Proficiency Testing of Standardized Samples Shows High Interlaboratory Agreement for Clinical Next-Generation Sequencing–Based Hematologic Malignancy Assays With Survey Material–Specific Differences in Variant Frequencies

Alissa Keegan, MD, PhD; Julia A. Bridge, MD; Neal J. Lindeman, MD; Thomas A. Long, MPH; Jason D. Merker, MD, PhD; Joel T. Moncur, MD, PhD; Nathan D. Montgomery, MD, PhD; Rakesh Nagarajan, MD, PhD; Paul G. Rothberg, PhD; Mark J. Routbort, MD, PhD; Patricia Vasalos, BS; Rena Xian, MD, PhD; Annette S. Kim, MD, PhD

Context.—As laboratories increasingly turn from single-analyte testing in hematologic malignancies to next-generation sequencing–based panel testing, there is a corresponding need for proficiency testing to ensure adequate performance of these next-generation sequencing assays for optimal patient care.

Objective.—To report the performance of laboratories on proficiency testing from the first 4 College of American Pathologists Next-Generation Sequencing Hematologic Malignancy surveys.

Design.—College of American Pathologists proficiency testing results for 36 different engineered variants and/or allele fractions as well as a sample with no pathogenic variants were analyzed for accuracy and associated assay performance characteristics.

Results.—The overall sensitivity observed for all variants was 93.5% (2190 of 2341) with 99.8% specificity (22 800 of 22 840). The false-negative rate was 6.5% (151 of 2341), and the largest single cause of these errors was difficulty in identifying variants in the sequence of CEBPA that is rich in cytosines and guanines. False-positive results (0.18%; 40 of 22 840) were most likely the result of preanalytic or postanalytic errors. Interestingly, the variant allele fractions were almost uniformly lower than the engineered fraction (as measured by digital polymerase chain reaction). Extensive troubleshooting identified a multifactorial cause for the low variant allele fractions, a result of an interaction between the linearized nature of the plasmid and the Illumina TruSeq chemistry.

Conclusions.—Laboratories demonstrated an overall accuracy of 99.2% (24 990 of 25 181) with 99.8% specificity and 93.5% sensitivity when examining 36 clinically relevant somatic single-nucleotide variants with a variant allele fraction of 10% or greater. The data also highlight an issue with artificial linearized plasmids as survey material for next-generation sequencing.

(T Arch Pathol Lab Med. 2020;144:959–966; doi: 10.5858/arpa.2019-0352-CP)
genes, with no gene or no specific mutation pathognomonic for a specific disease entity, with rare exceptions. However, the patterns may suggest a diagnosis, have significant implications for prognosis, and/or direct therapeutic strategies. Accordingly, NGS panel testing is rapidly becoming an integral part of the workup of hematologic neoplasms.

With this shift from single-analyte testing to panel testing for hematologic neoplasms in molecular diagnostic laboratories, new methods of ensuring the accuracy and reproducibility of this NGS testing are required.3,5,6 Accordingly, the College of American Pathologists Molecular Oncology Committee developed a next-generation sequencing hematologic malignancy proficiency testing survey. The purpose of this manuscript is to summarize the performance of laboratories on this survey.

METHODS

Data were derived from the first 2 years of mailings (4 mailings: A and B each year) of the College of American Pathologists Next-Generation Sequencing Hematologic Malignancy Survey (2016A through 2017B). Laboratories were concurrently sent 3 independent specimens containing linearized plasmids with recurrent somatic variants mixed with genomic DNA derived from the GM24385 cell line to achieve variant allele fractions (VAFs) ranging from 11.8% to 48.4%. Specimens were generated by a commercial reference material vendor under Good Manufacturing Practice.7 The synthetic DNA inserts contain a somatic variant with approximately 500 base pairs (bp) of flanking genomic sequence on each side of the variant. The flanking genomic DNA sequence was matched to the diluent genomic DNA (ie, GM24385). The VAFs were orthogonally confirmed by the insert vendor using digital polymerase chain reaction (dPCR) (Table). The VAFs were calculated as mutant copies divided by the sum of mutant copies and wild-type copies for each variant in specimens 1, 2, and 3.

