Two Distinct De novo Pathogenic Sequence Variants in the SCN2A Gene in Children with Early Infantile Epileptic Encephalopathy/Ohtahara Syndrome

Marketa Wayhelova 1,2, Jan Oppelt 3,4, Jan Smetana 1,2, Hana Filkova 2, Eva Hladilkova 2, Jana Soukalova 2, Renata Gaillyova 2 and Petr Kuglik 1,2 *

1 Faculty of Science, Institute of Experimental Biology, Masaryk University, Brno, Czech Republic
2 Department of Medical Genetics, University Hospital, Brno, Czech Republic
3 Faculty of Science, National Centre for Biomolecular Research, Masaryk University, Brno, Czech Republic
4 Centre for Applied Genomics, Netherlands Cancer Institute, Amsterdam, The Netherlands

Abstract

Early infantile epileptic encephalopathy/Ohtahara syndrome represents a group of genetically heterogeneous disorders affecting normal brain development and functioning. In this work, we present a case of two unrelated children diagnosed with early infantile epileptic encephalopathy and associated multiple congenital abnormalities. To obtain genetic diagnosis, these children were examined through multi-step diagnostic algorithm including G-banded karyotype analysis and whole-genomic screening using array-based comparative genomic hybridization (array-CGH) with negative results. Additionally, these children and their unaffected parents were enrolled for our pilot study of targeted next-generation sequencing technology (NGS) using commercial panel ClearSeq Inherited DiseaseXT (Agilent Technologies) and consequent validation by Sanger sequencing. Our analysis detected two distinct de novo pathogenic sequence variants in the SCN2A gene, resulting in p.Met1545Val and p.Ala263Val changes on the SCN2A protein level and explaining their pathological phenotypes.

In conclusion, our findings indicate that NGS can become the optimal approach to the genetics diagnostics of disorders of central nervous system (CNS) which can lead to the discovery of new candidate genes.

Keywords: Intellectual disability; Epileptic encephalopathy; Ohtahara syndrome; array-CGH; NGS; SCN2A gene

Introduction

Epileptic encephalopathies (EE) represent a group of epileptic conditions characterized by an abnormal cerebral electrical activity which result in progressive cerebral dysfunction. Based on the age of onset, eight EE syndromes are defined: early myoclonic encephalopathy and Ohtahara syndrome (neonatal period), West syndrome and Dravet syndrome (infancy), myoclonic status with nonprogressive encephalopathies, Lennox-Gastaut syndrome, Landau-Kleffner syndrome and epilepsy with continuous spike waves during slow wave sleep (childhood and adolescent period) [1,2]. The most severe forms of them are characterized by extreme abnormalities in electroencephalographic (EEG) pattern which cause global developmental delay and intellectual disabilities due to irreversible and fatal brain damage.

Early infantile epileptic encephalopathy (EIEE)/Ohtahara syndrome (OS) represents the earliest form of neonatal EEs, because symptoms appear within the three months after delivery, often within 10 days of life [2,3]. Infants acutely develop tonic spasms with variable progression, localization, frequency per day, and with independence of the sleep cycle. OS typically result in severe structural brain malformations and is often accompanied by variety of metabolic disorders which may modify clinical features and their severity. The broad phenotypic heterogeneity of epileptic encephalopathies reflects a broad genetic heterogeneity with monogenic and polygenic forms of EE and all types of inheritance (autosomal dominant, autosomal recessive and X-linked) [4]. Whole-genome and whole-exome sequencing techniques have led to the identification of pathogenic sequence variants in a large number of genes which are essential for normal brain development and functioning. These genes play an important role in the formation of functional synapses, signal generation and its transmission.

In our work, we present two cases of unrelated children with epileptic encephalopathies resulting in severe intellectual disability and multiple congenital abnormalities. They were examined by the multi-step genetic diagnostic algorithm including G-banded karyotype and-CGH, however, we did not identify any pathogenic copy-number variants (CNVs) explaining their pathological phenotype. Additionally, these children and their unaffected parents were involved in our pilot study using targeted NGS with a commercially available panel of 2742 gene known in pathogenesis of human inherited diseases. Here, we present the findings after applying this panel to our probands.

