Diagnostic Quality Assurance Pilot

White Paper – December 2021

Implications of performance variation in next-generation sequencing-based laboratory-developed tests for oncology: Stakeholder views

A companion discussion document to: Reference Samples to Compare Next Generation Sequencing Test Performance for Oncology Therapeutics and Diagnostics, published in *American Journal of Clinical Pathology*, December 2021

The below-named individuals, comprising select members of and liaisons (informal observers) to the Diagnostic Quality Assurance Pilot's steering committee, have reviewed and agreed to this document in its entirety and serve as its joint coauthors. All coauthors reviewed the final draft and reaffirmed their authorship immediately prior to publication. Tapestry Networks drafted this document based on interviews, held under a modified version of the Chatham House Rule,[1] conducted in 2019 and early 2020. Tapestry Networks synthesized and edited all comments prior to author review and approval. The views expressed in this article are the personal views of the authors and may not be understood or quoted as being made on behalf of or reflecting the position of the organization or agency or any of the committees or working parties of the organizations or agencies with whom the authors are affiliated. Author affiliations are current as of September 2021; some of the authors participated in the pilot steering committee through previous roles not listed here.

Jeff Allen, President & CEO, Friends of Cancer Research; Naomi Aronson, Executive Director, Clinical Evaluation, Innovation and Policy, Blue Cross and Blue Shield Association (BCBSA); Gabriel Bien-Willner, Medical Director, MolDx, Chief Medical Officer, Palmetto GBA; Paul Gerrard, Vice President, McDermott+Consulting; Lindee Goh, PhD, Partner, Tapestry Networks; Karen Gutekunst, VP, Regulatory Affairs, Illumina; Daniel F. Hayes, Stuart B. Padnos Professor of Breast Cancer Research, University of Michigan Comprehensive Cancer Center; Jennifer Keynton, R&D Manager, Diagnostic Reference Standards, Horizon Discovery; Robert D. Loberg, Vice President, Lifecycle Leader, Oncology & Genetics, Roche; Catherine Lofton-Day (independent/retired); Lisa M. McShane, Associate Director, Division of Cancer Treatment and Diagnosis, Biometric Research Program, National Cancer Institute; John Pfeifer, Vice Chair for Clinical Affairs, Pathology & Immunology, Washington University School of Medicine; Blase Polite, Associate Professor of Medicine, Deputy Section Chief for Clinical Operations, and Executive Medical Director for Cancer Accountable Care, University of Chicago Medicine; Girish Putcha, Chief Medical Officer and Clinical Laboratory Director, Freenome; Elizabeth Shaughnessy, Principal, Tapestry Networks; Barbara Zehnbauer, PhD, Adjunct Professor of Pathology, Emory University School of Medicine

Diagnostic Quality Assurance Pilot

Table of Contents

Abbreviations

AO	Accreditation organizations
CAP	College of American Pathologists
CDx	Companion diagnostic
CLIA	Clinical Laboratory Improvement Amendments
CMS	Centers for Medicare and Medicaid Services
FDA	US Food and Drug Administration
HER2	Human epidermal growth factor receptor 2
LDT	Laboratory developed test
LOD	Limit of detection
NGS	Next-generation sequencing
NYSDOH	New York State Department of Health
PT	Proficiency testing
SPOT/Dx	Sustainable Predictive Oncology Therapeutics and Diagnostics
STWG	Scientific and Technical Working Group
VAF	Variant allele frequency

Executive Summary

Important Note

This white paper represents a synthesis of discussions among steering committee members and liaisons in 2019 and early 2020. It was developed in parallel with the revision of the companion technical publication, a process that took several years. The coauthors recognize that molecular diagnostics is a rapidly changing field and encourage the community to view this paper as a set of perspectives offered at a specific point in time. Additional contributors not listed as authors provided relevant commentary reflected in the paper.

Furthermore, steering committee members recognize that stakeholders have had diverse perspectives on the pilot and its results. Some of these are reflected in the white paper's appendices.

Executive Summary

In 2013, the Sustainable Predictive Oncology Therapeutics and Diagnostics (SPOT/Dx) working group[2] committed to developing a new approach to assess laboratory test validation and performance. Participants aimed to ensure that oncologists could confidently select appropriate targeted therapies for treatment, regardless of which molecular diagnostic test was used.

Launched in 2016, the Diagnostic Quality Assurance Pilot provided clinical laboratories with digital images of tissue-section slides, engineered wet-lab reference samples (the "wet-lab challenge"), and in-silico sequence data file samples (the "in-silico challenge") and developed an evaluation methodology to assess the analytical performance of validated laboratory-developed tests (LDTs) relative to a US Food and Drug Administration (FDA)-approved companion diagnostic (CDx) for a targeted cancer therapy.

The pilot's results were published in *American Journal of Clinical Pathology* on December 2021. The pilot's steering committee (SC) has released this white paper as a companion document to the technical article to describe the pilot's impetus, design, and outcomes, as well as the SC members' and liaisons' reflections on the pilot's implications. This summary highlights SC member conclusions and recommended next steps.

The pilot's impetus, design, and technical outcomes

SPOT/Dx participants launched the pilot to test a way to assess comparability of analytical performance across advanced molecular diagnostic tests. Personalized medicine relies on the accurate detection of genetic variations that are targeted by pharmaceutical agents. LDTs do not receive the same scrutiny by the FDA as commercially marketed testing kits; therefore, questions about comparative performance informed a strategy to develop a new, reference sample–based approach to deploy in a pilot of diagnostic testing laboratories. The pilot's design embraced real-world practicalities as much as possible—for example, by choosing reference samples that could be sustained on a commercial basis and closely mimic patient tissue. Pilot data indicated that the pilot's process and reference samples worked as intended and can be used to evaluate the performance of next-generation sequencing (NGS) testing. Incidentally, through using the

reference samples provided in the pilot, most laboratories demonstrated that the performance of their assays met that of the CDx. The pilot results also indicated some variability in accuracy at some laboratories.

Stakeholder reflections on the pilot's implications

SC members voiced diverse views on the pilot's implications. Those involved in drug development noted that the pilot's data demonstrates the value of using a more uniform diagnostic in trials and research. Others suggested that the data could inform conversations about future advanced-clinical-diagnostics regulatory frameworks and help educate laboratories about the possible sources of error that the pilot identified. Some involved in the clinical care of patients found the results to be *"concerning"* but believe more work remains to better understand the direct impact of variant-detection errors on clinical decision-making and health outcomes. Consideration of whether a test meets clinically important performance thresholds was important to payers, though they had nuanced views about payers' role in incentivizing laboratories' further adoption of quality assurance. Finally, members recommended engaging with select patient organizations that are well versed in diagnostic policy to appropriately communicate the pilot's outcomes to patients.

Conclusions and recommended next steps

The SC shared common views on the pilot's major takeaways, the priority questions the pilot raised that should be addressed with the community, and the next steps stakeholders should consider.

Key takeaways

- Well-defined reference samples and customized in-silico files can be used by laboratories to compare the performance of their assays in detecting and reporting variants, including (but not necessarily limited to) those specified in an FDA-approved CDx. The pilot's approach could offer a new way for laboratories to validate and monitor their assays' performance as a complement to existing processes, especially for tests for selection of targeted therapies.
- Some variability in the detection of genetic variants was expected; the degree of variability the pilot revealed was a surprise to some SC members. *For an updated discussion on this variability and its meaning, please refer to the white paper's Annex.*
- Bioinformatics and sequence interpretation software must become essential areas of focus for test validation in laboratories, many believed. The in-silico challenge revealed that the root cause of several laboratories' errors might be attributed to the laboratory's interpretation software for a category of variants already known to be problematic for NGS platforms.
- Designing and engineering the reference samples and implementing the pilot's in-silico challenge were complex; however, vendors affirmed the replicability of the technology and general approach employed in the pilot for both the wet and in-silico challenges.

- The Centers for Medicare and Medicaid Services (CMS) regulate laboratory testing through the Clinical Laboratory Improvement Amendments (CLIA). CLIA certification of a laboratory may lack sufficiently defined quality criteria to ensure consistent performance of high-complexity tests, such as the NGS-based oncology panels evaluated as part of this pilot.
- In line with recommended good practices,[3] treating oncologists and pathologists should communicate before and/or after testing to better understand the clinical decision at hand, describe the test's potential limitations, and clarify nuances about the test's results.

Priority questions for further exploration

- Many laboratories demonstrated accurate performance within the scope of this pilot. What constitutes acceptable and obtainable analytical performance thresholds, especially for high-risk tests such as those for selection of targeted therapies? How can and should information on test performance and limitations be communicated to providers and their patients?
- Some experts noted that the types of variants analyzed as part of this pilot are not uncommon across the genome and are generally not among the most technically challenging when viewed in the context of recent advancements in biomarker-based sciences. Nonetheless, the likelihood of different types of variants being encountered in clinical practice varies depending on the specific markers being tested and the gene panel used. Therefore, how might laboratory performance compare for other types of variants that are rarer and/or difficult to detect (e.g., copy-number alterations and structural variants)?
- The laboratories involved were volunteers and may not be representative of all laboratories performing clinical NGS. Some members asked whether more variability would be found outside of such a self-selected group. Should this be evaluated and, if so, how?
- Although some members surmised that some errors in the participating laboratories could lead to inappropriate treatment decisions, the pilot did not solicit mock clinical reports, so the direct impact on patient care was beyond the scope of this study. What is the direct impact of variant detection and reporting errors on clinical decision-making and patient health outcomes, and how could future initiatives better assess this?
- A pharmaceutical manufacturer volunteered a candidate CDx and provided expertise and resources for the pilot's implementation, working with the SC to ensure transparency of outcomes. Moving forward, who should pay for the generation and sustained availability of reference samples? Who should incentivize their utilization and ensure the transparency and accurate reporting of results, and how would they do so?

Next steps

- **Educate laboratories about the trouble spots** that the pilot identified, especially those relating to data-interpretation software capabilities and limitations.

- **Inform relevant stakeholders**, such as oncologists, about potential variability in laboratory results relating to specific variant types, with appropriate caveats and discussion around the pilot's limitations.
- **Collaborate with platform manufacturers** to enable more seamless in-silico file analyses for quality assurance, given the value many SC members afforded to the in-silico component.
- **Explore additional steps with key healthcare stakeholders** as follows:
    - *Gather more data.* Expand the pilot to more laboratories or replicate its approach for other biomarkers or more difficult variants (e.g., larger indels, copy-number alterations or translocations). More closely assess the impact of variant-detection errors on clinician decision-making.
    - *Expand the use of reference samples and in-silico processes*. Institutionalize the pilot's process as an enhanced form of test performance validation or a complement to proficiency testing (PT) with accreditation organizations like the College of American Pathologists (CAP). Encourage payers to consider ways to support laboratories with quality assurance. Integrate development of well-defined reference samples as part of drug developers' clinical-development processes to assure appropriate postmarket treatment selection.
    - *Develop and standardize new approaches and best practices.* Improve transparency (i.e., awareness and understanding) about test performance, potentially through a multistakeholder body that could evaluate quality measures and validation tools and develop performance thresholds of clinical importance. Advance the importance of communication between oncologists and pathologists to improve test utilization and results interpretation or test limitations.

SC members valued the pilot's multistakeholder structure and are committed to broader engagement with stakeholders outside the SC to help them understand the pilot's outcomes and limitations.