Laboratories were instructed to perform NGS using the methodology routinely performed on clinical samples in their laboratory for the detection of somatic single-nucleotide variants, small insertions, and small deletions commonly found in hematologic malignancies. This could include targeted gene panels or whole-exome or whole-genome sequencing. As is standard for College of American Pathologists surveys, laboratories were instructed that they could confirm the variant by a secondary methodology. However, confirmation had to follow the laboratory’s standard procedure for testing clinical specimens and could not be referred to another laboratory. Each mailing contained 3 samples in which were represented anywhere from 0 to 6 different positive variants. Following testing, laboratories reported the variants detected in each specimen by selecting from a master variant list containing 17 genes (ASXL1, BRAF, CALR, CEBPA, DNMT3A, FLT3, IDH1, IDH2, JAK2, KIT, MPL, MYD88, NOTCH1, NPM1, SF3B1, TET2, and TP53) for which participants were given 6 different possible variant options. Because not all laboratories test for all genes or all positions within a gene, participants who did not test for a given locus were excluded from the analysis. In addition, laboratories provided coverage depth and variant allele fraction for each reported variant as well as the laboratory’s assay characteristics. The term VAF standard deviation index (VAF SDI) is used to define the magnitude and direction of the deviation from engineered VAF for each variant: [dPCR VAF – (mean participant VAF)]/standard deviation of all participant VAFs. For the investigation of the VAF underestimation, the GC content (the percentage of nucleotides in the strand that possess either cytosine or guanine bases) for each variant was calculated for ±150 bp of flanking sequence from corresponding gDNA.

All figures were generated using Prism (GraphPad Software) and Excel (Microsoft, Inc.). Statistical comparisons of VAFs by various NGS laboratory practices were tested by using analysis of variance for each positive variant from each mailing. A 2-pass 3-SD outlier screen was used to remove outliers prior to testing.

Because the sequence coverages were not Gaussian distributed, nonparametric Kruskal-Wallis tests were used to compare differences in sequence coverage distributions across different categories of NGS laboratory practices.8,9 A Bonferroni correction factor was applied for multiple-test comparison across the positive variants and mailings.10,11 Defined as the type 1 error (0.05) divided by the number of positive variants (36); therefore, the threshold for statistical significance was set to P = .001. On a per-variant basis, various parameters were compared using a paired t test performed by Prism. Linear regression was performed for pre-post–shipping comparisons using Prism.

RESULTS

During the period of this study (4 mailings), participation in the survey increased from 57 to 88 laboratories. Each of 18 positive variants was assessed twice across the 4 mailings, for a total of 2190 true-positive variant results from the participants across all the mailings included in this study. Of note, using data from the 2017A survey, the majority of laboratories, 78.8% (63 of 80), used some form of target enrichment prior to sequencing, with more laboratories using amplicon-based (55.7%; 45 of 80) compared with hybrid capture–based (22.5%; 18 of 80) methods. The remainder of the responses were not accompanied by sufficient methodologic detail to allow their categorization as an amplicon or hybrid capture method. The percentage of laboratories that correctly identified each variant ranged from 65.7% for the CEBPA p.H24fs variant (46 of 70) to 100% for the KIT p.D816V variant (56 of 56) (Table). The overall analytical sensitivity for all variant responses was 93.5% (2190 of 2341). The analytical specificity of the results was 99.8% (22 800 of 22 840). When true positives and true negatives are considered together, the overall accuracy was 99.2% (24 990 of 25 181).

The mean coverage (variably defined by laboratories as either total or unique reads mapping to a base) reported at the variant sites ranged from 1493 to 12 367 (Table) and did not appear to correlate with accuracy of detection, given the variable coverage for variants with a range of detection rates such as KIT p.D816V (196–17 890; 100% accuracy) and CEBPA p.H24fs (55–30 264; 65.7% accuracy).

