Genetics inMedicine | ARTICLE

Targeted gene sequencing in 6994 individuals with neurodevelopmental disorder with epilepsy

Henrike O. Heyne, MD[1,2,3,4], Mykyta Artomov, PhD[1,2,3], Florian Battke, PhD[5], Claudia Bianchini, PhD[6], Douglas R. Smith, PhD[7], Nora Liebmann[4], Vasisht Tadigotla, PhD[7], Christine M. Stanley, PhD[7], Dennis Lal, PhD[1,2,3,8], Heidi Rehm, PhD[9], Holger Lerche, MD[10], Mark J. Daly, PhD[1,2,3], Ingo Helbig, MD[11], Saskia Biskup, MD PhD[5], Yvonne G. Weber, MD[10] and Johannes R. Lemke, MD[4]

Purpose: We aimed to gain insight into frequencies of genetic variants in genes implicated in neurodevelopmental disorder with epilepsy (NDD+E) by investigating large cohorts of patients in a diagnostic setting.

Methods: We analyzed variants in NDD+E using epilepsy gene panel sequencing performed between 2013 and 2017 by two large diagnostic companies. We compared variant frequencies in 6994 panels with another 8588 recently published panels as well as exome-wide de novo variants in 1942 individuals with NDD+E and 10,937 controls.

Results: Genes with highest frequencies of ultrarare variants in NDD+E comprised *SCN1A*, *KCNQ2*, *SCN2A*, *CDKL5*, *SCN8A*, and *STXBP1*, concordant with the two other epilepsy cohorts we investigated. In only 46% of the analyzed 262 dominant and X-linked panel genes ultrarare variants in patients were reported. Among genes with contradictory evidence of association with epilepsy, *CACNB4*, *CLCN2*, *EFHC1*, *GABRD*, *MAGI2*, and *SRPX2* showed equal frequencies in cases and controls.

Conclusion: We show that improvement of panel design increased diagnostic yield over time, but panels still display genes with low or no diagnostic yield. With our data, we hope to improve current diagnostic NDD+E panel design and provide a resource of ultrarare variants in individuals with NDD+E to the community.

Genetics in Medicine (2019) 21:2496–2503; https://doi.org/10.1038/s41436-019-0531-0

Keywords: epilepsy; gene panel design; Mendelian genetics; clinical genetics, neurodevelopmental disorder

INTRODUCTION

In recent years, genetic research has gained novel biological insights into the etiology of neurological disorders, particularly in epilepsy.[1,2] Neurodevelopmental disorders with epilepsy (NDD+E) are a rare group of disorders frequently caused by de novo events in protein-coding genes.[3,4] Precise genetic diagnosis influences genetic counseling but may also guide treatment decisions by enabling medication or treatment tailored to the patient's underlying genetic defect.[2,4] Examples include treatment with sodium channel blockers in *SCN2A*- and *SCN8A*-related NDD+E,[5,6] ezogabine in *KCNQ2*-related NDD+E,[7] or a ketogenic diet in *SLC2A1*-related GLUT1 deficiency.[8] Up to 28% of de novo variants (DNVs) being found in NDD+E-related genes are associated with such targeted treatment approaches.[4] However, assessments of how often NDD+E-associated genes are mutated are currently insufficient due to lack of large-scale genetic analyses in NDD+E.

Targeted sequencing of specific disease-related gene panels has been part of the diagnostic workup of highly prevalent heterogeneous disorders such as breast cancer,[9] cardiomyopathy,[10] and epilepsy.[11–13] Multiple genes are sequenced in parallel enabling lower sequencing cost, higher coverage, and near-absence of secondary findings compared with exome sequencing.[14] However, high heterogeneity of epilepsy gene panel content has been observed.[4,15] This is likely due to

[1]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; [2]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA; [3]Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA; [4]Institute of Human Genetics, University of Leipzig Medical Center, Leipzig, Germany; [5]Praxis für Humangenetik and CeGaT GmbH, Tuebingen, Germany; [6]Pediatric Neurology, Neurogenetics and Neurobiology, Unit and Laboratories, Neuroscience Department, A Meyer Children's Hospital, University of Florence, Florence, Italy; [7]Courtagen Life Sciences Inc., Woburn, MA, USA; [8]Cleveland Clinic Genomic Medicine Institute and Neurological Institute, Cleveland, OH, USA; [9]Department of Medicine, Massachusetts General Hospital, Boston, MA, USA; [10]Department of Neurology and Epileptology, Hertie Institute for Clinical Brain Research, and Department of Neurosurgery, University of Tuebingen, Tuebingen, Germany; [11]Division of Neurology, Children's Hospital of Philadelphia, Philadelphia, PA, USA. Correspondence: Henrike O. Heyne (hheyne@broadinstitute.org) or Johannes R. Lemke (johannes.lemke@medizin.uni-leipzig.de)

Submitted 2 January 2019; accepted: 19 April 2019
Published online: 6 May 2019

Table 1 Cohort description

Cohort	Phenotype	PMID	*n*	Type of variants	Number of genes analyzed
CeGaT	NDD+E	Unpublished	3177	Ultrarare, DNV	645
Courtagen	NDD+E	Unpublished	3817	Ultrarare, DNV	645
Lindy et al.[17] (GeneDx)	NDD+E	29655203	8565	Likely pathogenic (ACMG)	70
Heyne et al.[4]	NDD+E	29942082	1942	DNV	18,228
Exomes from different cohorts; see "Materials and methods"	Controls	See "Materials and methods"	10,937	Ultrarare	645

ACMG variant interpretation guidelines by the American College of Medical Genetics and Genomics, *DNV* de novo variants, *NDD+E* neurodevelopmental disorders with epilepsy, *PMID* PubMed ID.

the dramatically growing number of genes associated with epilepsy and diverse integration in the established panels, often without robust statistical evidence.[4,16] To increase yield in diagnostic sequencing panels, it is essential to consider genes with proven disease association as well as a reasonable frequency of pathogenic variants among affected individuals.

