ANALYSIS
https://doi.org/10.1038/s41588-018-0143-7

nature genetics

De novo variants in neurodevelopmental disorders with epilepsy

Henrike O. Heyne[1,2,3,4]*, Tarjinder Singh[2,4], Hannah Stamberger[5,6,7], Rami Abou Jamra[1], Hande Caglayan[8], Dana Craiu[9], Peter De Jonghe[5,6,7], Renzo Guerrini[10], Katherine L. Helbig[11], Bobby P. C. Koeleman[12], Jack A. Kosmicki[2,4], Tarja Linnankivi[13], Patrick May[14], Hiltrud Muhle[15], Rikke S. Møller[16,17], Bernd A. Neubauer[18], Aarno Palotie[2], Manuela Pendziwiat[15], Pasquale Striano[19], Sha Tang[20], Sitao Wu[20], EuroEPINOMICS RES Consortium[21], Annapurna Poduri[22], Yvonne G. Weber[23], Sarah Weckhuysen[5,6,7], Sanjay M. Sisodiya[24,25], Mark J. Daly[2,4], Ingo Helbig[11,15], Dennis Lal[2,4,26] and Johannes R. Lemke[1]*

Epilepsy is a frequent feature of neurodevelopmental disorders (NDDs), but little is known about genetic differences between NDDs with and without epilepsy. We analyzed de novo variants (DNVs) in 6,753 parent–offspring trios ascertained to have different NDDs. In the subset of 1,942 individuals with NDDs with epilepsy, we identified 33 genes with a significant excess of DNVs, of which *SNAP25* and *GABRB2* had previously only limited evidence of disease association. Joint analysis of all individuals with NDDs also implicated *CACNA1E* as a novel disease-associated gene. Comparing NDDs with and without epilepsy, we found missense DNVs, DNVs in specific genes, age of recruitment, and severity of intellectual disability to be associated with epilepsy. We further demonstrate the extent to which our results affect current genetic testing as well as treatment, emphasizing the benefit of accurate genetic diagnosis in NDDs with epilepsy.

Epilepsies, defined as recurrent, unprovoked seizures, affect approximately 50 million people worldwide (World Health Organization (see URLs), March 2017). A substantial subset of severe and intractable epilepsies start in infancy and childhood and pose a major clinical burden to patients, families, and society[1]. Early-onset epilepsies are often comorbid with NDDs, such as developmental delay (DD), intellectual disability (ID), and autism spectrum disorders (ASDs)[2–4], whereas up to 26% of individuals with NDDs have epilepsy, depending on the severity of intellectual impairment[4–6]. Several genes have been implicated in both NDDs and epilepsy disorders[7,8]. The epileptic encephalopathies (EEs) comprise a heterogeneous group of epilepsy syndromes characterized by frequent and intractable seizures that are thought to contribute to developmental regression[3,9]. Phenotypic categorization of clinically recognizable EE syndromes has enabled identification of several associated genes[1,2,10]. However, the phenotypic spectrum of these disease-associated genes has been found to be broader than expected[11,12], ranging from EE (for example, *SCN1A*[13] and *KCNQ2* (ref. [14])) to unspecific NDDs with or without epilepsy (for example, *SCN2A*[15] and *STXBP1* (ref. [16])). Although clinically distinguishable entities exist, the clinical presentation of many patients with NDDs and epilepsy is not easily classified into specific EE syndromes[1,12]. Consequently, EE is often used synonymously with NDDs with epilepsy[17]. Targeted sequencing of disease-specific gene panels is commonly used in diagnostics of epilepsies[12,18,19]. However, the epilepsy gene-panel designs used by diagnostic laboratories differ substantially in gene content[19].

[1]University of Leipzig Hospitals and Clinics, Leipzig, Germany. [2]Program in Medical and Population Genetics, and Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, USA. [3]Integrated Research and Treatment Center (IFB) Adiposity Diseases, University of Leipzig Hospitals and Clinics, Leipzig, Germany. [4]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. [5]Neurogenetics Group, Center for Molecular Neurology, VIB, Antwerp, Belgium. [6]Laboratory of Neurogenetics, Institute Born–Bunge, University of Antwerp, Antwerp, Belgium. [7]Division of Neurology, University Hospital Antwerp, Antwerp, Belgium. [8]Department of Molecular Biology and Genetics, Bogaziçi University, Istanbul, Turkey. [9]'Carol Davila' University of Medicine Bucharest, Department of Clinical Neurosciences (No. 6), Pediatric Neurology Clinic, Alexandru Obregia Hospital, Bucharest, Romania. [10]Pediatric Neurology and Neurogenetics Unit and Laboratories, A. Meyer Children's Hospital–University of Florence, Florence, Italy. [11]Division of Neurology, Children's Hospital of Philadelphia, Philadelphia, PA, USA. [12]Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands. [13]Department of Pediatric Neurology, Children's Hospital, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. [14]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg. [15]Department of Neuropediatrics, University Medical Center Schleswig–Holstein, Christian–Albrechts University, Kiel, Germany. [16]Danish Epilepsy Centre, Dianalund, Denmark. [17]Institute for Regional Health Services, University of Southern Denmark, Odense, Denmark. [18]Department of Pediatric Neurology, University Hospital Giessen, Giessen, Germany. [19]Pediatric Neurology and Muscular Diseases Unit, Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, and Maternal and Child Health, University of Genoa 'G. Gaslini' Institute, Genoa, Italy. [20]Division of Clinical Genomics, Ambry Genetics, Aliso Viejo, CA, USA. [22]Epilepsy Genetics Program, Department of Neurology, Division of Epilepsy and Clinical Neurophysiology, Boston Children's Hospital, Boston, MA, USA. [23]Department of Neurology and Epileptology, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany. [24]Department of Clinical and Experimental Epilepsy, NIHR, University College London Hospitals Biomedical Research Centre, UCL Institute of Neurology, London, UK. [25]The Epilepsy Society, Chalfont-St-Peter Bucks, UK. [26]Cologne Center for Genomics (CCG), Cologne, Germany. [46]A list of members and affiliations appears at the end of the paper *e-mail: hheyne@broadinstitute.org; johannes.lemke@medizin.uni-leipzig.de

Application of a mutational model[18] to detect enrichment for DNVs has proven to be a powerful approach for identifying disease-associated genes in NDDs including ID, congenital heart disease, schizophrenia, and ASD[20–23]. For EE, the largest exome-wide DNV burden study to date comprised 356 parent–offspring trios with two classic EE syndromes: (infantile/epileptic spasms and Lennox–Gastaut syndrome) and identified seven genes with exome-wide significance[24]. To identify genes that are significantly associated with NDDs with epilepsy, we analyzed 6,753 parent–offspring trios of NDDs, focusing on 1,942 cases with epilepsy, including 529 individuals with EE. We compared the rates of DNVs among EE, NDDs with unspecified epilepsies, and NDDs without epilepsy to identify genetic differences among these phenotypic groups. We further investigated the potential implications of our findings for the design of genetic-testing approaches and assessed the extent of therapeutically relevant diagnoses.

Results

Description of dataset. We analyzed DNVs in parent–offspring trios of eight published[7,20,23–27], one partly published[28], and three unpublished cohorts including a total of 6,753 individuals with NDDs, stratifying for the 1,942 cases with epilepsy (description of cohorts in Supplementary Table 1 and Methods; DNVs in Supplementary Table 2). These 1,942 patients were classified as having either EE or NDDs with unspecified epilepsy (DD[21], ASD[11] with ID, or ID[20]). We defined those two phenotypic groups as $NDDs_{EE}$ ($n=529$) and $NDDs_{uE}$ ($n=1,413$), respectively (visualization in Supplementary Fig. 1). We later compared DNVs in NDDs with epilepsy ($NDDs_{EE+uE}$) against DNVs in NDDs without epilepsy ($NDDs_{woE}$, $n=4,811$). For genotype–phenotype comparisons, we restricted our analysis to regions that were adequately captured across different capture solutions (Methods and Supplementary Fig. 2). For ASD data from the Simon Simplex Consortium[29], we included only individuals with intelligence quotient (IQ) <70 (defined as ID), because different studies have found DNVs associated with only low-IQ ASD[6,30]. Individuals with $NDDs_{EE}$ were diagnosed with the following specific syndromes: infantile/epileptic spasms ($n=243$), Lennox–Gastaut syndrome ($n=145$), electrical status epilepticus in sleep ($n=42$), myoclonic-atonic epilepsy ($n=39$), Dravet syndrome ($n=16$), and unspecified EE ($n=44$). Six of eight NDD cohorts ($n=6,037$) included individuals with as well as without epilepsy[7,20,23,25–27]. Among these, 20.3% of individuals had epilepsy. In cohorts with more severe ID, a higher proportion of individuals had epilepsy (Spearman rank correlation, $P=0.012$, rho = 0.89; Supplementary Fig. 3), in line with the previous literature[4,6]. We considered DNVs of 1,911 healthy siblings of individuals with ASD as a control group.