There were 151 false-negative results (6.5%; 151 of 2341) reported across all 18 variants (Table; Figure 1). These false negatives did not appear to result from VAFs falling below the limit of detection of the assay, as there was no correlation between the percentage of false negatives and the engineered VAF (Supplemental Figure 1, A; see supplemental digital content, containing 6 figures and 1 table at www.archivesofpathology.org in the August 2020 table of contents). The percentage of false negatives also did not correlate with the presence or size of insertions or deletions, although the maximum size of a variant was only 6 nucleotides (Supplemental Figure 1, B). There was some correlation of the false-negative rate with GC content derived from 300 bp of gDNA sequence centered on the variant (Supplemental Figure 1, C). Using 2017B survey data, there was no observable association of false negatives with any particular sequencing platform or library preparation method.

Only 40 false-positive variants (0.18%; 40 of 22 840) were identified across all 4 surveys. Exactly 50% of these erroneous responses (20 of 40) resulted from the laboratory selecting the variant in the same gene that either was located adjacent to the correct response or had very similar nomenclature. Another 22.5% (9 of 40) of the errors were
Figure 1. Pie chart of true-positive (TP), false-positive (FP), and false-negative (FN) results with an analysis of the etiology of the FP results. True negatives were excluded because of their large number (n = 22 800) in order to emphasize the FP and FN results.

Figure 2. Variant allele fractions (VAFs) across all 4 mailings comparing amplicon-based (red) with hybrid capture (green) enrichment methodology. Red box indicates digital polymerase chain reaction–confirmed VAF. Abbreviations: *, P < .001 by 2-tailed t test; ns, did not pass significance threshold.

Figure 3. Variant allele fractions (VAFs) across all 4 mailings comparing Illumina TruSeq chemistry (blue, TruSight myeloid or custom TruSeq) with all others (orange) reporting a different enrichment methodology. Results where no method was listed are excluded. Red box indicates digital polymerase chain reaction–confirmed VAF. Abbreviations: *, P < .001 by 2-tailed t test; ns, did not pass significance threshold.
due to sample swaps (either submitting the results for one sample twice or swapping samples between the 2 years—submitting data for 2016A-01 instead of 2017A-01, etc). To determine the accuracy of the VAFs reported by participants, the mean participant VAFs per variant were compared with the engineered VAF determined by dPCR and divided by the standard deviation of the participant VAFs to provide a quantified VAF SDI (as described in the Methods section) (Table). The average of the VAFs reported by the laboratories differed from the engineered dPCR VAF by as much as 13.31% (1.66 VAF SDI for NOTCH1 p.L1600P, survey 2016B) (Table). The highest SDI was 3.03 for a low–BRAF p.V600E variant engineered to be 16.2% VAF, but with a mean participant VAF of only 8.46%. The mean reported VAF across all the samples underestimated the engineered VAF in 33 of 36 variants, for an overall average 0.68 VAF SDI. The VAF underestimation tended to be more pronounced with amplicon-based compared with hybrid capture–based target enrichment methods and reached statistical significance in 6 of the 36 variants (Figure 2; \( P < .001 \) for each of the other cases.

