The Journal of Molecular Diagnostics, Vol. 20, No. 5, September 2018

the Journal of Molecular Diagnostics
jmd.amjpathol.org

Determining Performance Metrics for Targeted Next-Generation Sequencing Panels Using Reference Materials

Megan H. Cleveland,* Justin M. Zook,* Marc Salit,*† and Peter M. Vallone*

From the Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, Maryland; and the Joint Initiative for Metrology in Biology,† Stanford, California*

Accepted for publication April 26, 2018.

Address correspondence to Megan H. Cleveland, Ph.D., National Institute of Standards and Technology, 100 Bureau Dr., Mail Stop 8314, Gaithersburg, MD 20899-8314. E-mail: megan.cleveland@nist.gov.

The National Institute of Standards and Technology has developed reference materials for five human genomes. DNA aliquots are available for purchase, and the data, analyses, and high-confidence small variant and homozygous reference calls are freely available on the web. These reference materials are useful for evaluating whole-genome sequencing methods and also can be used to benchmark targeted sequencing panels, which are used commonly in clinical settings. This article describes how to use the Genome in a Bottle samples to obtain performance metrics on any germline-targeted sequencing panel of interest, as well as the limitations of the reference materials. These materials are useful for understanding the limitations of, and optimizing, targeted sequencing panels and associated bioinformatics pipelines. Example figures are presented to illustrate ways to access the performance metrics of targeted sequencing panels, and a table of best practices is included. *(J Mol Diagn 2018, 20: 583—590; https://doi.org/10.1016/j.jmoldx.2018.04.005)*

In 2015, The National Institute of Standards and Technology (NIST, Gaithersburg, MD) released the first Genome in a Bottle (GIAB) reference material (RM): 8398. To generate this reference material, human genomic DNA from a large batch of GM12878 cells was extracted and aliquoted at the Coriell Institute for Medical Research (Camden, NJ). These homogeneous DNA aliquots were sequenced by multiple unique technologies, each with different capabilities and biases, to obtain a high-confidence truth set of small variant and homozygous reference calls.[1] In 2016, NIST released four additional human genomes as reference materials, a son-father-mother trio of Ashkenazi Jewish ancestry (RMs 8391 and 8392) and a son in a trio of Chinese ancestry (RM 8393), along with high-confidence calls and regions.[2,3] All five genomes used for these NIST RMs also are available publicly from the Coriell Institute for Medical Research as cell lines. Together, these DNA samples and truth sets can be used as reference materials to evaluate assays and analytic pipelines. When the results of a pipeline ("query") are compared with the truth set, most false positives and false negatives should be errors in the query set. The current high-confidence calls and regions cover approximately 90% of the sequence in GRCh37 and GRCh38 (the two most recent human releases from the Genome Reference Consortium), but tend to exclude large variants, long tandem repeats, and regions difficult to map with short reads. Ongoing work in GIAB is using new methods to characterize these more challenging variants and regions. The raw data, analyses, and high-confidence calls and regions are freely available online (Genome in a Bottle, *www.genomeinabottle.org*, last accessed March 12, 2018). These genomes and associated data have been widely used in the next-generation

Supported by internal funding from the National Institute of Standards and Technology and from the US Department of Commerce.

Disclosures: Certain commercial equipment, instruments, and materials are identified to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose. Information presented does not necessarily represent the official position of the National Institute of Standards and Technology. All work presented has been reviewed and approved by the National Institute of Standards and Technology Human Subjects Protections Office.

https://doi.org/10.1016/j.jmoldx.2018.04.005

sequencing (NGS) community to obtain performance metrics on whole-genome and whole-exome sequencing methods.[4–6] The Global Alliance for Genomics and Health (GA4GH) Benchmarking Team has standardized performance metrics and developed sophisticated variant comparison tools to compare variant calls and output these metrics.[7]

In addition to their use in evaluating whole-genome and whole-exome sequencing methods,[8–10] the GIAB reference materials also can be used with targeted sequencing panels. Next-generation targeted sequencing panels increasingly are being used for clinical purposes because of the higher number of targets that can be covered, relative to Sanger sequencing. Targeted sequencing also has several advantages relative to whole-genome sequencing or exome sequencing, including higher coverage for genes of interest at lower cost, and faster analysis time. Targeted sequencing panels have been used clinically for a wide variety of conditions including cystic fibrosis, epilepsy, cardiomyopathies, inherited cancers, disorders of sex development, autoinflammatory diseases, ataxia, retinal disorders,[11–17] and many others. To ensure the accuracy of these tests, laboratories need well-characterized reference materials and associated data sets for test development, validation, and quality control.