Here, we report likely damaging variants in 645 epilepsy panel genes sequenced at two molecular diagnostic companies, CeGaT (Germany) and Courtagen (USA). In total, 6994 patients with NDD+E of suspected monogenic cause underwent diagnostic sequencing at the respective companies, the majority as first-tier diagnostic test. We compare this large cohort of panels in NDD+E patients with another study of similar design ($n = 8565$) (ref. [17]), 10,937 controls, as well as with a cohort of exome-wide DNV in NDD+E[4] ($n = 1942$) investigating variant frequencies in confirmed and putative NDD+E genes in NDD+E panels (Table 1).

MATERIALS AND METHODS

General information

Informed consent for genetic testing was obtained from each subject or their legal guardian prior to any investigation. This study is covered by institutional review board (IRB) approvals of the ethics committee of the University of Leipzig, Germany (224/16-ek, 402/16-ek).

Gene panel sequencing data

We analyzed gene panel sequencing data of 6994 individuals diagnosed with NDD+E or related disorders of suspected monogenic origin. The data was generated during routine diagnostic sequencing by two different commercial companies, Courtagen (US, $n = 3817$ cases) and CeGaT (Germany, $n = 3177$ cases), with similar overall approach and design.[11] Information on cognitive outcome was available in about 59% of cases, revealing fractions of individuals with intellectual disability (ID) of 96% (2176/2266, Courtagen) and 97.8% (1833/1875, CeGaT). In the majority of cases, epilepsy was early-onset (before three years of age). Analysis was performed between 2013 and 2017, during which time up to ten different but vastly overlapping NDD+E panel designs were used by each company. Panels contained a median of 471 and 498 confirmed or suspected epilepsy genes at each respective company and a median 4870 individuals were sequenced per gene (Fig. 3, Supplementary Fig. S1, Supplementary Table S1). We decided to analyze the 645 genes (see Supplementary Table S2) that were sequenced in at least 2000 individuals. As the first systematic guideline for diagnostic variant interpretation was not published until 2015,[18] we decided to focus on functional (null variants as well as missense variants predicted to be deleterious by in silico tools; see "Materials and methods") ultrarare variants without pathogenicity labels that are not present in the general population.[19] In this setting, functional variants in genes not ordered by the respective clinician were not consistently reported, while we cannot access which genes were ordered by clinicians. Consequently, we identify few genes with significantly lower variant frequencies in cases than controls (Supplementary Fig. S3), suggesting underreporting of variants in these genes.

Data processing

A more detailed description of Courtagen's analysis pipeline has been published.[20] A brief overview of analysis steps is described as follows. Courtagen and CeGaT employed custom-designed Agilent Haloplex and SureSelect enrichment kits, to enrich patients' genomic DNA for target regions of epilepsy (candidate) genes. This was followed by paired-end sequencing (250 or 200 bp, respectively) on Illumina platforms (miSeq and HiSeq). Adapter sequences were then trimmed, and the sequencing reads were aligned to the human reference genome hg19 (GRCh37) with bwa-mem (bio-bwa.sourceforge.net). Reads that mapped equally well to more than one genomic position were discarded. Quality checks were performed ensuring adequate distributions of various quality control metrics such as insert size distribution, mismatch rates, GC bias, etc. Subsequent variant calling was done with different pipelines. Variants were filtered for population frequencies <1% (ExAC, EVS, 1000 Genomes) and platform-specific sequencing artifacts. Follow-up Sanger sequencing was then performed on most variants available to us.

In case of available parental samples, the de novo status of individual variants was tested by Sanger sequencing. For one of the companies, segregation testing was partially documented. Of 1173 ultrarare damaging variants, 162 (14%) were verified as de novo, 36 (3%) segregated with disease, and for 975 (83%) segregation was unknown.

Reannotation and filtering

All variants reported to patients as well as variants in controls were reannotated with the following pipeline. Variants were annotated with Ensembl's Variant Effect Predictor[21] (VEP) 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 variants (missense or null [premature stop codon, essential splice site, frameshift]) were included in further analyses. Variants that were present in a subset of ExAC (v0.3), an aggregation of exome sequences from adult individuals without severe childhood-onset diseases and without psychiatric diseases ($n = 45{,}376$) (ref. [19]), were excluded, as these have been shown to convey no detectable risk to disease on a group level.[22] To increase power for variants that were not tested for segregation, we filtered missense variants predicted to be damaging by PolyPhen[23] (v2.2.2) or Sift[24] (v5.2.2). In total, 42% of individuals had zero, 34% had one, 15% had two, and 8% had three or more ultrarare variants (either damaging missense or null variant). We labeled ultrarare variants for which we had no information on segregation as putative de novo variants when they had previously been reported as confirmed DNVs in individuals with NDD+E[4] and/or ClinVar[25] (date August 2017).