### Summary of Proficiency Testing Results for the Next-Generation Sequencing Hematologic Malignancy Survey

| Gene   | Transcript | Nucleotide Change | Protein Change | Genomic Description (hg19) | Survey |
|--------|------------|-------------------|----------------|---------------------------|--------|
| BRAF   | NM_004333.4| c.1799T>A         | p.V600E        | chr7:140453136A>T         | 2016B  |
|        |            |                   |                |                           | 2017B  |
| CALR   | NM_004343.3| c.1154_1155insTGTGTGTC | p.K385fs      | chr19:13054627delAinsCTTGTC | 2016B  |
|        |            |                   |                |                           | 2017A  |
| CEBPA  | NM_004364.4| c.68dupC          | p.H24fs        | chr19:33793253dupG        | 2016A  |
|        |            |                   |                |                           | 2017B  |
| DNMT3A | NM_002552.4| c.2645G>A         | p.R882H        | chr2:25457242C>T          | 2016A  |
|        |            |                   |                |                           | 2017B  |
| FLT3   | NM_004119.2| c.2503G>T         | p.D835Y        | chr13:28596242C>A         | 2016B  |
|        |            |                   |                |                           | 2017A  |
| IDH1   | NM_005896.2| c.395G>A          | p.R132H        | chr2:209113112C>T         | 2016B  |
|        |            |                   |                |                           | 2017A  |
| IDH2   | NM_002168.2| c.419G>A          | p.R140Q        | chr15:90631934C>T         | 2016A  |
|        |            |                   |                |                           | 2017A  |
| JAK2   | NM_004972.3| c.1624_1629delAATGAA | p.N542_E543del | chr9:5070035_5070040delAATGAA | 2016B  |
|        |            |                   |                |                           | 2017B  |
| KIT    | NM_000222.2| c.2447A>T         | p.V617F        | chr9:5073770C>T           | 2016A  |
|        |            |                   |                |                           | 2017B  |
| MPL    | NM_005373.2| c.1544G>T         | p.W515L        | chr1:43815009G>T          | 2016B  |
|        |            |                   |                |                           | 2017B  |
| MYD88  | NM_002468.4| c.794T>C          | p.L265P        | chr3:38182641T>C          | 2016B  |
|        |            |                   |                |                           | 2017B  |
| NOTCH1 | NM_017617.3| c.4799T>C         | p.L1600P       | chr9:139399344A>G         | 2016B  |
|        |            |                   |                |                           | 2017B  |
| NPM1   | NM_002520.6| c.860_863dupTCTGT | p.W288fs       | chr5:170837544_170837547dupTCTGT | 2016A  |
|        |            |                   |                |                           | 2017A  |
| SF3B1  | NM_012433.2| c.2098A>G         | p.K700E        | chr2:198266834T>C         | 2016B  |
|        |            |                   |                |                           | 2017B  |
| TET2   | NM_001127208.2| c.1648C>T       | p.R550*        | chr4:106156747C>T         | 2016A  |
|        |            |                   |                |                           | 2017B  |
| TP53   | NM_000546.5| c.743G>A          | p.R248Q        | chr17:7577538C>T          | 2016A  |
|        |            |                   |                |                           | 2017A  |

No variant^a^  

Abbreviations: dPCR, digital polymerase chain reaction; NA, not applicable; VAF, variant allele fraction; VAF SDI, VAF standard deviation index = \([(\text{dPCR VAF}) - (\text{mean VAF})]/(\text{VAF SD})\); ΔVAF, difference between the dPCR VAF and the mean participant VAF.  

^a^ This sample contained no pathogenic variants. Therefore, the number and percentage of correct responses are true negatives instead of true positives as in all the other cases, and the number and percentage of incorrect responses are false positives instead of false negatives as in all the other cases.
the 6 variants). Among amplicon-based enrichment methods, the panels using TruSeq chemistry (TruSight myeloid or custom TruSeq, Illumina, Inc), in use by 62.2% (28 of 45) of laboratories based upon the 2017A proficiency testing survey data, resulted in VAFs that were lower than those for other library preparation methods in 32 of 36 variants across all surveys, whereas all other methods showed VAFs closer to the engineered fraction (Figure 3).

Illumina TruSeq panels led to higher depth of sequence coverage, passing the significance threshold in 23 of 36 variants (Figure 4). Of laboratories using TruSeq library preparation methods, some participants used a commercial kit, the Illumina TruSight Myeloid Panel, whereas other used a TruSeq custom amplicon panel. There was no significant difference in VAFs produced by the commercial compared with the custom-designed panels, both producing similarly low VAFs (Figure 5). When we examined VAF versus coverage across all methods and respondents, VAF underestimation, as measured by VAF SDI on a per-variant basis, was seen across a wide range of coverage (Figure 6).