The recent publication by Roy et al,[18] "Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines," recommends the use of reference materials. In this article, we describe how the GIAB reference materials can be used to benchmark specific targeted sequencing panels. As an example, germline sequencing panels were selected based on two different library preparation techniques: hybrid capture, which uses oligo probes to capture regions of interest; and amplicon-based, which uses PCR to amplify the regions of interest. This work is not intended to be a comprehensive performance assessment of these methods or a comparison between platforms.

Materials and Methods

DNA Samples

This study used the five genomes contained within three NIST RMs: RM 8398, RM 8392, and RM 8393 (NIST). Each RM contains a 50-μL DNA aliquot at a concentration of approximately 200 ng/μL. RM 8398 contains extracted DNA from a large homogeneous batch of the GM12878 cell line. RM 8392 contains three separate tubes of DNA extracted from homogeneous large batches of three cell lines (GM24143, GM24149, and GM24385) derived from a mother-father-son Ashkenazim Jewish Trio, which is part of the Personal Genome Project. RM 8393 contains extracted DNA from a cell line (GM24631) derived from a male individual of Chinese descent, who is also part of the Personal Genome Project.

Library Preparation and Sequencing

Hybrid-Capture Library Preparation and Sequencing

Library preparation for the hybrid-capture method was performed with the TruSight Rapid Capture kit (FC-140-1104; Illumina, San Diego, CA) and the TruSight Inherited Disease Sequencing Panel (FC-121-0205; Illumina) according to the manufacturer's protocol.

Briefly, DNA was tagmented (a combination of DNA fragmentation and end-polishing, using transposons), adapters and barcodes were added, and then three to eight libraries were pooled for hybridization (varying numbers of libraries were pooled to obtain a broad range of sequencing depths). The library pool was hybridized twice with Inherited Disease Panel Oligos at 58°C. After library preparation, the library was checked on a 2100 Bioanalyzer high-sensitivity DNA chip (5067-4626; Agilent, Santa Clara, CA) to assess the quality before sequencing. The DNA concentration was measured with the Qubit high-sensitivity DNA assay (Q32851; Thermo Fisher Scientific, Waltham, MA), diluted to 4 nmol/L, and denatured with 0.2 mol/L NaOH. PhiX DNA (FC-110-3001; Illumina) was spiked in at 5% v/v. The denatured library then was sequenced with a MiSeq Reagent Kit (MS-102-3003; Illumina) for 300 cycles (2 × 150 bp) on an Illumina MiSeq or an Illumina ForenSeq.

Amplicon Library Preparation and Sequencing

For the amplicon sequencing, the Ion AmpliSeq Library Kit 2.0 (4475345; Thermo Fisher Scientific) and the AmpliSeq Inherited Disease Panel (4477686; Thermo Fisher Scientific) were used in accordance with the manufacturer's protocol. DNA from each genome was amplified in three separate primer pools, and then these PCR products were combined for barcoding and library preparation. The concentration of the final library was measured with the Ion Library TaqMan Quantification Kit (4468802; Thermo Fisher Scientific), and then two libraries were adjusted to a concentration of 40 pmol/L and combined before chip loading. Then 318v2 BC chips (4488146; Thermo Fisher Scientific) were loaded using the Ion Chef and sequenced on the Personal Genome Machine using the Ion PGM Hi-Q Chef kit (A25948; Thermo Fisher Scientific).

Variant Calling

Sequence variants were identified and stored in variant call format (VCF) files using the included commercial software. MiSeq Reporter software (BWA Enrichment version 2.5.1.3; Illumina) was used to generate the VCF files for the hybrid capture targeted sequencing.

Torrent Suite software version 5.0.5 (Thermo Fisher Scientific) was used to generate VCF files for the amplicon sequencing.

Original data of this study are at available at Mendeley Data (*https://doi.org/10.17632/7h9gmhnrdg.1*).

Data Analysis

After generation, the VCF files were compared with the GIAB high-confidence VCF files using the GA4GH Benchmarking application (PrecisionFDA, registration required, *http://precision.fda.gov*, last accessed March 12, 2018). The