White Paper

Introduction

Starting in 2013, a group of leading oncology stakeholders convened the SPOT/Dx working group, which aimed to assess how the US healthcare system could better adapt the diagnosis and treatment of cancer to the emerging precision-medicine era. Participants identified a significant yet overlooked challenge: targeted oncology therapeutics rely on the availability of advanced molecular diagnostic assays, many of which use NGS to identify appropriate therapies for patients with cancer, yet the performance of some of these tests is unknown to many stakeholders.

Some healthcare leaders have asked whether there needs to be greater transparency about clinical diagnostic performance given the recent growth and dynamism of the market, particularly for tests that directly inform treatment recommendations for patients.[4] An inappropriate treatment recommendation can have a tremendous impact on outcomes, particularly for patients with advanced cancers.

During SPOT/Dx discussions, representatives from industry, the laboratory community, patient-advocacy groups, medical professional societies, payers, and liaisons (informal observers) from relevant government agencies worked collaboratively to conceptualize a new approach to quality assurance that supported the interests, mandates, concerns, and resources of all parties. A subgroup of SPOT/Dx members subsequently launched the Diagnostic Quality Assurance Pilot in 2016 and formed the pilot's multistakeholder steering committee (SC) to oversee its implementation.

The data from the pilot were released in a technical publication in *American Journal of Clinical Pathology*. This white paper is a companion publication intended to share the SC's perspectives at a specific point in time on the pilot's design, outcomes, implications, and potential next steps to inform further discussion and action by the healthcare community. The paper provides a synthesis of the following:

- The pilot's impetus, design, and technical outcomes
- Stakeholder reflections on the pilot's implications
- Conclusions and recommended next steps

This white paper's content is based on group and individual discussions with various SC members throughout the course of the pilot's implementation, as well as focused conversations on the pilot's outcomes from May 2019 to March 2020. The white paper incorporates references to external viewpoints, literature, and analyses when relevant.

The pilot's impetus, design, and technical outcomes

Many critical decisions were taken from the pilot's launch in 2016 to its finalization in 2019. The project involved collaborative thinking, adaptation to the evolving technological and regulatory landscape, and incorporation of lessons learned. The pilot's inception, its design and

organizational structure, and the outcomes and lessons achieved are detailed in the following sections.

Impetus for a diagnostic quality assurance pilot

The regulatory framework for molecular diagnostic testing has been debated across the healthcare community and on Capitol Hill, with some stakeholders raising questions about whether the current framework sufficiently ensures the accuracy of advanced molecular diagnostic tests.[5] Against this backdrop, the pilot aimed to develop a process whereby validated LDTs could demonstrate the performance of their assays using well-defined reference samples that referred to an FDA-approved CDx.

Many individual US laboratories create their own LDTs; these are not typically regulated by the FDA in the same way as medical devices, which include in vitro diagnostic tests[6] marketed as kits to laboratories.[7] The FDA has exercised enforcement discretion over LDTs because in earlier years, these tests were deemed neither widespread nor complex, nor used in high-risk applications such as treatment selection.[8]

In addition to FDA oversight, the CMS CLIA program certifies clinical laboratories, reviewing various aspects of laboratory operations including, for example, quality control. CLIA authorizes specific third-party accreditation organizations (AOs) to aid in overseeing laboratories' compliance with CLIA regulations as the minimum standard that laboratories must meet. AOs, which include CAP, may have more stringent and specific requirements beyond those needed to assure compliance with CLIA regulations. CAP programs also inspect laboratories and provide PT services, which helps assess the accuracy and reliability of laboratory testing.[9] Cumulative PT findings across many laboratories have been publicly documented;[10] however, some stakeholders have observed that under current CMS CLIA mandates, the results of individual laboratory performance are primarily reviewed during biannual accreditation inspections and are not available in the public domain.

Similarly, some state regulatory agencies have developed their own requirements. For example, the New York State Department of Health (NYSDOH) now requires laboratories that are licensed to perform testing on New York State residents to perform PT and to submit their registered LDTs for approval, including evidence to support clinical validity or the accuracy with which a test identifies a patient's clinical status,[11] for each test.[12] Neither the results of PT nor the findings from the test review process are made public, though listings of such approved tests are available on NYSDOH's website.

With the expansion of targeted therapies, multimarker NGS-based LDTs have proliferated and far exceed the number of CDx tests approved by the FDA for those therapies. This has underscored the FDA's long-standing position that there is a public health need for greater oversight of LDTs. The FDA issued a discussion paper in 2017 on potential paths forward for LDT oversight that further details its perspective.[13]

Diagnostic Quality Assurance Pilot

Given the ongoing and often divisive discussions about regulation of advanced clinical testing, SPOT/Dx participants discussed a viable, voluntary way to assess performance across diagnostic assays that would not require a new law or regulatory paradigm to be enacted. One payer commented on the urgency for such an approach: *"This is needed yesterday. It should be self-evident that patients and providers want high-quality testing."*

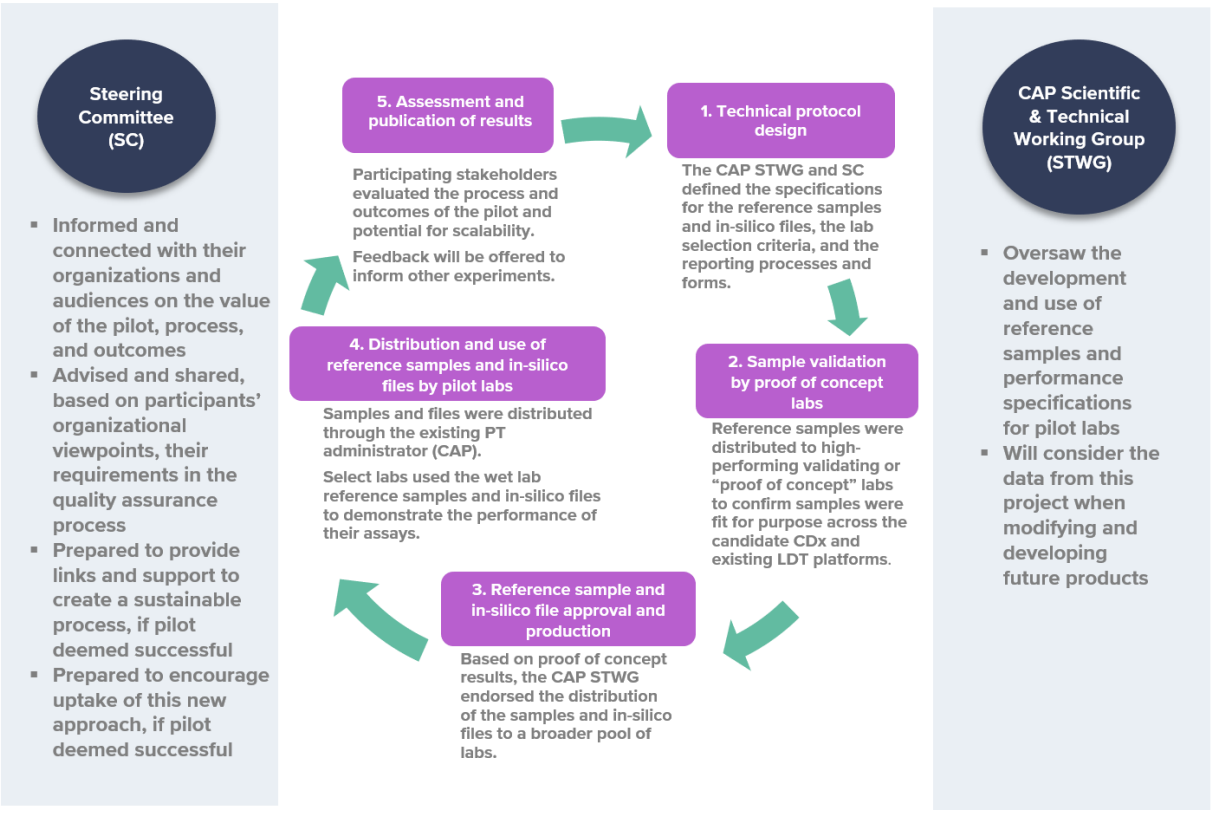
Designing and launching the pilot

The pilot compared LDT performance with specifications validated for an FDA-approved CDx, the Praxis™ Extended *RAS* Panel. The pilot also evaluated the efficiency and scalability of using genetically engineered reference samples and in-silico files to accomplish this comparison.

Biopharmaceutical developer Amgen and Amgen's CDx partner, Illumina, voluntarily proposed Praxis, which is a two-gene, multiple-variant NGS panel that helps identify patients with colorectal cancer who are eligible for treatment with Vectibix®, an *EGFR* inhibitor. This therapy is ruled out as a potential treatment option when *RAS* gene mutations are present because the *EGFR* inhibitor will be ineffective.

As is detailed in Figure 1, CAP helped implement the pilot with oversight from the pilot's SC and input from the pilot's scientific and technical working group (STWG). The latter guided the technical components of the pilot and comprised leading experts in the field of molecular pathology and NGS. Some members of the STWG served as lead authors for the pilot's technical paper.

Figure 1. Pilot workflow and organizational structure



The pilot's leadership embraced several principles, which are defined in the following box. *More detail on these principles and how they guided decisions about the pilot's design and implementation can be found in the white paper's Appendix.*

The pilot's principles

- **Sustainability.** Develop reference samples that can be manufactured and produced on a commercial scale.
- **Accelerated reference-sample creation.** Complete pilot prior to FDA approval of the candidate CDx and corresponding targeted therapy to ensure reference samples are ready for postmarket laboratory utilization.
- **Quick and effective action.** Complete proof-of-concept phase as rapidly as possible with select expert laboratories to ensure the reference samples are fit for purpose.
- **Transparency.** Promote visibility of pilot's outcomes.
- **Collaborative dialogue.** Build a pilot governance structure that ensures a diversity of perspectives.
- **Efficiency.** Use existing pathways and infrastructure as much as possible.

The pilot's technical outcomes and lessons, in brief

Throughout 2018 and early 2019, the selected laboratories received the challenge reference samples, conducted NGS testing and data interpretation, and provided their results to CAP. All participating laboratories were asked to report the sequence variants they detected. Laboratories also indicated how those variants would influence therapy selection using the CDx label as the model. CAP provided a brief clinical scenario explaining that the test was "being performed in the context of an oncologist who is treating a patient with metastatic colorectal cancer and is considering panitumumab (Vectibix®) therapy for [the] patient."[14] However, the pilot did not assess actual clinical laboratory reports, which may have more contextual explanations beyond detection of a sequence variant. The lack of clinical laboratory reports is an acknowledged limitation of the pilot.

Technical outcomes

The pilot's data are detailed in the companion technical paper; in summary, the pilot demonstrated that the type of reference samples created for the effort can be effectively used to evaluate performance across laboratories. Approximately two-thirds of laboratories demonstrated performance comparable with the FDA-approved CDx, with one-third of laboratories identifying all variants correctly. Multiple errors were identified in another one-third of laboratories; incidentally,

these primarily involved di- and trinucleotide substitutions (indels) and variants near each laboratory's LDT's lower limit of detection.

The pilot data signals that laboratories have more difficulty accurately identifying di- and trinucleotide sequence changes than single nucleotide variants. These errors were problematic in analyses of both the wet-lab and in-silico reference samples, signaling that insufficient expertise or implementation of postsequencing bioinformatics data-analysis tools were a contributor to the overall error rate. This prompted the pilot's technical leadership to suspect that some laboratories may have used outdated bioinformatics tools for sequence data interpretation. Such tools are highly customized and diverse across laboratories, making it challenging to identify any common root problems inherent in the bioinformatics data analyses.

