The Journal of Molecular Diagnostics, Vol. 18, No. 6, November 2016

the Journal of Molecular Diagnostics
jmd.amjpathol.org

Sanger Confirmation Is Required to Achieve Optimal Sensitivity and Specificity in Next-Generation Sequencing Panel Testing

Wenbo Mu, Hsiao-Mei Lu, Jefferey Chen, Shuwei Li, and Aaron M. Elliott

From Ambry Genetics, Aliso Viejo, California

Accepted for publication July 19, 2016.

Address correspondence to Aaron M. Elliott, Ph.D., Ambry Genetics, Research and Development, 15 Argonaut, Aliso Viejo, CA 92656. E-mail: aelliott@ambrygen.com.

Next-generation sequencing (NGS) has rapidly replaced Sanger sequencing as the method of choice for diagnostic gene-panel testing. For hereditary-cancer testing, the technical sensitivity and specificity of the assay are paramount as clinicians use results to make important clinical management and treatment decisions. There is significant debate within the diagnostics community regarding the necessity of confirming NGS variant calls by Sanger sequencing, considering that numerous laboratories report having 100% specificity from the NGS data alone. Here we report our results from 20,000 hereditary-cancer NGS panels spanning 47 genes, in which all 7845 nonpolymorphic variants were Sanger-sequenced. Of these, 98.7% were concordant between NGS and Sanger sequencing and 1.3% were identified as NGS false-positives, located mainly in complex genomic regions (A/T-rich regions, G/C-rich regions, homopolymer stretches, and pseudogene regions). Simulating a false-positive rate of zero by adjusting the variant-calling quality-score thresholds decreased the sensitivity of the assay from 100% to 97.8%, resulting in the missed detection of 176 Sanger-confirmed variants, the majority in complex genomic regions ($n = 114$) and mosaic mutations ($n = 7$). The data illustrate the importance of setting quality thresholds for panel testing only after thousands of samples have been processed and the necessity of Sanger confirmation of NGS variants to maintain the highest possible sensitivity. *(J Mol Diagn 2016, 18: 923—932; http://dx.doi.org/10.1016/j.jmoldx.2016.07.006)*

Rapid technical advances in target enrichment and next-generation sequencing (NGS) technologies have revolutionized clinical diagnostic testing. The ability to simultaneously analyze all genes involved in a disease phenotype at a reduced cost has enabled NGS to surpass conventional capillary Sanger sequencing as the method of choice in most high-throughput diagnostics laboratories.[1] Targeted NGS panel tests have become frontline assays for a wide variety of inherited disorders. In accordance with the accessibility of NGS, a multitude of diagnostics laboratories have recently launched a broad range of inherited cancer-susceptibility gene panels that test a wide variety of characterized genes associated with hereditary cancers.[2–6] The accuracy of these tests can vary based on a number of variables including the target enrichment platform, sequencing technology, bioinformatics pipeline used, and variant classification experience. The technical sensitivity and specificity of a hereditary-cancer test are paramount as physicians use the results to make important clinical management decisions such as costly specialized surveillance and preventive surgery.

There is significant debate within the diagnostic gene-sequencing space regarding the necessity of confirming NGS-determined variant calls using a secondary technology such as Sanger sequencing to avoid the possibility of reporting out a false-positive result. Many laboratories are resistant to Sanger-sequencing confirmation as it requires additional primer design, validation, clinical workflow, and licensed personnel for chromatogram analysis, which increase both the cost and turnaround time of testing.

Supported by Ambry Genetics.
W.M. and H.-M.L. contributed equally to this work.
Disclosure: All authors are employees of Ambry Genetics. A.M.E. and H.-M.L. are shareholders of Ambry Genetics. A.M.E. is an officer of Ambry Genetics.

http://dx.doi.org/10.1016/j.jmoldx.2016.07.006

Guidance on NGS laboratory standards from the College of American Pathologists leaves it up to the laboratory performing the assay to determine whether confirmation testing is appropriate.[7] In contrast, the American College of Medical Genetics and Genomics recommends secondary confirmation testing for all NGS-reported variants.[8] To date, the data concerning the need for secondary confirmation of NGS calls in germline testing are severely limited by the use of small data sets. Strom et al[9] analyzed 110 variants from 144 exome samples and concluded that single-nucleotide variants (SNVs) meeting a quality (Q) score of $\geq$500 should not require Sanger confirmation. Similarly, Baudhuin et al[10] recently concluded, after their study in 77 samples, that Sanger confirmation of NGS SNVs with coverage of $\geq$100$\times$ and a Q score of $\geq$20 was "unnecessarily redundant."

Validation data published and marketed by several commercial and academic laboratories offering hereditary-cancer NGS panel testing illustrated an analytical specificity of 100%, detecting zero false-positives.[3–5] However, it is likely that these assays sacrifice analytical sensitivity by missing true variants in complex genomic regions where false-positives occur. Most disease-causing variants are rare variants with minor allele frequencies of $<$0.1%.[11] Therefore, the appropriate balance between specificity and sensitivity cannot be accurately defined until thousands of samples are processed. Here we report our results from 20,000 samples, the largest study to date, to determine the necessity of Sanger-sequencing confirmation of NGS variants and to accurately define the appropriate sensitivity and specificity for a high-volume gene-panel diagnostic assay.

Materials and Methods

Samples

A consecutive series of 20,000 patient samples referred to Ambry Genetics (Aliso Viejo, CA) for NGS-based multigene hereditary-cancer testing were included in the study (Solutions Institutional Review Board protocol number 1OCT14–93). Data from multiple hereditary cancer–test offerings [BreastNext, ColoNext, PGLNext, RenalNext, PancNext, GYNplus, OvaNext, and CancerNext (all from Ambry Genetics)] were included in this study (Table 1). During initial test validation of these panels on over 200 characterized samples, the sensitivity and specificity of the assay were 100% and 99.99%, respectively. At least 6 to 7 μg of genomic DNA was extracted from whole blood or saliva using the QiaSymphony instrument (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Isolated DNA was quantified using a NanoDrop UV spectrophotometer (Thermo Fisher Scientific, Waltham, MA) and/or Qubit Fluorometer (Life Technologies, Carlsbad, CA), with quality metrics of A260/280 from 1.8 to 2.0 and A260/230 of $\geq$1.6.