Material and Methods

Two families with affected children were diagnosed at the Department of Medical Genetics (University Hospital Brno, Czech Republic) in years 2015 and 2016. Before the procedure of any genetic analyses, parents signed the informed consent. The informed consent (institutional number 4-288/15/2) was approved by the Ethical Committee, University Hospital Brno. The peripheral blood samples were collected in sterile heparinized tube for cytogenetic analyses. Genomic DNA samples were obtained from peripheral blood in EDTA according to the standard DNA isolation process using MagNaPure system (Roche s.r.o, Basel, Switzerland). We evaluated their quality and quantity by DNA agarose electrophoresis with TrackIt™ 1 Kb Plus DNA Ladder (Thermo Fisher Scientific, Thermo Scientific, Waltham, MA, USA). NanoDrop® ND-1000 (Thermo Fisher Scientific) and Qubit® 1.0 (Thermo Fisher Scientific).

Received November 22, 2017; Accepted December 13, 2017; Published December 15, 2017


Copyright: © 2017 Wayhelova M, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
G-banded karyotype analysis and array-CGH technique

Cytogenetic analysis of karyotype was accomplished using standard G-banding procedure and twenty mitoses per patient were evaluated. After G-banded karyotype analyses with negative results (46,XY in patient 1 and 46,XX in patient 2, respectively) we performed whole-genomic screening chromosomal rearrangements by array-CGH using a SurePrint G3 CGH Microarray 4x180K (Agilent Technologies, Santa Clara, CA, USA) according to the manufacturer's recommendations. Samples were matched with Human Genomic DNA reference (Promega, Madison, WI, USA). Array-CGH slides were scanned with Agilent Microarray Scanner. Data obtained using Agilent Feature Extraction software 12.0.2.2 and Agilent Cytogenomics 4.0, respectively, and visualized using Agilent Genomic Workbench Software 7.0.4.0. CNVs were detected using ADM-2 algorithm with filters of at least five neighbouring probes in region, minimal size of 100 kb in region and minimal absolute average log ratio of 0.25 as cut-off. All genomic positions were estimated on the human reference sequence GRCh37/hg19.

Some additional analyses and characterization of detected CNVs in patients and their parents were required to confirm their origin and localization and to specify their impact on the patient's phenotype. For these purposes we used targeted method of relative quantitative PCR (qPCR).

Relative quantitative PCR

Relative qPCR was designed and performed with two pairs of DNA primers (targeted to CNVs region and out of the CNVs region, respectively). The amount of 10 ng diluted genomic DNA samples (7.5 µl), specific DNA primers (5 µl of forward and reverse 50µM primers) and 12.5 µl reaction mix Power SYBR Green® PCR Master mix (Thermo Fisher Scientific) were prepared to the final reaction volume of 25 µl. Relative quantitation of target DNA regions (performed in duplicates) were run on StepOne using StepOne Software v2.3 (Thermo Fisher Scientific) following a recommended thermal protocol (95°C for 10 min followed by 40 cycles with 95°C for 15 sec, 60°C for 60 sec, and including final melt curve analysis up to 95°C (mode Step and Hold). Ct values were subtracted for the ERH gene (housekeeping gene) and for each tested DNA region to calculate 2-ΔΔCT [5,6]. After the 2-ΔΔCT calculation to get R-values we were able to confirm or to exclude the known variants. Coverage of targeted regions was explored using bed nome AnalysisTK and dbSNP (v147) variants were used as a set of known variants. Coverage of targeted regions was explored using bed tools (v2.23.0) [18]. Additional quality checks and statistics were obtained using Picard.

