

## Comparison of two next-generation sequencing kits for diagnosis of epileptic disorders with a user-friendly tool for displaying gene coverage, DeCovA



Sarra Dimassi<sup>a,b,c,1</sup>, Thomas Simonet<sup>d,e,1</sup>, Audrey Labalme<sup>a</sup>, Nadia Boutry-Kryza<sup>a</sup>, Amandine Campan-Fournier<sup>a</sup>, Raphaëlle Lamy<sup>a</sup>, Claire Bardel<sup>c,f,g</sup>, Mad-Hélénie Elsensohn<sup>c,f,g</sup>, Florence Roucher-Boulez<sup>c,g,h</sup>, Nicolas Chatron<sup>a</sup>, Audrey Putoux<sup>a,b,c</sup>, Julitta de Bellescize<sup>i</sup>, Dorothee Ville<sup>j</sup>, Laurent Schaeffer<sup>c,d,e</sup>, Pascal Roy<sup>c,f,g</sup>, Soumaya Mougou-Zerelli<sup>h,i</sup>, Ali Saad<sup>h,i</sup>, Alain Calender<sup>a,c</sup>, Damien Sanlaville<sup>a,b,c,\*</sup>, Gaetan Lesca<sup>a,b,c,\*</sup>

<sup>a</sup> Department of Medical Genetics, Lyon University Hospital, Lyon, France

<sup>b</sup> CNRL, CNRS UMR 5292, INSERM U1028, Lyon, France

<sup>c</sup> Claude Bernard Lyon I University, Lyon, France

<sup>d</sup> Department of Cell Biotechnology, Lyon, France

<sup>e</sup> Ecole Normale Supérieure de Lyon, Lyon, France

<sup>f</sup> Service de Biostatistique, Hospices Civils de Lyon, Lyon, France

<sup>g</sup> CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, Equipe Biostatistique-Santé, Villeurbanne, France

<sup>h</sup> Cytogenetics and Reproductive Biology Department, Farhat Hached University Teaching Hospital, Sousse, Tunisia

<sup>i</sup> Common Service Units for Research in Genetics, Faculty of Medicine of Sousse, University of Sousse, Sousse, Tunisia

<sup>j</sup> Department of Neuropediatrics, Reference Center for Tuberous Sclerosis and Rare Epileptic Syndromes, Lyon University Hospital, Lyon, France

### ARTICLE INFO

#### Article history:

Received 17 July 2015

Received in revised form 25 September 2015

Accepted 6 October 2015

#### Keywords:

Next-generation sequencing

Gene panel

Epilepsy

Coverage analysis

### ABSTRACT

In recent years, molecular genetics has been playing an increasing role in the diagnostic process of monogenic epilepsies. Knowing the genetic basis of one patient's epilepsy provides accurate genetic counseling and may guide therapeutic options. Genetic diagnosis of epilepsy syndromes has long been based on Sanger sequencing and search for large rearrangements using MLPA or DNA arrays (array-CGH or SNP-array). Recently, next-generation sequencing (NGS) was demonstrated to be a powerful approach to overcome the wide clinical and genetic heterogeneity of epileptic disorders. Coverage is critical for assessing the quality and accuracy of results from NGS. However, it is often a difficult parameter to display in practice. The aim of the study was to compare two library-building methods (Haloplex, Agilent and SeqCap EZ, Roche) for a targeted panel of 41 genes causing monogenic epileptic disorders. We included 24 patients, 20 of whom had known disease-causing mutations. For each patient both libraries were built in parallel and sequenced on an Ion Torrent Personal Genome Machine (PGM). To compare coverage and depth, we developed a simple homemade tool, named DeCovA (Depth and Coverage Analysis). DeCovA displays the sequencing depth of each base and the coverage of target genes for each genomic position. The fraction of each gene covered at different thresholds could be easily estimated. None of the two methods used, namely NextGene and Ion Reporter, were able to identify all the known mutations/CNVs displayed by the 20 patients. Variant detection rate was globally similar for the two techniques and DeCovA showed that failure to detect a mutation was mainly related to insufficient coverage.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### 1. Introduction

Epilepsy is one of the most common neurological conditions, with a prevalence of  $\approx 1\%$  (Hauser et al., 1996). Although the genetic basis

of epilepsies have long remained elusive, recent technological breakthroughs such as array comparative genomic hybridization (aCGH) and next-generation sequencing (NGS), which include gene panels and whole-exome sequencing (WES), have led, in the past few years, to the identification of an increasing number of genes (Lesca et al., 2013; Thevenon et al., 2014; Dimassi et al., 2015). Mutations or copy-number variations (CNV) in those genes are responsible or predispose to various, familial or sporadic, epileptic disorders (Lesca and Depienne, 2015). Genetic molecular diagnosis is an important step to

\* Corresponding author at: Service de Génétique, Groupement Hospitalier Est, 59 Boulevard Pinel, 69677 Bron, France.

E-mail addresses: [damien.sanlaville@chu-lyon.fr](mailto:damien.sanlaville@chu-lyon.fr) (D. Sanlaville), [gaetan.lesca@chu-lyon.fr](mailto:gaetan.lesca@chu-lyon.fr) (G. Lesca).

<sup>1</sup> These authors contributed equally to this work.

resume etiological investigations and to allow accurate genetic counseling, especially for childhood-onset epileptic disorders. It will probably also allow the use of specific therapies in the future.

Sanger sequencing combined with detection of copy-number variations with semi-quantitative techniques such as MLPA (Multiple Ligation-Probe Amplification) have been considered as the gold standard for molecular diagnosis for many years. However, the use of these techniques becomes increasingly time consuming and costly as the number of genes of interest increases. Recent studies have demonstrated the efficacy of detecting genetic alterations in patients with epileptic syndromes, in a diagnostic setting using an NGS panel approach or WES (Lemke et al., 2012; Martin et al., 2014; Della Mina et al., 2015). Although coverage is critical for assessing the quality and accuracy of results from NGS, it is often a difficult parameter to display in practice. The aim of the study was to compare two library-building technologies for a gene panel for the diagnosis of genetic epilepsies, as well as two software tools for variant analysis. To facilitate this comparison we developed DeCovA (Depth and Coverage Analysis), a user-friendly tool that displays the sequencing depth of each base and the coverage of target genes.

