UC Irvine
UC Irvine Previously Published Works

Title

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

Permalink

https://escholarship.org/uc/item/8w4650gd

Journal

Genetics in Medicine, 16(7)

Authors

Strom, Samuel
Lee, Hane
Das, Kingshuk
et al.

Publication Date

2014-07-01

DOI

10.1038/gim.2013.183

Peer reviewed

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

Samuel P. Strom, Ph.D.[1,2,*], **Hane Lee, Ph.D.**[1], **Kingshuk Das, M.D.**[1], **Eric Vilain, M.D., Ph.D.**[2,3], **Stanley F. Nelson, M.D.**[1,2], **Wayne W. Grody, M.D., Ph.D.**[1,2,3], and **Joshua L. Deignan, Ph.D.**[1]

[1]Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles

[2]Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles

[3]Department of Pediatrics, David Geffen School of Medicine, University of California Los Angeles

Abstract

Purpose—Sanger sequencing is currently considered the gold standard methodology for clinical molecular diagnostic testing. However, next generation sequencing (NGS) 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 NGS 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—103 of 110 total SNVs analyzed had a quality score ≥Q500, 103 (100%) of which were confirmed by Sanger sequencing. Of the remaining 7 variants with quality scores <Q500, 6 were confirmed by Sanger sequencing (85%).

Conclusions—For single nucleotide variants, we predict we will be able to reduce our Sanger confirmation workload going forward 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 utilize NGS technology. However, Sanger confirmation of low quality single nucleotide variants and all indels (insertions or deletions less than 10 bp) remains necessary at this time in our laboratory.

695 Charles E. Young Dr. Gonda Building, Rm. 5309, Los Angeles, CA 90095, Phone: 310-794-2781, Fax: 310-794-2771, sstrom@mednet.ucla.edu.

Introduction

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 ("mismapping"), 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 ("10x 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. In addition to the count of reference and non-reference ("variant") nucleotides observed at a given position, valuable data amasses. These data include: mapping quality (MQ), 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 employing the Genome Analysis Toolkit ("GATK") [1, 17] implements a Bayesian genotype likelihood model (based on known polymorphic loci such as dbSNP variants) and variant quality score recalibration (VQSR) to estimate posterior probabilities for each variant call (with hapmap_3.3.b37.sites and 1000G_omni2.5.b37.sites for training resources).

While 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 due to 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) datasets from the end-to-end analytically validated procedures established in the UCLA Clinical Genomics Center, which is part of the UCLA Molecular Diagnostics Laboratories (both CLIA- and CAP-accredited).

For variants with high quality scores (>Q10,000) and high coverage (>100x), 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 (such as a sample swap). Thus, for high-quality NGS variants, Sanger confirmation serves almost exclusively as a sample quality control (QC) 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 (QC) measure in our laboratory.

Materials and Methods

Clinical Exome Sequencing

Exome sequencing was performed in the UCLA Clinical Genomics Center [http://pathology.ucla.edu/genomics] following validated protocols. Briefly, high molecular

genomic DNA was isolated from whole blood collected in a lavender-top tube (K_2EDTA or K_3EDTA) using a QIAcube (QIAGEN). For all of the clinical samples, exome sequencing was performed using the Agilent SureSelect Human All Exon 50mb for exome capture and Illumina HiSeq2000 for sequencing as 50bp paired-end runs using V3 chemistry. For the non-clinical samples, Agilent SureSelect Human All Exon 50mb XT kit (V2) was used for exome capture and Illumina HiSeq2000 for sequencing as 100bp paired-end runs using V3 chemistry.

Data analysis was performed using the analytical pipeline implemented and validated for clinical exome sequencing in the UCLA Clinical Genomics Center. All sequence reads were aligned to the human reference genome (Human GRCh37/hg19) using Novoalign. PCR duplicates were marked by Picard, and GATK 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 flanking junctions were annotated using Variant Annotator X (VAX), an in-house MySQL database using data from the publicly available Ensembl Variant Effect Predictor [18]. A detailed description of the bioinformatic methods used to analyze these data is presented as Supplemental Materials 1.

