



**Green Park Collaborative**  
*A partnership for innovation and effectiveness*

## **Demonstrating the Clinical Utility of Next Generation Sequencing in Clinical Oncology**

### **Meeting Summary**

July 7, 2014 – World Trade Center, Baltimore, MD



The Green Park Collaborative is a major initiative of the  
Center for Medical Technology Policy

**November 7, 2014 | Version 1.0**

BETTER EVIDENCE.  
BETTER DECISIONS.  
BETTER HEALTH.

Center for Medical Technology Policy  
World Trade Center Baltimore  
401 E Pratt Street, Suite 631  
Baltimore, Maryland 21202

410.547.CMTP (P)  
410.547.5088 (F)  
info@cmtpNet.org (E)  
@cmtp\_baltimore (Tw)

[www.cmtpNet.org](http://www.cmtpNet.org)

## **TABLE OF CONTENTS**

<b>Introduction</b> .....	<b>1</b>
<b>Opening Session: Welcome and Overview</b> .....	<b>1</b>
<b>Session 1: Opportunities and Challenges for NGS as an Emerging Clinical Tool</b> .....	<b>3</b>
Ms. Tamara Syrek Jensen CMS and Evolving Policies on NGS for Clinical Diagnostics.....	3
Key Discussion Points .....	6
Dr. Dane Dickson Next Generation Sequencing: “A Scientist, Clinician, and Payer’s Perspective” .....	6
Key Discussion Points .....	11
Dr. John Pfeifer Opportunities and Challenges for NGS as an Emerging Clinical Tool: Sources of Confusion from the Clinician’s Perspective .....	11
Key Discussion Points and Questions.....	14
Evidence issues:.....	15
Other issues and questions:.....	15
<b>Session 2: Current Evidence and Uses of NGS</b> .....	<b>16</b>
Dr. Keyur Patel MD Anderson’s Mixed Reimbursement/Research Model for Using NGS.....	16
Key Questions.....	19
Dr. John Pfeifer Washington University’s Reimbursement-Driven Model .....	20
Key Questions and Discussion.....	22
Dr. Gary Palmer Foundation Medicine: Commercial Development of NGS.....	23
Key Discussion Points .....	27
<b>Session 3: Future Evidence Frameworks</b> .....	<b>28</b>
Dr. David Litwack FDA: Challenges and a Potential New Approach.....	28
Key Discussion Points .....	30
Dr. Jeff Allen A Collaborative Approach to Clinical Trials .....	30
Dr. Bruce Quinn Strategic Approaches to Clinical Utility .....	32
<b>Ending Thoughts and Continuing Dialogue</b> .....	<b>35</b>

**List of Tables**

**Table 1.** Palmetto MoDX Clinical Trial Determination (mCTD) Evidence Schema ..... 10  
**Table 2.** Company-Recommended Reasonable Uses of FoundationOne® Testing ..... 27

**List of Figures**

**Figure 1.** Level of Evidence Associated with Change in Standard of Care, NSCLC Example ..... 9  
**Figure 2.** MD Anderson MTEC: Process for Reviewing and Adopting New Candidate Biomarkers . 18  
**Figure 3.** Sources of First 3,936 Samples Analyzed by FoundationOne® ..... 24  
**Figure 4.** Genetic Alterations across Tumor Types as Detected by FoundationOne® ..... 26  
**Figure 5.** Lung-MAP “Master Protocol” Schema ..... 31  
**Figure 6.** Potential Gains in Validity and Utility from New Tests ..... 33  
**Figure 7.** The Six Questions Illustrated in Graphic Form ..... 34  
**Figure 8.** Three Axes of Business Models for Diagnostic Tests ..... 35

**Appendices**

**Appendix A – Workshop Agenda** ..... A-1  
**Appendix B – Workshop Participants** ..... B-1

## **INTRODUCTION**

The Center for Medical Technology Policy (CMTP), through its multi-stakeholder forum called the Green Park Collaborative – USA (GPC-USA), held a meeting in Baltimore on July 7, 2014 as a first step to developing recommendations for the design of studies to evaluate the clinical utility of next generation sequencing (NGS)-based testing in oncology. This work will be accomplished in coordination with the Roundtable on Consensus Standards for Multiplex Cancer Genomic Testing, a workgroup hosted by the American Society of Clinical Oncology (ASCO), The Association for Molecular Pathology (AMP), and the College of American Pathologists (CAP). Through GPC-USA, the Roundtable stakeholders interested in consensus standards will be joined by representatives from health plans, CMS, NGS test developers, FDA, patient groups, oncology guideline organizations, clinical oncologists, drug/biologic companies, and other groups with a direct interest in NGS. Summarized in an Effectiveness Guidance Document (EGD), these standards will help to shape the generation of evidence needed to address the information needs of payers, health technology assessment organizations, clinical guideline developers, as well as patients and clinicians.

The ultimate aim of this nine month project is to enable evidence-based decision making regarding coverage and reimbursement for NGS-based tests in oncology thereby assuring patients access to high-quality information useful for their care. Its recommendations will be aligned with existing and emerging regulatory guidance where relevant. This effort will be informed by CMTP's recent work on evidentiary standards for studies of the clinical validity and clinical utility of molecular diagnostics in oncology.

## **OPENING SESSION: WELCOME AND OVERVIEW**

In the introductory session of the meeting, the agenda for the day was introduced by Donna Messner, PhD, CMTP Vice President and Senior Research Director, and Leader of the GPC-USA Oncology Consortium. Dr. Messner explained that this meeting was intended to serve as a kick-off for a planned effort to develop an EGD in coordination with the Roundtable on Multiplex Testing. An expert technical working group representing relevant expertise across multiple stakeholder perspectives will be formed to meet periodically over approximately a nine month period to develop recommendations for the type of evidence needed to demonstrate clinical utility. The purpose of this meeting, therefore, was to hear perspectives on key points of concern for further discussion in working group deliberations, and to identify areas of possible common ground that might be achieved between stakeholders on key evidence questions.

Following this overview, Dr. Messner introduced Sean Tunis, MD, MSc, President and CEO of CMTP and Chair of the GPC-USA Steering Committee, to provide his perspective and insights on the mission set before the group to find common ground on clinical utility. Dr. Tunis noted that, by design, the group in attendance was highly diverse. The hope was that stakeholders having very different backgrounds and points of view would be open to hearing differing perspectives on the challenges of NGS and learning from one another. He encouraged participants to actively engage in discussion throughout the day.

Dr. Tunis reiterated that one important goal of the meeting was to work towards a shared understanding of the evidence reasonable to expect for demonstrating clinical utility of next generation sequencing platforms. He noted the “different ends of spectrum” of views on NGS. On the one hand, some observers tout the promise of NGS, saying for example that this technology is “poised to revolutionize biomedical research and usher in a new era of individualized rational medicine.”<sup>1</sup> Other observers meanwhile point out that so far this revolution in individualized rational medicine has very little evidence-based support. In 2009, the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group wrote, “Test applications are being proposed and marketed based on descriptive evidence and pathophysiologic reasoning, often lacking well-designed clinical trials or observational studies to establish validity and utility, but advocated by industry and patient interest groups.”<sup>2</sup> Moreover in 2008, the Secretary’s Advisory Committee on Genetics, Health, and Society (SACGHS) wrote that “Information on clinical utility is critical for managing patients, developing professional guidelines, and making coverage decisions” and recommended that the U.S. Department of Health and Human Services “create a public private entity of stakeholders to . . . establish evidentiary standards and levels of certainty required for different situations.”<sup>3</sup> To Dr. Tunis’ knowledge, or that of anyone attending this meeting, this type of multi-stakeholder entity was never convened — thus remains to be accomplished.

Crucially, Dr. Tunis said, a tension exists between the level of certainty we would like to achieve for evidence of clinical utility—the rigor, quality, and quantity of scientific evidence for demonstrating well validated, clinically and personally meaningful benefits for patients—and the rapidity with which we would like to obtain access to new technologies, including both potentially beneficial access for patients and also investments for rapid development of new technologies and robust economic benefits. The decision of where to set the bar for standards of evidence will have far reaching implications. Yet this is not solely a scientific judgment. It is also a social judgment about the level of scientific uncertainty we are willing to accept to balance these multiple competing socially desirable goals. Technical expertise will not provide the ultimate answer to the question of evidence standards because the answer cannot be derived solely through technical means.

Given that this assessment is both social and scientific, a multi-stakeholder process is critical but also poses challenges. Dr. Tunis noted that among the many “amazingly talented, experienced, and smart” people of diverse background in attendance, many mutually incompatible solutions can be confidently asserted as the “right” way to go. He urged the group members to have open minds and a willingness to revise their initial thinking on the issue throughout the day—and throughout the following months, as

---

<sup>1</sup> Multiple similar quotations can be found. This one came from Yegnasubramanian, S. Explanatory chapter: Next generation sequencing. *S. Methods Enzymol.* 2013; 529: 201–208. doi: 10.1016/B978-0-12-418687-3.00016-1.

<sup>2</sup> Teutsch SM, Bradley LA, Palomaki GE, Haddow JE, Piper M, Calonge N, Dotson WD, Douglas WP, and Berg AO, Chair, on behalf of the EGAPP Working Group. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) initiative: methods of the EGAPP Working Group. *Genetics in Medicine* (2009) 11, 3–14; doi:10.1097/GIM.0b013e318184137c.

<sup>3</sup> SACGHS (Secretary’s Advisory Committee on Genetics, Health, and Society). US system of oversight of genetic testing: A response to the charge of the Secretary of Health and Human Services. Bethesda, MD: Department of Health and Human Services; 2008.

the working group delves into these questions more deeply and looks to this group for continued engagement.

One participant underscored the challenge of the multi-stakeholder process, commenting that he served on the SACGHS committee calling for a private entity of stakeholders to establish the standards and levels of certainty for differing situations. The participant noted that the committee's membership constantly negotiated that tension between robust evidence and rapid technological development without ever coming to a resolution to balance the two goals. There was, he said, "a constant ebb and flow of ideas at that time." The difficulty of finding the proper balance was clear.

## **SESSION 1: OPPORTUNITIES AND CHALLENGES FOR NGS AS AN EMERGING CLINICAL TOOL**

Dr. Messner introduced the session, explaining that while in fact most people share some sense of the potential opportunity for NGS, it is the challenges where most discussions founder. Accordingly, the opening discussion is intended to focus constructive discussion around the differing perspectives and concerns of payers versus those of clinicians in specialist centers who are eager to integrate NGS-based diagnostics into their practices as quickly as feasible. Tamara Syrek Jensen, JD, Director of the Coverage and Analysis Group, Centers for Medicare and Medicaid Services (CMS), would give an overview of Medicare's current policies regarding diagnostic testing and the discussion CMS would like to have around evidence for these technologies. Dane Dixon, MD, Director of Clinical Science, MoIDx, Palmetto GBA, will then provide his view of key concerns for developing evidence for NGS-based tests. Dr. Dixon was kind enough to share in advance the points he planned to raise with the third speaker, John Pfeifer, MD, PhD, Vice Chairman for Clinical Affairs, Pathology and Immunology, Professor of Pathology and Immunology, and of Obstetrics and Gynecology, Washington University School of Medicine, St. Louis, who would provide a response and alternative perspective.

To guide the discussion, one participant suggested adopting the following terminological conventions:

- Analytical validity: How accurately and reliably does the test measure the analyte(s) of interest?
- Clinical validity: How accurately and reliably does the test "measure" the clinical endpoint(s) of interest?
- Clinical utility: Does use of the test to guide medical decision-making consistent with its indication(s) for use in its intended use population improve patient outcomes or provide health economic benefits in a statistically significant manner?

### **MS. TAMARA SYREK JENSEN**

#### **CMS AND EVOLVING POLICIES ON NGS FOR CLINICAL DIAGNOSTICS**

Under Medicare's framework for covering diagnostic and screening tests, said Tamara Syrek Jensen, payment for diagnostic testing can be obtained in one of three ways:

- 1) Claim adjudication, where an applicable claim code exists
- 2) Local coverage determination (LCD) through a local contractor
- 3) National coverage determination (NCD)

While coverage for diagnostic tests can also be accomplished at the national level through rule-making (a formal rule-and-comment process carried out through publication in the Codes of Federal Regulation, or CFR), Medicare prefers the NCD process because it is more rapid and can be made effective on the date of posting.