## 2. Materials and methods

### 2.1. Patients

Informed consent was obtained from the patients or their parents, according to the French bioethics law. DNA samples were anonymized and provided a reference number (1 to 24) for the study. We included 20 patients with germline mutations of various types, indels or copy-number variations (CNV) previously characterized in one out of the 15 genes studied in our laboratory by Sanger sequencing and MLPA (*ATP1A3*, *CDKL5*, *CSTB*, *EPM2A*, *FOXG1*, *FOXP2*, *GRIN2A*, *KCNQ2*, *KCNQ3*, *LG11*, *MECP2*, *NHLRC1*, *PRRT2*, *SCN2A*, and *STXBPI1*). We also included 4 patients with epileptic disorder of unknown cause. Samples were anonymized before library construction, sorted, and the analyses performed on a blind basis. The flow chart is shown in Fig. 1.

### 2.2. Design of the gene panels

A list of 41 genes related to epilepsy were targeted in our custom panels: *ALDH7A1*, *ARHGEF9*, *ARX*, *ATP1A3*, *CDKL5*, *CHRNA2*, *CHRNA4*, *CHRNA2*, *CSTB*, *DEPDC5*, *EPM2A*, *FOXG1*, *FOXP2*, *GRIN2A*, *KCDT7*, *KCNJ10*, *KCNMA1*, *KCNQ2*, *KCNQ3*, *KCNT1*, *LG11*, *MECP2*, *MEF2C*, *NHLRC1*, *PCDH19*, *PLCB1*, *PNKP*, *POLG*, *PRICKLE1*, *PRRT2*, *TCF4*, *SCARB2*, *SCN1A*, *SCN2A*, *SCN8A*, *SLC2A1*, *SLC9A6*, *SLC25A22*, *SPTAN1*, *STXBPI1*, and *UBE3A*. Each exon of various isoforms flanked by 50 bp, as well as 3'- and 5'-UTR regions, were submitted for the design. The targeted region represented 390,339 bp. The Haloplex panel was created using the SureDesign interface (Agilent, Santa Clara, CA, USA), which predicted a 99.83% coverage. With the SeqCap EZ Enrichment System (Roche Nimblegen, Madison, WI, USA) the theoretical coverage of the region targeted was 99.9%.

### 2.3. Library preparation

Haloplex libraries were built according to the manufacturer's recommendations (Haloplex Target Enrichment System, Version D.3, March 2013). SeqCap EZ libraries were built following the procedure "Preparation of Ion Torrent Personal Genome Machine (PGM) System Fragment Libraries for Targeted Re-sequencing using SeqCap EZ Solution Based Sequence Capture". The same DNA samples were used to test the two library preparation methods. Quality and quantity were assessed using a Bioanalyzer High-Sensitivity DNA Assay kit (Agilent). The fragments generated by the Haloplex library had a size between 50 and 500 bp whereas those generated by SeqCap EZ had a mean size of 200 bp.

### 2.4. Ion Torrent PGM Sequencing and alignments

Emulsion PCR was done according to the manufacturer's instructions. Ion sphere particles were enriched using the ES module and sequenced with the ion PGM on a 318 v2 chip (Life Technologies, Saint Aubin, France). As the sizes of the library fragments were different and because many fragments generated with the Haloplex library were superior to 200 bp, we used the Ion PGM Sequencing 400 Kit and the Ion PGM Sequencing 200 Kit to sequence the SeqCap EZ library. Libraries were pooled at a concentration of 20 pM. The platform specific pipeline Torrent Suite software v3.6.2 was used to perform optimized signal processing and sequence alignment onto the human genome reference hg19.

### 2.5. Coverage analysis

For our needs, we developed a specific tool, DeCovA, which is a command line Perl script using bedtools (<http://bioinformatics.oxfordjournals.org/content/26/6/841.full.pdf+html>) to extract depth from aligned bam files, and R tools to draw corresponding graphs (available upon request to [thomas.simonet@chu-lyon.fr](mailto:thomas.simonet@chu-lyon.fr)). DeCovA first determines the exonic target regions from either a bed file of genomic intervals or a list of genes or transcripts, by looking for overlap with annotation data from a RefSeq file (eg RefGene.txt or EnsGene.txt, from UCSC). Secondly, it runs the coverage Bed tool from the Bedtools suite to extract the depth of sequencing at each base of these regions, for each bam file to be analyzed (provided by a list of files, or directories where to find them). Finally, these data are used to compute coverage of targeted regions, which are reported, for each gene or transcript and for each sample, as a graph highlighting with a color code the regions not covered at different depth thresholds (Fig. 2A) and a text file containing the fraction of each exon that is not covered and genomic position of not covered domains (Fig. 2C). Additionally, a global analysis, for all the targeted intervals taken together, is also outputted, as a bar plot of the fraction of region of interest covered for each depth bin. The script also prints a bed file reporting these uncovered intervals to quickly determine which exons would have to be Sanger-sequenced to achieve full coverage.