NGS Library Preparation and Sequencing

A customized target-enrichment oligonucleotide library was designed to capture 49 hereditary cancer–related genes (Integrated DNA Technologies, Coralville, IA). Genomic DNA was mechanically sheared to 300-bp fragments with an LE220 focused ultrasonicator (Covaris, Woburn, MA) and NGS library prepared on a Freedom EVO100 automated system (Tecan, Mannedorf, Switzerland) according to the manufacturer's recommendation (Kapa Biosystems, Wilmington, MA). Reactions were purified using AMPure XP beads (Beckman Coulter, Brea, CA) and quantified on a 2200 TapeStation Instrument (Agilent Technologies, Santa Clara, CA). Adapter-ligated DNA were hybridized in solution to customized biotinylated DNA oligonucleotides complementary to the target sequences of interest (Integrated DNA Technologies). After hybridization, Streptavidin Dynabeads (Life Technologies) were used to capture the regions of interest and were PCR-amplified on a Bio-Rad T100 thermal cycler (Bio-Rad Laboratories, Hercules, CA)

Table 1 Genes Included in Next-Generation Sequencing Multigene Cancer Panels

Cancer type	No. of genes	Gene list*
Breast cancer	17	*ATM, BARD1, BRCA1, BRCA2, BRIP1, CDH1, CHEK2, MRE11A, MUTYH, NBN, NF1, PTEN, RAD50, RAD51C, RAD51D, TP53, PALB2*
Colorectal cancer	17	*APC, BMPR1A, CDH1, CHEK2, EPCAM, GREM1, MLH1, MSH2, MSH6, MUTYH, PMS2, POLD1, POLE, PTEN, SMAD4, STK11, TP53*
Paragangliomas/Pheochromocytomas	12	*FH, MAX, MEN1, NF1, RET, SDHA, SDHAF2, SDHB, SDHC, SDHD, TMEM127, VHL*
Renal cancer	19	*MLH1, MSH2, MSH6, PMS2, PTEN, TP53, VHL, EPCAM, FLCN, TSC2, TSC1, SDHB, MET, MITF, SDHC, SDHD, SDHA, FH, BAP1*
Pancreatic cancer	13	*APC, ATM, BRCA1, BRCA2, CDKN2A, EPCAM, MLH1, MSH2, MSH6, PMS2, STK11, TP53, PALB2*
Ovarian cancer/uterine cancer	24	*ATM, BARD1, BRCA1, BRCA2, BRIP1, CDH1, CHEK2, EPCAM, MLH1, MRE11A, MSH2, MSH6, MUTYH, NBN, NF1, PMS2, PTEN, RAD50, RAD51C, RAD51D, STK11, TP53, PALB2, SMARCA4*

*CDK4 is included in the comprehensive CancerNext test.

with the following conditions: 98°C for 45 seconds; followed by a program of 98°C for 15 seconds, 65°C for 30 seconds, and 72°C for 30 seconds for 12 cycles; and ending with a 1-minute extension at 72°C. Final libraries were purified using AMPure XP beads (Beckman Coulter) and quantified on the 2200 TapeStation (Agilent Technologies) for normalization and pooling. Sequencing was conducted on HiSeq2500 or NextSeq500 using 150-bp paired-end conditions as described in the manufacturer's standard workflow (Illumina, San Diego, CA).

Data Analysis

Variants in all coding domains plus at least five bases into the 5′ and 3′ ends of all introns, 5′ untranslated region (UTR) and 3′UTR were used for analysis. Variants in *PMS2* exons 11 to 15 were excluded from analysis due to interference with >99% homologous pseudogene *PMS2CL*. Although *BMPR1A*, *CHEK2*, and *SDHA* also contain highly homologous pseudogenes (98% to 99%), there are enough intronic mismatches to differentiate them when using 150-bp paired-end sequencing. *EPCAM* and *GREM1* cancer susceptibility is limited to gross abnormalities and so was not included in the NGS data.[12,13] For *MITF*, only the status of the c.952G>A (p.E318K) was analyzed during testing.[14] Data excluded gross deletions and duplications as these were identified through a customized microarray for testing. Initial data processing and base calling, including extraction of cluster intensities, were done using Real Time Analysis version 1.17.21.3 (HiSeq Control Software version 2.0.10; Illumina). Output files were processed and demultiplexed using Casava version 1.8.2 (Illumina). Sequence reads were aligned to the reference human genome (GRCh37) using NovoAlign version 3.02.07 (Novocraft Technologies, Selangor, Malaysia) and variant calls generated using Genome Analysis Toolkit version 3.2.2 (Broad Institute, Cambridge, MA). Duplicate reads were removed before variant calling. Variant calling was determined by multiple factors, including base quality, mapping quality, haplotype, coverage, number of nearby mismatches, and heterozygous ratio. A minimal base-calling Q score of 17 on each base was required. The variant-calling Q score of each variant is a measurement indicating the confidence of the specific variant called. A minimal coverage of 10× and a Q score of 30 were required for candidate variants to be called. Any region with a depth of coverage of <10× was filled in by Sanger sequencing. Data were annotated with the customized Variant Analyzer tool version 3.1 (Ambry Genetics), including nucleotide and amino acid conservation, biochemical nature of amino acid substitutions, and population frequency in Exome Sequencing Project,[15] 1000 Genomes Project,[16] and internal Ambry data. Each variant was classified as a pathogenic mutation, variant likely pathogenic, variant of unknown significance, likely benign, or benign, following the methods of classification consistent with the guideline from the American College of Medical Genetics and Genomics.[17] For variant classification, "likely" referred to a >95% likelihood of a variant being disease causing or benign, as proposed by the International Agency for Research on Cancer.[18] This classification algorithm incorporated multiple lines of evidence including co-segregation, co-occurrence, population frequency, phenotype and family history, variant location, structural impact, *in silico* predictions, and functional studies. According to the American College of Medical Genetics and Genomics guideline,[17] truncating mutations that occur close to the COOH-terminus and in-frame deletions/insertions may retain protein function; thus their interpretation requires extra caution and additional evidence such as functional assays for proper classification. All variants with likely clinical relevance (pathogenic variants, likely pathogenic variants, and variants of unknown significance) were confirmed by Sanger sequencing in the sense and antisense directions regardless of whether they had been detected previously. If a variant was not confirmed, Sanger sequencing was repeated with a second set of unique PCR primers to avoid allele dropout. Variants that failed to be confirmed by Sanger sequencing were identified as false-positive variants and were cleared from reporting. A negative report was signed out if a sample had a mean base Q score of >30, a percentage of perfect index of >85%, a percentage of bases over Q30 of >75%, and no detected clinically relevant variants.