The pilot's incidental findings also observed variation among laboratories in other areas, including the following:

- Identifying the appropriate level of neoplastic cellularity (amount of tumor cells) in a digital image of a stained tissue slide[15]
- Interpreting sequence changes that have been described in the *KRAS* and *NRAS* literature as potentially pathogenic but are *not* included in the CDx
- Some laboratories recorded variants with variant allele frequencies (VAFs) that were below the validated lower limit of detection of their respective LDT; however, the pilot did not collect data about interpretive comments that may have been included with clinical reports in these cases

The pilot had several limitations in addition to the lack of clinical reports acknowledged above. For example, because of the small sample size and the degree of variation across laboratories' approaches, it was not feasible to perform subgroup analyses to determine whether specific patterns of errors were dependent on specific characteristics of LDTs, such as specific NGS platforms or bioinformatics tools. Further details on the pilot's size, limitations, and outcomes can be found in the technical paper.

Lessons from the process

The SC learned several lessons about implementing a reference sample–based approach for assessing test performance. Others who may want to replicate this or a similar process in the future should consider the following:

- **Start-up took longer than anticipated.** CAP and other pilot participants spent considerable time in the early months of the project developing detailed contracts and request for proposal specifications for wet-lab and in-silico file vendors. Contracting and requests for proposals might be accelerated in subsequent iterations of the pilot or a similar approach.
- **The in-silico challenge required additional time and guidance from CAP and its vendors.** The in-silico challenge used a customized approach where laboratories generated sequencing files from a parent cell-line sample through their own pipelines and sent them to the in-silico vendor

via CAP's file-sharing portal. The vendor mutagenized the files and returned them to the portal, and laboratories then reintroduced the files into their bioinformatics pipelines. This process was challenging for some labs because many vendor platforms did not readily recognize externally introduced data files, which lacked a corresponding wet-lab sample. While work-arounds were available, they were complex to follow and time consuming for a subset of pilot labs.

- **The production of the wet lab's formalin-fixed paraffin-embedded reference samples, although integral to the pilot design, was costly and encountered technical delays.** The manufacture of custom human cell lines incurred several times the cost of the in-silico challenge. In addition, wet-lab sample development experienced slow cell-line growth that contributed to timeline delays.
- **The pilot's technical design was complex but scalable.** Vendors affirmed that the pilot's process was replicable; some also affirmed it could be streamlined in the future. One said, *"There were a couple of complexities in this project, but nothing entirely novel about the methodology and nothing that could not be worked around in the future."*

Stakeholder reflections on the pilot's implications

For this white paper, SC members and liaisons voiced nuanced and wide-ranging views about the pilot's outcomes and their meaning at a specific point in time. These opinions reflected members' broader stakeholder roles in the preclinical and clinical environment for diagnostics and therapeutics; some represent the views of technical experts, while others reflect those of payers, patient advocates, clinicians, and others. The following section presents details of stakeholder perspectives as they relate to various pre- and postmarket aspects of the pilot.

Premarket: Drug development and trials

The ongoing importance of targeted therapies within cancer treatment programs[16] and the corresponding growth of CDx's prompted discussion among SC members about the variability in results and its meaning. Specifically, some of the stakeholders involved in drug development, regulation, and other preclinical activities noted that it might be prudent to consider using a uniform diagnostic process in clinical trials. One said, *"This pilot lends support for having more centralized testing processes to confirm results for ongoing trials, and it brings to light that in clinical trials, there may be implications for companies that use local testing."* Another opined, *"The results confirm our worst fears. They show there are disagreements in terms of calls made."*

Some acknowledged, however, that financial constraints associated with early exploration or government-supported trials may necessitate the use of diverse local laboratories. One participant provided an example from a current large-scale clinical trial in which funding for centralized screening was exhausted and compelled a transition to community testing. The same participant underscored that using engineered reference samples like those developed in the pilot would

enable those running clinical trials reliant on these types of sequencing assays to better implement assay validation and quality-monitoring processes.

Others noted that biopharmaceutical companies could integrate well-defined reference samples into late-stage clinical-development pipelines. This would improve uniformity across diagnostics in the clinical-trial setting and prepare the materials to be available following approval.

Postmarket: Regulatory framework for clinical diagnostics

Some SC stakeholders opined that the pilot, while small, provided evidence that more oversight of NGS testing is needed, and that the pilot's data could be used as an informational tool for regulators, legislators, and others who are exploring the issue of regulatory oversight for clinical laboratory tests. Some asked whether the findings of this pilot might relate to the National Technology Transfer and Advancement Act of 1995, which required US federal agencies to use cooperatively developed standards and encouraged partnership with the private sector to do so.[17]

SC members shared diverse views on how the pilot's results, while limited, might inform the ongoing debates about optimal regulatory frameworks for laboratory testing:

- **The pilot's data could inform discussions on FDA oversight.** Some members urged that the FDA should have a more direct role in LDT oversight and noted that generating awareness within the FDA about the pilot's results may be a logical next step.[18] One explained, *"We've always favored FDA having more access to data about this issue, as it's hard to quantify the unknown. No one has any depth of information on nonregulated tests, and in the absence of information is where nervousness happens. This pilot confirmed the nervousness."*

> *"CLIA regulations involve things like adequately trained personnel and maintaining records of quality-control procedures, but there's a lot of room within that for what you really do."*
>
> —Subject matter expert

- **The pilot's outcomes raise questions about the way forward for CLIA, but opinions are nuanced.** Several professional organizations have long called for "CLIA modernization"[19] and some members emphasized that the existing 1988 CLIA accreditation regulations do not include specialty requirements for genetics or molecular pathology. Several were skeptical about the feasibility of CLIA updating regulations for the current environment, given its mandate, the time and resources involved in updating regulations, and the rapid pace at which molecular diagnostic technology is changing. However, some differed, saying that CLIA may simply need more time to *"work with the regulations."*

> *"CLIA is already looking at what they need to update. Given some time, CLIA will probably take care of it."*
>
> —Subject matter expert

- **The pilot's outcomes suggest that laboratories may need to provide more rigorous oversight of the software they use.** Presuming that software issues may be the root cause of some laboratories' errors in the pilot, some members were concerned that laboratories without deep bioinformatics expertise were relying too heavily on software to generate results, to the exclusion of considerations of assay design and library generation. Some thought that laboratories' understanding of their software and its potential vulnerabilities might need to improve and that the ways that their oversight of software is integrated with other quality measures might warrant further assessment. Some again called for a more prominent role for external regulatory oversight (i.e., via the FDA) on this issue; however, others underscored caveats to a regulatory solution given the FDA's mandate and rapid pace of software development. A simpler, more direct solution may be to develop better communication channels between NGS platform manufacturers and laboratory professionals to improve understanding of variant calling functions and external data-file analysis procedures.

> *"FDA has had to dance around this issue on how they should be regulating software associated with these systems. I'm sure FDA is not looking at every change in software; they can't possibly."*
>
> —Subject matter expert

Postmarket: Clinical laboratories

For many of the stakeholders involved in the pilot, what is of utmost importance is using the pilot's data to constructively support the laboratory community in addressing problems that may affect test quality—or, as one participant put it, *"How can we use the results of the pilot to help laboratories perform better?"* SC members involved in the laboratory community considered several implications of the pilot's outcomes for their colleagues. These are as follows:

- **There is a need to educate laboratories about risks for inaccurate variant calls and false positives and negatives, focusing on variants that are known to be problematic.** Members supported opportunities for relevant professional associations like CAP or the Association for Molecular Pathology to use the pilot's data to focus educational outreach on problematic variants. Future educational materials might identify the variation among laboratories that the pilot found around the accurate detection of single nucleotide variants with low variant allele fractions and di- and trinucleotide substitutions, and propose strategies for monitoring and adjusting lab processes.
- **Assessing laboratories' software limitations and utilization is important.** Echoing the concerns about software noted above, laboratory stakeholders want to better understand how laboratory staff are using bioinformatics software in practice. Laboratories might lack the expertise to carefully assess the analysis their systems generate and ensure the right calls are made. Some also cautioned that laboratories often customize their software, so making broad comparisons of potential trouble spots in software is challenging. However, some believed that these issues only further support the need for the pilot's approach: moving forward,

laboratories could use reference samples to more rigorously validate the accuracy of their bioinformatics pipelines.

- **The community could further prepare and support laboratories in implementing in-silico challenges.** Some believed that it was imperative to communicate with the laboratory community about the issues encountered during the in-silico challenge so that implementing similar approaches could be easier in the future. In addition, some believed that laboratories should engage sequencing-platform manufacturers to streamline the process for introduction of external files into laboratories' pipelines. More laboratories could then overcome a lack of bioinformatics expertise to participate more readily in in-silico–based challenges.
- **Laboratory leadership should further consider adopting and/or expanding the pilot's approach.** One potential route for evolving the pilot's approach into a business model could be the development of PT to replicate the pilot's methodology as a complementary performance-comparison service. One stakeholder said, *"This supports the need to have some type of uniform review process of tests that are out there so that everyone is being measured with [the] same measuring stick."* But realizing this vision, another noted, *"is an internal business decision for CAP"* or a similar AO.

> *"In-silico files provide a level of speed, flexibility, and economy that you can't get with exclusively wet-lab specimens. In-silico is a useful adjunct to the wet lab. You still need the wet-lab challenge, but if you want to ask if laboratories can find a robust array of mutations and different variant classes in different combinations, it's hard to imagine a wet-lab challenge scaling to that."*
>
> —Laboratory stakeholder

Postmarket: Treatment

Basing treatment decisions on reliable information is critical to an oncologist's ability to serve patients. Indeed, some SC members with a clinical background found the results to be *"concerning"* but believed that more work needs to be done to evaluate the direct impact of errors in variant detection and reporting on clinical decision-making.

Engagement between the oncology and laboratory communities about testing accuracy has precedent, particularly regarding known limitations of specific assays. Some stakeholders involved in the pilot drew parallels with the human epidermal growth factor receptor 2 *(HER2)* testing variability examined several years ago, which may have resulted in ineffective treatment for women with breast cancer.[20] As was the case with *HER2*, the pilot's outcomes may instill skepticism in oncologists for laboratory test results. One clinician said, *"When a pathologist tells us there are no* KRAS *or* NRAS *mutations, I assume that's true. These results suggest I should not be so confident. I don't know what to do now, and I'm not sure who to trust."*

The pilot did not collect mock clinical reports from laboratories and, ultimately, the requesting physician decides how to act on a given laboratory test's findings. Therefore, the pilot's leadership can only speculate on whether patients would have been declared eligible for Vectibix and whether they would experience beneficial responses or not, based on pilot laboratories' test errors. Further reflections about clinical impact are presented in the text box below.

Considerations on clinical impact

In the case of Vectibix, the Praxis Extended *RAS* Panel detects mutations that signal a patient is ineligible for therapy (referred to as a "rule-out" test). Errors where a variant was reported but reported incorrectly may have minimal effect, as a clinician will still observe that a mutation is present in the laboratory report even if the specific nucleotide base change is incorrect. In the case of a false negative, a patient could be declared eligible for a therapy for which he or she will have an ineffective response, thereby delaying more efficacious treatment options.