Several steps are taken to reduce the probability of sample swap errors in our laboratory including: 1) assays are performed by appropriately licensed technologists with experience in next generation sequencing workflows; 2) at least two unique identifiers are used to label all reaction vessels and worksheets at all pre-analytical stages; 3) 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. Also - when available - prior 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 as well. Some additional steps that laboratories could employ to reduce sample errors include running samples in duplicate, running the CES assay in parallel with a SNP 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 the plasmid sequence occurs in the final result.

Variant Selection

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

All SNVs selected, regardless of report status, are predicted to be non-synonymous 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 [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 sequenced by automated capillary gel electrophoresis (ABI 3730, Life Technologies Corp., Carlsbad, CA). An ABMG board-certified clinical molecular geneticist manually analyzed the resulting sequence traces using Sequence Scanner (ABI).

Cost Analysis

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

Results

Exome sequencing results were confirmed for 103/103 (100%) of SNVs with quality scores Q500 (Table 1). The coverage depth for these variants ranged from 5x–250x with a mean of 116x. The correlation between quality score and coverage depth is positive and statistically significant (R=0.56, P<0.0001) (Figure 1). Of the 7 SNVs with quality scores <Q500, only one was not corroborated by the Sanger sequencing data (Table 1, Figure 2).

From the first 144 signed out reports, the average number of reported variants per report is approximately 1 (range: 0–5 variants). With an estimated cost of $240 USD per confirmation and a sample volume of 40 reports per month, the total cost to the laboratory performing the test is $9,600 per month ($115,200 per year). Furthermore, the number of clinically relevant (non-incidental) 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. Notably, turn-around time for reports requiring variant confirmation were delayed at least one week on average compared to reports with no reported variants.

Discussion

For each UCLA Clinical Exome Sequencing 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 (such as the effect 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 a variant is worthy of reporting, the laboratory then considers the technical validity of the finding. Prior to May 2013, Sanger sequencing was used as an alternate methodology for validation of each reported variant. Since that time, only indels and SNVs with quality scores <Q500 are validated by Sanger sequencing.

As it has been considered the "gold standard" for over 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 alternate methodology prior to reporting is not often employed 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. While 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. While the error rate for a single base is relatively higher in NGS than Sanger sequencing, 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 non-informative beyond sample QC, as the vast majority of results will be concordant and the remaining negative results will not be interpretable. Thus, such high-quality next generation sequencing 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

GATK [17] 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 (approximately 40x coverage) for the Clinical Exome Sequencing test in our laboratory, above which all 103 single nucleotide variants detected were confirmed by Sanger sequencing.

Manual inspection of variant calls using a visualization tool such as the Integrative Genomics Browser (IGV) [21, 22] by a genomics expert is a potential alternative to our quality score threshold approach. While 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 <10bp) are also detected by Clinical Exome Sequencing and 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 Sanger-confirm all such reported variants.

The perceived benefits of performing Sanger confirmation on all NGS-detected SNVs lies in quality control and risk avoidance. This must be weighed against increased test cost, delayed turn-around time, and the potentially paralyzing failure to confirm a very high quality variant of clinical significance. While 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 upon 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 lab 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 Clinical Exome Sequencing results, the slight probability of a high quality variant being an analytical false positive is typically a minor consideration compared to 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 Clinical Exome Sequencing results with quality scores >Q500 (SNVs only). However, other laboratories wishing to follow this paradigm must establish their own quality thresholds for each assay and provide empiric evidence to support those decisions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. DePristo MA, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43(5):491–8. [PubMed: 21478889]

2. Kiezun A, et al. Exome sequencing and the genetic basis of complex traits. Nat Genet. 2012; 44(6): 623–30. [PubMed: 22641211]