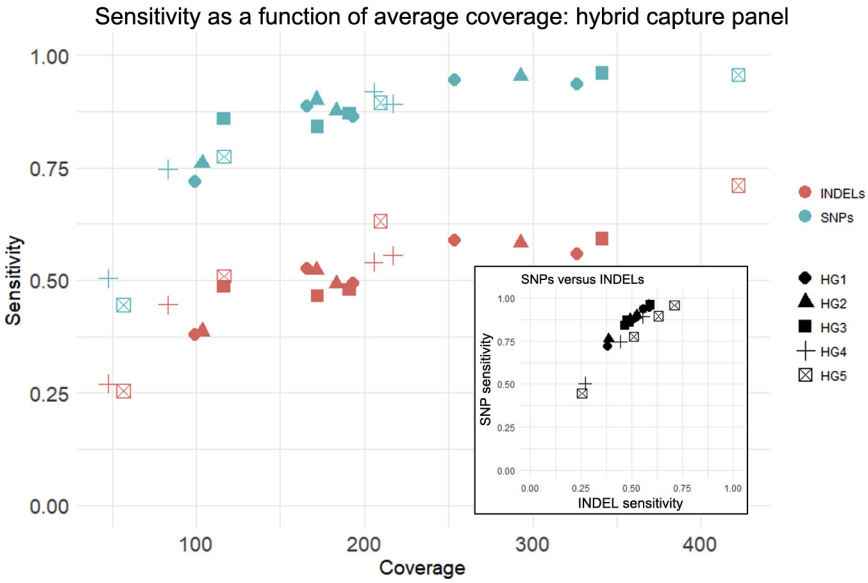


Figure 1 Effect of average coverage on sensitivity. As average coverage increases, sensitivity for both single-nucleotide polymorphisms (SNPs) and insertion/deletions (INDELs) increases. SNP sensitivity is higher than INDEL sensitivity. SNP sensitivity and INDEL sensitivity are correlated strongly (**inset**).

GA4GH Benchmarking Team developed standardized performance metrics for genomic variant calls as well as sophisticated variant comparison tools to robustly compare different representations of the same variant, and a set of standard browser extensible data (BED) files describing difficult genome contexts to stratify performance. The GA4GH Benchmarking application requires a truth VCF file (the GIAB high-confidence VCF file), the truth confident regions (the GIAB high-confidence BED file), the query VCF file (generated by the included commercial software), and the target regions (the BED file provided by the manufacturer for the targeted sequencing panel). All GIAB files (VCF files and BED files) are available (Genome in a Bottle, *https://github.com/genome-in-a-bottle*, last accessed March 12, 2018). The