Therefore, some SC members noted that true false negatives—the failure to detect genetic variants—could be considered more detrimental than some of the other types of errors observed in this pilot, as *"all the clinician really needs/wants to know is if a mutation is present or not."* However, others noted this may not be the case for other types of diagnostic tests, such as those where an identified variant deems a patient eligible for therapy (a "rule-in" test).

Members involved in clinical treatment shared various thoughts about potential implications of the pilot:

- **Oncologists should be informed about the pilot's results, but their ability to lead on this topic may be limited.** Some SC members proposed that the American Society of Clinical Oncology and CAP could jointly issue an interim memorandum to oncologists about published data on the performance of laboratories in molecular diagnostics. This memorandum could address studies that have indicated both excellent performance and the variability in detecting specific types of variants demonstrated in this pilot. Other approaches beyond information sharing may be challenging to implement. A guidelines-based approach, which was employed

> *"Pharma companies are held to developing a diagnostic test for a targeted therapy and getting it approved as a CDx for these programs in a very stringent way. The concern I have is, once that approval happens, then the relevant drug goes into a marketplace that is just the Wild West."*
>
> —Industry representative

as the solution to the *HER2* situation,[21] may not be scalable as a general strategy for all biomarker tests given the significant resources this would entail. Furthermore, oncologists do not have the expertise nor the mandate to assess the quality of a laboratory's performance and, as such, other stakeholders need to lead in this area.

- **In the longer term, there is a need to advance good practices such as pretesting and/or posttesting communications between oncologists and laboratory professionals, especially when a potential targeted therapy selection is being considered.** These could convey the indications for testing or limitations of testing and improve interpretation of results by oncologists to support clinical decision-making. For example, it remains unclear how laboratories should report variants that are present at a very low VAF, especially those that may be below the clinical thresholds determined by the CDx. Should pathologists indicate that they had detected a variant but that the VAF would not necessarily preclude the patient from treatment, as might be indicated on the CDx label? Furthermore, laboratorians underscored that understanding the potential treatment decision from the treating oncologist would be more helpful than simply receiving a test order for "cancer genetic testing," because they could tailor test result interpretations accordingly and identify nuances, such as the challenge of very low VAFs described above. Oncologists may need to offer more specific indications of the clinical decision under consideration, but constraints on their time may impede this strategy. Several recommended further examination of this issue.
- **Pharma may be able to integrate development of reference samples into their clinical-development pipelines to ensure accurate postmarket treatment decisions for patients.** In addition to oncologists, pharmaceutical companies have a vested interest in the treatment process: they need to ensure that the correct patients are selected for the targeted therapies they manufacture, especially as value-based outcomes become more influential. As one industry participant explained, *"In the outcomes-based world we're moving to in the United States, knowing that the right patient gets the right drug is important."* Pharmaceutical companies could invest in wet-lab samples during the premarket process to confirm that samples are commercially manufactured and ready for laboratories to use to validate their assays once the drug is on the market. Others, however, were skeptical that pharma should play a leadership role in supporting laboratory validation and instead favored third-party or professional association leadership.

Postmarket: Reimbursement

Laboratory services are supported by reimbursements from healthcare insurers. The reimbursement landscape for NGS tests is constantly evolving, and as more large-scale panels emerge, cost-benefit assessments on these technologies are murky and complex.[22] Against this backdrop, the pilot's results led payer representatives to raise many questions. In general, payers view a test's conformity to clinically important accuracy thresholds as important to paying for

services that provide clinical benefit to beneficiaries and avoiding services that provide limited or no benefit.

Reflecting on the pilot's results, some payers agreed that laboratories need more education about the bioinformatics pipeline vulnerabilities the pilot identified. Some also called for enhanced transparency of individual laboratory results in PT. One emphasized, *"It's transparency—that's a significant problem. Whether it's proficiency testing or reviews that New York State or anybody does, they are generally not in the public domain, and we should do better than that."*

The degree of leadership that payers could undertake regarding laboratory quality assurance was a topic of debate, especially given the size, scale, and limitations of the pilot. Of note, other stakeholders pointed to payers as being in a unique position to craft incentives that could encourage further adoption of quality-assurance measures in the laboratory community. One payer representative said, *"I'm not sure we can rely on providers and guideline bodies to advance quality. They could do something, but I'm not sure if they will. So you look to labs and others to self-govern and to payers to use sticks if carrots don't work; sticks are coverage policies and contracts on the commercial side."*

Others had different opinions about whether they would consider employing such tools at this time. The establishment of industry consensus standards for validation—and the inclusion of the approaches used in this pilot in those standards—would support some payers' willingness and ability to require the use of such validation techniques prior to covering a test. Other payers pointed to analogous strategies from other areas of healthcare that they might consider. These could include the establishment of centers of excellence or other ways to recognize high-quality services, such as the National Quality Forum's endorsement process, whereby it endorses outcomes metrics and government healthcare programs use these metrics to assess quality.

Payers also made the following caveats about their leadership on this topic:

- **Financial context is important.** One explained, *"This issue doesn't percolate much to the top for most payers. Even though results of tests gate a larger portion of spend for a health insurance plan, genetic testing falls at 1%–2% of spending overall. But the issue is there's an implicit belief that the results are correct and accurate. It doesn't put up a red flag–from a purely overall spend perspective–unless we are able to tie in impact, implications of discrepancies, and so forth. Most payers are not that aware or have that as a top concern in terms of quality of test results."*
- **Professional society leadership remains preferable, though not all agree.** Some payers indicated they would prefer to lean on associations such as the American Society of Clinical Oncology and CAP to play a role as a first step for further action. Others, as noted in some of the commentary above, were more skeptical about society leadership.

Postmarket: Patient engagement

Patient interests were at the center of the pilot's vision to create an approach that would ensure that those undergoing cancer treatment could expect consistently accurate testing for targeted therapy selection.

Some members of the SC envisioned a more focused approach to patient engagement in sharing and explaining the results of the pilot. Because the pilot was small, had several limitations, and had nuanced results, a grassroots communications campaign may be ineffective. Members recommended instead that the pilot's leadership inform select patient organizations that are well versed in diagnostic policy and testing.

Conclusions and recommended next steps

Drawing from the above reflections, some SC members shared common views on significant takeaways from the pilot, the questions it generated, and the next steps the community should consider. Together, these elements may serve as a framework for further discussion with other members of the community.

Key takeaways

- Well-defined reference samples and customized in-silico files can be used by laboratories to compare the performance of their assays in detecting and reporting variants, including (but not necessarily limited to) those specified in an FDA-approved CDx. The pilot's approach could offer a new way for laboratories to validate and monitor their assays' performance as a complement to existing processes, especially for tests used to select targeted therapies.
- Some variability in the detection of genetic variants was expected, but the degree of variability the pilot revealed was a surprise to some SC members. *For an updated discussion on this variability and its meaning, please refer to the white paper's Annex.*
- Bioinformatics and sequence-interpretation software must be essential areas of focus for test validation in laboratories moving forward, many believed. The in-silico challenge revealed that the root cause of several laboratories' errors might be attributed to the laboratory's interpretation software for a category of variants already known to be problematic for NGS platforms.
- Designing and engineering the reference samples and implementing the pilot's in-silico challenge were complex; however, vendors affirmed the replicability of the technology and general approach employed in the pilot for both the wet-lab and in-silico challenges. The in-silico process's potential for flexibility and scope is appealing, and there are opportunities to make in-silico files more manageable for interpretation by laboratories with varying informatics expertise.

- CLIA certification of a laboratory may lack sufficiently defined quality criteria to ensure consistent performance of high-complexity tests, such as the NGS-based oncology panels evaluated as part of this pilot.
- In line with recommended good practices,[23] treating oncologists and pathologists should communicate before and/or after testing to better understand the clinical decision at hand, describe the test's potential limitations, and clarify nuances about the test's results.

Priority questions for further exploration

The pilot also prompted several questions that some SC members suggested were important to discuss with the broader medical community. These were as follows:

- **What is an acceptable level of error for a diagnostic test used to inform treatment decisions for a targeted therapy?** The performance bar the pilot set—vis-à-vis the CDx—was achievable. Two-thirds of laboratories were able to detect and report all variants accurately or close to the positive and negative percent agreement rates of the CDx. Furthermore, studies summarizing CAP proficiency testing results have demonstrated excellent performance of NGS assays for single nucleotide variants and indels, which comprise the majority of actionable variants in colorectal cancer.[24] However, the pilot also showed that some laboratories may not perform as well on certain types of variants that, while more challenging to detect than hot-spot mutations, are not, in the views of some experts, uncommon across the genome and on variants near the laboratories' lower limit of detection. This percentage of error rates by laboratories is higher than previously reported in similar surveys.[25]

> *"I don't think people will bat an eye if 50% of laboratories are missing calls if these calls happen in every 1 in a million patients. So how big of a deal is this?"*
>
> —Payer

> *"For the one patient you miss, it's a very big deal. Mutations may be rare, but the denominator—i.e., the "million" others have referenced—is not the number of patients but the number of tested patients, in whom this frequency is much less rare."*
>
> —Payer

- **How does laboratory performance differ for other variants that are more rare or difficult to detect?** More complex mutations exist than those addressed in the pilot, such as larger indels, copy-number alterations, and fusions or translocations. More complex, newer diagnostic technologies and biomarkers, such as tumor mutational burden panels and microsatellite instability, are also becoming more important as science advances. For several SC members, the variability of laboratories' performance in this two-gene *RAS* panel raises concerns about accuracy with these other types of mutations and tests.

- **Would more variability be observed in a broader survey of additional laboratories?** Participating pilot laboratories were already a self-selected group. While the pilot data set is small, the existence of variation within a volunteer group signaled to several SC members that more variation was likely to be found in the *"real world,"* which compounds some members' concerns.
- **What is the impact of variant detection and reporting errors on clinical decisions?** As noted earlier, some surmised that participating laboratories' errors could lead to inappropriate treatment decisions. However, the pilot did not solicit mock clinical reports, so the direct impact on patient care and clinical outcomes was beyond the scope of this pilot study and could be an area for further exploration.
- **What is the long-term business model for sustainability or extension of this pilot's approach?** While the technical approach of the pilot was proven sound, nearly all SC members questioned how the process would be sustained: Who, for example, would ultimately be responsible for paying for the generation and sustained availability of reference samples? Who would incentivize their utilization and ensure transparency of outcomes?

> *"It's concerning that in such a small and controlled environment there was such diversity in the answers. This resulted in the realization that, oh my gosh, we have problems, and what do we do about them? You can't have problems and not do something about it. What do we do, knowing we have this information where we've got pretty disparate quality signs here?"*
>
> —Anonymous liaison

Next steps

Broadly, some stakeholders continue to ask whether more regulation is an appropriate solution to encourage standardization in complex NGS tests. Although the pilot's results were small in scale, the outcomes suggest there is variation across laboratories for certain types of variants. While SC members are not able to offer consensus-based recommendations about regulation based on this pilot, members *will* openly discuss the pilot's outcomes with all stakeholders, recognizing there may be a diversity of views. Many underscored the importance of the results' nuances: the pilot did not provide black-and-white answers but provided directional insights on critical questions the community needs to consider, as listed in the previous section.

Throughout this white paper are several ideas for next steps advanced by SC members and liaisons. Many consistently prioritized the following immediate actions:

- **Educate laboratories on the trouble spots** identified by the pilot, especially those relating to interpretation software.
- **Inform relevant stakeholders beyond laboratories** (e.g., oncologists) about the potential variability in laboratory results relating to specific variant types, with appropriate caveats and discussion around the pilot's limitations.