Population controls

We used controls as a population reference of ultrarare variant frequencies per gene. The population control data set was assembled at the Broad Institute from multiple exome sequencing projects. It included data from the National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project (for details, see http://evs.gs.washington.edu/EVS/), T2D-Genes study (http://www.type2diabetesgenetics.org/projects/t2dGenes), ATVB cohort (dbGAP accession phs000814.v1.p1), and Ottawa Heart study (dbGAP accession phs000806.v1.p1). All control samples were jointly processed through one alignment and variant calling pipeline. Samples of European ancestry were identified using principal component analysis. All first-degree relatives and duplicated samples were removed from downstream analysis with pairwise identical by descent analysis in PLINK.[26] From this data, a subset of genes present in diagnostic epilepsy panels was then used as control data in this study, excluding samples with a genotype call rate <95% totaling 10,937 individuals with mean age of 65 with no evidence of psychiatric/neurodegenerative disorder. We subjected genotypes for quality checks, keeping only genotypes with >30× coverage and genotype quality (GQ, estimated in GATK pipeline[27]) >30. On average, the number of sites with nonreference genotypes in controls that were excluded due to coverage <30× for this analysis was 1.03% (see Supplementary Figure S2). In one company, this number is on average 0.2% (personal communication). Due to the targeted approach, we expect this to be similarly low in the other company. Diagnostic panels may be at an advantage to identifying variants compared with exomes because they have higher average coverage and were subjected to initially lower GQ cutoffs. On the other hand, variants have been validated by Sanger sequencing in some of the controls, but all of the cases and variants in certain genes were systematically underreported in panels (see Supplementary Fig. S3). Controls are of non-Finnish European origin while cases are mostly of non-Finnish European origin, with few exceptions (personal communication). While controls and cases were not matched for more specific population structure we expect this to have no significant influence in singleton rates as these are relatively consistent in different (particularly non-Finnish European) populations in the 1000 Genomes Project (https://www.nature.com/articles/nature15393/figures/1) and we also show that many singletons in cases are likely of de novo origin.

Statistical analyses

We assessed individual gene tolerance to null or missense variants in the general population by using the probability of loss-of-function intolerance (pLI) score, missense *z*-score (*z*-score of observed vs. expected missense variants),[19] or s_{het} score (selective effects for heterozygous protein null variants).[28] We defined a gene as constrained with the cutoffs >0.9 for pLI, >3.09 for missense *z*-scores, and >0.05 for s_{het} based on recommendations of the score developers. We compared pLI and shet scores of variants using Wilcoxon rank tests because the data appeared not normally distributed upon inspection. As disease gene reference, we used a curated list of disease genes compiled by clinicians as part of the Deciphering Developmental Disorders (DDD) study (http://www.ebi.ac.uk/gene2phenotype/downloads/DDG2P.csv.gz, version 11/7/2018). We subset the list to genes associated with any descending Human Phenotype Ontology (HPO) terms[29] of epilepsy (HP:0001250) or intellectual disability (HP:0001250) or brain/cognition and only included dominant/X-linked disease genes labeled as "confirmed" or "probable". We also annotated missense badness, PolyPhen-2, and constraint (MPC) scores, a pathogenicity score that leverages regional depletion of missense variants in the general population as well as amino acid deleteriousness (K. Samocha et al., bioRxiv, 2017) to compare ultrarare variants and DNV.

RESULTS

Genes with ultrarare variants in NDD+E include DEE but also NDD genes

We assessed frequencies of likely protein-altering (missense or null) ultrarare variants in 6994 individuals with NDD+E (Fig. 1). While we did not assess variant pathogenicity with all American College of Medical Genetics and Genomics (ACMG) criteria,[18] this class of variants should be enriched for likely pathogenic variants. We analyzed 645 genes that were sequenced in at least 2000 individuals with NDD+E, with a median of 4870 individuals sequenced per gene. Of these, 215 genes were annotated as acting in an autosomal

