



OPEN

Sanger sequencing is no longer always necessary based on a single-center validation of 1109 NGS variants in 825 clinical exomes

A. Arteché-López^{1,2}, A. Ávila-Fernández¹, R. Romero¹, R. Riveiro-Álvarez¹, M. A. López-Martínez¹, A. Giménez-Pardo¹, C. Vélez-Monsalve¹, J. Gallego-Merlo¹, I. García-Vara¹, Berta Almoguera¹, A. Bustamante-Aragonés¹, F. Blanco-Kelly¹, S. Tahsin-Swafiri¹, E. Rodríguez-Pinilla¹, P. Minguez¹, I. Lorda¹, M. J. Trujillo-Tiebas¹ & C. Ayuso¹✉

Despite the improved accuracy of next-generation sequencing (NGS), it is widely accepted that variants need to be validated using Sanger sequencing before reporting. Validation of all NGS variants considerably increases the turnaround time and costs of clinical diagnosis. We comprehensively assessed this need in 1109 variants from 825 clinical exomes, the largest sample set to date assessed using Illumina chemistry reported. With a concordance of 100%, we conclude that Sanger sequencing can be very useful as an internal quality control, but not so much as a verification method for high-quality single-nucleotide and small insertion/deletions variants. Laboratories might validate and establish their own thresholds before discontinuing Sanger confirmation studies. We also expand and validate 23 copy number variations detected by exome sequencing in 20 samples, observing a concordance of 95.65% (22/23).

Next-generation sequencing (NGS) enables multiple genes to be analyzed simultaneously and cost-effectively, thus making it a widely applied and useful tool for the diagnosis of rare genetic disorders^{1–3}. However, it is widely accepted that NGS variants need to be validated with the gold standard Sanger sequencing technique prior to reporting, even though both the costs and turnaround time of this approach are considerable. Several studies have reported that the validation of high-quality NGS variants adds little value, although most have been performed with targeted gene panels and small sample sets^{4–9}. To our knowledge, only 3 studies have validated exome data^{10,11}, including the recent extended validation of 845 variants in 383 genes from 342 whole exomes¹².

Here, we present our experience in a single-center study in which we extensively assessed the need to further validate all NGS variants with an alternative technique in 1109 variants of 425 genes from 825 clinical exome samples. We also demonstrate the importance of appropriate internal quality controls to detect inevitable human errors and thus improve the reliability of laboratory results, and expand the validation of copy number variations (CNVs) detected by clinical exome panels.

Methods

Patients and samples. All patients (N = 825) included in this study were referred to the Clinical Genetics Service of Hospital Universitario Fundación Jiménez Díaz (HU-FJD) with the suspicion of a genetic disease. Written informed consent was obtained from each participant in accordance with institutional requirements. The study was reviewed and approved by the Research Ethics Committee of HU-FJD and fulfilled the principles of the Declaration of Helsinki.

¹Department of Genetics, Health Research Institute–Jimenez Diaz Foundation University Hospital (IIS-FJD), Avda. Reyes Católicos 2, 28040 Madrid, Spain. ²Department of Genetics, University Hospital, 12 de Octubre, Madrid, Spain. ✉email: cayuso@fjd.es

Whole peripheral blood samples were collected from probands. For the analysis of prenatal cases, amniotic fluid or chorionic villus samples were collected. Genomic DNA was extracted following standard procedures.

Clinical exome panel sequencing and data analysis. Clinical exome panels were sequenced consecutively using 2 different technologies, namely, the TruSight One Sequencing panel (TSO; *Illumina Inc, San Diego, CA, USA*) (N = 304) followed by the Clinical Exome Solution panel (CES; *Sophia Genetics*) (N = 521). These panels capture, respectively, 4813 and 4900 genes related to inherited diseases. More than 95% of the captured genes overlapped in both approaches.

