Assessing the necessity of confirmatory testing for exome-sequencing results in a clinical molecular diagnostic laboratory

Samuel P. Strom, PhD^{1,2}, Hane Lee, PhD¹, Kingshuk Das, MD¹, Eric Vilain, MD, PhD^{2,3}, Stanley F. Nelson, MD^{1,2}, Wayne W. Grody, MD, PhD^{1,2,3} and Joshua L. Deignan, PhD¹

Purpose: Sanger sequencing is currently considered the gold standard methodology for clinical molecular diagnostic testing. However, next-generation sequencing has already emerged as a much more efficient means to identify genetic variants within gene panels, the exome, or the genome. We sought to assess the accuracy of next-generation sequencing variant identification in our clinical genomics laboratory with the goal of establishing a quality score threshold for confirmatory Sanger-based testing.

Methods: Confirmation data for reported results from 144 sequential clinical exome-sequencing cases (94 unique variants) and an additional set of 16 variants from comparable research samples were analyzed.

Results: Of the 110 total single-nucleotide variants analyzed, 103 variants had a quality score $\geq Q500$, 103 (100%) of which were confirmed by Sanger sequencing. Of the remaining seven variants with

quality scores $< Q500$, six were confirmed by Sanger sequencing (85%).

Conclusion: For single-nucleotide variants, we predict that going forward, we will be able to reduce our Sanger confirmation workload by 70–80%. This serves as a proof of principle that as long as sufficient validation and quality control measures are implemented, the volume of Sanger confirmation can be reduced, alleviating a significant amount of the labor and cost burden on clinical laboratories wishing to use next-generation sequencing technology. However, Sanger confirmation of low-quality single-nucleotide variants and all insertions or deletions < 10 bp remains necessary at this time in our laboratory.

Genet Med advance online publication 9 January 2014

Key Words: exome; molecular diagnostics; next-generation sequencing; variant confirmation

Next-generation sequencing (NGS) technologies require probabilistic algorithms for the conversion of uniquely aligned short sequence reads into genotypes. These algorithms are sensitive to multiple sources of error, including sequencing errors, incorrect alignment (“mismatching”), and random sampling.^{1–8} False-positive results due to sequencing errors are particularly prevalent when read depth is below 10 reads per base on average (“10× coverage” by convention).³ Due to this uncertainty, amplification-based dye terminator dideoxy DNA (“Sanger”) sequencing has been used routinely to confirm NGS results.^{9–16}

However, as read depth increases and additional samples are tested using a consistent experimental protocol and analytical pipeline, more information is available to interrogate the validity of a given variant call. Valuable data, in addition to the count of reference and nonreference (“variant”) nucleotides observed at a given position, are amassed. These data include mapping quality, strand origin, base call quality, position of the variant within a sequence read, haplotype information, and cross-sample comparisons. The commonly used genotype-calling pipeline using the Genome Analysis Toolkit^{1,17} implements a Bayesian genotype likelihood model (based on known polymorphic loci such as dbSNP variants) and variant quality score

recalibration to estimate posterior probabilities for each variant call (with `hapmap_3.3.b37.sites` and `1000G_omni2.5.b37.sites` for training resources).

Although technically these final quality scores (“Qscores”, or “QN”, where N is a value greater than zero) are reported as log-scaled probabilities, comparison across experiment types is not advisable because of the large degree of variability of data volume, data quality, and options between NGS analytical pipelines. In this study, Qscores are considered to be relative measures and are compared only between clinical exome-sequencing (CES) data sets from the end-to-end analytically validated procedures established in the University of California, Los Angeles (UCLA) Clinical Genomics Center, which is part of the UCLA Molecular Diagnostics Laboratories (accredited by both the Clinical Laboratory Improvement Amendments (CLIA) and the College of American Pathologists).

For variants with high quality scores ($> Q10,000$) and high coverage ($> 100\times$), the amount of information supporting the genotype call is overwhelming. For such variants, failure to replicate the finding by Sanger sequencing is highly indicative of human error (e.g., sample swap). Thus, for high-quality NGS variants, Sanger confirmation serves almost exclusively as a

¹Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA; ²Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA; ³Department of Pediatrics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA. Correspondence: Samuel P. Strom (sstrom@mednet.ucla.edu)

Submitted 10 July 2013; accepted 18 October 2013; advance online publication 9 January 2014. doi:[10.1038/gim.2013.183](https://doi.org/10.1038/gim.2013.183)

sample quality control measure. Therefore, it is the goal of this study to establish a conservative internal quality score cutoff, above which Sanger confirmation of CES-identified variants will no longer be a necessary quality control measure in our laboratory.

