Detection rate of causal variants in severe childhood epilepsy is highest in patients with seizure onset within the first four weeks of life

David Staněk1*, Petra Laššuthová1, Katalin Štěrbová2, Markéta Víčková3, Jana Neupauerová1, Marcela Krůtová4 and Pavel Seeman1

Abstract

Background: Epilepsy is a heterogeneous disease with a broad phenotypic spectrum and diverse genotypes. A significant proportion of epilepsies has a genetic aetiology.

In our study, a custom designed gene panel with 112 genes known to be associated with epilepsies was used. In total, one hundred and fifty-one patients were tested (86 males / 65 females).

Results: In our cohort, the highest probability for the identification of the cause of the disease was for patients with a seizure onset within the first four weeks of life (61.9% clarification rate) – about two times more than other groups. The level of statistical significance was determined using a chi-square analysis.

From 112 genes included in the panel, suspicious and rare variants were found in 53 genes (47.3%). Among the 151 probands included in the study we identified pathogenic variants in 39 patients (25.8%), likely pathogenic variants in three patients (2%), variants of uncertain significance in 40 patients (26.5%) and likely benign variants in 69 patients (45.7%).

Conclusion: Our report shows the utility of diagnostic genetic testing of severe childhood epilepsies in a large cohort of patients with a diagnostic rate of 25.8%. A gene panel can be considered as a method of choice for the detection of pathogenic variants within patients with unknown origin of early onset severe epilepsy.

Keywords: Epilepsy, Epileptic encephalopathy, Targeted gene panel testing, MPS, Phenotype, KCNQ2
Patients
One hundred and fifty-one unrelated patients with severe childhood epilepsy were included in the study (86 males / 65 females). In the majority of cases, the epilepsy occurred sporadically (138/152), while in the remaining 13 patients the occurrence was familial.

Probands were referred for genetic analysis over a period from March 2015 to December 2016. The patients’ legal representatives all gave informed consent and the study was approved by the local ethics committee. Brain MR imaging revealed no structural abnormalities in any of the patients. DNA samples from both parents were collected for the interpretation of variants; in two exceptional cases this was not possible.

Optional: previously tested with array CGH – to exclude chromosomal aberrations as the cause of epileptic seizures.

Design
A custom gene panel design was created with SureDesign (SureDesign release 3.5.x, Agilent, California, USA) application. Genes were chosen according to these criteria:

1. “Known” epilepsy genes: At least two published reports describing a causal relationship between variants in the gene and epilepsy.

OR

2. At least one published report describing a causal relationship between variants in the gene and epilepsy in two or more unrelated patients.

At first, according to these criteria and a literature search, we included 97 relevant genes in the gene panel design (07/2015). Then, the second version (in 03/2016) was enriched by newly reported genes up to a final number of 112. Genes included in the designs are listed in the Additional file 1: part II.

Sequencing was performed on a MiSeq Desktop Sequencer from Illumina (Illumina, California, USA) with 2 × 150 bp sequencing kit (20 samples per run).

Data analysis
Data were analysed by two independent software tools - NextGene (NextGENe 2.41, Softgenetics, Pennsylvania, USA) and SureCall (SureCall 3.0.3.x, Agilent, California, USA).

NextGene analysis was performed with default settings. SureCall analysis was performed on default settings except for “SNP Read depth filter”. This value was set to value 10. The aim of this analysis was to increase the sensitivity of the whole process.

Afterwards, Alamut Batch (Alamut Batch 1.5.2, Interactive Biosoftware, Rouen, France) was used for annotating merged VCFs into a tabular file.

Variant evaluation
An Alamut Batch annotated file in tabular format was used as the input for evaluation. Variants were then filtered according to the following workflow:

