

ARTICLE

Improving molecular diagnosis in epilepsy by a dedicated high-throughput sequencing platform

Erika Della Mina¹, Roberto Ciccone¹, Francesca Brustia², Baran Bayindir¹, Ivan Limongelli², Annalisa Vetro¹, Maria Iacone³, Laura Pezzoli³, Riccardo Bellazzi^{4,5}, Gianfranco Perotti⁶, Valentina De Giorgis², Simona Lunghi², Giangennaro Coppola⁷, Simona Orcesi², Pietro Merli⁶, Salvatore Savasta⁶, Pierangelo Veggiotti² and Orsetta Zuffardi^{*,1,2}

We analyzed by next-generation sequencing (NGS) 67 epilepsy genes in 19 patients with different types of either isolated or syndromic epileptic disorders and in 15 controls to investigate whether a quick and cheap molecular diagnosis could be provided. The average number of nonsynonymous and splice site mutations per subject was similar in the two cohorts indicating that, even with relatively small targeted platforms, finding the disease gene is not an univocal process. Our diagnostic yield was 47% with nine cases in which we identified a very likely causative mutation. In most of them no interpretation would have been possible in absence of detailed phenotype and familial information. Seven out of 19 patients had a phenotype suggesting the involvement of a specific gene. Disease-causing mutations were found in six of these cases. Among the remaining patients, we could find a probably causative mutation only in three. None of the genes affected in the latter cases had been suspected *a priori*. Our protocol requires 8–10 weeks including the investigation of the parents with a cost per patient comparable to sequencing of 1–2 medium-to-large-sized genes by conventional techniques. The platform we used, although providing much less information than whole-exome or whole-genome sequencing, has the advantage that can also be run on ‘benchtop’ sequencers combining rapid turnaround times with higher manageability.

European Journal of Human Genetics (2015) 23, 354–362; doi:10.1038/ejhg.2014.92; published online 21 May 2014

INTRODUCTION

Epilepsy is one of the most common neurological disorders in humans with a prevalence of 1% and a lifetime incidence of up to 3%.¹ Epilepsies present with a broad range of clinical features and their genetic causes remain unknown in the vast majority of cases, although several genes have been identified in rare Mendelian forms, either heritable or sporadic. Finding the disease genes can be challenging, as the same epileptic phenotype may be associated with several genes. A molecular diagnosis of epilepsy is important especially in a pediatric setting in order to (1) establish the recurrence risk in following pregnancies, (2) stop the diagnostic odyssey that is frequently restless for undiagnosed epilepsies, and (3) provide, at least in some cases, specific therapies. Recently, genome-wide association studies revealed a few regions harboring high-ranking candidate genes, although these studies still necessitate further replication efforts.² Genomic arrays had been more successfully, as they allowed to identify several possible pathogenic copy-number variants not present in controls in about 9% of the cases.³ Presently, high-throughput sequencing is becoming the most promising approach to improve molecular diagnosis of this condition,^{4,5} although the interpretation of the results is far from being a standardized process.⁶ Indeed, next-generation sequencing (NGS) does not magically make diagnoses but typically provides a handful of possibilities requiring further studies on the function of each candidate gene. To overcome these problems, we composed a

panel containing most epilepsy genes, covering several relevant phenotypes. With this NGS platform, we studied 19 index patients suffering from a range of seizures, either familial or sporadic. Although initially we performed a blind study trying to interpret the sequencing data without any knowledge of the clinical history, we then realized that no analysis was possible in absence of detailed phenotypes and familial information. This study allowed us to evaluate not only the diagnostic capability of this approach but also the cost and the time required to report the final result to the family.

MATERIALS AND METHODS

Patient cohort

The 19 index cases ranged from few days to 4 years of age at the time of the first clinical examination. Most of them have been then followed-up for several years. This cohort has been randomly selected from patients afferent to our epileptic center for children and adolescents. All available family members have been enrolled for segregation analysis. Informed consent was obtained from each family and clinical evaluation and genetic testing were carried out in accordance with the ethics approval granted (11017C-RC2011 IRCCS C. Mondino, Diagnosis and therapy of epileptic syndromes).

In Table 1, a brief report of each case is provided including clinical diagnosis and history, and presence of the epilepsy in relatives (family history), electroencephalograms, magnetic resonance findings, and administered anti-epileptic drugs. In order to assess the diagnostic capability of our approach, we collected patients presenting with a wide range of epilepsy phenotypes: 11 are sporadic cases, whereas the remaining eight have a history of familial epilepsy.

¹Department Molecular Medicine, University of Pavia, Pavia, Italy; ²National Neurological Institute C. Mondino, Pavia, Italy; ³USSD Medical Genetics Laboratory of Ospedale Riuniti, Bergamo, Italy; ⁴Department of Industrial and Information Engineering, University of Pavia, Pavia, Italy; ⁵IRCCS Foundation Salvatore Maugeri, Pavia, Italy; ⁶IRCCS Foundation Policlinico San Matteo, Pavia, Italy; ⁷Child and Adolescent Neuropsychiatry, Medical School, University of Salerno, Salerno, Italy

*Correspondence: Professor O Zuffardi, Department of Molecular Medicine, University of Pavia, Via Forlanini 14, 27100 Pavia, Italy. Tel: +39 0382 987 733; Fax: +39 0382 525 030; E-mail: zuffardi@unipv.it

Received 5 December 2012; revised 17 March 2014; accepted 4 April 2014; published online 21 May 2014

Table 1 Patients' clinical data, subdivided in two subcohorts (A and B), including sex, age, AEDs administered, EEG, MRI results, and family history