To simulate a scenario of zero false-positives, a high Q-score cutoff could be set to exclude all low-confidence variants. Therefore, by adjusting this metric, one could control the number of false-positives that could be tolerated in variant calling. In the data set, 20 of 47 genes had false-positives; the maximal Q-scores of false-positives are listed in Supplemental Table S1. We gradually simulated 20 scenarios by using 20 different Q-score thresholds, which were uniformly distributed between the initial Q-score threshold of 30 and a maximal Q score of known false-positives. For example, ATM had a maximal Q score of 695; thus 20 different Q-score thresholds, in increments of 33, between 30 and 695 were simulated to gradually achieve a scenario of zero false-positives. All variants having a Q-score above a specific threshold were chosen for each analysis. Thus, the highest sensitivity occurred when the minimal Q score was used as the threshold in which the most false-positives were included. The highest specificity occurred when the maximal Q score was used as a threshold in which no false-positives were included. Due to variant-calling difficulties based on specific genomic content, the simulation was repeated for all 47 genes individually (Supplemental Figure S1).

Sanger Sequencing

Universally tagged primers (Integrated DNA Technologies) were designed using Vector VNTI Advance version 11.5.1

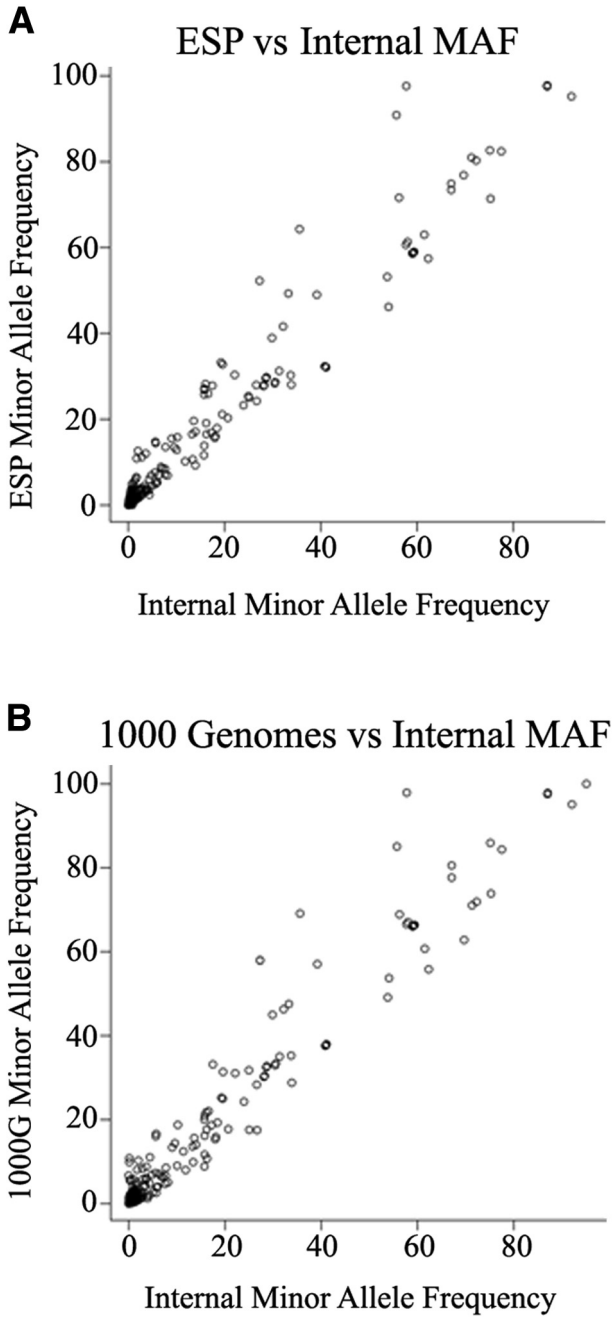


Figure 1 Polymorphic minor allele frequency (MAF) concordance between the data sets from the Ambry 20,000-sample study and Exome Sequencing Project (ESP)[15] and the 1000 Genomes Project[16] (1000G). **A** and **B:** Comparison of minor allele frequency of polymorphic variants between ESP (**A**) and 1000G (**B**) and the Ambry 20,000-sample study population.

(Life Technologies) and compared with those from public databases to avoid SNPs and nonspecific amplification. Genomic DNA was amplified using the HotStarTaq Master Mix Kit (Qiagen) with the following conditions: 95°C for 15 minutes; followed by a program of 94°C for 30 seconds, 60°C for 30 seconds, and 72°C for 45 seconds for 35 cycles; and ending with a 10-minute extension at 72°C. Amplicons were purified using HT ExoSAP-IT (Affymetrix, Santa Clara, CA) and bidirectionally sequenced using Big Dye Terminator version 3.1 on an ABI3730xl DNA analyzer (Applied Biosystems, Foster City, CA). Chromatogram analysis was conducted using Sequence Pilot version 4.2.1 (JSI Medical Systems, Boston, MA).

Results

Clinical NGS testing of 47 hereditary-cancer genes in 20,000 consecutive patients identified 1,130,361 benign polymorphic variants, 1773 pathogenic mutations, 204 likely pathogenic variants, and 5868 variants of unknown significance. Overall, the variants detected had a mean heterozygous ratio of 66% (range, 10% to 100%), mean coverage of 441× (range, 12× to 1204×), and a mean Q score of 8719 (range, Q30 to Q31,853). SNVs with Phred-scaled P value of >60 and small insertions and deletions (indels) with Phred-scaled P value of >200 were determined as strand bias variants in Genome Analysis Toolkit version 3.2.2. Of detected variants, 99.99% of SNVs and 100% of indels did not have strand bias. Although benign polymorphisms are not Sanger-confirmed and are filtered out for diagnostic-testing purposes, the frequency of common variants is a good internal quality-control metric to track. To do so, the internal population frequencies of the 692 unique polymorphism variants of the total of 1,130,361 benign polymorphisms identified were compared with the frequencies observed in publicly available data sets from the Exome Sequencing Project[15] and the 1000 Genomes Project.[16] Using the Pearson correlation coefficient and the Kolmogorov–Smirnov test, there was significant concordance between Ambry Genetics' internal data and the data from the Exome Sequencing Project[15] (Pearson correlation coefficient = 0.978, Kolmogorov–Smirnov test P = 2.2e-16) and the 1000 Genomes Project[16] (Pearson correlation coefficient = 0.978, Kolmogorov–Smirnov test P = 2.998e-15) (Figure 1).

