The Spectrum of *de novo* Variants in Neurodevelopmental Disorders with Epilepsy

Henrike O. Heyne[1,2], EuroEPINOMICS RES Consortium*, Rami Abou Jamra[1], Hande Caglayan[3], Dana Craiu[4], Peter De Jonghe[5,6,7], Renzo Guerrini[8], Katherine L. Helbig[9], Bobby P. C. Koeleman[10], Jack A. Kosmicki[11,26,27], Tarja Linnankivi[12,13], Patrick May[14], Hiltrud Muhle[15], Rikke S. Møller[16,17], Bernd A. Neubauer[18], Aarno Palotie[11], Manuela Pendziwiat[15], Hannah Stamberger[5,6,7], Pasquale Striano[19], Sha Tang[9], Sitao Wu[9], Mark Daly[11,26,27], Annapurna Poduri[20], Yvonne G. Weber[21], Sarah Weckhuysen[5,6,7], Sanjay M. Sisodiya[22,23], Ingo Helbig[15,24], Dennis Lal[11,25,26], Johannes R. Lemke[1]

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

[1] Institute of Human Genetics, University of Leipzig Hospitals and Clinics, Leipzig, Germany
[2] IFB AdiposityDiseases, Leipzig University Medical Center, Leipzig, Germany
[3] Department of Molecular Biology and Genetics Bogaziçi University Istanbul, Turkey
[4] "Carol Davila" University of Medicine Bucharest, Department of Clinical Neurosciences (No.6), Pediatric Neurology Clinic, Alexandru Obregia Hospital, Bucharest, Romania
[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] Pediatric Neurology and Neurogenetics Unit and Laboratories A. Meyer Children's Hospital-University of Florence, Florence, Italy
[9] Division of Clinical Genomics, Ambry Genetics, Aliso Viejo, California, USA
[10] Department of Medical Genetics University Medical Center Utrecht, Utrecht, The Netherlands
[11] Broad Institute, Stanley Center for Psychiatric Research, Cambridge, MA, USA
[12] Folkhälsan Institute of Genetics Helsinki Finland
[13] Research Programs Unit, Molecular Neurology, University of Helsinki, Helsinki, Finland
[14] Luxembourg Centre for Systems Biomedicine, Esch-sur-Alzette, University of Luxembourg, Luxembourg
[15] Department of Neuropediatrics, University Medical Center Schleswig-Holstein, Christian-Albrechts University, Kiel, Germany
[16] The 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, Genova, Italy
[20] Epilepsy Genetics Program, Department of Neurology, Division of Epilepsy and Clinical Neurophysiology, Boston Children's Hospital, Boston, MA, USA

[21] Department of Neurology and Epileptology, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany
[22] Department of Clinical and Experimental Epilepsy NIHR University College London Hospitals Biomedical Research Centre UCL Institute of Neurology, London, United Kingdom
[23] The Epilepsy Society Chalfont-St-Peter Bucks, United Kingdom
[24] Division of Neurology, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA
[25] Cologne Center for Genomics (CCG), Cologne, Germany
[26] Analytic and Translational Genetics Unit (ATGU), Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA
[27] Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA.

Abstract:	127 words
Main text:	2978 words
Figures:	5
Tables (box):	1
Methods:	2832 words

Corresponding authors:

Prof. Johannes Lemke (johannes.lemke@medizin.uni-leipzig.de)
and
Dr. Henrike Heyne (henrike.heyne@medizin.uni-leipzig.de)

Institute of Human Genetics
Philipp-Rosenthal-Str. 55
University of Leipzig Hospitals and Clinics
D-04103 Leipzig
Germany

Tel: +49-341-97-23800
Fax: +49-341-97-23819

Acknowledgements

We like to thank all patients and their families who participated in this study. We thank Johannes Krause for support in figure design and helpful discussions. Supported by the Eurocores program EuroEPINOMICS, the Fund for Scientific Research Flanders (FWO), the International Coordination Action (ICA) grant G0E8614N, and the University of Antwerp (research fund). H.O.H was supported by the Federal Ministry of Education and Research (BMBF), Germany, FKZ: 01EO1501. H.S. is PhD fellow of the Fund for Scientific Research Flanders (1125416N). R.G. received funding through the EU 7th Framework Program (FP7) under the project DESIRE grant N602531.

Abstract

Neurodevelopmental disorders (NDD) with epilepsy constitute a complex and heterogeneous phenotypic spectrum of largely unclear genetic architecture. We conducted exome-wide enrichment analyses for protein-altering de novo variants (DNV) in 7088 parent-offspring trios with NDD of which 2151 were comorbid with epilepsy. In this cohort, the genetic spectrum of epileptic encephalopathy (EE) and nonspecific NDD with epilepsy were markedly similar. We identified 33 genes significantly enriched for DNV in NDD with epilepsy, of which 27.3% were associated with therapeutic consequences. These 33 DNV-enriched genes were more often associated with synaptic transmission but less with chromatin modification when compared to NDD without epilepsy. On average, only 53% of the DNV-enriched genes were represented on available diagnostic sequencing panels, so our findings should drive significant improvements of genetic testing approaches.

Introduction

Epilepsy disorders, defined by recurrent, unprovoked seizures, affect about 50 million people worldwide (www.who.int, 03/2017). Onset of severe epilepsy is most common in infancy and early childhood[1,2] similar to neurodevelopmental disorders (NDD), such as developmental delay (DD), intellectual disability (ID) or autism spectrum disorders (ASD). Moreover, epilepsies in infancy and early childhood are often comorbid with NDD and vice versa: approximately 20% of individuals with NDD have seizures[3-5].

We refer to epilepsy comorbid with any diagnosis of ID, DD or ASD (Figure 1) as 'NDD with epilepsy'. The heterogeneous spectrum of NDD with epilepsy comprises epileptic encephalopathy (EE) defined as a progressive electroclinical syndrome associated with specific epileptic activity and/or EEG patterns and DD or regression[2] as well as a majority of nonspecific epilepsy phenotypes[1]. Phenotypic categorisation of epilepsies has enabled delineation of clinically-recognizable syndromes and identification of numerous associated genes. However, for multiple genes implicated in EE, genetic variation may cause different subtypes of EE (e.g. *SCN1A*, *KCNQ2, GABRA1*), nonspecific NDD with seizures or even NDD without seizures (e.g. *SCN2A, STXBP1, SYNGAP1*)[6]. Many affected genes in both NDD and EE are involved in processes of synaptic transmission[7-9]

Application of a mutational model[10] to detect enrichment for *de novo* variants (DNV) has proven to be a powerful approach for identification of disease-associated genes, as in ID, congenital heart disease and ASD[11-13]. Due to phenotypic and genetic heterogeneity of such disorders, these hypothesis-free exome-wide analyses typically demand large cohort sizes. For EE, the largest exome-wide DNV burden study published to date[8] comprises 356 parent-offspring trios revealing seven genes with exome-wide significance.

Several studies demonstrated that the spectrum of genes affected by DNV overlaps between different NDD, including EE phenotypes[14,15]. However, there is currently no systematic investigation on the spectrum of affected genes in NDD with epilepsy. We explored the DNV spectrum in 2151 parent-offspring trios of patients with NDD with epilepsy, searching for over-burdened genes and aiming for insights into the underlying genetic architecture. Furthermore, we investigated the potential impact of our findings on designing genetic testing approaches and assessed the extent of therapeutically relevant diagnoses.

Results

Description of dataset

We performed an exome-wide DNV enrichment analysis of parent-offspring trios of nine published[4,8,11,12,16-20], two partly published[21,22] and three unpublished cohorts of 7088 individuals with NDD stratifying for the 2151 cases with epilepsy (Supplementary Table 1, Figure 1). These 2151 patients were ascertained for either EE (n = 849) or nonspecific NDD (DD, low-IQ ASD, ID) comorbid with seizures (n = 1302) and exhibited 1896 protein-altering (i.e. missense and truncating, including essential splice site) DNV (Supplementary Table 2). We re-annotated the original data files using the same pipeline to ensure uniformity with respect to variant annotation. To improve power, we excluded DNV present in ExAC[23], as these have been shown to confer no detectable risk to NDD[24]. In the ASD cohort (phenotype data from the Simon Simplex Consortium[25]) we included only individuals with IQ < 80 (defined as "low IQ") as several studies have found DNV only associated with low-IQ ASD[10,26].