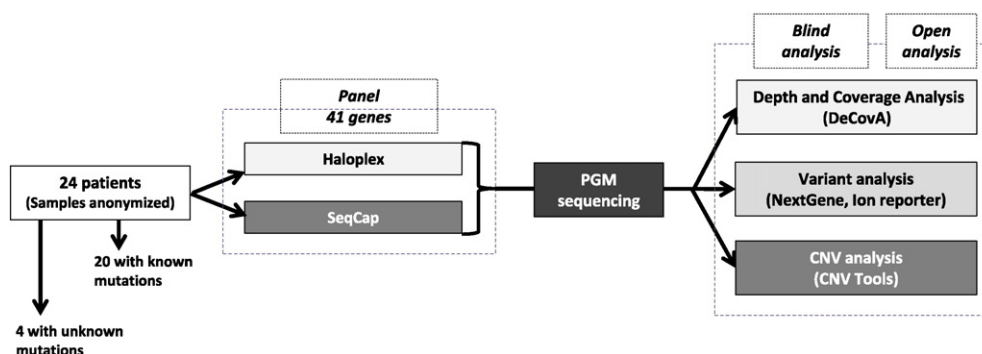
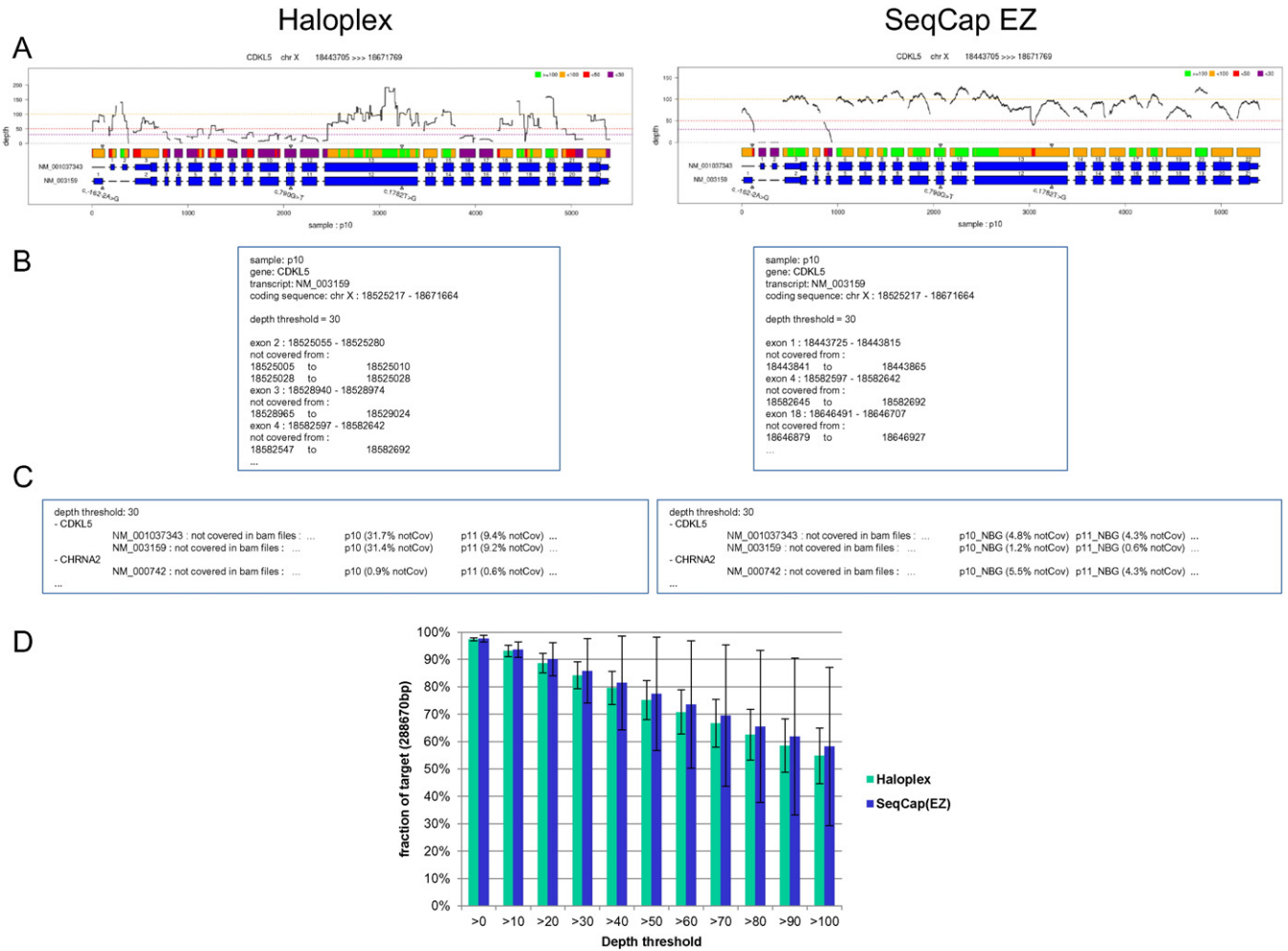


Fig. 1. Design of the study. DNA from patients was anonymized and the first phase of the analysis was performed on a blind basis.



**Fig. 2.** DeCoV allows for a quick overview of depth and coverage data, either on genes of interest, or on the whole genomic target. **A.** Graph by gene (here *CDKL5* as an example) and by sample, showing the depth with a black solid line, above the target regions, depicted with a color code: here, green is for regions  $>100\times$ , yellow between  $50\times$  to  $100\times$ , orange between  $30\times$  to  $50\times$  and red below  $30\times$ ; below are the exons, in blue, with coding regions widened, and the known mutations shown with small triangles. **B.** Part of the text file, giving, for each transcript, each depth threshold and each sample, the coordinates of uncovered domain, relative to the exons. **C.** Part of the text summary, giving, for each gene and transcript, the samples not fully covered (here shown for patients 10 and 11 samples, and for *CDKL5* and *CHRNA2* as examples). **D.** Bar plots of global coverage (here means and sd from all the samples), on the whole target region from the bed file, showing the similar capture efficiency for both methods.

DeCoV was conceived as a flexible tool: i) different depth thresholds can be set (up to 5), ii) length beyond ends of exons can be changed, either to a set value, or to reach the positions specified in the bed file, iii) the analysis includes or not non-coding transcripts, and iv) a file containing known mutations can be provided: the depth at these mutation positions will be reported and they will be plotted on the graphs.

Our tool can perform a combined analysis of a group of samples, which allows a quick comparison between them. In this case, a unique picture containing the analyses of all samples can be provided, for the graph per gene and the global bar plots, as well as a summary text file listing, for all the genes and transcripts, the samples not covered (Fig. 2B). In order to allow for the comparison of different enrichment and sequencing methods, the script provides a bar plot of means and standard deviations of coverage from all samples (Fig. 2D), and more importantly, a graph plotting the number of covered samples at each target position, above the gene scheme, for each depth threshold, which elicits the regions that constantly escape the method (Fig. 3).