MATERIALS AND METHODS

CES procedure

Exome sequencing was performed in the UCLA Clinical Genomics Center (<http://pathology.ucla.edu/genomics>) following validated protocols. Briefly, high-molecular-weight genomic DNA was isolated from whole blood collected in a lavender top tube (ethylene diamine tetraacetic acid dipotassium (K_2 EDTA) or ethylene diamine tetraacetic acid tripotassium (K_3 EDTA)) using a QIAcube (Qiagen, Venlo, The Netherlands). For all of the clinical samples, exome sequencing was performed using the Agilent SureSelect Human All Exon 50mb (Agilent Technologies, Santa Clara, CA) for exome capture and Illumina HiSeq2000 (Illumina, San Diego, CA) for sequencing as 50-bp paired-end runs using V3 chemistry. For the nonclinical samples, Agilent SureSelect Human All Exon 50mb XT kit (V2) was used for exome capture and Illumina HiSeq2000 for sequencing as 100-bp paired-end runs using V3 chemistry.

Data analysis was performed using the analytical pipeline implemented and validated for CES in the UCLA Clinical Genomics Center. All sequence reads were aligned to the human reference genome (Human GRCh37/hg19) using Novoalign (Novocraft, Selangor, Malaysia). Polymerase chain reaction (PCR) duplicates were marked by Picard (<http://picard.sourceforge.net/>), and Genome Analysis Toolkit (Broad Institute, Cambridge, MA) was used to realign indels, recalibrate the quality scores, call, filter, recalibrate, and evaluate the variants. All variants called across the protein-coding regions and the flanking junctions were annotated using Variant Annotator X, an in-house MySQL database using data from the publicly available Ensembl Variant Effect Predictor (European Bioinformatics Institute, Hinxton, UK).¹⁸ A detailed description of the bioinformatic methods used to analyze these data is presented in **Supplementary Materials** online.

Several steps are taken to reduce the probability of sample swap errors in our laboratory, including the following: (i) assays are performed by appropriately licensed technologists with experience in NGS workflows; (ii) at least two unique identifiers are used to label all reaction vessels and worksheets at all preanalytical stages; and (iii) samples are alternated by gender. In addition, when related individuals are tested as part of a trio, Mendelian errors are analyzed by counting the number of inconsistent genotypes. For instance, from internal experience, the proband should not have more than five *de novo* amino acid-altering rare variants, and approximately half of the heterozygous variants present in the proband should be inherited from the mother and the other half from the father. Moreover, when available, previous genetic testing results (such as variants described in clinical reports from

individual gene assessments or regions of homozygosity from chromosomal microarray analyses) are cross-referenced with the CES data also. Some additional steps that laboratories could use to reduce sample errors include running samples in duplicate, running the CES assay in parallel with a single-nucleotide polymorphism array or genotyping identity panel for concordance analysis (if not previously performed as mentioned earlier), or spiking the blood sample with a unique plasmid during extraction and confirming that the plasmid sequence occurs in the final result.

Variant selection

All clinically reported variants, both clinically significant findings and variants of uncertain significance, were selected for confirmation. In total, 110 unique single-nucleotide variants (SNVs) were selected for Sanger confirmation (**Table 1**), and a subset (16 SNVs) of these was randomly selected for assessment from a pool of variants with quality scores <Q2000. These additional variants had not been clinically reported because they are not in genes that are known to cause any clinical condition.

All SNVs selected, regardless of report status, are predicted to be nonsynonymous and are rare (with an average minor allele frequency <1% in the Exome Variant Server).¹⁹

Sanger sequencing

PCR primers were designed for each target locus using the Web-based Primer3Plus software (Andreas Untergasser and Harm Nijveen).²⁰ Targets were amplified using PCR and subjected to agarose gel electrophoresis for size analysis of resulting amplicons. If no amplicon was observed, multiple amplicons were observed, or an amplicon of improper size was observed, a second independent set of PCR primers was designed and tested in a similar fashion.