3. Abecasis GR, et al. A map of human genome variation from population-scale sequencing. Nature. 2010; 467(7319):1061–73. [PubMed: 20981092]

4. Brockman W, et al. Quality scores and SNP detection in sequencing-by-synthesis systems. Genome Res. 2008; 18(5):763–70. [PubMed: 18212088]

5. Li M, Nordborg M, Li LM. Adjust quality scores from alignment and improve sequencing accuracy. Nucleic Acids Res. 2004; 32(17):5183–91. [PubMed: 15459287]

6. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25(16): 2078–9. [PubMed: 19505943]

7. Li H, Homer N. A survey of sequence alignment algorithms for next-generation sequencing. Brief Bioinform. 2010; 11(5):473–83. [PubMed: 20460430]

8. Li Y, et al. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. Nat Genet. 2010; 42(11):969–72. [PubMed: 20890277]

9. Ng S, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009; 461(7261):272–6. [PubMed: 19684571]

10. Ng S, et al. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet. 2010; 42(1):30–5. [PubMed: 19915526]

11. Sikkema-Raddatz B, et al. Targeted next-generation sequencing can replace Sanger sequencing in clinical diagnostics. Hum Mutat. 2013; 34(7):1035–42. [PubMed: 23568810]

12. Peng G, et al. Rare variant detection using family-based sequencing analysis. Proc Natl Acad Sci U S A. 2013; 110(10):3985–90. [PubMed: 23426633]

13. McCourt CM, 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(7):e69604. [PubMed: 23922754]

14. Dames S, 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(4): 526–34. [PubMed: 23665194]

15. Sivakumaran TA, et al. Performance evaluation of the next-generation sequencing approach for molecular diagnosis of hereditary hearing loss. Otolaryngol Head Neck Surg. 2013; 148(6):1007–16. [PubMed: 23525850]

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. [PubMed: 23418865]

17. McKenna A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20(9):1297–303. [PubMed: 20644199]

18. McLaren W, et al. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics. 2010; 26(16):2069–70. [PubMed: 20562413]

19. NHLBI Exome Sequencing Project (ESP). Exome Variant Server. [cited 2012 December]; Available from: http://http://evs.gs.washington.edu/EVS/

20. Untergasser A, et al. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res. 2007; 35(Web Server issue):W71–4. [PubMed: 17485472]

21. Robinson JT, et al. Integrative genomics viewer. Nat Biotechnol. 2011; 29(1):24–6. [PubMed: 21221095]

22. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013; 14(2):178–92. [PubMed: 22517427]

23. Rehm HL, et al. ACMG clinical laboratory standards for next-generation sequencing. Genet Med. 2013; 15(9):733–47. [PubMed: 23887774]
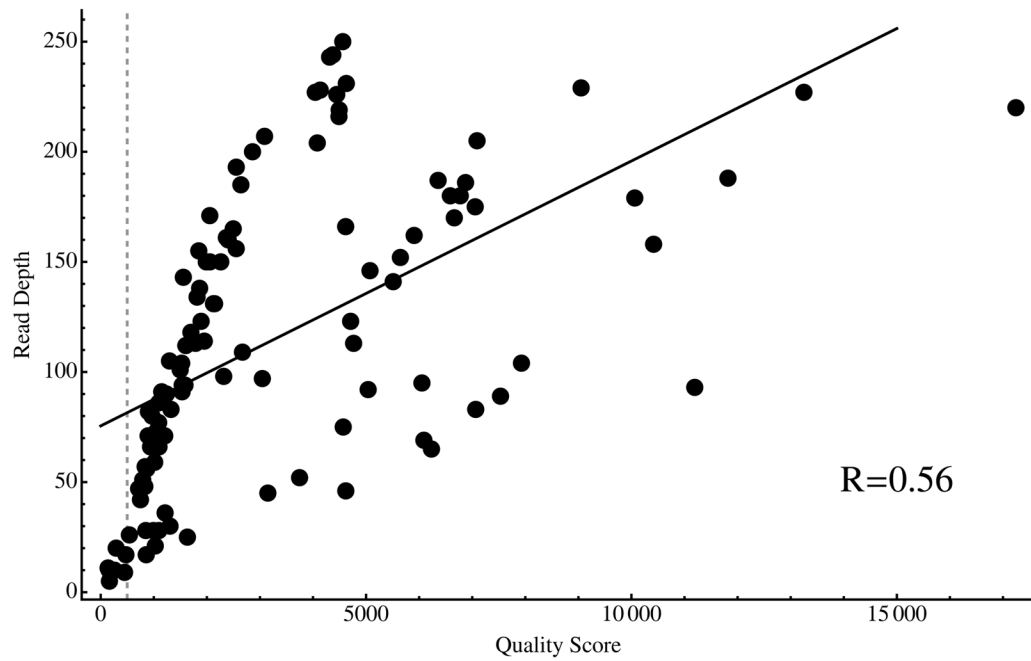
Figure 1. Correlation between Quality Scores and Depth of Coverage
Individual quality scores are plotted against read depth for 110 SNV loci tested. Quality score threshold of Q500 is marked by a dashed grey vertical line. The correlation is positive and significant (Pearson Correlation Significance Test, $P<10^{-13}$).