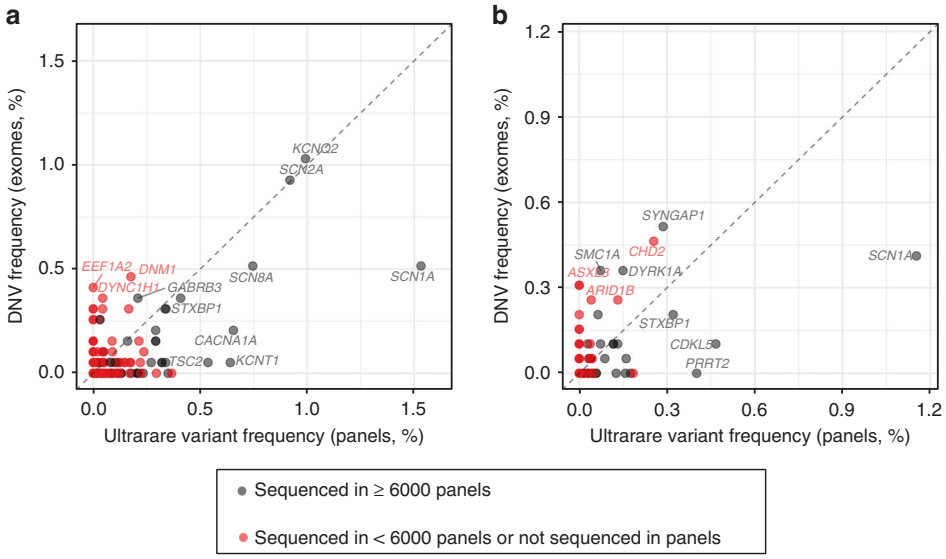

Fig. 1 Ultrarare variants in panels compared with de novo variants (DNVs) in exomes. a Damaging missense variants; **b** null variants. Genes that were sequenced in ≥6000 panels are labeled black; others are in red. The dotted line represents equal frequency of de novo variants in exomes and ultrarare variants in panels. Only variants in confirmed disease genes are shown (see "Materials and methods"). DNV frequency in *SCN1A* should be depleted as it is only occasionally prescreened prior to panel, but is usually prescreened prior to exome sequencing. Frequencies of DNVs in exomes and ultrarare variants in panels are correlated when considering highly covered panel genes (black dots). Missense variants: p value $= 3 \times 10^{-9}$, rho $= 0.63$; null variants: p value $= 4 \times 10^{-6}$, rho $= 0.53$, method: Spearman correlation.

dominant, 47 in an X-linked, 329 in an autosomal recessive, and 54 in an unknown inheritance mode. It has been shown repeatedly[22,30] that genes contributing to severe childhood-onset diseases with high penetrance are depleted for missense/null variants in the general population, measured by pLI/missense z-score.[19] Genes classified as constrained by a significant pLI/missense z-score likely contribute to NDD+E in a dominant/X-linked mode. Of 262 dominant/X-linked genes, 85 genes were constrained and carried at least two ultrarare variants in our data set. Forty-one of these 85 genes were previously described as developmental and/or epileptic encephalopathy (DEE/EE) and NDD+E genes,[4,31,32] while other frequently mutated genes were associated with other well-known genetic syndromes (e.g., *BRAF*, *KMT2D*, *TCF4*) or structural brain abnormalities (e.g., *ARX*, *CASK*, *FLNA*, *TUBB4A*). We compared per-gene ultrarare variant frequencies to 10,937 controls assessing the general population background rate (Supplementary Figure S3). Ultrarare variant frequencies (missense and null) of the top genes were *SCN1A* (2.7%), *KCNQ2* (1.2%), *SCN2A* (1.0%), *CDKL5* (0.8%), *SCN8A* (0.8%), *STXBP1* (0.7%), and *CACNA1A* (0.7%). Reassuringly, ranks of top genes were in concordance with a recently published study of similar design (gene panel sequencing in 8565 epilepsy patients;[17] see Fig. 2).

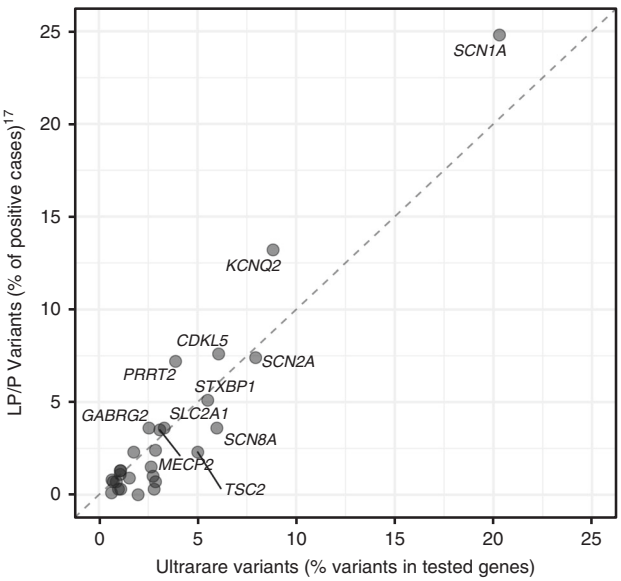

Fig. 2 Pathogenic variants in an independent panel cohort. Ultrarare variants in panels (damaging missense + null) versus pathogenic variants in a panel cohort of 8585 individuals including damaging missense + null+ copy-number variant (CNV) (CNVs constitute about 9% of pathogenic variants).[17] Adapted to the format of Lindy et al.,[17] the fraction of pathogenic variants in each gene is given as the proportion of variants in all positive cases. Only genes included in Lindy et al.[17] are shown. Correlation of data shown: p value $= 4 \times 10^{-7}$, rho $= 0.79$, method: Spearman correlation. *LP/P* likely pathogenic/pathogenic.

