NEW ARTICLE

This article was posted on the *Archives* Web site as a New Article. New Articles have been peer reviewed, copyedited, and reviewed by the authors. Additional changes or corrections may appear in these articles when they appear in a future print issue of the *Archives*. New Articles are citable by using the Digital Object Identifier (DOI), a unique number given to every article. The DOI will typically appear at the end of the abstract.

The print version of this manuscript will replace the New Article version at the above DOI once it is available.

SPOT/Dx Pilot Reanalysis and College of American Pathologists Proficiency Testing for *KRAS* and *NRAS* Demonstrate Excellent Laboratory Performance

Ahmet Zehir, PhD; Valentina Nardi, MD; Eric Q. Konnick, MD, MS; Christina M. Lockwood, PhD; Thomas A. Long, MPH; Nikoletta Sidiropoulos, MD; Rhona J. Souers, MS; Patricia Vasalos, BS; Neal I. Lindeman, MD; Joel T. Moncur, MD, PhD, MS

● **Context.**—The Sustainable Predictive Oncology Therapeutics and Diagnostics quality assurance pilot study (SPOT/Dx pilot) on molecular oncology next-generation sequencing (NGS) reportedly demonstrated performance limitations of NGS laboratory-developed tests, including discrepancies with a US Food and Drug Administration–approved companion diagnostic. The SPOT/Dx pilot methods differ from those used in proficiency testing (PT) programs.

Objective.—To reanalyze SPOT/Dx pilot data using PT program methods and compare to PT program data.

Design.—The College of American Pathologists (CAP) Molecular Oncology Committee reanalyzed SPOT/Dx pilot data applying PT program methods, adjusting for confounding conditions, and compared them to CAP NGS PT program performance (2019–2022).

Results.—Overall detection rates of *KRAS* and *NRAS* single-nucleotide variants (SNVs) and multinucleotide variants (MNVs) by SPOT/Dx pilot laboratories were 96.8% (716 of 740) and 81.1% (129 of 159), respectively. In CAP PT programs, the overall detection rates for the same SNVs and MNVs were 97.2% (2671 of 2748) and 91.8% (1853 of 2019), respectively. In 2022, the overall detection rate for 5 *KRAS* and *NRAS* MNVs in CAP PT programs was 97.3% (1161 of 1193).

Conclusions.—CAP PT program data demonstrate that laboratories consistently have high detection rates for *KRAS* and *NRAS* variants. The SPOT/Dx pilot has multiple design and analytic differences with established PT programs. Reanalyzed pilot data that adjust for confounding conditions demonstrate that laboratories proficiently detect SNVs and less successfully detect rare to never-observed MNVs. The SPOT/Dx pilot results are not generalizable to all molecular oncology testing and should not be used to market products or change policy affecting all molecular oncology testing.

(Arch Pathol Lab Med. doi: 10.5858/arpa.2023-0322-CP)

P roficiency testing (PT) constitutes a critical component of clinical diagnostics, ensuring laboratories can identify the biomarkers that their validated assays are designed to detect. It is also an essential part of quality management systems designed to support continual improvements in laboratory performance. For these purposes, the College of American Pathologists (CAP) Molecular Oncology Committee (Mol Onc) has been designing and administering PT programs for next-generation sequencing (NGS) assays since 2015. These PT programs include genetically engineered cell lines and in silico mutagenized sequencing data files designed to contain a variety of mutations in multiple genes at a range of variant allele fractions (VAFs).

In 2016, a diagnostic quality assurance pilot study was initiated by the Sustainable Predictive Oncology Therapeutics and Diagnostics (SPOT/Dx) working group to investigate the standardization of precision medicine laboratory testing. The pilot focused on testing of *RAS* mutations in colorectal cancer, which is recommended by the National Comprehensive Cancer Network (NCCN) for advanced or metastatic disease to determine eligibility for anti-EGFR therapy (NCCN Clinical Practice Guidelines in Oncology, version 2.2023, April 25, 2023). The pilot was organized by Tapestry Networks, a

private professional services company, and was primarily supported by Amgen. The working group participants included clinical providers, policy experts, regulators, payers, and patient advocates.[1] They sought to evaluate the feasibility of using genetically engineered cell lines and in silico mutagenized files, generated by methods already in use for CAP PT programs, as adjunctive materials to improve the standardization of NGS assays performed by molecular diagnostic laboratories using laboratory-developed tests (LDTs). They also attempted to compare the performance of NGS LDTs to a US Food and Drug Administration (FDA)–approved companion diagnostic (CDx) assay, Praxis Extended RAS Panel (Illumina, San Diego, California), designed to identify 56 variants in *KRAS* and *NRAS* for targeted colorectal cancer therapy determination.

In 2021, the results from the SPOT/Dx pilot were published.[2] Although the published article reaffirmed the feasibility of standardized sample preparation methods already used by the CAP Mol Onc to evaluate the laboratory performance of NGS assays, the article reported performance limitations of molecular oncology LDTs in general, including discrepancies between the results obtained with the NGS LDTs and those reported by the manufacturer of the FDA-approved CDx assay. The CAP Mol Onc identified multiple differences in the SPOT/Dx pilot design and analysis, compared with PT programs, that affected the results and their generalizability. Here, we present our reanalysis of the SPOT/Dx pilot data applying methods modeled after PT programs, highlight design and methodologic issues that may have confounded the pilot study, and compare SPOT/Dx pilot results to laboratory performance for the same variants as assessed through the CAP PT programs that survey hundreds of clinical laboratories.

MATERIALS AND METHODS

In the SPOT/Dx pilot, human cell lines ("wet" samples) with engineered single-nucleotide variants (SNVs) and multinucleotide variants (MNVs) in *KRAS* and *NRAS* were used to assess end-to-end NGS test performance, and electronic sequence data files were engineered with mutations ("dry" samples) used to evaluate the accuracy of the bioinformatic algorithms ("pipelines"), as previously reported.[2,3] CAP representatives participated in the Scientific Technical Working Group and provided logistical and statistical support for the SPOT/Dx pilot. CAP representatives were not involved with final decisions about pilot VAFs, analyses, and reporting of results.

The variants included in the SPOT/Dx pilot are listed in Supplemental Tables 1 and 2 (see supplemental digital content, containing 2 tables and supplemental material). Twenty-one unique variants in *KRAS* (10) and *NRAS* (11) were used for the pilot. A total of 14 of 17 SNVs were used 2 to 3 times each, and all 4 MNVs were used 3 times each across the wet and dry samples. The VAFs for the dry samples were determined by the in silico engineering process. The VAFs for the wet samples were determined by droplet digital polymerase chain reaction (ddPCR). Twenty-one volunteer laboratories submitted results as part of the pilot. None of these 21 laboratories used the FDA-approved Praxis comparator assay. The CDx was only used by the assay manufacturer during the proof-of-concept portion of the study, when 3 laboratories pretested the standardized samples before they were sent to the other participating laboratories. For each variant detected, laboratories reported the total coverage depth at that position and the percent VAF as a whole number.