Examining survey 2017A data, the VAFs were comparable across all ranges of read length, sequencing platforms, and analysis software (Supplemental Figures 2 through 4). This was true despite the tendency of Illumina TruSeq chemistry users to analyze with MiSeq reporter (Illumina) software (12 of 14 respondents; 85.7%); 14.3% (2 of 14) used NextGENe software (SoftGenetics, LLC) and the remainder of TruSeq chemistry laboratories did not provide details of their analysis software. The VAF SDI did not appear to associate with GC content in gDNA sequence surrounding

| No. of Laboratories That Detected Variant (% True-Positive) | No. of Laboratories That Did Not Detect Correct Variant (% False-Negative) | Engineered VAF as Provided by Vendor Using dPCR | Mean VAF, % | VAF SD, % | Δ VAF | VAF SDI | Mean Coverage | Range Coverage |
|-----------------------------------------------------------|-------------------------------------------------|-----------------------------------------------|-------------|-----------|-------|---------|---------------|---------------|
| 50 | 46 (92.0) | 4 (8.0) | 16.2 | 8.46 | 2.55 | 7.74 | 3.03 | 6592.07 | 264–48 958 |
| 81 | 79 (97.5) | 2 (2.5) | 27 | 21.68 | 7.93 | 5.32 | 0.67 | 9490.49 | 87–41 346 |
| 43 | 37 (86.0) | 6 (14.0) | 45.2 | 36.16 | 8.78 | 9.04 | 1.03 | 5485.03 | 57–63 078 |
| 64 | 61 (95.3) | 3 (4.7) | 35.2 | 27.60 | 9.70 | 7.60 | 0.78 | 5447.13 | 142–62 373 |
| 42 | 28 (66.7) | 14 (33.3) | 48.4 | 36.31 | 14.01 | 12.09 | 0.86 | 1493.14 | 15–796 |
| 70 | 46 (65.5) | 24 (34.3) | 29 | 24.07 | 9.37 | 4.93 | 0.53 | 2639.57 | 55–30 264 |
| 51 | 49 (96.1) | 2 (3.9) | 21.2 | 18.71 | 5.76 | 2.49 | 0.43 | 4970.22 | 85–23 511 |
| 85 | 81 (95.3) | 14 (4.7) | 14 | 13.57 | 8.30 | 0.43 | 0.05 | 5205.23 | 166–37 616 |
| 54 | 52 (96.3) | 2 (3.7) | 25.9 | 23.69 | 4.29 | 2.21 | 0.52 | 8611.54 | 164–34 693 |
| 77 | 74 (96.1) | 3 (3.9) | 17.2 | 15.09 | 3.37 | 2.11 | 0.62 | 7381.03 | 129–56 637 |
| 56 | 55 (98.2) | 1 (1.8) | 23.4 | 19.07 | 6.60 | 4.33 | 0.66 | 4919.96 | 459–19 291 |
| 79 | 78 (98.7) | 1 (1.3) | 17.7 | 12.96 | 5.10 | 0.43 | 0.05 | 779.84 | 33–59 608 |
| 55 | 54 (98.2) | 1 (1.8) | 15.0 | 16.23 | 3.56 | −1.23 | −0.34 | 7036.92 | 113–3 563 |
| 86 | 85 (98.8) | 1 (1.2) | 37 | 37.76 | 5.29 | −0.76 | −0.14 | 9229.28 | 143–55 232 |
| 56 | 54 (96.4) | 2 (3.6) | 37.2 | 26.37 | 10.07 | 7.63 | 0.96 | 5551.17 | 68–11 166 |
| 53 | 46 (86.8) | 7 (13.2) | 17 | 13.19 | 5.13 | 3.81 | 1.08 | 4019.30 | 31–28 064 |

Arch Pathol Lab Med—Vol 144, August 2020

PT for NGS Hematologic Malignancy Assays—Keegan et al 963
the variant (±150 nucleotides of the variant; Supplemental Figure 5).