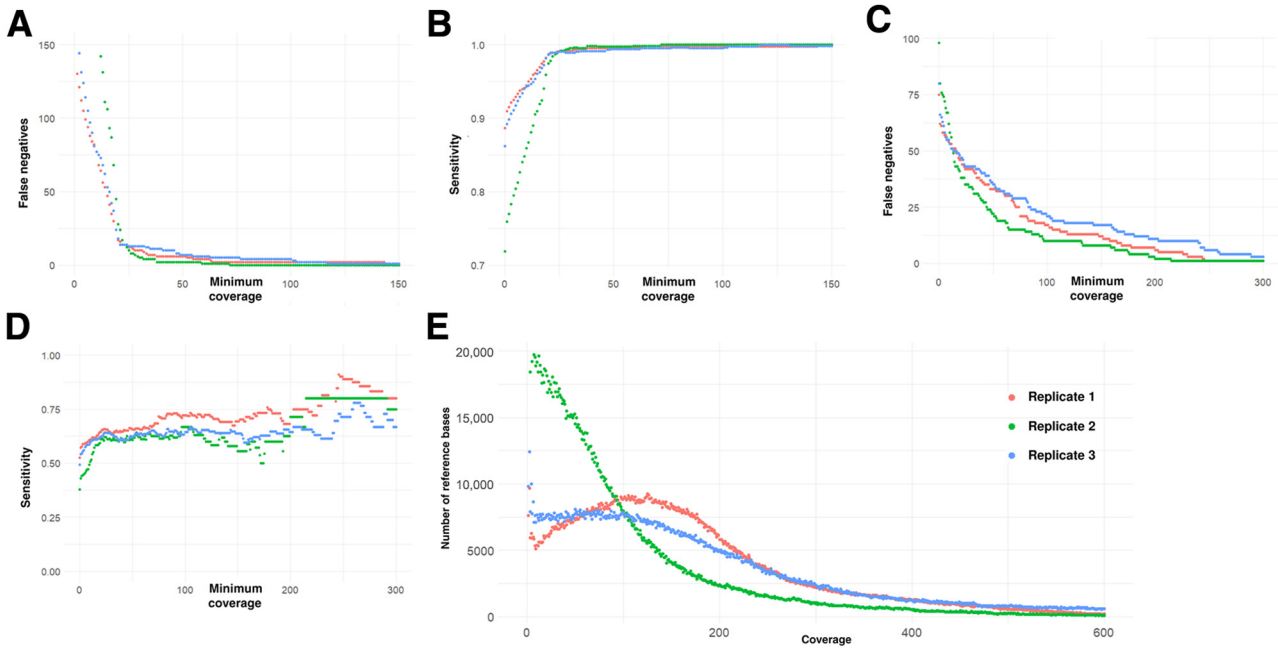


Figure 2 False negatives, sensitivity, and coverage at each site inside targeted regions. **A:** Single-nucleotide polymorphism (SNP) false-negative sites with minimum locus coverage at or below the coverage on the *x* axis; there are a few SNP false negatives at locus coverages greater than 50×. **B:** SNP sensitivity with minimum locus coverage at or below the coverage on the *x* axis; SNP sensitivity is greater than 98% for loci with a coverage of 25× or higher. **C:** Insertion/deletion (INDEL) false negatives with minimum locus coverage at or below the coverage on the *x* axis; there are still a significant number of false negatives at loci with coverages greater than 100×, indicating that read depth is not the only factor affecting INDEL detection. **D:** INDEL sensitivity with minimum locus coverage at or below the coverage on the *x* axis; even for loci covered at 200×, INDEL sensitivity does not exceed 75%. **E:** This histogram shows the distribution of read depths over the total number of reference bases in the manufacturer's BED file. In replicate 2 (green), the average coverage is lower and more reference bases are covered at less than 25× compared with replicates 1 and 3.

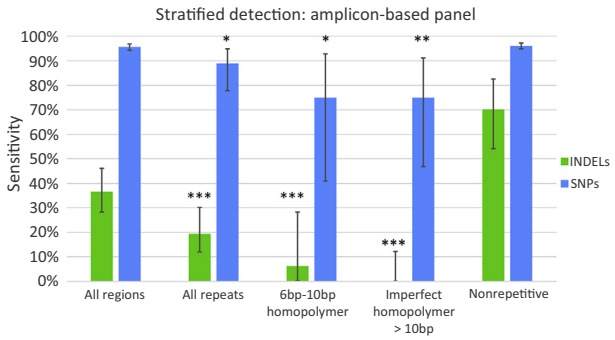


Figure 3 Insertion/deletion (INDEL) and single-nucleotide polymorphism (SNP) sensitivity stratified by region type. Overall, the INDEL sensitivity for this replicate of the amplicon-based panel assay with 273× coverage was 36% for INDELS and 96% for SNPs. INDEL sensitivity was significantly higher in nonrepetitive regions compared with all repetitive regions, 6- to 10-bp homopolymer regions, and imperfect homopolymer regions **Vertical black lines** indicate 95% CIs. SNP sensitivity also was significantly higher in nonrepetitive regions compared with all repetitive regions, 6- to 10-bp homopolymer regions, and imperfect homopolymer regions. $*P < 0.05$, $**P < 0.005$, and $***P < 0.0005$ (Benjamini-Hochberg procedure).

GA4GH application returns the count of false negatives (FNs), false positives (FPs), and true positives (TPs) in both standardized VCF and comma-separated value formats. Performance metrics follow the GA4GH standardized definitions, in which genotyping errors are counted both as FP and FN. In addition, the GA4GH application stratifies performance metrics by variant type, size, and genome context to enable understanding strengths and weaknesses of a method. Sensitivity was calculated using the following formula: sensitivity = TP/(TP + FN).

Sensitivity greater than the specific minimum coverage, X, was calculated by only including TPs and FNs at sites with coverage greater than or equal to X. Coverage analysis of each locus and common false negatives shared among replicates were determined using BEDtools.[19] The precisionFDA output VCF was first split into three files: the FNs, FPs, and TPs. Next, the BEDtools "coverage" command was used to determine the coverage at each FN and TP location. The BEDtools "multiinter" command was used to identify FNs shared between different replicates of the same genome. The number of common FNs are represented using the Venn Diagram Plotter (Pacific Northwest National Laboratory, Richland, WA).

CIs for stratified regions were calculated using the Fisher exact test in R.

FNs and FPs in the binary alignment map and VCF files were visualized using Golden Helix GenomeBrowse version 2.1.2 (Golden Helix Inc., Bozeman, MT).

Results

Effect of Average and Locus Coverage on Sensitivity

Sensitivity increased with increasing average coverage for both single-nucleotide polymorphisms (SNPs) and insertions/deletions (INDELs). With the hybrid capture sequencing, a maximum SNP sensitivity rate of 96% and a maximum INDEL sensitivity rate of 71% was observed at the highest mean coverage (422×). For all genomes examined, INDEL sensitivity was lower than SNP sensitivity (Figure 1). The average coverage and sensitivity were similar across replicates for the amplicon-based panel (Supplemental Figure S1). Depending on the sample, 82% to 93% of INDELs in the truth set were 1 to 5 bp in size, therefore analysis of larger INDELs was limited. For each sample, there were only 7 to 17 INDELs between 6 and 15 bp in size, and 1 to 6 INDELs larger than 15 bp in size.

The number of FNs, TPs, and sensitivity within individual data sets also was analyzed when excluding loci below a varying coverage threshold (Figure 2). For SNPs, almost all FNs occurred at locus coverage of 50× or less. For INDELs, many FNs remained even with high locus coverages (>100×). Further stratification by genome context therefore was performed to gain insight into the causes for these FNs.