To reanalyze SPOT/Dx pilot data, the Mol Onc calculated the mean of the reported VAFs for each variant across all participating laboratories and used those values as the criterion standards to determine if a given laboratory successfully identified a given variant. This differs slightly from methods routinely used for PT programs, which use the mean reported VAF minus 2 SDs to determine the criterion standards. Laboratories were only graded for variants with a calculated mean VAF at or above their laboratory's validated limit of detection (LOD). This differed from the SPOT/Dx pilot, which graded participants as incorrect if they did not report a variant with an engineered/ddPCR determined VAF at or above the reported LOD of the FDA-approved Praxis comparator assay according to its package insert, without consideration of the validated LOD for each laboratory's assay or the calculated mean VAFs.

To provide comparative results for the pilot reanalysis, we also analyzed CAP PT program performance data for variants that were used in the SPOT/Dx pilot. The CAP PT programs included NGS for Solid Tumors (NGSST), NGS for Hematologic Malignancies (NGSHM), and NGS Bioinformatics (NGSB1/2). These PT programs use the mean of laboratories' reported VAFs minus 2 SDs for each variant, rounded to the nearest tenth, as the criterion standard. Laboratories are expected to detect and report variants with criterion standard VAFs at or above their laboratory's assay LOD, according to Clinical Laboratory Improvement Amendments standards.[4]

Figures were generated using the ggplot package (version 2.2.1, RStudio, Boston, Massachusetts; 2016; http://github.com/tidyverse/ggplot2/) loaded on R (version 3.4.1, R Core Team, Vienna, Austria; 2017; http://www.R-project.org/) and SAS 9.4 (SAS Institute, Cary, North Carolina).

RESULTS

Real-World Prevalence and Allelic Frequencies of SPOT/Dx Pilot Variants

We characterized the real-world prevalence of the variants used in the SPOT/Dx pilot and used the American Association for Cancer Research (AACR) Genomics Evidence Neoplasia Information Exchange (GENIE) data set (public release v12), specifically the colorectal cancer subset (n = 14 328).[5] Among the 21 unique variants in the pilot project, only 1 mutation, *KRAS* p.Gly12Asp, had a colorectal cancer patient population prevalence greater than 1.0% (n = 1826; 12.72%). The remaining 20 variants were present in less than 1.0% of samples from colorectal cancer patients, indicating the rare nature of 95.2% (20 of 21) of the variants. Of note, one-third (7 of 21; 33.3%) of the unique variants, 3 SNVs and all 4 MNVs, have never been reported in this large data set (Table 1).

The VAFs of the SPOT/Dx pilot samples were designed to be at low and challenging levels. VAFs for the dry samples were engineered at 5.0% or 15.0%. Target VAFs for the wet samples were 6.0% to 8.0%, with final VAFs determined by ddPCR to be between 5.1% and 8.6%. Of the 54 variants in the pilot, 35.2% (19) were engineered with VAFs of 5.0%, which was equivalent to the validated LOD of 76.2% (16 of 21) of laboratories for SNVs and at or below the validated LOD of 95.2% (20 of 21) of laboratories for MNVs. The median engineered VAF for the wet and dry samples in the pilot was 6.2% for both *KRAS* and *NRAS*. For real-world comparison, in the AACR GENIE data set the overall median VAF for variants in *KRAS* was 29.3% (n = 5993) and for *NRAS* it was 30.0% (n = 601). Figure 1 demonstrates the engineered and ddPCR-determined VAFs of the SPOT/Dx pilot dry and wet samples, respectively, compared with the VAFs reported in the AACR GENIE data set of 14 328 colorectal cancers.

In our review of the SPOT/Dx results, the mean VAF reported by participating laboratories using NGS differed from the engineered/ddPCR VAF for dry and wet samples, respectively (Figures 2 and 3). Overall, mean reported VAFs from NGS were lower than engineered/ddPCR VAFs for 47 of 54 variants
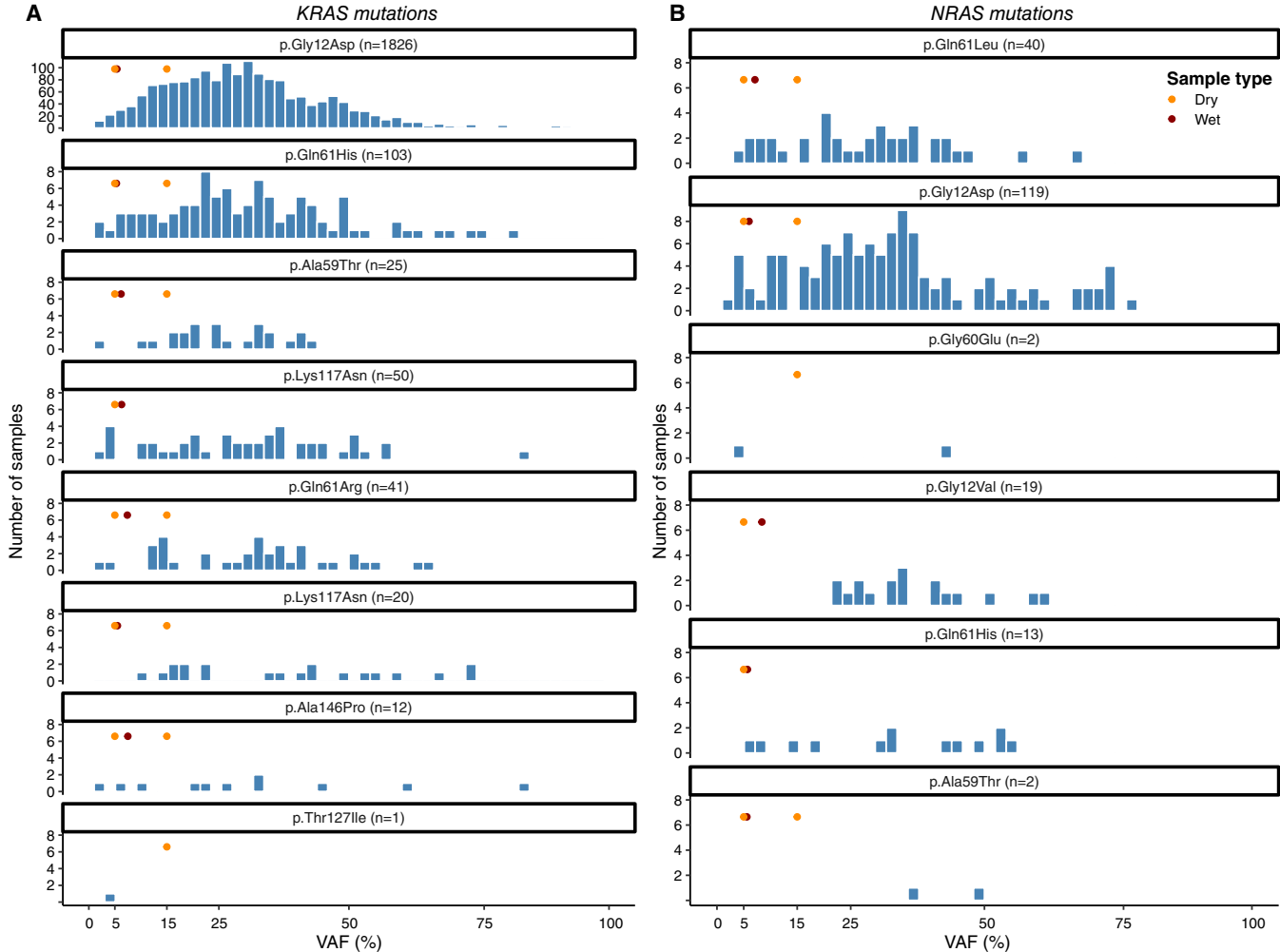
Table 1. Variants Included in the Sustainable Predictive Oncology Therapeutics and Diagnostics Quality Assurance Pilot Study (SPOT/Dx) That Have Never Been Reported in Colorectal Cancer in the American Association of Cancer Research Genomics Evidence Neoplasia Information Exchange Data Set of 14 328 Cases (Public Release v12)