DNVs in known EE-associated genes in patients with different NDD diagnoses. We first compared DNVs in known EE-associated genes among $NDDs_{EE}$, $NDDs_{uE}$, $NDDs_{woE}$, and control cohorts. We investigated missense and truncating DNVs ($DNV_{mis+trunc}$) in 50 known autosomal-dominant or X-linked EE-associated genes (updated list from ref. [19]; Supplementary Table 3). We excluded DNVs present in the Exome Aggregation Consortium (ExAC)[31] to improve power, because these have been shown to confer no risk for childhood-onset NDDs on a group level[32]. The frequency of $DNV_{mis+trunc}$ in EE-associated genes was not significantly different between $NDDs_{EE}$ (13.0% ± 3.1%, mean ± 95% confidence interval (CI) and $NDDs_{uE}$ (11.5% ± 1.8%, mean ± 95% CI, $P=0.4$, Fisher's exact test, Fig. 1a; individual cohorts in Supplementary Fig. 4) but was significantly greater than those in $NDDs_{woE}$ (2.7% ± 0.5%, mean ± 95% CI, $P=4.4\times10^{-46}$) and in healthy controls (0.3% ± 0.2%, mean ± 95% CI)[20]. Within three different NDD diagnoses (ID, ASD (with and without ID), and DD), we detected more DNVs in EE-associated genes in individuals with epilepsy than without epilepsy (Cochran–Mantel–Haenszel test, $P=3.5\times10^{-43}$, common odds ratio (OR) = 4.6, 95% CI = 3.7–5.9, Fig. 1b). This result suggests a markedly overlapping genetic spectrum of $NDDs_{EE}$ and $NDDs_{uE}$. We subsequently performed DNV enrichment analyses on the combined cohort of $NDDs_{EE+uE}$.

Discovery of genes with exome-wide DNV burden in NDDs with epilepsy. We compared the numbers of DNVs in the combined cohort of NDDs with epilepsy ($NDDs_{EE+uE}$) against the number of DNVs expected from a mutational model[30] and found a global enrichment of truncating (2.3-fold, $P_{trunc}=1\times10^{-47}$, Poisson exact test; Methods) and missense (1.6-fold, $P_{mis}=2\times10^{-33}$) but not synonymous DNVs (0.6 fold, $P_{syn}=1.0$). We identified 33 genes with an exome-wide-significant burden of DNV_{mis}, DNV_{trunc} or $DNV_{mis+trunc}$ (Table 1), of which *KCNQ2* ($n=21$), *SCN2A* ($n=20$), and *SCN1A* ($n=19$) were most frequently mutated. *GABRB2* and *SNAP25* previously had no statistical evidence of disease association (Supplementary Note). Beyond the 33 genes with exome-wide-significant DNV burden, 114 genes had at least two $DNV_{mis+trunc}$ in our cohort (Supplementary Table 4). After DNV enrichment analysis, we excluded DNVs in ExAC[31] to improve specificity[32].

After collectively analyzing all patients with NDDs with or without epilepsy ($n=6,753$), we found 101 genes with exome-wide DNV burden (Supplementary Table 5). Among these 101 genes, five were mutated in at least one individual with EE and at least two other individuals with epilepsy with DNVs in the same variant class. Of these, *SMARCA2*, *DYNC1H1*, and *SLC35A2* were formerly associated with NDDs with epilepsy. *KCNQ3* had previously been shown to have a limited association with NDDs with epilepsy, and *CACNA1E* had no previous significant evidence for disease association (genes further described in the Supplementary Note).

Phenotypic, biological, and therapeutic properties of genes with DNV burden in NDDs with epilepsy. We aimed to explore whether the 33 genes with DNV burden in NDDs with epilepsy ($NDDs_{EE+uE}$) were associated with specific phenotypes. Analyses of human phenotype ontology[33] (HPO) terms showed the most significant enrichment in genes associated with 'epileptic encephalopathy' (Methods and Supplementary Table 6). After exclusion of the 529 patients diagnosed with EE from the DNV enrichment analysis, the most significantly enriched HPO term was still 'epileptic encephalopathy' (Bonferroni $P=3.6\times10^{-14}$), thus confirming our previous findings (Fig. 1). For each DNV-enriched gene, we plotted the distribution of EE phenotypes, sex, and phenotypes of generalized, focal, febrile, or spasm seizure types (Supplementary Figs. 5 and 6).

Because the disease onset of NDDs with epilepsy typically occurs in infancy and early childhood, we evaluated expression of the 33 genes with DNV burden in the developing infant brain (expression data from BrainSpan (see URLs; Methods). At a group level, these genes showed high levels of infant brain expression (Supplementary Fig. 7a). The DNV-enriched genes were also substantially depleted in truncating and missense variants in the ExAC control data (Supplementary Fig. 7b,c). Genes with at least two DNVs in $NDDs_{EE+uE}$ but no significant DNV burden showed similar patterns.

We finally evaluated whether genes with $DNV_{mis+trunc}$ in NDDs with epilepsy were associated with therapy. For each gene, we used criteria from the Centre for Evidence-Based Medicine (CEBM)[34] to evaluate the evidence for targeted treatments. Five of the 33 DNV-enriched genes (*SCN1A*, *SCN2A*, *SCN8A*, *KCNQ2*, and *MECP2*) had evidence of therapeutic relevance (CEBM Grade of Recommendation A and B; Methods and Supplementary Table 7). These five genes accounted for 28% of all $DNV_{mis+trunc}$ in the significantly implicated genes. Three additional genes (*PTEN*, *CACNA1A*, and *SLC2A1*) with at least two $DNV_{mis+trunc}$, which were also known disease-associated genes, also had therapeutic relevance according to CEBM criteria. In total, 5% (84/1,587) of $DNV_{mis+trunc}$ in

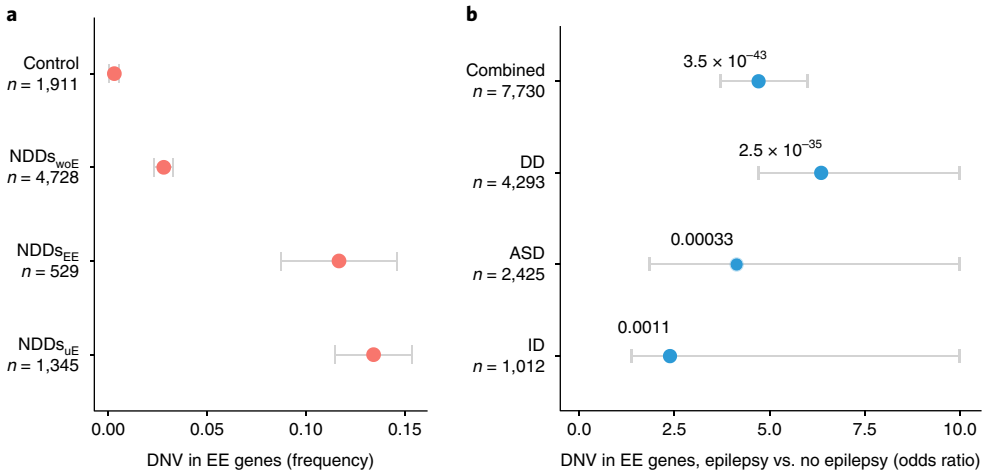

Fig. 1 | $DNV_{mis+trunc}$ in EE-associated genes in different cohorts of NDDs. a, The frequency of $DNV_{mis+trunc}$ in EE-associated genes is not significantly different between patients with $NDDs_{EE}$ and $NDDs_{uE}$, but is higher than in $NDDs_{woE}$ or healthy controls. **b**, The frequency of DNVs in EE-associated genes in individuals with versus without epilepsy across different NDDs (DD, ASD, and ID). *P* values are plotted next to respective odds ratios (two-sided Fisher's exact test for individual cohorts, Cochran–Mantel–Haenszel test for combined cohorts). Bars represent the 95% CI of the point estimates. Low-coverage exons and cohorts were excluded from the analysis.