Paired-end sequencing (2 × 150 bp) was carried out on a NextSeq 500 (-Illumina-). Bioinformatics analysis for the detection of single-nucleotide variants (SNVs) and small insertions/deletions (indels) in TSO samples was performed using the Burrows-Wheeler alignment (BWA) Enrichment App, which provides industry-standard alignment. Variant calling was based on the Genome Analysis Toolkit (GATK). An in-house Sophia Genetics pipeline was used for the analysis of the CES samples.

Detection of CNVs was only possible in CES samples and was performed using an in-house Sophia Genetics method based on a hidden Markov model. Specifically, the method evaluates the coverage levels of the target regions across all samples within the same sequencing run. The coverage is normalized by sample and target region, thus enabling CNV calling. A minimum of 8 samples per run is required.

Data was analyzed using a large number of virtual panels that included a wide variety of genes associated with genetic conditions, such as inherited retinal dystrophies, neuropathy, deafness, skeletal dysplasia, ataxia, myopathy, intellectual disability, and kidney disease. Variants were interpreted and classified by a team of experienced molecular geneticists following the criteria of the American College of Medical Genetics and Genomics¹³. We applied the Interactive Genomics Browser to visualize pathogenic variants, likely pathogenic variants, and variants of uncertain clinical significance in disease-related dominant genes or in recessive genes when found together with another pathogenic variant, likely pathogenic variant, or variant of uncertain significance. Variants were further validated using Sanger sequencing prior to reporting.

Quality metrics. High-quality SNVs and indel variants were defined as (1) FILTER = PASS, (2) QUAL ≥ 100, (3) Depth coverage ≥ 20X, and (4) Variant fraction ≥ 20%. Variants that did not fulfil these requirements were considered low-quality variants, with the exception of 22 related-disease variants classified as pathogenic or likely pathogenic according to the ACMG criteria that strictly and although close to the established threshold, met at least two of these four requirements.

The filter PASS depends on the pipeline used and is therefore different in TSO and CES samples. According to the manufacturer's recommendations, the metrics used to filter out SNVs and indels in TSO and CES samples were, respectively, "LowGQ;LowGQX;LowMQ;LowQD;SB" and "homopolymer_region;off_target;low_coverage;low_variant_fraction". Once again, according to the manufacturer's recommendations, the CNV analysis was based on the following metrics: "coverage uniformity; median absolute deviation of read depth and off-target rate of reads". More information is available under request.

Validation of SNVs and indels detected by NGS. Mixed samples or samples affected by laboratory contamination were defined as "misleading samples". Given that they were caused by pre-analytical but not technical errors, we did not take them into consideration for the validation analysis. Low-quality variants were also excluded, as they did not meet the criteria for discontinuing Sanger sequencing.

High-quality variants were further validated. With the exception of 217 outsourced variants (170 samples), all high-quality SNVs and indel variants were bi-directionally confirmed by Sanger sequencing in our laboratory following standard procedures. Primers were designed manually or by using the ExonPrimer in silico tool. All primers were always checked in the *SNPchecker* program to avoid common SNPs within the primer design. They were also checked in the In-silico PCR tool of the UCSC Genome Browser to confirm that they were specific for the region to be amplified.

Validation of CNVs detected by NGS. All CNVs detected in CES samples were confirmed in our laboratory either by multiplex ligation-dependent probe amplification (MLPA) (N = 16) or by comparative genomic hybridization (CGH) array (60 k or 400 k custom array) (N = 5), according to the availability of the MLPA technique and the size of the CNV.

A better approximation of the CNV breakpoints could only be determinate in samples validated by CGH array.

Results

NGS variant quality and distribution. The coverage depth for the variants analyzed ranged from 20x to 361x, with a mean of 107.74x.

A total of 1109 variants in 425 genes were validated in 825 samples, including 872 SNVs, 214 indels, and 23 CNVs (Table 1). Of these, 79.6% were heterozygous (883), 17.85% homozygous (198), and 2.52% hemizygous (28).