- **Discuss simpler workflows with platform manufacturers** to enable more seamless in-silico file processes, given the value SC members afforded to the in-silico process.

Other potential next steps that SC members raised for further discussion with external stakeholders are as follows:

- *Gather more data.* Expand the pilot to more laboratories or replicate its approach for other biomarkers or difficult variants (e.g., larger indels, copy number alterations or translocations). More closely assess the impact of variant detection errors on clinician decision-making.
- *Expand the use of reference samples and in-silico processes*. Institutionalize the pilot's process as an enhanced form of test performance validation or a complement to PT with accreditation organizations like CAP. Consider engaging with payers to further support laboratories with quality assurance. Integrate the development of well-defined reference samples as part of drug developers' clinical-development processes to assure appropriate postmarket treatment selection.
- *Develop and standardize new approaches and best practices.* Improve transparency (i.e., awareness and understanding) about test performance, potentially through a multistakeholder body that could evaluate quality measures and validation tools and develop performance thresholds of clinical importance. Advance the importance of communication between oncologists and pathologists to improve test utilization and results interpretation or test limitations.

Although the next steps advocated are diverse, several SC members believe the data from the pilot may be compelling enough to warrant action on potential follow-on initiatives. Many reiterated their support for the pilot's process and the value of having a multistakeholder SC during its implementation. Others emphasized that the pilot's results should not be viewed as a final outcome but as a first step in evolving quality-assurance processes as biomarker-based science rapidly accelerates. One said, *"The collaboration of this was great. It's a question of whether it's this or something else—not 'How do we do [the pilot] better?' but 'How do we do it faster so we can iterate faster?' We want to have an impact."*

About this document

Appendix: The pilot's design and guiding principles

The pilot's design team considered several core principles as it developed the pilot. These are discussed below.

Sustainability

The pilot sought to develop reference samples that could be manufactured and produced on a commercial scale. During early conversations, some SC members expressed a desire to compare test performance among laboratories using *both* reference samples and patient samples. Such an approach would shed light on the level of commutability[26] across engineered reference samples and real patient tissue. However, patient samples are difficult to access and distribute on a large-scale basis and are not readily sustainable in practice. Therefore, members agreed that adding a commutability component was outside the scope of the current pilot.

Sustainability was also a factor in prompting the STWG to add an in-silico challenge component. Members believed that comparing laboratory performance through a file-sharing platform could be faster, cheaper, and allow for more creativity and flexibility with respect to the variants tested than by using engineered wet-lab samples alone.

Timely reference material creation and availability

The original intent was to complete the pilot prior to FDA approval of the candidate CDx and the corresponding targeted therapy prior to clinical implementation. By making high-quality reference samples available *before* regulatory approval, the pilot would provide samples that were ready for use in a new test-validation/quality-assurance approach, should laboratories wish to utilize them.

The FDA approved the candidate drug and CDx before the pilot's completion. This did not directly affect the results of the pilot. Some members were concerned that participating laboratories may have switched to using the CDx in their clinical testing rather than their own LDTs; however, none did, as the approved CDx was not available for sale in the United States before participating laboratories completed the pilot.

Quick and effective action

The pilot strived to complete a proof of concept as rapidly as possible and adjust the proposed process as needed. Therefore, the pilot began with a dedicated proof-of-concept phase during which CAP sent all reference samples to three expert laboratories and the CDx manufacturer Illumina's own laboratory. These proof-of-concept-phase participants ensured that the reference samples were fit for purpose and appropriate for sharing with the larger group of pilot laboratories.

The SC and STWG deemed this phase highly valuable, as it helped identify potential trouble spots in the processes before wider distribution. For example, it helped the implementation team anticipate that some laboratories might have difficulties with the file-sharing process involved in the in-silico challenge.

Transparency of results

Several stakeholders within the SC have expressed concern that they typically have been unable to obtain detailed information on the performance of LDTs, such as the results of PT surveys from CAP or test evaluations by NYSDOH, as CMS does not currently mandate that this information be made publicly available outside of laboratories' AOs. In designing the pilot, the SC was committed to promoting the visibility of its outcomes, and members agreed from the start to make the results of the pilot public. However, the types and levels of detail which the pilot would share was a topic of discussion. Some members advocated that the identities of individual pilot laboratories be kept confidential; otherwise, few would be willing to participate, especially given that the pilot would be a learning process for all involved.

To ensure that CAP could recruit enough laboratories, the SC agreed to keep individual laboratories' identities confidential, including to the SC itself, though information on the percentage of labs able to achieve high levels of concordance and data on the types of platforms used could be shared.

Collaborative dialogue

The pilot aimed to ensure a diversity and balance of perspectives among stakeholders involved in its leadership. The SC that oversaw the pilot's implementation represented a variety of institutions relevant to diagnostics in both premarket and postmarket settings, which at the time of the pilot's launch was unique, at least as far as SC members are aware. This ensured that if the pilot's approach were eventually expanded, broader views and potential implications could be discussed in advance.

Furthermore, the SC recognized that the pilot was not operating in a vacuum and the regulatory framework was in flux as various proposals for clinical-diagnostics regulatory reform were being considered by legislators. Although the pilot explicitly did not aim to influence policy, the multistakeholder composition of the group helped ensure that the pilot's approach remained relevant throughout the process and that external efforts or changes in the landscape did not nullify the need for the pilot.

Finally, the pilot continued to engage with stakeholders outside of the SC and STWG throughout the process. Members frequently delivered presentations describing the pilot's progress at various forums, including the annual meetings of the Association for Molecular Pathology and the American Association for Cancer Research. Ongoing updates were also made available to the public on the Tapestry Networks website.[27] Through these public presentations and other materials, the pilot's leadership apprised the broader diagnostic community of the pilot and its objectives.

Efficiency

The pilot aimed to work within existing mandates and use existing communication pathways and distribution infrastructure as much as possible. CAP emerged as the leading technical

implementation partner for the pilot given CAP's existing network of laboratories participating in PT. CAP recruited volunteer laboratories for the pilot from among labs enrolled in CAP's NGS Oncology PT and leveraged its new cloud-based file-sharing platform to conduct the in-silico challenge.

The overarching principle of working as close to a real-world environment as possible underlay other key decisions in the pilot and its methodology. For example, the STWG decided to use isogenic human cell lines engineered with CRISPR technology for the wet-lab reference samples to mimic patient samples as closely as possible. In a similar vein, the pilot laboratory selection process aimed to model the diversity in the laboratory market. Although CAP recruited laboratories on a volunteer basis, the final group intentionally represented a mix of laboratory profiles (e.g., a laboratory's status as an academic or community laboratory, the type of platforms it uses, and its annual volume of relevant testing). As noted earlier, all participating laboratories used laboratory-developed assays and none of the participants used the FDA CDx.

Endnotes

[1] This white paper reflects the use of a modified version of the Chatham House Rule whereby comments are not attributed to individuals, corporations, or institutions. Comments by members and liaisons of the steering committee and other stakeholders appear in italics.

[2] The SPOT/Dx working group was a multistakeholder network facilitated by Tapestry Networks that addressed the emerging precision medicine paradigm in the diagnosis and treatment of cancer. It was operational from 2013 to 2015. "SPOT/Dx Working Group," Tapestry Networks, accessed August 10, 2021.

[3] See, for example, Bin Chen et al., "Good Laboratory Practices for Molecular Genetic Testing for Heritable Diseases and Conditions," *MMWR Recommendations and Reports* 58, no. RR06 (2009), 1–29; Erin P. Balogh et al., *Improving Diagnosis in Healthcare* (Washington, DC: National Academies Press, 2015), 153–154; and John Hickner et al., "Primary Care Physicians' Challenges in Ordering Clinical Laboratory Tests and Interpreting Results," *Journal of the American Board of Family Medicine* 27, no. 2 (2014), 268–274.

[4] For discussion on the debate that followed the FDA's guidance on a proposed framework for LDT regulatory oversight, see Jonathan R. Genzen, "Regulation of Laboratory-Developed Tests: A Clinical Laboratory Perspective," *American Journal of Clinical Pathology* 152, no. 2 (2019), 122–131.

[5] Congresswoman Diana DeGette, "DeGette, Colleagues Release Draft Legislation to Modernize FDA Regulation of Diagnostic Tests," news release, December 6, 2018.

[6] In-vitro diagnostic tests are tests conducted on samples such as blood or tissue that, among other uses, can be used to identify patients that are likely to benefit or not benefit from specific treatments. "In Vitro Diagnostics," FDA, accessed August 11, 2021.

[7] Jeff Schreier, Robert Feeney, and Peter Keeling, "Diagnostics Reform and Harmonization of Clinical Laboratory Testing," *Journal of Molecular Diagnostics* 21, no. 5 (September 2019), 737.

[8] "Laboratory Developed Tests," US Food and Drug Administration, accessed August 11, 2021.

[9] Centers for Medicare & Medicaid Services, *Clinical Laboratory Improvement Amendments (CLIA): Proficiency Testing and PT Referral Dos and Don'ts* (CMS, 2017).

[10] Ann Moyer et al., "Genotype and Phenotype Concordance for Pharmacogenetic Tests Through Proficiency Survey Testing: An Update," *Archives of Pathology and Laboratory Medicine* 144, no. 9 (2020), 1057–1066.

[11] Wylie Burke, "Genetic Tests: Clinical Validity and Clinical Utility," *Current Protocols in Human Genetics* 81 (2014), 9.15.1–9.15.8.

[12] Jonathan Genzen, "Regulation of Laboratory-Developed Tests: A Clinical Laboratory Perspective," *American Journal of Clinical Pathology* 152, no. 2 (2019), 122–131; "Modernization of CLIA: LDTs," American Association of Clinical Chemistry, November 30, 2018.

[13] Food and Drug Administration, *Discussion Paper on Laboratory Developed Tests (LDTs)* (Food and Drug Administration, January 13, 2017), and "Laboratory Developed Tests," US Food and Drug Administration, accessed August 10, 2021.

[14] The College of American Pathologists Quality Pilot Scientific and Technical Working Group, "Results: Steering Committee Update, Part 3," (slide presentation, Quality Pilot Steering Committee Meeting, Washington, DC, May 28, 2019), slide 3.

[15] Irina Heid et al., "Co-clinical Assessment of Tumor Cellularity in Pancreatic Cancer," *Clinical Cancer Research* 23, no. 6 (2016).

[16] IQVIA Institute,*Global Oncology Trends 2018* (IQVIA, May 24, 2018).

[17] "National Technology Transfer and Advancement Act of 1995," NIST, accessed August 11, 2021.

[18] Several FDA staff participated in the pilot as liaisons (i.e., informal observers) throughout the course of its implementation.

[19] See, for example, Association for Molecular Pathology,*Proposal for Modernization of CLIA Regulations for Laboratory Developed Testing Procedures (LDPs)* (AMP, August 2015).

[20] Antonio C. Wolff et al., "Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update,"*Archives of Pathology & Laboratory Medicine* 142 (November 2018), 1364–1382.

[21] Wolff et al., "Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update."