Unlike molecular diagnostic tests, by law, screening tests cannot be covered through an LCD or claim adjudication, but must be evaluated under an NCD. Unlike private insurers which must cover screening tests having a Grade A or B rating from the U.S. Preventative Task Force, Medicare must consider the following criteria: first, whether the test is appropriate for the Medicare population; and then, whether it is reasonable or necessary for the prevention or early detection of an illness or disability.

Under the federal regulations, diagnostic tests can only be covered by Medicare if they are to be used in the management of a specific medical problem (42 CFR 410.32). Such tests must be reasonable and necessary for the diagnosis or treatment of illness or injury or to improve the functioning of a malformed body member. The assessment of “reasonable and necessary” can be extremely challenging. The term is not defined in statute. Two previous unsuccessful attempts have been made to define it via rulemaking (in 1989 and 2000). As a result, CMS has operationalized “reasonable and necessary” to mean “adequate evidence to conclude that the item or service improves clinically meaningful health outcomes for the Medicare population.”

In answer to a question from a participant, Ms. Jensen said that health outcomes are the “ultimate outcome” for CMS, as opposed to changes in patient management, but she added that health outcomes can mean many things. Ultimately, she said, her interpretation of Social Security Act 1862(a)(1), which prohibits coverage for items and services not “reasonable and necessary for the diagnosis or treatment of illness or injury or to improve the functioning of a malformed body member,” is that the outcomes have to be clinically meaningful.

Over 19,000 tests have been registered in the National Institutes of Health’s (NIH) Genetic Testing Registry for 4,500 conditions,<sup>4</sup> and some significant proportion of these will be potentially relevant to the Medicare population. Under the current system, especially with the volume of genetic and genomic tests being developed, it has not been clear what CMS has been paying for and whether it is beneficial to patients. Complicating matters is the fact that few of these products are FDA-approved. FDA review provides information on clinical validity that would not necessarily be available for laboratory-developed tests (LDTs) not reviewed by the FDA. Accordingly, CMS has an added measure of confidence for tests undergoing FDA-CMS parallel review, since CMS is able to see the FDA’s rigorous review hear the thoughts of reviewers. For LDTs, by contrast, CMS has to consider not merely clinical utility (beneficial outcomes for the patient), but also clinical validity (the correlation of the test with a specific phenotype).

Increasingly for NCDs, CMS is concerned with questions such as: Does the test add anything substantively new to the clinical tools available? What are the advantages and disadvantages to making

---

<sup>4</sup> See <https://www.ncbi.nlm.nih.gov/gtr/>

an NCD decision in either direction? What impact would it have on the need for additional testing or treatment? What impact on quality-of-life? (see Box 1)

In coverage with evidence development decisions, particularly, quality-of-life outcomes are typically needed. This can mean many different things depending on clinical context and cannot be the sole outcome considered, but CMS does consider quality-of-life outcomes. In addition, she said, NCD assessments are not made specifically in relation to the standard of care; comparative effectiveness cannot be the sole basis for the decision, but the test does need to demonstrate clinically relevant benefits to patients.

A new challenge comes with the Protecting Access to Medicare Act (PAMA), which was signed into law on April 1, 2013. PAMA changes the payment paradigm for what it calls “advanced diagnostic laboratory tests.” One provision requires that all of these tests go through the LCD process but does not define what that process is. The currently existing LCD process is not defined anywhere in statute; it is simply a manual with instructions from CMS to local contractors. So CMS has issued a notice asking for public comment on a proposed new process for LCDs (see 79 FR 133, July 11, 2014)—one designed for more efficient handling of the increasing volume of diagnostic tests to be considered under the LCD process. The proposed process preserves public comment and appeals, while making optional the previously required carrier advisory committees and open meetings.

Another notable feature of PAMA is that the Secretary may designate between one and four Medicare administrative contractors (MACs) to establish coverage and payment processes. If Medicare used only one MAC, it would essentially constitute a national coverage decision. However most observers express a preference for the involvement of more MACs, not fewer. One meeting participant commented that involvement of multiple MACs will mean inconsistency in the decisions and policies made. In response, Ms. Jensen noted that if Congress did not want local coverage determinations, the provisions would not be included in the statute. It can be advantageous to limit a new product with a less mature track record to a local decision rather than allowing it to come to a national level immediately.

Ms. Jensen closed by stressing that an important goal for her as Director of the Coverage and Analysis Group would be to work out the best approach to molecular diagnostic tests. She indicated that discussion and information-sharing with the NIH (which generates a great deal of evidence for tests) and FDA could be part of the solution. However, the first priority is to establish the LCD process under which CMS policy decisions will take place (this step is underway through notice and comment, as noted earlier). The second step is then making the process more predictable. For predictability to be

### Box 1. Example Diagnostic Questions for NCDs

- What data does it generate?
- What clinical information do the data provide?
- (How) Have we gotten this information before (from other sources)?
- What (different) decisions are made based on this information?
- What are the advantages and disadvantages of basing decisions on this new test?



achieved, consistent standards of evidence have to be applied. The challenge is in achieving some common ground on where that evidence bar should be set.

### **Key Discussion Points**

A participant noted that many different types of LDTs exist. A laboratory may make a minor adaptation of an FDA-approved kit (e.g., replacing the type of deionized water used) or a laboratory may create its own procedure for identifying a particular genomic signature. Both could be considered LDTs. What differences in evidence might be required for each situation?

In answer to this question, Ms. Jensen said that specific nuances of LDTs will need to be taken into account. It is a question Jeffrey Roche has started to review.<sup>5</sup> However currently most tests do not undergo FDA review and must be taken on a case-by-case basis.

Another participant comments that “there are many ways to catch a BRAF,” meaning many valid approaches one could use in a laboratory to identify this marker, and improvements are made periodically. “We change the test every year or so based on new information and on what the tumor board is saying has clinical utility for patients. We are a small academic laboratory and can’t go to FDA for a 510(k) every time we make a change. These tests are rapidly changing.”

Finally, a comment was made that under PAMA, the Healthcare Common Procedure Coding System (HCPCS) could potentially be used to add fields beyond pricing information, and could conceivably be used for measuring test outcomes. Ms. Jansen noted that because CMS is in the middle of rule-making, she could not comment on this specific issue.<sup>6</sup>

## **DR. DANE DICKSON**

### **NEXT GENERATION SEQUENCING: “A SCIENTIST, CLINICIAN, AND PAYER’S PERSPECTIVE”**

In one sense, NGS represents a paradigm shift in analytical technology. Yet in another sense, said Dr. Dane Dickson, it is not unique because it is only one among many significant advances in technology we have seen in the past or are likely to see in the future. So while this is an important advance, to the extent possible, we need to develop approaches that are transferrable to new technologies in the future.

A key challenge with NGS currently amounts to a lack of standardization. Dr. Dixon illustrated the point with outdated units of measure: the cubit (the distance from the elbow to the tip of the middle finger) and the league (the distance someone can walk in an hour). Measurements of this type can be repeated by the same person to achieve the same results, but results will vary from person to person. Dr. Dickson said that NGS-based tests have analogous variability due to lack of standardization, including differences in laboratory sample handling, analytical technologies and capabilities, and informatics software. Before Medicare can consider the utility of the information coming from these platforms, these types of

---

<sup>5</sup> Jeffrey Roche, M.D., M.P.H., Medical Officer, Coverage and Analysis Group, Centers for Medicare and Medicaid Services.

<sup>6</sup> See <http://www.cms.gov/Medicare/Coding/MedHCPCSGenInfo/index.html?redirect=/medhpcpsgeninfo/>

## Box 2. Levels of Evidence

### +++++ Randomized Controlled Trial

- Defined patients
- Defined treatment
- Predefined outcomes
- Control group
- Low levels of associated bias.

### +++ Prospective Interventional Trial

- Defined patients
- Defined treatment
- Predefined outcomes
- Medium levels of associated bias

### + Retrospective Observational Trial

analytic differences have to be reconciled to assure that results from one laboratory will be comparable to those from another.

When one considers clinical utility, an additional challenge is that not all mutations in a specific gene are the same in terms of effect on phenotype, and multiple genes of unknown relative importance may be involved. NGS platforms may identify mutations or variants not identified by an approved companion diagnostic test for the same indication, providing extra information of unknown significance. Dr. Dickson suggested that in some cases the detected mutation may actually represent the composition of only a small fraction of tissue sample taken, so is not representative of the overall composition of the tumor (not a “dominant clone”) and thus is not necessarily informative regarding patient management. These differences in analytical methods could result in situations where, for example, a targeted therapy is used on a patient when another therapy would have been more appropriate simply because the method of detection was more sensitive than the method associated with the FDA-approved companion diagnostic test.

An additional question is the detection of variants traditionally associated with one indication in a new context. For example, does a HER2 mutation imply the same therapy in gastric cancer that it does in breast cancer? Dr. Dickson pointed out other gaps in the knowledge of the science, including the role of epigenetic factors and the changing genetic composition of tumors over time.

In summary, Dr. Dickson said that until standardization across laboratories can be achieved, tests from individual laboratories have to be considered individually for assessment of utility.

### ***Evidence Needed for Coverage***

To discuss the evidence needed from his perspective, Dr. Dickson began with a general hierarchy of the types of studies potentially used in developing clinical evidence, assigning ratings based on the quality of the resulting data (see Box 2). The “five plus” level represents evidence from randomized controlled trials, which is “wonderful data” that has been used to set the standard of care for decades. At the “three plus” level is what he termed “prospective international trials,” in which patients are prospectively identified, treated, and followed for outcomes, with comparison to historical controls. At the “one plus” level were retrospective observational trials. According to Dr. Dickson, after having

reviewed hundreds of trials, he has concluded that retrospective studies “rarely” provide adequate information on the patients treated and the interventions used.

Finally, while case reports can be potentially informative, there is a tendency to draw inferences improperly from individual case reports. Dr. Dickson noted a phenomenon he termed “standard of care migration” in which care patterns established on the basis of high quality evidence can become modified on the basis of lower quality evidence and nevertheless be accepted as if the new care pattern were associated with the same level of certainty.

A useful framework for assessing clinical utility can be found in a 1991 paper by Fryback and Thornbury. In it, the authors define an evidence hierarchy for efficacy based not on the study design, but on the type of information generated (see Box 3).<sup>7</sup> In this adaptation of the framework to diagnostic test applications,<sup>8</sup> the lower levels refer to test validation parameters (Level 2) and the ability of the test to change a clinician’s diagnostic thinking about the patient (Level 3). Level 4 is achieved if the change in diagnostic thinking results in a demonstrated change in patient management. Level 5 is achieved if the change in patient management results in a measurable positive impact on patient outcomes.

Dr. Dickson noted that CMS used this framework to consider positron emission tomography (PET) scanning for dementia and neurodegenerative disease.<sup>9</sup> Responding to comments in the Decision Memo (comments saying that evidence of “improved health outcomes” should not be a factor for a coverage determination on amyloid PET), CMS wrote that they “*generally consider the evidence in the hierarchical*

### Box 3. Levels of Diagnostic Efficacy (Fryback & Thornbury)

#### Level 2

Diagnostic accuracy, sensitivity, and specificity of the test

#### Level 3

Does the information produce a change in the physician's diagnostic thinking?

#### Level 4

Does the information produce an effect on the patient management plan?

#### Level 5

What is the effect of the information on patient outcomes?