Variants were detected either from the analysis of the TSO (395 variants) or analysis of the CES samples (714 variants) (Table 1).

NGS variant validation. Misleading samples and low-quality variants were excluded from the NGS data validation. They represented 0.7% (6/825) of the total variants analyzed (Fig. 1). Of the remaining 819 samples

Panel	Samples	Variants analyzed	No. of genes with analyzed variants	SNVs	Indels	CNVs
TSO (4813 genes)	304	395	167	316	77	0
CES (4900 genes)	521	714	258	556	137	23
Total	825	1109	425	872	214	23

Table 1. Clinical exome panels and variant distribution.

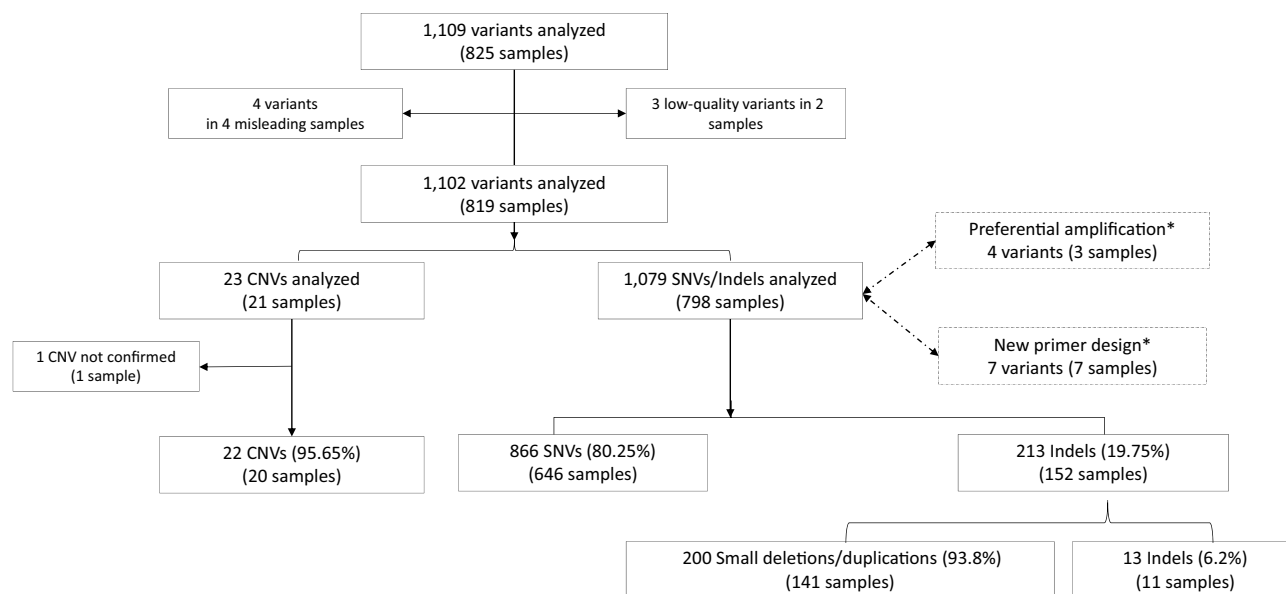


Figure 1. Validation process in the present study. (*) Samples with discordance encountered during the Sanger validation.

considered for validation, a total of 1102 variants were analyzed: 1079 SNVs/Indels in 798 samples and 23 CNVs in 21 samples (Fig. 1).

High-quality SNVs and indel variants were checked by our team of experienced geneticists and manually visualized using the Integrative Genomics Browser tool before being further confirmed by Sanger sequencing. No false-positive SNVs ($n = 866$) or false-positive indel variants ($n = 213$) were detected, thus yielding a concordance of 100%.