1. Variants found in three or more patients from the same run were filtered out

![Fig. 1](image-url) The number of pathogenic variants in individual genes stratified by inheritance mode. AD = autosomal dominant, AR = autosomal recessive, XL = X-linked. Legend: X axis: all genes Y axis: Number of pathogenic variants
| Gene       | Ref Seq      | DNA-level          | Protein level        | AD/AR | DN/DN | Prediction (SIFT, PolyPhen2, ClinVar) | ExAC all |
|------------|--------------|--------------------|----------------------|-------|-------|---------------------------------------|----------|
| **Pathogenic AD variants**                      |              |                    |                      |       |       |                                       |          |
| CACNA1A    | NM_001127221.1 | c.13319826G > T    | AD                   | DN    |       |                                       |          |
| CACNA1A    | NM_001127221.1 | c.2663A > T        | p.Gln888Leu          | AD    | DN    |                                       | 0.0005   |
| dup chr9   |              |                    |                      |       |       |                                       |          |
| GABRB3     | NM_000814.5   | c.841A > G         | p.Thr281Ala          | AD    | DN    |                                       |          |
| GABRG2     | NM_000816.3   | c.968G > A         | p.Arg323Gln          | AD    | DN    |                                       |          |
| GRIN1      | NM_007327.3   | c.2443G > A        | p.Gly815Arg          | AD    | DN    |                                       |          |
| GRIN1      | NM_007327.3   | c.1643G > A        | p.Arg548Gln          | AD    | DN    |                                       |          |
| HCN1       | NM_021072.3   | c.1189A > G        | p.Ile397Leu          | AD    | DN    |                                       |          |
| KCNQ2      | NM_172107.2   | c.826A > C         | p.Thr276Pro          | AD    | DN    |                                       |          |
| KCNQ2      | NM_172107.2   | c.1004C > G        | p.Pro335Arg          | AD    | DN    |                                       |          |
| KCNQ2      | NM_172107.2   | c.701C > T         | p.Thr234Ile          | AD    | DN    |                                       |          |
| KCNQ2      | NM_172107.2   | c.913_915delTTC    | p.Phe305del          | AD    | DN    |                                       |          |
| SCN1A      | NM_0011202435.1 | c.1244 T > A      | p.Ile415Lys          | AD    | DN    |                                       |          |
| SCN1A      | NM_001156963.1 | c.5384A > G       | p.Glu1795Gly         | AD    | DN    |                                       |          |
| SCN1A      | NM_001156963.1 | c.4384dup         | p.Tyr1462Leufs*24    | AD    | DN    |                                       |          |
| SCN1A      | NM_001165963.1 | c.1178G > A       | p.Arg393His          | AD    | DN    |                                       |          |
| SCN1A      | NM_001156963.1 | c.1525C > T       | p.Gln509*            | AD    | DN    |                                       |          |
| SCN2A      | NM_0010404142.1 | c.2774 T > C      | p.Met925Thr          | AD    | DN    |                                       |          |
| SCN2A      | NM_0010404142.1 | c.5009C > T      | p.Thr1862Ile         | AD    | DN    |                                       |          |
| SCN8A      | NM_014191.3   | c.4921C > G       | p.Leu1641Val         | AD    | DN    |                                       |          |
| SCN8A      | NM_014191.3   | c.2549G > A       | p.Arg850Gln          | AD    | DN    |                                       |          |
| SCN8A      | NM_014191.3   | c.4850G > T       | p.Arg1617Leu         | AD    | DN    |                                       |          |
| STXBP1     | NM_003165.3   | c.1654 T > C      | p.Cys552Arg          | AD    | DN    |                                       |          |
| UBEE3A     | NM_130838.1   | c.1149G > C       | p.Glu383Asp          | AD    | INH   |                                       |          |
| **Pathogenic AR variants**                      |              |                    |                      |       |       |                                       |          |
| ALOH7A1    | NM_001182.4   | c.1318-1G > C     | AR                   | INH   |       |                                       | 0.00041  |
| ALOH7A1    | NM_001182.4   | c.518-14_518delinsCA | AR               | INH   |       |                                       |          |
| SLC13A5    | NM_177550.3   | c.425C > T        | p.Thr142Met          | AR    | INH   | D                                       | P        |
| TREX1      | NM_016381.3   | c.10621072del     | p.Leu354Phefs*22     | AR    | UNK   |                                       |          |
| TREX1      | NM_016381.3   | c.1072A > C       | p.Thr358Pro          | AR    | INH   | T                                       | P        | 0.00016  |
| **Pathogenic X-linked variants**                |              |                    |                      |       |       |                                       |          |
| CDKL5      | NM_003159.2   | c.2578C > T       | p.Gln860*            | XL    | DN    |                                       |          |
| CDKL5      | NM_003159.2   | c.463 + 5G > A    |                         | XL    | DN    |                                       |          |
| CDKL5      | NM_003159.2   | c.1247_1248del    |                         | XL    | DN    |                                       | P        |
| IQSEC2     | NM_001111125.2 | c.3206G > C      | p.Arg1069Pro         | XL    | INH   | D                                       | PD       |
| MECP2      | NM_004992.3   | c.1219_1229del    | p.Asp407Glnfs*25     | XL    | DN    |                                       |          |
| WDR45      | NM_007075.3   | c.654del          | p.Arg219Alafs*69     | XL    | DN    |                                       |          |
| WDR45      | NM_007075.3   | c.970_971del      | p.Val324Hisfs*17     | XL    | DN    |                                       |          |
| WDR45      | NM_007075.3   | c.511C > T        | p.Gln171*            | XL    | DN    |                                       |          |
2. Variants with a higher percentage (over 1%) in population databases (ExAC, 1000G) were deprioritized.