NDDs with epilepsy were in genes with therapeutic consequences. According to the guidelines of the American College of Medical Genetics (ACMG)[35], all DNVs that are not in ExAC and that are in known disease-associated genes or genes with DNV burden in our dataset were categorized as 'likely pathogenic'; however, we did not apply all ACMG criteria to individual DNVs (Methods).

Comparing DNVs between NDDs with and without epilepsy. We compared the frequencies of $DNV_{mis+trunc}$ in NDDs with epilepsy ($NDDs_{EE+uE}$) and $NDDs_{woE}$ across all 107 DNV-enriched genes (logistic regression; Methods). The likelihood of epilepsy increased with the age at the time of recruitment (3-year OR = 1.11, 95% CI = 1.04–1.18, $P = 3 \times 10^{-3}$; individual genes in Supplementary Fig. 8). Sex was not associated with epilepsy status ($P = 0.5$). Individuals with DNV_{mis} were more likely to have epilepsy than individuals with DNV_{trunc} (Fig. 2; $OR_{mis} = 2.1$, 95% CI = 1.6–2.8, $P = 2 \times 10^{-7}$). In line with previous reports[15], we observed this pattern on a single-gene level for *SCN2A* (Firth regression, $OR_{mis} = 23.5$, 95% CI 3.8–277, $P = 0.0003$; Table 1). Confirming previous findings[24,36], we found that DNVs in ion-channel-encoding genes were associated with epilepsy (OR = 6.0, 95% CI = 3.9–9.2, $P = 1 \times 10^{-16}$). Notably, 83% (110/133) of DNVs in ion-channel-encoding genes were DNV_{mis}. However, in the subset of 910 DNVs not in ion channel-encoding genes, DNV_{mis} were still associated with epilepsy (OR = 1.5, $P = 0.005$, 95% CI = 1.1–2.1), thus suggesting that the effect of DNV_{mis} on epilepsy was not entirely driven by ion-channel-encoding genes. We also performed gene set enrichment analyses with all DNVs with 1,766 curated gene sets previously described, separately for DNV_{mis} and DNV_{trunc} (Supplementary Note)[22]. We observed significant differences between individuals with and without epilepsy in 64 gene sets, 62 of them in DNV_{mis}. Moreover, 59 of them were enriched in epilepsy and were mostly related to ion-channel and neuronal function (Supplementary Note and Supplementary Table 8). We observed a higher rate of DNV_{mis} in $NDDs_{EE}$ than in $NDDs_{uE}$, though only with nominal significance (Fisher's exact test, OR = 1.8, 95% CI = 1.04–3.4, $P = 0.03$; Supplementary Fig. 9b). Four genes were more frequently mutated in NDDs with epilepsy ($NDDs_{EE+uE}$) than $NDDs_{woE}$ (Fisher's exact test; Fig. 2a,b, Table 1 and Supplementary Table 9). With the exception of *SCN1A*, the frequencies of DNVs were not significantly different per gene between $NDDs_{EE}$ and $NDDs_{uE}$ for DNV_{mis} or DNV_{trunc} (Supplementary Fig. 9 and Supplementary Table 10).

Evaluation of diagnostic gene panels for epilepsy disorders. Targeted sequencing of disease-specific gene panels is widely used in diagnostics of epilepsies[18,19]. We compared our results to 24 diagnostic panels for epilepsy or EE (Methods; full list in Supplementary Table 11). In total, the 24 different panels covered 358 unique genes (81.5 ± 8.8 genes per panel, mean ± s.d.). Applying these 24 diagnostic panels on our dataset would have detected on average only 59% of $DNV_{mis+trunc}$ in the 33 DNV-enriched genes (Supplementary Fig. 10). However, similarly to most other research studies involving clinical whole-exome sequencing (WES)[7], we were not able to fully assess the extent of potential prescreening. We investigated whether genes in the 24 panels had some evidence of disease association, given the following features that we (and others[23,32]) observed in genes with DNV burden in NDDs: depletion of truncating and missense variants in ExAC[31] controls as well as brain expression (Methods and Supplementary Fig. 8). We restricted this analysis to autosomal-dominant and X-linked panel genes ($n_{dominant+X\text{-}linked} = 191$, Supplementary Table 12). 95% (52/55) of panel genes that had two or more $DNV_{mis+trunc}$ in our study were both constrained and brain expressed. However, only 63% (86/136) of panel genes with one or zero $DNV_{mis+trunc}$ in our study were constrained and brain expressed (Fisher's exact test, OR = 10.2, 95% CI = 3.0–53.0, $P = 2.3 \times 10^{-6}$). We applied evidence of disease association, as defined by the ClinGen Gene Curation Workgroup[37], to those 50 panel genes lacking two of the following criteria: DNV, brain expression, or constrained. We found that nine of the 50 genes had no, eight had limited, and seven had conflicting published evidence of disease association (Supplementary Table 13). Fourteen genes showed moderate, strong, or definitive evidence of association with entities of which neither NDDs nor epilepsy were major features, a result that may partly be explained by a panel design containing genes associated with diseases beyond the spectrum of NDDs (further details in Methods and Supplementary Fig. 10).

Discussion

In this study, we systematically investigated DNVs in NDDs with and without epilepsy. In NDDs with epilepsy, we could scarcely distinguish individuals ascertained to have EE and NDDs with unspecified epilepsy on a genetic level. Thus, we conclude that these phenotypic groups share a spectrum of disease-associated genes predominantly including genes initially reported as EE-associated genes. We identified 33 genes with DNV burden in NDDs with

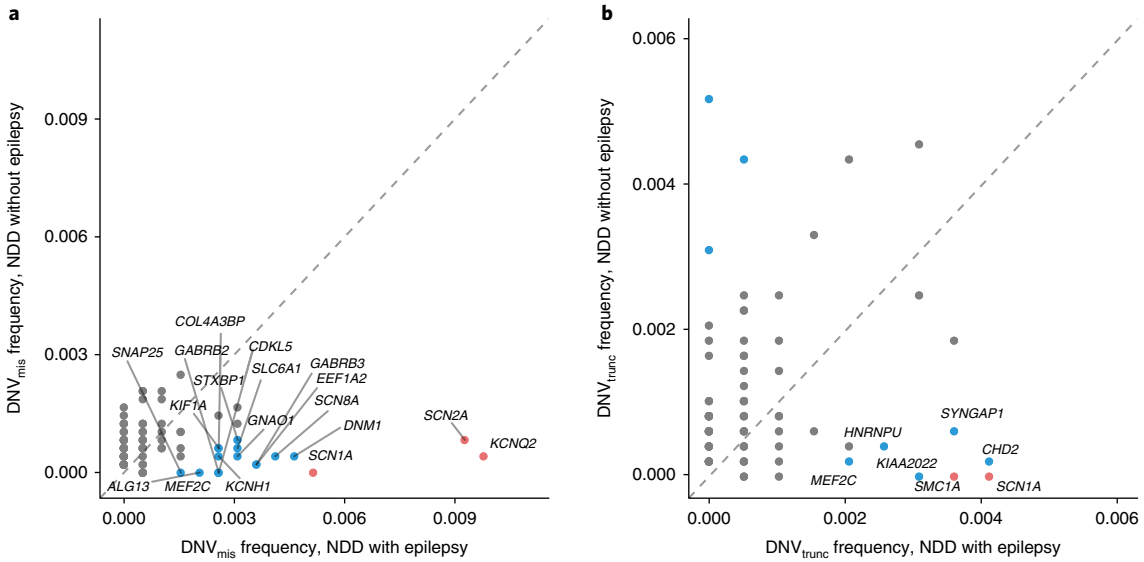

Fig. 2 | DNVs in NDDs with epilepsy ($NDDs_{EE+uE}$, $n = 1{,}874$) versus without epilepsy ($NDDs_{woE}$, $n = 4{,}728$) in 107 genes with significant DNV burden. a, 524 DNV_{mis}. **b**, 561 DNV_{trunc}. Genes with different DNV frequencies in individuals with versus without epilepsy are labeled (two-sided Fisher's exact test; blue, nominal significance, $P < 0.05$; red, significant after correction for 266 tests). The dotted line represents equal frequency of DNVs in NDDs with and without epilepsy. Low-coverage exons and cohorts were excluded from the analysis.

epilepsy, most of which were expressed in the infant brain and depleted in functional variation in ExAC[31], as previously described for NDD genes[23,32]. We found a statistically robust disease association for *SNAP25*, *GABRB2*, and *CACNA1E*, which was previously lacking (Supplementary Note).