However, even though they all resulted in true-positive cases of the NGS, we found 3 Sanger sequencing discrepancies (see Supplementary Figures S1–S3): (a) The heterozygous variant c.1819C>T (p.Arg607Cys) in the *NOTCH3* gene (NM_000435.2) was not detected by Sanger sequencing; (b) The heterozygous variant c.1256C>T (p.Pro419Leu) in the *TPRN* gene (NM_001128228) was detected as an homozygous variant by Sanger sequencing; (c) The heterozygous variant c.489C>G (p.Ser263Arg) in the *CIQTNF5* dominant gene (NM_015645.4) was detected as an homozygous variant by Sanger sequencing. None of these samples had rare variants within the primer regions or were located in a region with homopolymers or a region with pseudogenes. Because we were able to confirm both heterozygous variants with the redesign of the *NOTCH3* and *TPRN* primers, we concluded that a preferential amplification during Sanger sequencing was the most likely explanation. Regarding the *CIQTNF5* variant, the NGS analysis of the buccal cells, together with the familial segregation of the variant, also revealed a likely preferential amplification of the mutant allele by Sanger sequencing in this sample. However, the transcription of this *CIQTNF5* gene can be bicistronic—including the upstream membrane frizzled-related protein gene (MFRP)—or monocistronic, thus preventing us from ruling out the possibility that this peculiarity might somehow influence Sanger results.

Moreover, in at least 7 different samples out of the 170 total outsourced cases, the mutant allele was not detected during the first round of Sanger sequencing. A new primer design and a second round of Sanger sequencing were required to validate these samples.

Overall, we recorded problems in 10 Sanger sequences: 3 discrepancies were most likely due to preferential amplifications, and 7 discordances with the outsourced samples were due to problems with the specificity of primers or with the PCR conditions.

CNV validation. During CNV validation, we found a heterozygous deletion of approximately 18 kb (chr12:88454782–88472858), presumably spanning exons 40 to 46 of the *CEP290* gene (NM_025114), that could not be further confirmed by a 400 k custom CGH-array including 12 probes covering the deleted region. MLPA

ID	Sex	CNV	Estimated size (Kb) by NGS	Zygoty	Minimal involved regions: Genes	Confirmed by
1	Male	Deletion	2.91	Het	Exons 8 to 10: <i>KCNH2</i> (NM_172057.2)	MLPA
2	Male	Deletion	3.31	Homo	Exons 7 and 8: <i>CLN3</i> (NM_001042432.1)	MLPA
3	Male	Deletion	4.32	Het	Complete gene (NM_019105): <i>TNXB</i> (NM_019105)	MLPA
4	Female	Duplication	10.65	Het	Complete gene: <i>FBN1</i> (NM_000138.4)	MLPA
5	Male	Deletion	12.56	Het	Exons 7 to 12: <i>LDLR</i> (NM_001195798.2)	MLPA
6	Female	Deletion	18.08	Het	Exons 40 to 46: <i>CEP290</i> (NM_025114.4)	Not confirmed by 400 k custom array
7	Female	Deletion	20.03	Het	Exons 11 to 16: <i>NIPBL</i> (NM_015384.5)	MLPA
8	Male	Deletion	104.90	Het	Exons 22 to 29: <i>USH2A</i> (NM_206933)	MLPA
9	Female	Deletion	109.28	Het	Exons 3 and 4: <i>MECP2</i> (NM_004992)	MLPA
10	Male	Deletion	117.39	Het	Exons 39 to 47: <i>USH2A</i> (NM_206933)	MLPA
11	Female	Deletion	206.66	Homo	Complete gene and exons 7 to 13 of: <i>STRC</i> (NM_153700) - <i>CASTPER2</i> (NM_172095.4)	MLPA
12	Male	Deletion	207.38	Het	Exon 1 to 9: <i>ITGB4</i> (NM_00213.5)	MLPA
13	Female	Deletion	224.40	Het	Exons 2 to 13: <i>PRPF31</i> (NM_015629.3)	MLPA
14	Male	Deletion	267.78	Hem	Exon 5: <i>RP2</i> (NM_006915.2)	MLPA
15	Female	Deletion	273.71	Het	Exons 3 to 10: <i>ANKRD11</i> (NM_013275.5)	a-CGH 60 k
16	Male	Deletion	449.14	Het	Exons 9 to 24: <i>DSP</i> (NM_004415.4)	MLPA
17	Male	Duplication	557.45	Het	chr10:102,957,798–103,443,479: <i>LBX1</i> , <i>BTRC</i> , <i>POLL</i> , <i>DPCD</i> , <i>FBXW4</i>	a-CGH 60 k
18	Female	Deletion	1982.87	Het	Complete gene: <i>ABCC6</i> (NM_001171)	MLPA
19	Male	Deletion	2050.09	Het	Complete gene: <i>STS</i> (NM_00351.5)	MLPA
20	Male	Duplication	2064.91	Het	chr17:34,816,424–36,207,539: <i>LHX1</i> , <i>ACACA</i> , <i>HNF1B</i>	a-CGH 60 k
21	Male	Deletion	3509.57	Het	chr22:19,438,161–19,956,753: <i>UFD1L</i> , <i>GP1BB</i> , <i>TBX1</i> , <i>GNB1L</i> , <i>TXNRD2</i> , <i>COMT</i>	a-CGH 60 k