DNV in EE genes in NDD patients with different primary diagnoses

We first asked whether EE and NDD with epilepsy might both be over-burdened for mutations in genes associated with EE. Comparing protein-altering DNV in 50 established autosomal dominant or X-linked EE genes (updated from[27], Supplementary Table 3) normalized to cohort size, we found no significant differences between cohorts ascertained for EE (15.7%±4.2, mean ± sd) and non-specific NDD with epilepsy (13.4%±4.7, Wilcoxon-Rank test, p-value = 0.3, Figure 2A, Supplementary Figure S1). In contrast, this proportion was significantly reduced in three different NDD cohorts (DD[21], ASD[11], and ID[20]) when only considering patients without epilepsy (2.7%±0.7, two-sided Wilcoxon rank sum test, p-value = 0.003) and in 1911 healthy controls (0.3%)[12]. Within these three NDD cohorts[21,11,20], we also detected about 5-fold more DNV in EE genes in 1203 patients with seizures compared to 4937 individuals without seizures (Cochran–Mantel–Haenszel test, p-value 1.6×10^{-44}, common OR 4.98, 95%-CI: 3.9 to 6.35, Figure 2B). This suggests a markedly overlapping genetic spectrum of EE and nonspecific NDD with epilepsy. We subsequently performed DNV enrichment analyses on the combined cohort of NDD with epilepsy.

Enrichment of protein-altering DNV in 2151 individuals with NDD with epilepsy

By comparing the observed versus the expected numbers of protein-altering DNV, we revealed a global enrichment of truncating (2.5 fold, p= 1×10^{-65}) and missense DNV (1.05 fold, p=0.038) and identified 33 exome-wide significantly burdened genes (Box 1A), of which *SCN1A* (n=22), *SCN2A* (n=22), and *KCNQ2* (n=20) were most frequently mutated. Among these 33 genes, *GABRB2, SNAP25, HNRNPU, ARID1B, KCNH1* and *COL4A3BP* had hitherto only limited published evidence for association with NDD with epilepsy. For *SNAP25* (discussed in Supplementary Note) there has only been one report on a single case with a DNV[28] (details of patients with DNV in *SNAP25* see Supplementary Table 4). The distribution of missense versus truncating variants within the 33 genes largely agrees with pathogenic DNV reported in ClinVar[29] (Supplementary Figure S2).

With the exception of *ARID1B*, we observed that the 33 significant genes were more frequently mutated in NDD with epilepsy (n=2151) versus NDD without epilepsy (n=4937). In total, for 23 genes the different burden was at least nominally significant, and 11 of these surpassed multiple testing correction (Fisher's Exact test, Bonferroni correction, Figure 3). For the same genes, patients without epilepsy were younger at time of sequencing (generalized linear model, p-value = 0.0002) than patients with epilepsy, leaving open the possibility that some individuals might have been recruited prior to seizure onset (Supplementary Figure S3, Supplementary Note).

SCN1A was the only gene with significantly different DNV frequencies (Fisher's Exact test, Bonferroni p-value corrected for 33 tests: 0.01, Supplementary Table 5) between EE (17/841 individuals) and NDD with epilepsy (5/1310 individuals).

Beyond the 33 exome-wide significant genes, another 155 genes were nominally enriched (not surpassing the multiple testing correction for 18,225 genes) and affected by at least two protein-altering DNV in our cohort (Supplementary Table 6). After removal of 262 patients with DNV in any of the 50 established EE genes (Supplementary Table 3), two additional genes (*PPP2R5D, SMARCA2*) surpassed exome-wide significance for association with NDD with epilepsy (Supplementary Table 7).

By collectively analysing all patients with NDD with or without seizures (n= 7088), we found exome-wide enrichment for DNV in 84 genes with *BTF3, CNOT3, FBXO11* and *RAC1* with previously no, and *CACNA1E, CHD3, PPM1D, TCF7L2, TLK2* and *TRIP12* with previously limited evidence for association with NDD (criteria of ClinGen Gene Curation Workgroup[30], Supplementary Note, Supplementary Table 8).

Nominating candidate genes among nominally DNV-enriched genes

Among the 155 nominally DNV-enriched genes (Supplementary Table 6), many may truly be associated with NDD with epilepsy, but failed to surpass the multiple testing

correction threshold due to sample size. We therefore further investigated biological and evolutionary properties of these 155 genes, using the 33 exome-wide significant genes as positive, and genes nominally enriched for DNV in controls as negative references. Since disease onset of NDD with epilepsy is typically in infancy and early childhood, we evaluated expression levels of genes associated with NDD with epilepsy in the developing infant brain. At a group level, the 33 significant genes showed highest brain expression levels, compared to slightly lower expression of the 155 genes nominally enriched in NDD with epilepsy, followed by significantly lower brain expression of genes nominally enriched in controls (Figure 4A). Furthermore, the 155 nominally-enriched genes in cases displayed significantly higher scores for loss-of-function (pLI score) and missense variant intolerance (missense z-score) compared to genes nominally enriched in controls (Figures 4B and 4C). Based on these observations, we selected 94 out of the 155 nominally enriched genes using criteria of brain expression values (log2+1-transformed RPKM > 1) and reduced tolerance to nonsynonymous genetic variation (pLI score > 0.9 or missense z-score > 3.09). In a protein–protein interaction network[31], in addition, 43 of these 94 genes were part of a functional module ("gene community"[32]) including at least one exome-wide significant gene. Thus, we consider these 43 genes as high-priority candidates for association with NDD with epilepsy (Box 1B, Supplementary Table 9, Supplementary Table 6).

Box 1,

A, Genes exome-wide DNV-enriched in NDD with epilepsy
ALG13, ARHGEF9, ARID1B, ASXL3, CDKL5, CHD2, COL4A3BP, DNM1, DYRK1A, EEF1A2, FOXG1, GABRB2, GABRB3, GNAO1, GRIN2A, GRIN2B, HNRNPU, KCNH1, KCNQ2, KIAA2022, MECP2, MEF2C, PURA, SCN1A, SCN2A, SCN8A, SLC35A2, SLC6A1, SMC1A, SNAP25, STXBP1, SYNGAP1, WDR45
B, Candidate genes for NDD with epilepsy
ANK3, ATP1A3, CACNA1A, CACNA1E, CAMK2B, CAPRIN1, CASK, CHD8, CNOT3, DDX3X, DYNC1H1, EHMT1, G3BP1, GABRA1, GATAD2B, GRIN1, HUWE1, KCNB1, KIF1A, KIF5C, LRP1, LRP4, MAP2, MAP2K1, MAP3K1, MAP4K4, MAST1, MED12, MED13L, PACS2, PAPD5, PRPF8, RAC1, SCN3A, SLC1A2, SMARCA2, SON, TAF1, TBL1XR1, TRRAP, U2AF2, YWHAG, ZMYND8

Functional enrichment analyses

Functional enrichment analyses[33] with the 33 exome-wide DNV-enriched genes revealed "Epileptic Encephalopathy" (HP:0200134, Bonferroni p-value 8×10^{-22}) as the most significant human phenotype ontology (=HPO) term, even when excluding 849 patients that were diagnosed with EE (Bonferroni p-value 1.5×10^{-18}). Most other highly significant HPO terms were seizure- or behavior-related (Supplementary Table 10, Supplementary Figure S4A). Biological pathway terms (Gene Ontology = GO, Kegg [Kyoto Encyclopedia of Genes and Genomes] and Reactome) enriched in the 33 DNV-enriched genes were predominantly connected to synaptic function (GO Biological Process "synaptic signalling" 6.4×10^{-9}, Supplementary Table 10, Supplementary Figure S4A), as previously observed for EE[8], and for ASD and ID[7,9].