Raw variant calls were performed using VarScan2 (v2.4.2) [19] with default settings except: I) Minimal variant frequency 0.2. II) Variants with p-value 0.05. dbSNP ID was added if a match between dbSNP and raw variants was found using snpSift (v4.2) [20]. Filtering of the raw variants was done using VarScan2 with default settings except minimal p-value which was set to 0.05. SNPs in very close proximity to indels were discarded. Preprocessed sequencing reads were mapped to a reference genome (hg19, UCSC Genome FTP http://hgdownload.cse.ucsc.edu/goldenPath/hg19/chromosomes/) [10,11] by BWA aln/ sampe (v0.7.15) [12]. Alignments were further processed with Stampy (v1.0.29) [13]. Aligned bam position and mate information sorted files was corrected using Same tools (v1.3) [14]. Since each library was sequenced in two separate sequencing runs, bam files were merged into one using Picard (v2.1.0) [15] for each sample. PCR duplicates were marked and removed using Picard. Duplicates clean bam files were indel realigned using Genome Analysis TK (v3.6) [16]. Base Quality Score Recalibration (BQSR) was performed in two steps using Genome AnalysisTK and dbSNP (v147) [17] variants were used as a set of known variants. Coverage of targeted regions was explored using bed tools (v2.23.0) [18]. Additional quality checks and statistics were obtained using Picard.

Bioinformatic NGS data processing

Quality of raw sequencing files was checked using FastQC (v0.11.5) [7]. Presence of adapters was scanned using minion and swan (Kraken package, v15-065) [8]. Preprocessing of the raw sequencing files was performed using cutadapt (v1.11) [9] in following steps: I) Very low-quality ends were trimmed (Phred<5), II) Adapters from both reads of a pair were removed with minimal overlap of 3 bp and maximum of 10 mismatchs in a matched sequence (removed adapters: R1- AGATCGGAAGAGCACACGTCT- GAACTCAGTCAC, R2 - AGATCGGAAGAGCGTCGTGTAGGAAAGAGTGTAGAT- CCGTGTCGCCGTATCATCATT), III). Very short (<35 bp) and unpaired reads after the trimming were discarded. Preprocessed sequencing reads were mapped to a reference genome (hg19, UCSC Genome FTP http://hgdownload.cse.ucsc.edu/goldenPath/hg19/chromosomes/) [10,11] by BWA aln/ sampe (v0.7.15) [12]. Alignments were further processed with Stampy (v1.0.29) [13]. Aligned bam position and mate information sorted files was corrected using Same tools (v1.3) [14]. Since each library was sequenced in two separate sequencing runs, bam files were merged into one using Picard (v2.1.0) [15] for each sample. PCR duplicates were marked and removed using Picard. Duplicates clean bam files were indel realigned using Genome Analysis TK (v3.6) [16]. Base Quality Score Recalibration (BQSR) was performed in two steps using Genome AnalysisTK and dbSNP (v147) [17] variants were used as a set of known variants. Coverage of targeted regions was explored using bed tools (v2.23.0) [18]. Additional quality checks and statistics were obtained using Picard.

Raw variant calls were performed using VarScan2 (v2.4.2) [19] with default settings except: I) Minimal variant frequency 0.2. II) Variants with p-value 0.05. dbSNP ID was added if a match between dbSNP and raw variants was found using snpSift (v4.2) [20]. Filtering of the raw variants was done using VarScan2 with default settings except minimal p-value which was set to 0.05. SNPs in very close proximity to indels were removed from the calls. Effect of the variants, ClinVar (downloaded

Table 1: The overview of clinical data and genomic analyses in patient 1 and patient 2. The patients were examined through multi-step diagnostic algorithm including clinical genetic evaluation and genetic counselling.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Genomic analyses</th>
<th>Clinical data</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-banded karyotype array-CGH and relative qPCR</td>
<td>Targeted NGS</td>
<td>Sanger sequencing</td>
</tr>
<tr>
<td>Patient 1</td>
<td>46,XY arr[GRCh37] 1q21.32(57291428_57730158)x3 de novo origin</td>
<td>SCN2A gene NC_000002.11: g.166166923C&gt;T (p.Ala263Val) de novo origin</td>
</tr>
<tr>
<td>Patient 2</td>
<td>46,XX arr[GRCh37] 1p36.32(2694731_3202657)x3 maternally inherited</td>
<td>SCN2A gene NC_000002.11: g.166243337A&gt;G (p.Met154Val) de novo origin</td>
</tr>
</tbody>
</table>

J Neurol Disord, an open access journal
ISSN: 2329-6895

Volume 5 • Issue 6 • 1000369

Table 2: The verification of CNVs using R-values calculation based on relative qPCR results. R-values were calculated for both the regions of CNVs and control regions in families 1 and 2. Threshold R-values: <0.7 means loss of genetic material in targeted DNA region, >1.3 means gain of genetic material in targeted DNA region.