Table 2. CNVs analyzed in this study. *Het* heterozygous, *Homo* homozygous, *Hemi* hemizygous.

was not available. Considering this as a false positive, we had a concordance of 95.65% (22/23) during the CNV validation (Table 2).

Of the remaining 22 validated CNVs in 20 samples, all but 2 were detected in different genes and all but 3 were heterozygous CNVs. Only 2 duplications were detected. The minimal and maximal CNV size detected by NGS in our samples was 2.91 Kb and 3.6 Mb respectively, with a medium of 207.02 Kb.

Discussion

Validation data from various targeted panels have already been reported in several studies, with increasingly large sample sets^{6–8} (see summary Table 3). In 2013, Sikkema-Raddatz et al.⁷ evaluated 155 SNVs and 13 indels from 48 genes in 84 individuals. In 2015, Baudhuin et al.⁸ validated 797 SNVs and 122 unique indel variants from 117 captured genes in 84 patients, and, one year later in 2016, Mu et al.⁶ analyzed 6912 SNVs and 933 indels from 47 genes. Larger panels capturing 568 and 4813 genes have also been validated in 200 and 100 samples, respectively⁴, in which, excluding 4 low-quality variants and/or sample mix-ups, discrepancies were not observed. Interestingly, all of the abovementioned studies reported a concordance of 100% for high-quality NGS variants, concluding that Sanger sequencing is unnecessary in such cases.

Validation data from exomes—in which lower-quality parameters are expected owing to the higher number of genes captured—have also been reported. Beck et al.¹⁰ evaluated over 5800 variants from 19 genes in 684 exomes, from which only 2 low-quality variants were discrepant. In 2014, Strom et al.¹¹ validated 94 unique variants from 110 genes in 144 exomes, again with only 1 discordant low-quality variant. Recently, Zheng et al.¹² extended this validation to 342 whole exomes and 4848 targeted panels, of which 3911 were analyzed using large Mendelian disease panels (1130 and 2182 genes). To our knowledge, this was the first study to validate exomes using BGI technology. Nevertheless, once again, all high-quality variants were further confirmed (100%).

