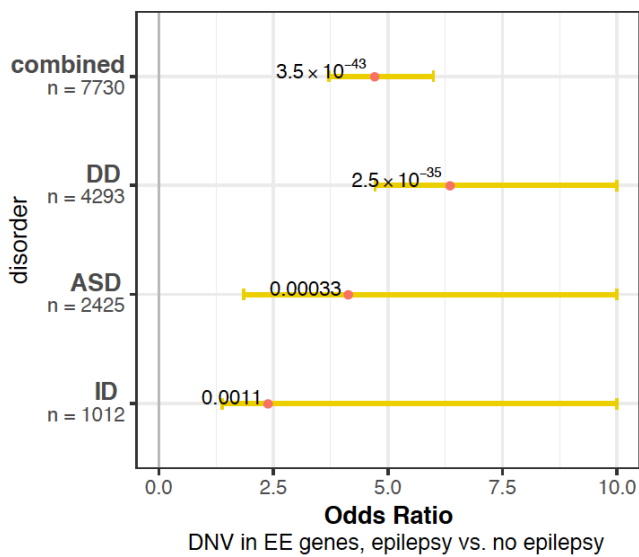
671

672 **A,**



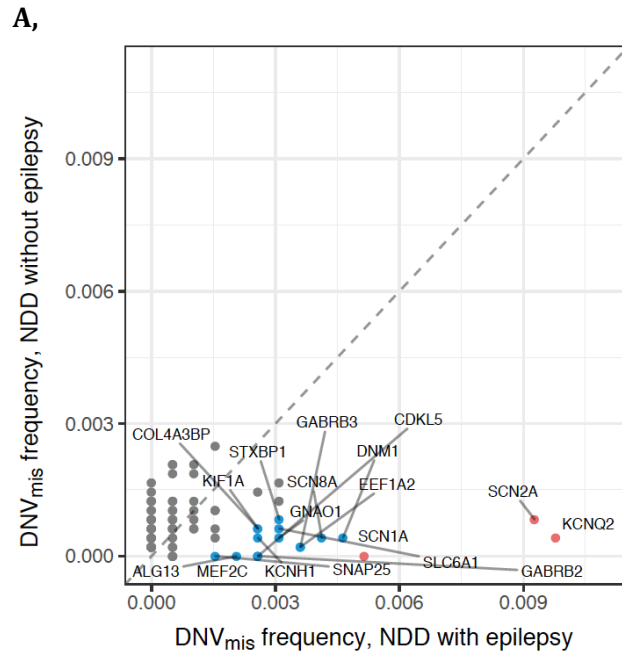
673

674 **B,**

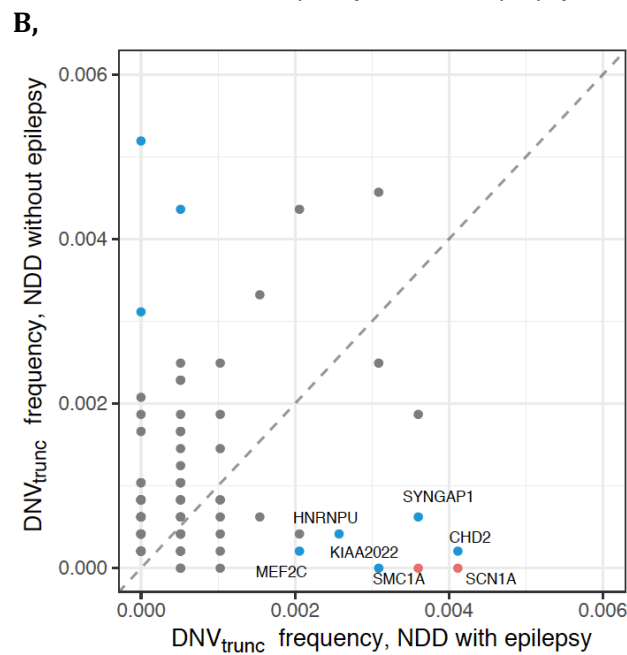


675

676 **Figure 2.** DNV in patients with epilepsy ( $NDD_{EE+uE}$ ) versus without epilepsy ( $NDD_{woE}$ )  
677 in 107 genes with significant DNV burden. **A**,  $DNV_{mis}$ , **B**,  $DNV_{trunc}$ . Genes with different  
678 frequencies are labeled (method: Fisher's Exact test, blue: nominal significance, p-value  
679 < 0.05, red: significant after correcting for 266 tests). The dotted line represents equal  
680 frequency of DNV in NDD with and without epilepsy.  
681  
682



683  
684



685  
686

## 687 References

688

- 689 1. McTague, A., Howell, K.B., Cross, J.H., Kurian, M.A. & Scheffer, I.E. The  
690 genetic landscape of the epileptic encephalopathies of infancy and  
691 childhood. *Lancet Neurol* **15**, 304-16 (2016).
- 692 2. Myers, C.T. & Mefford, H.C. Advancing epilepsy genetics in the genomic  
693 era. *Genome Med* **7**, 91 (2015).
- 694 3. Scheffer, I.E. *et al.* ILAE classification of the epilepsies: Position paper of  
695 the ILAE Commission for Classification and Terminology. *Epilepsia* (2017).
- 696 4. McGrother, C.W. *et al.* Epilepsy in adults with intellectual disabilities:  
697 prevalence, associations and service implications. *Seizure* **15**, 376-86  
698 (2006).
- 699 5. Thomas, S., Hovinga, M.E., Rai, D. & Lee, B.K. Brief Report: Prevalence of  
700 Co-occurring Epilepsy and Autism Spectrum Disorder: The U.S. National  
701 Survey of Children's Health 2011-2012. *J Autism Dev Disord* **47**, 224-229  
702 (2017).
- 703 6. Robinson, E.B. *et al.* Autism spectrum disorder severity reflects the  
704 average contribution of de novo and familial influences. *Proc Natl Acad Sci*  
705 *USA* **111**, 15161-5 (2014).
- 706 7. Deciphering Developmental Disorders, S. Prevalence and architecture of  
707 de novo mutations in developmental disorders. *Nature* **542**, 433-438  