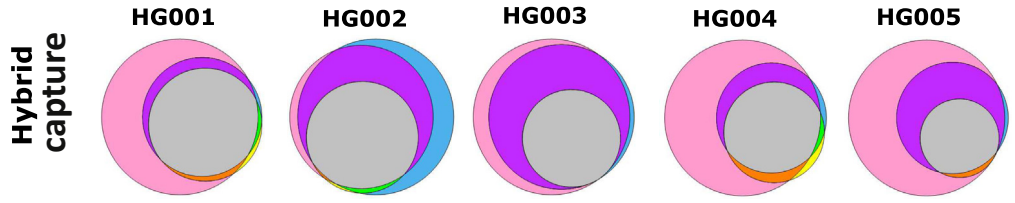
SNP sensitivity and INDEL sensitivity were examined over various stratified regions (Figure 3) in the amplicon-based sequencing. Compared with regions with higher complexity, INDEL sensitivity decreased significantly in repetitive regions, imperfect homopolymers >10 bp, and perfect homopolymers >10 bp (with no INDEL detection at all in the latter). The overall INDEL detection rate at 273× coverage was 36%; however, this increased to 70% in regions that were of higher complexity (with no repeats, homopolymers, or imperfect homopolymers). SNP sensitivity followed a similar pattern. The overall SNP sensitivity was 96%, but SNP sensitivity decreased to 75% in 6- to 10-bp homopolymer regions.

Consistency of FNs between Replicates

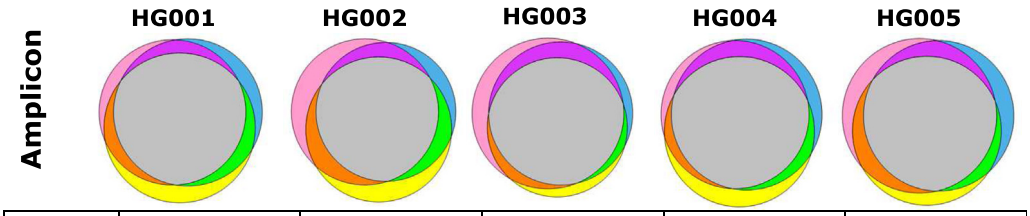
For both amplicon and hybrid-capture panels, the locations of FN calls were similar between replicates (Figure 4). The number of total FNs varied more in the hybrid capture assay because there was more variability in average coverage; however, almost all FNs contained within the higher-coverage replicates also occurred in the lower-coverage replicates. For amplicon-based sequencing, the average coverage was very similar between replicates, and approximately 40% of FNs were shared by all replicates.

Causes of FPs

FPs were less common than FNs and tended to occur near actual variants, in repetitive regions, and near the ends and beginnings of reads. In the example shown (Figure 5), the true variant is a complex compound heterozygous mutation. The GIAB high-confidence VCF shows that for the Ashkenazi son, there was a 2-bp insertion on the paternal allele and a 4-bp insertion followed by a G to A SNP on the maternal allele. The variant caller incorrectly called this location as simply having a heterozygous (G/A) SNP.



	HG001		HG002		HG003		HG004		HG005	
	Unique FNs for replicate, %	Average coverage for replicate	Unique FNs for replicate, %	Average coverage for replicate	Unique FNs for replicate, %	Average coverage for replicate	Unique FNs for replicate, %	Average coverage for replicate	Unique FNs for replicate, %	Average coverage for replicate
Replicate 1	41.1	193×	4.3	172×	15.6	172×	47.4	83×	48.2	116×
Replicate 2	0.4	326×	18.2	183×	2.1	191×	1.0	218×	0.8	209×
Replicate 3	0.7	253×	0.4	293×	0.0	293×	1.2	206×	0.0	422×



	HG001		HG002		HG003		HG004		HG005	
	Unique FNs for replicate, %	Average coverage for replicate	Unique FNs for replicate, %	Average coverage for replicate	Unique FNs for replicate, %	Average coverage for replicate	Unique FNs for replicate, %	Average coverage for replicate	Unique FNs for replicate, %	Average coverage for replicate
Replicate 1	5.5	273×	12.5	277×	12.3	264×	4.4	258×	9.2	272×
Replicate 2	6.8	256×	6.3	245×	5.3	239×	6.7	269×	8.4	244×
Replicate 3	9.6	231×	10.4	191×	6.1	316×	10.4	281×	5.9	225×

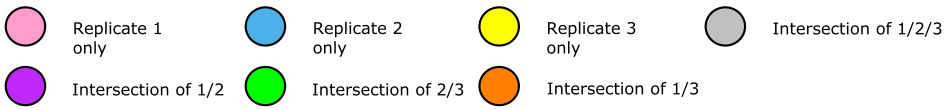


Figure 4 False negatives (FNs) shared across replicates. Venn diagrams show the overlap of FNs between replicates. In addition, a table that shows the average coverage for each replicate and the percentage of FNs that were unique to that replicate is included. For both amplicon and hybrid-capture panels, FNs appear to be nonrandom; a high number of the same FNs appear in multiple replicates, with lower-coverage replicates having most of the same FNs as higher-coverage replicates, plus additional FNs.