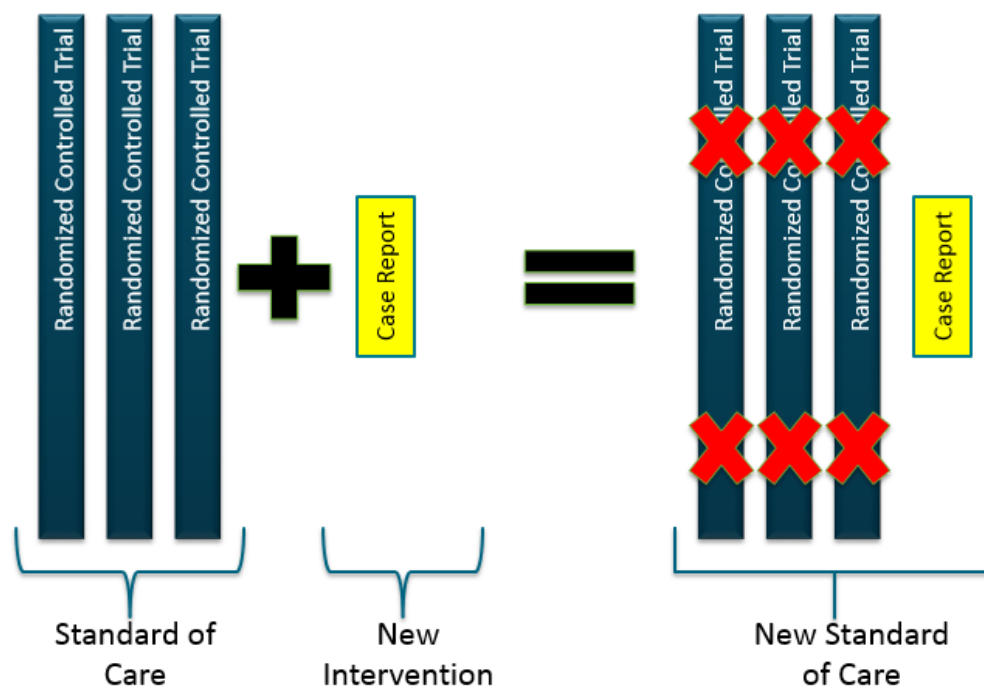
<sup>7</sup> DG Fryback and JR Thornbury. The Efficacy of Diagnostic Imaging. *Med Decis Making* 1991; 11(2): 88-94. This paper was designed to describe levels of efficacy of diagnostic imaging. In the original paper, the levels were conceived as cumulative, with lower levels being necessary but not sufficient to assure efficacy in the higher levels. In the adaptation to other diagnostic tests, levels 1 and 6 are dropped. Level 1 refers to the technical quality of the images and level 6 refers to societal costs and benefits of the technology.

<sup>8</sup> The paper describes levels of efficacy of diagnostic imaging. In the original paper, the levels were conceived as cumulative, with lower levels being necessary but not sufficient to assure efficacy in the higher levels. In the adaptation to other diagnostic tests, levels 1 and 6 are dropped. Level 1 refers to the technical quality of the images and level 6 refers to societal costs and benefits of the technology.

<sup>9</sup> Decision Memo for Beta Amyloid Positron Emission Tomography in Dementia and Neurodegenerative Disease (CAG-00431N). <http://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=265> (accessed July 28, 2014).

*framework of Fryback and Thornbury (1991)” and have “generally found evidence of efficacy at Level 5 more persuasive to support unconditional coverage. We believe that coverage supported by that level or higher evidence results in the greatest benefit for Medicare beneficiaries.”*

Many laboratories developing diagnostics will object that this higher level of evidence is more difficult to achieve than the lower levels, and that it is not feasible to insist on this standard for diagnostics. Nevertheless, Dr. Dickson asked, what level of evidence should be needed to change the standard of care? He described a meeting of a “major group” in which clinicians in attendance made consensus recommendations to address non-EGFR and non-ALK mutations in non-small cell lung cancer (NSCLC) examined by NGS. The group recommended that NSCLC patients with *RET*-positive lung cancer should be treated with crizotinib (even in first line, before platinum-based chemotherapy), and that HER2-positive lung cancer should be treated with trastuzimab (even in first line before platinum-based chemotherapy). Upon investigation, Dr. Dixon found that the level of evidence actually existing to support the recommendation for crizotinib was a case report of three individuals; evidence supporting treatment by trastuzimab was a case report of one individual. These recommendations effectively ignore previously collected high-quality evidence and supplant it with a few case reports to set a new standard of care (see Figure 1). This type of decision-making represents a “leap of faith” driven perhaps by a perception of the technology as “cutting edge.” Even so, and despite the lack of analytic standardization between laboratories, this group recommended that payers cover the drug when given under these circumstances. Although leaps of faith may need to be taken in some instances, “collecting good data along the way” is crucial to verify the validity of that leap.



**Figure 1. Level of Evidence Associated with Change in Standard of Care, NSCLC Example**

The MoIDX program at Palmetto is currently considering strength of evidence on a hierarchy in which the highest levels of evidence (level 3) comes from prospective controlled trials and prospective-retrospective trials based on previously archived tissue samples (see Table 1-A). Intermediate levels of evidence in this hierarchy (level 2) involve prospective observational studies using the test in the care plan and complex data modeling. Retrospective observational studies represent the lowest level of clinical evidence (level 1) while the lowest overall level of evidence (level 0) refers to preclinical studies. Under the Palmetto framework, the body of existing evidence is reviewed and categorized according to this hierarchy. To be considered for traditional approval, the strongest trial is expected to be a prospective design at mCTD level 3A, 3B or 2A with supporting studies at levels 3 or 2 (prospective trial or data modeling). Retrospective studies and preclinical designs fall out of the decision tree as non-supportive for approval. The reasoning behind this framework is that these are the only study designs that can adequately address the types of questions that need to be asked for clinical utility.

**Table 1. Palmetto MoIDX Clinical Trial Determination (mCTD) Evidence Schema**

**mCTD 3A - Prospective Controlled Trial (PCT)** directly addressing the molecular test as the actionable item leading to significant improvement compared to a current accepted standard of care. End points of the trial must be something widely considered as being significant by the respected medical community (e.g. overall survival). The trial must be adequately powered to address the outcome of the intervention based on the test.

**mCTD 3B - Prospective-Retrospective Trial (PRT).** Previously reported prospective trial using archived sample, looking at how a given molecular test can be shown to improve outcomes in a very specific patient population based on the results from the original trial. Often only samplings of patients from the original trial are evaluated. The samples must be well defined as to associated patient characteristics and treatments so as to adequately determine exactly what type of patients benefit from the given test.

**mCTD 2A - Prospective Observational Study (POS)** where patients are prospectively enrolled in a registry, and then treated according to a defined pathway using the molecular test as an integral part of the care plan.

**mCTD 2B - Data Modeling (RDM)** complex data modeling to determine risk-benefit of a given test using large data sets to estimate impact of a given molecular test on the standard of care approach.

**mCTD 1 - Retrospective Observational Study (ROS)** where there no stipulation of treatment or follow up based on the molecular study.

**mCTD 0 - Preclinical Studies (PS)** Preclinical data or related studies or trials.

### **Palmetto Initiatives Going Forward**

Finally, to develop the MoIDx program further and enhance its ability to address NGS-based testing, Dr. Dickson noted that the following initiatives are underway:

- 1) Completed a draft check list of criteria for standardizing NGS that is currently being circulated among academic centers and industry leaders for feedback and comment
- 2) Building transparency and consistency around the science by establishing a national academic advisory panel to MoIDX and standardizing its policies
- 3) Working towards the ability to implement coverage with evidence development (CED) at the local level through MAC for products with early but compelling clinical evidence
- 4) Developing a framework for creating a registry in which cancer patients can be molecularly screened and either treated by standard care, enrolled in a clinical trial, or treated according to an identified molecular pathway, as appropriate, with clinical outcomes data captured in the registry. The initiative, called MED-C, is in the concept stage and would be coordinated with academic medical centers.

### Key Discussion Points

- Should levels of acceptable evidence vary according to clinical context? What about the example of a HER2 active mutation in a lung cancer patient with no other treatment options? Should the threshold of evidence be somewhat lower in this case than in instances where well established treatment options exist?
- While there are differences between platforms, in the published, peer-reviewed literature, there is an overlap of between 0.85 and 0.95 in the 3 million or so variants identified by different platforms using different library preparations. That figure is actually better than it is for radiology.
- Whether we can say that one lab's mutation finding is comparable to the next lab's for the same sample comes down to a matter of proficiency testing under CAP and holding laboratories to the high standards set by CAP.

### DR. JOHN PFEIFER

#### OPPORTUNITIES AND CHALLENGES FOR NGS AS AN EMERGING CLINICAL TOOL:

##### SOURCES OF CONFUSION FROM THE CLINICIAN'S PERSPECTIVE

When an oncologist or pediatrician orders a test, the cost of the test is incurred by the clinical laboratory performing it. As a result, said Dr. John Pfeifer, the clinical laboratory staff need to have clarity on the evidence needed for testing so that they can confer with clinical colleagues as to whether a test is warranted and provide advice on whether it will be reimbursed. To help achieve that clarity, some common misperceptions need to be addressed.

The first point of clarification to be made is that **NGS is a method, not a test**. Any number of different methods can be used to examine genetic material, including traditional cytogenetics, interphase FISH (fluorescence in situ hybridization), microarray testing, Sanger sequencing, and NGS. The question is: what is the diagnostic, predictive, or prognostic information that the clinician wants? Depending on the information desired, the question is which method is best or most appropriate to run the test that provides the needed information.

It also helps to clarify the circumstances under which NGS would typically be used. Many people point to the plummeting cost of doing whole genome sequencing – from a high of \$100 million to make a genome to a cost now maybe a few thousand dollars. But in the clinical setting, whole genome

sequencing is not typically done; the goal is to examine more limited sequences of DNA in genetic regions that have predictive or prognostic utility. Currently, NGS for sequencing in clinical laboratories is only cost effective for examining larger quantities of DNA (in Dr. Pfifer's laboratory, that threshold falls at about 2.5 kilobases). Beneath this threshold, it is more cost effective to use Sanger sequencing.

Dr. Pfeifer offered a real-life example to illustrate some of the confusion around this distinction between tests and methods (see Case Example 1).

---

*Case Example 1: The patient is a man in his 70s with a history of head and neck squamous cell cancer who has undergone surgical therapy and now has recurrent disease. The oncologist orders "[Lab] testing to guide therapy." The payer denies coverage because it is not clear what test is ordered and how the results will impact therapy. The payer also cites its policy: "The Genetic Cancer Susceptibility Panel using next generation sequencing are [sic] considered experimental, investigational and unproven."*

---

What are the confusions in this case?

- 1) The oncologist's order is vague and does not, for example, indicate whether the patient has comorbidities that could limit potential targeted treatments and thus would not need certain targets to be tested. Is the test for guiding office therapy or to enroll the patient in a clinical trial? It is not clear how the test will actually impact the care of the patient.
- 2) The payer's policy relates to a susceptibility panel, which does not apply here; the patient does not have a genetic susceptibility, he has recurrent cancer.
- 3) The payer's policy indicates that coverage was denied based on the test's methodology rather than what the test was designed to do. This is a novel basis for denying coverage! What kind of evidence is required to demonstrate that NGS methods are as good as or superior to other methods?

*The context determines what kind of test is being done using a given method.* For example, a diagnostic test requiring sequencing of 10 genes might be done by NGS simply because it is less expensive than Sanger sequencing. In terms of testing purpose, the two methods are interchangeable.

In predictive testing, a distinction needs to be made between:

- 1) analysis of a set of genes in a tumor type for which there is a medical literature demonstrating that therapy based on variants improves outcomes;
- 2) analysis of a set of genes for which there is a medical literature demonstrating that therapy based on variants improved outcomes in *another* tumor type;
- 3) analysis of a set of genes for which there is, as of now, no evidence-based literature demonstrating that therapy based on variants improves outcomes.

In the latter case, the issue is not whether NGS or Sanger is used for the sequencing of the genes, but the fact that the literature does not provide evidence-based support for decision-making based on the result.

A second source of confusion is **technical standards**. There are different levels of methodological rigor even among College of American Pathologists (CAP)-certified labs. Dr. Pfeifer said, for example, that his laboratory will not perform testing on samples having less than a certain quantity of DNA material because the analytic sensitivity and specificity of the assay “absolutely falls apart.” Nevertheless, a competing lab is specifically marketing its capability to run assays on samples that Dr. Pfeifer’s lab rejects due to under-threshold quantities of DNA material. What sensitivity and specificity does the competing laboratory achieve when they run these samples? What should be the threshold of analytic sensitivity and specificity to say that a test is valid?