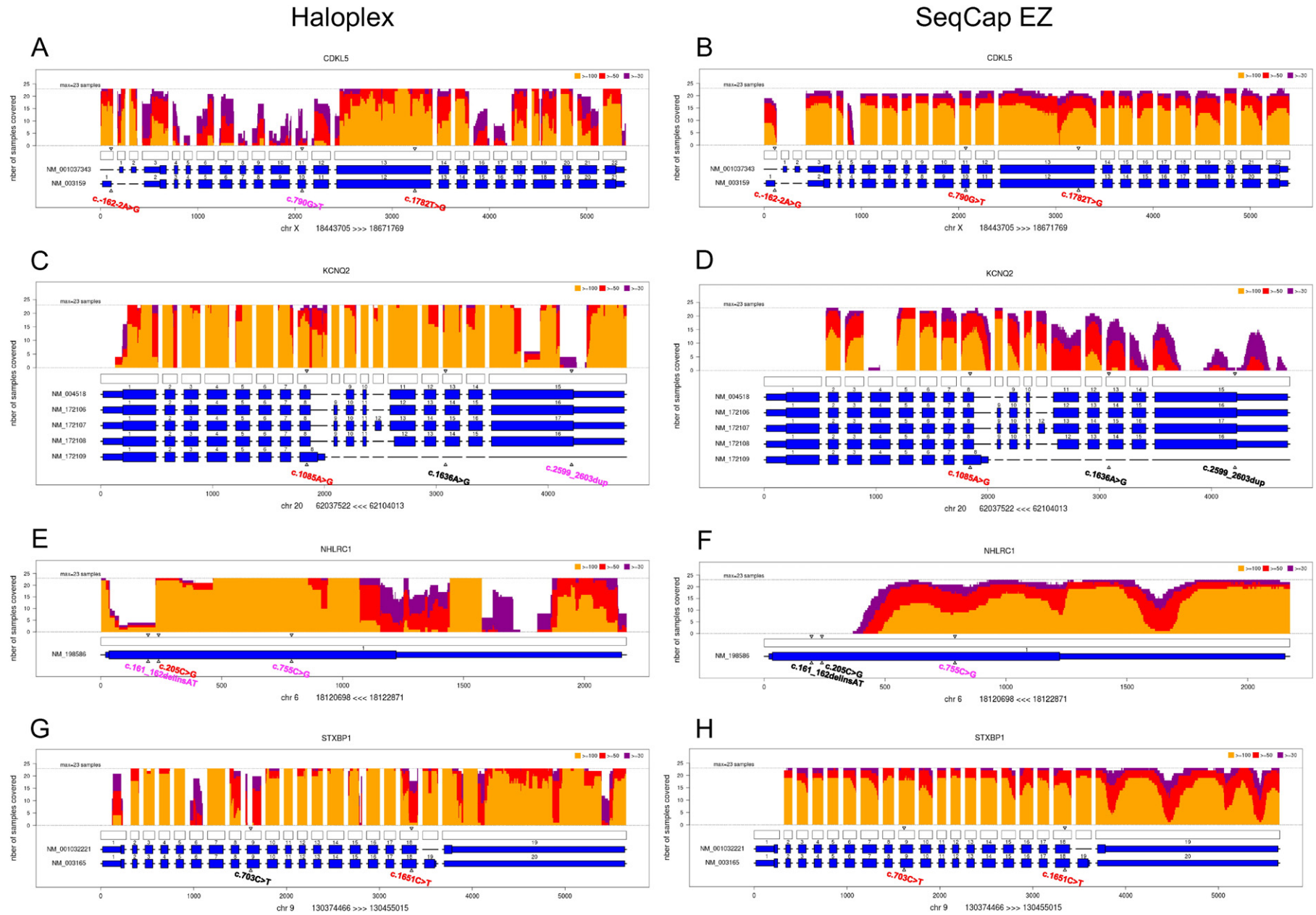
## 2.6. Mutation identification

Variant calling was performed by the Torrent Suite software V3.6.2, using germline parameters and low stringency settings. The vcf files obtained were annotated and filtered using the Ion Reporter Software 4.4

with custom parameters (inclusion of variants located in exons, splice site and UTR regions, not synonymous, with a minimum coverage of  $20\times$ , and not reported as common SNP in the USCS website). We also tested the NextGene software (Softgenetics, State College, PA, USA) to re-analyze a new variant calling from BAM files, using personal filter settings: inclusion of variants within coding sequences ( $\pm 12$  bp) and intronic splice-sites with a minimum coverage of  $20\times$ , and discarding of synonymous single-nucleotide variations. Each mutation identified was visualized using Alamut Visual (Integrative Biossoftwares, Rouen, France). In addition, CNV Tools using the SNP-Based Normalization with smoothing algorithm, with a  $\log_2$  ratio filter  $\leq -0.7$  and  $\geq 0.3$  proposed by NextGene was used to look for copy-number variations (CNV) in the patients.

## 2.7. Results

Sequencing of libraries prepared with Haloplex generated an average of 485,210 (SD 154,231) reads per patient with a mean read length of 135 bp (SD = 7). Sequencing of libraries prepared with SeqCap EZ generated a similar average number of reads: 469,283 (SD = 273,083) per patient with a mean read length of 199 bp (SD = 9) (Table 1). Approximately  $2.67\% \pm 0.55\%$  of our target were not sequenced ( $0\times$ ) with the Haloplex kit and  $2.03\% \pm 1\%$  with SeqCap EZ kit. 88.67%



**Fig. 3.** Illustration of the sum coverage obtained by Haloplex and SeqCap EZ for *CDKL5* (A, B), *KCNQ2* (C, D), *NHLRC1* (E, F), and *STXBP1* (G, H) genes using DeCovA for our 23 samples (sample 13 was removed for further analysis). The vertical axis shows the number of patients with a coverage  $\geq 100\times$  (yellow),  $50\times$  (orange), and  $30\times$  (red). The horizontal axis shows the different exons (large blue rectangles), and UTR regions (thin blue rectangles) for the corresponding isoforms, and the unfilled rectangles stand for target regions from bed file. Triangles point to the positions of mutations: those written in black were found by neither NextGene nor IonReporter analysis, those written in pink were found only with IonReporter variant caller, and those in red were identified by both.