Variant	No. of Times Used in the SPOT/Dx
KRAS c.38_39GC>AA, p.Gly13Glu	3
KRAS c.34_36GGT>TGG, p.Gly12Trp	3
NRAS c.38_39GT>AA, p.Gly13Glu	3
NRAS c.34_36GGT>TGG, p.Gly12Trp	3
NRAS c.436G>A, p.Ala146Thr	3
NRAS c.351G>T, p.Lys117Asn	2
NRAS c.443C>G, p.Thr148Ser	1

analyzed (87%). The mean VAF detected by participants was, on average, 0.4 percentage points below the engineered/ddPCR VAF for dry and wet samples combined. In addition, the mean reported VAF for 8 *KRAS* SNVs (7 dry, 1 wet), 8 *NRAS* SNVs (all dry), 3 *KRAS* MNVs (2 dry, 1 wet), and 1 *NRAS* MNV (dry) fell below 5%. Importantly, 76.2% (16 of 21) of the participating laboratories reported a 5% VAF as their LOD for SNVs, and 95.2% (20 of 21) of laboratories reported a LOD at or above 5% for MNVs. We applied the mean reported VAF, rather than the engineered/ddPCR VAF, as the criterion standard for determining if a laboratory's reported detection status could be included in the variant-specific analysis.

Performance for SNVs

SNVs are the most common class of actionable mutations of *KRAS* and *NRAS* in cancer. Using the mean reported VAF as the criterion standard, the overall detection rate of 42 *KRAS* and *NRAS* SNVs by pilot laboratories was 96.8% (716 of 740; Supplemental Table 1). For 16 of the 42 SNVs in the SPOT/Dx pilot, the detection rates increased to 100% compared with the detection rates in the original study (Figure 4). Of these 16 SNVs, 8 involved *KRAS* and 8 involved *NRAS*. Most of these

Figure 1. *Variant allele fraction (VAF) distributions of* KRAS *(A) and* NRAS *(B) variants reported in the American Association for Cancer Research (AACR) Genomics Evidence Neoplasia Information Exchange (GENIE) data set (public access v12), which overlap with the variants included in the Sustainable Predictive Oncology Therapeutics and Diagnostics quality assurance pilot study (SPOT/Dx). Orange and red dots indicate the engineered/droplet digital polymerase chain reaction–determined VAFs of the dry and wet samples for each variant in the SPOT/Dx pilot, respectively. Seven unique variants used in the SPOT/Dx pilot are not shown because they were never described as being present in the AACR GENIE data set of 14 328 colorectal cancers.*
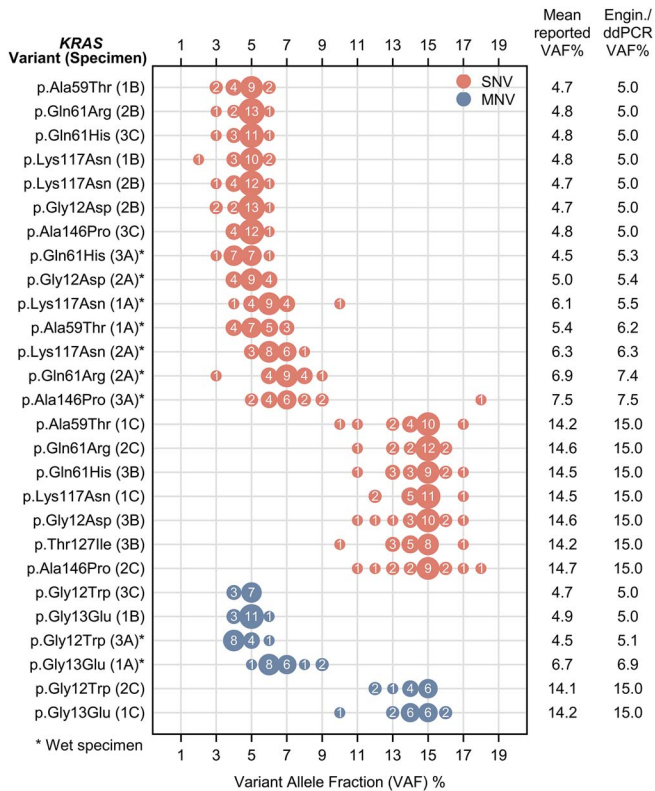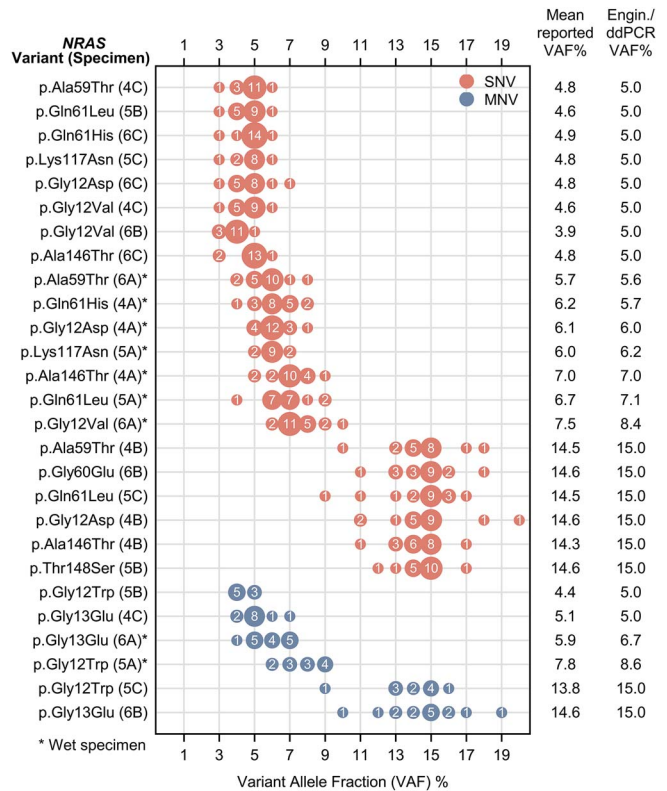
Figure 2. *Variant allele fraction (VAF) distributions reported by Sustainable Predictive Oncology Therapeutics and Diagnostics quality assurance pilot study (SPOT/Dx) laboratories for KRAS single-nucleotide variants (SNVs) and deletion-insertion (delins)/multinucleotide variants (MNVs). The bubble plot shows SPOT/Dx participating laboratories reported KRAS VAFs as whole numbers (x-axis) for each variant (y-axis), shown with the amino acid change and specimen number (in parentheses) used in the SPOT/Dx study. Mean VAFs rounded to 1 decimal place and engineered/droplet digital polymerase chain reaction (ddPCR)–determined VAFs are also shown to the side. The bubble size is proportional to the number of laboratories reporting the same VAF, which is indicated in the bubble. Orange and blue colors are used to indicate SNVs and delins/MNVs, respectively.*