We found that individuals with DNV_{mis} were generally more likely to have epilepsy than individuals with DNV_{trunc}. This association was largely driven by ion-channel-encoding genes, thereby confirming longstanding statements that many epilepsy disorders act as channelopathies[2,24,36]. Heterozygous DNV_{mis} have been shown to cause epilepsy via dominant-negative (for example, *KCNQ2* (ref. [38])) or gain-of-function (for example, *SCN8A*[39]) effecNDD with epilepsy compared with NDD without epilepsy, and the strongest enrichment was seen in genes associatedts on ion channels. On the individual gene level, missense variants in *SCN2A*[15] and *SCN8A*[40] are more strongly implicated in epilepsy than protein-truncating variants, as we statistically confirmed for *SCN2A*. Yet, we found that DNV_{mis} were also associated with epilepsy independently of ion-channel-encoding genes. This result may suggest that alterations in protein function may quantitatively play a larger role than haploinsufficiency[41] in the pathophysiology of NDDs with epilepsy than that of NDDs without epilepsy. We also found that 62 gene sets had a greater burden of DNV_{mis} in NDD with epilepsy compared with NDD without epilepsy, and the strongest enrichment was seen in genes associated with ion-channel and neuronal function (Supplementary Note). However, biological interpretations should be made with caution, given that previous studies have found that many of these gene sets share a large number of underlying genes[22], and a subset of genes are disproportionately represented in public gene set databases[42]. We further replicated a previous finding that the rate of epilepsy is correlated with the severity of intellectual disability[4–6], thus suggesting that impaired brain function may contribute to epileptogenesis, or that genetic variants may cause both epilepsy and NDDs. Alternatively, severe epileptic activity may also damage brain function and thereby contribute to NDDs, thus constituting the original definition of EE[9,17]. This possibility is supported by many cases of clinical regression after the onset of epilepsy and improvement of NDDs through seizure control.

In NDDs with epilepsy, we found no genetic differences between unspecified epilepsy and EE, with the exception of *SCN1A* (Supplementary Note). Phenotypic heterogeneity has been described for most EE-associated genes[1,11]; i.e., variants in the same gene can lead to a spectrum of different phenotypes. Due to pleiotropy, individuals who carry a pathogenic DNV in an EE-associated gene and meet the diagnostic criteria of EE may also be eligible for another NDD diagnosis and thus may by chance be assigned to an ASD, DD, or ID and not an EE screening cohort. In line with this hypothesis, we found typically EE-associated seizure types (for example, epileptic spasms) in cohorts with unspecified epilepsy. Some of the diagnostic criteria for EE[1,10] may present ambiguously and consequently lead to uncertainty in terminology[17]. Thus, 43% (21/49) of individuals diagnosed with EE in the Epi4K-E2 (ref. [24]) study initially presented with DD before seizure onset, thus conflicting with the original definition of EE[3,17]. Clear phenotypic distinction between encephalopathic and nonencephalopathic epilepsies may therefore be difficult. Accordingly, mechanisms that result in an encephalopathic course of a genetic NDD remain elusive.

Restricting DNA sequencing or DNA-sequence analysis to panels of known disease-associated genes is widely used in diagnosis for genetic diseases, including epilepsy[19] (100,000 Genomes Project (see URLs)). We confirmed that epilepsy gene panels from diagnostic laboratories differ substantially in gene content[18] and include at least 24 genes with low evidence of disease association (ClinGen criteria[37]). Statistically nonrobust gene–disease associations occasionally result in false-positive reports of causality, thus posing challenges to correct diagnosis in research and clinical settings[11,43]. Our data provide grounds for replacing genes with limited evidence with genes with higher evidence in the design of gene panels for NDDs with epilepsy.

Therapeutic approaches tailored to each patient's underlying genetic variant have successfully been applied for several EEs[2], including treatment with ezogabine in *KCNQ2* encephalopathy[44] or a ketogenic diet in *SLC2A1*-related disorders[45]. In our study, 5.3% of $DNV_{mis+trunc}$ were in eight genes (Supplementary Table 8), for which we confirmed therapeutic consequences with established evidence-based-medicine criteria[34]. This finding reinforces the urgency of making genetic diagnoses in NDDs with epilepsy. We expect that

Table 1 | Genes with exome-wide DNV burden in NDDs with epilepsy

	DNV_{trunc}		DNV_{mis}		$DNV_{mis+trunc}$
Epilepsy	**Yes**	**No**	**Yes**	**No**	**Yes + no**
Gene					
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

Genes are listed in order of decreasing numbers of $DNV_{mis+trunc}$ in NDDs with epilepsy.

with increasing understanding of the underlying pathomechanisms, the group of genetic epilepsies with relevant therapeutic consequences will continue to grow.

URLs. World Health Organization, http://www.who.int/; BrainSpan, http://www.brainspan.org/; 100,000 Genomes Project, http://www.genomicsengland.co.uk/.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-018-0143-7.

Received: 20 March 2017; Accepted: 9 April 2018;
Published online: 25 June 2018

References

1. McTague, A., Howell, K. B., Cross, J. H., Kurian, M. A. & Scheffer, I. E. The genetic landscape of the epileptic encephalopathies of infancy and childhood. *Lancet Neurol.* **15**, 304–316 (2016).
2. Myers, C. T. & Mefford, H. C. Advancing epilepsy genetics in the genomic era. *Genome Med.* **7**, 91 (2015).
3. Scheffer, I. E. et al. ILAE classification of the epilepsies: position paper of the ILAE commission for classification and terminology. *Epilepsia* **58**, 512–521 (2017).
4. McGrother, C. W. et al. Epilepsy in adults with intellectual disabilities: prevalence, associations and service implications. *Seizure* **15**, 376–386 (2006).
5. Thomas, S., Hovinga, M. E., Rai, D. & Lee, B. K. Brief report: prevalence of co-occurring epilepsy and autism spectrum disorder: the U.S. National Survey of Children's Health 2011–2012. *J. Autism Dev. Disord.* **47**, 224–229 (2017).
6. Robinson, E. B. et al. Autism spectrum disorder severity reflects the average contribution of de novo and familial influences. *Proc. Natl. Acad. Sci. USA* **111**, 15161–15165 (2014).
7. Deciphering Developmental Disorders Study. Prevalence and architecture of de novo mutations in developmental disorders. *Nature* **542**, 433–438 (2017).
8. Li, J. et al. Genes with de novo mutations are shared by four neuropsychiatric disorders discovered from NPdenovo database. *Mol. Psychiatry* **21**, 290–297 (2016).
9. Berg, A. T. et al. Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005–2009. *Epilepsia* **51**, 676–685 (2010).
10. Berg, A. T. & Millichap, J. J. The 2010 revised classification of seizures and epilepsy. *Continuum (Minneap. Minn.)* **19**, 571–597 (2013).
11. Helbig, I. & Tayoun, A. A. Understanding genotypes and phenotypes in epileptic encephalopathies. *Mol. Syndromol.* **7**, 172–181 (2016).
12. Trump, N. et al. Improving diagnosis and broadening the phenotypes in early-onset seizure and severe developmental delay disorders through gene panel analysis. *J. Med. Genet.* **53**, 310–317 (2016).
13. Depienne, C. et al. Spectrum of *SCN1A* gene mutations associated with Dravet syndrome: analysis of 333 patients. *J. Med. Genet.* **46**, 183–191 (2009).
14. Weckhuysen, S. et al. Extending the *KCNQ2* encephalopathy spectrum: clinical and neuroimaging findings in 17 patients. *Neurology* **81**, 1697–1703 (2013).
15. Wolff, M. et al. Genetic and phenotypic heterogeneity suggest therapeutic implications in *SCN2A*-related disorders. *Brain* **140**, 1316–1336 (2017).
16. Stamberger, H. et al. STXBP1 encephalopathy: a neurodevelopmental disorder including epilepsy. *Neurology* **86**, 954–962 (2016).
17. Howell, K. B., Harvey, A. S. & Archer, J. S. Epileptic encephalopathy: use and misuse of a clinically and conceptually important concept. *Epilepsia* **57**, 343–347 (2016).
18. Chambers, C., Jansen, L. A. & Dhamija, R. Review of commercially available epilepsy genetic panels. *J. Genet. Couns.* **25**, 213–217 (2016).
19. Lemke, J. R. et al. Targeted next generation sequencing as a diagnostic tool in epileptic disorders. *Epilepsia* **53**, 1387–1398 (2012).
20. Iossifov, I. et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).
21. Zaidi, S. et al. De novo mutations in histone-modifying genes in congenital heart disease. *Nature* **498**, 220–223 (2013).
22. Singh, T. et al. The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. *Nat. Genet.* **49**, 1167–1173 (2017).
23. Lelieveld, S. H. et al. Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nat. Neurosci.* **19**, 1194–1196 (2016).
24. EuroEPINOMICS-RES Consortium. Epilepsy Phenome/Genome Project & Epi4K Consortium. De novo mutations in synaptic transmission genes including DNM1 cause epileptic encephalopathies. *Am. J. Hum. Genet.* **95**, 360–370 (2014).
25. Hamdan, F. F. et al. De novo mutations in moderate or severe intellectual disability. *PLoS Genet.* **10**, e1004772 (2014).
26. de Ligt, J. et al. Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.* **367**, 1921–1929 (2012).
27. Rauch, A. et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* **380**, 1674–1682 (2012).
28. Helbig, K. L. et al. Diagnostic exome sequencing provides a molecular diagnosis for a significant proportion of patients with epilepsy. *Genet. Med.* **18**, 898–905 (2016).
29. Fischbach, G. D. & Lord, C. The Simons Simplex Collection: a resource for identification of autism genetic risk factors. *Neuron* **68**, 192–195 (2010).
30. Samocha, K. E. et al. A framework for the interpretation of de novo mutation in human disease. *Nat. Genet.* **46**, 944–950 (2014).
31. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).