Discussion

Targeted sequencing panels are increasingly used in clinical settings because they offer higher coverage depth at a lower cost, relative to whole-genome and whole-exome sequencing. There are two main types of targeted sequencing panels: probe capture–based and amplicon based. Selection of efficient probes or primers, careful library preparation, and appropriate bioinformatic pipelines all have impacts on panel sensitivity.[20] Potential pathogenic variants typically are confirmed using Sanger sequencing,[21] which can identify NGS FPs. These NGS FPs often occur near true variants, which often may be identified by follow-up Sanger sequencing if the FP was flagged for follow-up evaluation. There is some debate about whether Sanger sequencing is necessary when specific conditions are met by the NGS sequencing.[22] Minimizing FNs also is important, and it is critical both to ensure sufficient coverage at every locus and assess whether the pipeline can detect more difficult variants even at high coverage.

As shown in this work, GIAB benchmark genomes can be used to evaluate a targeted sequencing panel of interest. Multiple sequencing replicates, with five different genomes, were performed on both hybrid-capture and amplicon-based

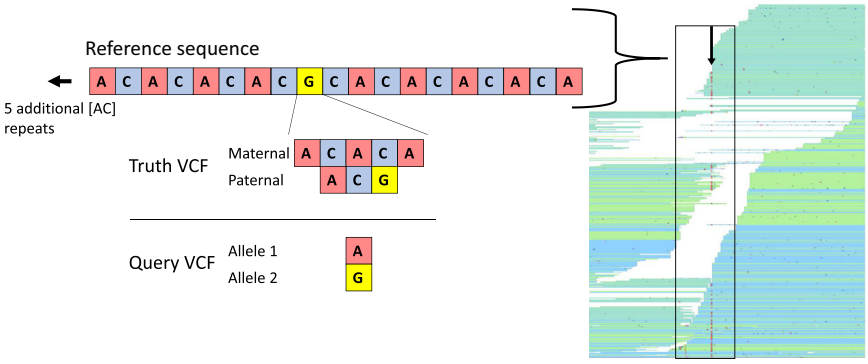


Figure 5 False positive (and false negative) call near a true variant. False positives were most likely to occur in repetitive regions, near the ends of reads and near true variants. In the example shown here, there is a 4-bp insertion on the maternal allele, followed by a G to A SNP, and a 2-bp insertion on the paternal allele. The variant caller incorrectly identified only a G to A SNP. The region has 10 [AC] repeats preceding the variant and 5 [CA] repeats after the variant. The read pileup on the **right** is shown to indicate the location of the miscalled variant (**arrow**, marked in red), which is near the beginnings and ends of the reads. The miscalled variant resulted from misaligned reads that do not encompass the entire repeat and its flanking sequences. VCF, variant call format.

sequencing panels. This allowed comparison of the results of the targeted sequencing panels with the GIAB high-confidence calls, using freely available bioinformatic data and tools. Overall sensitivity, site-specific sensitivity, FNs, and FPs were examined. The results were similar for both types of panels.

In the tested targeted sequencing panels, average coverage was found to be the main determinant of sensitivity, with the individual genome having no noticeable effect. Replicates with similar coverage have mostly the same FNs, with lower-coverage replicates having additional FNs. For the targeted sequencing panels examined in this study, low-coverage regions are not random—they likely are caused by either inefficient PCR primers in amplicon sequencing or inefficient capture probes in hybrid capture sequencing.

On a per site level, in the assays tested, it was observed that most SNP FNs were caused by low coverage; this shows the usefulness of evaluating the effect of coverage on the FN rate. If one excludes all regions with low coverage, SNP sensitivity is very high. For instance, if only loci with coverage greater than 50× are considered, the SNP sensitivity is greater than 99% for all genome replicates examined. In contrast, only approximately half of all INDEL FNs appeared to be caused by low coverage. The GA4GH benchmarking tool's stratification functionality, which showed that INDEL FNs with high coverage mostly occurred in repetitive regions, was therefore used. Although there were few FPs in the targeted sequencing panels examined here, similar analyses and figures could be generated for FPs when more FPs occur.