Case <i>n</i> -subcohort	Sex	Age (years)	Diagnosis	Clinical features	AEDs ^a	EEG	MRI	Family history
1-A	F	17	Lennox-Gastaut syndrome in a subject with double cortex	Severe mental retardation. Tonic, atonic, and myoclonic seizures during sleep and awake	VPA + CLB + ETS	Lennox like	Double cortex	Questionable parental consanguinity
2-A	F	14	Pyridoxine-dependent epilepsy	Neonatal seizures and epileptic status at 6 months of age responsive to pyridoxine. Normal IQ	LTG + B6	Currently normal	Normal	Negative
3-A	M	14	Pyridoxine-dependent epilepsy	Intractable seizures during the first year and epileptic status pyridoxine responsive. Normal IQ	B6	Currently normal	Normal	Progressive motor and language disability (mother's paternal uncle)
4-A	M	4	Dravet syndrome	Typical evolution of Dravet syndrome. Actually IQ 46	VPA + TPM + CLB + STP	Typical of Dravet syndrome	Normal	Negative
5-A	F	0.8	Benign familial neonatal seizures	Intractable tonic-clonic seizures during the first 25 days of life spontaneously remised	PB + TPM + MDL + Rivotril	Normal	Normal	Father with benign neonatal seizures
6-A	M	7	Epileptic encephalopathy and pervasive developmental disorder	Neonatal seizures, followed by focal seizures. Severe mental retardation. Pervasive developmental disorder	VPA	Sharp and sharp-wave on central and frontal regions	Normal	Negative
7-A	F	7	Febrile seizures plus	Febrile and afebrile GTC. Normal IQ	LEV	Large amplitude and diffuse sharp-wave abnormalities	Normal	Migraine in the maternal line
8-B(i)	M	6	Myoclonic epilepsy	Epileptic negative myoclonus. Normal IQ	VPA + LTG	Spike-wave discharges	Not performed	Brother of 8-B(ii)
8-B(ii)	M	14	Generalized epilepsy	One myoclonic seizure. Normal IQ.	LTG	Spike-slow wave discharges	Normal	Brother of 8-B(i)
9-B	F	3	Febrile seizures	Numerous febrile seizures during the first year.	VPA	Normal	Not performed	Family history of febrile seizures
10-B	M	6	Benign epilepsy of childhood with centrotemporal spikes (BECTS)	Normal psychomotor development (GQ = 11.6) Few typical rolandic seizures. Borderline learning difficulties	None	Atypical BECTS	Normal	Family history of BECTS and mild intellectual disabilities
11-B	M	8	Myoclonic epilepsy	Cognitive impairment from 4 years (GQ 90.4; IQ 53)	VPA + LTG	Currently normal	Not performed	Mother, maternal uncle, and grandfather with generalized epilepsy
12-B	M	16	Febrile seizures plus	Febrile seizures from 2 to 6 years. Normal IQ	VPA, suspended in 2006	Currently normal	Not performed	Family history of febrile afebrile seizures
13-B	F	5	Febrile seizures	Febrile seizures, normal IQ	None	Normal	Not performed	Family history of febrile seizures (father) and FS and absences (mother)
14-B	M	12	Febrile seizures plus	Febrile seizures, then afebrile GTC. Normal IQ	VPA	Generalized abnormalities	Hippocampal asymmetry, incomplete rotation of hippocampal structure	Family history febrile and afebrile seizures (father and sister)
15-B	M	9	Lennox-Gastaut syndrome	Severe cognitive impairment, myoclonic and tonic seizures or GTC Mental retardation	VPA + ETS + NTZ + HC + PB	Spikes and polyspikes tonic activity, slow waves during sleep and awake	Normal	Not reported
16-B	M	3	Epileptic encephalopathy	GTC and partial seizures. Severe cognitive impairment	LTG + VPA + HC	Generalized sharps and sharp waves. Poor organization.	Normal	Consanguineous parents (first cousins)
17-B	M	4	Focal epilepsy	Febrile seizures, status epilepticus. Normal psychomotor development. QS 90 at 2 years	VPA + TPM	Poor organization, slow, not epileptic abnormalities	Mesial temporal sclerosis and temporal lobe dysplasia left	Not reported
18-B	M	4	Epileptic encephalopathy	Neonatal seizures, spasms, and hypsarrhythmia Severe cognitive impairment, hypotonia, strabismus, and motor stereotypies	VPA + LEV + NTZ	Poor organization, slow, slow waves, and polyspikes during sleep	Cerebral atrophy	Not reported
19-B	M	5	Epileptic encephalopathy and hypothyroidism	Severe cognitive impairment, spasms in flexion, and partial seizures, Hypothyroidism	VPA + TPM + LTG	Poor organization. Multifocal abnormalities	Modest diffuse cerebral atrophy	Not reported

Abbreviations: AEDs, antiepileptic drugs; EEG, electroencephalogram; F, female; M, male; MRI, magnetic resonance imaging.

^aAEDs abbreviations: B6, pyridoxine; CBZ, carbamazepine; CLB, clobazam; ETS, ethosuximide; HC, hydrocortisone; LEV, levetiracetam; LTG, lamotrigine; MDL, midazolam; NTZ, nitrazepam; PB, phenobarbital; STP, stiripentol; TPM, topiramate; VPA, valproic acid.

Patients have been subdivided in two groups (Table 1): the first one, subcohort A, was constituted by seven patients whose clinical features were either strongly or more loosely suggestive for a syndrome associated with a specific gene; the second group, subcohort B, included 12 subjects with very different types of epilepsy, presumably heterogeneous in their genetic basis and not suggestive of any or a single specific gene. In both subgroups other clinical features such as language impairment, psychomotor delay, or autism spectrum disorder were present in several patients. Magnetic resonance abnormalities were detected in some of them (Table 1). All cases had been previously analyzed by array comparative genomic hybridization and a few (6-A, 8-B(i), 8-B(ii)) by Sanger Sequencing for specific genes without any positive result.

Control cohort

Control subjects (nine females and six males), ranging from 18 to 35 years of age, were recruited among blood donors as controls for both this and a cardiovascular study. Besides, they were requested to answer a structured general medical questionnaire with specific emphasis on neurological and cardiac symptoms, control subjects had to answer two specific questions: (1) have you ever suffered from epilepsy or seizures, and (2) have you become acquainted with any seizure disorders or EEG abnormalities present in some of your family members? Only those who answered negatively were recruited.

Platform design

A custom-designed target enrichment library for 67 genes (see Supplementary Information) has been designed by using the Agilent eArray website (<https://earray.chem.agilent.com/earray/>). This library contains unique baits covering the exons, the UTRs, and the intron–exon junctions of the selected genes. The estimated base coverage of the library is 0.45 Mb (Supplementary Table 1S).

The selection was made on the basis of the following criteria: (1) genes associated with idiopathic epilepsy; (2) genes associated with syndromic epilepsy; (3) genes associated with epilepsy and cerebral malformations excluding holoprosencephaly; (4) genes that appeared to be the best candidates for epilepsy in microdeletion syndromes. Selected genes are reported in Supplementary Table 1.

Sample preparation

DNA (5 µg) extracted from peripheral blood by standard methods were diluted in 700 µl of nebulization buffer (Illumina, San Diego, CA, USA) and sheared using a nebulization technique (Invitrogen, Carlsbad, CA, USA), which breaks up DNA into pieces <500 bp, through the application of 60–70 psi (pound force per square inch) of purified air for 4 min. This process generates double-stranded DNA fragments containing 3' or 5' overhangs that were cleaned up using QIAquick spin columns (Qiagen, Hilden, Germany). A quality control step on the recovered DNA was then performed using Nanodrop 1000 to quantify the DNA by a 260-nm reading and Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA) to check the size of the fragments.

According to the Agilent SureSelectXT protocol, sheared DNA overhangs were end-repaired and then purified using the magnetic bead-based Agencourt AMPure XP purification system (Beckman Coulter Genomics, Brea, CA, USA). Then we performed the adding of 'A' bases to the 3' end of the DNA fragments and the ligation of indexing-specific paired-end adapters.