32. Kosmicki, J. A. et al. Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. *Nat. Genet.* **49**, 504–510 (2017).
33. Reimand, J. et al. g:Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.* **44**, W83–W89 (2016).
34. OCEBM Levels of Evidence Working Group. *The Oxford Levels of Evidence* (Oxford Centre for Evidence-Based Medicine, Oxford, 2009).
35. Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
36. Mulley, J. C., Scheffer, I. E., Petrou, S. & Berkovic, S. F. Channelopathies as a genetic cause of epilepsy. *Curr. Opin. Neurol.* **16**, 171–176 (2003).
37. Rehm, H. L. et al. ClinGen: the Clinical Genome Resource. *N. Engl. J. Med.* **372**, 2235–2242 (2015).
38. Weckhuysen, S. et al. *KCNQ2* encephalopathy: emerging phenotype of a neonatal epileptic encephalopathy. *Ann. Neurol.* **71**, 15–25 (2012).
39. Wagnon, J. L. et al. Pathogenic mechanism of recurrent mutations of *SCN8A* in epileptic encephalopathy. *Ann. Clin. Transl. Neurol.* **3**, 114–123 (2015).
40. Blanchard, M. G. et al. De novo gain-of-function and loss-of-function mutations of *SCN8A* in patients with intellectual disabilities and epilepsy. *J. Med. Genet.* **52**, 330–337 (2015).
41. Wilkie, A. O. The molecular basis of genetic dominance. *J. Med. Genet.* **31**, 89–98 (1994).
42. Haynes, W. A., Tomczak, A. & Khatri, P. Gene annotation bias impedes biomedical research. *Sci. Rep.* **8**, 1362 (2018).
43. MacArthur, D. G. et al. Guidelines for investigating causality of sequence variants in human disease. *Nature* **508**, 469–476 (2014).
44. Millichap, J. J. et al. *KCNQ2* encephalopathy: features, mutational hot spots, and ezogabine treatment of 11 patients. *Neurol. Genet.* **2**, e96 (2016).
45. De Giorgis, V. & Veggiotti, P. GLUT1 deficiency syndrome 2013: current state of the art. *Seizure* **22**, 803–811 (2013).

Acknowledgements

We thank all patients and their families who participated in this study, as well as the teams who were involved in recruiting patients and gathering samples and data at the respective study sites. We thank L. Vissers and C. Gilissen for epilepsy and age phenotypes from the cohort of Lelieveld et al.[23] and J. McRae for useful discussions on the DDD cohort[7]. We are grateful to members of the ATGU and the Institute for Human Genetics in Leipzig for insightful discussions. We thank J. Krause for support in figure design and helpful conversations. This work was supported by the Eurocores program EuroEPINOMICS, the Fund for Scientific Research Flanders (FWO), International Coordination Action (ICA) grant G0E8614N, and the University of Antwerp (research fund). H.O.H. was supported by stipends from the Federal Ministry of Education and Research (BMBF), Germany, FKZ: 01EO1501 and the German Research Foundation (DFG): HE7987/1-1. H.S. was supported as a PhD fellow of the Fund for Scientific Research Flanders (1125416N). I.H. and Y.G.W. were supported by DFG grants WE4896/3-1 and HE5415/6-1. R.G. received funding through the EU Seventh Framework Programme (FP7) under the project DESIRE grant N602531. The DDD study presents independent research commissioned by the Health Innovation Challenge Fund (grant HICF-1009-003), a parallel funding partnership between the Wellcome Trust and the Department of Health, and the Wellcome Trust Sanger Institute (grant WT098051). The views expressed in this publication are those of the authors and not necessarily those of the Wellcome Trust or the Department of Health. The research team acknowledges the support of the National Institute for Health Research, through the Comprehensive Clinical Research Network.

Author contributions

H.O.H. performed the analyses and drafted the manuscript. H.O.H. and J.R.L. conceived the study. H.O.H., J.R.L., M.J.D., T.S., D.L. and H.S. contributed to analysis concepts and methods. H.O.H., J.R.L., D.L., I.H., T.S., M.J.D., S.M.S., and S. Weckhuysen interpreted the results. T.S., H.S., R.A.J., H.C., D.C., P.D.J., R.G., K.L.H., B.P.C.K., J.A.K., D.L., T.L., P.M., H.M., R.S.M., B.A.N., A. Palotie, M.P., P.S., S.T., S. Wu, the EuroEPINOMICS RES Consortium, S.T., A. Poduri, Y.G.W., S. Weckhuysen, and I.H. provided patient data or analysis tools. All authors revised and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0143-7.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to H.O.H. or J.R.L.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

EuroEPINOMICS RES Consortium

Zaid Afawi[27], Carolien de Kovel[12,28], Petia Dimova[29], Tania Djémié[5,6,7], Milda Endziniene[30], Dorota Hoffman-Zacharska[31], Johanna Jähn[15], Christian Korff[32], Anna-Elina Lehesjoki[33,34,35], Carla Marini[10], Stefanie H. Müller[15], Deb Pal[36], Niklas Schwarz[23], Kaja Selmer[37], Jose Serratosa[38,39], Ulrich Stephani[15], Katalin Štěrbová[40], Arvid Suls[5,6,7], Steffen Syrbe[41], Inga Talvik[42], Shan Tang[36], Sarah von Spiczak[43,44] and Federico Zara[45]

[21]Division of Clinical Genomics, Ambry Genetics, Aliso Viejo, CA, USA. [27]Department of Physiology and Pharmacology, Tel Aviv University Medical School, Ramat Aviv, Israel. [28]Department of Language and Genetics, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. [29]Epilepsy Center, Department of Neurosurgery, University Hospital 'St. Ivan Rilski', Sofia, Bulgaria. [30]Neurology Department, Medical Academy, Lithuanian University of Health Services, Kaunas, Lithuania. [31]Department of Medical Genetics, Institute of Mother and Child, Warsaw, Poland. [32]Neuropediatrics Unit, University of Geneva, Geneva, Switzerland. [33]Folkhälsan Institute of Genetics, Helsinki, Finland. [34]Neuroscience Center, University of Helsinki, Helsinki, Finland. [35]Research Programs Unit, Molecular Neurology, University of Helsinki, Helsinki, Finland. [36]Department of Basic and Clinical Neurosciences, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK. [37]Department of Medical Genetics, Oslo University Hospital, Oslo, Norway. [38]Neurology Laboratory and Epilepsy Unit, Department of Neurology, IIS-Fundacion Jimenez Diaz, UAM, Madrid, Spain. [39]Centro Investigacion Biomedica en Red de Enfermedades Raras (CIBERER), Madrid, Spain. [40]Department of Pediatric Neurology, Second Faculty of Medicine, Charles University in Prague and University Hospital Motol, Prague, Czech Republic. [41]Department of General Paediatrics, Division of Child Neurology and Inherited Metabolic Diseases, Centre for Paediatrics and Adolescent Medicine, University Hospital Heidelberg, Heidelberg, Germany. [42]Tallinn Children's Hospital, Tallinn, Estonia. [43]Department of Neuropediatrics, Christian-Albrechts-University, Kiel, Germany. [44]Northern German Epilepsy Center for Children & Adolescents, Schwentinental/OT, Raisdorf, Germany. [45]Laboratory of Genetics and Neuroscience, Institute of G. Gaslini, Genoa, Italy.