We observed less significant enrichment in synaptic GO-categories (GO "synaptic signalling" p-value 0.0024) and high significance of chromatin-related categories in genes DNV-enriched in NDD without epilepsy (n = 4937, Supplementary Figure S4B). Searching for keywords "synaptic" and "chromatin" among all enriched GO categories, we observed more genes affected by DNV related to synaptic functions (Fisher's Exact test, p-value = 3.6×10^{-14}, OR 2.3, 95%- CI 1.8 to 2.8) in patients with NDD with epilepsy,

but more genes related to chromatin organization in patients with NDD without epilepsy (p-value = 2.3×10^{-9}, OR 2.2, 95%- CI 1.7 to 3.0).

The 43 candidate genes were enriched in similar HPO categories (e.g. "Febrile seizures", HP:0002373, Bonferroni p-value 4.3×10^{-4}) and biological pathway terms as the 33 significant genes (e.g. "Transmission across Chemical Synapses", Bonferroni p-value 0.000125, Supplementary Table 10, third sheet, Supplementary Figure S4C). As a negative control, we also tested i) 51 nominally DNV-enriched genes tolerant to loss-of-function and missense variation or not expressed in the infant brain and not annotated in a protein network with exome-wide significant genes and ii) 44 genes nominally DNV-enriched in healthy controls. Neither gene set was significantly enriched for any functional category.

Evaluation of diagnostic gene panels for seizure disorders using results from this study

Targeted sequencing of disease-specific gene panels is widely employed in diagnostics of epilepsies[27]. We compared our results to 24 diagnostic panels for epilepsy or EE, identified through the Genetic Testing Registry[34] and internet research (full list in Supplementary Table 11).

In total, the 24 panels covered 358 unique genes (81.5 ± 8.8 genes/panel, mean ± sd). Of the 33 DNV-enriched genes, on average 17.6 ± 1.1 were included per panel with CDKL5, SCN1A, SCN2A, SCN8A and STXBP1 covered by all 24 providers (Figure 5A).

We investigated whether genes within the 24 panels lacked evidence for association with NDD with epilepsy, according to characteristics of exome-wide DNV enriched genes described above (Figure 4). We restricted this analysis to autosomal dominant and X-linked acting genes ($n_{dominant+x\text{-}linked}$=191, Supplementary Table 12). Of these genes, 131 were not DNV-enriched in our study (nominal significance), of which 49 failed at least one of the two characteristics brain expression or functional constraint. Among DNV-enriched genes only four genes failed these criteria (Fisher's exact test, p-value 3.7×10^{-6}, OR 8.3). This result remained significant when excluding exome-wide significant genes from the analysis (p-value = 0.00083, OR 8.3). When applying evidence of disease association as defined by the ClinGen Gene Curation Workgroup[30], we found that ten of the 49 genes had no, seven limited and seven conflicting published evidence for disease association. Thirteen genes showed moderate, strong or definitive evidence for association to entities where neither NDD nor epilepsy were major features which may partly be explained by a panel design containing genes associated with diseases beyond NDD with epilepsy (for further details see Figure 4B and Supplementary Table 13).

Discussion

Comparing individuals with a primary diagnosis of EE versus patients with primary diagnoses of other types of NDD with epilepsy, we unexpectedly observed similar proportions of DNV in known EE genes, substantially differing from patients with NDD without seizures as well as healthy controls. Consequently, in the subsequent functional enrichment analyses, "Epileptic encephalopathy" and other epilepsy-related categories were the most significant HPO terms in DNV-enriched genes in NDD with epilepsy regardless of whether or not patients with EE were excluded from the analysis. Thus, we conclude that EE and non-specific NDD with epilepsy at least partly share a spectrum of disease genes corresponding to the known spectrum of EE genes. This also agrees with our phenotypic categorisation of EE as a subtype of NDD with epilepsy (Figure 1). This phenomenon can be explained by patients with multiple symptoms being eligible for different primary diagnoses (such as ASD, EE, DD, ID). An example for this is *SCN1A*, where all respective individuals (n_{ID}=1, n_{DD}=2, n_{ASD}=2) with protein-altering DNV in this gene have epilepsy (Supplementary Table 15). Secondly, variants in the same gene can

lead to a spectrum of different phenotypes, such as for *SCN1A*[35,36], *KCNQ2*[37], *SCN2A*[38] or *STXBP1*[39], where some patients fulfil diagnostic criteria of EE while others present non-specifically with NDD with epilepsy.

When comparing NDD with and without epilepsy, we find an overrepresentation of genes related to synaptic function in epilepsy and an overrepresentation of genes related to chromatin modelling in NDD without epilepsy. This points towards different pathomechanisms contributing to either NDD with or without epilepsy. As synaptic function has been linked to ASD and ID in multiple studies[7,9], it currently remains unknown to what degree the signal may come from patients with epilepsy within these cohorts.

While we find evidence for a distinct genetic spectrum in NDD with epilepsy, we also find genetic overlap with NDD without epilepsy. When comparing variant frequencies in the 33 DNV-enriched genes in NDD with epilepsy to NDD without seizures, 23 genes were more frequently affected in NDD patients with epilepsy (of these 11 survived multiple testing, Figure 2). This suggests that DNV in these 11 genes are more likely to result in the development of epilepsy than DNV in other NDD genes. *SCN1A* is the only gene with significantly different DNV frequencies between EE (17/841 individuals) and NDD with epilepsy (5/1310 individuals). This is not unexpected, since *SCN1A* is the most frequently mutated gene in Dravet syndrome, usually presenting as a clinically recognisable EE entity[35,36]. In conclusion, the genetic spectrum of nonspecific NDD with epilepsy appears markedly similar to EE while also showing some overlap with NDD without epilepsy. In agreement with this, a recent broadening of the EE term to "developmental and epileptic encephalopathy" has been suggested[40].

We identified an exome-wide significant burden of protein-altering DNV for 33 genes in NDD with epilepsy (see Box 1A). Excluding patients with DNV in established EE genes from the enrichment analysis resulted in two additional genes surviving multiple testing (*PPP2R5D* and *SMARCA2*). Our data thus statistically confirmed limited or moderate evidence for association with NDD with epilepsy for *SNAP25* (discussed in Supplementary Note), *GABRB2*, *ASXL3*, *HNRNPU*, *SMC1A*, *ARID1B*, *KCNH1*, *COL4A3BP* and *PPP2R5D* for which previously several small scale and single case studies supported association to NDD with epilepsy. By collectively analysing all patients with NDD (n= 7088) with and without epilepsy, we found genome-wide enrichment for 84 genes (for details see results and Supplementary Table 7). Based on observations in exome-wide significant genes, we selected 43 candidate genes for NDD with epilepsy out of 155 nominally DNV-enriched genes with at least two DNV that were brain expressed, under evolutionary constraint and showed evidence for protein–protein interaction with exome-wide significant genes.

To investigate the impact of our results on diagnostics, we evaluated gene panels from 24 diagnostic labs. Likely due to lack of unbiased data on the underlying mutational spectrum of NDD with epilepsy, gene panels of diagnostic providers for this entity differ substantially in gene content (Supplementary Table 6). We provide evidence for 33 genes exome-wide significantly associated with NDD with epilepsy. At present, 17.6 ± 1.1 of these 33 genes (53.3%) are included per panel for EE or NDD with epilepsy by 24 different academic and commercial providers.

For some genes, evidence for association with NDD with epilepsy is low and sometimes only based on isolated case reports[30]. Failure to detect mutations in alleged disease genes in large cohorts may be due to chance, especially for genes with low mutation rates, but may also result from incorrect gene-disease association[41]. We show that genes on diagnostic panels that were enriched for protein-altering DNV in our study more often fulfil criteria for disease-association (brain expression, functional constraint) and may therefore be more likely truly disease associated than genes not DNV-enriched. Of 49 panel genes that were not DNV-enriched and failed one or two disease criteria, 24

had limited, conflicting or no published evidence for disease association (ClinGen criteria[30], details Supplementary Table 8, Figure 4B). We hope that our data may provide grounds for replacing genes backed by limited evidence with higher evidence genes in panels focussing on NDD with epilepsy, which could significantly improve the diagnostic yield of targeted sequencing approaches.

Finally, we assessed to what extent the results of our study might influence therapeutic decisions by initiating individualized treatments or avoiding inappropriate therapy with low degree of benefit and/or high risk of side effects. Therapeutic approaches, tailored to the patient's underlying genetic defect, have successfully been applied for several EE[6] (e.g. treatment with ezogabine in *KCNQ2* encephalopathy[42] or ketogenic diet in *SLC2A1*-related disorders[43]).