In this study, we assessed the validation of the largest sample set of clinical exomes using Illumina technology. Remarkably, we found 3 discrepancies (3/789; 0.3%), all of which were resolved as false negatives of Sanger sequencing, thus demonstrating that this technique might not always be the best approach for confirming NGS variants, since, under specific conditions, the probability of detecting a false positive by NGS seems to be lower than that of detecting a false negative by Sanger sequencing¹¹. Additionally, Sanger sequencing had to be repeated in 7 samples, leading to unnecessary consumption of time and money. Regardless of these drawbacks, we confirmed all of the 866 SNVs and 213 indel variants we had defined and filtered out as high-quality. Therefore, and consistent with these results, we decided to discontinue the validation of these variants, thus significantly reducing costs and turnaround times. Remarkably, this decision did not apply to variants that generated uncertainty for our geneticists by being associated either with the region where the variant was detected (i.e. homopolymers, pseudogenes) or with the nomenclature of the indel variant. Similarly, it did not apply to prenatal cases, where a double check is always performed, or to low-quality variants associated with the clinical suspicion, where validation is also recommended to decrease false-negative diagnosis. In this study, we do not evaluate the rate of false

References	Targeted panel or exome analyzed	Analyzed patients	Number of Sanger analyzed variants	Variants quality conditions for which 100% concordance is obtained
Sikkema-Raddatz et al. ⁷	48-genes targeted panel (SureSelect)	84	168 (155 SNVs, 13 indels)	Depth coverage > 30× Het ratio > 20%
Baudhuin et al. ⁸	117-genes targeted panel(SureSelect)	84	919 (797 SNVs, 122 indels)	Depth coverage ≥ 100×, Variant quality score > 20 Flanking region mean base quality score > 15
Mu et al. ⁶	49-genes targeted panel	20,000	7845 (6912 SNVs, 933 indels)	Depth coverage > 100× Het ratio > 40%
Nelson et al. ⁴	568-genes inherited disorder panel (Sure Select) 4813- genesMendelian disease panel (Tru Sight One)	200 100	296 (255 SNVs, 41 indels)	PASS Depth coverage > 20× GQ score > 300 VAF 0.3–0.6 for het or > 0.9 or hom/hem
Beck et al. ¹⁰	Clinical exome (SureSelect All Exon/ TruSeq)	684	5800 variants	MPG score > 10
<i>Strom</i>	Clinical exome (SureSelect All Exon)	144	110 SNVs (94 unique SNVs)	Quality score ≥ Q500
Zheng et al. ¹²	Whole exome (BGI)	342	855 (828 SNVs, 27 indels)	Depth coverage ≥ 35× Het ratio ≥ 35%
	1130 and 2181 targeted panel (SeqCap EZ)	3569	4777 (3684 SNVs, 1093 indels)	
	115 and 313-gene targeted panel (SeqCap EZ)	1279	1969 (1444 SNVs, 525 indels)	
		5190	7601 (5956 SNVs, 1645 indels)	
Current Study	167 (Tru Sight One)	304	393 (316 SNVs, 77 indels)	PASS QUAL ≥ 100 Depth coverage ≥ 20×, Variant fraction ≥ 20%
	258 (Clinical Exome Solution)	521	693 (556 SNVs, 137 indels)	
		825	1086 (872 SNVs, 214 indels)	

Table 3. Summary of the NGS validation studies. *VAF* variant allele frequency, *Het* heterozygous, *Hom* homozygous, *Hem* Hemizygous, *SNV* single nucleotide variants, *MPG* most probable genotype.

negative results in NGS. We have found 22 variants that, although didn't strictly meet all the established quality parameters, were further confirmed by Sanger Sequencing for being pathogenic or likely pathogenic related-disease variants, supporting the importance of lowering the quality thresholds of these variants to decrease false-negative NGS results. Undoubtedly, the use of more sophisticated and machine-learning approaches would help to reduce these errors. Further studies are required to evaluate and establish the most appropriate quality threshold to obtain the best sensitivity of NGS in the clinical practice.

To our knowledge, there are no specific NGS guidelines that establish the quality parameters required for a variant to be considered high-quality, as this depends mainly on issues associated with the individual laboratory and on the NGS approach used¹. For clinical exome panels, and based on our experience, we recommended the following: (1) FILTER = PASS, (2) QUAL ≥ 100, (3) depth coverage ≥ 20X, and (4) variant fraction ≥ 20%. However, laboratories should check and establish their thresholds before discontinuing the confirmation analysis.