Figure 3. *Variant allele fraction (VAF) distributions reported by Sustainable Predictive Oncology Therapeutics and Diagnostics quality assurance pilot study (SPOT/Dx) laboratories for NRAS single-nucleotide variants (SNVs) and deletion-insertion (delins)/multinucleotide variants (MNVs). The bubble plot shows SPOT/Dx participating laboratories reported NRAS VAFs as whole numbers (x-axis) for each variant (y-axis), shown with the amino acid change and specimen number (in parentheses) used in the SPOT/Dx study. Mean VAFs rounded to 1 decimal place and engineered/droplet digital polymerase chain reaction (ddPCR)–determined VAFs are also shown to the side. The bubble size is proportional to the number of laboratories reporting the same VAF, which is indicated in the bubble. Orange and blue colors are used to indicate SNVs and delins/MNVs, respectively.*

SNVs (93.8%; 15 of 16) were in silico and engineered to be at a 5% allelic fraction.

Performance for Deletion-Insertions/MNVs

MNVs consist of multiple adjacent SNVs that, according to the Human Genome Variation Society guidelines, should be classified as a single deletion-insertion (delins) mutation event.[6] The SPOT/Dx pilot included 12 MNVs, of which 6 were in *KRAS* and 6 in *NRAS*, representing 4 unique mutations across dry and wet samples with differing engineered/ddPCR-determined VAFs. The overall detection rate of *KRAS* and *NRAS* MNVs by pilot laboratories was 81.1% (129 of 159). Across these variants, our reanalysis shows increased detection rates for 7 variants compared with SPOT/Dx pilot analysis detection rates (Figure 4), all with engineered VAFs less than 10%. We also observed 15 instances when laboratories detected and reported a different mutation at the same codon as the intended MNV for a total of 7 variants—generally the first nucleotide of the intended MNV. This suggests bioinformatics pipelines at the time of the study interpreted the 2 or

3 adjacent nucleotide substitutions as separate, adjacent SNVs rather than merging them into a single MNV.

Additional Sources of Laboratory Errors

In addition to variants not being detected (false-negative results) and the "miscalls" (a false negative resulting from reporting 1 variant as a different variant) reported above, other sources of laboratory errors included variants detected below the validated LOD but not reported, and clerical errors. Table 2 shows a summary of false-negative results for SNVs and MNVs in relation to each laboratory's LOD. More than half (75 of 129; 58.1%) of the false-negative results reported in the original pilot article were for variants with mean reported VAFs below the laboratories' validated LOD, which were excluded from the reanalysis.

Table 3 shows, where possible, a breakdown of laboratory errors and the number of errors for each laboratory. For both wet and dry results, the number of false-negative results in the reanalyzed SPOT/Dx data was 6.0% (54 of 899 total participant results), with 24 false negatives for SNVs (3.2%; 24 of 740 total results) and 30 false
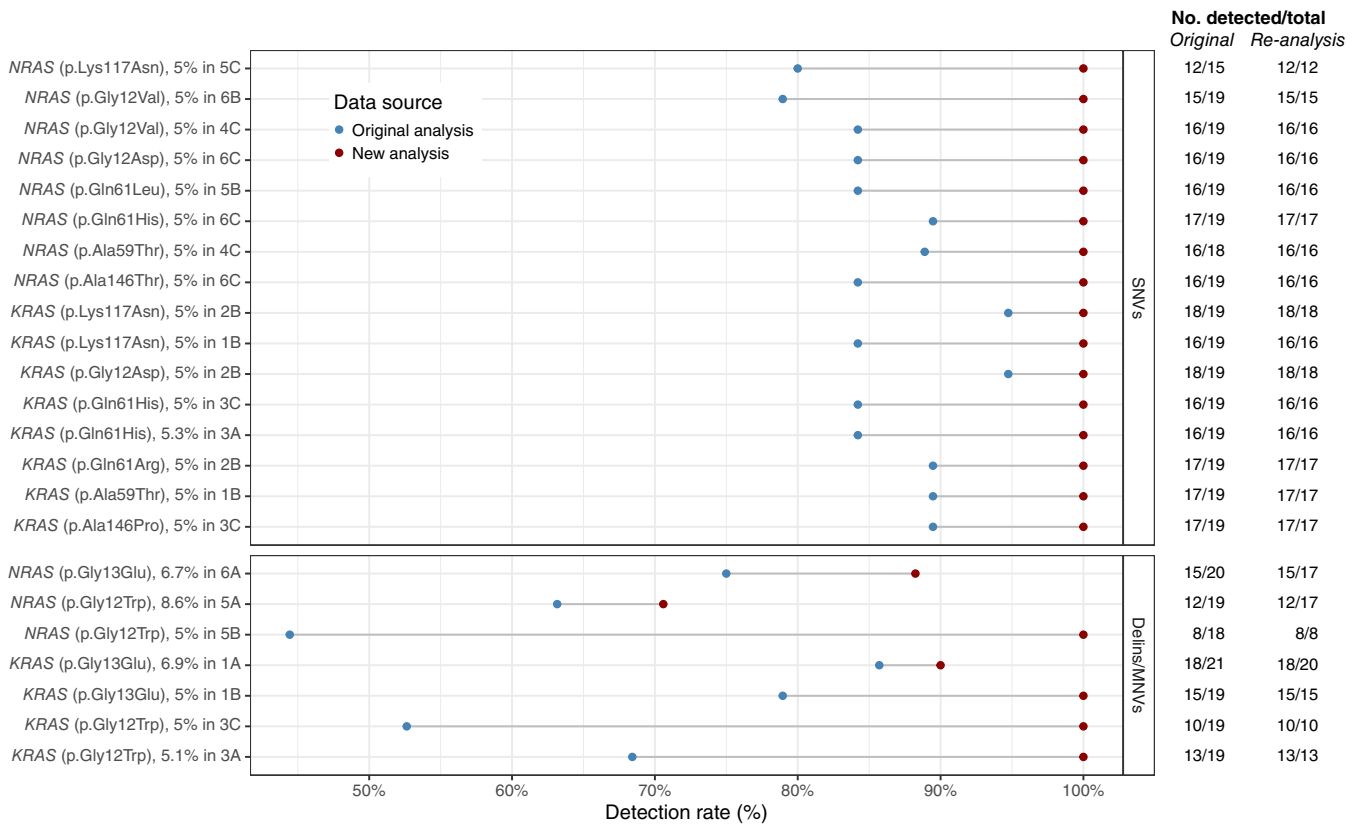
Figure 4. *Comparison of detection rates between original and new analyses. Detection rate of each single-nucleotide variant (SNV) and deletion-insertion (delins)/multinucleotide variant (MNV) across participating laboratories in dry and wet samples for variants with different detection rates in the original analysis and reanalysis. Colors indicate whether the data source is the original (prior) Sustainable Predictive Oncology Therapeutics and Diagnostics quality assurance pilot study (SPOT/Dx) analysis[2] (blue) or the reanalysis presented in this article (red). The y-axis labels indicate the gene, amino acid change, the engineered VAF, and the specimen number. The number of variant-specific responses included in the original pilot analysis and the reanalysis, respectively, are also shown to the side.*

negatives for MNVs (18.9%; 30 of 159 total results). Of the 54 total false-negative results, miscalls accounted for 15 (27.8%), detected but nonreported due to VAF below validated LOD accounted for 5 cases (9.3%), and clerical errors (ie, transcription errors) for 12 (22.2%); for the remaining 22 cases (40.7%), the reason for a false-negative result could not be determined.