Confirmatory Sanger-sequencing analysis was performed on all 7845 nonpolymorphic variants, including pathogenic

Table 2 The Variant Type Distribution of Nonpolymorphic Variants that Were Concordant between Next-Generation Sequencing and Sanger Sequencing

Genetic variant	Variant type	No. of variants	% of variants
SNV	Synonymous	48	0.6
	Missense	5817	75.1
	Splicing	140	1.8
	Nonsense	302	3.9
	Methionine change	14	0.2
	5′UTR/3′UTR/intron	497	6.4
Indel	In-frame	148	1.9
	Frameshift	721	9.3
	Splicing	10	0.1
	5′UTR/3′UTR/intron	49	0.6

SNV, single-nucleotide variant; UTR, untranslated region.

variants, likely pathogenic variants, and variants of unknown significance, which consisted of 6912 (3481 unique) SNVs and 933 (432 unique) indels. Of these, 98.7% (n = 7746) were concordant between NGS and Sanger sequencing, with 3876 of them being unique. The variant types and distribution of concordant variants are listed in Table 2. The Sanger-confirmed true-positives had a mean heterozygous ratio of 46% (range, 10% to 100%), mean coverage of 415× (range, 11× to 1204×), and a mean Q score of 2247 (range, Q38 to Q22,458). The remaining 1.3% of variants (94 SNVs, 5 indels) were cleared by Sanger sequencing and identified as NGS false-positives, with 46 of them unique variants (Figure 2A, Table 3, and Supplemental Table S2). The false-positives detected had a mean heterozygous ratio of 16% (range, 10% to 40%) and mean coverage of 223× (range, 10× to 689×).

False-positives were not evenly distributed across all genes as would be expected if they were random PCR or sequencing artifacts. For example, 27 genes being tested (*APC, BAP1, BMPR1A, CDK4, CDKN2A, FH, FLCN, MAX, MEN1, MET, MITF, MLH1, MSH6, NBN, POLD1, RAD51C, SDHAF2, SDHB, SDHC, SDHD, SMARCA4, STK11, TMEM127, TP53, TSC1, TSC2,* and *VHL*) had no false-positives (Table 3 and Figure 2B). Complex genomic regions including homopolymer stretches, pseudogene homology, and G/C-rich (≥65% G/C bases) and A/T-rich (≥65% A/T bases) regions can impact false-positive rates.[19–21] Of the 99 false-positives detected in the study, 46.7% were located in A/T-rich regions; 20.0%, in G/C-rich regions; 14.1%, in homopolymer stretches; and 2.0%, in pseudogene regions, and 17.2% had no discernable genomic cause (Supplemental Table S2). Importantly, 26% (n = 26) of false-positives identified represent

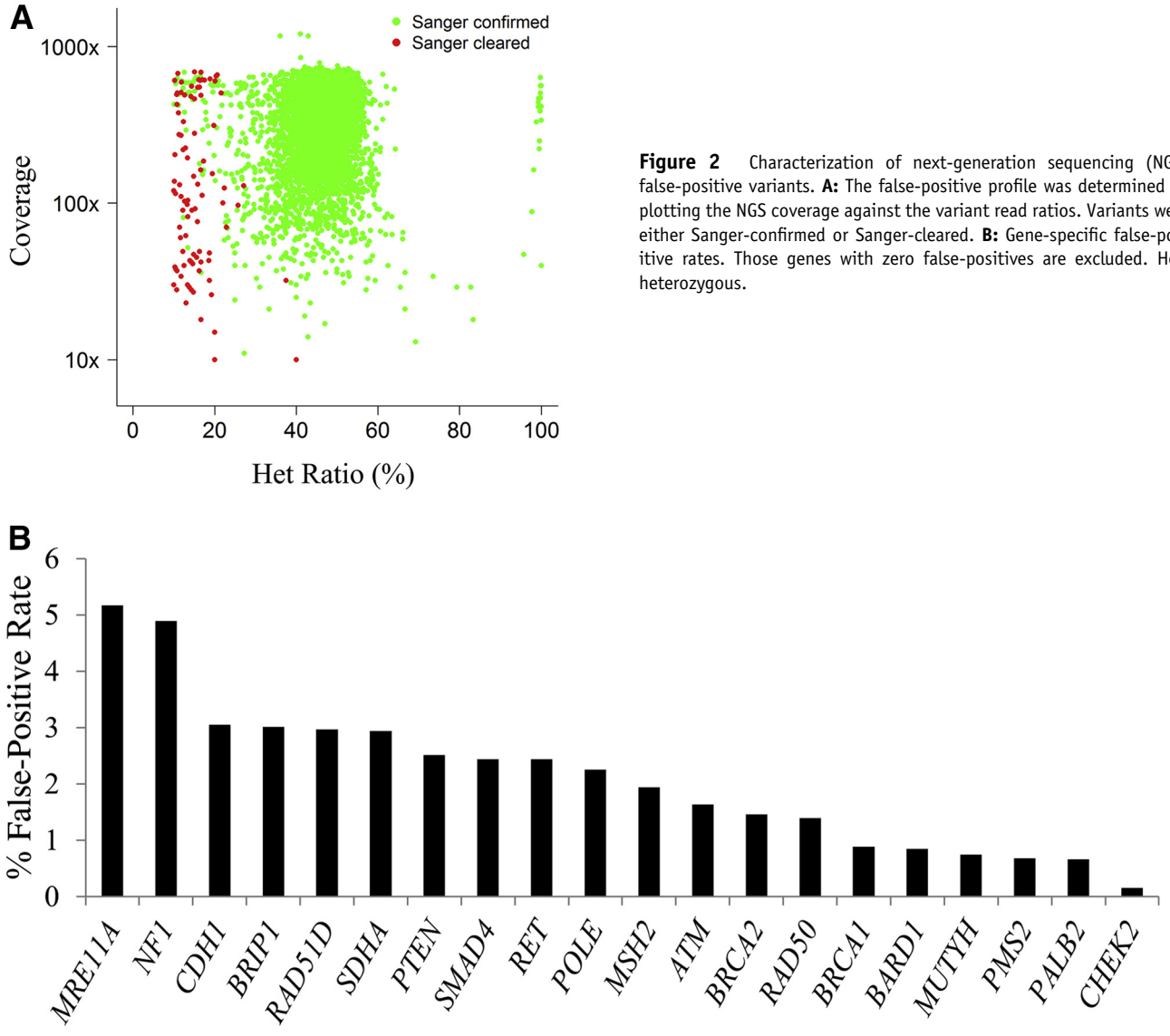


Figure 2 Characterization of next-generation sequencing (NGS) false-positive variants. **A:** The false-positive profile was determined by plotting the NGS coverage against the variant read ratios. Variants were either Sanger-confirmed or Sanger-cleared. **B:** Gene-specific false-positive rates. Those genes with zero false-positives are excluded. Het, heterozygous.