Within the 33 exome-wide significant genes of our meta-analysis, 79 out of 289 protein-altering DNV (27.3%) were found in genes for which precision medicine approaches have been described with a Centre for Evidence-Based Medicine level of evidence[44] of IIb or higher (Supplementary Table 15). This corresponds to 4.2% of all DNV in NDD with epilepsy. However, the extent of pathogenic or likely pathogenic variants among those 79 DNV remains unclear, as we did not investigate pathogenicity of individual protein-altering DNV according to ACMG criteria[45]. We expect that with further gain of knowledge on underlying pathomechanisms, diagnoses with relevant therapeutic consequences and benefit in patient care will keep increasing.

Methods

Patient cohorts

In this study, we included 7088 patients with the following intellectual and/or developmental disorders: developmental delay (DD), intellectual disability (ID), epileptic encephalopathy (EE) and autism spectrum disorder (ASD), which we summarize as neurodevelopmental disorders (NDD). In every cohort, we categorized patients for presence or absence of seizures or epilepsy while we did not classify for specific epilepsy syndromes. We obtained this information from the respective patients' HPO categories, phenotype tables or phenotypic descriptions. Epilepsy is defined by more than one unprovoked seizure occurring greater than 24 hours apart, increased probability of further unprovoked seizures and/or diagnosis of an epilepsy syndrome (International League Against Epilepsy, www.ilea.org). We assume that seizures were due to epilepsy and not reversible causes, such as high fever or hypoglycaemia. Further, diagnosis of epilepsy may not have been documented for all patients. Also, some patients with DNV in EE genes may develop epilepsy after inclusion in a study (Supplementary Figure S3, Supplementary Note). Finally, we do not distinguish brain malformations as a secondary cause of epilepsy in our cohort. For our main analyses, we stratified this combined cohort of patients with NDD for patients comorbid with epilepsy (n = 2151), by a primary diagnosis of EE[8,18,20-22], or by presence of epilepsy/seizures[11,12,16,17,19]. Patients in the latter group thereby expressed the two key features of EE, namely NDD and epilepsy[1,2], while having a different primary diagnosis than EE. The largest cohort contained 863 patients with a primary diagnosis of severe DD, of which approximately 87% had ID[4]. Across 7 cohorts, 849 patients were primarily diagnosed with EE and 267 patients from 5 cohorts with ID. We also included one cohort of patients with ASD and epilepsy[12] who were phenotyped within the Simon Simplex Consortium[25]. Among individuals with ASD, we only included patients with low IQ as it has been shown that DNV in normal IQ ASD patients play only a minor, if any role, in disease etiology[10,26]. In total, 180 patients had ASD, epilepsy and "low" (IQ < 80) or "very low" IQ (IQ < 70, n = 124) as defined by Differential Ability Scales, the most frequently

applied IQ test in 77% (138/180) of these patients. Two EE cohorts and one ID cohort comprising a combined 226 patients were not previously published; two cohorts were only partly published[21,22].

We also conducted DNV enrichment analyses in patients with NDD without epilepsy (n = 4937). Here, we included patients from the three largest NDD cohorts with primary diagnosis DD[4](n=3430), low-IQ ASD[12,25] (n = 847) and ID[16] (n = 660). Previously sequenced trios (n = 1911)[12], from unaffected siblings of a child with autism spectrum disorder, served as control trios.

In all cohorts, both patients and their unaffected parents underwent whole exome sequencing (WES). Variants that were not present in either parent were considered *de novo* variants (DNV). The study was approved by the ethics committee of the University of Leipzig (224/16-ek, 402/16-ek) and additional local ethics committees. A list of all published and unpublished cohorts used in this paper can be found in Supplementary Table 1.

Sequencing pipelines of previously unpublished or partly published cohorts (cohorts 10 - 14)

Libraries were prepared from parents' and patients' DNA, exome captured and sequenced on Illumina sequencers. Raw data was processed and technically filtered with established pipelines at the respective academic or diagnostic laboratories. DNV data from all cohorts was re-annotated for this study (see below). Specific pipelines of cohorts 10 to 14 are described below.

Cohort 10 (Ambry Genetics):

Diagnostic whole exome sequencing was performed on parent-offspring trios at Ambry Genetics (Aliso Viejo, CA) in 228 patients with a history of seizures who have been previously described[21]. Genomic DNA extraction, exome library preparation, sequencing, bioinformatics pipeline, and data analyses were performed as previously described[46]. Briefly, samples were prepared and sequenced using paired-end, 100 cycle chemistry on the Illumina HiSeq 2500 sequencer. Exome enrichment was performed using either the SureSelect Target Enrichment System 3.0 (Agilent Technologies) or SeqCap EZ VCRome 2.0 (Roche NimblGen). The sequencing reads were aligned to human reference genome (GRCh37) and variants were called by using CASAVA software (Illumina). The following variants filters are applied to generate a list of high confident de novo variant calls: 1) at least one member of trio with mutation base coverage >40x and all trio member with at least 20x base coverage; 2) heterozygous read ratio in proband >30%; 3) heterozygous read ratio in parent <10%; and 4) exclusion of known sequencing artefacts (based on Ambry Genetics' internal databases).

Cohort 11 (Boston children's hospital).

The sequencing pipeline has been described previously[22]. In short, DNA extracted from blood or saliva underwent whole exome sequencing capture using either the Agilent SureSelect XTHuman All Exon v4 or Illumina Rapid Capture Exome enrichment kit. Sequencing of 100bp paired-end reads was obtained using Illumina HiSeq (Illumina, SanDiego, CA). Coverage was >90% or >80% meeting 20x coverage with the two methods respectively. The pipeline for data analysis in both cases included alignment and quality score recalibration with Burrows-Wheeler Aligner (BWA, bio-bwa.sourceforge.net/) and variant calling with GATK (www.broadinstitute.org/gatk/). De novo variant calling sites lacking parental coverage as well as variants present in the parents were excluded regardless of read frequency in the parents.

Cohorts 12 (EuroEPINOMICS RES) and 13 (DFG atypical EE):

Exonic and adjacent intronic sequences were enriched from genomic DNA using the NimbleGen SeqCap EZ Human Exome Library v2.0 enrichment kit. WES was performed

using a 100bp paired-end read protocol due to the manufacturer's recommendations on an Illumina HiSeq2000 sequencer by the Cologne Center for Genomics (CCG), Cologne, Germany. Reads were mapped on the human hg19 reference genome (bwa-aln software, bio-bwa.sourceforge.net/). The UnifiedGenotyper (GATK, www.broadinstitute.org/gatk/) and Mpileup (Samtools, http://samtools.sourceforge.net/) software were used to call variants. The paired sample feature from the DeNovoGear software was further used to examine potential de novo mutations in twin pairs. Data analysis and filtering of mapped target sequences was performed with the 'Varbank' exome and genome analysis pipeline v.2.1 (unpublished; https://varbank.ccg.uni-koeln.de). In particular, we filtered for high-quality (coverage of more than six reads, fraction of allele carrying reads at least 25%, a minimum genotype quality score of 10, VQSLOD greater than -8) and rare (Caucasian population allele frequency < 0.5%) variations on targeted regions + flanking 100bp. In order to exclude pipeline specific artifacts, we also filtered against an in-house cohort (AF < 2%) of variations, which were created with the same analysis pipeline. The filter conditions were set to be more sensitive following manual inspections of aligned reads. In particular, we looked for typical patterns of false positive variations (e.g., more than two haplotypes, base quality or mapping quality bias, strand bias, allele read position bias, low complexity region, alignment errors).