Comparison with CAP PT Results

Finally, we compared the results of the SPOT/Dx pilot with results from CAP PT programs, which have been

conducted for a longer time and have been used by a larger number of laboratories across a broader range of NGS platforms. Although the SPOT/Dx pilot had 21 participating laboratories, between 2019 and 2022 the number of enrolled participants in the CAP NGSST, NGSHM, and NGSB1/2 PT programs ranged from 252 to 393, from 147 to 225, and from 40 to 56, respectively. We summarized the CAP PT program performance for variants that overlap between the SPOT/Dx pilot and the CAP PT programs in Table 4.

Overall, for *KRAS* and *NRAS* SNVs, variants were observed across a range of mean VAFs (8.9%–45.0%), with an overall detection rate of 97.2% (2671 of 2748). For *KRAS* and *NRAS* MNVs, variants were observed between mean VAFs of 9.2% to 18.9%, with a total detection rate of 91.8% (1853 of 2019). The overall detection rate of MNVs was highest in 2022, with laboratories detecting 97.3% (1161 of 1193) of variants across 5 MNVs. In addition to *KRAS* and *NRAS* MNVs, which are exceptionally uncommon in colorectal cancer, CAP NGS PT programs include MNVs that are prevalent enough to be included in variant data sets like AACR GENIE. For instance, *BRAF* c.1798_1799delGTinsAA, p.V600K is present in 3.6% of malignant melanomas according to AACR GENIE. In the CAP NGSST-A-2022 mailing, this MNV was present at a mean reported VAF of 11.6% and was detected by 98.1% (255 of 260) of participating laboratories. Similarly,

Table 2. False-Negative Results With Mean Reported Variant Allele Fractions (VAFs) at or Above and Below Participant Laboratories' Limit of Detection (LOD) for Single-Nucleotide Variants (SNVs) and Multinucleotide Variants (MNVs)[a]

	Mean Observed VAF		
Variant Type	At or Above Participant LOD	Below Participant LOD	Total
SNV	24	40	64
MNV	30	35	65
Total	**54**	**75**	**129**

[a] For the SPOT/Dx pilot data reanalysis, only variants with mean reported VAFs at or above a participant's LOD are interpreted as false negatives.

Table 3. Summary of False Negatives by Participant for the Original Sustainable Predictive Oncology Therapeutics and Diagnostics Quality Assurance Pilot Study (SPOT/Dx) Analysis and Reanalysis by the College of American Pathologists Molecular Oncology Committee

| | SPOT/Dx Pilot Analysis | | Reanalysis of SPOT/Dx Pilot Data | | | | |
| | | | | False Negative Reason Breakdown | | | |
Laboratory	False Negatives	Miscalls[a]	False Negatives	Miscalls[a]	Nonreported Identified Variants[b]	Clerical Error	Unknown
1	2	1	2	1			1
2	3		1		1		
3	0		0				
4	0		0				
5	2	1	1		1		
6	9		2				2
7	7	7	3	3			
8	7	7	3	3			
9	4	3	0				
10	0		0				
11	33		13			12	1
12	0		0				
13	8	2	5	1			4
14	28	4	12	1	3		8
15	5	5	3	3			
16	15	3	5				5
17	0		0				
18	6	6	3	3			
POC1	3	1	1				1
POC2	0		0				
POC3	0		0				
Total	132[c]	40	54	15	5	12	22

[a] Miscalls, that is, single-nucleotide variants instead of a dinucleotide variant or wrong nomenclature.
[b] Variants detected below the validated limit of detection were not reported according to laboratory policy.
[c] Three false-negative results were manually added after original grading in the SPOT/Dx.

in the CAP NGSHM-A-2022 mailing, *BRAF* p.V600K was present at a mean reported VAF of 21.6% and was detected by 100% (158 of 158) of laboratories.[7,8]

DISCUSSION

The SPOT/Dx pilot aimed to assess the consistency and accuracy of NGS LDTs across the landscape of molecular oncology testing. The pilot authors concluded that there was, in general, variable accuracy in the detection of genetic variants used to identify patients for targeted therapy.[2] The pilot strengths include its inclusion of multidisciplinary stakeholders and its focus on 2 challenging areas of molecular testing: LOD and MNVs. However, the SPOT/Dx pilot had multiple differences in its design and analysis compared with established PT programs, which inflated the appearance of variability in NGS performance and limited the generalizability of the results.

Our analysis of the SPOT/Dx pilot results using methods modeled after established PT programs shows that, contrary to the reported conclusions of the original SPOT/Dx pilot, laboratory performance for *KRAS* and *NRAS* SNVs was excellent, both in wet and dry engineered samples. The overall detection rate for SNVs was 96.8%. The reanalysis confirmed that MNVs, although exceptionally rare or never observed in colorectal cancer, were detected at

lower rates than SNVs, with an overall detection rate of 81.1%.

We also assessed the performance of the SPOT/Dx pilot variants that were also used in CAP NGS PT programs. The CAP PT programs have evaluated hundreds of participant laboratories, have a greater diversity of platforms and pipelines, assess laboratories at multiple time points, and assess variants and conditions that reflect the scope of clinical practice, which make the data more generalizable. CAP PT data confirmed that NGS laboratories demonstrate excellent performance identifying the *KRAS* and *NRAS* SNVs used in the SPOT/Dx pilot, with an overall detection rate of 97.2%. CAP PT data also demonstrated that MNVs in *KRAS* and *NRAS* are detected at an overall rate of 91.8%. The lower rate of detection of MNVs compared with SNVs is primarily attributable to earlier versions of bioinformatics pipelines miscalling MNVs as SNVs, which has improved over time. This is supported by the higher overall detection rate of 5 *KRAS* and *NRAS* MNVs (97.3%) in the most recently analyzed CAP PT data from 2022. MNVs are so uncommon in colorectal cancer (prevalence <.007%) that an analytic false-negative rate of 10% to 20% for these variants would result in a false negative or a miscall in fewer than 1 in every 70 000 colorectal cancer patients.[5] CAP PT performance data also demonstrate that a commonly encountered MNV (*BRAF* p.V600K) was detected by 98.1% to 100% of CAP PT program participants in 2022.