Table 3 Gene-Specific Next-Generation Sequencing FPRs and Simulated Sensitivity Analysis when FPR Adjusted to Zero

Gene	No. of Sanger-sequenced variants*	FPR	Sensitivity when FPR = 0%†
APC	309	0.00	100.00
ATM	857	1.63	94.90
BAP1	28	0.00	100.00
BARD1	235	0.85	97.00
BMPR1A	43	0.00	100.00
BRCA1	341	0.88	93.49
BRCA2	548	1.46	100.00
BRIP1	332	3.01	100.00
CDH1	131	3.05	96.06
CDK4	29	0.00	100.00
CDKN2A	40	0.00	100.00
CHEK2	653	0.15	97.70
FH	29	0.00	100.00
FLCN	33	0.00	100.00
MAX	4	0.00	100.00
MEN1	10	0.00	100.00
MET	38	0.00	100.00
MITF	5	0.00	100.00
MLH1	145	0.00	100.00
MRE11A	174	5.17	100.00
MSH2	309	1.94	83.50
MSH6	327	0.00	100.00
MUTYH	540	0.74	98.32
NBN	249	0.00	100.00
NF1	286	4.90	99.63
PALB2	302	0.66	100.00
PMS2	147	0.68	99.32
POLD1	91	0.00	100.00
POLE	133	2.26	94.62
PTEN	239	2.51	95.71
RAD50	359	1.39	99.44
RAD51C	181	0.00	100.00
RAD51D	135	2.96	96.18
RET	41	2.44	100.00
SDHA	34	2.94	100.00
SDHAF2	2	0.00	100.00
SDHB	29	0.00	100.00
SDHC	4	0.00	100.00
SDHD	15	0.00	100.00
SMAD4	41	2.44	100.00
SMARCA4	151	0.00	100.00
STK11	64	0.00	100.00
TMEM127	7	0.00	100.00
TP53	94	0.00	100.00
TSC1	25	0.00	100.00
TSC2	48	0.00	100.00
VHL	8	0.00	100.00

*All of the nonpolymorphic variants were confirmed/cleared by Sanger sequencing.

†Sensitivity when pipeline false-positive rate is adjusted to zero.

FPR, false-positive rate.

variants present in the Ambry Genetics internal database that had been previously Sanger-confirmed and reported out, illustrating that the filtering of recurrent false-positives as artifacts is not feasible (Supplemental Table S2). For example, the founder pathogenic *MUTYH* mutation c.1187 G>A, which is located in a G/C-rich region and responsible for *MUTYH*-associated polyposis, was identified as a false-positive in three samples and Sanger-confirmed 211 times in our data set. It is also evident that setting heterozygous read ratio and coverage filters for variant calling that would completely eliminate the need for Sanger-confirming of variants would sacrifice sensitivity, as confirmed variants and false-positives are intermixed at low coverage and heterozygous ratios (Figure 2A). Without affecting sensitivity and specificity, setting a conservative threshold of a >40% heterozygous read ratio and >100× coverage for variants that do not need secondary confirmation would decrease the Sanger-sequencing burden by 87%.

To verify the importance of having a highly sensitive assay for diagnostic testing, one that can detect true variants at low coverage and low heterozygous ratios in which false-positives generally reside, we gradually adjusted the variant-calling threshold to achieve 100% specificity, eliminating all false-positives from the data. The sensitivity [True-positives/(True-positives + False-negatives)] and false-positive rate [False-positives/(True calls + False-positives)] of the test were calculated at each incremental adjustment. As the false-positive rate decreased from 1.3% to 0%, the sensitivity of the assay decreased from 100% to 97.8%, resulting in the missed detection of 176 Sanger-confirmed variants (142 unique variants) (Table 3, Supplemental Table S3, and Figure 3). The false-negatives were concentrated in 13 genes (*ATM, BARD1, BRCA1, CDH1, CHEK2, MSH2, MUTYH, NF1, PMS2, POLE, PTEN, RAD50, RAD51C,* and *RAD51D*), ranging from 0.37% for *NF1* to 16.50% of confirmed calls missed for *MSH2* (Table 3, Figure 3, and Supplemental Figure S1). For example, in the 20,000-sample cohort, the pathogenic *MSH2* splicing mutation c.942+3A>T, located in the 3′ end of exon 5 in a difficult-to-sequence homopolymer stretch of 27 adenines, was missed in 83% of patients harboring the mutation (five of six) when the false-positive rate was adjusted to zero (Figure 4). Likewise, *CHEK2*, which has several highly homologous pseudogenes located on chromosomes 2, 7, 10, 13, 15, 16, X, and Y, had a false-negative rate of 2.7% in the simulated study.[22] Similar to false-positives, many of the false-negatives identified in the study were located in complex genomic regions, with 45.5% located in A/T-rich regions, 8.5%, in G/C-rich regions; 2.8%, in homopolymer stretches; 8%, in pseudogene regions; and 35.2% had no discernable genomic region cause. In addition, there were seven mosaic mutations missed in *ATM* ($n = 4$), *BRCA1* ($n = 1$), *CHEK2* ($n = 1$), and *RAD51D* ($n = 1$) (Supplemental Table S3). Variants were identified as mosaic based on NGS low heterozygous read ratios and Sanger-confirmation chromatogram analysis (Supplemental Figure S2).