3. Classification into four groups based on criteria in the annotated file (such as: ACMG classification [4], prediction programs – SIFT [5], Polyphen2 [6], Mutation taster [7], Clinvar [8], conservation, inheritance, X-linked disease). Groups were defined by pathogenicity of the variant: Pathogenic, likely pathogenic, variants of uncertain significance, benign and likely benign were grouped together.

**CNV analysis**

We used tools integrated into NextGENe for the detection of copy number variations. For the analysis, samples were compared with healthy control samples. NextGENe CNV tools perform a detection based on the Hidden Markov Model [9].

The resulting report shows INDELs in a tabular file. For precise analysis, this CNV Tool was performed against a different group of healthy controls, see Additional file 1: part III.

**Parental testing**

For each pathogenic or likely pathogenic variant, Sanger sequencing and segregation analysis was performed and evaluated.

**Results**

After analysis of the whole dataset, we selected 99 SNVs and one CNV for further analysis. From 112 genes included in the panel, suspicious and rare variants were found in 53 genes (47.3%).

Most frequently, variants were found in **SCN1A** (eight occurrences) and **KCNQ2** (five occurrences). All variants found in the project were then stratified into these four classes:

- Pathogenic
- Likely pathogenic
- Variants of uncertain significance (VUS)
- Benign and likely benign

For the next step, Sanger sequencing of variants, classified as Pathogenic or Likely pathogenic, was used; the total was 42 SNVs in 22 genes out of 112 in the panel (19.6%). Furthermore we identified one CNV on chr 9. (Fig. 1 and Table 1).

The rest of the variants (VUS and Benign) are summarized in Additional file 1: part IV.

Among the 151 probands included in the study, we identified pathogenic variants in 39 patients (25.8%), likely pathogenic variants in three patients (2%), variants of uncertain significance in 40 patients (26.5%) and likely benign variants in 69 patients (45.7%) (Fig. 2). Two patients were carriers of two pathogenic variants – one was a compound heterozygote for two variants in trans in the **ALDH7A1**, the second was a compound heterozygote for two variants in trans in the **TREX1**.
Inheritance patterns
Our results showed that the majority of pathogenic or likely pathogenic variants were found in genes that follow autosomal dominant pattern of inheritance (27 SNVs in 14 genes), also CNV on chr9 was found to be acting with AD inheritance. Another 5% were found in genes associated with autosomal recessive inheritance (five SNVs in three genes), and finally 24% were found in genes with X-linked inheritance (11 SNVs in five genes).

For 40 patients, out of 42 classified with pathogenic or likely pathogenic variants, DNA samples from both parents were available for segregation analysis by Sanger sequencing.

In 34 patients, including 33 with SNVs and one with CNV, these variants arose de novo. These de novo variants were found in 15 genes and CNV was found in chromosome 9.

Nine variants in seven genes (detected in 7 patients) were inherited.

In two cases only one parental sample was available. In one of these cases the detected variant was inherited from the mother.

Information about maternal or paternal inheritance of inherited (and unknown) variants are available in Additional file 1: part V.