Table 4. *KRAS* and *NRAS* Single-Nucleotide Variants (SNVs) and Multinucleotide Variants (MNVs): Reanalyzed Sustainable Predictive Oncology Therapeutics and Diagnostics Quality Assurance Pilot Study (SPOT/Dx) Data Compared to College of American Pathologists (CAP) Next-Generation Sequencing Proficiency Testing (PT) Program Data

Gene	Variant Type	Transcript	Nucleotide Change Protein Change	Reanalyzed SPOT/Dx Data Mean Reported VAF, %	Detected, %	No. Detected/Total	CAP PT Program Data Mailing	Mean Reported VAF, %	Detected, %	No. Detected/Total
KRAS	SNV	chr12:25380275T>G	c.183A>C p.Gln61His	4.5 / 14.5 / 4.8	100.0 / 100.0 / 100.0	16/16 / 19/19 / 16/16	NGSST-B-2020	9.8	99.1	221/223
KRAS	SNV	chr12:25398284C>T	c.35G>A p.Gly12Asp	5.0 / 4.7 / 14.6	81.0 / 100.0 / 100.0	17/21 / 18/18 / 19/19	NGSST-A-2019 / NGSHM-A-2019 / NGSHM-A-2019 / NGSHM-B-2020	10.1 / 23.5 / 8.9 / 14.8	97.9 / 96.4 / 94.6 / 99.3	188/192 / 107/111 / 105/111 / 145/146
NRAS	SNV	chr1:115252204C>T	c.436G>A p.Ala146Thr	7.0 / 4.8 / 14.3	100.0 / 100.0 / 100.0	19/19 / 16/16 / 19/19	NGSHM-B-2020 / NGSHM-B-2022	18.5 / 14.6	88.7 / 97.2	110/124 / 139/143
NRAS	SNV	chr1:115258747C>T	c.35G>A p.Gly12Asp	6.1 / 4.8 / 14.6	100.0 / 100.0 / 100.0	20/20 / 16/16 / 19/19	NGSHM-A-2020 / NGSST-B-2020	15.1 / 20.3	100.0 / 98.7	132/132 / 223/226
NRAS	SNV	chr1:115256529T>A	c.182A>T p.Gln61Leu	6.7 / 4.6 / 14.5	90.0 / 100.0 / 94.7	18/20 / 16/16 / 18/19	NGSB1/2-B-2019 / NGSST-A-2019 / NGSHM-A-2019 / NGSHM-A-2020 / NGSHM-A-2021 / NGSST-A-2022	29.0 / 27.4 / 15.5 / 9.7 / 9.2 / 10.8	97.7 / 96.4 / 96.5 / 100.0 / 98.1 / 97.7	42/43 / 185/192 / 109/113 / 131/131 / 159/162 / 258/264
NRAS	SNV	chr1:115258747C>A	c.35G>T p.Gly12Val	4.6 / 7.5 / 3.9	100.0 / 100.0 / 100.0	16/16 / 21/21 / 15/15	NGSHM-A-2020 / NGSHM-B-2020	9.8 / 10.5	99.2 / 86.4	131/132 / 127/147
NRAS	SNV	chr1:115256528T>G	c.183A>C p.Gln61His	4.9 / 6.2	100.0 / 95.0	17/17 / 19/20	NGSHM-A-2022	45.0	99.4	159/160
KRAS	MNV	chr12:25398281_25398280 CG>TT	c.38_39GC>AA p.Gly13Glu	6.7 / 4.9 / 14.2	90.0 / 100.0 / 89.5	18/20 / 15/15 / 17/19	NGSHM-A-2019 / NGSST-B-2022 / NGSHM-B-2022	13.7 / 18.0 / 12.6	76.6 / 98.3 / 98.2	85/111 / 344/350 / 166/169
KRAS	MNV	chr12:25398285_25398283 CCA>ACC	c.34_36GGT>TGG p.Gly12Trp	4.5 / 4.7 / 14.1	100.0 / 100.0 / 68.4	13/13 / 10/10 / 13/19	NGSST-A-2021 / NGSST-B-2022	18.2 / 12.2	87.3 / 95.9	255/292 / 330/344
NRAS	MNV	chr1:115258744_115258743 CA>TT	c.38_39GT>AA p.Gly13Glu	5.1 / 5.9 / 14.6	80.0 / 88.2 / 83.3	12/15 / 15/17 / 15/18	NGSHM-A-2022	11.1	99.4	159/160
NRAS	MNV	chr1:115258748_115258746 CCA>ACC	c.34_36GGT>TGG p.Gly12Trp	7.8 / 4.4 / 13.8	70.6 / 100.0 / 61.1	12/17 / 8/8 / 11/18	NGSHM-B-2020 / NGSST-A-2021 / NGSHM-B-2022	18.9 / 9.2 / 16.3	82.6 / 83.5 / 95.3	119/144 / 233/279 / 162/170

Abbreviations: NGSB1/2, NGS bioinformatics; NGSHM, NGS for hematologic malignancies; NGSST, NGS for solid tumors; VAF, variant allele fraction.

SPOT/Dx Pilot Instructions

The SPOT/Dx pilot had several methodologic and analytic differences compared with established PT programs that may have affected the validity and generalizability of the pilot results. One difference was the failure to provide study participants with clear instructions. Participants were informed that:

> [t]his pilot is being performed in the context of an oncologist who is treating a patient with metastatic colorectal cancer and is considering panitumumab Vectibix therapy for this patient. Vectibix is indicated for the treatment of patients with wild-type *RAS* (defined as wild-type in both *KRAS* and *NRAS* as determined by an FDA-approved test for this use). Based on the sequence variants detected in each specimen, indicate on the result form how your laboratory would interpret the results in the context of testing for Vectibix therapy by selecting a response for "Mutation Reported." (Supplemental Materials)

From these statements, the pilot authors expected participants to report on the same variants with the same LOD as the FDA-approved Praxis assay linked to Vectibix, and not using the LOD that was validated in their laboratories, which contradicts standard clinical laboratory regulations.[4] In addition, there were no directions for reporting VAFs, which could only be reported as whole numbers. As a result, it is unknown whether laboratories rounded or truncated the VAFs they were reporting.

Another omission in the pilot instructions was the neoplastic cellularity for each specimen. Laboratories routinely use percent neoplastic cellularity to identify samples at risk of false negativity due to low VAFs. In fact, many laboratories will not test a sample with low neoplastic cellularity based on thresholds determined by their validation.[9] Even the FDA-approved Praxis assay requires laboratories to reject samples with tumor cellularity less than or equal to 50%.[10] The omission of neoplastic cellularity deprived pilot laboratories of critical information they needed for quality assurance and likely contributed to the appearance of variable performance, particularly for the large percentage of pilot samples with VAFs near standard LOD levels. In addition, the original SPOT/Dx pilot article stated that "[t]he cell lines in the wet samples were nominally 100% neoplastic," which is practically incorrect because the ddPCR measured VAFs.[2] Moreover, in clinical practice, the clinical relevance of a variant found at an allele fraction of 5% to 15% in a sample with 50% to 100% tumor cellularity is unclear; such an alteration would likely represent a subclonal alteration rather than an oncogenic driver for the cancer as a whole. This would be unusual for *RAS* variants in untreated colorectal cancer and represents a disparity between the pilot samples and real-world patient care.