708 (2017).
- 709 8. Li, J. *et al.* Genes with de novo mutations are shared by four  
710 neuropsychiatric disorders discovered from NPdenovo database. *Mol*  
711 *Psychiatry* **21**, 290-7 (2016).
- 712 9. Berg, A.T. *et al.* Revised terminology and concepts for organization of  
713 seizures and epilepsies: report of the ILAE Commission on Classification  
714 and Terminology, 2005-2009. *Epilepsia* **51**, 676-85 (2010).
- 715 10. Berg, A.T. & Millichap, J.J. The 2010 revised classification of seizures and  
716 epilepsy. *Continuum (Minneap Minn)* **19**, 571-97 (2013).
- 717 11. Helbig, I. & Tayoun, A.A. Understanding Genotypes and Phenotypes in  
718 Epileptic Encephalopathies. *Mol Syndromol* **7**, 172-181 (2016).
- 719 12. Trump, N. *et al.* Improving diagnosis and broadening the phenotypes in  
720 early-onset seizure and severe developmental delay disorders through  
721 gene panel analysis. *J Med Genet* **53**, 310-7 (2016).
- 722 13. Depienne, C. *et al.* Spectrum of SCN1A gene mutations associated with  
723 Dravet syndrome: analysis of 333 patients. *J Med Genet* **46**, 183-91 (2009).
- 724 14. Weckhuysen, S. *et al.* Extending the KCNQ2 encephalopathy spectrum:  
725 clinical and neuroimaging findings in 17 patients. *Neurology* **81**, 1697-703  
726 (2013).
- 727 15. Wolff, M. *et al.* Genetic and phenotypic heterogeneity suggest therapeutic  
728 implications in SCN2A-related disorders. *Brain* (2017).
- 729 16. Stamberger, H. *et al.* STXBP1 encephalopathy: A neurodevelopmental  
730 disorder including epilepsy. *Neurology* **86**, 954-62 (2016).
- 731 17. Howell, K.B., Harvey, A.S. & Archer, J.S. Epileptic encephalopathy: Use and  
732 misuse of a clinically and conceptually important concept. *Epilepsia* **57**,  
733 343-7 (2016).
- 734 18. Chambers, C., Jansen, L.A. & Dhamija, R. Review of Commercially Available  
735 Epilepsy Genetic Panels. *J Genet Couns* **25**, 213-7 (2016).