Unique, properly sized amplicons were purified using standard techniques. BigDye Terminator DNA-sequencing reactions were then performed on eluted amplicons, and sequencing was conducted by automated capillary gel electrophoresis (ABI 3730, Life Technologies, Carlsbad, CA). A clinical molecular geneticist certified by the American Board of Medical Genetics manually analyzed the resulting sequence traces using Sequence Scanner (Applied Biosystems, Foster City, CA).

Cost analysis

The reagent cost per validation is estimated to be US\$20 per variant on average. The personnel cost for designing PCR primers, running the assay, analyzing the data, and interpreting the results is estimated to be US\$120. Overhead (including facilities, maintenance, instrument costs, and other considerations) contribute approximately US\$100 per test. Combined, the estimated cost of performing Sanger confirmation of a single SNV is thus approximately US\$240. These values were calculated based on standard clinical molecular genetics practices and average licensed medical technologist salaries in the UCLA Molecular Diagnostics Laboratories.

Table 1 Variant information

Gene	Zyg.	QUAL	Conf.	Depth	Alt.	VF
<i>OBSCN</i>	Het	139	Yes	11	6	54.5
<i>BEAN1</i>	Het	157	Yes	10	6	60
<i>SURF1</i>	Homo	164	Yes	5	5	100
<i>ZNHIT2</i>	Het	258	No	10	4	40
<i>CACNA1H</i>	Het	292	Yes	20	7	35
<i>SHB</i>	Homo	449	Yes	9	9	100
<i>GAA</i>	Het	475	Yes	17	10	58.8
<i>KCNT1</i>	Het	540	Yes	26	17	65.4
<i>MYLK</i>	Het	714	Yes	47	24	51.1
<i>FGFR1</i>	Het	749	Yes	42	24	57.1
<i>BEAN1</i>	Het	791	Yes	51	25	49
<i>RANBP3</i>	Het	831	Yes	48	26	54.2
<i>TGM6</i>	Het	837	Yes	57	27	47.4
<i>OTOF</i>	Het	850	Yes	28	14	50
<i>NDUFS8</i>	Homo	858	Yes	17	17	100
<i>BCOR</i>	Het	867	Yes	56	27	48.2
<i>CACNA1A</i>	Het	892	Yes	71	30	42.3
<i>TCF4</i>	Het	898	Yes	83	35	42.2
<i>OBSCN</i>	Het	938	Yes	66	31	47
<i>SYNE1</i>	Het	966	Yes	80	33	41.3
<i>MECP2</i>	Hemi	995	Yes	28	28	100
<i>FBN3</i>	Het	1,017	Yes	59	33	55.9
<i>SCML4</i>	Het	1,028	Yes	21	9	42.9
<i>WWP2</i>	Het	1,052	Yes	73	39	53.4
<i>AFG3L2</i>	Het	1,092	Yes	77	39	50.6
<i>MYH6</i>	Het	1,094	Yes	86	38	44.2
<i>COL20A1</i>	Het	1,096	Yes	28	13	46.4
<i>SCN5A</i>	Het	1,098	Yes	66	32	48.5
<i>NPAP1</i>	Het	1,149	Yes	91	40	44
<i>AGTR1</i>	Het	1,207	Yes	71	37	52.1
<i>MTATP6</i>	Homo	1,213	Yes	36	35	97.2
<i>USP21</i>	Het	1,238	Yes	90	43	47.8
<i>LMNA</i>	Het	1,296	Yes	105	43	41
<i>SLC9A2</i>	Het	1,306	Yes	30	18	60
<i>ATP8A2</i>	Het	1,324	Yes	83	44	53
<i>POLR3A</i>	Het	1,498	Yes	101	48	47.5
<i>CHRNA7</i>	Het	1,526	Yes	104	49	47.1
<i>ABCA4</i>	Het	1,533	Yes	91	48	52.7
<i>C1QTNF5</i>	Het	1,535	Yes	94	48	51.1
<i>MYH7</i>	Het	1,557	Yes	143	57	39.9
<i>CACNA1D</i>	Het	1,588	Yes	94	54	57.4
<i>COL6A2</i>	Het	1,603	Yes	112	61	54.5
<i>KDM6A</i>	Hemi	1,633	Yes	25	25	100
<i>KCNQ3</i>	Het	1,699	Yes	118	60	50.8
<i>FANCG</i>	Het	1,728	Yes	114	54	47.4
<i>SYNE1</i>	Het	1,787	Yes	113	58	51.3
<i>SLC37A1</i>	Het	1,815	Yes	134	65	48.5
<i>OPA1</i>	Het	1,848	Yes	155	61	39.4
<i>GJB2</i>	Het	1,864	Yes	138	59	42.8
<i>WFS1</i>	Het	1,891	Yes	123	61	49.6
<i>MYBPC3</i>	Het	1,952	Yes	114	59	51.8
<i>SCN8A</i>	Het	1,988	Yes	150	150	100
<i>RUNX2</i>	Het	2,055	Yes	171	74	43.3
<i>RAD21</i>	Het	2,058	Yes	150	66	44
<i>SETX</i>	Het	2,125	Yes	131	65	49.6
<i>TNXB</i>	Het	2,145	Yes	131	67	51.1
<i>RET</i>	Het	2,262	Yes	150	76	50.7