Labs should have to provide this information to payers, and should also have to provide them with information about the capabilities of their test to find all four major classes of variants: single-nucleotide variants (SNVs), insertions and deletions (indels), copy number variants (CNVs), and structural variations (SVs). Some labs have NGS-based tests that find only subsets of mutations, offer the tests at a low price, and then use other types of tests (e.g., FISH) to fill in the gaps and increase their reimbursement. A lab with a more comprehensive (and expensive) NGS test would actually save overall costs because ancillary testing would not be necessary. Yet a lab doing all-NGS testing in this manner would receive only the lower reimbursement level.

**Test interpretation** is another area of uncertainty. It is a goal that every lab in the U.S. would get exactly the same test answer and interpret the result the same way regardless of where the testing was done and what specimen type was used. The significance of a given mutation, however, may vary with different types of cancer and sometimes be unknown. How do labs interpret the significance of the mutation and how can those interpretations be standardized?

Panels present another challenge. There is precedent that payers will reimburse for panels (generally) but they seem to be reluctant when it comes to NGS panels. Consider an analogous situation. In clinical chemistry testing there is a basic seven-analyte metabolic panel that is routinely reimbursed. Yet, by design, the chemistry instrument is capable of producing data for many additional analytes. The reagent cost of these additional analytes is minimal. Yet labs do not routinely suggest to payers that the additional testing be done. Similarly, with an immunoassay machine that tests thyroid stimulating hormone, labs do not argue that for only \$7 more the patient could be screened for diabetes and have baseline values established for two biologic markers. There is an understanding between the payers and the labs that test panels are necessary but there will be no payment for other tests that may also have been useful. Similarly in the context of NGS, more information can be produced than is directly ordered, and some of this information may have utility. However, payers worry that larger panels will find numerous incidental results that cost a lot of money to investigate and may turn out to be irrelevant.

Finally, there is confusion about the level of evidence needed to show utility. In discussions of evidence, there is a tendency to talk as if the same level of evidence should apply to all testing,



regardless of whether the purpose of the test is diagnosis, prediction, or prognosis. A test for an inherited disease, for example (constitutional testing), is profoundly different from a test on cancer, and the evidence required for the tests should reflect those differences. The greatest problem, however, is the logical paradox of personalized medicine: the more our ability to individualize diagnosis and treatment, the greater the difficulty of substantiating these diagnoses and treatments with evidence.

Dr. Pfeifer illustrated this paradox with a real-life example (see Case Example 2):

---

*Case Example 2: A patient with a very rare cancer, thymic carcinoma of squamous morphology, was found to have a c-KIT mutation. Had the c-KIT mutation appeared in a patient with a gastrointestinal stromal tumor, the drug imatinib would be indicated. However, in this very rare squamous cell carcinoma, imatinib had not been tested. When the payer was approached with the suggestion that imatinib should be tried in this case, he requested evidence in the form of a prospective randomized double-blind trial that imatinib would work in this situation.*

---

The confusion here is that a randomized trial to study this rare tumor type, rare morphology, and probably also rare mutation would take an unfeasibly long time. As Dr. Pfeifer put it, “our progeny can give the answer.”

In fact, given that each patient has a unique genetic background and no two tumors are exactly the same, Dr. Pfeifer noted that the next logical question is: “For that matter, what level of evidence do we have for *any* drug in *any* patient?” This is the crux of the paradox. How do we know that a drug will work in any particular cancer patient absent clinical trials performed on hundreds of clones with the same cancer? We have a lot of population-based data collected from people with largely unknown genetic backgrounds. How relevant are those data really to the next person who walks through the door with colon cancer? The net quandary arising from this paradox is: what is the appropriate evidence level to support reimbursement in many of these situations and what can we do for patients until this is sorted out?

### **Key Discussion Points and Questions**

How can information on rare cases and treatment responses be captured so it becomes part of the evidence? It is difficult to judge how often targeted therapies for rare tumors are successful. It is possible that only the successfully treated cases get publicized while tumors that do not respond to targeted therapy do not receive the same attention.

- Data-sharing among academic institutions would be helpful, or through national-level organization might allow for accumulation of meaningful evidence.
- One barrier to data-sharing is that academic institutions have proprietary interests in their tests and bioinformatics pipelines that they are reluctant to give away. Also, even gathering this data

within a single institution is difficult due to institutional review board requirements and other factors.

- Another barrier to data-sharing is that data from different sources are not necessarily comparable; not all assays identify all four major classes of variants, for example, and different tests may examine different parts of the coding sequence. Until criteria are set so that all labs are performing the same test, it will be difficult to combine information into a data set that has much utility.
- One participant noted that ASCO has proposed a prospective registry to capture data off-label use of targeted therapies that is driven or informed by the results of a molecular diagnostic test.
- The answer will depend on the characteristics of the test performed; more comprehensive tests will probably find more potentially actionable variants than less comprehensive ones, for example. It will also depend on the variety of tumors being tested; a laboratory testing only lung cancer samples will have a different answer than one taking all comers.
- One participant referred to the 80/20 rule (the idea that for many events, 80% of the effects come from 20% of the causes) to suggest using two simple questions for a quick evaluation of different centers' testing programs: (1) of the last 100 large gene panels you gave for cancer somatic mutation, how many patients received a targeted therapy that was based on the test, and (2) how many of those patients had some significant clinical response?

**Evidence issues:**

- When a case of successful treatment in a rare cancer is recorded, what does that mean for a larger body of evidence?
- How does one distinguish between situations in which it is most appropriate to go ahead and do studies and situations so rare that studies are not worth pursuing?
- Anatomic site and morphology are both important in diagnosing patients (a squamous cell carcinoma of the rectum is not the same as a squamous cell carcinoma, which is different from that of the lung). Molecular information must be considered in connection with these two other factors. Therefore, drug trials for cancer agents need to start being based on genetic information in addition to tumor type and morphology
- Ultimately we may need to look not just at individual mutations driving disease, but a spectrum of mutations

**Other issues and questions:**

- While some sophisticated payers are attending this meeting, payers are generally not as sophisticated about these issues as laboratories are. An evidence guidance document that incorporates the perspectives of payers, clinicians, and patients about what sort of information is minimally sufficient for making judgments about clinical utility would help in bridging the knowledge gap. Better convergence around what constitutes adequate evidence of clinical benefit to patients is also a prerequisite to coordination among the academic community.
- Until evidence is developed, what guidelines should be used for compassionate care in the interim?

- NGS has been extensively hyped and patient expectations are unrealistically high despite the lack of evidence. We are in the “irrational exuberance” phase of new technologies, and will have to pass through a “collapse” (disappointment that comes with the recognition that the technology is not a magic bullet) before we reach “clinical acceptance” and understand exactly how NGS fits best into medicine.
- Clinicians know the questions they need help with. They do not necessarily know, however, the clinical utility of various tests. MRIs to evaluate back pain, for example, are one of the biggest areas of unnecessary imaging and their clinical utility has never been demonstrated. Similarly, coronary CT angiography has not been shown to lead to any sort of improved outcome. Showing that a test changes clinician behavior does not settle the issue of the test’s utility.

## **SESSION 2: CURRENT EVIDENCE AND USES OF NGS**

The second session of the meeting involved presentations from people representing three different organizations that are using NGS in a clinical setting: two academic centers and a private (now public) company. First, Dr. Keyur Patel, Medical Director of the Molecular Diagnostics Laboratory (MDL) at the University of Texas M.D. Anderson Cancer Center (MDACC) explained how their institution created a system for molecular testing and adapted it when NGS was developed. Their model for funding focuses on being able to provide data to payers to support reimbursement as well as relying on institutional resources and grants to develop investigational biomarkers. Dr. John Pfeifer returned to the podium to present the Washington University School of Medicine’s approach, which concentrates on maximizing the potential for reimbursement and seeks program self-sufficiency without reliance on institutional or grant funds. Finally, Dr. Gary Palmer from Foundation Medicine, Inc. in Cambridge, MA, presented commercial model which he indicated was centered on maximizing patient benefit rather than reimbursement. The experiences of these three organizations illustrate how the policies and expectations of payers and accessibility of alternative funding sources significantly shape the development and use of NGS testing for clinical purposes.

### **DR. KEYUR PATEL**

#### **MD ANDERSON’S MIXED REIMBURSEMENT/RESEARCH MODEL FOR USING NGS**

Dr. Patel opened with some historical background. MDACC recognized in 2008 that it needed to provide a framework for clinical care and clinical trials that were based on personalized, precision, or individualized biomarker-directed therapy. At that time, they developed a Clinical Cancer Genomics Initiative to develop a framework for both clinical trials and clinical care coordinating the activities around myriad specimens representing various tumor types, sample types, collection sites within institutes, methods for analysis, and laboratories doing the analyses.

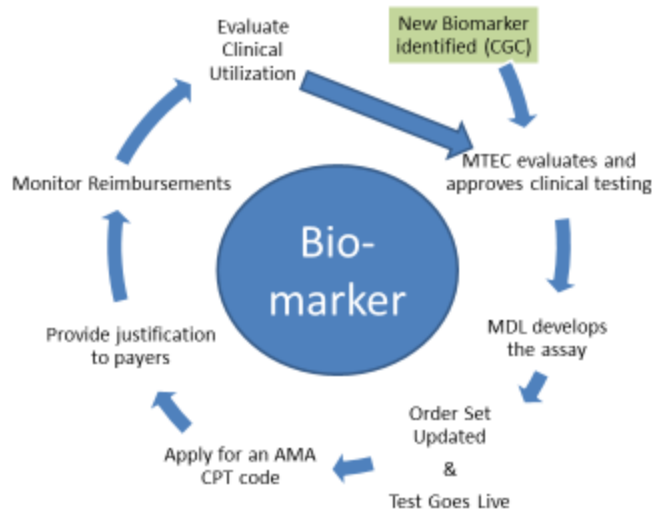
At this time (in 2008) a rapid increase in molecular testing took place, particularly for solid tumors. The demand for testing began to overtake the clinical laboratory capacity. NGS had not yet been developed and the limited available tests were not highly multiplex-capable. The rapid growth of molecular testing without internal regulation gave rise to concerns about billing practices and patient charges and ultimately resulted in unreimbursed laboratory costs, Dr. Patel explained. In response to this situation,

in 2010 MDACC established its multi-disciplinary Molecular Testing Evaluation Committee (MTEC) to address these and other concerns. The charges to the MTEC were extensive:

- *Define criteria and establish processes* for determining that a CLIA-compliant molecular diagnostics test is considered “*standard of care*” in a specific clinical setting at MDACC
- If a test is not considered “*standard of care,*” determine whether it is of sufficient scientific and clinical interest to merit *investment of institutional funds to develop clinical data* to achieve that status
- Determine whether to maintain a specific biomarker on the *roster of services for routine clinical testing* and *vet organ-site specific electronic order entry sets* of molecular diagnostics tests
- *Develop documentation* for use by administrators as evidence of standard of care to use in negotiations with payers
- *Monitor reports* on documentation of medical necessity for, billing compliance for, and utilization of molecular diagnostics tests by physicians
- *Review outcomes and clinical effectiveness* studies to provide input for the lab medicine roster of services

Currently, candidate new biomarkers potentially undergo three levels of review. Tumor-specific committees of clinicians meet to discuss evidence for panels or mutations relevant for their specific tumor of interest. Their conclusions are then passed to an intra-divisional Clinical Genomics Council and the institutional MTEC for further review (see Figure 2). Through these processes, mutations have been organized into disease-specific order sets for physicians to use when ordering molecular diagnostic testing services. So, for example, one set of biomarkers would be made available to clinicians diagnosing stomach and esophageal cancers while another set of options for ordering biomarker testing would be available for intestinal and colorectal cancers. Both sets of options would be based on what the MTEC considered to be clinically actionable and standard-of-care for MDACC. In response to a question, Dr. Patel clarified that the inclusion of a biomarker test on an order set does not mean that payers have agreed to reimburse for the test. Inclusion means that test utilization is considered appropriate in their setting and an evidence base exists to provide to payers.