( $\pm 3.53$ ) of our target were covered at  $20\times$  with the Haloplex design and  $90.08\%$  ( $\pm 6.04$ ) with the SeqCap EZ design.

## 2.8. Depth and coverage analysis

DeCovA used these alignment data to provide a friendly graphic illustration showing the depth line, per gene included in the design and per sample, and the stretches of DNA not covered at  $30\times$ ,  $50\times$ , or  $100\times$  thresholds, colored over rectangles representing the bed intervals, above schemes of exons and intronic regions of all transcripts (Fig. 2A). In our experiments, Haloplex data from patient 13 displayed insufficient coverage and was excluded for further analyses. Bar plots of means of coverage (Fig. 2D) showed a similar coverage for both library-building techniques. For example,  $84\% \pm 5$  of the targeted regions were covered at  $30\times$  with Haloplex and  $80\% \pm 11$  with SeqCap EZ (but note that we performed deduplication for SeqCap EZ data, to avoid artificial increase of depth).

Graphs with the sum of patients show the regions that are fully uncovered by either one or both library-building techniques: above the schemes of exons and bed intervals is plotted, by base position, the number of covered sample, at  $30$ ,  $50$ , and  $100\times$  (Fig. 3). For example, exon 4 of *KCNQ2* was covered with Haloplex (Fig. 3C) but not with SeqCap EZ (Fig. 3D), whereas large portions of the last exon are uncovered with both methods.

Coverage was globally similar for both library-building techniques, as shown for four genes (*CDKL5*, *KCNQ2*, *NHLRC1*, and *STXBPI1*) in Fig. 3. We could easily see the differences in coverage and depth between the two methods: the global number of regions that were not covered was similar and consistent for the 23 patients for each method (patient 13 has been excluded). However, depth values showed higher variability among patients as well as among adjacent regions for a given patient with Haloplex than with SeqCap EZ (Fig. 2A).

## 2.9. Mutation identification

Data were first analyzed on a blind basis, with NextGene and Ion Reporter. We did not consider variants in the 5'- or 3'-UTR (except for the first 30 bp) as well as intronic variants that were not located within the

acceptor or donor splice sites. With the filter settings we defined, NextGene analysis generated an average of 159 variants per sample for the Haloplex panel and an average of 82 variants for the SeqCap EZ panel.

Detection of single nucleotide variation (SNV) analysis was the most efficient, allowing the identification of 10/15 of expected SNVs with the Haloplex panel and 11/15 with the SeqCap EZ panel (Table 1). The large number of potential indels detected by NextGene analysis (average of putative indels per patient = 151 (SD = 33) for Haloplex and 72 (SD = 10) for SeqCap EZ) made it impossible to select true indels from artifacts mimicking indels. By contrast, Ion Reporter analysis displayed an average of 21 (SD = 6) variants per sample for the Haloplex kit and 3 (SD = 1) for the SeqCap EZ kit. Detection of indels became possible due to the limited numbers of variants detected: 3/4 indels were correctly found with Haloplex and 1/4 with SeqCap EZ. After the removal of anonymity, we were able to identify the causative mutations (SNV + indels) in 48% of patients with the Haloplex kit and in 52% with the SeqCap EZ kit, using NextGene (Table 1). By contrast, none of the 2 CNVs included in our series could be correctly identified (data not shown).

In order to understand the reason why some SNVs or indels had not been found during the blind phase of the study, we directly looked at the position of each mutation/CNV with Alamut visual. All the missense or indels that were not detected during the blind phase were located in regions that were not covered or that had a depth below the  $20\times$  threshold, as shown by DeCovA. The *MECP2* c.916C>T mutation of patient 16 was present in the variant pool generated by the SeqCap EZ sequencing, but was filtered out by both NextGene and Ion Reporter because the coverage of this region was  $14\times$  (Table 2).

Fig. 3 shows some example of genes that were better covered by either Haloplex (Fig. 3C,D) or SeqCap EZ (Fig. 3A,B), as shown with DeCovA. Some regions failed to be properly covered by either technique because they were GC-rich. Fig. 3 illustrates this point for the last exon of *KCNQ2* or the beginning of the coding sequence of *NHLRC1* (Fig. 3, Table 2). In these two examples, SeqCap EZ did not capture any sequence whereas Haloplex could provide amplification products above the  $20\times$  threshold (Fig. 3C–F). By contrast, the *STXBPI1* mutation of

**Table 1**

Summary of data of sequencing and of mutation identification, for the different combinations of library-building techniques and analysis software.