For the targeted NGS panels examined, FPs did not occur randomly but instead were most likely to occur at or around complex variants, in repetitive regions, and near the beginnings and ends of reads. In contrast to whole-genome sequencing, targeted sequencing has reads that begin and end near the same location; the start and stop points are either centered around the capture probe, or occur at the ends of the PCR primer regions. For this reason, although a region within a targeted sequencing panel and region from whole-genome sequencing may be sequenced at the same coverage, it is more likely that reads in the targeted panel

Table 1 Best Practices for Using Reference Materials to Assess Performance of Targeted Assays

Recommendation	Explanation
Manual curation	Manually curate false positives and false negatives to help understand their source (eg, they are located near true variants, in repetitive regions, or at the edges of reads)
Identify low coverage	Determine how many false negatives are associated with low-coverage regions
Stratify	Stratify false negatives and false positives according to variant type and genome context (eg, homopolymers, tandem repeats, difficult-to-map regions)
CIs	Calculate CIs for performance metrics for variants of different types in different genome contexts because some variant types and genome contexts may have limited numbers of examples in targeted regions
Use additional samples	The GIAB samples are not intended to be used as the only validation method for clinical tests because there are a limited number of variants in the targeted regions of most clinical assays, and the variants in the GIAB samples are likely not representative of the variants of interest clinically
Use high-confidence BED file	The GIAB samples are useful for benchmarking, but comparisons generally should be made only within the high-confidence BED file
Most difficult regions are outside the BED file	The high-confidence regions are not yet comprehensive so they exclude the most difficult regions and variants

GIAB, Genome in a Bottle.

will have more nonrandom start and end points. When these start and end points occur in repetitive regions, it can be difficult for the variant caller to align the read properly and make the correct call. This might be eliminated with multiple primer sets and capture probes. These observations were true of the vendor-supplied pipelines used, but variance in performance between pipelines is expected. Example figures (Figures 1, 2, 3, 4, and 5) comparing the targeted panel calls with the GIAB benchmark calls are shown; these figures can help to highlight which factors are important for any pipeline.

One limitation of this work is that the GIAB high-confidence calls are biased toward the relatively simple calls. The high-confidence regions include a relatively small number of larger INDELs, especially in coding regions, and no structural variant or copy number variation calls. A panel may perform well over the GIAB high-confidence regions and still perform poorly on more difficult variants and difficult regions of the genome. Ideally, one should test a large number of variants of different types, sizes, and sequence contexts; this usually is possible for whole-genome sequencing with only a small number of benchmark genomes, but this small number of genomes is unlikely to contain enough variants for targeted sequencing tests. This is particularly important because some clinical tests are enriched for more difficult variants.[23]

The available GIAB genomes and bioinformatics data are a resource for benchmarking the performance of targeted clinical gene sequencing panels. These performance benchmarks then can be used to inform practical recommendations for the use of particular targeted sequencing panels (eg, necessary target coverage levels and the identification of regions where variant calls can be made with sufficient confidence). Finally, benchmark observations can suggest principles that may be used in the design of probes or primers for targeted sequencing panels, such as the need to avoid placing read boundaries in repetitive regions or the importance of knowing the limitations of the test in these regions. Table 1 outlines our recommendations for best practices.

Supplemental Data

Supplemental material for this article can be found at *https://doi.org/10.1016/j.jmoldx.2018.04.005*.