Table 1 Continued

Gene	Zyg.	QUAL	Conf.	Depth	Alt.	VF
<i>CPT2</i>	Het	2,317	Yes	98	37	37.8
<i>STAT1</i>	Het	2,367	Yes	161	74	46
<i>SYNE1</i>	Het	2,402	Yes	160	75	46.9
<i>ERCC5</i>	Het	2,497	Yes	165	80	48.5
<i>CNTNAP2</i>	Het	2,553	Yes	156	79	50.6
<i>APC</i>	Het	2,554	Yes	197	91	46.2
<i>PTPN11</i>	Het	2,641	Yes	185	80	43.2
<i>CLN8</i>	Het	2,671	Yes	109	50	45.9
<i>SYNE1</i>	Het	2,860	Yes	200	90	45
<i>OTOF</i>	Het	3,045	Yes	97	50	51.5
<i>TSHR</i>	Het	3,087	Yes	207	99	47.8
<i>G6PD</i>	Hemi	3,149	Yes	45	45	100
<i>G6PD</i>	Hemi	3,746	Yes	52	52	100
<i>SCN2A</i>	Het	4,043	Yes	227	100	44.1
<i>ZEB2</i>	Het	4,080	Yes	204	101	49.5
<i>SETX</i>	Het	4,136	Yes	228	109	47.8
<i>TTN</i>	Het	4,312	Yes	243	119	49
<i>TTN</i>	Het	4,373	Yes	243	110	45.3
<i>TTN</i>	Het	4,447	Yes	226	112	49.6
<i>SMCHD1</i>	Het	4,489	Yes	216	115	53.2
<i>ACVR1</i>	Het	4,493	Yes	219	114	52.1
<i>ITPR1</i>	Het	4,560	Yes	250	137	54.8
<i>NDUFA1</i>	Hemi	4,569	Yes	75	75	100
<i>TBC1D24</i>	Het	4,615	Yes	166	84	50.6
<i>TBC1D25</i>	Hemi	4,618	Yes	46	46	100
<i>CHD7</i>	Het	4,628	Yes	231	124	53.7
<i>RP1</i>	Het	4,709	Yes	123	60	48.8
<i>CERKL</i>	Het	4,765	Yes	116	58	50
<i>KDM6A</i>	Hemi	5,039	Yes	92	92	100
<i>MET</i>	Het	5,073	Yes	146	71	48.6
<i>GJB2</i>	Homo	5,511	Yes	141	141	100
<i>RP1</i>	Het	5,646	Yes	152	72	47.4
<i>TBC1D24</i>	Het	5,906	Yes	163	73	44.8
<i>KCNJ13</i>	Homo	6,052	Yes	95	95	100
<i>PIK3C2G</i>	Homo	6,089	Yes	69	69	100
<i>SMG5</i>	Homo	6,232	Yes	65	65	100
<i>GOSR2</i>	Het	6,357	Yes	187	79	42.2
<i>CERKL</i>	Het	6,587	Yes	180	83	46.1
<i>RAPGEF6</i>	Het	6,662	Yes	171	81	47.4
<i>USH2A</i>	Homo	6,775	Yes	180	180	100
<i>GOSR2</i>	Het	6,874	Yes	186	83	44.6
<i>AHI1</i>	Homo	7,056	Yes	175	175	100
<i>SLC35C1</i>	Homo	7,066	Yes	83	83	100
<i>OPA1</i>	Het	7,090	Yes	205	92	44.9
<i>POFUT1</i>	Homo	7,533	Yes	89	89	100
<i>GLDC</i>	Homo	7,927	Yes	104	104	100
<i>MYOM2</i>	Homo	9,051	Yes	228	228	100
<i>MAGEA1</i>	Het	10,063	Yes	179	97	54.2
<i>ATP7A</i>	Hemi	10,417	Yes	158	158	100
<i>FAM98B</i>	Homo	11,193	Yes	93	92	98.9
<i>HTR2C</i>	Hemi	11,818	Yes	186	186	100
<i>PDE6C</i>	Homo	13,249	Yes	227	227	100
<i>ALDH18A1</i>	Homo	17,244	Yes	220	220	100