After a few cycles of PCR amplification, 500 ng of DNA from the resulting libraries were hybridized to the bait set using the SureSelectXT MP Capture Library Kit (Design no. 5190-0312931—Agilent) at 65 °C for 24 h. Hybrids capture was performed according to the manufacturer's protocol with Streptavidin-coated Dynal magnetic beads (Invitrogen). Captured samples were further purified through Agencourt AMPure XP beads and subjected to a PCR-based amplification reaction to add index tags (each is a sequence of six bases in length allowing to identify samples after pooling), according to the Agilent SureSelectXT protocol. For each step of library preparation, all samples were quantified on a Bioanalyzer 2100 (Agilent). We performed a multiplexed run on the Illumina Genome Analyzer IIx, where nine multiple samples were sequenced in a single lane of a flow cell; number of samples to be pooled has been calculated on the basis of the enriched target's size, according to Agilent's instructions. The sample libraries from nine individuals were denatured with NaOH and loaded on a single lane of a Illumina Flowcell v4 where DNA

clusters were generated through a one-step workflow (according to Illumina protocol) on the Cluster Station using TruSeq PE Cluster Kit v5 (Illumina).

One percent volume of a PhiX control library (Illumina) was used as internal control and loaded in each lane of the flowcell.

The capture was considered successful if at least 99% of our target regions were covered by more than eight reads of high quality (that is a Phred-scaled mapping quality score of at least 20 for each).

Annotation and interpretation of data

Sequences for each sample were generated by Illumina software CASAVA v1.8.1. Reads were filtered by quality relying on the standard Illumina quality filter test. Reads were aligned to the most recent version of human genome (GRCh37/hg19) using BWA software package v0.6.0⁷ and filtered out for PCR duplicates by Samtools v0.1.18.⁸

Reads were realigned around inferred indels and their base qualities were recalibrated taking into account the context of alignment by the Genome Analysis Toolkit (GATK) v1.6 suite.⁹

SNPs and short indels were called using GATK UnifiedGenotyper module and the resulting variants were filtered using GATK Variant Filtration module and specific Perl scripts, as such variants were probably owing to alignment errors and, in general, they cannot be considered reliable variants. Several filtering constraints were also applied, such as minimum variant quality (50, Phred scaled), a minimum of five reads supporting variant, number of ambiguous mapped reads overlapping variant, neighborhood of each reliable indel or homopolymer excluding those single-nucleotide variations that overlap within.

Variants annotation was performed on the resulting data set by in-house genomic database application. Prediction tracks for each annotated variant were generated by automatic remote calling procedures to Mutation Taster¹⁰ and Polyphen2 (version 2.2.2).¹¹

In order to identify potential causative mutations, we applied the so-called discrete filtering approach.¹² We first excluded synonymous out of target and UTR-overlapping variants. We then excluded the variants present in dbSNP135 and Exome Sequencing Project Databases (ESP) with a frequency higher than 1%. Moreover, we discarded variants reported in our in-house database (66 whole exomes) that were identified in at least two individuals without epilepsy or other neurological disorders.

We then took into account only variants predicted to alter the protein structure or function by at least one of the three prediction tools we used (Mutation Taster, SIFT, Polyphen2) as well as variants for which all prediction tools failed (see Supplementary Table 2S). At the end, we excluded all variants occurring in at least three cases and/or at least two subjects of the control cohort.

We prioritized the candidate alterations on the base of the expression and function of the altered gene, the type of mutation and its effect at protein level, presence of the variant in the Human Gene Mutation Database 2012.3 (HGMD) or in the literature, the metabolic pathway involved, and obviously the clinical features of the patient/family. A manual inspection of the variants eliminated by the prediction tools filtering step allowed us to reconsider them on the base of possible correlations with the patients' phenotype. For example, this permitted reconsidering a variant of *ALDH7A1*, which was then ascertained as causative (see Results, case 2). The entire protocol of data analyses is illustrated in the flowchart reported in Figure 1.

The final subset of mutations was confirmed by Sanger sequencing followed by segregation analysis in each family. Only a close collaboration with the specialist allowed us to find a specific genotype–phenotype correlation discarding those variants that were either pathogenic in recessive state in families where the condition segregated in a dominant manner or those that correlated to a neurological phenotype totally different from that of the patient. This point was taken with extreme caution as it could not be excluded, *a priori*, that a novel mutation might cause a totally different phenotype with respect to the ones known to be associated with the same gene.¹³ Some of the remaining alterations, even if predicted damaging by at least one tool, have been discarded when they did not segregate with the epileptic phenotype in familial cases (ie, *MAGI2* in case 5-A that was inherited by the healthy mother, whereas *KCNQ2* was considered causative because it was inherited by the affected father).

We applied the same filtering strategies to the control cohort in order to perform a statistical test to assess whether the number of deleterious variants in cases was significantly higher than in controls.

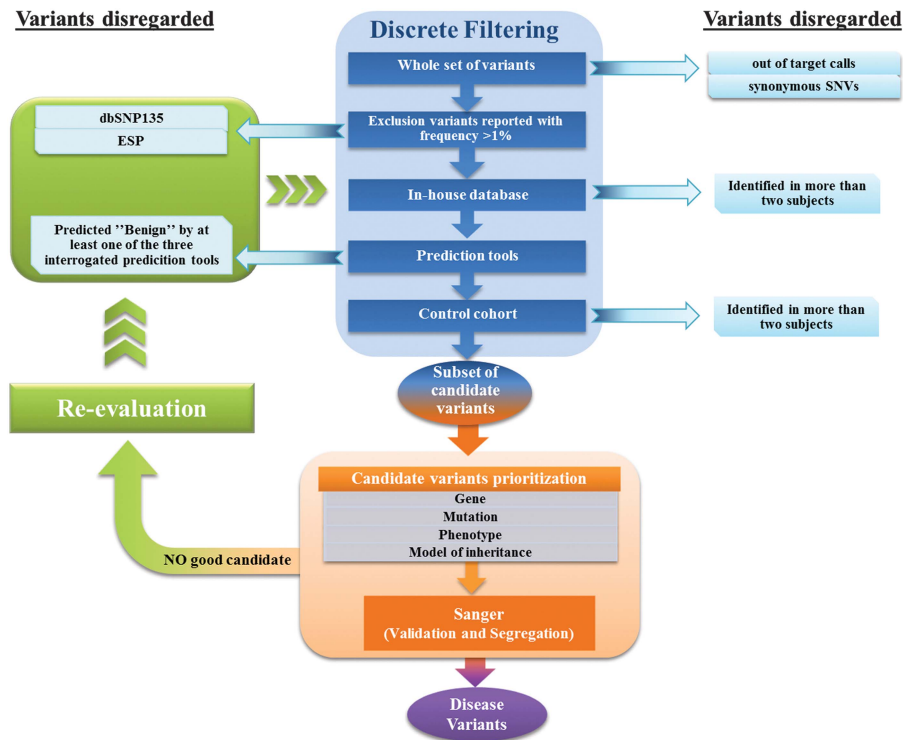


Figure 1 Flow chart representing the strategy adopted to analyze sequencing data. Discrete filtering, prioritization, and re-evaluation steps are highlighted in blue, orange and green, respectively.

RESULTS

Targeted massive parallel sequencing of patient and control cohorts produced for each subject about 180 Mb of sequence, which yielded an average coverage of about $400 \times$ at each targeted base. On average, 96% of target bases exceeded the $15 \times$ coverage threshold required for confident analysis, defined as 99% power to detect a variant.