Sequencing and mapping stats	Haloplex		SeqCap EZ	
Total sequenced reads	485,210 (154,231)		469,283 (273,083)	
Total sequenced bases	7.97E + 07 ( $\pm 1.85E + 07$ )		10.31E + 07 ( $\pm 4.95E + 07$ )	
Mean read length	135 ( $\pm 7$ )		199 ( $\pm 9$ )	
Bases $\geq$ Q20	88.78% ( $\pm 1.23$ )		89.04% ( $\pm 0.49$ )	
Bases $\geq$ Q30	52.50% ( $\pm 1.78$ )		48.98% ( $\pm 1.00$ )	
%GC	50.9% ( $\pm 1.1$ )		43.2% ( $\pm 0.9$ )	
Mapping rate	97.79% ( $\pm 1.45$ )		99.61% ( $\pm 0.06$ )	
Duplicate rate	NS		50.50% ( $\pm 8.24$ )	
Aligned reads $\geq$ mapQ20	90.22% ( $\pm 4.53$ )		93.23% ( $\pm 0.35$ )	
Aligned reads $\geq$ mapQ30	88.98% ( $\pm 4.65$ )		91.52% ( $\pm 0.43$ )	
Capture efficiency	76.50% ( $\pm 0.50$ )		81.08% ( $\pm 7.03$ )	
Mean target coverage	159 $\pm$ 37		125 $\pm$ 53	
Not sequenced target regions	2.67% ( $\pm 0.55$ )		2.35% ( $\pm 1.2$ )	
Target regions covered $> 20\times$	88.67% ( $\pm 3.53$ )		90.08% ( $\pm 6.04$ )	
Target regions covered $> 30\times$	84.21% ( $\pm 4.87$ )		85.85% ( $\pm 11.74$ )	
Variant calling stats	NextGene	Ion Reporter	NextGene	Ion Reporter
Total (filtered)	214 ( $\pm 38$ )	203 ( $\pm 16$ )	99 ( $\pm 8$ )	169 ( $\pm 14$ )
Indels	77.3% ( $\pm 3.3$ )	26.3% ( $\pm 5.1$ )	58.7% ( $\pm 3.3$ )	7.5% ( $\pm 1.2$ )
Known SNV (dbSnp144)	83.0% ( $\pm 7.2$ )	97.3% ( $\pm 1.1$ )	94.3% ( $\pm 6.4$ )	97.8% ( $\pm 1.1$ )
Ti/Tv	2.66 ( $\pm 0.83$ )	3.17 ( $\pm 0.33$ )	3.74 ( $\pm 1.85$ )	2.85 ( $\pm 0.26$ )
Mutation identification	NextGene	Ion Reporter	NextGene	Ion Reporter
SNV (n = 15)	10	12	11	13
Indel (n = 4)	0	3	0	1
CNV (n = 2)	0	0	0	0
Total (n = 21)	10 (48%)	15 (71%)	11 (52%)	14 (67%)

**Table 2**

List of the mutations included in the present study and the results obtained with the different combinations of capture technique and analysis software, during the blind phase.

Patient	Gene	Nucleotide change	Reference sequence	Protein effect	Haloplex		SeqCap EZ	
					NextGene	Ion Reporter	NextGene	Ion Reporter
1	CDKL5	c.-162-2A>G (hemizygous)	NM_003159.2	—	F	F	F	F
2	KCNQ2	Deletion of exons 16 and 17	NM_172107.2	—	NF	NF	NF	NF
3	NHLRC	c.755C>G (homozygous)	NM_198586.2	p.Arg252Pro	NF	F	NF	F
4#	CHRNA2	c.845T>G	NM_000748.2	p.Leu282Arg	F	F	F	F
5	KCNQ3	c.950T>C	NM_004519.2	p.Ile317Pro	F	F	F	F
6	Unknown	Unknown	—	—	NF	NF	NF	NF
7	MECP2	c.397C>T	NM_001110792.1	p.Arg133Cys	F	F	F	F
8	KCNQ2	c.2599_2603dup	NM_172107.2	—	NF	F	NF	NF
9	CDKL5	c.1782T>G	NM_003159.2	p.Tyr549*	F	F	F	F
10	GRIN2A	Deletion of exon 1 and 2	NM_000833.3	—	NF	NF	NF	NF
11	STXBPI	c.1651C>T	NM_003165.3	p.Arg551Cys	F	F	F	F
12	FOXP1	c.647delC	NM_0055249.4	—	NF	F	NF	F
13	KCNQ2	c.1636A>G	NM_172107.2	p.Met546Val	NF	NF	NF	NF
14	Unknown	Unknown	—	—	NF	NF	NF	NF
15	SCN1A	c.602+1G>A	NM_001202435.1	Exon skipping	F	F	F	F
16	MECP2	c.916C>T	NM_001110792.1	p.Arg306Cys	NF	NF	F	F
17	STXBPI	c.703C>T	NM_003165.3	p.Arg235*	NF	NF	F	F
18	CSTB	c.202C>T	NM_000100.3	p.Arg68*	F	F	F	F
19	KCNQ2	c.1085A>G	NM_172107.2	p.Tyr362Cys	F	F	F	F
20	CDKL5	c.790G>T	NM_003159.2	p.Gly264*	NF	F	F	F
21	NHLRC1	c.205C>G	NM_198586.2	p.Pro69Ala	F	F	NF	NF
		c.161_162delinsAT	NM_198586.2	—	NF	F	NF	NF
22	MECP2	c.1157_1200del44	NM_001110792.1	—	NF	NF	NF	NF
23	Unknown	Unknown	—	—	NF	NF	NF	NF
24	GRIN2A	c.4161C>A	NM_000833.3	p.Tyr1387*	F	F	F	F

F = mutation found; NF = mutations not found; # = mutation identified during the present study in a patient with previously unknown diagnosis.

patient 17 was located in a region that was sufficiently covered with SeqCap EZ but not with Haloplex.

In the 20 patients for whom mutations had been identified by Sanger targeted sequencing prior to the present study, according to the electro-clinical presentation (Table 2), we did not find any additional potentially pathogenic variant in other genes.

Among the 4 patients with previously unknown diagnosis, a missense mutation (c.845T>G/p.Leu282Arg) of *CHRNA2* was identified in patient 4, by both SeqCap EZ and Haloplex library building techniques. Sanger sequencing for the patient and his parents showed that the mutation had occurred *de novo*. *In silico* prediction tools (SIFT, Polyphen-2, and Mutation Taster) were in favor of a deleterious effect. It was absent from the ExAC databases of control individuals. Patient 4 had frequent frontal seizures with sleep predominance, consistent with Nocturnal Frontal Lobe Epilepsy (OMIM #600513).

### 3. Discussion

In this paper, we compared two library-building techniques to develop an NGS gene panel for molecular diagnosis of common epilepsies in a clinical setting. To our knowledge, this is the first study comparing two different technologies for a diagnostic gene panel for epileptic disorders. Only rare studies have compared commercially available kits (e.g. Bodi et al., 2013). These studies compared technical characteristics but did not compare the efficiency in a diagnostic setting.

The full process was first performed on a blind basis, with no *a priori* about the genes involved and the positions and types of the mutations. To facilitate coverage analysis, we developed a user-friendly software that we named DeCovA. DeCovA displays coverage analysis and depth parameters for each gene of the panel, on a figure. It can be used individually and inserted in the patient report, to provide information about quality and technical limitations (coverage and depth) that are important for an accredited diagnostic laboratory (Fig. 2). In addition, DeCovA can also be used for group analysis to provide information about the properties and limitations of the panel during the technical validation phase (Fig. 3). In the present study, the two library-building techniques, Haloplex and SeqCap EZ, showed globally similar coverages (Figs. 2, 3), despite a globally more homogeneous coverage for SeqCap EZ.