The mean false-negative heterozygous ratio was 40% (range, 10% to 69%) and mean coverage, 123× (range,

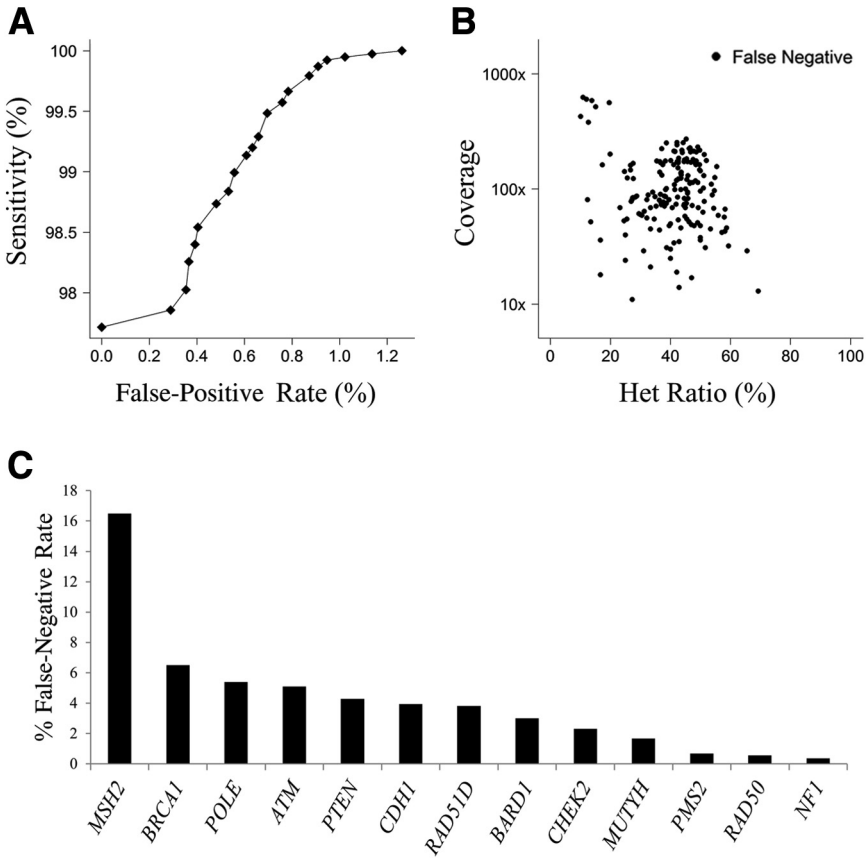


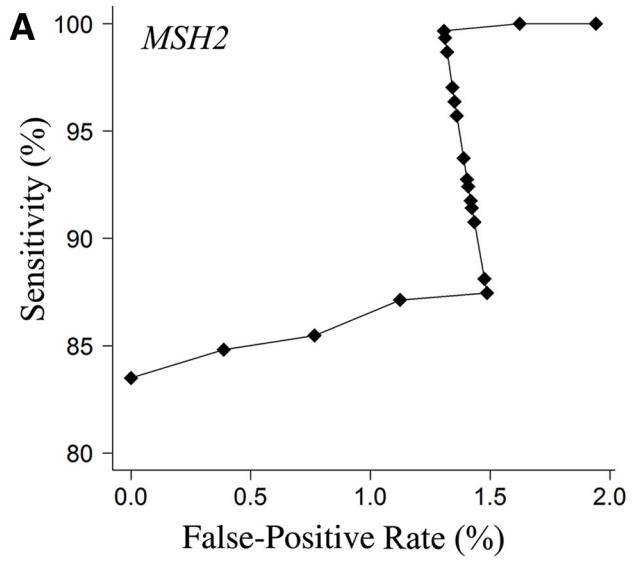
Figure 3 Simulated analysis illustrating the loss of assay sensitivity when the false-positive rate is adjusted to zero. **A:** Sensitivity versus false-positive rate. As the false-positive rate is reduced from a baseline value of 1.26% to 0%, the sensitivity of the assay is reduced from 100% to 97.8%. **B:** The coverage and heterozygous read ratio profile of false-negatives when the false-positive rate is reduced to zero. **C:** Gene-specific false-negative rates when the false-positive rate is reduced to zero. Those genes in which no false-negatives occur in the simulated study are excluded. Het, heterozygous.

11× to 625×) (Figure 3). Importantly, approximately 18% ($n = 32$) of false-negatives had heterozygous read ratios below 30% and 56% ($n = 98$) and coverage under 100×, metrics similar to those attributed to false-positives (Figure 3). Therefore, true variants can often have low coverage and low heterozygous ratios, and therefore are indistinguishable, at many loci, from false-positives. Having a bioinformatics pipeline with 100% specificity will also filter out true variants with balanced heterozygous ratios and good coverage. In the study, 30.7% ($n = 54$) of false-negatives had a heterozygous ratio of ≥40% and coverage of ≥100× (Figure 3). Due to the rare population frequency of suspected disease-causing variants, identifying the balance between specificity and sensitivity can be achieved only after a significant number of samples have been processed and confirmed. For example, 94% ($n = 166$) of false-negatives in the adjusted data set from our internal database had a variant frequency of ≤0.1%, or 1 in 1000 samples tested in a high-risk cohort. Moreover, 68% ($n = 119$) were observed in only <1 in 10,000 samples, or ≤0.01%.

Discussion

The ability to simultaneously sequence large numbers of genes implicated in disease quickly and in a cost-effective manner has enabled NGS to rapidly replace Sanger sequencing as the method of choice in diagnostic-testing laboratories. As a result, a multitude of gene-sequencing panels have entered the market, with varying degrees of sensitivity and specificity depending on numerous factors, such as the technologies used for sample enrichment and sequencing, variant-confirmation methods, bioinformatics pipeline, and variant interpretation. For hereditary-cancer testing, high sensitivity and specificity of an assay are essential as clinicians rely on the data to make important clinical management decisions such as costly specialized surveillance, treatment selection, and preventive surgery. Therefore, missing any cancer-related mutation or reporting out a false-positive result can have drastic health care consequences for patients and their families.

There has been a lot of debate regarding the necessity of confirming variants detected by NGS with a secondary technology such as Sanger confirmation. Currently, there are no specific guidelines on how diagnostic-testing laboratories should incorporate confirmatory methods. Several studies have been published recently detailing that Sanger-sequencing confirmation of NGS variants should not be required for variants with high coverage and Q scores. However, these studies have been very small, analyzing Sanger-confirmation results in only around 100 samples.