In line with our results, Sanger sequencing is very useful as an internal quality control, as reflected by the 4 misleading samples we would not have detected otherwise. Unfortunately, even with high-quality measures, a laboratory cannot detect 100% of human errors. However, we believe these errors can be significantly minimized using proper controls such as the use of commercial sample tracking kits. Moreover, we propose the controls already implemented in our laboratory. First, the proband sample is always used as a positive control during segregation analysis. Because this is not possible for samples coming from external laboratories/hospitals, all external reported variants are further validated. Second, within the same sequencing run, we ensure that Sanger sequencing is performed in at least 4 different samples (at least 1 sample per pool). And third, the sex of each clinical exome analyzed is always checked prior to reporting, thus enabling us to detect at least 50% of errors. A counter to this measure is the detection (incidentally or not) of sex chromosome aneuploidy or a sex discrepancy due to an allogenic transplant in the patient. Urine or buccal cells, but not peripheral blood samples, would be the recommended samples to be analyzed in this last case.

To date, several studies have validated CNVs detected by targeted NGS panels^{14–16} and whole exomes¹⁷, with a wide variation in the false negatives and false positives reported in the analyses^{16, 18}. Undoubtedly, their recent implementation in clinical practice has considerably increased our diagnostic yield, with the result that it is important to implement their detection during NGS. Here, we expand the validation of 22 CNVs detected by CES in 20 different samples. Our approach revealed wide variability in the CNV size detected within our cohort. Interestingly, one of the smallest CNVs detected (3.31 Kb) was homozygous, reflecting that a CNV is easily detected when the normal allele is not present. Interestingly, with the clinical exome approach, it is not possible to accurately establish the CNV detection limit of our NGS approach (CES), as this depends not only on the number of reads per sample, but also on the region and, specifically, on the number of probes within that specific region. In this study, it was not possible to accurately define the CNV breakpoints or size of most samples, nor was it possible to evaluate the incidence of false negatives within our samples.

We found a false-positive CNV in the *CEP290* gene, that is, a concordance of 95.65% (22/23); however, considering the small sample set analyzed, few conclusions can be drawn. Based on our experience, and despite the need for further validation, we believe that if a CNV is clearly related to the clinical suspicion, then it is very likely to be real. All suspected CNVs should be further validated with additional technology, which, in some

cases, would enable us to determine specific breaking points. This is not always possible with exome sequencing technology. Change may come about with advances in genome sequencing techniques, as this robust approach makes it possible to detect specific CNVs¹⁹. However, many more large-scale validation studies are still needed.

Conclusions

NGS is a robust technique for the detection of SNVs and indels that obviates the need for additional validation. Laboratories should establish their own quality thresholds before deciding to discontinue confirmation studies, and should implement internal quality controls in order to prevent human error and obtain reliable results. Our study enabled us to expand the validation data of 22 CNVs detected by NGS and highlight the importance of its implementation. Nevertheless, further validation of CNVs continues to be necessary.