Comparing variant frequencies per gene in 6994 panels and 1942 trio exomes

We compared ultrarare variant frequencies in our panel data set to DNV frequencies in a large recent exome-wide trio study of 1942 individuals with NDD+E.[4] Restricting our data set to genes sequenced in 6000 to 6994 individuals, we found correlation between the data sets for both missense and null variants (missense variants: p value $= 3 \times 10^{-9}$, rho $= 0.63$; null variants: p value $= 4 \times 10^{-6}$, rho $= 0.53$, method: Spearman correlation, see Fig. 1). This suggests that a large fraction

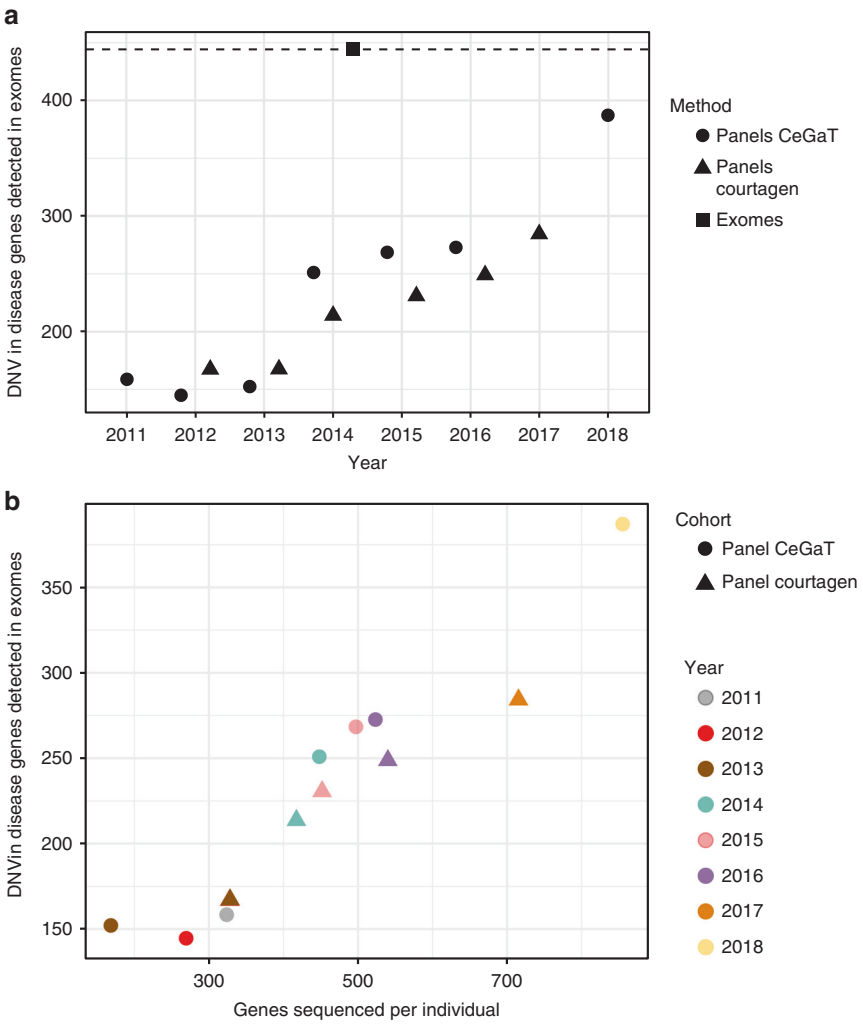

Fig. 3 Number of disease genes in panels increasing over time. We determined how many de novo variants (DNVs) in disease genes (as reported by the Deciphering Developmental Disorders study,[30] see "Materials and methods") would have been found in 1942 individuals with NDD+E as part of an exome-wide study[4] when using panels instead of exomes. For example, of the 444 exome-wide DNVs detected in Heyne et al.,[4] the panel designs in our current study of the years 2011 to 2013 would have covered fewer than 200. After improvement of panel design over the years, up to 300–400 DNVs would have been detected in 2017 to 2018. The increasing number of DNVs in disease genes is correlated with increasing panel sizes over time. As we do not evaluate individual variant pathogenicity and do not include all disease genes, true diagnostic yields would be different and likely higher (usually up to 40% for exomes[33,34] and up to 26% for panels in 2018). **a** Number of individuals with DNVs in disease genes in panels over time (triangles and dots) versus exomes (square). **b** Total number of genes sequenced per patient and number of diagnoses.

of ultrarare variants in our data set arose de novo even if only a fraction of them were tested for segregation. However, there was no or negative correlation (missense variants: p value = 0.05, rho = −0.17; null variants: p value = 0.7, rho = −0.04, method: Spearman correlation) between panels and exome sequencing when considering all genes, as many genes were not included in the diagnostic gene panels. Assuming gene panel sequencing identifies 100% of the DNVs found in trio exome sequencing for a given gene, we investigated how many likely protein-altering DNVs in curated disease genes (see "Materials and methods") would have been found in 1942 individuals with NDD with epilepsy as part of the exome-wide study when using panels instead of exomes. We found 444 DNVs in the exome sequencing data in those genes, while panels would have identified on average 245 DNVs. The proportion of identified DNVs in panels significantly increased over time, however, as panels were continuously updated according to the literature (Fig. 3). In this approach we only consider damaging missense and null DNVs and do not evaluate pathogenicity of individual variants. Therefore, and as the set of disease genes is more strictly defined, true diagnostic yields are likely higher (usually up to 40% for clinical exome sequencing[33,34] and up to 26% in most recent panel diagnostics).