<table>
<thead>
<tr>
<th>Variables</th>
<th>R-values</th>
<th>Duplication 18q21.32</th>
<th>Control region 18q12.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient 1</td>
<td>1.46</td>
<td>1.13</td>
<td></td>
</tr>
<tr>
<td>Mother</td>
<td>0.88</td>
<td>1.08</td>
<td></td>
</tr>
<tr>
<td>Father</td>
<td>1.00</td>
<td>1.16</td>
<td></td>
</tr>
<tr>
<td>Family 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient 2</td>
<td>1.48</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>Mother</td>
<td>1.60</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td>Father</td>
<td>1.05</td>
<td>1.03</td>
<td></td>
</tr>
</tbody>
</table>

Patient 2 is a girl (*2014) born from the 1st gravidity to unrelated parents. She was delivered 5 days post-term, 3400 g/50 cm. After the delivery she suffered from tachycardia, hypotonia and myoclonia. After magnetic resonance imaging and EEG, she was diagnosed with early infantile myoclonic encephalopathy with poor response to anti-epileptic therapy. Suspected metabolic disorders, myotonic dystrophy type 1 and TORCH syndrome were excluded. She was diagnosed severe psychomotor retardation with epilepsy, axial hypotonia and limb spasticity. She has apparent facial stigmatization (narrow forehead, long palpebral fissures, low-set ears, carp mouth) and abnormal dermatoglyphic pattern of left palm. She is fed through percutaneous endoscopic gastrostomy (PEG) feeding tube.

Whole-genomic analyses and their verification

Both patients showed a normal karyotype by G-banding. Consequently array-CGH on oligonucleotide DNA microarrays was performed with CNVs findings in both patients. 539-kb duplication of 18q21.32 chromosome: arr[GRCh37] 18q21.32(57291428–57830158) x 2 in patient 1 and 508-kb duplication of 1p36.32 chromosome: arr[GRCh37] 1p36.32(2694731_3202657) x 3 in patient 2 were detected using 4X180K CGH Microarray. The CNVs findings were verified in both patients and their parents using relative quantitative PCR analysis. The presence of 18q21.32 duplication was detected only in patient 1 and it was shown to have de novo origin. We evaluated this CNV as a variant of undetermined significance. The 1p36.32 duplication in patient 2 was shown to be maternally inherited. Therefore, we evaluate it as a likely benign variant. The R-values for both the CNVs regions and regions outside the CNVs in our parents and their parents are presented in Table 2.

Due to the unexplained severe pathological phenotypes which were highly suspected to have genetic causes our patients were involved to the pilot project using targeted NGS approach with capture design ClearSeq Inherited Disease. We preferentially focused on exonic variants and we filtered only non-synonymous exonic variants (SNVs and insertions/deletions) unique for the given patient. These variants were analysed in detail including the search in ClinVar [21], OMIM [27,28], dbSNP [17] and UniProtKB/Swiss-Prot [29] variant pages. Finally, our findings were correlated to patient's phenotype and were verified by independent analysis of Sanger sequencing in patients and their unaffected parents.