Alt., number of independent reads supporting the alternate allele; Conf., variant confirmed by Sanger sequencing; Hemi, hemizygous (X-linked observed in a male); Het, heterozygous; Homo, homozygous; QUAL, quality score of clinical exome-sequencing locus; VF, variant frequency (equal to $100 \times \text{Alt./depth}$); Zyg, zygosity. Grey shading indicates QUAL <500, variant not confirmed, coverage <40x, respectively.

RESULTS

Exome-sequencing results were confirmed for 103/103 (100%) of SNVs, with quality scores $\geq Q500$ (Table 1). The coverage depth for these variants ranged from $5\times$ to $250\times$, with a mean of $116\times$. The correlation between quality score and coverage depth is positive and statistically significant ($R = 0.56$; $P < 0.0001$; Figure 1). Of the seven SNVs with quality scores $< Q500$, only one was not corroborated by the Sanger sequencing data (Table 1 and Figure 2).

From the first 144 signed-out reports, the average number of reported variants per report is ~ 1 (range: 0–5 variants). With an estimated cost of US\$240 per confirmation and a sample volume of 40 reports per month, the total cost to the laboratory performing the test is US\$9,600 per month (US\$115,200 per year). Furthermore, the number of clinically relevant (nonincidental) variants reported per case is not expected to decline over time. Instead, as more disease–gene associations are made, we expect the number of cases with at least one potentially causal variant to increase. Thus, the cost of Sanger confirmation will scale at least linearly with this increased sample volume.

Figure 1 Correlation between quality scores and depth of coverage. Individual quality scores are plotted against read depth for 110 single-nucleotide variant loci tested. Quality score threshold of Q500 is marked by a dashed gray vertical line. The correlation is positive and significant (Pearson correlation significance test, $P < 10^{-13}$).

Figure 2 Validation results sorted by quality score. Each single-nucleotide variant (SNV) tested is represented by a point, sorted by ascending quality score. Red points represent SNVs with quality scores $< Q500$ (horizontal red dashed line). Vertical red bars indicate failure to confirm.

Notably, turnaround times for reports requiring variant confirmation were delayed at least 1 week on average compared with reports with no reported variants.

DISCUSSION

For each UCLA CES test, the decision to report a variant begins with interpretation by a group of diverse experts at a Genomic Data Board. This interpretation considers the molecular genetic evidence (e.g., the effect that a DNA change is predicted to have on its corresponding protein product) as it relates directly to the primary clinical concern(s) noted by the ordering physician. At present, incidental findings are not reported. If the board decides that a variant is worthy of reporting, the laboratory then considers the technical validity of the finding. Before May 2013, Sanger sequencing was used as an alternative methodology for validation of each reported variant. Since that time, only indels and SNVs with quality scores $< Q500$ are validated by Sanger sequencing.

Because it has been considered the gold standard for more than two decades, using capillary-based Sanger sequencing for confirmation of all NGS results is a safe choice. However, taken out of context, this is highly unusual; technical confirmation of results from a validated assay using an alternative methodology before reporting is not often done for other types of molecular testing. Additionally, there are several specific reasons to suspect that Sanger confirmation of all clinically relevant SNVs detected by NGS is an unnecessarily conservative approach with significant drawbacks.