Figure 2. Validation results sorted by quality score

Each 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.

Table 1

Variant information.

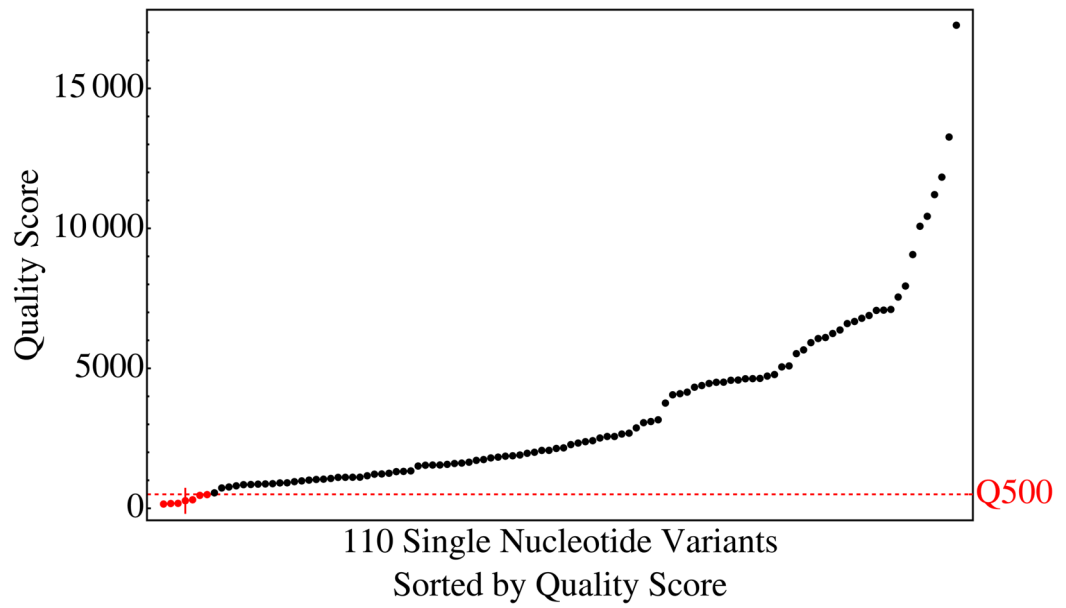| #	Gene	Zyg.	QUAL	Conf.	Depth	Alt.	V.F.
1	OBSCN	Het	139	Yes	11	6	54.5
2	BEAN1	Het	157	Yes	10	6	60
3	SURF1	Homo	164	Yes	5	5	100
4	ZNHIT2	Het	258	No	10	4	40
5	CACNA1H	Het	292	Yes	20	7	35
6	SHB	Homo	449	Yes	9	9	100
7	GAA	Het	475	Yes	17	10	58.8
8	KCNT1	Het	540	Yes	26	17	65.4
9	MYLK	Het	714	Yes	47	24	51.1
10	FGFR1	Het	749	Yes	42	24	57.1
11	BEAN1	Het	791	Yes	51	25	49
12	RANBP3	Het	831	Yes	48	26	54.2
13	TGM6	Het	837	Yes	57	27	47.4
14	OTOF	Het	850	Yes	28	14	50
15	NDUFS8	Homo	858	Yes	17	17	100
16	BCOR	Het	867	Yes	56	27	48.2
17	CACNA1A	Het	892	Yes	71	30	42.3
18	TCF4	Het	898	Yes	83	35	42.2
19	OBSCN	Het	938	Yes	66	31	47
20	SYNE1	Het	966	Yes	80	33	41.3
21	MECP2	Hemi	995	Yes	28	28	100
22	FBN3	Het	1,017	Yes	59	33	55.9
23	SCML4	Het	1,028	Yes	21	9	42.9
24	WWP2	Het	1,052	Yes	73	39	53.4
25	AFG3L2	Het	1,092	Yes	77	39	50.6
26	MYH6	Het	1,094	Yes	86	38	44.2
27	COL20A1	Het	1,096	Yes	28	13	46.4