## Clinical Biomarker Life-Cycle



**Figure 2. MD Anderson MTEC: Process for Reviewing and Adopting New Candidate Biomarkers**

Although initially MDACC relied on Sanger sequencing, pyrosequencing, or high resolution melting screening followed by Sanger confirmation, they have now moved to NGS panel testing because they believe it provides advantages for physicians, patients, the laboratories, and the institution (Box 4). It improved his lab's efficiency drastically, consolidating many processes and tests, reducing costs, streamlining work flows, and allowing the use of smaller amounts of tissue samples. Over the last three to four years, their use of single-plex testing has dwindled to almost nothing and the vast majority of their testing is now performed using NGS panels, often commercially available ones.

To focus on clinical and patient-care issues raised by NGS, MDACC established a multidisciplinary Molecular Tumor Board to bring together oncologists, laboratories, and clinical cancer geneticists, among others. The Board discusses topics such as how to approach incidental findings, understanding the implications of test results for clinical care, and provides a mechanism for educating the laboratories about clinical issues associated with test utilization.

Thus far, the laboratory has run more than 5000 solid tumor samples by NGS. Through the MTEC process, approximately 30% to 40% of these tests were judged as not meeting the standard-of-care criteria but were deemed appropriate for clinical trials and research. Accordingly, these tests were not sent to payers and instead were paid for by the MDACC's Institute for Personalized Cancer Therapy.

At the same time, Dr. Patel noted that they have been working to improve the panels they use for testing, and are moving towards the creation of custom panels. For example, the majority of the 53 genes included in their commercial hematologic “hot spot” panel did not show any mutations in more than 1,000 clinical samples. So they developed a custom hematologic panel testing only 28 genes examining entire coding sequences (not merely hot spots). In the future they will develop more cost-efficient high-yield panels with specific panels for early/known tumors and a common all-inclusive panel for advanced/rare tumors.

The Institute is also funding a clinical trial to investigate a comprehensive cancer panel having 400 genes for solid tumors and hematologic malignancies. This is being done since, while the order sets contain genes for common tumor types, a cancer center needs to be able to address uncommon tumors where biomarkers are less well studied or advanced refractory metastatic disease where the patient has failed multiple therapies and is looking for other options.

#### Key Questions

A participant asked Dr. Patel what definition MDACC uses to say that tests are actionable or can be used as “standard of care”? Is each gene in the panel considered to have an indication? For example, would MDACC oncologists prescribe vemurafenib for BRAF-mutant colon cancer even though that drug has only been approved for melanoma? Dr. Patel responded that, no, BRAF is considered prognostic in colon cancer. He added that the judgment is site-specific, tumor-specific, and genotype-specific.

Other questions that arose but were not directly addressed were:

- How is MDACC evaluating the 400-gene panel and how does that evaluation differ from the evaluation of smaller, disease-specific panels?
- A large gene panel will contain genes that are both predictive and prognostic. How are the actionable genes picked out under these circumstances? What is considered actionable?

### Box 4. Advantages to NGS at MD Anderson

#### For Oncologists

- Simultaneous screening for multiple actionable targets
- Timely availability of results
- Detection of previously undiagnosed genetic abnormality
- Sensitivity to subclones with resistance

#### For Patients

- Comprehensive testing
- Declining costs, fast turnaround
- Potentially life-saving

#### For Institute

- State-of-the-art testing
- Value relative to costs of targeted agents
- Better clinical outcomes
- Data for assessing clinical effectiveness for future practices

#### For Laboratory

- Improved efficiency
- Consolidation of workflows
- Smaller sample amount
- Improved turnaround time

## **DR. JOHN PFEIFER**

### **WASHINGTON UNIVERSITY'S REIMBURSEMENT-DRIVEN MODEL**

In presenting how Washington University St. Louis was adopting NGS, Dr. Pfeifer emphasized that he was not warranting their approach as necessarily the best way. The program they established and the choices they made were designed for the context of their institution and were conditioned by the financial environment in which they adopted NGS.

Like MDACC, the adoption of NGS did not take place in a vacuum. They were already performing clinical molecular testing by other methods and wondered whether NGS could be harnessed to help them achieve results faster, better, or cheaper. Since they did not have grants, philanthropic support, or other independent sources of research funding, they did not design their program to ask academic research questions (for example, they are not looking at the roles of whole exomes or genomes and establishing a variety of clinical trials). Instead, their focus was on the usefulness of NGS to do existing clinical tests that clinicians were already ordering and payers were already covering.

Dr. Pfeifer and his staff found that NGS was particularly useful (informative and cost effective) in some circumstances but not necessarily in others. As noted earlier, in their lab NGS was ideally suited when more than 2.5 kilobases of DNA needed to be sequenced. However, with the hybrid capture technology they primarily use for NGS, Sanger indirect sequencing is better for small target regions and interphase FISH is better to evaluate specific structural rearrangements. They therefore developed the concept of "use cases" to identify circumstances in which NGS testing would be appropriate.

Over the last four or five years, as more genes of interest began to be identified in the literature and accepted in various types of cancer, large enough genetic regions were in need of analysis that NGS panels became more practicable. Again, their NGS testing program had to be self-sustainable. They believed that payers would require three things before providing coverage: utility (data showing improved outcomes), widespread adoption, and evidence-based literature. From the beginning, therefore, the design of panels was based on criteria for medical necessity. Initially, focused on the genes for which they were already getting paid, they developed an NGS panel testing 25 genes in solid tumors, leukemia and lymphoma. After evidence accumulated and acceptance grew for additional genes, the next version of the panel that expanded to 40 genes, all with an established role in patient care for diagnosis, prognosis, or treatment. This version of the panel actually examines an additional 108 genes that are not reported, but which are emerging in the literature as possible future targets of clinical relevance. Physicians are not allowed to have access to the additional gene results unless they have an IRB-approved protocol. Including these additional genes in the panel allows them to validate a very large set at one time and gives them flexibility to move quickly to begin reporting additional genes as the evidence develops.

A third version is now being developed, this time substratified based on disease and organ site with around 5 to 15 genes per subpanel. Increasingly, Dr. Pfeifer said, Wash U is moving towards designing panels to focus more on the clinical utility for specific organ sites. Dr. Pfeifer also noted that their panels are designed to detect all four classes of mutations (indels, structural variants, SNVs, and CNVs) and they employ top notch metrics in their validation process.

Their move to smaller panels has also been driven by clinician need. Their clinical colleagues have explained that they have only limited time to review the reports and would prefer simplified reports that are focused on the clinically actionable results (about 40% of their cases). Clinical interest in mutations that are not clinically actionable is limited, as is interest in variants of uncertain significance (VUSs). As panel sizes increase, more VUSs are observed and clinicians may spend a lot of time and effort following up a number of incidental findings that turn out not to have clinical significance.

How does reimbursement work for this approach of analyzing more variants than they routinely report to clinicians? They were aware of precedents involving panel tests where payers would pay for panels so long as there was a line between where the established clinical utility of the panel ended and the more investigational part began. Their local CMS contractor has agreed to reimburse if they separate the costs of testing the 40 genes from the marginal cost of testing the other 108 and bill appropriately. They intend to follow a similar approach for the version three panel. They will test for all the genes in one test platform but will bill for and report only the subset that was ordered.

To assist in obtaining reimbursement, they also developed a “white paper” to educate payers. It serves to explain NGS testing, define the concept of “use case scenarios,” and provide cost savings analysis. They have learned that Genomic and Pathology Services has to perform the payer contact because they have the most expertise and interest. Thus far, precertification is required less than half the time and usually can be handled with a phone call. While reimbursement varies, testing is being reimbursed by private insurance carriers for 80% to 90% of cases, with some patient groups and payers up to 95%. These high rates reflect the way they designed the panels – small panels testing genes that they were already getting paid to test. In sum, using NGS to perform tests that are being ordered for established reasons can result in cost savings, technical savings, and reimbursement.

Although they have achieved high reimbursement rates, the model is not perfect, Dr. Pfeifer observed. Since they have designed their program specifically to satisfy requirements for reimbursement, there are a number of procedures they are not using (because they would not be reimbursed) but which arguably should be done to improve the quality of information obtained or to support new knowledge generation. These procedures are:

- **Tumor-normal pairs.** First, it is not clear whether they should be sequencing matched tumor-normal pairs (i.e., comparing tumor DNA to that of the patient’s normal tissue – constitutional DNA -- to differentiate between germline and somatic variants, potentially simplifying analysis and improving error control). They have not tried to get reimbursement for such testing; only for tests looking solely at tumors.
- **Heterogeneity in primary tumor.** Second, it is well established that the genetic composition of primary tumors is highly heterogeneous. Hence, if one samples a primary tumor in multiple locations, one will find different mutations. There is presently scientific debate over whether tumors should be sampled more widely from six or eight different areas and the data somehow pooled. In any event, at this time reimbursement is not available for testing the same tumor six to eight times.



- **Heterogeneity of metastatic foci.** Similarly, metastatic foci have different patterns of sequential development and mutation versus the primary tumor. If only one sample of metastatic tissue can be taken, then the current practice is to sample a location thought to be more advanced in development. However, it can be argued that if a patient has three or four foci of metastatic disease, then each one should be biopsied and analyzed to look for shared mutations which might suggest a rational therapeutic strategy. Currently, reimbursement would not be available for this approach.
- **Characterization of escape mutations and RNAseq.** In addition, a growing body of literature is emerging on the value of taking a biopsy of recurrent tumors, not just to identify potentially new targets for therapy, but also to gain prognostic information based on characterizing how the tumor escaped therapy. Their model does not accommodate such testing, nor does it address emerging technologies such as RNA sequence analysis or methylation analysis.
- **Clinical trial support.** Finally, the model does not provide clinical trial support. It has no mechanism for identifying patients who might be good candidates for clinical trials.

### Key Questions and Discussion

Dr. Pfeifer was asked to define what he means by “actionable.” He replied that he thinks of actionable in terms of three different categories.

- 1) The first category is mutations for which there is an FDA-approved therapy, targeted based on histologic type of tumor and anatomic site. These are most often reimbursed by payers who are already paying for the result by some other methodology.
- 2) The second is mutations identified through the literature that are emerging and that his group believes will become standard-of-care relatively quickly. These types of variants are included in the panel because they can be useful to clinicians. A good example of this type of variant is the p10 mutation in patients with activating EGFR mutations in non-small cell lung cancer. A number of papers have shown that patients with an activating mutation in PTEN do not respond to EGFR inhibitors. For clinical colleagues, this is “emerging actionable intelligence” – it provides a rationale for moving on when certain patients are not responding to EGFR inhibitors. Very soon this type of information is likely to become an accepted part of clinical decision-making, but payers are reluctant to reimburse for these tests, which is one of the reasons they are developing more focused panels.
- 3) Finally, there are mutations that one could argue are actionable based on pathophysiologic pathways and studies in human cell lines, but for which there are no clinical data. Dr. Pfeifer’s laboratory does not regard these variants as actionable and does not include them in reports. This approach is not ideal from an academic perspective, since a lot of potentially relevant information is lost. But the testing model has to be supported by reimbursement.

In response to this description of actionability, Dr. Dickson from Palmetto commented that he uses the term “adaptive panels” for the type of panel development Dr. Pfeifer is doing, adding that he applauds this approach because of the way it depends on evidence and builds on emerging knowledge to anticipate future needs. “There are some questions,” he said, “but I really appreciate what you’re doing.”

**DR. GARY PALMER**

**FOUNDATION MEDICINE: COMMERCIAL DEVELOPMENT OF NGS**

Dr. Gary Palmer presented a significantly different approach to NGS testing compared to the two academic centers. Dr. Palmer explained that Foundation Medicine was set up as a private company (now public) to maximize the patient benefit from NGS. Thus, their definition of “actionable” is considerably more liberal and includes not only situations in which they believe a drug therapy is available (whether on or off label), but also situations in which a patient might be eligible for a clinical trial and even situations where there is a pathophysiologic rationale but no direct evidence. Concern about reimbursement has not been a primary consideration in their test development.