The majority of genes contained no or fewer ultrarare variants in epilepsy cases than in controls

Comparing variants in cases and controls, we noticed that 255 of 645 panel genes (39.5%) did not display any ultrarare variants in >2000 NDD+E cases (Fig. 1, Supplementary

Figure S3, Supplementary Table S2). Further 247 genes (38.3%) had lower frequencies of ultrarare variants in cases compared with population controls. The majority of these in total 502 rarely mutated genes were of autosomal recessive inheritance (60%, 300/502), for which we would not expect higher variant frequencies in cases. However, 30% (149/502 genes) were of autosomal dominant/X-linked inheritance (Supplementary Table S2). For the remaining 10%, inheritance was unclear. A limitation of this study is that we cannot guarantee that variants in genes not ordered by clinicians were consistently reported and therefore we cannot exclude that some missed variants were in true disease genes. However, the 149 not or rarely mutated dominant/X-linked genes had lower constraint scores (pLI: 0.93 [median], missense *z*-score: 2.2 ± 2.6 [mean, SD], s_{het} score: 0.08 [median]) on a group level than the 119 dominant/X-linked genes with higher ultrarare variant frequencies (pLI: 0.99 [median], missense *z*-score: 3.5 ± 2.1 [mean, SD], s_{het} score: 0.16 [median]) in cases than in controls (see "Materials and methods"; respective *p* values pLI score: 5×10^{-4}, missense z-score: 2×10^{-5}, s_{het} score: 1×10^{-4}). Additionally, they were not significantly different in estimated mutation rate[35] (two-sided *t* test, *p* values for missense mutation rate 0.76, null mutation rate 0.86). This suggests that a lower mutation rate is not the reason for low frequencies of ultrarare variants in most of these genes among NDD+E cases; instead it is more likely that most of these genes are not true NDD+E genes.

Confirmed and putative de novo variants

Among 6994 epilepsy cases, we revealed 333 DNVs that were not in ExAC as well as either damaging missense or null DNVs. Ninety-five percent (317/333) of DNVs were in 54 constrained genes, 32 genes displayed at least two DNVs (Supplementary Table S1, Fig. **1**, Supplementary Figure S4), and 4.7% (331/6994) of cases had a total number of 333 damaging DNVs. As segregation testing was not performed systematically in the overall cohort, this number is certainly an underestimate (see "Materials and methods"). It has been documented for many disease genes, including genes associated with NDD+E,[36] that disease-causing missense variants cluster in particular functionally relevant protein domains. We annotated MPC scores, a pathogenicity score that considers if missense variants in the general population are depleted in particular regions of a gene (K. Samocha et al., bioRxiv, 2017). Higher MPC scores indicate increased deleteriousness of missense variants. We found a median MPC score of 2.3 for 333 DNVs and 0.76 for 11,233 ultrarare variants for which disease segregation was unknown (Wilcoxon rank sum test, *p* value 1×10^{-76}). Also within constrained genes, we found a median MPC of 2.13 in DNVs and 1.03 for variants with unknown segregation status (Wilcoxon rank sum test, *p* value 2×10^{-45}). These results confirm the increased likelihood of pathogenicity of DNVs in comparison with ultrarare variants with unknown disease segregation.

DISCUSSION

Gene panel analysis is widely used in genetic diagnostics of NDD+E. However, panel designs vary substantially across companies[4,15] and over time. Here, we report a large cohort of individuals with NDD+E (6994 cases) that underwent gene panel sequencing in a diagnostic setting.

Frequencies of ultrarare variants in our cohort were compared with two other large NDD+E cohorts: (1) DNVs in 1942 trio exomes[4] and (2) likely pathogenic variants in 8565 gene panels.[17] Of the top 20 disease genes with the highest numbers of DNVs in exomes, 16 were also present in our panel data. *ARID1B*, *ASXL3*, *EEF1A2*, and *SLC6A1* were the genes missing in panels. Considering the top 35 disease genes in exomes (at least 4 DNVs in exomes), the following genes were missing in panels: *KCNH1*, *PURA*, *COL4A3BP*, *KIF1A, ANKRD11*, *DDX3X*, *MED13L*, and *PPP2R5D*. We suggest those genes could be added in subsequent panel designs. Of the top 20 exome genes, only 8 were present in Lindy et al.[17] These results illustrate the high genetic heterogeneity of NDD+E. The most frequently mutated genes in exomes as well as panels were *SCN1A, SCN2A*, and *KCNQ2*. Following at about half their frequency were *CDKL5*, *SCN8A*, *STXBP1*, *SYNGAP1*, *TSC2*, and *CACNA1A*. These genes are consistently present at high diagnostic yield in NDD+E.[11,13,33] Of note, *GABRG2*, *TSC2*, and *PRRT2* had high frequencies of ultrarare variants in our panel study and in Lindy et al.[17] but barely displayed DNVs in exomes (*GABRG2* and *TSC2*: 1 DNV, *PRRT*: 0 DNV) suggesting that many of the variants in *TSC2*, *GABRG2*, and *PRRT2* may be inherited rather than de novo.