For further investigation of the source of the VAF underestimation, 3 laboratories using TruSight panels returned leftover survey material to the vendor for dPCR requantification. Variant allele fractions were highly correlated before and after shipping, thus excluding sample degradation as a potential source of error (Supplemental Figure 6). Variant allele fractions were also consistent when the FASTQ files from 2 laboratories that reported VAF underestimation were reanalyzed with the most up-to-date Illumina software (Supplemental Table), excluding antiquated software as the cause of the issue.

The 2016–2017 NGS survey material was derived from plasmids bearing 1000 bp of insert sequence centered on the variant of interest, with digital PCR marker sequences and unique restriction enzyme sites on either end of the insert (Figure 7, A). The plasmid was linearized with a single restriction enzyme digestion and mixed with diluent genomic DNA, and the engineered VAF level was quantified with dPCR. To explore whether the dPCR marker sequences and/or the plasmid backbone was the source of interference, the vendor prepared isolated inserts free from plasmid backbone with or without the dPCR marker sequences for 5 variants. The variants included one variant, NRAS p.Q61K, from the College of American Pathologists NGS solid tumor survey for comparison. These inserts were diluted in genomic DNA, akin to the original survey material, and analyzed by 4 laboratories. Again, VAF underestimation was present, with greater underestimation for most inserts bearing dPCR marker sequences (Figure 7, B). Interestingly, VAFs of single-nucleotide polymorphisms in the diluent

Figure 4. Coverage depth across all 4 mailings comparing Illumina TruSeq chemistry (blue, TruSight myeloid or custom TruSeq) with all others (orange) reporting a different enrichment methodology. Abbreviations: *, P < .001 by 2-tailed t test; ns, did not pass significance threshold.

Figure 5. Variant allele fraction (VAF) depth across all 4 mailings comparing Illumina TruSight (blue) and Illumina TruSeq Custom (red) enrichment methods. Red box indicates digital polymerase chain reaction–confirmed VAF. Abbreviation: ns, did not pass significance threshold of P < .001 by 2-tailed t test.
genomic DNA from the GM24385 cell line consistently approached expected germline allele frequencies of 50% or 100% in all 3 laboratories (Figure 7, A through C). Because these endogenous variants are seemingly not affected, we hypothesized that the source of the underestimation is the interaction of the engineered material and the enrichment method, with some additional contribution of the marker sequences.

DISCUSSION

One hundred laboratories reported proficiency testing results from their NGS-based assay for the identification of somatic variants in hematologic malignancies. The laboratories were provided with 3 engineered specimens containing a total of 18 unique recurring somatic variants with a VAF between 11.8% and 48.4%. The overall accuracy observed for all variants for laboratories submitting at least 5 positive and 5 negative results per survey was very high (99.2%), with 93.5% sensitivity and 99.8% specificity. Although the accuracy for each individual variant ranged from 65.7% to 100%, CEBPA accounted for the only results with less than 85% accuracy, despite being at high

---

**Figure 6.** Relationship of variant allele fraction (VAF) to depth of coverage across all variants in the 4 mailings. Coverage (log scale) is plotted against the VAF standard deviation index (SDI) for the respective variant (the number of standard deviations from the mean VAF per variant).