In patient 1 we identified total of 17 012 DNA sequence variants, when 16 120 of them (94.76%) with known reference single-nucleotide polymorphism identifier (rsID). As a result of multi-step analytical process we identified a heterozygous NC_000002.11:g.166166923>C>T transition affecting the SCN2A gene resulting in a p.Ala263Val substitution (Figure 1). This variant has been previously detected in patients with neonatal seizures and EIEE/OS [30,31]. To support our assumption of its pathogenicity we performed confirmative analysis by targeted Sanger sequencing in patient 1 and his parents. We proved the heterozygosity for p.Ala263Val substitution in patient 1 and confirm its de novo origin as presented in the Figure 2.
In patient 2 we identified a total of 18,277 DNA sequence variants and 17,436 of them (95.40%) were classified with known rsID. After NGS data analysis and variant filtering, we detected a heterozygous NC_000002.11:g.166243337A>G transition affecting the SCN2A gene resulting in a p.Met1545Val substitution (Figure 3). To our knowledge, less than ten cases of this likely pathogenic variant have been reported so far [32,33]. Consequently, we performed targeted Sanger sequencing in patient 2 and her parents. We confirmed de novo origin and heterozygous state for p.Met1545Val in patient 2, as shown in Figure 4. This result gives a strong evidence for its pathogenic or likely pathogenic effect, respectively.

Discussion

We present two cases of children with epileptic encephalopathies. They have been examined through multi-step diagnostic procedure including G-banded karyotype and array-CGH with negative results. Consequently, they were examined by targeted NGS using commercial panel ClearSeq Inherited DiseaseXT including 2742 genes associated to human congenital diseases, followed by verification by Sanger sequencing in their parents. Here, we focused on a detailed characterization and clinical consequences of two unique pathogenic sequence variants in the SCN2A gene, p.Ala263Val in patient 1 and p.Met1545Val in patient 2.

The SCN2A gene is located at 2q24.3 chromosome region and encodes a sodium voltage-gated channel alpha subunit 2, which belongs to nine-member sodium channel alpha gene family [34]. Voltage-gated sodium channels are transmembrane glycoprotein complexes which function in generation and propagation of action potentials in neurons and muscles. They comprise of one pore-forming alpha subunit which may be associated with either one or two beta subunits [35]. The SCN2A protein comprises of four internal homology repeats, each of them contains eight potential transmembrane segments, and multiple glycosylation and phosphorylation sites [36]. The SCN2A gene is predominantly expressed in brain tissue, especially in neurons and glia. The encoded SCN2A protein is distributed along the surface of granule cells in cerebellum and it is suggested its primary involvement in the transmission of information to Purkinje cells [37]. It plays a fundamental role in controlling electrical excitability during CNS development and plasticity [38]. Recent studies highlighted its important role in psychiatric diagnoses and neurologic diseases through the identification of multiple SCN2A pathogenic sequence SNVs. Predominantly, they were identified and confirmed as one of...
important causes in the wide spectrum of epileptic encephalopathies exhibiting a highly variable clinical picture [39,40].

SCN2A-related disorders represent a group of clinically variable pathological phenotypes. Allelic variants in the SCN2A gene have been identified as pathogenic in many children with severe clinical condition defined as early infantile epileptic encephalopathy type (EIEE) and in children with less severe benign familial neonatal infantile seizures (BFNIS) [41]. EIEE/OS belongs to a group of genetically and clinically distinct epileptic encephalopathies. Approximately 75% of EIEE may progress to West syndrome, a diagnosis with an adverse prognosis including the arrest of psychomotor development and abnormal EEG pattern negatively affecting brain functions [42]. Wolff et al. defined two groups of patients with distinct seizure onset (before or after three months of age). These patients exhibit variable additional phenotypic features related to epilepsy, different impact of pathogenic variants in the SCN2A gene to encoded protein resulting in different response to treatment by Na+ channel blockers [43].

The variant p.Ala263Val has been already evaluated as pathogenic in a large amount of patients with EIEE/OS [44]. The amino acid alanine in the position 263 is located in the hydrophobic S5 segment of transmembrane domain I of the SCN2A protein. Functional studies on cell culture experiments have indicated that the alteration in ion channel causes gain-of-function neuronal hyperexcitability with a 3-fold increase in persistent Na+ current [30]. Schwarz et al. presented a cohort of patients with distinct SCN2A pathogenic sequence variants to observe common clinical characteristics [45]. The seizures recurred more frequently, and they persisted longer than in less severe BFNIS. There were presented also cases of patients with the same variant p.Ala263Val that was detected in our patient 1. The patients were diagnosed an apparent hypotonia and sleeping difficulties. They exhibited neonatal-onset complex episodic neurologic phenotype, including ataxia and myoclonia. The Na+ channel blocker treatment may reduce the negative symptoms of EIEE/OS and improve the prognosis of significant proportion of affected children. Wolff et al. (2017) reviewed a large cohort of 201 patients with SCN2A-related disorders including 71 patients not reported previously [43]. They reported six cases of recurrent p.Ala263Val variant. Three of them were diagnosed BFNIS with late onset episodic ataxia while three others exhibited more severe phenotypes (epileptic encephalopathy/Ohtahara syndrome) with adverse prognosis, as presented in our patient 1.