First, NGS can be sampled to generate dozens or hundreds of independent reads across a locus, whereas increased sampling of Sanger sequencing requires technical replicates. Although a Sanger sequencing peak does represent a large number of individual DNA molecules, these are clonal and arise from an unknown number of original template molecules. At heterozygous positions sequenced bidirectionally, the minimum number of original template molecules required to produce a signal is only four: forward reference, forward alternate, reverse reference, and reverse alternate. Although it is likely that a larger number of template molecules are typically amplified, it is not possible to assess or confirm this number due to the clonal nature of PCR amplification. Moreover, the error rate for a single base is relatively higher in NGS than in Sanger sequencing; however, high read depth (“coverage”) of a locus can overcome this issue.

Additionally, PCR-based amplification is susceptible to allele dropout due to cryptic variation within primer-binding sites, whereas the target enrichment techniques used in exome sequencing are not. Additionally, some genomic intervals are extremely difficult to amplify and may not yield high-quality Sanger results despite multiple attempts. Being unable to report a clinically significant variant due to a failure of the Sanger technology introduces a challenging obstacle if the NGS assay is analytically validated.

NGS variant identification is not without error. However, above a certain hypothetical quality threshold, the probability

of observing a false-positive NGS result is lower than the false-negative rate of Sanger sequencing (which itself is not perfect). This means that for variants meeting this threshold, performing Sanger sequencing is noninformative beyond sample quality control because the vast majority of results will be concordant and the remaining negative results will not be interpretable. Thus, such high-quality NGS results, when routinely obtained using a method validated by a clinical laboratory, should be considered an equally defensible “gold standard.”

The difficulty then is in determining a high-confidence quality threshold. Coverage depth is a useful guide, but probabilistic genotyping algorithms—such as those implemented within the Genome Analysis Toolkit¹⁷—provide highly informative quality scores. Because quality scores are assay specific and relative, it is not possible to calculate an a priori threshold value. Rather, based on a sample of 110 SNV confirmation tests, we have established a conservative in-house quality score threshold of Q500 (~40× coverage) for the CES test in our laboratory, above which all 103 SNVs detected were confirmed by Sanger sequencing.

Manual inspection of variant calls using a visualization tool such as the Integrative Genomics Browser^{21,22} by a genomics expert is a potential alternative to our quality score threshold approach. Although our experiences generally support this as a valid potential solution, we do not have sufficient data to broadly assess the efficacy of this approach.

Small insertions and deletions (“indels,” defined here as being <10bp long) are also detected by CES and are reported if clinically significant. At this time, we do not have sufficient data to propose a quality score threshold for confirmation of indels and will thus continue to confirm all such reported variants by the Sanger method.

The perceived benefits of performing Sanger confirmation on all NGS-detected SNVs lie in quality control and risk avoidance. These must be weighed against increased test cost, delayed turnaround time, and the potentially paralyzing failure to confirm a very high-quality variant of clinical significance. Although current professional practice guidelines recommend confirmatory testing of all clinical NGS results, they also allow for laboratories to reduce the amount of confirmatory testing performed as long as suitable validation studies have been completed.²³ Follow-up testing of identified variants in additional family members for carrier or pre-symptomatic status by Sanger sequencing is performed in our laboratory on request for an additional fee. However, in practice, this has been a rare occurrence; for the majority of our exome-sequencing cases, the original proband is the only family member tested, which also argues against the need to have a pair of Sanger sequencing primers available in the laboratory for every variant detected.

All genetic tests introduce uncertainty. At the genomic level, it is the exception, not the rule, when a causal relationship between a genetic variant and a clinical condition can be made absolutely. Thus, when counseling for CES results, the slight probability of a high-quality variant being an analytical false positive is typically a minor consideration compared

Table 2 Summary of Sanger confirmation results, split by quality score threshold of Q500

	Clinical SNVs	Additional SNVs	Total
<Q500	5/6	1/1	6/7
≥Q500	88/88	15/15	103/103
Total	93/94	16/16	109/110

SNV, single-nucleotide variant.

with the uncertainty of genotype–phenotype relationships. This argues against devoting large amounts of resources to confirmatory testing for variants of high confidence, especially when the testing laboratory is conservative in the ascertainment and reporting of “causative” variants, as ours is.