#	Gene	Zyg.	QUAL	Conf.	Depth	Alt.	V.F.
28	SCN5A	Het	1,098	Yes	66	32	48.5
29	NPAP1	Het	1,149	Yes	91	40	44
30	AGTR1	Het	1,207	Yes	71	37	52.1
31	MTATP6	Homo	1,213	Yes	36	35	97.2
32	USP21	Het	1,238	Yes	90	43	47.8
33	LMNA	Het	1,296	Yes	105	43	41
34	SLC9A2	Het	1,306	Yes	30	18	60
35	ATP8A2	Het	1,324	Yes	83	44	53
36	POLR3A	Het	1,498	Yes	101	48	47.5
37	CHRNA7	Het	1,526	Yes	104	49	47.1
38	ABCA4	Het	1,533	Yes	91	48	52.7
39	C1QTNF5	Het	1,535	Yes	94	48	51.1
40	MYH7	Het	1,557	Yes	143	57	39.9
41	CACNA1D	Het	1,588	Yes	94	54	57.4
42	COL6A2	Het	1,603	Yes	112	61	54.5
43	KDM6A	Hemi	1,633	Yes	25	25	100
44	KCNQ3	Het	1,699	Yes	118	60	50.8
45	FANCG	Het	1,728	Yes	114	54	47.4
46	SYNE1	Het	1,787	Yes	113	58	51.3
47	SLC37A1	Het	1,815	Yes	134	65	48.5
48	OPA1	Het	1,848	Yes	155	61	39.4
49	GJB2	Het	1,864	Yes	138	59	42.8
50	WFS1	Het	1,891	Yes	123	61	49.6
51	MYBPC3	Het	1,952	Yes	114	59	51.8
52	SCN8A	Het	1,988	Yes	150	150	100
53	RUNX2	Het	2,055	Yes	171	74	43.3
54	RAD21	Het	2,058	Yes	150	66	44
55	SETX	Het	2,125	Yes	131	65	49.6
56	TNXB	Het	2,145	Yes	131	67	51.1