In contrast to p.Ala263Val variant, the p.Met1545Val variant in the SCN2A gene has been detected in less than ten patients with EIEE so far. The amino acid methionine in the position 1545 is located in the
hydrophobic S1 segment of transmembrane domain IV of the SCN2A protein. However, its alterations and their functional impact have not been estimated so far. In 2016, Dilena et al. presented a case of newborn male with a manifestation of epileptic encephalopathy, characterized by abnormal EEG pattern and tonic-clonic migrating partial seizures [33]. The neurological examination showed hypotonia, poor movement and feeding difficulties as were noticed in our patient 2. The condition was poorly responding to a wide range of anti-epileptic drugs, but Na+ channel blocker carbamazepine therapy as prevention to seizure relapses had shown to be effective. Due the fact the developmental delay is notable; it probably indicates a significant contribution of epilepsies and seizures to abnormal brain functional development. Gokben et al. presented a study of early-onset epileptic encephalopathies using targeted NGS panel including sixteen genes [46]. They detected twelve pathogenic or likely pathogenic sequence variants and defined de novo heterozygous p.Met1545Val in the SCN2A gene as a novel pathogenic variant with a causative effect in patients with EIEE/OS.

As described in clinical data, both our patients share key features including epileptic encephalopathy with poor response on first-choice anti-epileptic therapy associated with severe psychomotor delay and hypotonia, however some apparent features are specific for each of them (the absence of apparent facial stigmatization and feeding difficulties in patient 1). Both the detected SCN2A pathogenic variants (p.Ala263Val in patient 1, p.Met1545Val in patient 2) change the structure of transmembrane domains in the protein, therefore we expect that some variants in other genes and non-coding sequence variants may modify the phenotypes and can explain the existing phenotypic variability between our patients and among patients with epileptic encephalopathies with causative variants in the SCN2A gene.

Conclusion

In our study we present two cases of children with epileptic encephalopathies, intellectual disabilities and multiple congenital abnormalities that were enrolled in our pilot study using targeted NGS as a novel step in diagnostic algorithm at the Department of Medical Genetics (University Hospital, Brno). We detected two de novo distinct pathogenic sequence variant in the SCN2A gene which change the structure of transmembrane domains in the SCN2A protein. Thus, we proved that SCN2A pathogenic sequence variants play an important role in the pathogenesis of EE with adverse prognosis.

In conclusion, our study confirmed that targeted NGS with a commercially available gene panel could represent a suitable alternative to whole-exome or whole-genome sequencing, respectively. Its advantage especially consists in more targeted analysis of disease-causing genes; therefore, we suggest targeted NGS as an additional step after whole-genome screening of CNVs using array-CGH. This combination of two sensitive and precise methods could significantly improve the quality of genetic counselling in affected families. Furthermore, recent findings show that targeted NGS may contribute to the discovery of rare pathogenic sequence variants in the molecular genetic diagnostics of CNS disorders.

Acknowledgments

This work was Supported by Ministry of Health, Czech Republic - conceptual development of research organization (FNBr, 6529705). We would also like to thank patients and their parents for their enrolment in our pilot study. We acknowledge the CF New Generation Sequencing Bioinformatics supported by the CIISB research infrastructure (LM2015043 funded by MEYS CR) for their support with obtaining scientific data presented in this paper. Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the programme “Projects of Large Research, Development, and Innovations Infrastructures” (CESNET LM2015042), is greatly appreciated. We declare no conflicts of interest.

References

24. R-Development Core Team (2011) R: A language and environment for statistical