Methods

Patient cohorts. For this study, we identified 8,529 patients with the following NDDs: DD[7] ($n=4,293$), ASD[20] ($n=2,508$), EE[24,28] ($NDDs_{EE}$, $n=529$), ID[23,25–27] ($n=1,035$), and NDDs with epilepsy[28] ($n=164$). From these cohorts, we selected 6,753 individuals, for whom the presence or absence of epilepsy was ascertained, ~88% of whom had ID (on the basis of the assumption of 81.7% ID in the DDD study[7], 89.8% ID in a diagnostic cohort from AmbryGenetics[28], and 100% ID in all other cohorts). Among individuals with ASD who were phenotyped within the Simon Simplex Consortium[29], we restricted our analysis to patients with ID (IQ <70), because DNVs have been shown to play only a minor role in normal-IQ ASD[6,30]. Previously sequenced trios ($n=1,911$) from unaffected siblings of a child with ASD[20,29] served as control trios. For our main analyses, we stratified this combined cohort of individuals with NDDs for patients comorbid or primarily diagnosed with epilepsy ($NDDs_{EE+uE}$, $n=1,942$)[20,29]. Two EE cohorts and one ID cohort comprising a combined 144 patients were not previously published; one cohort was only partly published (Supplementary Table 1). Medical doctors—mostly clinical geneticists, but also neurologists, pediatricians, and, for ASD[29], some primary-care physicians—reported phenotypes, including the presence of epilepsy, in all patients. Our analysis was based on the assumption that medical professionals are sufficiently qualified to diagnose the presence or absence of epilepsy correctly.

Subphenotypes. We obtained information on specific EE syndromes in 98% (518/529) of individuals with $NDDs_{EE}$. We obtained data on specific seizure types (febrile, focal, spasms, and generalized) for 55% (140/256) and the age of seizure onset for 30% (77/256) of individuals with $DNV_{mis+trunc}$ in genes with DNV burden in $NDDs_{EE+uE}$ (Supplementary Figs. 5 and 6). We did not obtain EEG data per patient. Some patients may have developed epilepsy after inclusion in the study, so we assessed the age at recruitment, which we obtained for 94% (1,087/1,157) of all individuals with NDDs with $DNV_{mis+trunc}$ in DNV-enriched genes (median age at recruitment, 74.8 months). We obtained the age of seizure onset for 30% (77/256) of individuals with epilepsy and $DNV_{mis+trunc}$ in DNV-enriched genes (Supplementary Fig. 5). We identified 30 individuals with potentially epilepsy-relevant brain malformations (abnormalities of neuronal migration, structural abnormalities of the corpus callosum, midbrain, and brainstem as schizencephaly, megalencephgaly, and holoprosencephaly) in individuals with $DNV_{mis+trunc}$ in DNV-enriched genes (29 from DDD[7], 1 from Hamdan et al.[25]). 11 of the individuals (37%) also had seizures.

Whole-exome sequencing of parent–patient trios. In all cohorts, both affected individuals and their unaffected parents underwent WES. Variants that were not present in either parent were considered DNVs. 1,942 individuals with NDDs with epilepsy ($NDDs_{EE+uE}$) had 1,687 DNV_{mis} and 396 DNV_{trunc} (i.e., stop-gain, frameshift, and essential-splice-site mutations). 4,811 individuals with $NDDs_{woE}$ had 4,227 DNV_{mis} and 1,120 DNV_{trunc} (Supplementary Table 2; individual cohorts in Supplementary Fig. 3). The study was approved by the ethics committee of the University of Leipzig (224/16-ek and 402/16-ek) and additional local ethics committees. The DDD cohort of this study was approved by the UK Research Ethics Committee (10/H0305/83, granted by the Cambridge South Research Ethics Committee and GEN/284/12, granted by the Republic of Ireland Research Ethics Committee). A list of all published and unpublished cohorts used in this study is provided in Supplementary Table 1.

Sequencing pipelines of previously unpublished/partly published cohorts (cohorts 8–11). Libraries were prepared from parent and patient DNA, and exomes were captured and sequenced on Illumina sequencers. Raw data were processed and technically filtered with established pipelines at the respective academic or diagnostic laboratories. For Cohort 8 (Ambry Genetics), diagnostic WES was performed on parent–offspring trios at Ambry Genetics (Aliso Viejo, California, USA) in 216 individuals with a history of seizures, as previously described[28]. Genomic-DNA extraction, exome library preparation, sequencing, bioinformatics-pipeline filtering, and data analyses were performed as previously described[46]. The following variant filters were applied to generate a list of high-confidence de novo variant calls: (i) mutation base coverage ≥20× in all members of the trio; (ii) heterozygous read ratio in probands >30% and <80%; (iii) heterozygous read ratio in parents <2%; (iv) genotype quality cutoffs SNV > 100 and indels >300; and (v) exclusion of known sequencing artifacts (on the basis of Ambry Genetics' internal databases). For cohorts 9 (EuroEPINOMICS RES) and 10 (DFG atypical EE), genomic-DNA extraction, exome library preparation, sequencing, bioinformatics-pipeline filtering, and data processing were performed as previously described[47]. The following filtering criteria were applied: read depth >30, frequency of alternative allele between 30% and 70% for the child and not present in the parents, a minimum VQSLOD genotype quality score of −8, Caucasian-population allele frequency <1%, variations on targeted regions + flanking 100 bp. To exclude pipeline-specific artifacts, variants were also filtered against an in-house cohort of variations, which were created with the same analysis pipeline. For cohort 11 (University of Leipzig), genomic-DNA extraction, exome library preparation, sequencing, bioinformatics-pipeline filtering, and data analyses were performed as previously described[48]. Quality filtering of sequencing reads in both parents and children was done according to the following criteria: read depth >20, genotype quality >500, frequency of alternative allele between 30% and 70% for the child and not present in the parents, frequency <1% in internal database, and variant called by at least two different genotype callers.

False-positive rates of DNVs. In cohorts 1 to 4, all DNVs were validated by Sanger sequencing to eliminate false-positive calls. In cohorts 5 to 7, through random selection of variants for Sanger validation, the false-positive rate was estimated to be approximately 1.4% and <5%, respectively. In clinical cohorts 8 to 11, variants defined as being worth reporting back to patients (i.e., variants that have unknown significance or are likely to be pathogenic) are normally validated by Sanger sequencing. Given this experience, the false discovery rates in these cohorts were estimated to be <5%.

Annotation and filtering. DNV files were generated and quality filtered by the individual groups. All DNVs were reannotated with the following pipeline. Variants were annotated with Ensembl's Variant Effect Predictor (http://grch37.ensembl.org/Homo_sapiens/Tools/VEP/), version 82, using database 83 of GRCh37 as the reference genome. Per variant, the transcript with the most severe consequence, as predicted by VEP, was selected for further analyses. Variant impacts, in decreasing order, were 'high', 'moderate', 'modifier', and 'low'. Only protein-altering DNVs (DNV_{mis} or DNV_{trunc} (premature stop codon, essential splice site, or frameshift)) were included in further analyses. Synonymous DNVs (DNV_{syn}) were analyzed as negative controls, because most DNV_{syn} have no effect on the amino acid sequence in the protein. Variants that were present in ExAC[31], an aggregation of 60,706 exome sequences from adults without severe childhood-onset diseases, were excluded after DNV enrichment, because these have been shown to convey no detectable risk for NDDs on a group level[32]. DNV rates per cohort are shown in Supplementary Fig. 2. We did not investigate the pathogenicity of individual DNVs according to the guidelines of the ACMG. However, ACMG criteria PS2 (de novo occurrence, with maternity and paternity confirmed) and PM2 (absence from controls) applied to all DNVs in our cohort. The combination of PS2 and PM2 classifies a variant as at least 'likely pathogenic'. ACMG criteria are applicable only to variants in disease-associated genes[35]. Therefore, all DNVs in known disease-associated genes and genes with genome-wide DNV burden in our dataset were presumed to be likely pathogenic DNVs.