Cohort 14 (University of Leipzig):
Exome capture was carried out with Illumina's Nextera Rapid Capture Exome Kit (Illumina, Inc., San Diego, CA, USA). DNA Sequencing was done with NextSeq500 or HiSeq4000 sequencers (Illumina, Inc.) to 2 × 150bp reads at the Centogene AG, Rostock, Germany. Raw sequencing reads were converted to standard fastq format using bcl2fastq software 2.17.1.14 (Illumina, Inc.), and fed to a pipeline at Centogene AG based on the 1000 Genomes Project (1000G) data analysis pipeline and GATK best practice recommendations. Sequencing reads were aligned to the GRCh37 (hg19) build of the human reference genome using bwa-mem (bio-bwa.sourceforge.net/). In addition to GATK HaplotypeCaller (www.broadinstitute.org/gatk/)), variant calling was performed with freebayes (https://github.com/ekg/freebayes) and samtools (http://samtools.sourceforge.net/). Quality filtering of sequencing reads in both parents and children was done according to the following criteria: read depth > 20, quality > 500, frequency of alternative allele between 30 and 70% for the child and not present in the parents, frequency < 1% in internal database, variant called by at least two different genotype callers.

False positive rates of DNV
In subcohorts 1 to 6, all DNV were validated by Sanger sequencing to eliminate false positive calls. In cohorts 7 and 8 not all variants have been confirmed by Sanger sequencing. Through random selection of variants for validation, the false positive rate was estimated to be approximately 1.4% and < 5 %, respectively. In the clinical cohorts 10 to 14, variants defined as variants worth reporting back to patients (variants of unknown significance or [likely] pathogenic) are normally validated by Sanger sequencing. With this experience, false discovery rates in these cohorts were estimated to be < 5% (personal communications).

Annotation and Filtering
DNV files were generated and quality-filtered by the individual groups. All DNV were re-annotated with the following pipeline. Variants were annotated with Ensembl's Variant Effect Predictor (http://grch37.ensembl.org/Homo_sapiens/Tools/VEP) of version 82 using database 83 of GRCh37 as reference genome. Per variant, the transcript with the most severe impact, as predicted by VEP, was selected for further analyses. The decreasing order of variant impacts was HIGH, MODERATE, MODIFIER, LOW. Only

protein – altering (missense or protein truncating DNV [premature stop codon, essential splice site, frameshift]) were included in further analyses. Variants that were present in ExAC[23], an aggregation of 60,706 exome sequences from adult individuals without severe childhood-onset diseases, were excluded, as these have been shown to convey no detectable risk to NDD[24]. Per cohort, there was a mean 0.9 ± 0.06 protein – altering DNV per individual.

Statistical analysis

All statistical analyses were done with the R programming language (www.r-project.org). Fisher's Exact Test for Count Data, Wilcoxon rank sum test, Cochran-Mantel-Haenszel test, generalized linear model, Spearman correlation, Welch two-sided t-test and calculation of empirical p-values were performed as referenced in the results. For datasets assumed to be normally distributed after visual inspection, mean and standard deviation (sd) are written as mean ± sd. The R code used to perform the statistical analyses and figures is available upon request.

DNV enrichment analyses

To identify genes with a significant mutational burden, we compared the number of observed with the number of expected number protein-altering DNV for each gene using an established framework of gene-specific mutation rates[10]. The analysis was done with the R package denovolyzer[47]. We corrected the obtained p-values for the number of genes for which gene specific mutation rates[10] were available (n = 18225). Genes that were significant with a Bonferroni corrected p-value < 0.05 were termed "exome-wide significant" and considered as disease associated. All genes that had at least two protein altering DNV and that were nominally significant with a p-value < 0.05 were termed "nominally significant". DNV enrichment analyses were carried out in the cohort of all patients with NDD (n = 7088) as well as in two sub cohorts stratified for patients with epilepsy (n = 2151) or without epilepsy (n = 4937).

Functional enrichment analyses

Significantly enriched pathways were computed with the R package of g:Profiler[33], using ordered enrichment analysis on significance-ranked proteins. Different gene sets were queried using a background gene set of all 18225 genes annotated in[10]. Human phenotype ontology, biological processes from Gene Ontology and pathways from the KEGG and Reactome databases were evaluated while other functional annotations were ignored. To correct for multiple testing the built-in Bonferroni correction was used. Only terms that were statistically significant with a Bonferroni corrected p-value < 0.01 were presented, as our negative controls (i: nominally DNV-enriched genes tolerant to loss-of-function and missense variation or not expressed in the infant brain and not annotated in a protein network with exome-wide significant genes and ii: genes nominally DNV-enriched in healthy control) were not enriched for any functional categories below this p-value.

Acquisition and processing of brain gene expression data

We downloaded the Developmental Transcriptome dataset of 'BrainSpan: Atlas of the Developing Human Brain' (www.brainspan.org, funded by ARRA Awards 1RC2MH089921-01, 1RC2MH090047-01, and 1RC2MH089929-01, 2011). Details about tissue acquisition and sequencing methodology can be found in the BrainSpan website's documentation. The atlas includes RNA sequencing data generated from tissue samples of developing post-mortem brains of neurologically unremarkable donors covering 8 to 16 brain structures. We extracted brain expression data from the 5 donors that were infants aged 0 to 12 months. 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 (log2 + 1)-transformed RPKM values. We defined infant brain gene expression as median (log2 + 1)-transformed RPKM value > 1.

Evaluation of genes' intolerance to protein altering variants

We assessed individual gene tolerance to truncating or missense variants in the general population by using the pLI score (probability of being LoF intolerant) and missense z-score (z-score of observed versus expected missense variants), respectively[23]. We used as gene constrained cut-offs >0.9 for pLI and >3.09 for missense-z scores based on recommendations of the score developers (as suggested on www.exac.broadinstitute.org).

We calculated empirical p-values to evaluate if pLI scores of exome-wide and nominally DNV enriched genes were significantly higher compared to pLI scores of random gene sets as described in[16]. Briefly, we computed the expected pLI for a given gene set with size n by randomly drawing gene sets with size n from the total 18,225 pLI annotated genes calculating their respective median pLI. We repeated this random sampling 1,000,000 times. 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.

Protein network

Protein-protein interaction (PPI) network analysis is a powerful method to detect the functional relationship between disease-related genes and can be used to map genetic variation to molecular processes[48]. We sought to identify functional (PPI) modules enriched for DNV in NDD with epilepsy. To increase the likelihood of the candidate genes to be truly associated with the exome-wide significant genes for NDD with epilepsy we used all the 188 genes that were at least nominally significantly enriched for DNV in NDD with epilepsy to search for functionally enriched modules. We analysed the 188 genes with GeNets (http://apps.broadinstitute.org/genets, time of analysis: 02/2017), a biological pathway analysis and visualization platform. We used protein annotations by InWeb_InBioMap[31], a computational framework to leverage curated PPI of 87% of reviewed human UniProt IDs. In total, 87 genes were part of a PPI network with an overall connectivity p-value of 0.05. Within the identified PPI network, we defined functional modules called gene "communities" using a method presented in [32]. Ten "communities" were identified each spanning 3 to 21 genes (Supplementary Figure S5), each with a connectivity p-value < $2x10^{-3}$. 7 of the 10 communities included an exome-wide significant gene. Consequently, we prioritized the nominally DNV enriched genes in these 7 communities as candidate genes for NDD with epilepsy, if they also fulfilled other criteria for disease association like functional constraint and brain expression.

Diagnostic gene panels for epileptic encephalopathy/ comprehensive epilepsy from 24 academic/ 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[34] (GTR) of NCBI (date: 01/2017) for providers of tests for "Epileptic encephalopathy, childhood-onset" and identified 16 diagnostic epilepsy panels. We excluded 3 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 at the respective providers' websites.

We calculated to what proportion the 33 genes that were exome-wide significantly enriched for protein-altering DNV were covered in 24 commercial or

academic providers of gene panels for epileptic encephalopathy/comprehensive epilepsy. For each gene, we counted the number of times it was included in any of the 24 panels and divided that by the total number of 24 gene panels.

We investigated if there were genes in the 24 diagnostic gene panels without evidence for implication in NDD with epilepsy based on our analyses. We focused on dominant or x-linked genes (listed in Supplementary Table 12). We tested these genes for three criteria of association with NDD with epilepsy: Firstly, if genes had at least nominal significance for enrichment in protein altering DNV in this meta-analysis; secondly, whether genes were expressed in the infant brain defined by a median RPKM of all samples and brain regions > 1; thirdly, whether genes had a pLI > 0.9 or missense z-score > 3.09[23] indicating intolerance to truncating or missense variants. We intersected these lists to nominate genes that may not be truly associated with NDD with epilepsy.