We compared the number of variants per subject between patients and controls using the non-parametric statistical hypothesis test of WMW (see Supplementary Information). We did not find any significant difference between the two cohorts after filtering (P -value = 0.4928). This might be owing to the limited number of subjects analyzed by this platform. In Table 2, we listed all the variants remaining after the filtering and prioritization processes (ranging from one to three per subject). In bold we highlighted those variants considered as having the main effect on the patient's phenotype.

By this approach we were able to identify candidate SNVs, very likely causative of the epileptic phenotype, in nine out of 19 patients, six of which belong to subcohort A and three belonging to subcohort B (Figure 2). In the remaining 10 cases, we could not find any potential causative alterations even after a careful re-evaluation by manual inspection of variants filtered out. Causative mutations were validated by Sanger sequencing (data not shown). All variants thereafter reported have been submitted to Leiden Open Variation Database 3.0 (Supplementary Table 3S).

Subcohort A

Case 1-A, a female, showed a *de novo* transition in exon 3 of the *DCX* gene (NM_000555.3) leading to a nonsense mutation. The reading frame was interrupted by a premature stop codon possibly leading to nonsense-mediated decay (NMD) of the mRNA as suggested by the prediction tools (Polyphen2, Mutation Taster, and SIFT). Leger *et al*¹⁴ has reported

the same mutation. A paternally inherited *SHANK3* mutation, considered probably damaging by Polyphen2, was also present.

Case 2-A showed two mutations both in *ALDH7A1* gene, one inherited from the father and the other one from the mother. The paternal allele had a transition reported in ESP (MAF 0.0077%) and HGMD (CM087549).¹⁵ The maternal allele had a mutation occurring in intron 16, 5 bp upstream to the previous exon. This mutation was present in ESP (MAF 0.0231%) in heterozygosis and was also reported in HGMD (CS091873).¹⁶ The intronic variant was predicted to alter the splicing by skipping of exon 16.

Case 3-A had a maternally inherited missense mutation at *GPR98* gene that was predicted as damaging by one prediction tool.

Case 4-A, showed a *de novo* splice site mutation in the *SCN1A* gene. A heterozygous missense mutation of *POLG*, was also detected both in the patient and in his normal father. This alteration was given as damaging by prediction tools.

Case 5-A showed a 1-bp deletion, in the *KCNQ2* gene creating a frame shift with a premature stop 15 codons downstream. The mutation was inherited from the father who suffered from the same type of epilepsy. A missense mutation in *MAGI2*, predicted as damaging and inherited from her normal mother was also present.

Case 6-A showed a *de novo* missense substitution in the *KCNQ2* gene. Other two heterozygous missense mutations were detected in *GPR98* and *TBC1D24*, both inherited from the normal father.

Case 7-A had two missense mutations at *SCN1A* and *SCN1B*, both predicted as damaging. *SCN1A* mutation was *de novo*, whereas *SCN1B* alteration was inherited from the father.

Subcohort B

Case 9-B showed a 1-bp exonic duplication in the *GABRG2* gene. This duplication creates a frame shift starting at codon

Table 2 Selected variants after the filtering protocol. In bold are reported those variants that fulfill the criteria for disease-causing mutations: affected genes already associated with patient's phenotype, exhibit complete segregation with the disease, and are absent in healthy controls

Case	UCSC gene	Genotype	Region ^a	Inheritance	gDNA level (GRCh37)	cDNA ^b	Protein	Mutation taster	Polyphen2	SIFT
1-A	SHANK3	het	Exon 8	pat	Chr22: g.51121780C>T	NM_001080420.1:c.898C>T	p.(Arg300Cys)	NP	Probably damaging	
	DCX	het	Exon 3	dn	ChrX: g.110653572G>A	NM_000555.3:c.298C>T	p.(Arg100*)	Disease causing	NP	Probably damaging
2-A	ALDH7A1	het	Exon 6	pat	Chr5: g.125912837T>C	NM_001201377.1:c.500A>G	p.(Asn195Ser)	Disease causing	Probably damaging	Deleterious
	ALDH7A1	het	Intron	mat	Chr5: g.125885616C>T	NM_001182.2:c.1405 + 5G>A	—	Polymorphism	NP	NP
3-A	GPR98	het	Exon 85	mat	Chr5: g.90281186T>C	NM_032119.3:c.17999T>C	p.(Val6000Ala)	Disease causing	NP	Neutral
4-A	SCN1A	het	Splice site	dn	Chr2: g.166904280T>C	NM_006920.4:c.1029-2A>G	—	Disease causing	NP	NP
	POLG	het	Exon 17	pat	Chr15: g.89864482C>T	NM_001126131.1:c.2608G>A	p.(Val870Ile)	Disease causing	Probably damaging	Neutral
5-A	MAGI2	het	Exon 16	mat	Chr7: g.77789528G>A	NM_012301.3:c.2659C>T	p.(Arg887Cys)	Disease causing	Probably damaging	Deleterious
	KCNQ2	het	Exon 5	pat	Chr20: g.62073806del	NM_172107.2:c.769del	p.(Glu257Argfs*16)	Disease causing	NP	NP
6-A	GPR98	het	Exon 33	pat	Chr5: g.89990155C>T	NM_032119.3:c.7582C>T	p.(Pro2528Ser)	Disease causing	Benign	Neutral
	TBC1D24	het	Exon 2	pat	Chr16: g.2546790G>A	NM_020705.2:c.641G>A	p.(Arg214His)	Disease causing	Probably damaging	Neutral
	KCNQ2	het	Exon 6	dn	Chr20: g.62070955G>A	NM_172107.2:c.923C>T	p.(Pro308Leu)	Disease causing	Probably damaging	Deleterious
7-A	SCN1A	het	Exon 1	dn	Chr2: g.166829960C>T	NM_006920.4:c.172G>A	p.(Gly58Arg)	Disease causing	Possibly damaging	Deleterious
	SCN1B	het	Exon 2	pat	Chr19: g.35523542G>A	NM_199037.3:c.151G>A	p.(Ala51Thr)	Disease causing	Possibly damaging	Deleterious
8-B	No candidate mutations									
9-B	RELN	het	Exon 24	mat	Chr7: g.103244917G>A	NM_005045.3:c.3022C>T	p.(Arg1008Cys)	Disease causing	Possibly damaging	Deleterious
	GABRG2	het	Exon 4	pat	Chr5: g.161524667dup	NM_198904.2:c.351dup	p.(Ala118Cysfs*6)	Disease causing	NP	Deleterious
10-B	RELN	het	Exon 11	pat	Chr7: g.103322621G>T	NM_005045.3:c.1231C>A	p.(Leu411Ile) (rs144978163)	Disease causing	Benign	Neutral
	GRIN2A	het	Exon 11	mat	Chr16: g.9892164C>A	NM_001134407.1:c.2326G>T	p.(Asp776Tyr)	Disease causing	Probably damaging	Deleterious
11-B	No candidate mutations									
12-B	No candidate mutations									
13-B	SCN2A	het	Exon 11	mat	Chr2: g.166172087G>A	NM_001040142.1:c.1490G>A	p.(Ser497Asn)	Disease causing	Benign	Neutral
14-B	No candidate mutations									
15-B	No candidate mutations									
16-B	GPR98	het	Exon 19	pat	Chr5: g.89948228C>G	NM_032119.3:c.3482C>G	p.(Ser1161Cys)	Disease causing	Possibly damaging	Probably damaging
	GRIN2A	het	Exon 8	pat	Chr16: g.9927982C>T	NM_001134407.1:c.1757G>A	p.(Arg586Lys)	Disease causing	Probably damaging	Neutral
17-B	No candidate mutations									
18-B	No candidate mutations									
19-B	SCN9A	het	Exon 21	mat	Chr2: g.167089942G>C	NM_002977.3:c.3799C>G	p.(Leu1267Val)	Disease causing	Probably damaging	Deleterious