Foundation Medicine’s approach to testing is based on the following three hypotheses:

- 1) Genomic testing should be done across tumor histologies
- 2) Evidence is mounting that target activity is transferrable across tumor types
- 3) The best chance to find clinically relevant alterations is by deep sequencing entire coding regions of cancer-related genes

They perform exactly the same test for 236 genes across the entire solid tumor spectrum. They use this approach because the same alterations show up in different tumors, although the frequencies may differ. They believe, and have evidence to support, that target activity is generally transferrable across tumor types. In this regard, they are moving in the opposite direction of those institutions that are collapsing their tests into organ-specific panels. They also believe that “hot spot” testing (testing designed to detect only commonly occurring mutations in known cancer-related genes, as opposed to deep sequencing of genes) will inevitably miss identifying actionable alterations.

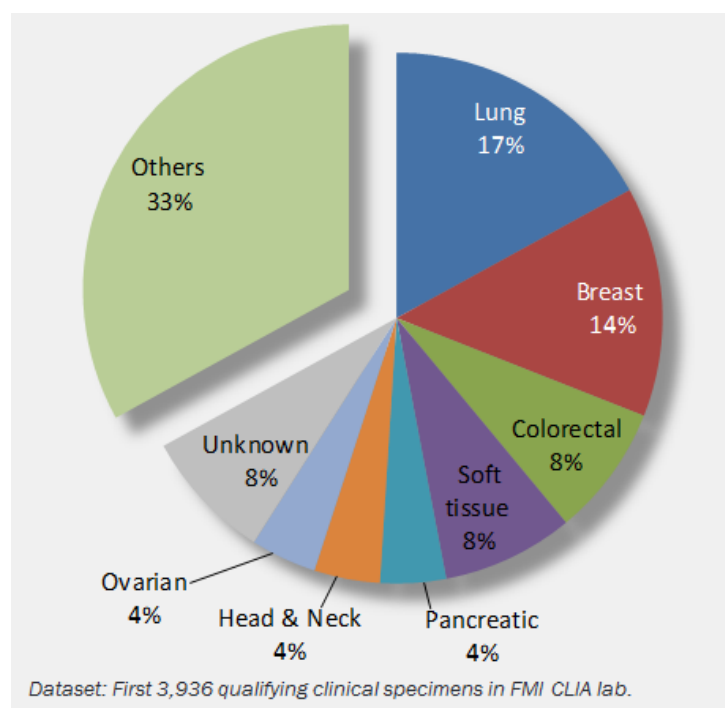
Their NGS test, FoundationOne®, has a published validation paper and is commercially available. It sequences the complete coding regions and selected introns of 236 genes that have known somatic mutations and can find all types of alterations. Two other notable aspects of the test are that it can be used on paraffin-embedded tissue and that it uses customized validated computational biology algorithms to narrow down the 1,000-1,500 discovered abnormalities per tumor to a manageable number that can be reported to the physician.

Regarding hot spot testing, in Foundation Medicine’s experience there are three to four times as many actionable alterations outside the hot spots as there are in the hot spots and they have presented relevant data in papers and several abstracts. Due to the difficulty of predicting which alterations a tumor may have, they have designed their test to be comprehensive. Although alterations in a few genes (e.g., p53) are relatively common, the frequency of alterations tails off quite quickly for other genes and it is very hard to predict which of the less common alterations will be present.

Their experience with clinicians accords with the observations of previous speakers: clinicians want the actionable information highlighted on the front of the report and do not want a lot of extraneous information. Thus a FoundationOne® report lists on the front page (1) all the actionable alterations found, (2) any FDA-approved drug for that tumor type for that type of alteration, (3) any FDA-approved

drug in a different tumor type that they believe could potentially be effective, and (4) any clinical trials for which the patient might be eligible based on the alteration.

In assessing the clinical utility of the test, Dr. Palmer noted, one of the threshold questions is whether physicians use the test in making decisions. Outcomes are the most important issue but if physicians are not using the test it will not be of any benefit. When the test first became available it had no specific indication for use. It was left to physicians to decide where the test might be of value and Foundation Medicine collected data on how physicians used it. Such data would be useful in designing trials to develop indications for use for payers.



Not surprisingly, the most common uses were for the most common cancers: lung, breast, and colorectal (Figure 3). In those instances, physicians were using the test for patients who had exhausted standard treatment options but were still candidates for additional treatment. For those patients, clinicians had the choice of picking a chemotherapy regimen “off the shelf,” which would have a very low probability of success, or using the test to look for a target. The clinical context is a critically important factor, Dr. Palmer asserted. Whether one would act on an alteration is related not only to the amount of evidence behind it but also to where in the treatment paradigm the patient happens to be.

**Figure 3. Sources of First 3,936 Samples Analyzed by FoundationOne®**

The 33% labeled “Others” consisted primarily of rare tumors at metastatic presentation. Because the tumors were rare there were no standard therapies and clinicians might have thought it more reasonable to look for a target to treat rather than to try a chemotherapy regimen. Pancreatic cancer presented a different clinical context. There are some therapies recommended by various groups but they are not very good. Physicians were using the test to identify targets at *de novo* presentation of pancreatic carcinoma. Overall, approximately 80% of their cases have at least one alteration that is actionable.

Dr. Palmer observed that when their test first became available it was used almost exclusively by academics but at present there are more community doctors than academics ordering it. It continues to be used primarily at the back end of the clinical paradigm when standard treatments have been exhausted although there are arguments for using it up front. Instead of ordering, for example, a lung panel, a physician could order their broader test and obtain at one time all the information in the lung panel as well as additional information that could help guide treatment if the lung panel is negative.

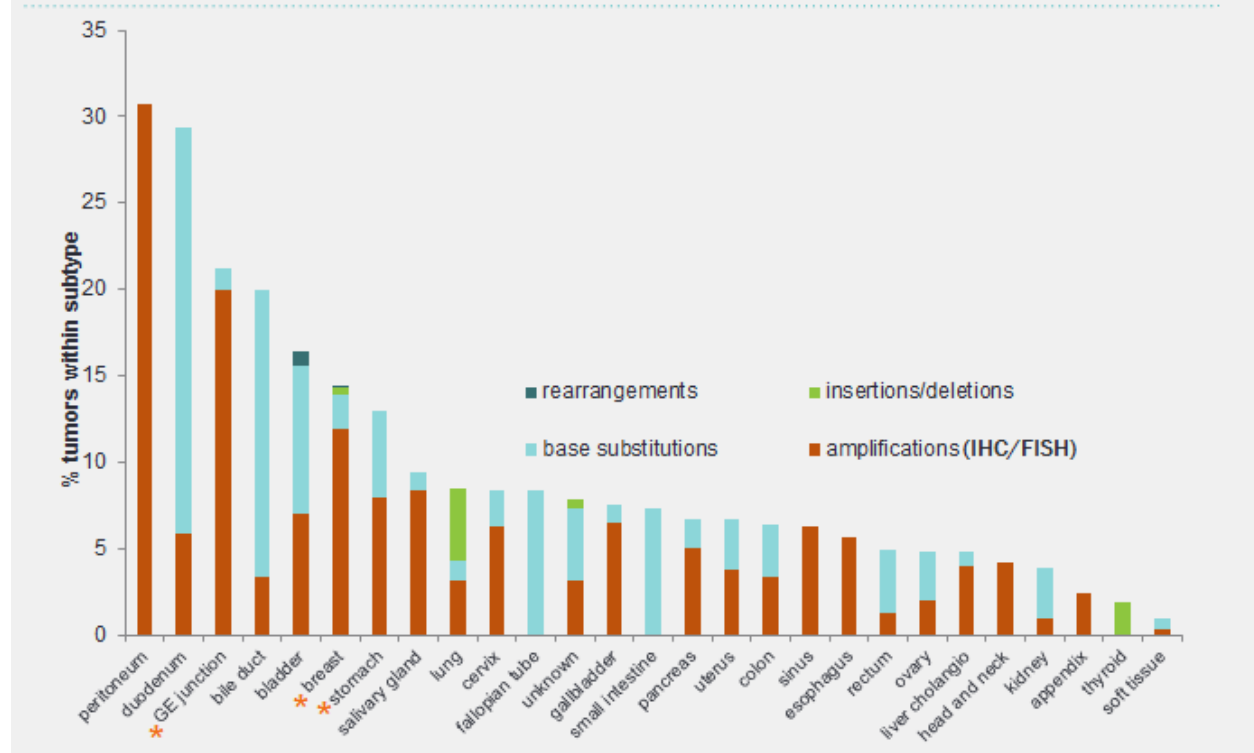
A study at Sloan-Kettering looked at this issue by using FoundationOne® to test 31 patients with lung adenocarcinoma who had tested negative for all alterations that were part of the Sloan-Kettering work-up. NGS testing identified 96 genomic alterations in the patients and 65% of patients had potentially treatable alterations.

Foundation Medicine has a study underway with US Oncology Research that quantitatively examines the decision impact of their test. In second-line solid tumor cases, physicians had to go on record explaining what their treatment plan would be when their patients progressed. After the plan had been documented, the test results were released to the physicians and nearly 30% of the time the physicians changed their recommendations. A second part of the study is examining the outcomes of the chosen treatments.

What about other data on outcomes? Dr. Palmer described two large plans to collect outcome data. First, they have established a registry that will be up to 3,000 patients followed for a year that will contain data on the targeted therapy the patient received and the outcome, among other things. In response to a question, Dr. Palmer explained that sites have to choose whether or not to participate in the registry. Those sites that do participate have to go through an IRB and should be providing information on 100% of their patients. Second, they are working with Google to develop a two-way portal system that can provide test results to physicians as well as collect data from the physicians on treatment decisions and outcomes. Moreover, physicians will be able to search the database to find cases comparable to their own and see treatments and outcomes. Although these systems are not perfect, the data are important to capture and clinical trials are not the answer for examining uncommon tumor alterations.

Turning to the hypothesis that target activity is transferrable across tumor types, Dr. Palmer reviewed data Foundation Medicine has collected on the frequency of *HER2* and *ERBB2* alterations in 27 different tumor types (Figure 4). Current practice is for oncologists to order HER2 amplification testing for breast and gastro-esophageal cancers and there are FDA-approved anti-HER2 agents for those cancers.

## HER2/ERBB2 Alterations Across 27 Tumor Types



**Figure 4. Genetic Alterations across Tumor Types as Detected by FoundationOne®**

Yet Foundation Medicine has identified HER2 or ERBB2 amplifications in numerous other solid tumor types (shown in orange in Figure 4). Will those tumors also respond to the anti-HER2 agents? Although the answers are not yet clear, data are accumulating for many of the tumor types in the form of aggregate case reports or clinical trials that are underway. Interestingly, he said, there are almost as many cumulative activating HER2 mutations as there are amplifications, although the proportion changes from one tumor type to the next. Breast cancer has primarily amplifications, for example, while lung cancer displays primarily activations. In currently ongoing clinical trials of neratinib, an investigational pan-HER inhibitor, patients with activating HER2 alterations have been responding to the drug.

Additional information has come from a study they are performing on ALK rearrangements in lung cancer. Of all the ALK-positive cases their NGS testing identified, 32% were not identified by the standard FISH testing and thus would not have been treated with crizotinib, an FDA-approved ALK inhibitor. In fact, however, 70% of the FISH-negative NGS-positive patients did respond to crizotinib.

Dr. Palmer closed by reviewing what Foundation Medicine believes to be reasonable indications for use of its test (see Table 1-B). He stressed that the test is not for every metastatic cancer case. A patient newly diagnosed with metastatic breast cancer, for example, has a number of other good options. Again, the clinical context is a critical factor in determining where comprehensive testing will be more useful.

**Table 2. Company-Recommended Reasonable Uses of FoundationOne® Testing**

Patients newly diagnosed with Stage IV adenocarcinoma of the lung.

Patients newly diagnosed with carcinoma of unknown origin.

Patients newly diagnosed with Stage IV rare or uncommon solid tumors for whom no systemic treatment exists in clinical care guidelines and/or pathways.

Patients newly diagnosed with Stage IV solid tumors where the median overall survival is less than two years (e.g., pancreatic cancer)

Patients diagnosed with solid tumors whose only specimen was obtained via fine needle aspiration, thoracentesis, paracentesis, or endobronchial ultrasound yielding insufficient tissue to complete requisite molecular testing thereby placing the patient at risk for new invasive diagnostic procedure(s).