We next compared the variant analysis between the two library-building technologies, using two different software tools: NextGene and Ion Reporter. The analysis process was performed on a blind basis. After this phase, we lifted anonymity and compared data obtained by the blind analysis and the data previously obtained by Sanger sequencing and MLPA. None of the two techniques could identify all the mutations/CNVs displayed by the 20 patients (Table 1). The capacity of detection depended on the type of mutations. Unsurprisingly, SNV had the higher detection level and the pathogenic mutation could easily be distinguished from additional benign variants. Those SNVs that were not detected by one of the techniques were located in regions that were not covered or the depth was below the threshold of 20×. We failed to identify indels with NextGene because of the very high number (mean = 151, SD = 33) of artifacts that were generated. Artifacts are known to be frequently generated by the Ion Torrent sequencing, especially in homopolymeric stretches (Bragg et al., 2013; Tarabeux et al., 2014). The use of Ion Reporter dramatically reduced the number of putative indels and improved the detection of indels.

In a previous study based on a PGM-based diagnostic panel for mutation screening in *BRCA1* and *BRCA2* genes, a simulation suggested that the minimum depth range should be increased to 100–130× for reliable indel detection, depending on the context and bioinformatics, whereas the range for SNV detection was 20–30× (Tarabeux et al., 2014). The authors suggested taking into account the recurrence rate in several patients included in each run. In addition, the novel Hi-Q chemistry and for the PGM will probably improve the sequencing of homopolymeric stretches.

In the present study, we failed to identify any of the two CNVs that were present in our patients, using CNV Tools (NextGene). Large rearrangements or CNVs, including deletion or duplications of one or more exons, have been shown to account for about 10% of all mutations for many genes. Their identification has long been a major issue in molecular genetics. This issue has been resolved thanks to optimized semi-quantitative PCR-based techniques such as MLPA or array platforms (array-CGH and SNP array) which remain the most efficient techniques for CNV diagnosis. Identification of small intragenic CNVs by NGS is not straightforward and will need further improvement to be transferred to the diagnosis field. The fact that we failed to identify these two CNVs is

likely to be due to poor coverage of the corresponding exons by both library-building techniques. Patient 2 had a deletion of the last two exons of *KCNQ2* that were incompletely covered by either technique (Fig. 3C and D). It was the same for the first two exons of *GRIN2A* that were deleted in patient 10 (data not shown).

Previous studies have demonstrated that targeted NGS improves the diagnostic yield diseases with wide clinical and genetic heterogeneity, such as monogenic epilepsies (Lemke et al., 2012; Della Mina et al., 2015). Such techniques allow for detecting mutations in genes that account for rare disorders and that are usually not studied in diagnostic laboratories. In addition, this approach broadens the electro-clinical phenotype related to mutations in a given gene. In the present study, this is illustrated by the results from patient 4. A previously unknown *de novo* mutation of *CHRNA2* was found in patient 4. This gene was not tested by Sanger sequencing in our diagnostic laboratory. The patient had a sporadic presentation evocative of nocturnal frontal lobe epilepsy, but without family history (Phillips et al. 2001; Wang et al., 2014). He also had moderate intellectual disability, which has not been reported to date in patients with mutations in this gene.

We report here our first attempt to transfer the activity of our diagnostic laboratory from classical techniques (Sanger and MLPA) to targeted NGS, using a small-size Ion Torrent NGS sequencer. The main alternative to targeted NGS, which is widely discussed nowadays, is WES. Although targeted NGS provides less information than WES and does not allow identification of novel genes involved in the pathology, it is suitable in routine diagnostic practice. It provides a short turn-around time for several decades of genes in parallel, with a better coverage than WES (Lemke et al., 2012). Sensitivity to detect mutations or CNV but also coverage and depth are parameters of major importance for a diagnostic laboratory which is in charge of the complete analysis of a list of genes because the expected requirement and level of constraint are much higher in routine diagnosis than in the research setting (Tarabeux et al., 2014).

DeCovA allows for a fast and reliable coverage analysis. Thanks to this efficient tool we could assess each kit and establish that SeqCap EZ provides a better coverage for our entire panel and a better homogeneity of coverage. A plot can be displayed for a group of samples, allowing for group or method comparisons. Gaps per gene are highlighted and a warning can be defined in order to improve the design of a gene panel and, eventually, to return results according to the rating level of the test (Type A, B, and C) of the Eurogentest Guidelines. Miya et al. (2015) suggested an alternative strategy based on a combination of different NGS techniques to improve mutation detection. In addition, CNV testing still needs the use of custom DNA array (array-CGH or SNP-array) covering all coding exons of the genes included in the panel.

Statistics: data from patient 13 were excluded from calculation for means and SD due to their poor quality. Capture efficiency is defined as the percentage of bases on bait regions, from unique aligned reads. NS: not suitable; ND: not determined; SNV = single nucleotide variation; CNV = copy-number variation; indel = insertion or deletion.

## Acknowledgments

We thank the Next-Generation Sequencing platform of Lyon University Hospital (Hospices Civils de Lyon).