Harmonization of different cohorts. The core analysis of our study compared the observed versus expected enrichment of $DNV_{mis+trunc}$ by using a mutational model in individuals with $NDDs_{EE+uE}$. For this analysis, we were conservative in assuming that every gene was well captured across all cohorts. However, when comparing the DNV burden across different phenotypes, we aimed to separate technical from biological differences with the following methods. In exome sequencing, different capture solutions capture specific exonic regions with different efficiencies. These differences have been shown to be quite stable within and across different samples of the same capture kits[49]. We therefore generated a list of exons that displayed consistent high coverage across different capture solutions. We collected published and internal data, aiming for the highest possible variety of capture kits, by using 3,000 samples and five different capture kits, including NimbleGen SeqCap v2 and v3, and Agilent SureSelect v2, v3, and v5. We generated a list of exons for which at least 80% of all samples had at least 10× coverage. We excluded the oldest capture kits before calculating the high-coverage exons as well as excluding the two oldest cohorts[26,27] from our list of DNVs. Restriction to high-coverage regions resulted in a loss of ~11% of DNVs in DNV-enriched genes. We consequently performed all genotype–phenotype comparisons across cohorts (Figs 1a and 2, and Supplementary Figs. 6–10) with this restricted DNV set. Further, we compared the frequency of DNV_{syn} across all cohorts and excluded cohorts in which DNV_{syn} were not available. In the subset of DNVs in high-coverage exons, rates of supposedly neutral DNV_{syn} were not different between individuals with or without epilepsy (Poisson exact test, $P=0.48$, rate ratio (RR) = 0.99), $NDDs_{uE}$ and $NDDs_{EE}$ ($P=0.65$, RR = 0.94), or NDDs and controls ($P=0.58$, RR = 0.99). The frequency of $DNV_{mis+trunc}$ was also not different between individuals with or without epilepsy ($P=0.5$, RR = 1.02). Our likelihood of identifying $DNV_{mis+trunc}$ in EE-associated genes in the epilepsy cohort was therefore not inflated by a higher baseline rate of $DNV_{mis+trunc}$ in comparison to $NDDs_{woE}$. We reannotated all DNVs in the same manner as described above.

Statistical analysis. All statistical analyses were done with the R programming language (http://www.r-project.org/). Fisher's exact test for count data, Wilcoxon rank-sum test, Poisson exact test, Cochran–Mantel–Haenszel test, logistic regression, Firth regression, Spearman correlation, Welch two-sided *t* test and calculations of empirical *P* values were performed as indicated in the results. For datasets assumed to be normally distributed after visual inspection, data are reported as mean ± s.d. For Poisson exact tests, effect sizes were reported as RR, which is the quotient of the two rates compared in the test. For Fisher's exact test and logistic regression analyses, we reported ORs. The R code used to perform the statistical analyses and figures is available upon request.

DNV enrichment analyses. To identify genes with a significant DNV burden, we compared the numbers of observed with the numbers of expected missense, truncating, and synonymous DNVs per gene, by using an established framework of gene-specific mutation rates[30]. The analysis was done with the R package denovolyzer[50], which compares observed versus expected DNVs with a Poisson exact test. We corrected the obtained *P* values with the Bonferroni method for the number of genes for which gene-specific mutation rates[30] were available ($n = 18{,}225$) and six tests, thus resulting in a *P*-value significance threshold of 5×10^{-7}. Genes that passed that significance threshold for missense, truncating, or both missense and truncating DNVs were considered genes with an exome-wide DNV burden. To compare DNVs among disease groups, DNV enrichment analyses were carried out in the cohort of all patients with NDDs ($n = 6{,}753$) as well as in patients with epilepsy ($NDDs_{EE+uE}$, $n = 1{,}942$) and without epilepsy ($NDDs_{woE}$, $n = 4{,}811$), but only genes with a $DNV_{mis+trunc}$ burden in the NDDs with the epilepsy cohort and the combined NDD cohort are reported.

HPO enrichment analyses. Significantly enriched HPO terms were computed with the R package of g:Profiler[33], by using ordered enrichment analysis on significance-ranked proteins (Supplementary Table 8). Different gene sets were queried by using the background gene set of all 18,225 genes for which gene-specific mutation rates were available[30]. Only terms that were statistically significant with a Bonferroni-corrected $P < 0.01$ were reported, because our negative controls (genes with at least two $DNV_{mis+trunc}$ in healthy controls) were not enriched in any functional categories below this *P* value.

Therapeutic relevance. To assess whether DNVs in our cohort were in therapeutically relevant genes, we searched the literature for treatment recommendations for all established disease-associated genes with at least two $DNV_{mis+trunc}$ in our NDDs with epilepsy cohort. We rated the publications with the standardized score of the Oxford Centre for Evidence-Based Medicine[34]. We reported and considered only genes for which at least one treatment recommendation achieved a level of evidence of II or higher. A list of all genes and levels of evidence is shown in Supplementary Table 9.

Acquisition and processing of brain gene expression data. We downloaded the Developmental Transcriptome dataset of 'BrainSpan: Atlas of the Developing Human Brain' (http://www.brainspan.org/, funded by ARRA awards 1RC2MH089921-01, 1RC2MH090047-01, and 1RC2MH089929-01, 2011). The atlas includes RNA-sequencing data generated from tissue samples of developing postmortem brains of neurologically unremarkable donors, covering 8 to 16 brain structures. We extracted brain expression data from the five donors that were infants 0 to 12 months of age. Per gene, we obtained the median RPKM value of all infant individuals and across brain regions. In all calculations and figures, gene expression values are displayed as median $(\log_2 + 1)$-transformed RPKM values. We defined infant brain gene expression as median $(\log_2 + 1)$-transformed RPKM value > 1. More details about tissue acquisition and sequencing methodology can be found in the BrainSpan documentation online.

Evaluation of gene intolerance to protein-altering variants. We assessed individual gene tolerance to truncating or missense variants in the general population with the pLI score (probability of being loss-of-function intolerant) and missense *Z* score. These scores indicate depletion of truncating and missense variants in ExAC[31] (60,706 individuals without childhood-onset diseases), respectively. We used gene-constraint cutoffs > 0.9 for pLI and > 3.09 for missense *Z* scores, as recommended by the score developers[31]. We calculated empirical *P* values to evaluate whether the pLI scores of exome-wide and nominally DNV-enriched genes were significantly higher than the pLI scores of random gene sets, as previously described[23]. Briefly, we computed the expected pLI for a given gene set with size *n* by randomly drawing 1,000,000 gene sets with size *n* from the total 18,225 pLI annotated genes. We computed how many times the median pLI score of randomly sampled gene sets would exceed the median pLI of the gene set under investigation. To that number, we added 1 and divided by the number of total samplings + 1 to obtain the empirical *P* value.

Comparing DNV in $NDDs_{EE}$, $NDDs_{uE}$ and $NDDs_{woE}$. We investigated $DNV_{mis+trunc}$ in $NDDs_{EE+uE+woE}$ across all 107 genes that were DNV enriched in $NDDs_{EE+uE}$, $NDDs_{woE}$ and/or $NDDs_{EE+uE+woE}$. We restricted our analysis to DNVs not in ExAC and in high-coverage regions. To investigate whether age at the time of recruitment, sex or variant class (DNV_{mis}/DNV_{trunc}) influenced the presence of epilepsy, we tested these parameters as covariates in a logistic regression model with epilepsy as response variable. We aimed to explore whether DNVs in NDDs with epilepsy, compared with NDDs without epilepsy, might be associated with ion channels, given the long-standing hypothesis that many epilepsies are channelopathies[36]. We extracted a comprehensive gene set of 237 known ion-channel-encoding genes from 1,766 previously described[22] curated gene sets derived from public pathway databases and publications (Supplementary Note). To investigate whether ion-channel-encoding genes were associated with epilepsy, we included an annotation as an ion-channel-encoding gene as a categorical predictor in the logistic regression model. We used Firth regression to assess the effect of variant class on the presence of epilepsy for individual genes. We used Fisher's exact test to compare the frequencies of DNVs per gene between phenotype groups. To account for multiple testing, we corrected *P* values for the number of tests performed (Bonferroni method).