#	Gene	Zyg.	QUAL	Conf.	Depth	Alt.	V.F.
57	*RET*	Het	2,262	Yes	150	76	50.7
58	*CPT2*	Het	2,317	Yes	98	37	37.8
59	*STAT1*	Het	2,367	Yes	161	74	46
60	*SYNE1*	Het	2,402	Yes	160	75	46.9
61	*ERCC5*	Het	2,497	Yes	165	80	48.5
62	*CNTNAP2*	Het	2,553	Yes	156	79	50.6
63	*APC*	Het	2,554	Yes	197	91	46.2
64	*PTPN11*	Het	2,641	Yes	185	80	43.2
65	*CLN8*	Het	2,671	Yes	109	50	45.9
66	*SYNE1*	Het	2,860	Yes	200	90	45
67	*OTOF*	Het	3,045	Yes	97	50	51.5
68	*TSHR*	Het	3,087	Yes	207	99	47.8
69	*G6PD*	Hemi	3,149	Yes	45	45	100
70	*G6PD*	Hemi	3,746	Yes	52	52	100
71	*SCN2A*	Het	4,043	Yes	227	100	44.1
72	*ZEB2*	Het	4,080	Yes	204	101	49.5
73	*SETX*	Het	4,136	Yes	228	109	47.8
74	*TTN*	Het	4,312	Yes	243	119	49
75	*TTN*	Het	4,373	Yes	243	110	45.3
76	*TTN*	Het	4,447	Yes	226	112	49.6
77	*SMCHD1*	Het	4,489	Yes	216	115	53.2
78	*ACVR1*	Het	4,493	Yes	219	114	52.1
79	*ITPR1*	Het	4,560	Yes	250	137	54.8
80	*NDUFA1*	Hemi	4,569	Yes	75	75	100
81	*TBC1D24*	Het	4,615	Yes	166	84	50.6
82	*TBC1D25*	Hemi	4,618	Yes	46	46	100
83	*CHD7*	Het	4,628	Yes	231	124	53.7
84	*RP1*	Het	4,709	Yes	123	60	48.8
85	*CERKL*	Het	4,765	Yes	116	58	50

#	Gene	Zyg.	QUAL	Conf.	Depth	Alt.	V.F.
86	*KDM6A*	Hemi	5,039	Yes	92	92	100
87	*MET*	Het	5,073	Yes	146	71	48.6
88	*GJB2*	Homo	5,511	Yes	141	141	100
89	*RP1*	Het	5,646	Yes	152	72	47.4
90	*TBC1D24*	Het	5,906	Yes	163	73	44.8
91	*KCNJ13*	Homo	6,052	Yes	95	95	100
92	*PIK3C2G*	Homo	6,089	Yes	69	69	100
93	*SMG5*	Homo	6,232	Yes	65	65	100
94	*GOSR2*	Het	6,357	Yes	187	79	42.2
95	*CERKL*	Het	6,587	Yes	180	83	46.1
96	*RAPGEF6*	Het	6,662	Yes	171	81	47.4
97	*USH2A*	Homo	6,775	Yes	180	180	100
98	*GOSR2*	Het	6,874	Yes	186	83	44.6
99	*AHI1*	Homo	7,056	Yes	175	175	100
100	*SLC35C1*	Homo	7,066	Yes	83	83	100
101	*OPA1*	Het	7,090	Yes	205	92	44.9
102	*POFUT1*	Homo	7,533	Yes	89	89	100
103	*GLDC*	Homo	7,927	Yes	104	104	100
104	*MYOM2*	Homo	9,051	Yes	228	228	100
105	*MAGEA1*	Het	10,063	Yes	179	97	54.2
106	*ATP7A*	Hemi	10,417	Yes	158	158	100
107	*FAM98B*	Homo	11,193	Yes	93	92	98.9
108	*HTR2C*	Hemi	11,818	Yes	186	186	100
109	*PDE6C*	Homo	13,249	Yes	227	227	100
110	*ALDH18A1*	Homo	17,244	Yes	220	220	100

Abbreviations: Zyg., zygosity; Het, heterozygous; Hom homozygous; Hemi, hemizygous (X-linked observed in a male); QUAL, clinical exome sequencing locus Quality Score; Conf., variant confirmed by Sanger sequencing; Alt., number of independent reads supporting the alternate allele; V.F., variant frequency (equal to 100 * Alt./Depth). Red shading indicates, respectively: QUAL <500, variant not confirmed, coverage <40x.

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