## References

- Bodi, K., Perera, A.G., Adams, P.S., Bintzler, D., Dewar, K., Grove, D.S., Kieleczawa, J., Lyons, R.H., Neubert, T.A., Noll, A.C., Singh, S., Steen, R., Zianni, M., 2013. Comparison of commercially available target enrichment methods for next-generation sequencing. *J. Biomol. Tech.* 24, 73–86.
- Bragg, L.M., Stone, G., Butler, M.K., Hugenholz, P., Tyson, G.W., 2013. Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. *PLoS Comput. Biol.* 9, e1003031.
- Della Mina, E., Ciccone, R., Brustia, F., Bayindir, B., Limongelli, I., Vetro, A., Iascone, M., Pezzoli, L., Bellazzi, R., Perotti, G., De Giorgis, V., Lunghi, S., Coppola, G., Orcesi, S., Merli, P., Savasta, S., Veggioni, P., Zuffardi, O., 2015. Improving molecular diagnosis in epilepsy by a dedicated high-throughput sequencing platform. *Eur. J. Hum. Genet.* 23, 354–362.
- Dimassi, S., Labalme, A., Ville, D., Calender, A., Mignot, C., Boutry-Kryza, N., de Bellescize, J., Rivier-Ringenbach, C., Bourel-Ponchel, E., Cheillan, D., Simonet, T., Maignent, K., Rossi, M., Till, M., Ederly, P., Heron, D., Des Portes, V., Sanlaville, D., Lesca, G., 2015. Whole-exome sequencing improves the diagnosis yield in sporadic infantile spasm syndrome. *Clin. Genet.* <http://dx.doi.org/10.1111/cge.12636>.
- Hauser, W.A., Annegers, J.F., Rocca, W.A., 1996. Descriptive epidemiology of epilepsy: contributions of population-based studies from Rochester, Minnesota. *Mayo Clin. Proc.* 71, 576–586.
- Lemke, J.R., Riesch, E., Scheurenbrand, T., Schubach, M., Wilhelm, C., Steiner, I., Hansen, J., Courage, C., Gallati, S., Bürki, S., Strozzi, S., Simonetti, B.G., Grunt, S., Steinlin, M., Alber, M., Wolff, M., Klopstock, T., Prott, E.C., Lorenz, R., Spaich, C., Rona, S., Lakshminarasimhan, M., Kröll, J., Dorn, T., Krämer, G., Synofzik, M., Becker, F., Weber, Y.G., Lerche, H., Böhm, D., Biskup, S., 2012. Targeted next generation sequencing as a diagnostic tool in epileptic disorders. *Epilepsia* 53, 1387–1398.
- Lesca, G., Depienne, C., 2015. Epilepsy genetics: the ongoing revolution. *Rev. Neurol. (Paris)* 171, 539–557.
- Lesca, G., Rudolf, G., Bruneau, N., Lozovaya, N., Labalme, A., Boutry-Kryza, N., Salmi, M., Tsintsadze, T., Addis, L., Motte, J., Wright, S., Tsintsadze, V., Michel, A., Doummar, D., Lascelles, K., Strug, L., Waters, P., de Bellescize, J., Vrielynck, P., de Saint Martin, A., Ville, D., Ryvlin, P., Arzimanoglou, A., Hirsch, E., Vincent, A., Pal, D., Burnashev, N., Sanlaville, D., Szepietowski, P., 2013. *GRIN2A* mutations in acquired epileptic aphasia and related childhood focal epilepsies and encephalopathies with speech and language dysfunction. *Nat. Genet.* 45, 1061–1066.
- Martin, H.C., Kim, G.E., Pagnamenta, A.T., Murakami, Y., Carvill, G.L., Meyer, E., Copley, R.R., Rimmer, A., Barcia, G., Fleming, M.R., Kronengold, J., Brown, M.R., Hudspeth, K.A., Broxholme, J., Kanapin, A., Cazier, J.B., Kinoshita, T., Nabbout, R., WGS500 Consortium, Bentley, D., McVean, G., Heavin, S., Zaiwalla, Z., McShane, T., Mefford, H.C., Shears, D., Stewart, H., Kurian, M.A., Scheffer, I.E., Blair, E., Donnelly, P., Kaczmarek, L.K., Taylor, J.C., 2014. Clinical whole-genome sequencing in severe early-onset epilepsy reveals new genes and improves molecular diagnosis. *Hum. Mol. Genet.* 23, 3200–3211.
- Miya, F., Kato, M., Shiohama, T., Okamoto, N., Saitoh, S., Yamasaki, M., Shigemizu, D., Abe, T., Morizono, T., Boroevich, K.A., Kosaki, K., Kanemura, Y., Tsunoda, T., 2015. A combination of targeted enrichment methodologies for whole-exome sequencing reveals novel pathogenic mutations. *Sci. Rep.* 5, 9331.
- Phillips, H.A., Favre, I., Kirkpatrick, M., Zuberi, S.M., Goudie, D., Heron, S.E., Scheffer, I.E., Sutherland, G.R., Berkovic, S.F., Bertrand, D., Mulley, J.C., 2001. *CHRNA2* is the second acetylcholine receptor subunit associated with autosomal dominant nocturnal frontal lobe epilepsy. *Am. J. Hum. Genet.* 68, 225–231.
- Tarabeux, J., Zeitouni, B., Moncoutier, V., Tenreiro, H., Abidallah, K., Lair, S., Legoix-Né, P., Leroy, Q., Rouleau, E., Golmard, L., Barillot, E., Stern, M.H., Rio-Frio, T., Stoppa-Lyonnet, D., Houdayer, C., 2014. Streamlined ion torrent PGM-based diagnostics: *BRCA1* and *BRCA2* genes as a model. *Eur. J. Hum. Genet.* 22, 535–541.
- Thevenon, J., Milh, M., Feillet, F., St-Onge, J., Duffourd, Y., Jugé, C., Roubertie, A., Héron, D., Mignot, C., Raffo, E., Isidor, B., Wahlen, S., Sanlaville, D., Villeneuve, N., Darmency-Stamboul, V., Toutain, A., Lefebvre, M., Chouchane, M., Huet, F., Lafon, A., de Saint Martin, A., Lesca, G., El Chehadah, S., Thauvin-Robinet, C., Masurel-Paulet, A., Odent, S., Villard, L., Philippe, C., Faivre, L., Rivière, J.B., 2014. Mutations in *SLC13A5* cause autosomal-recessive epileptic encephalopathy with seizure onset in the first days of life. *Am. J. Hum. Genet.* 95, 113–120.
- Wang, M.Y., Liu, X.Z., Wang, J., Wu, L.W., 2014. A novel mutation of the nicotinic acetylcholine receptor gene *CHRNA4* in a Chinese patient with non-familial nocturnal frontal lobe epilepsy. *Epilepsy Res.* 108, 1927–1931.