Diagnostic gene panels for epileptic encephalopathy/comprehensive epilepsy from 24 academic or commercial providers. We set out to compare our results to diagnostic gene panels for epileptic encephalopathy of international commercial and academic providers. We searched the Genetic Testing Registry (GTR)[51] of NCBI (January 2017) for providers of tests for 'Epileptic encephalopathy, childhood-onset' and identified 16 diagnostic epilepsy panels. We excluded three panels with <20 or >200 genes, and added 11 additional diagnostic providers not registered at GTR to evaluate 24 diagnostic panels targeting epilepsy in general ($n = 11$) or EE specifically ($n = 13$). The gene content covered in each of the 24 gene panels can be found in Supplementary Table 11. Gene lists were freely available for download from the respective providers' websites. For each of the 33 genes with DNV burden in NDDs with epilepsy, we calculated the proportion at which they were included in 24 commercial or academic providers of gene panels for epileptic encephalopathy/comprehensive epilepsy. For each gene, we then multiplied the percentage of inclusion in any of the 24 panels by the total number of $DNV_{mis+trunc}$ of that gene in the cohort of 1,942 individuals with $NDDs_{EE+uE}$.

We investigated whether there were genes in the 24 diagnostic gene panels without evidence of implication in NDDs with epilepsy. We focused on 191 dominant or X-linked panel genes (listed in Supplementary Table 14). We tested these genes for three criteria of association with NDDs with epilepsy: first, whether genes had at least two $DNV_{mis+trunc}$ in our study; second, whether genes were expressed in the infant brain, as defined by a median RPKM of all samples and brain regions > 1; and third, whether genes had a pLI > 0.9 or missense *Z* score > 3.09, indicating intolerance to truncating or missense variants[31]. We intersected these lists to nominate genes that did not display features of DNV-enriched genes in this study. To these genes, we applied ClinGen criteria[37] for gene–disease association.

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability. All data used for computing results supporting the findings of this study are available within the paper and its supplementary information files. Raw sequencing data of published cohorts are referenced at the respective publications. Raw sequencing data for the EuroEPINOMICS RES cohort have been deposited in the European Genome-phenome Archive (EGA) under accession code EGAS00001000048. Raw sequencing data for cohort 10 (DFG atypical EE) will be deposited in a public repository after finalization of the individual project.

References

46. Farwell, K. D. et al. Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genet. Med.* **17**, 578–586 (2015).
47. Huppke, P. et al. Activating de novo mutations in *NFE2L2* encoding NRF2 cause a multisystem disorder. *Nat. Commun.* **8**, 818 (2017).
48. Trujillano, D. et al. Clinical exome sequencing: results from 2819 samples reflecting 1000 families. *Eur. J. Hum. Genet.* **25**, 176–182 (2017).
49. Chilamakuri, C. S. et al. Performance comparison of four exome capture systems for deep sequencing. *BMC Genomics* **15**, 449 (2014).
50. Ware, J. S., Samocha, K. E., Homsy, J. & Daly, M. J. Interpreting de novo variation in human disease using denovolyzeR. *Curr. Protoc. Hum. Genet.* **87**, 7.25 (2015).
51. Rubinstein, W. S. et al. The NIH genetic testing registry: a new, centralized database of genetic tests to enable access to comprehensive information and improve transparency. *Nucleic Acids Res.* **41**, D925–D935 (2013).

nature research

Corresponding author(s): Henrike Heyne, Johannes Lemke

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Experimental design

1. Sample size

   Describe how sample size was determined.

   This is the currently largest cohort aiming to discover genes associated with NDD with epilepsy with a hypothesis-free exome-wide approach. The largest cohort before this study with the same aim was the Epi4k-EuroEpinomics consortium. With their sample size of n=356 the study could identify 7 genes with an exome-wide burden of de novo variants. As our cohorts constitutes 5x that sample size, we consequently hoped to identify substantially more genes associated with NDD with epilepsy. Finding 33 disease-associated genes met these predictions

2. Data exclusions

   Describe any data exclusions.

   We excluded individuals, for whom the presence or absence of epilepsy could not be ascertained and individuals with IQ > 70 (when IQ was available). We excluded cohorts without synonymous de novo variants (DNV) or outlier rates of synonymous DNV. We excluded cohorts with old DNA capture kits and exons with inconsistent coverage between capture kits for all analyses involving comparisons between different cohorts.

3. Replication

   Describe the measures taken to verify the reproducibility of the experimental findings.

   Not applicable, as the study does not contain experimental data.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   Not applicable, as the study does not contain experimental data.

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   Not applicable, as the study does not contain experimental data.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☒	☐ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ Test values indicating whether an effect is present *Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.*
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

Post-sequencing data analysis pipelines of the four unpublished cohorts have been described previously. For Cohort 8 (Ambry Genetics), data analyses of diagnostic WES were performed with software previously described in Farwell, K.D. et al., Genetics in Medicine, 2015. For cohorts 9 (EuroEPINOMICS RES) and 10 (DFG atypical EE), data processing were performed with software described in Huppke, P. et al., Nature Communications, 2017. For cohort 11 (University of Leipzig), data analyses were performed with the software described in Trujillano, D. et al., European Journal of Human Genetics, 2017.
Variants were annotated with Ensembl's Variant Effect Predictor (http://grch37.ensembl.org/Homo_sapiens/Tools/VEP), version 82.
All statistical analyses were done with the R programming language, version 3.3.3 (www.rproject.org). The R code used to perform the statistical analyses and figures is available upon request. R packages used for specific analyses were denovolyzeR for de novo enrichment analyses and gProfileR for Human Phenotype Ontology enrichment analyses, as referenced in the manuscript.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

No unique materials were used in this study.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used in this study.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

No cell lines were used in this study.

b. Describe the method of cell line authentication used.

Describe the authentication procedures for each cell line used OR declare that none of the cell lines used have been authenticated OR state that no eukaryotic cell lines were used.

c. Report whether the cell lines were tested for mycoplasma contamination.

Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination OR state that no eukaryotic cell lines were used.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

Provide a rationale for the use of commonly misidentified cell lines OR state that no commonly misidentified cell lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.

No animals were used in this study.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

For this study, we ascertained 8,529 patients with the following neurodevelopmental disorders (NDD): developmental delay (DD, n=4293), autism spectrum disorder (ASD, n=2508), epileptic encephalopathy (EE, n=529), intellectual disability (ID, n=1035), and epilepsy with NDD (n=164). From this cohort, we selected 6753 individuals, for which the presence or absence of epilepsy was ascertained and of whom ca. 88% had ID (based on assumption of 81.7% ID in the DDD study, 89.8% ID in a diagnostic cohort from AmbryGenetics and 100% ID in all other cohorts.) Among individuals with ASD who were phenotyped within the Simon Simplex Consortium, we restricted our analysis to patients with ID (IQ < 70) as it has been shown that DNV play only a minor role, in normal IQ. Previously sequenced trios (n = 1911), from unaffected siblings of a child with ASD, served as control trios. For our main analyses, we stratified this combined cohort of patients with NDD for patients comorbid or primarily diagnosed with epilepsy (n=1942). Two EE cohorts and one ID cohort comprising a combined 144 patients were not previously published; one cohort was only partly published (see Supplementary Table 1). Medical doctors, mostly clinical geneticists, but also neurologists, paediatricians and for ASD some primary care physicians reported out phenotypes, including presence of epilepsy, in all patients. Our analysis is based on the assumption that medical professionals are sufficiently qualified to diagnose the presence or absence of epilepsy correctly. Individuals with EE were diagnosed with following specific syndromes: IS (n = 243), LGS (n = 145), electrical status epilepticus in sleep (ESES, n = 42), myoclonic-atonic epilepsy (MAE, n = 39), Dravet syndrome (DS, n = 16), unspecified EE (n = 44). Some patients may have developed epilepsy after inclusion in the study, so we ascertained age at recruitment (median of 74.8 months for individuals with DNV in DNV-enriched genes). Six of eight NDD cohorts (n = 6037) included individuals with as well as without epilepsy. Of these, 20.3% of patients had epilepsy.