References

1. Zook JM, McDaniel J, Parikh H, Heaton H, Irvine SA, Truty R, Mclean CY, La Vega FMD, Salit M: Reproducible integration of multiple sequencing datasets to form high-confidence SNP, indel, and reference calls for five human genome reference materials. BioRxiv 2018, doi:10.1101/281006
2. Zook JM, Chapman B, Wang J, Mittelman D, Hofmann O, Hide W, Salit M: Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. Nat Biotechnol 2014, 32: 246−251
3. Zook JM, Catoe D, McDaniel J, Vang L, Spies N, Sidow A, et al: Extensive sequencing of seven human genomes to characterize benchmark reference materials. Sci Data 2016, 3:160025
4. Laurie S, Fernandez-Callejo M, Marco-Sola S, Trotta JR, Camps J, Chacón A, Espinosa A, Gut M, Gut I, Heath S, Beltran S: From wet-lab to variations: concordance and speed of bioinformatics pipelines for whole genome and whole exome sequencing. Hum Mutat 2016, 37:1263−1271
5. Cornish A, Guda C: A comparison of variant calling pipelines using genome in a bottle as a reference. Biomed Res Int 2015, 2015:456479
6. Cleary JG, Braithwaite R, Gaastra K, Hilbush BS, Inglis S, Irvine SA, Jackson A, Littin R, Rathod M, Ware D, Zook JM, Trigg L, De La Vega FMM: Comparing variant call files for performance benchmarking of next-generation sequencing variant calling pipelines. BioRxiv 2015, doi:10.1101/023754
7. Krusche P, Trigg L, Boutros PC, Mason CE, De La Vega FM, Moore BL, Gonzalez-Porta M, Eberle MA, Tezak Z, Lababidi S, Truty R, Asimenos G, Funke B, Fleharty M, Salit M, Zook JM: Best practices for benchmarking germline small variant calls in human genomes. BioRxiv 2018, doi:10.1101/270157
8. Hwang S, Kim E, Lee I, Marcotte EM: Systematic comparison of variant calling pipelines using gold standard personal exome variants. Sci Rep 2015, 5:1−8
9. Meynert AM, Ansari M, FitzPatrick DR, Taylor MS: Variant detection sensitivity and biases in whole genome and exome sequencing. BMC Bioinformatics 2014, 15:1−11
10. Linderman MD, Brandt T, Edelmann L, Jabado O, Kasai Y, Kornreich R, Mahajan M, Shah H, Kasarskis A, Schadt EE: Analytical validation of whole exome and whole genome sequencing for clinical applications. BMC Med Genomics 2014, 7:20
11. Eggers S, Sadedin S, van den Bergen JA, Robevska G, Ohnesorg T, Hewitt J, et al: Disorders of sex development: insights from targeted gene sequencing of a large international patient cohort. Genome Biol 2016, 17:243
12. Wang X, Zein WM, D'Souza L, Roberson C, Wetherby K, He H, Villarta A, Turriff A, Johnson KR, Fann YC: Applying next generation sequencing with microdroplet PCR to determine the disease-causing mutations in retinal dystrophies. BMC Ophthalmol 2017, 17:157
13. Omoyinmi E, Standing A, Keylock A, Price-Kuehne F, Melo Gomes S, Rowczenio D, Nanthapisal S, Cullup T, Nyanhete R, Ashton E, Murphy C, Clarke M, Ahlfors H, Jenkins L, Gilmour K, Eleftheriou D, Lachmann HJ, Hawkins PN, Klein N, Brogan PA: Clinical impact of a targeted next-generation sequencing gene panel for autoinflammation and vasculitis. PLoS One 2017, 12:1−20
14. Iqbal Z, Rydning SL, Wedding IM, Koht J, Pihlstrøm L, Rengmark AH, Henriksen SP, Tallaksen CME, Toft M: Targeted high throughput sequencing in hereditary ataxia and spastic paraplegia. PLoS One 2017, 12:1−19
15. Celestino-Soper PBS, Gao H, Lynnes TC, Lin H, Liu Y, Spoonamore KG, Chen P-S, Vatta M: Validation and utilization of a clinical next-generation sequencing panel for selected cardiovascular disorders. Front Cardiovasc Med 2017, 4:11
16. Lucarelli M, Porcaro L, Biffignandi A, Costantino L, Giannone V, Alberti L, Bruno SM, Corbetta C, Torresani E, Colombo C, Seia M: A new targeted CFTR mutation panel based on next-generation sequencing technology. J Mol Diagn 2017, 19:788−800
17. LaDuca H, Pesaran T, Elliott AM, Speare V, Dolinsky JS, Gau CL, Chao E: Utilization of multigene panels in hereditary cancer predisposition testing. Next Generation Sequencing in Cancer Research, Volume 2: From Basepairs to Bedsides. Edited by Wu W, Choudhry H. New York, NY: Springer, 2015. pp. 459−482
18. Roy S, Coldren C, Karunamurthy A, Kip NS, Klee EW, Lincoln SE, Leon A, Pullambhatla M, Temple-Smolkin RL, Voelkerding KV, Wang C, Carter AB: Standards and guidelines for validating next-generation sequencing bioinformatics pipelines. J Mol Diagn 2017, 20:4−27
19. Quinlan AR: BEDTools: the Swiss-Army tool for genome feature analysis. Curr Protoc Bioinformatics 2014, 47:11.12.1−11.12.34

20. Jennings LJ, Arcila ME, Corless C, Kamel-Reid S, Lubin IM, Pfeifer J, Temple-Smolkin RL, Voelkerding KV, Nikiforova MN: Guidelines for validation of next-generation sequencing–based oncology panels. J Mol Diagn 2017, 19:341–365
21. 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
22. Mu W, Lu H, Chen J, Li S, Elliott AM: Sanger confirmation is required to achieve optimal sensitivity and specificity in next-generation sequencing panel testing. J Mol Diagn 2016, 18:923–932
23. Lincoln SE, Zook JM, Chowdhury S, Mahamdallie S, Fellowes A, Klee EW, Truty R, Huang C, Tomson FL, Cleveland MH, Vallone PM, Ding Y, Seal S, Desilva W, Garlick RK, Salit M, Rahman N, Lincoln SE: An interlaboratory study of complex variant detection. BioRxiv 2017, doi:10.1101/218529