Stemming from these theoretical and practical considerations, and based on data resulting from the confirmation of 110 SNVs, our group has decided to discontinue routine Sanger confirmation of reported CES results with quality scores >Q500 (SNVs only; **Table 2**). However, other laboratories wishing to follow this paradigm must establish their own quality thresholds for each assay and provide empirical evidence to support those decisions.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at <http://www.nature.com/gim>

ACKNOWLEDGMENTS

This work was made possible by the use of data generated from consented clinical testing participants, and we thank them for this valuable contribution. Technical assistance was provided by Nora Warschaw, Traci Toy, Robert Chin, Thien Huynh, and Jean Reiss at the UCLA Molecular Diagnostics Laboratories. Variant Annotator X software was used with the permission and guidance of its author, Michael Yourshaw, and computational assistance was provided by Bret Harry.

DISCLOSURE

S.P.S., H.L., K.D., S.F.N., W.W.G., and J.L.D. are employed by a fee-for-service laboratory at UCLA performing next-generation sequencing.

REFERENCES

- 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.
- Kiezun A, Garimella K, Do R, et al. Exome sequencing and the genetic basis of complex traits. *Nat Genet* 2012;44:623–630.
- Abecasis GR, Altshuler D, Auton A, et al. A map of human genome variation from population-scale sequencing. *Nature* 2010; 467:1061–1073.
- Brockman W, Alvarez P, Young S, et al. Quality scores and SNP detection in sequencing-by-synthesis systems. *Genome Res* 2008;18:763–770.
- Li M, Nordborg M, Li LM. Adjust quality scores from alignment and improve sequencing accuracy. *Nucleic Acids Res* 2004;32:5183–5191.
- Li H, Handsaker B, Wysoker A, et al.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–2079.

7. Li H, Homer N. A survey of sequence alignment algorithms for next-generation sequencing. *Brief Bioinformatics* 2010;11:473–483.
8. Li Y, Vinckenbosch N, Tian G, et al. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat Genet* 2010;42:969–972.
9. Ng SB, Turner EH, Robertson PD, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009;461:272–276.
10. Ng SB, Buckingham KJ, Lee C, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 2010;42:30–35.
11. Sikkema-Raddatz B, Johansson LF, de Boer EN, et al. Targeted next-generation sequencing can replace Sanger sequencing in clinical diagnostics. *Hum Mutat* 2013;34:1035–1042.
12. Peng G, Fan Y, Palculict TB, et al. Rare variant detection using family-based sequencing analysis. *Proc Natl Acad Sci USA* 2013;110:3985–3990.
13. McCourt CM, McArt DG, Mills K, et al. Validation of next generation sequencing technologies in comparison to current diagnostic gold standards for BRAF, EGFR and KRAS mutational analysis. *PLoS ONE* 2013;8:e69604.
14. Dames S, Chou LS, Xiao Y, et al. The development of next-generation sequencing assays for the mitochondrial genome and 108 nuclear genes associated with mitochondrial disorders. *J Mol Diagn* 2013;15:526–534.
15. Sivakumaran TA, Husami A, Kissell D, et al. Performance evaluation of the next-generation sequencing approach for molecular diagnosis of hereditary hearing loss. *Otolaryngol Head Neck Surg* 2013;148:1007–1016.
16. Chin EL, da Silva C, Hegde M. Assessment of clinical analytical sensitivity and specificity of next-generation sequencing for detection of simple and complex mutations. *BMC Genet* 2013;14:6.
17. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–1303.
18. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* 2010;26:2069–2070.
19. NHLBI Exome Sequencing Project (ESP). Exome Variant Server [cited 2012 December]. <http://evs.gs.washington.edu/EVS/>.
20. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* 2007;35(Web Server issue):W71–W74.
21. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol* 2011;29:24–26.
22. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinformatics* 2013;14:178–192.
23. Rehm HL, Bale SJ, Bayrak-Toydemir P, et al.; Working Group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Committee. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med* 2013;15:733–747.