Data availability

The authors declare that 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 of cohort EuroEPINOMICS RES have been deposited in the European Genome-phenome Archive (EGA) with the accession code EGAS00001000048 (https://www.ebi.ac.uk/ega/datasets/EGAD00001000021). Raw sequencing data of cohort 13 (DFG atypical EE) will be deposited in a public repository after finalizing of the individual project.

Figures

Figure 1. Schematic Venn diagram of neurodevelopmental disorders (NDD) and seizures. We define NDD as an umbrella term for the following disorders:

- ASD (= autism spectrum disorder, red circle). In this study, we only included individuals with low-IQ ASD.
- ID (= intellectual disability, yellow circle), used synonymous to DD (developmental delay), as approximately 87% of individuals with DD have ID[4]
- EE (= epileptic encephalopathy, brown circle)

Approximately 20% of individuals with NDD have seizures. Per definition, all individuals with EE have seizures. Thus, individuals with EE are a subcohort of "NDD with epilepsy" while not all individuals with "NDD with epilepsy" have been diagnosed with EE.

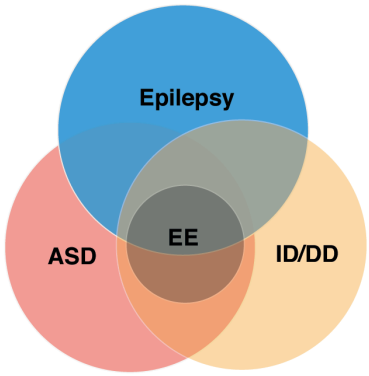
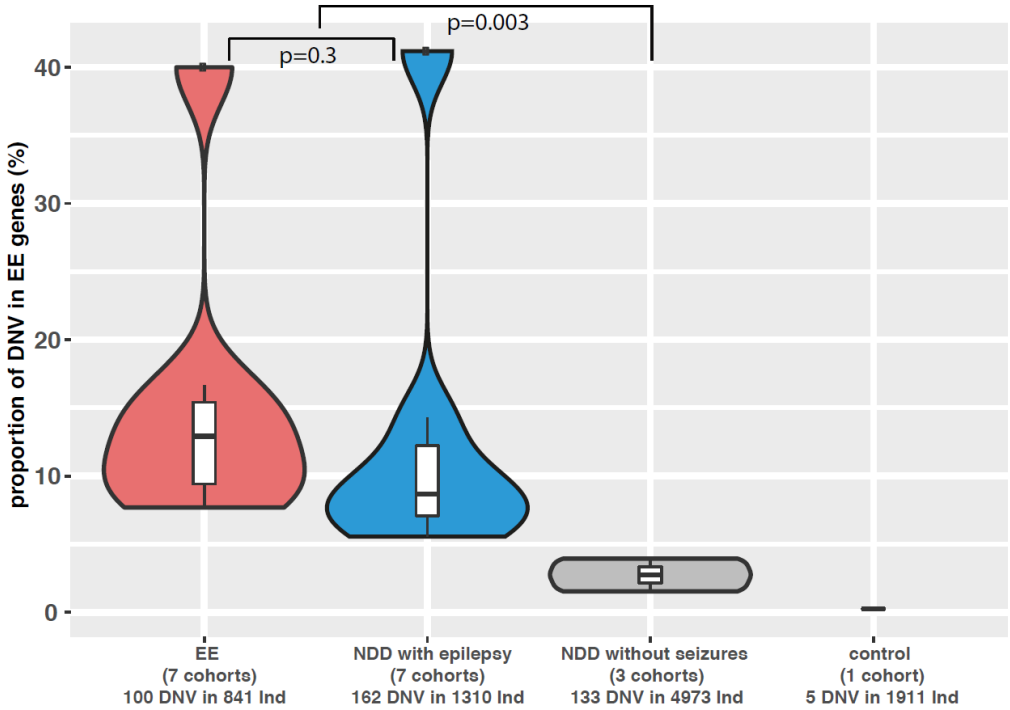


Figure 2. Protein altering DNV in EE genes in different patient cohorts, normalized for cohort size. **A**, The proportion of DNV in EE genes is similar in patients diagnosed with EE (red) and NDD (DD/low IQ- ASD/ID) with epilepsy (blue) vs NDD without epilepsy (grey). Cohort size is given as number of individuals (Ind). Two-sided Wilcoxon rank sum tests compared the frequency of mutated EE genes between patients with EE and NDD without epilepsy as well as patients with NDD with (EE & NDD combined) to without epilepsy. **B**, The proportion of DNV in EE genes is higher in patients with epilepsy versus without epilepsy within three different NDD cohorts (DD, low IQ- ASD, ID). P-values are plotted next to respective odds ratios (red dots), while 95%-confidence intervals are shown in yellow (Fisher's exact test for individual cohorts, Cochran-Mantel-Haenszel test for combined cohorts).

A,



B,

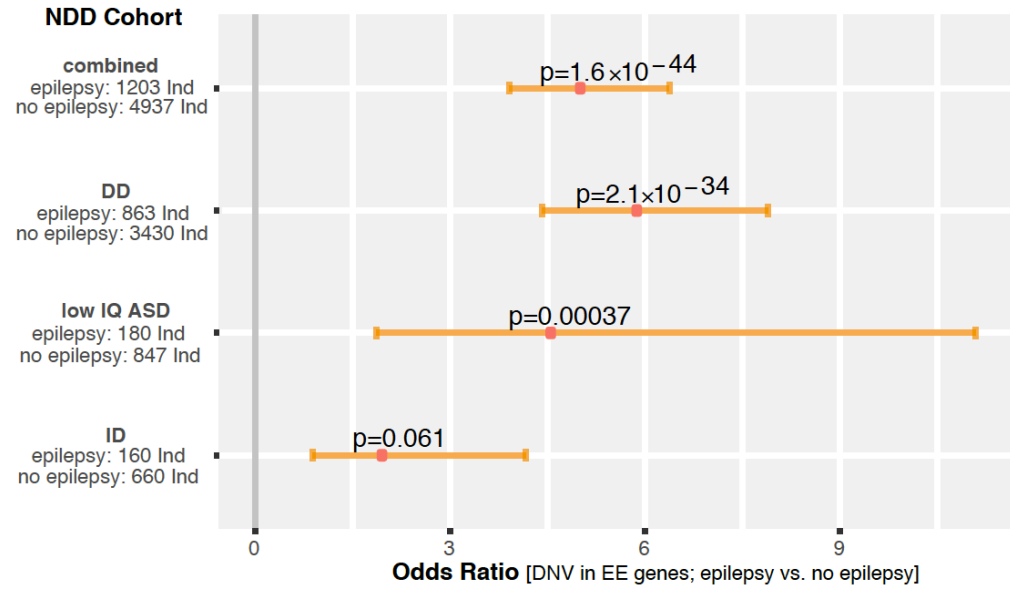


Figure 3. Forrest plot of 33 exome-wide significantly DNV enriched genes in NDD patients with and without epilepsy. Per gene, frequency of DNV in NDD with epilepsy vs NDD without epilepsy was compared using Fisher's Exact Test. Odds ratios per gene for enrichment of DNV in epilepsy vs no epilepsy were plotted on the x-axis as dots while their 95% confidence interval was represented by bars and labeled by respective p-values (-log10 transformed). For clarity, error bars were cut at an odds ratio of 26. Triangular dots and blue bars indicate multiple testing resistant significance (Bonferroni correction for n = 33 genes). In total, missense and truncating DNV of 7088 patients with different NDD diagnoses DD, low IQ- ASD, ID or EE with epilepsy (n = 2151) and without epilepsy (n = 4937) went into the analysis.