| Targeted Variants | dPCR | TruSight | I-PCR | TruSight | HM-PCR | HM-PCR | I-PCR | HM-PCR | I-PCR | HM-PCR | I-PCR | HM-PCR | I-PCR | HM-PCR | I-PCR | HM-PCR |
|-------------------|------|----------|-------|----------|--------|--------|-------|--------|-------|--------|-------|--------|-------|--------|-------|--------|
| NRAS p.G61K       | 35%* | ND       | 33%   | 26%      | 21%    | 10%    | 29%   | 11.00% | 13%    | 3.50%  | 17.8% | 13.24% | 35.50% | 18%    |
| IDH1 p.R132H      | 23.40%| 11.50%  | 26%   | 18%      | 2%     | 1%     | 5%    | 2.00%  | 5.40%  | 1.90%  | 5.73% | 25.10% | 12%    |
| KIT p.D816V       | 42.70%| 30.10%  | 41%   | 37%      | 5%     | 5%     | 9%    | 8.00%  | 8.60%  | 4.50%  | 10.34% | 5.80%  | 27.90% | 17.40% |
| KIT p.N822K       | 47.90%| 29.50%  | 71%   | 35%      | 4%     | 1%     | 9%    | 5.00%  | 5.10%  | 4.30%  | 11.17% | 27.60% | 17.60% |
| MPL p.W515L       | 45.80%| 26.50%  | 51%   | 59%      | 8%     | 7%     | 18%   | 26.00% | 15%    | 9.00%  | 9.31% | 7.22%  | 47.80% | 53.10% |

**Figure 7.** Results of investigation into the variant allele fraction (VAF) underestimation by laboratories using TruSeq chemistry. A, Table of the 5 engineered variants, the original survey engineered VAF of the linearized plasmid (* indicates that this VAF was from the next-generation sequencing solid tumor survey), the Next-Generation Sequencing Hematologic Malignancy Survey VAF average, the digital polymerase chain reaction (dPCR) results of the isolated insert (I) and the isolated insert with retained marker regions (I+M), and the results from 4 different laboratories (including a single-nucleotide polymorphism [SNP] covered by laboratory D (LabD) and 2 different platforms used by LabD: TruSight and Ion Torrent). B, VAF distribution of known SNPs covered by the panel used by laboratory B (LabB). C, VAF distribution of known SNPs (dbSNP database nomenclature) covered by the panel used by laboratory C (LabC). Abbreviations: bp, base pair; CAP, College of American Pathologists; ID, identifier; LabA, laboratory A; ND, not determined.
engineered VAFs (29% on the 2017B survey and 48.4% on the 2016A survey). More than 80% of variants (29 of 36) achieved greater than 90% accuracy. These highly accurate results are consistent with those previously observed for the equivalent survey in solid tumors and support the reliability of clinical NGS-based oncology testing, indicating very high interlaboratory agreement for the detection of somatic single-nucleotide variants and small indels.1,6

Only 40 potential false-positive results were reported; 50% (20) of these were the likely result of an erroneous manual entry process, and an additional 22.5% (9) were the result of sample swaps. Although preanalytic and post-analytic errors may have significant consequences for patient care, when these errors are excluded, the true analytic false-positive rate is as low as 0.04% (11 of 25 181 total responses).

The largest fraction of the false negatives (25.2%; 38 of 151) observed in these surveys were for the detection of the CEBPA c.68dupC (p.H24fs) variant, despite being at high engineered VAFs (29% on the 2017B survey, 48.4% on the 2016A survey). The lower rate of survey respondents testing and reporting for this CEBPA variant likely reflects respondent awareness of the challenges of reliably detecting these mutations in NGS panels. Biallelic CEBPA mutations are a positive prognostic factor in acute myeloid leukemia and represent a distinct category in the World Health Organization classification.12 With an extreme GC content of 75%, CEBPA is afflicted by suboptimal capture and amplification efficiency, leading to low read depths. There is also difficulty in mapping CEBPA variants owing to repeat sequences and because most significant CEBPA mutations are insertion/deletions. These technical challenges are reflected in the survey results with lower mean coverages and higher VAF standard deviation.