Abbreviations: dn, *de novo*; het, heterozygous; hom, homozygous; mat, maternal; pat, paternal; NP, not predicted. Bold entries indicate deleterious mutations. In addition, the associated amino-acid substitutions are located in evolutionarily highly conserved residues and are predicted to functionally affect the encoded protein. Results of segregation analyses by Sanger sequencing are reported too.

^aExon numbering refers to the NCBI transcript number reported in^b.

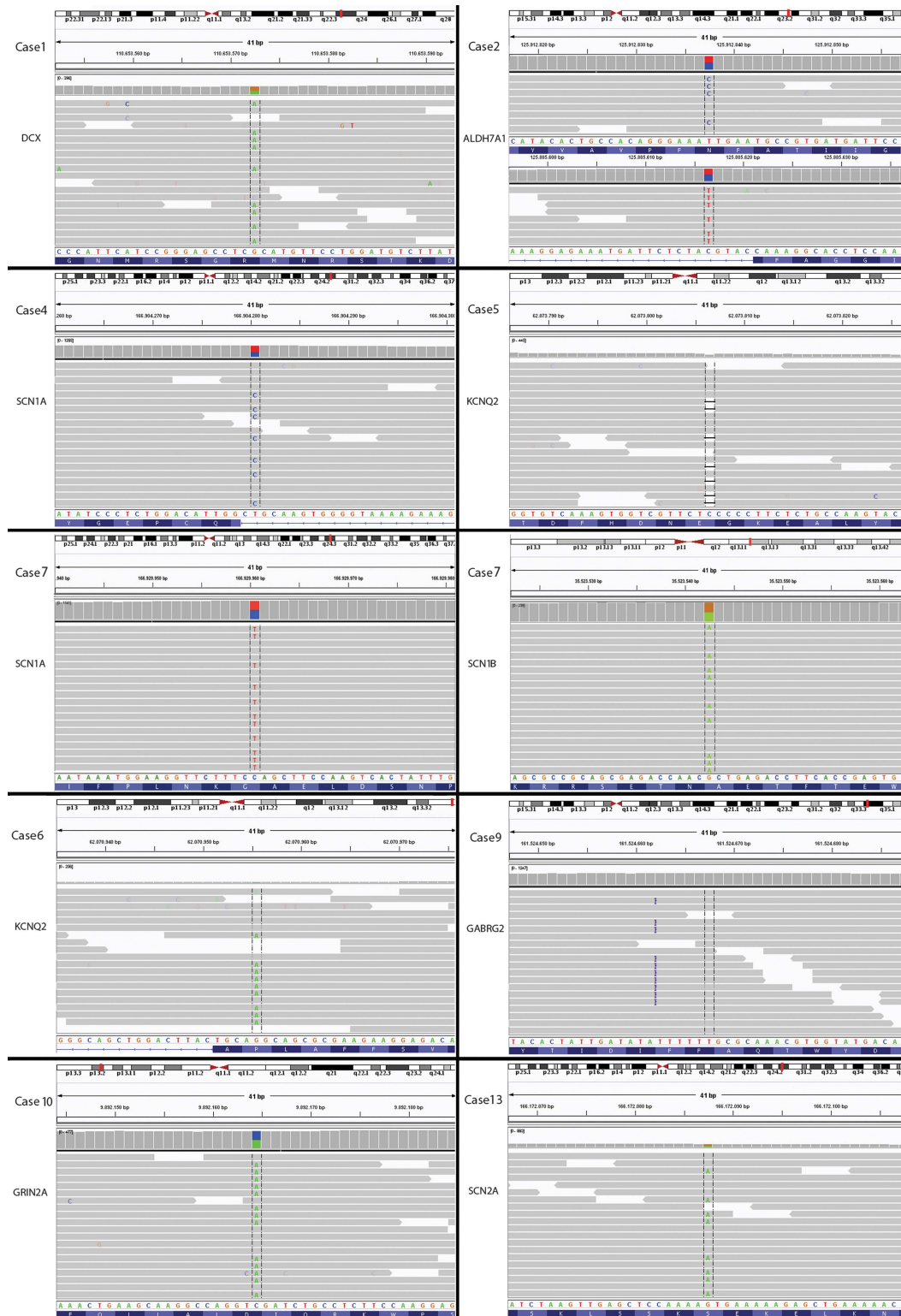


Figure 2 Alignments loaded on IGV 2.1(Integrative Genomics Viewer) for every causative mutation reported in bold in Table 2. Chromosomal view has the covered 41 bp delimited in red, followed by genomic coordinates, relative coverage for single base pair, and alignments covering 101 bp in average. At the bottom of every image the reference sequence and the corresponding amino-acid sequence are visible.

Ala118 with the new reading frame ending in a stop five codons downstream. The mRNA was predicted as target for NMD by Mutation Taster. The same duplication was found in the father who suffered from the same condition. A mutation of *RELN*,

predicted as damaging and inherited from her normal mother was also detected.

Case 10-B had a heterozygous missense mutation in the *GRIN2A* gene inherited from his mother who showed an overlapping

phenotype. This transversion occurred in a highly conserved nucleotide. A missense mutation in the *RELN* gene, predicted disease-causing by Mutation Taster, resulted to be inherited from the normal father.

Case 13-B showed a maternally inherited missense transition in the *SCN2A* gene, affecting a highly conserved nucleotide and predicted to be damaging by SIFT and Mutation Taster but not by Polyphen.

Case 16-B had two missense mutations at *GPR98* and *GRIN2A*, both predicted to be disease causing by at least one prediction tool. Both *GRIN2A* and *GPR98* mutations were inherited from the normal father.

Case 19-B had a predicted damaging missense mutation at *SCN9A*, inherited from the mother.

Cases 8-B, 11-B, 12-B, 14-B, 15-B, 17-B, and 18-B did not show any possible causative mutation.

Control cohort variants

The examination of the 22 variants found in controls after discrete filtering revealed that only 4 of these, all heterozygous, were potentially disease causing.