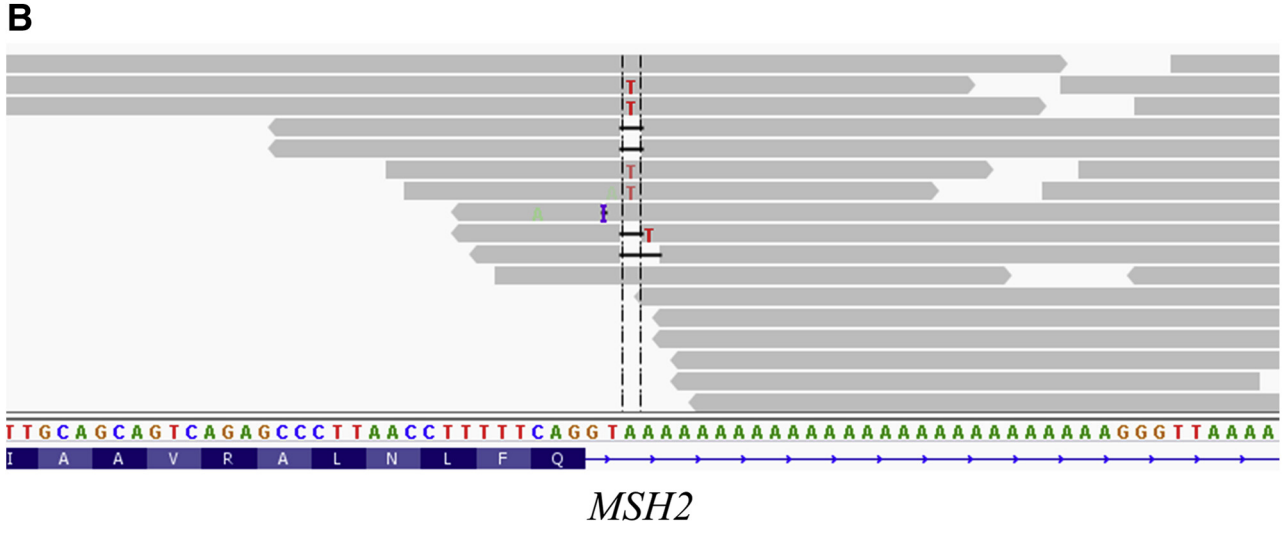


Figure 4 The ability to detect *MSH2* mutations is reduced when the false-positive rate is adjusted to zero. **A:** As the false-positive rate of *MSH2* is reduced from a baseline value of 2% to 0%, sensitivity is reduced from 100% to 83.5%. **B:** Integrative Genomics Viewer screenshot of a true pathogenic *MSH2* c.942+3A>T mutation located in a difficult-to-analyze homopolymer stretch of 27 adenine residues, which is frequently missed when the false-positive rate is adjusted to zero. **Horizontal black bars** represent false deletion calls.

Furthermore, these studies used variant Q score, which is ambiguous and highly specific to their workflow, as one of the main metrics to determine threshold settings.[9,10]

Here we describe the largest study to date, with 20,000 hereditary-cancer panel samples analyzed by NGS and nearly 8000 variants Sanger-confirmed. Our data illustrate that a conservative quality threshold for high-confidence NGS calls is a minimal read depth coverage of >100× and a variant allele frequency (or heterozygous ratio) of >40%. The calls satisfying these thresholds are confident to be reported out by NGS without secondary confirmation. Variant calls not meeting these metrics should be Sanger-sequenced to ensure the highest possible sensitivity and specificity of the assay. Importantly, these proper thresholds can be achieved only after the processing of thousands of samples because clinically relevant variants are extremely rare.

To limit Sanger-sequencing variant confirmation further, one could include manual Integrative Genomics Viewer (IGV) (Broad Institute) data analysis in the review process. For example, one could lower the heterozygosity ratio threshold to 25%, while keeping the coverage cutoff the same, and implement IGV review of variants >25% and <40%. However, we would advise caution using this approach extensively in a high-volume diagnostics setting as it requires extensive training and incorporates subjective judgment by an individual, which could result in additional errors and test-processing delays. Furthermore, reducing the threshold in regions known to have pseudogene homology that can confound sequencing results could be problematic. For example, *SMAD4* has a processed pseudogene found with a frequency of 0.26%, which can confound NGS alignment and result in erroneous variant calls.[23] Considering that the majority of false-positives occur in difficult-to-sequence regions, it could be challenging to reliably discern between false-positives and true calls by visual inspection of the IGV data.

The NGS false-positives are primarily due to false variants in regions difficult to sequence and align such as A/T- and G/C-rich regions, homopolymer stretches, and pseudogenes. The adjustment of our bioinformatics pipeline to reflect zero false-positives resulted in missing 2.2% of true clinically relevant mutations, with certain genes impacted more than others. Our data clearly illustrate the dangers of eliminating Sanger confirmation completely from diagnostic testing. Additionally, NGS false-positives can be dependent on capture chemistry, sequencing platform, and analytical pipeline. For example, false-positives can be more problematic in primer-based target enrichment than in probe-base enrichment due to the inability to remove PCR duplicates in the resulting data. In our data set, we did not see a difference in the false-positive rate when using different sequence aligners such as the Burrows-Wheeler Aligner version 0.7.12 (*https://sourceforge.net/projects/bio-bwa/files*) or Bowtie2 (Johns Hopkins University, Baltimore, MD). We also did not observe any significant difference in the results between the HiSeq2500 and NextSeq500 sequencing platforms, with false-positive rates of 1.3% and 1.2%, respectively.

It is concerning, the number of publications and commercially offered clinical diagnostic NGS tests that claim zero false-positives from the NGS panel data alone.[3–5] Our data illustrate that Sanger confirmation is needed to keep accuracy high in difficult-to-sequence regions. In addition, laboratories that keep sensitivity high but do not Sanger-confirm variants within specific parameters will report out false-positives. It is understandable why high-volume clinical diagnostics laboratories would want to limit Sanger confirmation of NGS variants as it requires significant time and resources. A slew of new diagnostics companies have entered the market recently, offering low-priced tests often at the expense of accuracy by avoiding high-priced steps such as Sanger-sequencing confirmations and microarrays for calling deletions and duplications. Likewise, as insurance companies continue to pressure reimbursement of NGS tests, laboratories could be forced to limit pricey confirmation steps, microarrays, and other side-assays that impact testing accuracy. It is estimated that Sanger confirmation costs ~$240 per sample and adds an additional approximately 2 days to test-turnaround times.[9] These costs are not trivial for high-volume laboratories. Based on these estimates, in our 20,000-sample data set described here, we incurred a cost of $1.9 million to include Sanger-sequencing confirmation. As a high-volume diagnostics laboratory, we recommend setting appropriate quality thresholds for Sanger confirmation only after processing thousands of samples to accurately determine problematic regions. To keep the sensitivity and specificity of the assay optimal, Sanger confirmation should be maintained for all variants not meeting these thresholds.