Patients with Stage IV solid tumors who have exhausted established guideline-driven systemic therapy(ies) and requisite molecular testing but who maintain adequate functional status as deemed by the treating physician.

### Key Discussion Points

- A federal regulation will be coming into effect that will require CLIA-approved laboratories to allow patients to directly access testing results and reports.<sup>10</sup> Developing a report on NGS testing that will be understandable by patients will be a significant challenge.
- Data that might help define clinical actionability from the perspective of combination therapies are just now beginning to be collected.
- How often there are discrepancies in diagnostic tests and what do they mean? One participant shared that they may see discrepancies in 20% to 40% of their cases, although there is usually a clear biologic mechanism for the discrepancy that they can find with investigation. Nevertheless, because they are learning about the limitations of established tests, they are using NGS to inform the results from companion diagnostics.
- Payers have to grapple with a number of complicated issues related to this model. Isn't it still more cost-effective to use a simpler methodology if one is examining only a small number of genes? How can a payer determine and compare the value of different platforms testing for anywhere from 5 to 100 genes? How can a payer know whether mutations picked up in non-hot spot testing are actionable? Is there value in performing NGS on top of companion diagnostics or can NGS replace them? What is a reasonable basis for determining the amount of reimbursement? How can a payer reach decisions other than case by case by case?

---

<sup>10</sup> CLIA Program and HIPAA Privacy Rule; Patients' Access to Test Reports; Final Rule. Federal Register Vol. 79 No. 25 (Thursday, February 6, 2014). The final rule became effective April 7, 2014 and all entities covered by the Health Insurance Portability and Accountability Act (HIPAA) must comply with the rule no later than October 6, 2014.

## **SESSION 3: FUTURE EVIDENCE FRAMEWORKS**

Dr. Tunis spoke briefly at the beginning of the session to explain that the focus of the next presentations would be problem solving and developing solutions to the issues already raised. He noted significant disparity of views and seemingly large gaps to bridge but emphasized that because the participants were all meeting together there was opportunity to begin to converge around some areas.

### **DR. DAVID LITWACK**

#### **FDA: CHALLENGES AND A POTENTIAL NEW APPROACH**

Dr. David Litwack, Personalized Medicine Staff at FDA/CDRH/OIR, began by outlining the challenges of NGS for FDA. As already noted, NGS tests can exist in many forms within a single lab, which means that they do not fit FDA's one test/one submission or amendment model. Often there is no clear intended use population. Particularly with whole genome sequencing, it is impossible to list all the variants that may be detected. FDA has difficulty deciding on specific test labeling, especially given the varying levels of sources of evidence. Finally, Dr. Litwack observed that the field is very dynamic with science rapidly driving new uses.

In talking about how FDA reviews *in vitro* diagnostics (IVDs), Dr. Litwack noted that FDA evaluates analytical and clinical validity, but not clinical utility. Clinical studies are the traditional method of collecting the necessary clinical evidence. An NGS test would generally be considered to be a companion diagnostic and evaluated by using a "locked down" version of the test in conjunction with a drug trial. With NGS panels, however, defining a locked down version is difficult. In the future, FDA will likely increasingly rely on "big data" concepts using literature, case studies, and particularly genetic databases to evaluate multiplex IVDs.

Before discussing genetic databases in more detail, Dr. Litwack addressed the challenge of local testing in clinical trials. Clinical trials may use local tests to select patients or identify subgroups for analysis, or they may use a patient's existing medical record that could contain results from a different test. When FDA does not know what test was used, what its technology was, or what its performance characteristics are, test results may not be comparable and grouping all the patients together will lead to heterogeneity. While the clinical trial results may then better reflect the real world practice of medicine, evaluation of the results is more difficult. Confirming local results with centralized testing is not an ideal solution because it results in selection bias, and using only central testing would limit the clinical utility of the trial results. Genetic databases will encounter the same problems. One way of dealing with the problem is through systems such as electronic health records if they capture information about what test was used and the laboratory that generated it.

Returning to genetic databases, Dr. Litwack noted that FDA recently took a non-traditional approach to evaluating clinical validity by using a cystic fibrosis database (CFTR2) as the basis for clearing the Illumina MiSeqDx Cystic Fibrosis 139-variant assay. The CFTR2 database had already been set up at Johns Hopkins and was a very high quality curated database containing mutation name as well as preclinical and clinical data including functional assay data. It was particularly thorough because it was supported by the cystic fibrosis patient community. FDA's only real request was to version the database to track changes. Because of the wealth of high quality data, including variants, FDA was able to evaluate the

clinical validity of the assay. Dr. Litwack suggested that FDA may more frequently rely on this model as it attempts to combine data from different entities to evaluate evidence.

This novel use of the CFTR2 database prompted FDA's Personalized Medicine Staff to obtain funding by the Critical Path Institute to develop a framework to guide thinking about databases from a regulatory standpoint. The objectives included: (1) developing a comprehensive list of databases; (2) understanding how the databases are validating, annotating, and curating data; (3) evaluating information on quality assessments, systems used, and process improvements; and (4) identifying best practices with respect to database quality, standards, and criteria. To identify best practices, FDA performed a series of structured interviews with a number of different entities. The following list of proposed best practices resulted:

- An SOP in place for the evaluation of different forms of evidence used in the determination of pathogenicity of a variant.
- Inclusion of clinical/phenotypic characteristics as part of the variant assessment process.
- Standard nomenclature for gene name, genomic coordinates, nucleotide change, amino acid change, etc.
- Doctorate-level full-time curators with a biology background with a second check by either another doctorate-level biologist or a genetic counselor, and an SOP in place for curation process and curator training.
- An SOP in place for evaluation of research literature, and when practical, a pre-curated literature knowledgebase.
- "Complete provenance tracking" of a start-to-finish analytical pipeline, a record of versions of databases and tools that are used in the analytical pipeline, parameters in those pipelines, and cutoffs used for quality filters for the particular assay, as well as any changes in versioning of annotation.

Dr. Litwack emphasized that these are only proposed and would certainly be discussed and expanded or modified based on input from stakeholders.

In closing, Dr. Litwack shared some ideas relating to genetic databases that FDA is considering for the future. One of the most important points was the possibility of moving from thinking about what the specific rules of evidence ought to be to more of a quality systems approach to ensure that genetic databases operate with a certain degree of quality. A related concept is developing a "regulatory grade" database that could be used to shorten review if it contained information on the genetic variant being studied. Finally, Dr. Litwack observed that 30% to 40% of the databases that FDA identified were no longer active, raising issues of how to use such evidence and how to track and preserve evidence in the future.



### **Key Discussion Points**

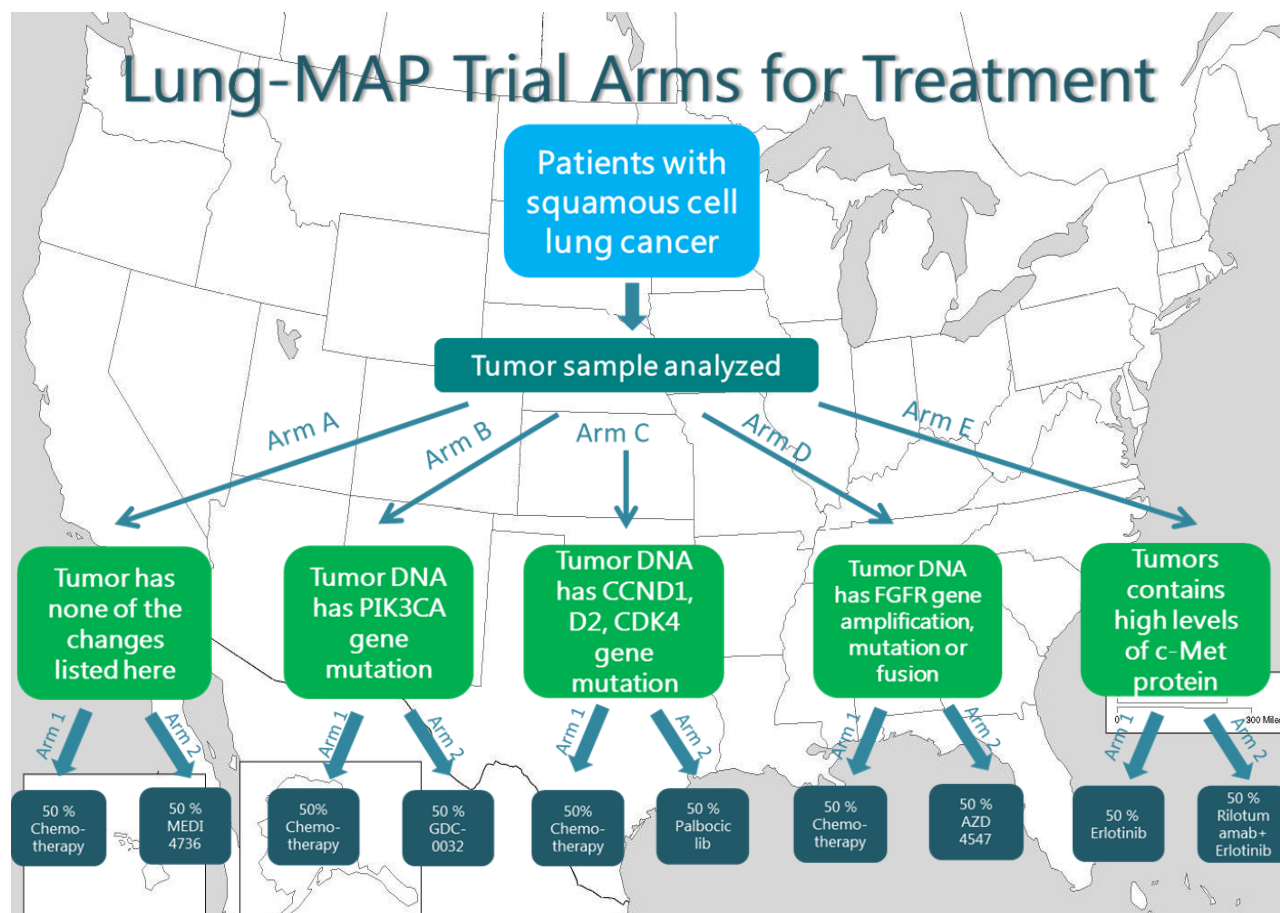
- Databases that are not well-managed or out of date will not provide useful information.
- Guidance from FDA on the criteria for a “regulatory grade” database would be important to standardization.
- Standardization of analytical validity would have many uses, including interpreting data from different databases.
- There is a lot of evidence available but there has to be a way of aggregating it to make it useful to patients.

### **DR. JEFF ALLEN**

#### **A COLLABORATIVE APPROACH TO CLINICAL TRIALS**

Dr. Jeff Allen, Executive Director, Friends of Cancer Research, explained the Lung Cancer Master Protocol (Lung-MAP), a cutting edge clinical trial that uses a multi-drug targeted screening method to match patients with promising new treatments. Lung-MAP was created to address several challenges in drug development. Current standard practice is for each new therapy to be tested independently from other therapies seeking to treat the same condition. Time is also an issue: it takes two to three years to start up and implement a large scale phase 3 trial and then patient accrual rates are a consistent challenge, particularly for rare genetic variants.

Building on an idea from the FDA, Friends of Cancer Research brought together a multi-stakeholder working group that has moved, in a year and a half, from concept paper to operationalizing a master protocol for a multi-drug clinical trial. The trial involves second-line squamous cell carcinoma of the lung and evaluates several different drugs simultaneously in different arms. The drugs being investigated were chosen by an independent drug selection committee and had to be able to demonstrate biological activity against a measurable target. The study is designed as a phase 2/3 trial with the phase 2 portion functioning as a screening process to identify which drugs show sufficient promise to proceed to the phase 3 portion. The clinical trial is shown schematically in Figure 3 below.