[22] As one example of this complexity, some studies have suggested that the use of multigene panels may not add significant costs to payers, but only when taking several factors into account, such as the presumed cost of multiple rounds of single-gene testing and the presumed benefit of targeted therapies in specific types of cancer. Tiffany M. Yu et al., "Budget Impact of Next-Generation Sequencing for Molecular Assessment of Advanced Non–Small Cell Lung Cancer,"*Value in Health* 21, no. 11 (November 2018), 1278–1285. Another recent cost-benefit assessment of multigene-panel testing found that it is "moderately" more cost effective than single-marker testing in lung cancer patients: L. Steuten et al., "Cost Effectiveness of Multigene Panel Sequencing for Patients With Advanced Non-Small-Cell Lung Cancer,"*JCO Clinical Cancer Informatics* (June 26, 2019).

[23] See, for example, Chen et al., "Good Laboratory Practices for Molecular Genetic Testing for Heritable Diseases and Conditions"; Balogh et al.,*Improving Diagnosis in Healthcare* ; and Hickner et al., "Primary Care Physicians' Challenges in Ordering Clinical Laboratory Tests and Interpreting Results."

[24] Jason Merker et al., "Proficiency Testing of Standardized Samples Shows Very High Interlaboratory Agreement for Clinical Next-Generation Sequencing-Based Oncology Assays,"*Archives of Pathology & Laboratory Medicine* 143, no. 4 (2018), 463–471, and Lea F. Surrey et al., "Next-Generation Sequencing (NGS) Methods Show Superior or Equivalent Performance to Non-NGS Methods on BRAF, EGFR, and KRAS Proficiency Testing Samples,"*Archives of Pathology & Laboratory Medicine* 143, no. 8 (2019), 980–984.

[25] Pfeifer et al., "Reference Samples to Compare Next Generation Sequencing Test Performance for Oncology Therapeutics and Diagnostics"*American Journal of Clinical Pathology, 2021;, aqab164.*

[26] Commutability is "the equivalence of the mathematical relationships between the results of different measurement procedures for a reference material and for representative samples from healthy and diseased individuals." "Commutability Study," National Institute of Diabetes and Digestive and Kidney Diseases, accessed August 11, 2021.

[27] "Diagnostic Quality Assurance Pilot," Tapestry Networks, accessed August 11, 2021.

Annex: Summary of Themes -Addressing lessons from the Diagnostic Quality Assurance Pilot

Previously published reflections from the December 2020 Molecular Diagnostic Quality Assurance Pilot Summit

Addressing lessons from the Diagnostic Quality Assurance Pilot

Introduction

In 2013, the Sustainable Predictive Oncology Therapeutics and Diagnostics (SPOT/Dx) working group committed to developing a new approach to assess laboratory test validation and performance. Consequently, the Diagnostic Quality Assurance Pilot was launched in 2016 with the aim of ensuring that oncologists could confidently select appropriate targeted therapies for treatment, regardless of which molecular diagnostic test was used to inform their recommendation.[1] A multistakeholder steering committee—which included representatives from the American Society of Clinical Oncology, the College of American Pathologists, Friends of Cancer Research, public and private payers (including Palmetto GBA, Blue Cross Blue Shield Association, and others), industry representatives from Amgen and Illumina, and observers from the National Cancer Institute, US Food and Drug Administration (FDA), and Centers for Medicare and Medicaid Services—helped inform the pilot's efforts. The Scientific and Technical Working Group, overseen by the College of American Pathologists and comprising leading molecular pathology experts, supported the pilot's technical design and execution. Steering committee and working group leadership convened in December 2020 with alumni from the original SPOT/Dx working group and other stakeholders for a summit to discuss the pilot's results and implications.

Over the course of implementation, the pilot provided clinical laboratories with engineered wet-lab reference samples, in-silico sequence data file samples, and digital images of tissue section slides, and it developed an evaluation methodology to assess the analytical performance of validated laboratory-developed tests (LDTs) relative to an FDA-approved companion diagnostic (CDx) for a targeted cancer therapy. During the summit, pilot leadership and others contended that well-defined reference samples and in-silico files could offer an enhanced quality assurance (QA) approach that could complement existing processes such as proficiency testing.[2] Many participants agreed there is value in understanding how performance compares across laboratories. Some stakeholders, however, emphasized the complexities and limitations of the pilot's approach. That said, even the pilot's skeptics recognized its value as a starting point to improve understanding about laboratory test validation and QA as large-scale, complex next-generation sequencing (NGS) panels become more prevalent.

This *Summary of Themes* provides further synthesis of the December virtual summit. *Please see the Diagnostic Quality Assurance Pilot website for additional background and details on the pilot. For a list of summit participants, please see the appendix on page 12.*

Stakeholders have diverse interpretations of the pilot's outcomes

The summit aimed to understand whether the community agreed on the pilot's findings, recognizing that the pilot itself was small scale. Participants debated the pilot's design and conclusions throughout the three-day summit. The pilot data have not yet been published but were presented to participants to inform the basis of summit discussions. Summit conversations demonstrated that the community has diverse perspectives on interpretation of the pilot's data, though some noted the pilot's approach could lay the groundwork for future pragmatic approaches to QA.

The pilot was designed to determine whether the reference samples and in-silico files outlined by the Scientific and Technical Working Group could provide a diverse group of laboratories with an opportunity to demonstrate the performance of their tests relative to a CDx. The candidate CDx used for comparison was a two-gene, multiple-variant NGS panel—the Praxis Extended RAS Panel—voluntarily proposed by biopharmaceutical developer Amgen and Amgen's CDx partner, Illumina. Praxis helps identify patients with colorectal cancer who are eligible for treatment with Vectibix (panitumumab), which, at the time of the pilot's launch, was undergoing FDA review for a new indication.[3] Pilot leadership aimed to develop the samples prior to FDA approval of Praxis and Vectibix so that they could be available for postmarket utilization by laboratories, should the pilot's approach be expanded.

Proof-of-concept data verified that the reference samples and in-silico files worked as planned. Pilot leadership distributed the first round of wet-lab samples in December 2018 and received results from laboratories in March 2019. Pilot data indicated that the reference samples and in-silico files worked as intended. As an incidental finding, most participating laboratories' performance met that of the CDx, although the pilot observed variability by some laboratories. The latter point prompted robust discussion among participants.

Across conversations about the pilot, some members of the pathology community dismissed the variability the pilot demonstrated, expressing concerns that the pilot was too difficult and, in short, *"intended to force labs to failure."* Leadership of the pilot vigorously countered this point, emphasizing that such an outcome was not the intention of the group and that concerns about the pilot's level of difficulty did not arise until after the pilot's data was revealed. Others emphasized that most laboratories demonstrated strong performance, which suggests the bar for the pilot was not set too high. One laboratory expert said, *"To insist [the project intended to force labs to failure] is to ignore the labs that did well."* Another, underscoring that the pilot occurred at a specific point in time in a technology landscape that is constantly changing, said, *"It's almost too easy to criticize the pilot's design in hindsight, and that would just be unfair."*

The following specific areas of disagreement featured prominently in summit discussions.

Clinical implications of the variability observed in the pilot on the patient population

Participants engaged in debate over whether the variability around laboratories' lower limit of detection (LOD) that the pilot identified was clinically meaningful. The pilot assessed, among other factors, performance around laboratories' LOD as defined by the CDx, which was set at 5% variant allele frequency (VAF) or "the percentage of sequence reads observed matching a specific DNA variant divided by the overall coverage at that locus."[4] Some participants agreed with the observation of one pathologist who said that the variants involved in the pilot *"were unusual variants seen at very low variant allele fractions."* Therefore, some concluded that the impact on the patient population would be limited.

> *"If these variants are placed on a specific CDx but there has only been five or fewer of such variants ever detected in clinic, how relevant is it to clinical practice?"*
>
> —Payer

> *"There is a premise to these questions that I think is concerning, which is basically to suggest that just because these variants are rare or because we tested at the limit of detection of 5%—which, again, was because it was pegged to the companion—that somehow miscalling these is okay."*
>
> —Subject matter expert

Others disagreed, highlighting that, rare or not, the variants included in the pilot were clinically meaningfully because of their inclusion in the CDx kit. *"These variants were selected based on the companion. These were not somehow picked to exploit weaknesses in bioinformatic pipelines or anything like that. They were picked because that's what was actually on the Praxis panel and were related to the selection of Vectibix,"* one participant said. Others noted that, on the whole, patients with variants at low VAFs may not be exceptionally rare, with one participant citing that *"studies show that between 10% and 15% of clinical samples for many actionable targets may be in the 5%–10% VAF range,"* depending on whether the patient has primary or recurrent disease.

Preanalytical and operational issues involved in targeting low VAFs

Some participants from the pathology community noted other issues related to the inclusion of low VAFs in the pilot. Many laboratories, some noted, do not report on specimens with low VAFs as a general policy. Pilot leadership underscored that only laboratories that incorrectly reported variants that were present above their stated LODs were cited as having made an error.

Additionally, some pathologists expressed concern about the inclusion of variants at laboratories' stated LODs because of preanalytical complexities. VAF is, several stakeholders emphasized, inextricably linked with tumor cellularity, or the amount of tumor cells in the specimen and their arrangement into clusters.[5] Assessments of cellularity can vary extensively in clinical practice and require pathologists' analysis of several factors upon examining a specimen, as summit participants reported. Therefore, some emphasized that because claims about a test's ability to detect certain VAFs cannot be isolated from a specific specimen's cellularity, the failure of laboratories to appropriately detect or report low VAFs near their LOD in engineered reference

> *"We do put allelic frequencies out there very, very low. We go down to 5%, 2.5%, and we've even put out 0.1% allelic frequency. And again, it comes back to giving the labs the opportunity to interrogate their data to see whether or not their assay would pick it up, regardless of whether or not they would report it clinically."*
>
> —European quality assurance laboratory expert

samples should not be afforded much weight.[6] A pathologist said, *"Clinical significance of VAF is entirely dependent on the variant and the percentage of tumor cellularity. By itself, VAF has little meaning."* Others acknowledged this complexity, highlighting that metrics that seamlessly integrate tumor cellularity and VAF remain elusive, but they urged that the community not *"make the perfect the enemy of the good."*

One stakeholder noted that *"many laboratories claim 5%"* VAF, and thus it is important to assess how those claims compare with practice, at least to the extent that practice can be mimicked through use of reference samples. Furthermore, others emphasized that visual assessments of tumor content and reporting on VAFs are already included in existing QA methods, and some European stakeholders noted that their QA assessments routinely test low VAFs for certain types of assays, including NGS.

The role of the CDx as the gold standard

The pilot aimed to assess LDT performance characteristics based on the specification of the FDA-approved Praxis Extended RAS CDx. Some participants emphasized challenges with doing so. One questioned the value of using the CDx as a performance standard given the rapid pace of innovation: *"In light of thinking of what we want, which is the right information for proper patient management, a CDx may or may not align with that. The CDx goal may be, as stated by a pharmaceutical company, to know whether that specific drug may be useful for that patient—but that might not be a broad enough diagnostic test to give the physician all of the information that they need, particularly in light of the technologies that we have at our disposal."*

Others disagreed, emphasizing the pragmatism of comparing laboratory performance to the CDx. An industry representative underscored, *"I certainly don't believe that the in vitro diagnostic is the gold standard, but it is the standard. Right or wrong, it's what was used to determine the clinical validation for that analyte paired with that drug. Is that perfect? As I just stated, absolutely not, but it's what's available."*

Finally, some stakeholders from the laboratory community did not question the value of using the CDx as the standard but instead called for enhanced dissemination of CDx performance characteristics for laboratories to more readily duplicate. Other stakeholders contested this point, indicating that information on performance characteristics is available—*"performance characteristics of FDA-approved companion diagnostics are always available on the FDA website, including the test's package insert,"* one said—which signals a need for greater clarity on the resources and appropriate level of detail available to laboratories.