Supplemental Data

Supplemental material for this article can be found at *http://dx.doi.org/10.1016/j.jmoldx.2016.07.006*.

References

1. Sikkema-Raddatz B, Johansson LF, de Boer EN, Almomani R, Boven LG, van den Berg MP, van Spaendonck-Zwarts KY, van Tintelen JP, Sijmons RH, Jongbloed JD, Sinke RJ: Targeted next-generation sequencing can replace Sanger sequencing in clinical diagnostics. Hum Mutat 2013, 34:1035–1042
2. Chong HK, Wang T, Lu HM, Seidler S, Lu H, Keiles S, Chao EC, Stuenkel AJ, Li X, Elliott AM: The validation and clinical implementation of BRCAplus: a comprehensive high-risk breast cancer diagnostic assay. PLoS One 2014, 9:e97408
3. Walsh T, Lee MK, Casadei S, Thornton AM, Stray SM, Pennil C, Nord AS, Mandell JB, Swisher EM, King MC: Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. Proc Natl Acad Sci U S A 2010, 107:12629–12633
4. Lincoln SE, Kobayashi Y, Anderson MJ, Yang S, Desmond AJ, Mills MA, Nilsen GB, Jacobs KB, Monzon FA, Kurian AW, Ford JM, Ellisen LW: A systematic comparison of traditional and multigene panel testing for hereditary breast and ovarian cancer genes in more than 1000 patients. J Mol Diagn 2015, 17:533–544
5. Judkins T, Leclair B, Bowles K, Gutin N, Trost J, McCulloch J, Bhatnagar S, Murray A, Craft J, Wardell B, Bastian M, Mitchell J, Chen J, Tran T, Williams D, Potter J, Jammulapati S, Perry M, Morris B, Roa B, Timms K: Development and analytical validation of a 25-gene next generation sequencing panel that includes the BRCA1 and BRCA2 genes to assess hereditary cancer risk. BMC Cancer 2015, 15:215. [Internet] doi:10.1186/s12885-015-1224-y
6. Hall MJ, Forman AD, Pilarski R, Wiesner G, Giri VN: Gene panel testing for inherited cancer risk. J Natl Compr Canc Netw 2014, 12:1339–1346
7. Aziz N, Zhao Q, Bry L, Driscoll DK, Funke B, Gibson JS, Grody WW, Hegde MR, Hoeltge GA, Leonard DG, Merker JD, Nagarajan R, Palicki LA, Robetorye RS, Schrijver I, Weck KE, Voelkerding KV: College of American Pathologists' laboratory standards for next-generation sequencing clinical tests. Arch Pathol Lab Med 2015, 139:481–493
8. Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Friez MJ, Funke BH, Hegde MR, Lyon E: ACMG clinical laboratory standards for next-generation sequencing. Genet Med 2013, 15:733–747
9. Strom SP, Lee H, Das K, Vilain E, Nelson SF, Grody WW, Deignan JL: Assessing the necessity of confirmatory testing for exome-sequencing results in a clinical molecular diagnostic laboratory. Genet Med 2014, 16:510–515
10. Baudhuin LM, Lagerstedt SA, Klee EW, Fadra N, Oglesbee D, Ferber MJ: Confirming variants in next-generation sequencing panel testing by Sanger sequencing. J Mol Diagn 2015, 17:456–461
11. Lek M, Karczewski K, Minikel E, Samocha K, Banks E, Fennell T, et al; Exome Aggregation Consortium: Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016, 536:285–291
12. Kuiper RP, Vissers LE, Venkatachalam R, Bodmer D, Hoenselaar E, Goossens M, et al: Recurrence and variability of germline EPCAM deletions in Lynch syndrome. Hum Mutat 2011, 32:407–414
13. Jaeger E, Leedham S, Lewis A, Segditsas S, Becker M, Cuadrado PR, Davis H, Kaur K, Heinimann K, Howarth K, East J, Taylor J, Thomas H, Tomlinson I: Hereditary mixed polyposis syndrome is caused by a 40-kb upstream duplication that leads to increased and ectopic expression of the BMP antagonist GREM1. Nat Genet 2012, 44:699–703

14. Bertolotto C, Lesueur F, Giuliano S, Strub T, de Lichy M, Bille K, et al: A SUMOylation-defective MITF germline mutation predisposes to melanoma and renal carcinoma [Erratum appeared in *Nature* 2016, 531:126]. Nature 2011, 480:94–98
15. Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, Gabriel S, Rieder MJ, Altshuler D, Shendure J, Nickerson DA, Bamshad MJ; NHLBI Exome Sequencing Project, Akey JM: Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature 2013, 493:216–220
16. 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA: An integrated map of genetic variation from 1,092 human genomes. Nature 2012, 491: 56–65
17. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL: 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 2015, 17:405–424
18. Tavtigian SV, Greenblatt MS, Goldgar DE, Boffetta P: IARC Unclassified Genetic Variants Working Group: Assessing pathogenicity: overview of results from the IARC Unclassified Genetic Variants Working Group. Hum Mutat 2008, 29:1261–1264
19. Benjamini Y, Speed TP: Summarizing and correcting the GC content bias in high-throughput screening. Nucleic Acids Res 2012, 40:e72
20. Ratan A, Miller W, Guillory J, Stinson J, Seshagiri S, Schuster SC: Comparison of sequencing platforms for single nucleotide variant calls in a human sample. PLoS One 2013, 8:e55089
21. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y: A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 2012, 13:341. [Internet] doi:10.1186/1471-2164-13-341
22. Sodha N, Williams R, Mangion J, Bullock SL, Yuille MR, Eeles RA: Screening hCHK2 for mutations. Science 2000, 289:359
23. Millson A, Lewis T, Pesaran T, Salvador D, Gillespie K, Gau CL, Pont-Kingdon G, Lyon E, Bayrak-Toydemir P: Processed pseudogene confounding deletion/duplication assays for SMAD4. J Mol Diagn 2015, 17:576–582