The distribution of the age at seizure onset and age at inclusion into the study
The distribution of the age at seizure onset and age at inclusion into the study is described in the Additional file 1: part I. The information was gathered from the patient’s documentation. The median age at seizure onset of the whole group was 14.5 months; the first quartile was 4 months and the third quartile was 36 months. The median age at inclusion into study was 93 months; the first quartile was 49.5 months and the third quartile was 169 months.

The probability of finding the pathogenic variant in relation to age at seizure onset
The gene panel testing indicated that the highest probability for finding the cause of epilepsy was in the cohort of patients with the earliest onset of seizures, i.e. within the first four weeks of life – 61.9% clarification rate (13/21 patients).

In other age groups the clarification rate was lower: 35.8% (19/53) in the group of patients with first seizure between four weeks and 12 months of age; 11.1% (5/45) in the group with first seizure between 12 months and 36 months of age; and 15.6% in patients with first seizure after the age of 36 months (5/32) (Fig. 3).

The dataset was analysed using the Chi squared statistic and the results were statistically significant ($p = 0.000052$; the result is significant at $p < 0.05$). Moreover, we always compared two groups with the Fisher’s exact test and significant relationships between the age at onset and clarification rate were observed for group 1 (<four weeks) and group 2 (4 weeks – 12 months) $p = 0.0673$; group 1 and group 3(12 months – 36 months) $p = 0.0001$; group 1 and group 4 (>36 months) $p = 0.0009$. However, no
significant difference was observed when comparing group 3 vs. group 4 (p = value 0.7330).

**Discussion**

We were able to identify the cause of severe childhood epilepsy in 25.8% of patients from our cohort and this finding concurs with previously published reports [1, 2]. From the 112 genes in the panel, pathogenic or likely pathogenic variants occurred in 22 of them.

MPS gene panel testing enables the testing of a large number of genes in parallel with very high coverage and low costs. More supportive information is presented in Additional file 1: part VI.

Over 80% of the pathogenic variants arose de novo, as they were not present in the parental samples (confirmed by Sanger sequencing).

**Special cases**

The first case was a patient with a CHD2 variant, where the father’s sample was not available. The second patient, with two heterozygous variants in gene TREX1, has one variant inherited from his/her mother (p.Thr358Pro) while the other variant is of unknown inheritance, as the father’s sample was not available. A female proband with Angelman syndrome has a variant in gene UBE3A inherited from her healthy mother. This is caused by imprinting, when the proband’s maternal allele is active and the paternal is not expressed. The mother inherited the variant from her father.

**Comparison with previously published reports about epilepsy gene panels**

In order to assess the sensitivity and specificity of our approach, we compared our data with previously published reports. The analysis is presented in Fig. 4.

In agreement with the study by Helbig, et al. [10], the diagnostic rate was approximately 30%. Moreover, the vast majority of the pathogenic variants in currently known genes are de novo. Our results are very similar to those shown by Trump, et al. [11] (94%). According the study by Kovel, et al. [12], in which the authors adopted a very different approach and designed a large panel consisting

![Fig. 4](image-url)
of 351 genes, the diagnostic efficiency of their panel did not increase and was much lower than the expected 30%.

Designing a large panel can help to find some rare variants which other panels cannot reveal; but it also involves higher costs, lower coverage, lower read depth and more difficult interpretations of the more variants that are multiplied by each sample. Based on our experience, we would recommend the design of a panel with approximately 100 well selected genes.

Age at seizure onset
Our results show that the probability of finding the pathogenic variant is the highest in patients with the earliest age at seizure onset (results in Fig. 3). Among our patients, the detection rate for pathogenic or likely pathogenic variant found in a group with seizure onset during the first four weeks of life was 61.9%.

In the groups where the onset of seizure was later, the number of pathogenic or likely pathogenic variants was found to be significantly lower. In a group where the age of seizure onset was between four weeks to 12 months it was 35.8%; between 12 months and 36 months, 11.1%; and after 36 months, 21.7%.

In effect, the earlier the phenotype is manifest then the chances of finding a pathogenic variant are significantly higher. This is shown in Fig. 3.

This trend has also been described previously by Helbig, et al. [10].

CNV testing
In our cohort two CNVs were found. The first classified as Pathogenic (on chr 9) for epilepsy and the second classified as VUS. These were detected using NextGENe CNV comparison tool based on Hidden Markov Model

Results and were further confirmed by an Array CGH.