While many disease genes affected in trio exomes were not included in panel designs, we show that gene panel content consistently improves over time. Many frequently mutated genes are associated with "classic" developmental and/or epileptic encephalopathies, whereas others are associated with more unspecific diagnoses of NDD. A too narrow target on "classic epilepsy genes" therefore neglects that NDDs are accompanied by epilepsy in approximately 20% of cases and therefore any NDD gene is principally potentially also associated with NDD+E.[4,30] Aptly, we recently showed that 24 diagnostic providers of panel sequencing also lacked a substantial fraction of NDD+E-associated genes in their panel designs.[4] In the early days of next-generation sequencing (NGS), small panel sequencing allowed the introduction of this new technology into clinical diagnostics. Today, panels still offer a cost-effective method to diagnose causal pathogenic variants in the most commonly affected genes as in some countries current reimbursement frameworks do not adequately cover the additional costs of exome sequencing. Yet exome sequencing covers far more of the genetic heterogeneity of NDD+E, and a recent US study found that panels are not necessarily more cost-effective than exome sequencing in the United States.[37] The number of NDD+E disease genes is continuously increasing, which has only become possible by wide adoption of in particular trio exome/genome sequencing approaches. Detection rates with panel

diagnostics are necessarily limited by medical knowledge at the time of panel design. On the other hand, higher coverage in panels than in exomes is superior in detecting low-grade mosaicism in a patient.

The majority of dominant/X-linked panel genes (502 of 645) either did not display any ultrarare variants in >2000 epilepsy cases or had even lower frequencies of ultrarare variants in cases than controls. This could be due to a low mutation rate of these genes or a phenotype rarely ascertained in our cohort. However, given the fact that these genes had no significantly different mutation rate but significantly lower constraint scores compared with all other dominant or X-linked genes in this study, it is likely that many of them are not disease-associated. This observation is paralleled by a study of similar design on 7855 individuals with childhood-onset cardiomyopathy, where several genes frequently sequenced in clinical routine could also not be convincingly associated with disease.[10] In our study, panel design originated in 2010, when multiple candidate genes for rare diseases were nominated without sufficient statistical evidence and could not be confirmed in a clinical setting.[38] This was also described specifically for epilepsy genetics.[4,16]

Of 645 panel genes in our study, 329 genes were associated with recessive inheritance. However, variants in recessive genes segregating with disease were only observed in approximately 1.7% (27/1633 cases) within a documented subfraction of this study. This is in concordance with rates of 1.3% (ref. [12]) ($n = 775$ cases) and 1.1% (ref. [17]) ($n = 8565$ cases) in two recent NDD+E studies using gene panels and 3.6% ($n = 7448$ cases) in an exome-wide study on developmental disorders with and without epilepsy from nonconsanguineous families.[39] Thus, panel designs usually display an imbalanced distribution of recessive genes (very few percent of diagnoses but approximately half of panel genes) versus dominant genes (vast majority of diagnoses but only half of panel genes).

Limitations of our study include inconsistent variant reporting in cases and that the different cohorts we compared were neither technically nor ancestry matched. However, we do not expect these technical limitations to alter the key conclusions of this study (see "Materials and methods").

We also evaluated the frequencies of ultrarare variants in five genes with contradictory evidence of gene–disease relationship, which thus had been classified as "disputed" by the formal criteria of the ClinGen Consortium[16] (*CACNA1H*, *CACNB4*, *EFHC1*, *MAGI2*, *SRPX2*) as well as two genes with contradictory susceptibility to epilepsy (*CLCN2*, *GABRD*). *CACNB4*, *EFHC1*, *MAGI2*, *SRPX2*, *CLCN2*, and *GABRD* showed identical frequencies of ultrarare variants in cases compared with controls (Supplementary Figure S5). Thus, our findings support the evidence that *CACNB4*, *EFHC1*, *MAGI2*, *SRPX2*, *CLCN2*, and *GABRD* may not be associated with epilepsy. Coverage of *CACNA1H* was too poor in controls from ExAC to allow a valid comparison of variant frequencies between cases and controls.

In summary, our data provide evidence to further improve the design of NDD+E panels by (1) including genes with highest burden of ultrarare variants, (2) adjusting the ratio of autosomal dominant and X-linked genes with high diagnostic yield versus autosomal recessive genes with low diagnostic yield, and (3) excluding genes with poor evidence for true disease association or very few ultrarare variants in epilepsy cases.

SUPPLEMENTARY INFORMATION

The online version of this article (https://doi.org/10.1038/s41436-019-0531-0) contains supplementary material, which is available to authorized users.

ACKNOWLEDGEMENTS

We thank the team of the Institute of Human Genetics of the University of Leipzig as well as the Analytic and Translational Genetics Unit, Boston, for helpful discussions. 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 and HE7987/1–2. Y.G. W. was supported by DFG grants WE4896/3–1 and WE4896/4–1.

CODE AVAILABILITY

All statistical analyses were done with the R programming language (www.r-project.org). The code is available upon request.

DISCLOSURE

The data presented here come from two commercial companies. C.M.S., V.T. and D.R.S. have been employees of Courtagen. S.B. is an owner of CeGaT, F.B. is an employee of CeGaT. The other authors declare no conflicts of interest.