In particular, we found a missense substitution p.(Gly216Ala) (c.647G>C) in *CHRNA4* gene (NM_000744) in one case and a heterozygous missense mutation p.(Gly1602Ser) (c.4804G>A) in *FLNA* gene (NM_001110556) in a female subject. A third control subject had a mutation of *UBE3A* (NM_000462) (c.1735G>A, p.(Val579Met)) and a second mutation of *SCN1B* (NM_199037) (c.178C>T, p.(Arg60Cys)).

DISCUSSION

We used a NGS-based approach to test 67 epilepsy genes in 19 patients with different types of epilepsy. Patients had been stratified in two groups according to their neurological phenotypes. In the group A, including seven patients whose clinical features were rather suggestive for a specific syndrome, we detected a likely causative mutation in six (cases 1-A, 2-A, 4-A, 5-A, 6-A, 7-A). In the group B, including 12 patients with a phenotype not distinctive for any specific gene, we have been able to find a plausible causative mutations only in three (cases 9-B, 10-B, 13-B), whereas the remaining cases were either negative (seven cases) or had mutations whose role was unclear (cases 16-B and 19-B). These results were not unexpected and emphasize the restriction of this approach is a lack of knowledge about the functional role of most variants, resulting in a large number of variants of uncertain significance.¹⁷ For this reason, the diagnostic yield of 47% (9/19) is quite high. The 12 cases who had at least one mutation are discussed below.

Patient 1-A had a typical clinical picture of Lennox–Gastaut syndrome and magnetic resonance imaging showed a very large double cortex overlapping with the subcortical band heterotopia syndrome. The *de novo* truncating mutation of *DCX* fits very well with her double cortex.¹⁴ The mutation of *SHANK3* was inherited from her normal father. *SHANK3* alterations are causative of Phelan–McDermid syndrome¹⁸ and are characterized by complete penetrance. Thus, we assumed that this missense mutation was likely benign.

Patient 2-A showed neonatal seizures and multifocal epileptiform discharges at EEG, which became normalized with pyridoxine. Presently the patient is 14 years old and the treatment with pyridoxine allowed a complete control of seizures with a normal psychomotor development. She had a compound heterozygous mutation for *ALDH7A1*. The intronic mutation would have been lost if we had disregarded all SNPs reported in public databases, whereas we have taken into consideration all the SNPs with a MAF < 1%. Actually,

both mutations have already been described in patients with pyridoxine-dependent epilepsy.^{15,16} The finding that her epileptic crisis ceased after pyridoxine treatment indeed demonstrated the causality of the *ALDH7A1* mutations.

Patient 3-A, who was also affected by pyridoxine-dependent epilepsy, had normal IQ. No mutations were detected in the candidate *ALDH7A1* gene, whereas a missense mutation was present in *GPR98* inherited from the normal mother. Alterations of *GPR98* have been associated with familial febrile seizures and autosomal recessive or digenic dominant Usher syndrome. However, the clinical phenotype of the patient was completely different from these conditions. Our findings suggest that pyridoxine-dependent seizure does not only rely on *ALDH7A1* mutations.¹⁹

Patient 4-A had a clinical diagnosis of Dravet syndrome, so the *de novo* splice site mutation of *SCN1A* fitted well with his phenotype.^{20–22} The heterozygous mutation in *POLG* was considered not associated with his condition because it was also present in the healthy father, whereas dominant *POLG* mutations are associated with adult onset progressive external ophthalmoplegia.²³

Patient 5-A had neonatal generalized tonic-clonic seizures, occurring on the second day of life, not responsive to any therapy. Seizures ceased spontaneously at the 25th day of life. Her father, who was found to carry the same mutation, had the same neonatal condition also ending at the 25th day of life. The frameshift mutation at *KCNQ2* well correlated with the phenotype.²⁴

The mutation in *MAGI2*, inherited from the healthy mother, was not considered disease causing, although it was predicted as probably damaging. Haploinsufficiency for *MAGI2* has been associated with hypsarrhythmia,²⁵ a condition different from the one we observed in this family.

Patient 6-A showed neonatal seizures since his 3rd day of life that ceased 1 month later. He then suffered from sporadic seizures episodes persisting until now (8 years old). He also presented a severe pervasive developmental disorder. The family history was negative. He had a *de novo* missense mutation of *KCNQ2*. This type of mutation has been reported in several patients with early onset epileptic encephalopathy^{24,26,27} in contrast to frameshift or nonsense mutations that are more frequently found among patients with BFNS.

Patient 7-A was affected by generalized epilepsy with febrile seizures plus (GEFSP) and had normal IQ. Our investigations revealed a *de novo* missense mutation of *SCN1A* and a missense mutation of *SCN1B* inherited by the father. The two mutations could explain her phenotype as GEFSP is extremely heterogeneous, and both *SCN1A* and *SCN1B* are among the genes associated with this condition.²⁸

Patient 9-B, during her first year of life suffered from numerous febrile seizures that diminished later on. She had a frameshift mutation at *GABRG2* present also in her father who suffered from febrile seizure until 5 years of age. In fact, mutations of *GABRG2*, either missense or truncating, cause a spectrum of seizure disorders, ranging from early-onset isolated febrile seizures to GEFSP, type 3, which represents the most severe phenotype.^{29–31} The type of febrile seizures present in this patient and in her father fits well with the milder phenotype. The heterozygous mutation of *RELN* has not been considered as causal of her phenotype because only recessive mutations of this gene are associated with a pathogenic condition characterized by lissencephaly³² not present in our patient who has a normal psychomotor development.

Patient 10-B showed a typical clinical and EEG's picture of benign childhood epilepsy with centrotemporal spikes. We detected a missense mutation at *GRIN2A* that was also present in his mother and aunt (the sister of the mother) showing the same clinical and

EEG's picture with remission of seizures in adolescence and borderline cognitive level. Our patient did not have overt seizures until the age of 7 years when rolandic epilepsy appeared. A series of mutations of this gene have been described in subjects/families with a phenotype overlapping that of our patient including learning disabilities.^{33,34}

Patient 13-B had a missense mutation of *SCN2A* inherited from her mother. Mutations of this gene are associated with noteworthy clinical variability ranging from familial benign seizures to generalized epilepsy with febrile seizures or epileptic encephalopathy.³⁵ The patient presented only febrile seizures, whereas her mother had benign generalized epilepsy with absences.

Patient 16-B had an epileptic encephalopathy with severe cognitive impairment. The two missense mutations highlighted in *GPR98* and *GRIN2A* did not seem related to his phenotype. As he was born from healthy consanguineous parents, an autosomal recessive condition has to be taken into consideration.

Finally, patient 19-B had a missense mutation of *SCN9A* predicted as damaging. However, mutations of this gene are usually associated with febrile seizure, GEFSP, and Dravet syndrome,³⁶ whereas the patient's phenotype was suggestive of an epileptic encephalopathy strongly resembling West syndrome with hypsarrhythmia, spasm, and psychomotor regression. Both his parents were healthy.