Conclusion
Our study has proven that MPS gene panel is a powerful tool for the DNA diagnosis of severe MRI negative childhood epilepsies. Today, a gene panel is an optimal method for the identification of pathogenic variants in highly heterogeneous disorders such as the genetically determined disorders including severe childhood epilepsies. In a cohort of 151 patients, we were able to identify the cause of epilepsy in 27.8% of patients (39 patients with pathogenic variant and three with likely pathogenic variants).

Abbreviations
AD: Autosomal dominant; AR: Autosomal recessive; CNV: Copy number variation; MPS: Massively parallel sequencing; SNV: Single nucleotide variant; VUS: Variant of uncertain significance; WES: Whole exome sequencing

Acknowledgements
We would like to thank the parents for their participation and consent for the analyses.

Funding
This work was supported by the Ministry of Health Czech Republic [AZV 15-33041].

Availability of data and materials
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Authors' contributions
DS participated in the bioinformatics analyses and drafted the manuscript. PL carried out the molecular genetic studies and evaluated the data. MK and JN participated in the molecular genetic studies and carried out the Sanger sequencing. KS performed the clinical evaluation, neurological consultation and recruited the patients. MV performed genetic consultations; PS recruited the patients and designed the study. All authors read and approved the final version of the manuscript.

Ethics approval and consent to participate
The study was approved by the ethics committee of University Hospital Motol.

Consent for publication
Informed consent was obtained from all patients or their legal representatives.

Competing interests
The authors whose names are listed in Authors name section certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers’ bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
1Department of Paediatric Neurology, DNA Laboratory, 2nd Faculty of Medicine, Charles University in Prague and University Hospital Motol, Prague, Czech Republic. 2Department of Paediatric Neurology, 2nd Faculty of Medicine, Charles University in Prague and University Hospital Motol, Prague, Czech Republic. 3Department of Biology and Medical Genetics, 2nd Faculty of Medicine, Charles University and University Hospital Motol, Prague, Czech Republic. 4Department of Medical Microbiology, 2nd Faculty of Medicine, Charles University and University Hospital Motol, Prague, Czech Republic.

Received: 13 October 2017 Accepted: 18 April 2018
Published online: 02 May 2018

References
1. Thomas RH, Berkovic SF. The hidden genetics of epilepsy-a clinically important new paradigm. Nat Rev Neurol. 2014;10(5):283–92.
2. Mercimek-Mahmutoglu S, Patel J, Cordeiro D, Hewson S, Callen D, Donner EJ, Hahn CD, Kanu P, Kobayashi J, Minassian BA, et al. Diagnostic yield of genetic testing in epileptic encephalopathy in childhood. Epilepsia. 2015; 56(5):707–16.
3. Nieh SE, Sherr EH. Epileptic encephalopathies: new genes and new pathways. Neurotherapeutics. 2014;11(4):796–806.
4. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, 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(5):405–24.
5. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4(7):1073–81.
6. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248–9.
7. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods. 2014;11(4):361–2.
8. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res. 2016;44(D1):D862–8.
9. Plagnol V, Curtis J, Epstein M, Mok KY, Stebbings E, Grigoriadou S, Wood NW, Hambleton S, Burns SO, Thrasher AJ, et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. Bioinformatics. 2012;28(21):2747–54.
10. Helbig KL, Farwell Hagman KD, Shinde DN, Mroske C, Powis Z, Li S, Tang S, Helbig I. Diagnostic exome sequencing provides a molecular diagnosis for a significant proportion of patients with epilepsy. Genet Med. 2016;18(9):898–905.
11. Trump N, McTague A, Brittain H, Papandreou A, Meyer E, Ngoh A, Palmer R, Morogh D, Boastred C, Hurst JA, 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(5):310–7.
12. de Kovel CG, Brilstra EH, van Kempen MJ, Van't Slot R, Nijman U, Afawi Z, De Jonghe P, Djemie T, Guerini R, Hardies K, et al. Targeted sequencing of 351 candidate genes for epileptic encephalopathy in a large cohort of patients. Mol Genet Genomic Med. 2016;4(5):568–80.