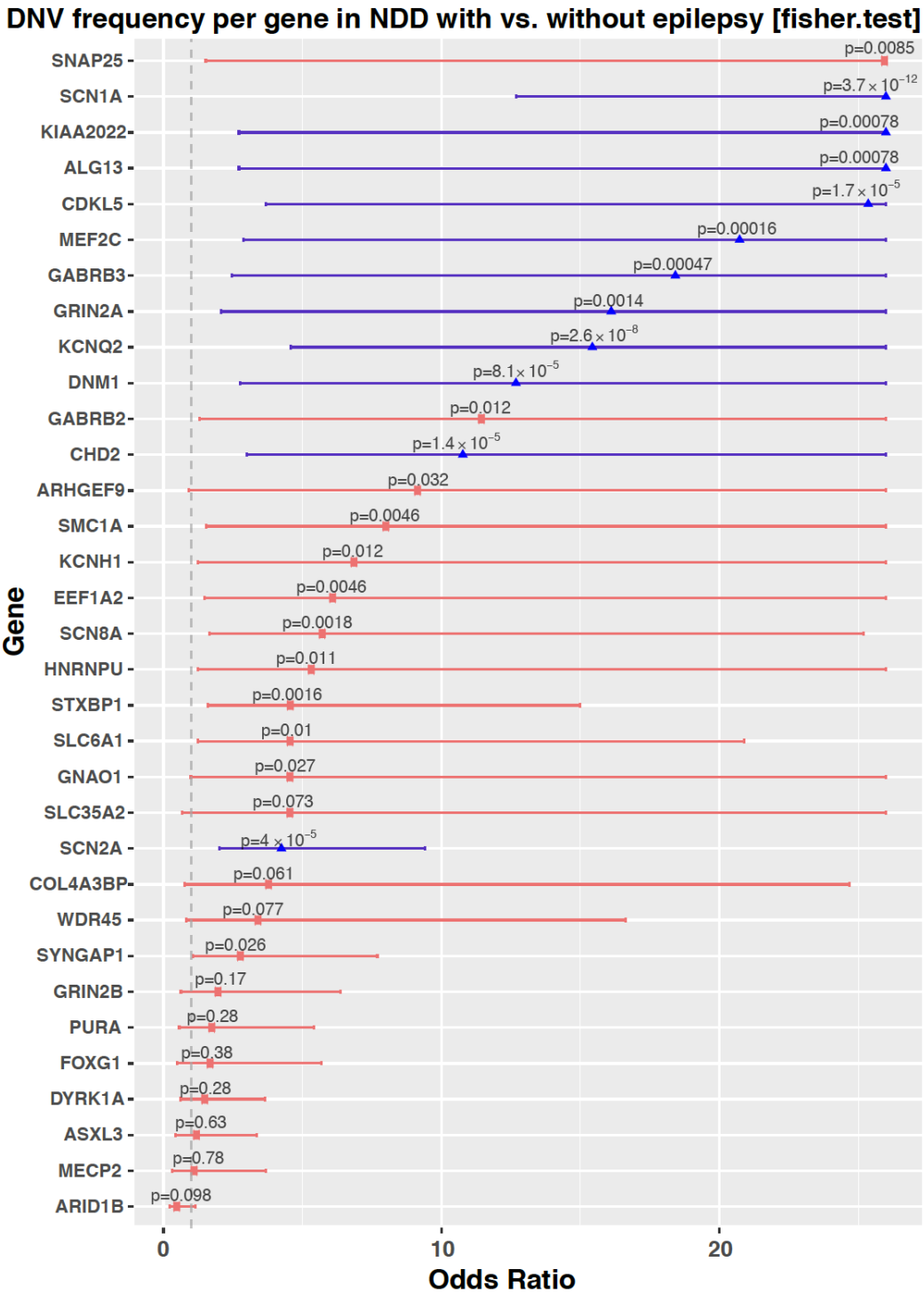


Figure 4. Genes exome-wide (red) or nominally (orange) significantly DNV enriched in NDD with epilepsy are compared to genes nominally significant in controls (yellow). **A**, Exome-wide and nominally DNV-enriched genes are higher expressed in the infant brain compared to genes nominally DNV-enriched in controls (two-sided t-tests, t = 4.2, df = 60 and t = 3.3, df = 70, respectively). **B, C,** Exome-wide and nominally DNV enriched genes have higher constraint scores than genes in the control group. Missense z-sore: two-sided t-test (t = 5.6, df = 64 and t = 4.5, df = 82, respectively). pLI score: empirical p-value compared to 1M same size random gene sets of all available pLI scores.

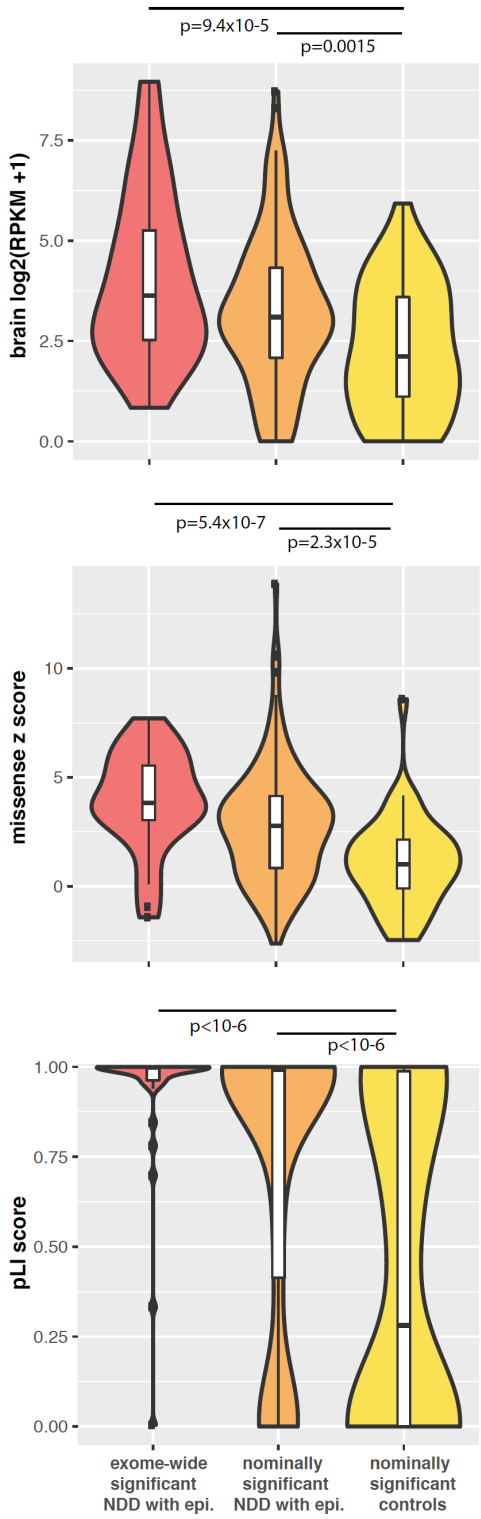
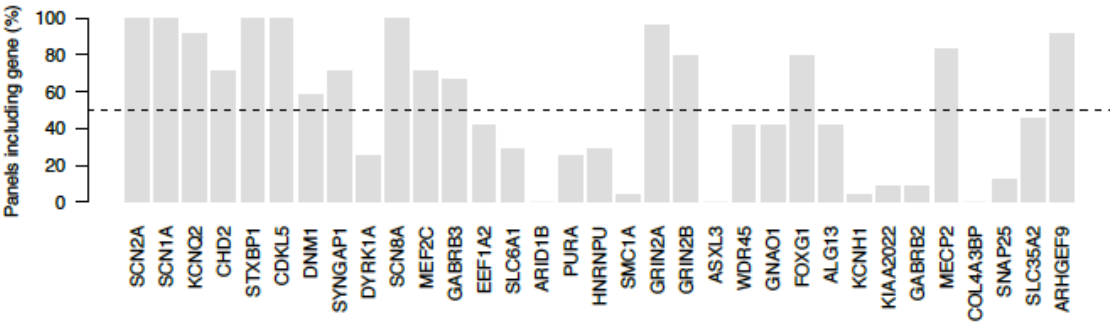


Figure 5. Comparing the findings of our meta-analysis to 24 commercial and academic providers of diagnostic gene panels for "epileptic encephalopathy" or "comprehensive epilepsy".