REFERENCES

1. Thomas RH, Berkovic SF. The hidden genetics of epilepsy—a clinically important new paradigm. Nat Rev Neurol. 2014;10:283–292.
2. Myers CT, Mefford HC. Advancing epilepsy genetics in the genomic era. Genome Med. 2015;7:91.
3. Allen AS, Berkovic SF, Cossette P, et al. de novo mutations in epileptic encephalopathies. Nature. 2013;501:217–221.
4. Heyne HO, Singh T, Stamberger H, et al. de novo variants in neuro developmental disorders with epilepsy. Nat Genet. 2018;50:1048–1053.
5. Wolff M, Johannesen KM, Hedrich UB, et al. Genetic and phenotypic heterogeneity suggest therapeutic implications in SCN2A-related disorders. Brain. 2017;140:1316–1336.
6. Møller RS, Johannesen KM. Precision medicine: SCN8A encephalopathy treated with sodium channel blockers. Neurotherapeutics. 2016;13:190–191.
7. Millichap JJ, Park KL, Tsuchida T, et al. KCNQ2 encephalopathy: features, mutational hot spots, and ezogabine treatment of 11 patients. Neurol Genet. 2016;2:e96.
8. De Giorgis V, Veggiotti P. GLUT1 deficiency syndrome 2013: current state of the art. Seizure. 2013;22:803–811.
9. Easton DF, Pharoah PD, Antoniou AC, et al. Gene-panel sequencing and the prediction of breast-cancer risk. N Engl J Med. 2015;372:2243–2257.
10. Walsh R, Thomson KL, Ware JS, et al. Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples. Genet Med. 2017;19:192–203.

11. Lemke JR, Riesch E, Scheurenbrand T, et al. Targeted next generation sequencing as a diagnostic tool in epileptic disorders. Epilepsia. 2012;53: 1387–1398.
12. Berg AT, Coryell J, Saneto RP, et al. Early-life epilepsies and the emerging role of genetic testing. JAMA Pediatr. 2017;171:863–871.
13. Trump N, McTague A, Brittain H, et al. Improving diagnosis and broadening the phenotypes in early-onset seizure and severe developmental delay disorders through gene panel analysis. J Med Genet. 2016;53:310–317.
14. Xue Y, Ankala A, Wilcox WR, Hegde MR. Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing. Genet Med. 2015;17:444–451.
15. Chambers C, Jansen LA, Dhamija R. Review of commercially available epilepsy genetic panels. J Genet Couns. 2016;25:213–217.
16. Helbig I, Riggs ER, Barry CA, et al. The ClinGen Epilepsy Gene Curation Expert Panel—bridging the divide between clinical domain knowledge and formal gene curation criteria. Hum Mutat. 2018;39:1476–1484.
17. Lindy AS, Stosser MB, Butler E, et al. Diagnostic outcomes for genetic testing of 70 genes in 8565 patients with epilepsy and neurodevelopmental disorders. Epilepsia. 2018;59:1062–1071.
18. Richards S, Aziz N, Bale 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. 2015;17:405–424.
19. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536:285–291.
20. Smith DR, Stanley CM, Foss T, Boles RG, McKernan K. Rare genetic variants in the endocannabinoid system genes CNR1 and DAGLA are associated with neurological phenotypes in humans. PLoS ONE. 2017;12: e0187926.
21. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17:122.
22. Kosmicki JA, Samocha KE, Howrigan DP, et al. Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. Nat Genet. 2017;49:504–510.
23. Adzhubei I, Jordan DM, Sunyaev SR Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013; Chapter 7: Unit7.20.
24. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31:3812–3814.
25. Landrum MJ, Lee JM, Benson M, et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res. 2016;44 (D1):D862–868.
26. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–575.
27. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43:11–33.
28. Cassa CA, Weghorn D, Balick DJ, et al. Estimating the selective effects of heterozygous protein-truncating variants from human exome data. Nat Genet. 2017;49:806–810.
29. Kohler S, Vasilevsky NA, Engelstad M, et al. The Human Phenotype Ontology in 2017. Nucleic Acids Res. 2017;45(D1):D865–D876.
30. Deciphering Developmental Disorders Study. Prevalence and architecture of de novo mutations in developmental disorders. Nature. 2017;542: 433–438.
31. EpiPM Consortium. A roadmap for precision medicine in the epilepsies. Lancet Neurol. 2015;14:1219–1228.
32. Epi4K Consortium, Epilepsy Phenome/Genome Project. Ultra-rare genetic variation in common epilepsies: a case-control sequencing study. Lancet Neurol. 2017;16:135–143.
33. Helbig KL, Farwell Hagman KD, Shinde DN, et al. Diagnostic exome sequencing provides a molecular diagnosis for a significant proportion of patients with epilepsy. Genet Med. 2016;18:898–905.
34. Srivastava S, Cohen JS, Vernon H, et al. Clinical whole exome sequencing in child neurology practice. Ann Neurol. 2014;76:473–483.
35. Samocha KE, Robinson EB, Sanders SJ, et al. A framework for the interpretation of de novo mutation in human disease. Nat Genet. 2014;46:944–950.
36. Traynelis J, Silk M, Wang Q, et al. Optimizing genomic medicine in epilepsy through a gene-customized approach to missense variant interpretation. Genome Res. 2017;27:1715–1729.
37. Sánchez Fernández I, Loddenkemper T, Gaínza-Lein M, Sheidley BR, Poduri A Diagnostic yield of genetic tests in epilepsy: A meta-analysis and cost-effectiveness study. Neurology. 2019.
38. MacArthur DG, Manolio TA, Dimmock DP, et al. Guidelines for investigating causality of sequence variants in human disease. Nature. 2014;508:469–476.
39. Martin HC, Jones WD, McIntyre R, et al. Quantifying the contribution of recessive coding variation to developmental disorders. Science. 2018; 362:1161–1164.