SPOT/Dx Pilot Sample VAFs

A major difference between the SPOT/Dx pilot and established PT programs was that most of the dry and wet samples were designed to have VAFs at or near the validated LOD of most participants' assays. Because of the expected variation in VAFs detected by NGS assays,[11] many laboratories detected variants below their LOD and did not report them. This design amplified minor differences in the quantitative aspects of NGS versus other methods and artifactually increased the appearance of variable laboratory performance in the pilot.

NGS quantifies variants by counting the number of reads containing a variant and dividing by the total number of reads at that position.[12] To be counted, the reads must align to the reference genome. Because variants by definition differ from the reference genome, the likelihood of variant alignment is lower than for reference/nonvariant sequences. Consequently, the VAFs determined by NGS can be slightly lower than VAFs engineered by in silico mutagenesis or those determined by ddPCR, which do not use alignment to a reference genome for quantitation. The results of the SPOT/Dx pilot are consistent with this phenomenon, with 47 of 54 dry and wet variants (87%) having mean observed VAFs below the engineered/ddPCR-determined VAF. The mean VAF detected by participants was, on average, 0.4 percentage points below the engineered/ddPCR-determined VAF for dry and wet samples combined. Although this slight difference in VAF has no implication for clinical management, if it crosses the threshold for an assay's validated reportable range, it can affect whether a variant is reported. This SPOT/Dx pilot design weakness was also noted by Harada and Mackinnon, who stated that using standardized samples with VAFs that are lower than the LDT's LOD may not accurately reflect a laboratory's analytical performance and interpretation.[13]

The SPOT/Dx pilot was further confounded by the design of the result form, which only permitted laboratories to report VAFs as whole numbers. This also contributed to the appearance of poor performance in the study because it failed to differentiate between laboratories that detected a variant just below most laboratories' LOD (ie, 4.9%) and 5.0%.

As previously noted, the reanalysis in the current study used methods modeled after established PT programs. The criterion standard for each variant was the mean VAF for laboratories that detected the variant. This method reduces the risk of grading laboratories based on technical differences between NGS and the non-NGS techniques used to determine the VAF criterion standard of the engineered/ddPCR determined samples. The fact that this small methodological difference so dramatically altered the results of the pilot highlights the problem of using the engineered/ddPCR VAF as the criterion standard, as in the SPOT/Dx pilot.

SPOT/Dx Pilot MNVs

Another important difference between the SPOT/Dx pilot and PT programs was the inclusion of a disproportionately high number of MNVs. *KRAS* or *NRAS* MNVs are so rare that there are no examples in the AACR GENIE dataset (public release v12) of 14 328 colorectal carcinomas, yet they comprised nearly a quarter (12/54; 22.2%) of the variants in the pilot.[5] MNVs are known to be problematic, particularly for early versions of bioinformatics pipelines, which miscalled them as individual and/or adjacent SNVs.[14–16] Moreover, these miscalls would still have been interpreted as variants of strong clinical significance for therapeutic decisions, suggesting that miscalls of specific variants may result in the same clinical therapeutic approach, and reducing the actual impact of miscalls of these extremely rare variants.[17]

The correct identification and annotation of MNVs in colorectal carcinoma is an important opportunity for improvement in bioinformatics pipelines. It also highlights the opportunity for laboratories to correctly annotate MNVs called as separate SNVs by involving trained practitioners to directly examine aligned sequence files, rather than relying solely upon computational processes. The overall detection rate

of 97.3% for *KRAS* and *NRAS* MNVs by 2022 CAP NGS PT program participants suggests that bioinformatics pipelines and/or practices for reviewing sequence files have already been modified to improve MNV detection. The improvement of MNV annotations over time has also been observed in submissions of variant datasets to cBioPortal, further suggesting that pipelines and analytical practices have been updated to correctly annotate these variants.[18] The improvement in MNV detection over time also supports the value of PT programs as part of a quality management system to help vendors and laboratories improve and/or monitor trends in performance.

SPOT/Dx Pilot Omissions

The SPOT/Dx pilot article omitted important information about the underlying reasons for several of the reported errors. The pilot article described one laboratory with 13 unacceptable results. This single laboratory made a clerical error on the result form for this study that accounted for 12/13 of their errors and 9.3% (12/129) of the total errors reported in the pilot article. The error was substituting decimals for whole numbers when reporting VAFs (ie, reporting 5.0% as 0.05 instead of as 5). This kind of error is related to the contrived data entry portal for this study and does not reflect the laboratory's routine process for reporting clinical results.

As previously noted, many of the MNVs were miscalled as SNVs that represented the first nucleotide of the MNV. These miscalls could be considered annotation errors and they accounted for 11.6% (15/129) of the errors reported in the original article. In addition, 5/129 (3.9%) of the originally reported errors were due to laboratories not reporting rare to never-observed variants included in the CDx labeling but not in their LDT. This is particularly problematic given the existence of multiple FDA-approved assays for *RAS* testing in colorectal cancer that do not include the same variants as the Praxis Extended RAS Panel CDx.[19] Altogether, these report omissions mislead readers and skewed conclusions regarding laboratory performance, as they are more related to the design and analysis of the SPOT/Dx pilot.

Comparisons to Companion Diagnostic

The SPOT/Dx pilot was further confounded by attempts to compare LDT performance with the reported performance of the Praxis Extended RAS Panel CDx. The SNVs in the validation of the CDx were at VAFs greater than 15%, whereas the VAFs in the SPOT/Dx pilot were all 15% or less.[10] In addition, the validation of the CDx to detect MNVs was based on the study of only 2 dinucleotide variants and did not include evidence of ability to detect many of the MNVs included in the pilot. Finally, no clinical laboratories performing the Praxis Extended RAS Panel CDx were included in the study, precluding a direct comparison of the CDx and NGS LDTs in a clinical setting. Given these important limitations, attempts to compare SPOT/Dx pilot data to the reported performance of the CDx are problematic.

Generalizability of SPOT/Dx Pilot Results

Results obtained from a small pilot study focused on one disease with rare or never-reported variants at low VAFs cannot be generalized to overall laboratory performance for all types of cancer. The pilot results only reflect laboratory performance for the study samples, which represent a minute percentage of samples encountered in routine clinical practice. As stated by Harada and Mackinnon,[13] although the MNVs in the pilot "... serve the purpose of challenging a laboratory's informatics pipeline, they do not simulate a real-world situation," and the study design "... does not fully align with pan-tumor genomic analysis, which most laboratories are currently implementing." Laboratories may even be unable to test the types of samples included in the pilot, given that the tumor cellularity may be insufficient based on the validated analytic sensitivity of their NGS assays. The conclusions of the SPOT/Dx pilot about variable accuracy in the detection of genetic variants among some LDTs only apply to the samples with uncommon mutations at low-level VAFs included in the pilot and not to the performance of NGS LDTs overall.

Limitations of the Reanalysis

This study reanalyzed SPOT/Dx pilot data using analytic methods modeled after established PT programs to address some of the confounding variables. However, many of the problematic aspects of the study design could not be mitigated. The lack of clear instructions, omission of preanalytic quality assurance data, and failure to clearly communicate to participants how the resulting data would be assessed and ultimately used likely contributed to numerous false-negative results. Similarly, the reporting of VAFs as whole numbers likely led to the appearance of false negatives (ie, rounding to 5% instead of reporting a result of 4.9%).