Participants proposed several high-level concepts for improving similar studies and QA efforts

The pilot's implementation did not occur in a vacuum: the technology landscape for diagnostics has evolved rapidly in recent years, especially in oncology. Data collected by the American Society of Clinical Oncology in 2017 indicated that 75% of oncologists used NGS test results to inform patient-care decisions.[7] Against this backdrop and despite an absence of consensus about the pilot's outcomes, participants discussed how the pilot, as an initial effort, could help inform or shape the future of molecular diagnostic QA.

Future research studies or pilots in this space should focus on clinical reporting and LOD complexities

Participants debated how the community could gather more or better data to help understand potential QA vulnerabilities. Future studies like the pilot should, some urged, focus on laboratory reporting practices, given that laboratory reports may communicate important claims, complexities, and/or test limitations. External QA programs in Europe have evaluated and scored the content of clinical reports for many years as an accompaniment to the technical performance demonstration.[8] Mock clinical reports were omitted from the pilot study, which pilot leadership and others acknowledged as a limitation. Some stakeholders cautioned that assessing clinical reports lies at the *"border of technical issues and practice of medicine";* however, several continued to advocate for closer understanding of laboratory reports. A regulatory representative emphasized their importance: *"We look very carefully at the reports, what they say, what the limitations are. And now FDA has very clear recommendations for what can be in a 510(k) report for an assay. What do you report? If there are certain variants or amplicons, depending on how the assay is set up, that you repeatedly in your validation cannot detect, then that just needs to be communicated in the report."* Similarly, for some stakeholders, future studies that assess how laboratories address LODs in their reports is of particular interest and may help inform future good practices. *"I would definitely push for a way to capture whether a response was below a lab's LOD—did they see it and just did not report it? So, I think that clearly needs to be added in some capacity."* The pilot provided this information in part through laboratories' responses in the pilot's data collection form, but some participants believe this issue warrants more detailed clarification and evaluation in the future.

> *"There are so many variables that we have to consider. And that professional practice piece of it—that is just as critical as the technical aspect. That's where having that clinical report, knowing how that pathologist, that molecular diagnostics expert can communicate, is super critical."*
>
> —Pathologist

The pilot's in-silico process could have merit for QA in the short term

The in-silico process was prioritized by several stakeholders as a valuable component of the pilot. It could be employed to inform QA efforts today or to educate laboratories about potential bioinformatics vulnerabilities. The promise of the pilot's in-silico process could be further realized, some emphasized, if operational bottlenecks for laboratories, especially the importation of in-silico mutagenized sequence files into vendor-supplied bioinformatics pipelines as is discussed below, were resolved.

The pilot took a novel approach to assessing laboratories' bioinformatics pipelines. One pilot leader explained, *"We have not seen studies until this one that have paired samples of wet-lab and in-silico data files that are on the same variants and assay, tested by the labs at the same time."* Pilot leadership underscored the value of the customized in-silico processes implemented as part of the pilot. Customized in-silico files, when compared with engineered wet-lab samples, are lower cost, flexible, and able to test a wide variety of variants. Furthermore, the in-silico process proved valuable in helping to isolate the root cause of some laboratories' difficulties.

> *"One of the best use cases for these data is to use it to arm the consumers—in this case, laboratories who are shopping for assays, for kits, for bioinformatic pipelines to understand that the thing that they're purchasing can detect the variants that it seeks or states that it can detect. It would be good for providers to have the pipelines tested in advance of them purchasing them to understand the limitations of what these tests can and cannot do. So that's one way that this could be used moving forward."*
>
> —Payer

Pilot leadership and other stakeholders also discussed several caveats to in-silico performance assessments. First, they would optimally be offered in parallel to a wet lab/reference sample–based challenge. Second, laboratories encountered several operational challenges for which they were *"really not prepared."* In particular, laboratories faced logistical challenges in introducing external files into vendor-supplied bioinformatics pipelines that were not linked to a corresponding clinical sample that had been assessed internally. One subject matter expert said, *"There is a bottleneck in laboratories understanding how to insert these files, especially laboratories that may not be in an academic or tertiary-care medical center. There are a lot of laboratories that don't necessarily have that expertise, and anything that we could do to help them use these files would go a long way in positioning labs to develop highly validated tests for higher quality."*

Finally, some participants recommended that platform manufacturers should play a role in identifying and sharing work-around procedures to facilitate a more seamless process for laboratory staff. One expert cautioned that *"there's always going to be a challenge by introducing an electronic signature of a variant into a data file, versus having those variants in the samples"*; the expert noted, however, that *"it's not insurmountable"* and suggested further conversations to advance this concept.

Reference samples have value, but how to incentivize and sustain their utilization remains an open question

Despite divergent viewpoints in the pilot's outcomes, most stakeholders agreed that engineered reference samples have a role to play in enhancing test validation and/or ongoing QA. However, some emphasized that more salient questions lie in understanding how to use reference samples in a way that is timely, meaningful, and cost effective for the community. Judging from the tenor of summit discussions, answering these questions in a consensus-based fashion may take time. One stakeholder said, *"To me, the single biggest challenge here is not, Can you use reference samples? Rather, it's, How do you do it in a way that generates sustainability and timeliness?"* Participants addressed several issues to resolve for development and implementation of reference samples and in-silico data files to enhance QA on a large-scale basis.

> *"Having standardized reference materials is absolutely the right way to go."*
>
> —Pathologist

Both engineered reference samples and patient specimens have a role to play, but the community lacks consensus on the optimal use cases for each

Broadly, many participants recognized that relying solely on patient specimens is not a scalable or sustainable solution. Some, however, emphasized that specific types of reference samples should be used for specific types of QA activities. A laboratory representative said, *"One thing that would be good to differentiate is the goals of providing either digital or wet-lab specimens and whether these would be intended for use primarily for proficiency testing or for test validation and development, because the kind of specimens and the variant frequencies that we would require for those two different scenarios could be very different."*

> *"We had a lot of discussion [in our breakout group] about NTRK [neurotrophic tyrosine receptor kinase fusions]. And there are going to be those variants that are even less frequent, and it's just very difficult for labs to get samples to validate their tests."*
>
> —Industry representative

Some pathologists specifically highlighted the preference for patient specimens, especially for initial test validation. If test validation is primarily conducted on engineered samples, laboratory systems can develop biases for detecting variants in reference samples that they may miss in real clinical specimens, some participants argued. Others disagreed, underscoring that reference samples can be helpful during validation to understand a test's performance with less common variants. One payer noted, *"When you're getting samples to validate a test, SNVs [single nucleotide variants] are everywhere—so it's no surprise that labs have no difficulty in finding them, whereas indels are more difficult to find and, especially if you don't do a thorough validation, you may not even be sampling them as a part of your validation."*

It remains unclear what institutions should define gold standards and develop reference materials

Some stakeholders underscored that the value of the pilot's approach lay in introducing an external standard that can allow laboratories to assess how their assays compare with one another. However, who sets that external standard and who pays for reference samples and in-silico file development remains a topic of debate. Participants offered several considerations:

- **Institutions with appropriate expertise should lead.** Some stakeholders continued to reiterate the need for third-party institutions with a core business relating to standards development to focus on crafting *"universally accepted"* QA standards and materials.
- **Regulators could help define and develop reference samples.** Some recommended an enhanced role for regulatory bodies in defining standards and even directly developing reference samples to compare performance across assays. Some participants pointed to the FDA's development of a well-characterized reference panel during the COVID-19 crisis, which enabled laboratories to compare the analytical performance of their COVID-19 assays, as a precedent that could be expanded.[9]
- **Pharmaceutical manufacturers may not support reference-sample development.** Summit participants discussed the role of pharmaceutical manufacturers in future enhanced QA approaches. In the pilot's methodology, a pharmaceutical manufacturer underwrote the cost of reference-sample development as part of a premarket development process. Industry representatives indicated that in the future, doing so would be challenging. Pathways for developing highly regulated products are already complex, and reference-sample development remains outside of the industry's core busines. An industry representative said, *"I don't know that we, as individual companies, are in the best position to do it successfully."* That said, industry players echoed support for ongoing involvement in similar efforts and noted that an industry role would be best suited as part of a broader multistakeholder consortium.

> *"In-silico samples are a useful means of identifying or being able to test a platform's ability to detect certain kinds of variants, particularly more difficult to come by variants."*
>
> —Payer

The community could consider incentives for investing in enhanced QA

Recognizing that there is always value in improving quality, participants at the summit discussed methods for incentivizing QA strategies, including those that involve reference samples. Specifically, for some participants, one of the principal challenges is how to encourage the use of reference samples on a sustainable basis. Other participants noted that efforts to incentivize quality must be carefully balanced with the need to increase access to biomarker testing, which is currently underutilized, to prevent any negative impact for patients in need of testing.

Some acknowledged that laboratories may not have the resources to invest accordingly in reference samples unless validation and ongoing QA processes are more systematically required and standardized by relevant stakeholders. The significant cost of these samples is a factor the community needs to consider. One subject matter expert said, *"Horizon, SeraCare, and other groups, they could make beautiful controls. A problem comes in an LDT laboratory space where people are going to think about money—and well-manufactured GMP [Good Manufacturing Practice] control material is expensive. And so people are still going to err on the side of using their own favorite. I think the pressure needs to be on the payers to demand that some sort of validation is given such that assays are testing the relevant genes."*

As noted in the comment above, some stakeholders considered whether payers have a more prominent role to play in encouraging test validation and assurance processes. For some payers, a laboratory *"is considered as a manufacturer of a service or test and still has to abide by all the same rules to demonstrate that the service that's being performed is a quality service and results are accurate."* One described his institution's specific experience: *"Upon reviewing validations, we're refusing to reimburse 50% of these tests because we don't think that the lab performance characteristics are good enough."* Some acknowledged the role of Palmetto GBA's MolDx Program as a leader in laboratory-related coverage policies, given that MolDx requires detailed validation data from laboratories for tests to receive coverage. However, participants acknowledged that some payers may not be sufficiently educated in laboratory related topics to be able to assess for quality and others emphasized the laboratory community should lead on QA-related issues.

More broadly, lack of consistency and standardization in testing can create undue burden on patients and inefficiencies across the healthcare system. Specifically, a payer noted that some cancer institutions require re-testing of patients within their own laboratories once a patient enters the system.

Additional lessons from the pilot can support enhanced QA for the technology landscape of tomorrow

In addition to conversations that considered short- and mid-term application of the pilot's outcomes, summit participants discussed the pilot's long-term relevance to an ever-changing molecular diagnostics landscape. More gene sequencing test panels are available to survey hundreds of genes and inform oncology treatment, including FDA-approved companion diagnostics, such as Foundation Medicine's FoundationOne CDx.[10] During the summit, participants discussed the importance of the pilot's QA approach for a two-gene panel, considering the advancement of larger panels since the pilot's launch. One payer described the present moment as a period of transition from *"an analyte-specific view of the world, where we have to look at specific analytes—one gene, one mutation for one drug—"* to a stage where *"we can look at literally everything all at once."*

The discussion among participants moved to acknowledge that as therapies increasingly target rare biomarkers and as the size of panels increases, the advancement of laboratory competency in detecting a range of mutations—including complex and rare ones—is of utmost importance. Thus, most emphasized the need for pragmatic approaches. Specifically, several participants called for a QA paradigm that focuses on assessing performance for classes of variants such as *"single nucleotide variants, indels, structural variants, copy number variations, etc."* Participants noted that such an approach underscores that the community cannot test for every single possible variant; rather, it would be a starting point that, over time, could be augmented by real-world evidence generated by laboratories that could contribute to the growing body of data linking variants to therapeutic approaches.