Our targeted platform was thought with the aim to provide a quick and cheap molecular diagnosis to most patients with an epileptic disorder. When we built the platform, we thought we could identify the causative mutation independently from any clinical information. In this sense we required the specialist to select the cases in a totally random way and without giving us any information about their phenotype and family history. The only request was to exclude cases with holoprosencephaly for which we have a dedicated NGS platform. Actually, the finding of multiple candidate mutations made clear that the culprit gene could be highlighted only by knowing in detail both the patient's phenotype and the family history. Moreover, predicted damaging mutations had been detected in the healthy controls as well. Eventually, in 9 of 19 patients we identified a very likely causative mutation (Table 2) with most of them detected in cohort A including patients whose phenotype indeed suggested the involvement of a specific gene. Among the 12 patients owing to cohort B, affected by different types of epilepsy not suggestive for any or a single specific gene, we could find the most likely causative mutation only in three and in all of them (cases 9-B, 10-B, 13-B) the alteration could be hardly suspected *a priori*.

The absence of any mutation in seven patients (cases 8-B, 11-B, 12-B, 14-B, 15-B, 17-B, 18-B) indicated that alterations in many other genes not present in our platform are associated with epilepsy, stressing the high genetic heterogeneity of this disorder.

The analysis of the control cohort revealed four potentially damaging mutations in three healthy individuals. None of these variants were previously reported in HGMD. One female subject had a *FLNA* mutation predicted to be damaging. We could not define whether this was a benign variant rather than a really damaging mutation with incomplete penetrance as reported for females with mutation of this gene and cardiac valvular dysplasia (OMIM #314400).³⁷ The interpretation of the *CHRNA4* mutation was also difficult as alterations of this gene can cause either nocturnal frontal lobe epilepsy type 1, although with incomplete penetrance, or nightmares and other sleep disorders that are often undiagnosed.³⁸ Variants in *SCN1B* and *UBE3A* were identified in the same subject. As in the case of *FLNA*, the *SCN1B* mutation could be either neutral or pathological with incomplete penetrance,²⁸ whereas the *UBE3A* mutation could be either a benign variant or disease causing but

inherited from the father.³⁹ To conclude, we were unable to interpret some of the genetic lesions we met in the control cohort, further stressing that our knowledge of genetic variants is presently limited, increasing the risk of false-positive and false-negative information.⁴⁰ If these lesions were indeed pathogenic, we should consider the hypothesis that common disease traits such as epilepsy are the result of different genetic components. In fact, the observation of deleterious mutations in 237 ion-channel genes with the same prevalence in individuals with epilepsy and control subjects suggested that, at least for these genes, the personal risk assessment in epilepsy depends more on the combination of the variants rather than specific deleterious variants.⁴¹ The lack of enrichment of protein-disrupting ion-channel mutations in individuals with epilepsy has been confirmed by Heinzen *et al.*⁴² These authors also demonstrated that single epilepsy-susceptible variants identified by exome sequencing in patients with idiopathic generalized epilepsy (juvenile myoclonic epilepsy and absence epilepsy), although rare, were possibly real risk factors, each of them accounting for only a small fraction of individuals with epilepsy. This burden of data makes evident the complex architecture of epilepsy with genetic heterogeneity much higher than expected. It is conceivable that in the near future the collection of clinical history, EEG, and imaging will be combined with the analysis of NGS-dedicated platforms. The negative cases will be analyzed for whole exome if not for whole genome. The advantages of this approach are evident both in the immediate (consulting for risk of recurrence) and in the long run when specifically targeted treatments will be adopted.

We have been able to conclude the analysis of nine patients (the number of patients we pool in a single lane), including the enlargement of the investigation to parents, in 8–10 weeks with a cost per patient comparable to sequencing 1–2 medium-to-large-sized genes by conventional techniques, a result overlapping that reported by Lemke *et al.*⁴ Our results suggest that using a single platform to sequence all or most of the epilepsy genes may increase the diagnostic yield. This is particularly true in absence of clinical signs suggestive for involvement of a specific gene like in three patients of our cohort. Obviously novel mutations require that their causative role is further confirmed by segregation or functional analyses. On the contrary, smaller platforms containing a limited number of genes may reduce the efficacy of the NGS-based approach, as epilepsy is extremely heterogeneous both under the genetic and phenotypic point of view. Our platform, as well as the one already described by Lemke *et al.*⁴ has the advantage that can also be run on 'benchtop' sequencers, which forego high yields in exchange for low capital costs, small physical footprints, and more rapid turnaround times, making them far more attractive to smaller biomedical laboratories.⁴³

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We are grateful to the patients and families for their participation in the study. This work was supported by the Ricerca Corrente RC11017C (Diagnosis and therapy of epileptic syndromes), Fondazione IRCCS C Mondino, Pavia, and the Telethon Foundation (project GGP10121C).

WEB RESOURCES

1. NIH Undiagnosed Diseases Program: <http://rarediseases.info.nih.gov/Resources.aspx?PageID=31>
2. dbSNP135: <http://www.ncbi.nlm.nih.gov/projects/SNP/>
3. 1000 Genomes Project Database: <http://www.1000genomes.org/data>