Received: 23 June 2020; Accepted: 22 February 2021

Published online: 11 March 2021

References

- Matthijs, G. *et al.* Guidelines for diagnostic next-generation sequencing. *Eur. J. Hum. Genet.* **24**, 2–5 (2016).
- Lee, H. *et al.* Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA* **312**, 1880–1887 (2014).
- Yang, Y. *et al.* Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N. Engl. J. Med.* **369**, 1502–1511 (2013).
- Nelson, A. C. *et al.* Criteria for clinical reporting of variants from a broad target capture NGS assay without Sanger verification. *JSM Biomark.* **2**, 1005 (2015).
- Yohe, S. *et al.* Clinical validation of targeted next-generation sequencing for inherited disorders. *Arch. Pathol. Lab. Med.* **139**, 204–210 (2015).
- Mu, W., Lu, H.-M., Chen, J., Li, S. & Elliott, A. M. Sanger confirmation is required to achieve optimal sensitivity and specificity in next-generation sequencing panel testing. *J. Mol. Diagn.* **18**, 923–932 (2016).
- Sikkema-Raddatz, B. *et al.* Targeted next-generation sequencing can replace sanger sequencing in clinical diagnostics. *Hum. Mutat.* **34**, 1035–1042 (2013).
- Baudhuin, L. M. *et al.* Confirming variants in next-generation sequencing panel testing by sanger sequencing. *J. Mol. Diagn.* **17**, 456–461 (2015).
- McCourt, C. M. *et al.* Validation of next generation sequencing technologies in comparison to current diagnostic gold standards for BRAF, EGFR and KRAS mutational analysis. *PLoS ONE* **8**, e69604 (2013).
- Beck, T. F. *et al.* Systematic evaluation of Sanger validation of next-generation sequencing variants. *Clin. Chem.* **62**, 647–654 (2016).
- Strom, S. P. *et al.* Assessing the necessity of confirmatory testing for exome-sequencing results in a clinical molecular diagnostic laboratory. *Genet. Med.* **16**, 510–515 (2014).
- Zheng, J. *et al.* A comprehensive assessment of next-generation sequencing variants validation using a secondary technology. *Mol. Genet. Genomic Med.* **7**, e748 (2019).
- Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–423 (2015).
- Yao, R., Yu, T., Qing, Y., Wang, J. & Shen, Y. Evaluation of copy number variant detection from panel-based next-generation sequencing data. *Mol. Genet. Genomic Med.* **7**, e00513 (2019).
- Kerkhof, J. *et al.* Clinical validation of copy number variant detection from targeted next-generation sequencing panels. *J. Mol. Diagn.* **19**, 905–920 (2017).
- Ellingford, J. M. *et al.* Validation of copy number variation analysis for next-generation sequencing diagnostics. *Eur. J. Hum. Genet.* **25**, 719–724 (2017).
- Rajagopalan, R., Murrell, J. R., Luo, M. & Conlin, L. K. A highly sensitive and specific workflow for detecting rare copy-number variants from exome sequencing data. *Genome Med.* **12**, 14 (2020).
- Pfundt, R. *et al.* Detection of clinically relevant copy-number variants by exome sequencing in a large cohort of genetic disorders. *Genet. Med.* **19**, 667–675 (2017).
- Gross, A. M. *et al.* Copy-number variants in clinical genome sequencing: deployment and interpretation for rare and undiagnosed disease. *Genet. Med.* **21**, 1121–1130 (2019).

Acknowledgements

This work was supported by grants from the Instituto de Salud Carlos III (ISCIII) of the Spanish Ministry of Health, including CIBERER (06/07/0036) and IIS-FJD Biobank 14 PT13/0010/0012, and from the regional government of Madrid, RARE Genomics-CM (CAM, B2017/BMD-3721), all partially supported by FEDER (European Regional Development Fund). The University Chair UAM-IIS-FJD of Genomic Medicine and the Ramon Areces Foundation also supported our work. The authors would also like to thank to the clinical staff of University Hospital Foundation Jimenez Diaz, who made this study possible. We are grateful to Thomas O’Boyle for the English review of this article.

Author contributions

Wrote the first draft of the manuscript: A.A.L. Contribution to data preparation and acquisition: M.A.L.M., A.G.P., C.V., J.G., I.G. R.R., P.M.P. Contribution to the design of the study and analysis of data: A.A.L., A.A.F., C.A. Critical review of the manuscript for important intellectual content: A.A.F., R.R.A., M.J.T., B.A., I.L., C.A. Agreement with manuscript results and conclusions: All authors. Supervision: C.A. All the authors have read the manuscript and agree on publication.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-85182-w>.

Correspondence and requests for materials should be addressed to C.A.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021