Another limitation of the reanalysis is that the methods differed slightly from those used in established PT programs. To calculate the criterion standard VAFs for the reanalysis, we used the mean of each reported VAF, whereas PT programs routinely use the mean reported VAF minus 2 SDs. The reasons for the difference are 2-fold: The number of pilot laboratories was small, and many of the SPOT/Dx pilot variants were engineered at or near the participating laboratories' LOD. Using the mean VAF minus 2 SDs would have excluded most responses in the pilot because of the known interassay and intraassay variability in the observed VAFs. Hence, engineering standardized samples with target VAFs at the LOD will predictably result in a subset of laboratories detecting a variant below the target VAF and LOD, as evidenced in the variants in the pilot (Figures 2 and 3).

Lastly, 3 to 4 years elapsed between the SPOT/Dx pilot data collection (2018–2019) and publication (2022). This time interval limited our ability to communicate with participating laboratories for clarification about some of the factors associated with the observed errors. For example, we could not ask all laboratories with false-negative results if they did not report variants that were detected below their validated LOD. This information could have provided the reason for at least some of the 22 unknown false-negative results in Table 3.

CONCLUSIONS

Reanalysis of the SPOT/Dx study data with methods modeled after established PT programs revealed an overall detection rate of 96.8% for SNVs and 81.1% for MNVs. The comparison with CAP PT program data, obtained from hundreds of laboratories at multiple time points, is a much more substantial representation of laboratory practice, and it demonstrated overall SNV and MNV detection rates of 97.2% and 91.8%, respectively, with the detection rate for *KRAS*

and *NRAS* MNVs increasing to 97.3% in 2022. The reanalysis revealed multiple design and analytic differences between the original SPOT/Dx pilot and established PT programs, including the lack of clear instructions to study participants, variants engineered with VAFs at or below many laboratories' LOD, a disproportionately high number of rare to never-observed MNVs, the omission of important information about the underlying reasons for inaccurate results, and comparisons to the reported performance of an FDA-approved CDx that was validated under different conditions. In addition, the SPOT/Dx pilot findings about colorectal cancer are not generalizable to all molecular oncology testing. The conclusions of the SPOT/Dx pilot report about variable accuracy in the detection of genetic variants among some LDTs are limited by the pilot study design and methods. The SPOT/Dx pilot article should not be used as the basis to market products or change policy, given that some of the results have multiple confounding variables and only apply to performance when testing the rare conditions in the pilot rather than the performance of NGS LDTs overall.

References

1. Tapestry Networks. SPOT/Dx Working Group project Web site. https://www.tapestrynetworks.com/our-work/healthcare/spotdx-working-group. Accessed August 19, 2023.

2. Pfeifer JD, Loberg R, Lofton-Day C, Zehnbauer BA. Reference samples to compare next-generation sequencing test performance for oncology therapeutics and diagnostics. *Am J Clin Pathol*. 2022;157(4):628–638. DOI: 10.1093/ajcp/aqab164

3. Furtado LV, Souers RJ, Vasalos P, et al. Four-year laboratory performance of the first College of American Pathologists in silico next-generation sequencing bioinformatics proficiency testing surveys. *Arch Pathol Lab Med*. 2023;147(2):137–142. DOI: 10.5858/arpa.2021-0384-CP

4. Clinical Laboratory Improvement Amendments of 1988. https://www.govinfo.gov/content/pkg/STATUTE-102/pdf/STATUTE-102-Pg2903.pdf. Accessed March 3, 2023.

5. AACR Project GENIE Consortium. AACR Project GENIE: powering precision medicine through an international consortium. *Cancer Discov*. 2017;7(8):818–831. DOI: 10.1158/2159-8290.Cd-17-0151

6. Sequence variant nomenclature, version 20.05. Human Genome Variation Society Web site. http://varnomen.hgvs.org/recommendations/DNA/variant/delins/. Accessed January 28, 2023.

7. *Next-Generation Sequencing Solid Tumor*, NGSST-A, 2022, Proficiency Testing Program, College of American Pathologists. 2022.

8. *Next-Generation Sequencing Hematologic Malignancies*, NGSHM-A, 2022, Proficiency Testing Program, College of American Pathologists. 2022.

9. Devereaux KA, Souers RJ, Graham RP, et al. Neoplastic cellularity assessment in molecular testing. *Arch Pathol Lab Med*. 2022;146(9):1062–1071. DOI: 10.5858/arpa.2021-0166-CP

10. PraxisTM Extended RAS Panel [package insert]. San Diego, CA: Illumina. https://www.accessdata.fda.gov/cdrh_docs/pdf16/p160038c.pdf. Accessed June 7, 2023.

11. Spencer DH, Tyagi M, Vallania F, et al. Performance of common analysis methods for detecting low-frequency single nucleotide variants in targeted next-generation sequence data. *J Mol Diagn*. 2014;16(1):75–88. DOI: 10.1016/j.jmoldx.2013.09.003

12. Strom SP. Current practices and guidelines for clinical next-generation sequencing oncology testing. *Cancer Biol Med*. 2016;13(1):3–11. DOI: 10.28092/j.issn.2095-3941.2016.0004

13. Harada S, Mackinnon AC. New approaches and strategies for proficiency testing for next-generation sequencing-based oncology assays. *Am J Clin Pathol*. 2022;157(4):478–479. DOI: 10.1093/ajcp/aqab175

14. Ghoneim DH, Myers JR, Tuttle E, Paciorkowski AR. Comparison of insertion/deletion calling algorithms on human next-generation sequencing data. *BMC Res Notes*. 2014;7:864. DOI: 10.1186/1756-0500-7-864

15. Nardi V, Tsuchiya KD, Kim AS, et al. Next-generation sequencing somatic and germline assay troubleshooting guide derived from proficiency testing data. *Arch Pathol Lab Med*. 2022;146(4):451–461. DOI: 10.5858/arpa.2020-0842-CP

16. Wakeling MN, Laver TW, Colclough K, Parish A, Ellard S, Baple EL. Misannotation of multiple-nucleotide variants risks misdiagnosis. *Wellcome Open Res*. 2019;4:145. DOI: 10.12688/wellcomeopenres.15420.2

17. Dean L, Kane M. Panitumumab therapy and RAS and BRAF genotype. In Pratt VM, Scott SA, Pirmohamed M, Equivel B, Kattman BL, Malheiro AJ, eds. *Medical Genetics Summaries*. National Center for Biotechnology Information (Bethesda, MD); 2012:463–475.

18. Srinivasan S, Kalinava N, Aldana R, et al. Misannotated multi-nucleotide variants in public cancer genomics datasets lead to inaccurate mutation calls with significant implications. *Cancer Res*. 2021;81(2):282–288. DOI: 10.1158/0008-5472.Can-20-2151

19. List of cleared or approved companion diagnostic devices (in vitro and imaging tools). US Food and Drug Administration Web site. https://www.fda.gov/medical-devices/in-vitro-diagnostics/list-cleared-or-approved-companion-diagnostic-devices-in-vitro-and-imaging-tools. Accessed June 11, 2023.