- 736 19. Lemke, J.R. *et al.* Targeted next generation sequencing as a diagnostic tool  
737 in epileptic disorders. *Epilepsia* **53**, 1387-98 (2012).
- 738 20. Iossifov, I. *et al.* The contribution of de novo coding mutations to autism  
739 spectrum disorder. *Nature* **515**, 216-21 (2014).
- 740 21. Zaidi, S. *et al.* De novo mutations in histone-modifying genes in congenital  
741 heart disease. *Nature* **498**, 220-3 (2013).
- 742 22. Singh, T. *et al.* The contribution of rare variants to risk of schizophrenia in  
743 individuals with and without intellectual disability. *Nat Genet* (2017).
- 744 23. Lelieveld, S.H. *et al.* Meta-analysis of 2,104 trios provides support for 10  
745 new genes for intellectual disability. *Nat Neurosci* **19**, 1194-6 (2016).
- 746 24. EuroEpinomics & Epi4K. De novo mutations in synaptic transmission  
747 genes including DNM1 cause epileptic encephalopathies. *Am J Hum Genet*  
748 **95**, 360-70 (2014).
- 749 25. Hamdan, F.F. *et al.* De novo mutations in moderate or severe intellectual  
750 disability. *PLoS Genet* **10**, e1004772 (2014).
- 751 26. de Ligt, J. *et al.* Diagnostic exome sequencing in persons with severe  
752 intellectual disability. *N Engl J Med* **367**, 1921-9 (2012).
- 753 27. Rauch, A. *et al.* Range of genetic mutations associated with severe non-  
754 syndromic sporadic intellectual disability: an exome sequencing study.  
755 *Lancet* **380**, 1674-82 (2012).
- 756 28. Helbig, K.L. *et al.* Diagnostic exome sequencing provides a molecular  
757 diagnosis for a significant proportion of patients with epilepsy. *Genet Med*  
758 **18**, 898-905 (2016).
- 759 29. Fischbach, G.D. & Lord, C. The Simons Simplex Collection: a resource for  
760 identification of autism genetic risk factors. *Neuron* **68**, 192-5 (2010).
- 761 30. Samocha, K.E. *et al.* A framework for the interpretation of de novo  
762 mutation in human disease. *Nat Genet* **46**, 944-50 (2014).
- 763 31. McRae, J.F. *et al.* Prevalence, phenotype and architecture of  
764 developmental disorders caused by de novo mutation. *bioRxiv* (2016).
- 765 32. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706  
766 humans. *Nature* **536**, 285-91 (2016).
- 767 33. Kosmicki, J.A. *et al.* Refining the role of de novo protein-truncating  
768 variants in neurodevelopmental disorders by using population reference  
769 samples. *Nat Genet* (2017).
- 770 34. Reimand, J. *et al.* g:Profiler-a web server for functional interpretation of  
771 gene lists (2016 update). *Nucleic Acids Res* **44**, W83-9 (2016).
- 772 35. OCEBM Levels of Evidence Working Group\* (Jeremy Howick, I.C.J.L.L.,  
773 Paul Glasziou, Trish Greenhalgh, Carl Heneghan, Alessandro Liberati, Ivan  
774 Moschetti, Bob Phillips, Hazel Thornton, Olive Goddard and Mary  
775 Hodgkinson). The Oxford Levels of Evidence. in *Oxford Centre for*  
776 *Evidence-Based Medicine*. (2009).
- 777 36. Richards, S. *et al.* Standards and guidelines for the interpretation of  
778 sequence variants: a joint consensus recommendation of the American  
779 College of Medical Genetics and Genomics and the Association for  
780 Molecular Pathology. *Genet Med* **17**, 405-24 (2015).
- 781 37. Mulley, J.C., Scheffer, I.E., Petrou, S. & Berkovic, S.F. Channelopathies as a  
782 genetic cause of epilepsy. *Curr Opin Neurol* **16**, 171-6 (2003).