Conclusion

Individual pathologists and pathology professional organizations have diverse interpretations of the pilot's outcomes and their generalizability. However, many stakeholders that participated in the summit agree that reference samples that enable enhanced test performance analysis and comparison are valuable. Several also agreed that bioinformatics pipelines and their potential vulnerabilities need greater attention. Some want to learn more from stakeholders outside the United States who frequently implement external QA approaches that are not dissimilar to the pilot's approach.

Despite the divergent viewpoints, all stakeholders acknowledged that science is changing rapidly and performance of tests for detecting rare, complex mutations will become increasingly important. As the science advances, some noted that quality assurance processes will continue to evolve in parallel. As this evolution occurs, participants agreed on the ongoing importance of patient-centricity in diagnostics and testing. One industry participant said, *"If we think in terms of patients, what really gets us to the bottom line is, Can we produce high-quality tests despite the evolution of the technologies?"*

Pilot leadership will consider comments from all stakeholders participating in the summit as it revises a technical manuscript detailing the pilot's methodology and data and a white paper discussing lessons learned and remaining questions for the community to consider.

About this document

This *Summary of Themes* reflects the use of a modified version of the Chatham House Rule whereby comments are not attributed to individuals, corporations, or institutions. Italicized quotations reflect comments made by participants before and during the meeting.

Tapestry Networks is a privately held professional-services firm. Its mission is to advance society's ability to govern and lead across the borders of sector, geography, and constituency. To do this, Tapestry forms multistakeholder collaborations that embrace the public and private sector, as well as civil society. The participants in these initiatives are leaders drawn from key stakeholder organizations who realize the status quo is neither desirable nor sustainable and are seeking a goal that transcends their own interests and benefits everyone. Tapestry has used this approach to address critical and complex challenges in corporate governance, financial services, and healthcare.

Appendix: Participants

- **Amgen:**
  - Megan Doyle, Global Regulatory and R&D Policy Lead, Diagnostics, Digital Health, and Combination Products
  - Greg Friberg, VP & Therapeutic Area Head, Global Development
  - Molly Martell, Global Lead, Diagnostic Payer Strategy
  - Dave Stanforth, Director, Clinical Biomarkers and Diagnostics, Head of Diagnostics Strategy and Development, Amgen
- **American Society of Clinical Oncology:** Tom Oliver, Director, Clinical Practice Guidelines Division
- **Association for Molecular Pathology:**
  - Mary Williams, Executive Director
  - Robyn Temple-Smolkin, Director, Clinical & Scientific Affairs
- **Blue Cross and Blue Shield Association:** Naomi Aronson, Executive Director, Clinical Evaluation, Innovation and Policy
- **Centers for Disease Control and Prevention:** Lisa Kalman, Health Scientist
- **Centers for Medicare and Medicaid Services:**
  - Joseph Chin, Deputy Director, Coverage and Analysis Group
  - Sarah Harding, Policy Analyst, Director Division of Ambulatory Services, Center for Medicare
  - Penny Keller, Division of Laboratory Services, Center for Clinical Standards and Quality, CLIA
  - Faye Valcarel, Center of Clinical Standards and Quality, Survey & Certification Group, Division of Laboratory Services, CLIA
  - Amy Zale, CLIA Policy Branch B, Branch Manager
- **College of American Pathologists:**
  - Dara Aisner, Associate Professor, Pathology, University of Colorado School of Medicine, CAP Genomic Medicine Resource Committee

  - Neal Lindeman, Vice Chair for Molecular Pathology at the Brigham and Women's Hospital, Associate Professor of Pathology, Harvard Medical School, Vice Chair, CAP Molecular Oncology Committee
  - Christina Lockwood, Associate Professor and Director of the Genetics and Solid Tumor Diagnostics Laboratory at the University of Washington, Medical Director of the Brotman Baty Precision Medicine Institute, and Clinical Director of the Northwest Genomics Center, CAP Molecular Oncology Committee
  - Eric Konnick, Pathologist, Seattle Cancer Care Alliance and University of Washington Medical Center, Assistant Professor of Laboratory Medicine (UW), Vice Chair, CAP Genomic Medicine Resource Committee
  - Patty Vasalos, Technical Director, Scientific Resources, CAP
- **Emory University School of Medicine:** Barbara Zehnbauer, Adjunct Professor of Pathology
- **European Medicines Agency:** Markus Paulmichl, Vice Chair, EMA Pharmacogenomics Working Party
- **eviCore:** Lon Castle, CMO, Laboratory and Specialty Drug Services
- **Food and Drug Administration:**
  - Steven Lemery, Acting Division Director, Division of Oncology 3, CDER
  - Michael Pacanowski, Associate Director, Genomics and Targeted Therapy, Office of Clinical Pharmacology, Office of Translational Sciences, CDER
  - Wendy Rubenstein, Director, Personalized Medicine
  - Julie A. Schneider, Regulatory Scientist, Office of Hematology and Oncology Products, CDER
  - Zivana Tezak, Associate Director for Science and Technology, Office of In Vitro Diagnostics and Radiological Health, CDRH
- **Frederick National Laboratory for Cancer Research/FNIH:** Mickey Williams, Director, Molecular Characterization Laboratory
- **Freenome:** Girish Putcha, Chief Medical Officer
- **Friends of Cancer Research:** Jeff Allen, President and CEO
- **Genentech-Roche:**
  - Katia Basset, Principal CDx Project Leader
  - Danelle Miller, VP, Global Regulatory Policy & Intelligence, Roche Diagnostics

    - Eric Peters, Director and Head, CDx
- **Genomics Quality Assessment:** Sandi Deans, Consultant Clinical Scientists and Director; also NHS England, National Laboratory & Scientific Lead (Genomics)
- **Gilead:**
    - Terrell Baptiste, Senior Manager, Sr. Manager Regulatory Policy and Intelligence
    - Neville Mehenti, Senior Director, Global Commercial Product Strategy
    - Scott Patterson, VP, Biomarker Sciences
- **Horizon Discovery:**
    - Keith Cannon, Director, Commercial Product Management, Diagnostics
    - Jennifer Keynton, Manager R&D Diagnostics
- **Humana:** Bryan Loy, Physician Lead, Oncology, Laboratory, and Personalized Medicine
- **Illumina:**
    - Phil Febbo, SVP and CMO
    - Karen Gutekunst, Vice President of Diagnostic Development
- **LabCorp:** Anjen Chenn, Discipline Director, Molecular Oncology
- **Len Lichtenfeld**, former Deputy Chief Medical Officer, the American Cancer Society
- **LUNGevity Foundation:** Andrea Ferris, President and CEO
- **Massachusetts General Hospital:** Keith Flaherty, Director of the Henri and Belinda Termeer Center for Targeted Therapy
- **National Cancer Institute:**
    - Tracy Lively, Chief, Diagnostics Evaluation Branch, and Deputy Associate Director, Cancer Diagnosis Program, Division of Cancer Treatment and Diagnosis
    - Lisa Meier McShane, Associate Director, Division of Cancer Treatment & Diagnosis, and Chief, Biometric Research Program
- **National Institute of Standards and Technology:** Justin Zook, Human Genomics Team Leader
- **NorthShore University Health System:** Karen Kaul, Chair of Pathology and Laboratory Medicine
- **New York State Department of Health:** Erasmus Schneider, Associate Director for Research and Technology, Wadsworth Center

- **Optum Genomics:** Jill Hagenkord, CMO
- **Pacific Business Group on Health:** Emma Hoo, Director, Pay for Value
- **Palmetto GBA:** Gabriel Bien-Willner, Medical Director and CMO, MolDx
- **Thermo Fisher:**
  - Garret Hampton, President, Clinical Sequencing and Oncology
  - Kelli Tanzella, Senior Director, Global Regulatory Affairs, Clinical, & Compliance
- **UnitedHealthcare:** Jennifer Malin, Senior Medical Director, Oncology & Genetics
- **University of Chicago Medicine:** Blase Polite, Associate Professor of Medicine, Deputy Section Chief for Clinical Operations, and Executive Medical Director for Cancer Accountable Care
- **University of Michigan Comprehensive Cancer Center:** Daniel F. Hayes, Stuart B. Padnos Professor of Breast Cancer Research
- **Washington University School of Medicine:** John Pfeifer, Vice Chair for Clinical Affairs, Pathology and Immunology and former Quality Pilot Scientific Technical Working Group Chair

Endnotes

[1] For additional details on the pilot, please see "Diagnostic Quality Assurance pilot," Tapestry Networks, accessed January 28, 2021.

[2] For a definition of proficiency testing and more information on its processes, please see Centers for Medicare and Medicaid Services, *Proficiency Testing and PT Referral Dos and Don'ts* (CMS: September 2017), 2.

[3] Vectibix and Praxis were approved by the FDA in 2017. "FDA Approves Vectibix® (Panitumumab) for Use in Wild-Type RAS Metastatic Colorectal Cancer," Amgen, news release, June 29, 2017.

[4] Samuel P. Strom, "Current Practices and Guidelines for Clinical Next-Generation Sequencing Oncology Testing," *Cancer Biology & Medicine* 13, no. 1 (March 2016), 5. Note that in the clinical trial for the Extended RAS panel and Vecitibix, clinical response was seen in patients with as low as 2.6% VAF.

[5] Irina Heid, et al., "Co-clinical Assessment of Tumor Cellularity in Pancreatic Cancer," *Clinical Cancer Research* 23, no. 6 (2017), 1462.

[6] Some stakeholders noted following the meeting that another critical challenge is that the manner in which laboratories determine their LODs can vary and is not transparent to external stakeholders. A participant said, *"Labs need to provide greater clarity about what their LOD VAF truly means. Is 5% VAF detected by mixing 5% variant DNA with 95% normal DNA? Or is 5% VAF detected from a starting sample of 50% or greater tumor-cell content? 30% tumor? 10% tumor?"*

[7] Andrew N. Freeman, et al., "Use of Next-Generation Sequencing Tests to Guide Cancer Treatment: Results From a Nationally Representative Survey of Oncologists in the United States," *JCO Precision Oncology* (November 13, 2018), 1.

[8] Lien Tembuyser, et al., "Higher Quality of Molecular Testing, an Unfulfilled Priority: Results from External Quality Assessment for KRAS Mutation Testing in Colorectal Cancer," *Journal of Molecular Diagnostics* 16, no. 3 (2014), 371-377; Susan D. Richman, et al., "*RAS* screening in colorectal cancer: a comprehensive analysis of the results from the UK NEQAS colorectal cancer external quality assurance schemes (2009–2016)," *Virchows Arch.* 471, no. 6 (2017), 721–729; Ellen Bellon, et al., "External quality assessment for KRAS testing is needed: setup of a European program and report of the first joined regional quality assessment rounds," *The Oncologist* 16, no. 4 (2011),467-478.

[9] "SARS-CoV-2 Reference Panel Comparative Data," US Food and Drug Administration, accessed January 28, 2021.

[10] "Genomic Profiling Tests Cleared by FDA Can Help Guide Cancer Treatment, Clinical Trial Enrollment," National Cancer Institute, December 21, 2017.