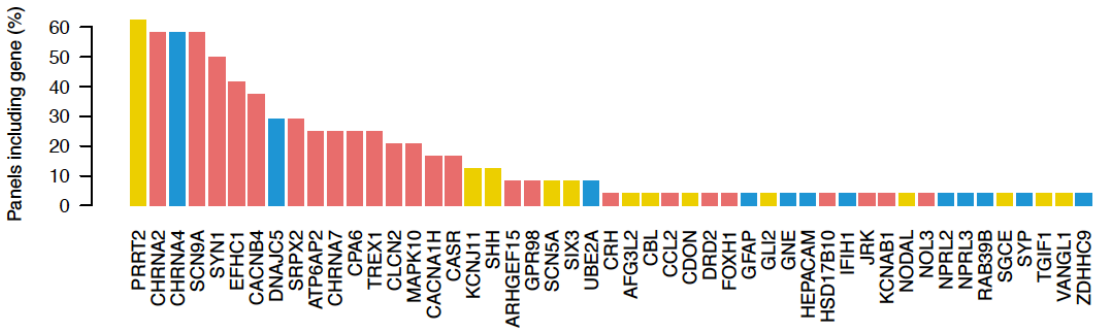
A, This bar plot shows to what proportion the 33 genes DNV enriched in NDD with epilepsy were covered by 24 diagnostic gene panels. *CDKL5*, *SCN1A*, *SCN2A*, *SCN8A*, *SPTAN1* and *STXBP1* were contained in all 24 providers, while no DNV in *SPTAN1* was found in our cohort. Genes are sorted by number of protein-altering DNV in our study.

B, Forty-nine genes on diagnostic panels failed two of the three criteria of association with NDD with epilepsy (at least nominally enriched for DNV, infant brain gene expression, intolerance to loss of function or missense variation). Twenty-four of these 49 genes also had no, limited, or conflicting evidence for disease association as defined by ClinGen[30] (red bars). Genes with moderate to definitive evidence for association with other phenotypes than NDD with epilepsy are shown as yellow bars, genes with at least moderate evidence for association with NDD with epilepsy as blue bars.

A,



B,



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**, 304-16 (2016).
2. 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**, 676-85 (2010).
3. McGrother, C.W. *et al.* Epilepsy in adults with intellectual disabilities: prevalence, associations and service implications. *Seizure* **15**, 376-86 (2006).
4. Deciphering Developmental Disorders, S. Prevalence and architecture of de novo mutations in developmental disorders. *Nature* **542**, 433-438 (2017).
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. Myers, C.T. & Mefford, H.C. Advancing epilepsy genetics in the genomic era. *Genome Med* **7**, 91 (2015).
7. De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209-15 (2014).
8. EuroEpinomics & Epi4K. De novo mutations in synaptic transmission genes including DNM1 cause epileptic encephalopathies. *Am J Hum Genet* **95**, 360-70 (2014).
9. Krumm, N., O'Roak, B.J., Shendure, J. & Eichler, E.E. A de novo convergence of autism genetics and molecular neuroscience. *Trends Neurosci* **37**, 95-105 (2014).
10. Samocha, K.E. *et al.* A framework for the interpretation of de novo mutation in human disease. *Nat Genet* **46**, 944-50 (2014).
11. de Ligt, J. *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med* **367**, 1921-9 (2012).
12. Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216-21 (2014).
13. Zaidi, S. *et al.* De novo mutations in histone-modifying genes in congenital heart disease. *Nature* **498**, 220-3 (2013).
14. Zhu, X., Need, A.C., Petrovski, S. & Goldstein, D.B. One gene, many neuropsychiatric disorders: lessons from Mendelian diseases. *Nat Neurosci* **17**, 773-81 (2014).
15. Li, J. *et al.* Genes with de novo mutations are shared by four neuropsychiatric disorders discovered from NPdenovo database. *Mol Psychiatry* **21**, 290-7 (2016).
16. Lelieveld, S.H. *et al.* Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nat Neurosci* **19**, 1194-6 (2016).
17. Hamdan, F.F. *et al.* De novo mutations in moderate or severe intellectual disability. *PLoS Genet* **10**, e1004772 (2014).
18. Michaud, J.L. *et al.* The genetic landscape of infantile spasms. *Hum Mol Genet* **23**, 4846-58 (2014).

19. Rauch, A. *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* **380**, 1674-82 (2012).
20. Veeramah, K.R. *et al.* Exome sequencing reveals new causal mutations in children with epileptic encephalopathies. *Epilepsia* **54**, 1270-81 (2013).
21. 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).
22. Olson, H.E. *et al.* Genetics and genotype-phenotype correlations in early onset epileptic encephalopathy with burst suppression. *Ann Neurol***81**, 419-429 (2017).
23. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-91 (2016).
24. Kosmicki, J.A. *et al.* Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. *Nat Genet* (2017).
25. Fischbach, G.D. & Lord, C. The Simons Simplex Collection: a resource for identification of autism genetic risk factors. *Neuron* **68**, 192-5 (2010).
26. Robinson, E.B. *et al.* Autism spectrum disorder severity reflects the average contribution of de novo and familial influences. *Proc Natl Acad Sci U S A* **111**, 15161-5 (2014).
27. Lemke, J.R. *et al.* Targeted next generation sequencing as a diagnostic tool in epileptic disorders. *Epilepsia* **53**, 1387-98 (2012).
28. Rohena, L. *et al.* Mutation in SNAP25 as a novel genetic cause of epilepsy and intellectual disability. *Rare Dis* **1**, e26314 (2013).
29. Landrum, M.J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* **44**, D862-8 (2016).
30. Rehm, H.L. *et al.* ClinGen--the Clinical Genome Resource, http://www.clinicalgenome.org/site/assets/files/2657/current_clinical_validity_classifications.pdf. *N Engl J Med* **372**, 2235-42 (2015).
31. Li, T. *et al.* A scored human protein-protein interaction network to catalyze genomic interpretation. *Nat Methods* **14**, 61-64 (2017).
32. Clauset, A., Newman, M.E. & Moore, C. Finding community structure in very large networks. *Phys Rev E Stat Nonlin Soft Matter Phys* **70**, 066111 (2004).
33. Reimand, J. *et al.* g:Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res* **44**, W83-9 (2016).
34. 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**, D925-35 (2013).
35. Brunklaus, A., Ellis, R., Reavey, E., Forbes, G.H. & Zuberi, S.M. Prognostic, clinical and demographic features in SCN1A mutation-positive Dravet syndrome. *Brain* **135**, 2329-36 (2012).
36. Claes, L. *et al.* De novo mutations in the sodium-channel gene SCN1A cause severe myoclonic epilepsy of infancy. *Am J Hum Genet* **68**, 1327-32 (2001).
37. Weckhuysen, S. *et al.* KCNQ2 encephalopathy: emerging phenotype of a neonatal epileptic encephalopathy. *Ann Neurol* **71**, 15-25 (2012).

38. Shi, X. *et al.* Clinical spectrum of SCN2A mutations. *Brain Dev* **34**, 541-5 (2012).
39. Stamberger, H. *et al.* STXBP1 encephalopathy: A neurodevelopmental disorder including epilepsy. *Neurology* **86**, 954-62 (2016).
40. Scheffer, I.E. *et al.* ILAE classification of the epilepsies: Position paper of the ILAE Commission for Classification and Terminology. *Epilepsia* (2017).
41. MacArthur, D.G. *et al.* Guidelines for investigating causality of sequence variants in human disease. *Nature* **508**, 469-76 (2014).
42. Millichap, J.J. *et al.* KCNQ2 encephalopathy: Features, mutational hot spots, and ezogabine treatment of 11 patients. *Neurol Genet* **2**, e96 (2016).
43. De Giorgis, V. & Veggiotti, P. GLUT1 deficiency syndrome 2013: current state of the art. *Seizure* **22**, 803-11 (2013).
44. OCEBM Levels of Evidence Working Group* (Jeremy Howick, I.C.J.L.L., Paul Glasziou, Trish Greenhalgh, Carl Heneghan, Alessandro Liberati, Ivan Moschetti, Bob Phillips, Hazel Thornton, Olive Goddard and Mary Hodgkinson). The Oxford Levels of Evidence 2. in *Oxford Centre for Evidence-Based Medicine.* (2011).
45. 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**, 405-24 (2015).
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**, 578-86 (2015).
47. 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 1-15 (2015).
48. Rossin, E.J. *et al.* Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet* **7**, e1001273 (2011).