- 783 38. Rehm, H.L. *et al.* ClinGen--the Clinical Genome Resource,  
784 [www.clinicalgenome.org/site/assets/files/2657/current\\_clinical\\_validity](http://www.clinicalgenome.org/site/assets/files/2657/current_clinical_validity)  
785 [\\_classifications.pdf](http://www.clinicalgenome.org/site/assets/files/2657/current_clinical_validity_classifications.pdf). *N Engl J Med* **372**, 2235-42 (2015).
- 786 39. Weckhuysen, S. *et al.* KCNQ2 encephalopathy: emerging phenotype of a  
787 neonatal epileptic encephalopathy. *Ann Neurol* **71**, 15-25 (2012).
- 788 40. Wagnon, J.L. *et al.* Pathogenic mechanism of recurrent mutations of  
789 SCN8A in epileptic encephalopathy. *Ann Clin Transl Neurol* **3**, 114-23  
790 (2016).
- 791 41. Blanchard, M.G. *et al.* De novo gain-of-function and loss-of-function  
792 mutations of SCN8A in patients with intellectual disabilities and epilepsy.  
793 *J Med Genet* **52**, 330-7 (2015).
- 794 42. Wilkie, A.O. The molecular basis of genetic dominance. *J Med Genet* **31**, 89-  
795 98 (1994).
- 796 43. Haynes, W., Tomczak, A. & Khatri, P. Gene annotation bias impedes  
797 biomedical research. *bioRxiv* (2017).
- 798 44. MacArthur, D.G. *et al.* Guidelines for investigating causality of sequence  
799 variants in human disease. *Nature* **508**, 469-76 (2014).
- 800 45. Millichap, J.J. *et al.* KCNQ2 encephalopathy: Features, mutational hot spots,  
801 and ezogabine treatment of 11 patients. *Neurol Genet* **2**, e96 (2016).
- 802 46. De Giorgis, V. & Veggiotti, P. GLUT1 deficiency syndrome 2013: current  
803 state of the art. *Seizure* **22**, 803-11 (2013).
- 804 47. Study, D.D.D. Prevalence and architecture of de novo mutations in  
805 developmental disorders. *Nature* **542**, 433-438 (2017).
- 806 48. Farwell, K.D. *et al.* Enhanced utility of family-centered diagnostic exome  
807 sequencing with inheritance model-based analysis: results from 500  
808 unselected families with undiagnosed genetic conditions. *Genet Med* **17**,  
809 578-86 (2015).
- 810 49. Chilamakuri, C.S. *et al.* Performance comparison of four exome capture  
811 systems for deep sequencing. *BMC Genomics* **15**, 449 (2014).
- 812 50. Ware, J.S., Samocha, K.E., Homsy, J. & Daly, M.J. Interpreting de novo  
813 Variation in Human Disease Using denovolyzeR. *Curr Protoc Hum Genet*  
814 **87**, 7 25 1-15 (2015).
- 815 51. Rubinstein, W.S. *et al.* The NIH genetic testing registry: a new, centralized  
816 database of genetic tests to enable access to comprehensive information  
817 and improve transparency. *Nucleic Acids Res* **41**, D925-35 (2013).
- 818