4. Exome Sequencing Project Database: <https://esp.gs.washington.edu/drupal/>
 5. HGMD 2012.3: <http://www.biobase-international.com/product/hgmd>
 6. Mutation Taster: <http://www.mutationtaster.org>
 7. Polyphen2: <http://genetics.bwh.harvard.edu/pph2/>
 8. SIFT: <http://SIFT.jcvi.org>
 9. LOVD: <http://www.lovd.nl/3.0/home>
-
- 1 Hauser WA, Annegers JF, Rocca WA: Descriptive epidemiology of epilepsy: contributions of population-based studies from Rochester, Minnesota. *Mayo Clin Proc* 1996; **71**: 576–586.
 - 2 EPICURE Consortium, EMINet Consortium, Steffens M *et al*: Genome-wide association analysis of genetic generalized epilepsies implicates susceptibility loci at 1q43, 2p16.1, 2q22.3 and 17q21.32. *Hum Mol Genet* 2012; **21**: 5359–5372.
 - 3 Mefford HC, Muhle H, Ostertag P *et al*: Genome-wide copy number variation in epilepsy: novel susceptibility loci in idiopathic generalized and focal epilepsies. *PLoS Genet* 2010; **6**: e1000962.
 - 4 Lemke JR, Riesch E, Scheurenbrand T *et al*: Targeted next generation sequencing as a diagnostic tool in epileptic disorders. *Epilepsia* 2012; **53**: 1387–1398.
 - 5 EPICURE Consortium, Epilepsy Phenome/Genome Project, Allen AS *et al*: De novo mutations in epileptic encephalopathies. *Nature* 2013; **501**: 217–221.
 - 6 Johansen Taber KA, Dickinson BD, Wilson M: The promise and challenges of next-generation genome sequencing for clinical care. *JAMA Intern Med* 2013; **174**: 275–280.
 - 7 Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**: 1754–1760.
 - 8 Li H, Handsaker B, Wysoker A *et al*: The sequence alignment/map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–2079.
 - 9 DePristo MA, Banks E, Poplin R *et al*: A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; **43**: 491–498.
 - 10 Schwarz JM, Rodelsperger C, Schuelke M, Seelow D: MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 2010; **7**: 575–576.
 - 11 Adzhubei IA, Schmidt S, Peshkin L *et al*: A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7**: 248–249.
 - 12 Bamshad MJ, Ng SB, Bigham AW *et al*: Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 2011; **12**: 745–755.
 - 13 Webb TR, Parfitt DA, Gardner JC *et al*: Deep intronic mutation in OFD1, identified by targeted genomic next-generation sequencing, causes a severe form of X-linked retinitis pigmentosa (RP23). *Hum Mol Genet* 2012; **21**: 3647–3654.
 - 14 Leger PL, Souville I, Boddaert N *et al*: The location of DCX mutations predicts malformation severity in X-linked lissencephaly. *Neurogenetics* 2008; **9**: 277–285.
 - 15 Bennett CL, Chen Y, Hahn S, Glass IA, Gospe Jr SM: Prevalence of ALDH7A1 mutations in 18 North American pyridoxine-dependent seizure (PDS) patients. *Epilepsia* 2009; **50**: 1167–1175.
 - 16 Striano P, Battaglia S, Giordano L *et al*: Two novel ALDH7A1 (antiquitin) splicing mutations associated with pyridoxine-dependent seizures. *Epilepsia* 2009; **50**: 933–936.
 - 17 Jacob HJ: Next-generation sequencing for clinical diagnostics. *N Engl J Med* 2013; **369**: 1557–1558.
 - 18 Bonaglia MC, Giorda R, Beri S *et al*: Molecular mechanisms generating and stabilizing terminal 22q13 deletions in 44 subjects with Phelan/McDermid syndrome. *PLoS Genet* 2011; **7**: e1002173.
 - 19 Mefford HC, Cook J, Gospe Jr SM: Epilepsy due to 20q13.33 subtelomere deletion masquerading as pyridoxine-dependent epilepsy. *Am J Med Genet A* 2012; **158A**: 3190–3195.
 - 20 Depienne C, Trouillard O, Saint-Martin C *et al*: Spectrum of SCN1A gene mutations associated with Dravet syndrome: analysis of 333 patients. *J Med Genet* 2009; **46**: 183–191.
 - 21 Escayg A, Goldin AL: Sodium channel SCN1A and epilepsy: mutations and mechanisms. *Epilepsia* 2010; **51**: 1650–1658.
 - 22 Dravet C, Bureau M, Oguni H, Fukuyama Y, Cokar O: Severe myoclonic epilepsy in infancy: Dravet syndrome. *Adv Neurol* 2005; **95**: 71–102.
 - 23 Kiechl S, Horvath R, Luoma P *et al*: Two families with autosomal dominant progressive external ophthalmoplegia. *J Neurol Neurosurg Psychiatry* 2004; **75**: 1125–1128.
 - 24 Weckhuysen S, Mandelstam S, Suls A *et al*: KCNQ2 encephalopathy: emerging phenotype of a neonatal epileptic encephalopathy. *Ann Neurol* 2012; **71**: 15–25.
 - 25 Marshall CR, Young EJ, Pani AM *et al*: Infantile spasms is associated with deletion of the MAGI2 gene on chromosome 7q11.23-q21.11. *Am J Hum Genet* 2008; **83**: 106–111.
 - 26 Saitsu H, Kato M, Koide A *et al*: Whole exome sequencing identifies KCNQ2 mutations in Ohtahara syndrome. *Ann Neurol* 2012; **72**: 298–300.
 - 27 Kato M, Yamagata T, Kubota M *et al*: Clinical spectrum of early onset epileptic encephalopathies caused by KCNQ2 mutation. *Epilepsia* 2013; **54**: 1282–1287.
 - 28 Bonanni P, Malcarne M, Moro F *et al*: Generalized epilepsy with febrile seizures plus (GEFS+): clinical spectrum in seven Italian families unrelated to SCN1A, SCN1B, and GABRG2 gene mutations. *Epilepsia* 2004; **45**: 149–158.
 - 29 Baulac S, Huberfeld G, Gourfinkel-An I *et al*: First genetic evidence of GABA(A) receptor dysfunction in epilepsy: a mutation in the gamma2-subunit gene. *Nat Genet* 2001; **28**: 46–48.
 - 30 Kananura C, Haug K, Sander T *et al*: A splice-site mutation in GABRG2 associated with childhood absence epilepsy and febrile convulsions. *Arch Neurol* 2002; **59**: 1137–1141.
 - 31 Macdonald RL, Kang JQ, Gallagher MJ: Mutations in GABAA receptor subunits associated with genetic epilepsies. *J Physiol* 2010; **588**: 1861–1869.
 - 32 Bahi-Buisson N, Guerrini R: Diffuse malformations of cortical development. *Handb Clin Neurol* 2013; **111**: 653–665.
 - 33 Carvill GL, Regan BM, Yendle SC *et al*: GRIN2A mutations cause epilepsy-aphasia spectrum disorders. *Nat Genet* 2013; **45**: 1073–1076.
 - 34 Lemke JR, Lal D, Reinthaler EM *et al*: Mutations in GRIN2A cause idiopathic focal epilepsy with rolandic spikes. *Nat Genet* 2013; **45**: 1067–1072.
 - 35 Nakamura K, Kato M, Osaka H *et al*: Clinical spectrum of SCN2A mutations expanding to Ohtahara syndrome. *Neurology* 2013; **81**: 992–998.
 - 36 Mulley JC, Hodgson B, McMahon JM *et al*: Role of the sodium channel SCN9A in genetic epilepsy with febrile seizures plus and Dravet syndrome. *Epilepsia* 2013; **54**: e122–e126.
 - 37 Kyndt F, Gueffet JP, Probst V *et al*: Mutations in the gene encoding filamin A as a cause for familial cardiac valvular dystrophy. *Circulation* 2007; **115**: 40–49.
 - 38 Derry CP, Duncan JS, Berkovic SF: Paroxysmal motor disorders of sleep: the clinical spectrum and differentiation from epilepsy. *Epilepsia* 2006; **47**: 1775–1791.
 - 39 Buiting K: Prader-Willi syndrome and Angelman syndrome. *Am J Med Genet C Semin Med Genet* 2010; **154C**: 365–376.
 - 40 Robinson PN, Krawitz P, Mundlos S: Strategies for exome and genome sequence data analysis in disease-gene discovery projects. *Clin Genet* 2011; **80**: 127–132.
 - 41 Klassen T, Davis C, Goldman A *et al*: Exome sequencing of ion channel genes reveals complex profiles confounding personal risk assessment in epilepsy. *Cell* 2011; **145**: 1036–1048.
 - 42 Heinzen EL, Depoind C, Cavalleri GL *et al*: Exome sequencing followed by large-scale genotyping fails to identify single rare variants of large effect in idiopathic generalized epilepsy. *Am J Hum Genet* 2012; **91**: 293–302.
 - 43 MacArthur DG, Lek M: The uncertain road towards genomic medicine. *Trends Genet* 2012; **28**: 303–305.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)