In addition, we identified a specific artifact of the engineered constructs used as proficiency materials that precluded an accurate assessment of VAF quantitation for laboratories using the Illumina TruSeq chemistry. This phenomenon was not observed in the corresponding solid tumor NGS survey, possibly because of the increased use of hybrid capture and fragmentation in these typically large panels.1 The investigation of this artifact hinged upon the identification of typical VAFs associated with polymorphisms covered by the panels, indicating that this would not have ramifications in normal clinical practice. Accordingly, the change to isogenic cell lines instead of plasmids has mitigated this survey-specific issue that likely has no bearing on clinical estimates of VAF (College of American Pathologists, unpublished data; participant summary report released May 8, 2019).

In summary, the overall analytic performance of NGS for detecting variants associated with myeloid and lymphoid disease is excellent. The few false-negative results point to the need for closer attention and perhaps more stringent reporting criteria for the CEBPA gene in general hem NGS panels. On the other hand, the rare false-positive results point toward the need for careful proofreading of result entry and careful sample designation. Finally, these survey results revealed an unexpected interaction of the artificial survey material and a particular amplification chemistry. These latter findings illustrate the challenges associated with the development of proficiency testing and quality control materials.

We thank Mark Christian, BS, and Thermo Fisher Scientific for contributions in variant allele frequency troubleshooting. We thank Emily Chen, PhD, and Illumina for contributions in variant allele frequency troubleshooting and reanalysis of FASTQ files.

References
1. Merker JD, Devereaux K, Lafort AJ, et al. Proficiency testing of standardized samples shows very high interlaboratory agreement for clinical next-generation sequencing–based oncology assays. Arch Pathol Lab Med. 2019;143(4):463–471. doi:10.5858/arpa.2018-0336-CP
2. Nagarajan R, Bartley AN, Bridge JA, et al. A window into clinical next-generation sequencing–based oncology testing practices. Arch Pathol Lab Med. 2017;141(12):1679–1685.
3. Bombard Y, Bach PB, Offit K. Translating genomics in cancer care. J Natl Compr Canc Netw. 2013;11(11):1343–1355. doi:10.6004/jncn.2013.0158
4. McClure RF, Ewalt MD, Crow J, et al. Clinical significance of DNA variants in chronic myeloid neoplasms: a report of the Association for Molecular Pathology. J Mol Diagn. 2018;20(6):717–737.
5. Moncur JT, Bartley AN, Bridge JA, et al. Performance comparison of different analytic methods in proficiency testing for mutations in the BRAF, EGFR, and KRAS genes: a study of the College of American Pathologists Molecular Oncology Committee. Arch Pathol Lab Med. 2019;143(10):1203–1211. doi:10.5858/arpa.2018-0396-CP
6. Surrey LF, Oakley FD, Merker JD, et al. Next-generation sequencing (NGS) methods show superior or equivalent performance to non-NGS Methods on BRAF, EGFR, and KRAS proficiency testing samples. Arch Pathol Lab Med. 2019;143(8):980–984. doi:10.5858/arpa.2018-0394-CP
7. US Food and Drug Administration. Current Good Manufacturing Practices (CGMPs). December 27, 2017. Accessed February 2, 2018.
8. Conover WJ. Practical Nonparametric Statistics. 3rd ed. New York, NY: John Wiley & Sons Inc; 1999.
9. Kruskal WH, Wallis WW. Use of ranks in one-criterion variance analysis. J Am Stat Assoc. 1952;47(260):583–621.
10. Bonferroni CE. Teoria statistica delle classi e calcolo delle probabilit`a [Statistical class theory and probability calculus]. Pubbl R Ist Super Sci Economiche Commerciali Firenze. 1936;8:1–62.
11. Dunn OJ. Multiple comparisons among means. J Am Stat Assoc. 1961; 56(293):52–64.
12. Arber DA, Bruning RD, Le Beau MM, et al. Acute myeloid leukemia with recurrent genetic abnormalities. In: Swerdlow SH, Campo E, Harris NL, et al, eds. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. 4th ed. Geneva, Switzerland: World Health Organization; 2017:130–149.