**Figure 5. Lung-MAP “Master Protocol” Schema**

While there is consistency between each of the arms in terms of the endpoints and how the drugs will be evaluated, each sub-study can be conducted independently of the others. Patients will accrue at different rates depending on the prevalence of the given biomarker. Foundation Medicine was selected as the entity to screen the patients for biomarkers and it is using NGS to do so. Dr. Allen explained that the goal is to have four to seven independent arms operating at any given time. The trial is flexible in that it would allow a company to bring its biomarker hypothesis into the trial, plug in and obtain the data needed to analyze its drug, and then plug out, leaving space for others to come along. The trial also should be attractive to patients who are considering clinical research as part of their treatment paradigm because they would undergo a single screen for many potential drug treatments rather than proceeding single gene test by single gene test to see whether they qualify for a potential trial.

In response to a question, Dr. Allen explained that there was no minimum gene mutation frequency to be eligible for the trial, but they hoped to have about 50% of the patients eligible for a biomarker arm. He then discussed the advantages of their trial design: enrollment efficiency by having more treatment options available; operational efficiency by having the ability to just amend the protocol for a new drug rather than starting from scratch; consistency; predictability in that following completion of phase 3 a drug would be eligible for regulatory approval; and patient benefit by having more options available.

[Dr.] Allen cautioned that in some instances NGS results would have to be bridged to a clinical diagnostic or regulators would have to address how to use an NGS test as a companion diagnostic.

Comments from other stakeholders were positive and supportive of the effort. Dr. Allen noted that in the future they may be able to increase efficiency by testing multiple drugs in one class or testing combination therapies but their primary goal was to get the study up and running quickly. While they are excited about the five drugs they are testing, perhaps the most important result could be development of a system that allows for rapid integration of new biomarker-driven drugs over time.

## **DR. BRUCE QUINN**

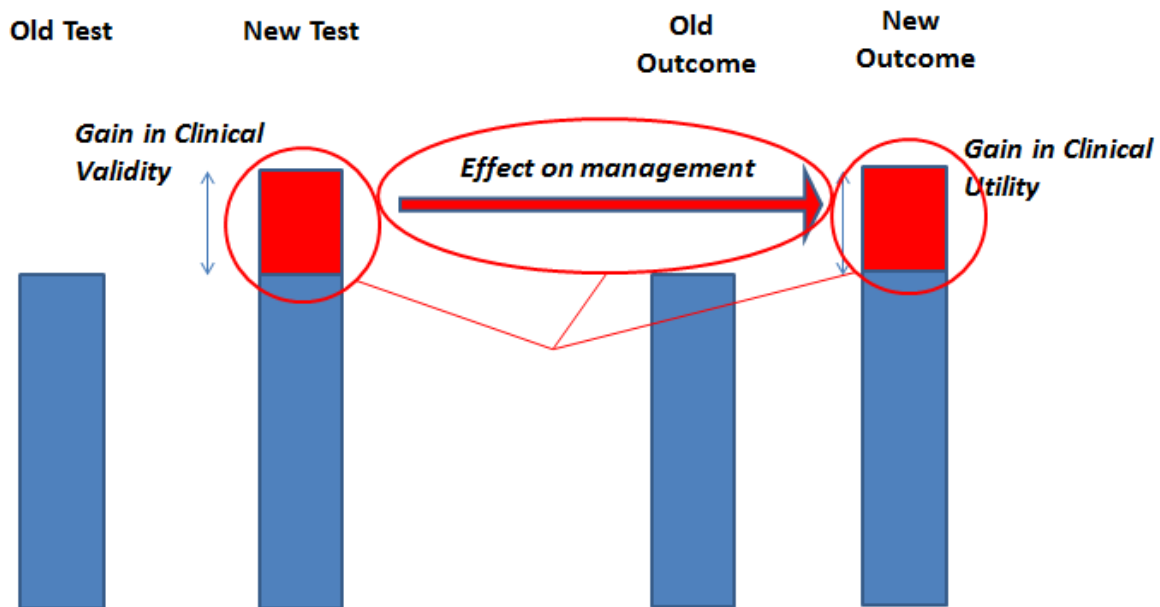
### **STRATEGIC APPROACHES TO CLINICAL UTILITY**

In the final session, Dr. Bruce Quinn, Senior Health Policy Advisor with Foley Hoag, considered the existing frameworks for thinking about clinical utility. He turned first to the concepts of analytic validity, clinical validity, and clinical utility developed by EGAPP to characterize diagnostic test evaluation. These concepts, he noted, constitute an information classification system, not a framework for decision-making. Such a classification system will not ultimately provide a basis for deciding the adequacy of a diagnostic test for clinical use. Likewise for the Fryback-Thornbury hierarchy (discussed in an earlier meeting session, Box 3), which can be readily cross-walked to the EGAPP classification system. As well the traditional hierarchy of evidence (such as that represented in Box 2, p.8 above), often referred to as “levels” of evidence, really is a trial design classification system. A study’s position on the hierarchy does not necessarily correspond to the value of the evidence it provides. Randomized clinical trials could suffer from flaws such as using the wrong end points or patient population, while studies lower on the hierarchy (e.g., cohort studies) may not be compromised by any specific identifiable bias.

Dr. Quinn asserted that the types of frameworks reviewed above can lead to tautological assessments of diagnostic tests: “Your test lacks sufficient clinical utility, so you need more clinical utility.” More structure and specificity are needed to help payers and test developers communicate meaningfully and effectively with each other about the value or usefulness of diagnostic tests.

Consider the FDA approach to risk-benefit. The benefits are usually understood far better than the risks in drug trials. Because there is no satisfactory quantitative method for comparing the risks and benefits, risk-benefit assessments are often complex rather than simple, difficult to explain, and experts may disagree. Nevertheless, the goal is effectively to subtract the risks from the benefits to a “reasonable assurance” that the product is safe and effective.

With diagnostic tests, the goal is to achieve a gain in clinical utility over whatever situation currently exists. Hence, clinical utility is always comparative against something, and this increase in utility is always achieved through an effect on patient management that is driven by gains in clinical validity (see Figure 4)<sup>11</sup>. The gains in both clinical validity and utility are comparative: it is always necessary to identify comparator(s), the units of comparison, and the associated uncertainty (statistical, pragmatic, and conceptual). The issue for coverage of diagnostic tests is whether the available data and rationale would convince a reasonable skeptic of the three features shown in red: the gain in clinical validity, the effect on management, and the gain in clinical utility.



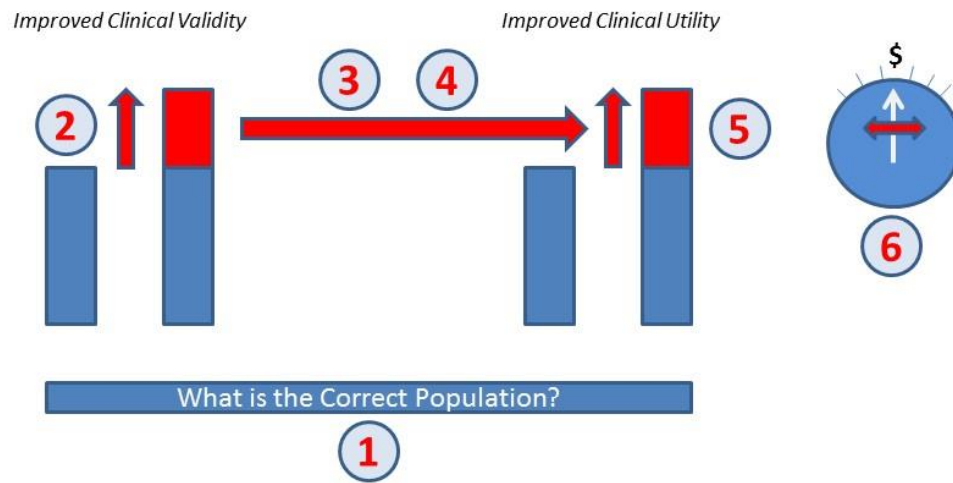
**Figure 6. Potential Gains in Validity and Utility from New Tests<sup>11</sup>**

Dr. Quinn then turned to a six-question framework developed by Felix Frueh for real tests and real decisions that he believes can provide a communications structure for clinical utility. They are granular enough to characterize real-world concerns but not so granular that they create unnecessary complication. They are:

- 1) Who should be tested and under what circumstances?
- 2) What does the test tell us, that we did not know?
- 3) Can we act on the information provided by the test?
- 4) Will we act on the information provided by the test?
- 5) Does the outcome change, in a way we find value in?
- 6) Can we afford it? (Is it a *reasonable* value?)

These questions can be graphically illustrated by superimposing them on the following figure (see Figure 5)<sup>11</sup>.

<sup>11</sup> Frueh FW, Quinn B. Molecular diagnostics clinical utility strategy: a six-part framework. Expert Review of Molecular Diagnostics 2014 14:7, 777-786



- 1) Who should be tested and under what circumstances?
- 2) What does the test tell us, that we did not know?
- 3) Can we act on the information provided by the test? (Ideal)
- 4) Will we act on the information provided by the test? (Real)
- 5) Does the outcome change, in a way we find value in?
- 6) Can we afford it? (Is it a *reasonable* value?)

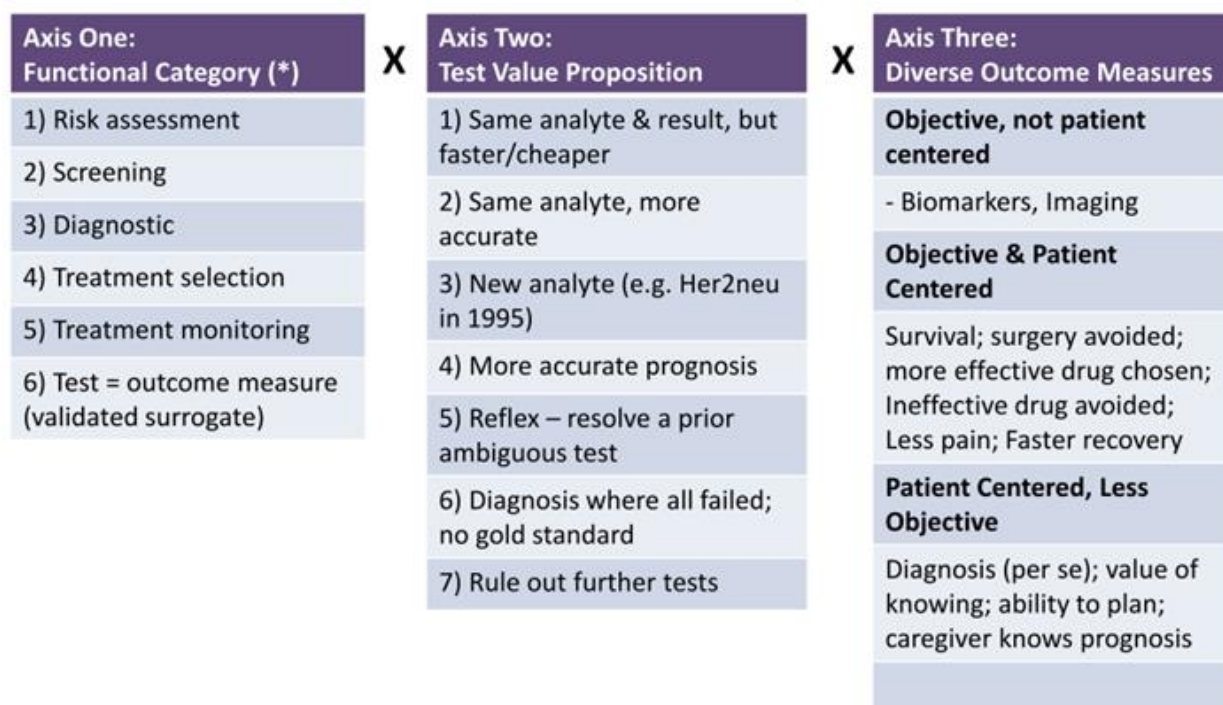
- Comparative Questions (2,5,6)**
- Against what comparator(s)
  - In what units?
  - With what uncertainty?
    - *Statistical, Pragmatic, Conceptual*

15

**Figure 7. The Six Questions Illustrated in Graphic Form**

To illustrate, Dr. Quinn applied this framework to two different real life examples: non-invasive prenatal testing (NIPT) for Down syndrome and amyloid PET scans for Alzheimer’s disease. NIPT performed well on all six questions and thus would be predicted to be adopted quickly by payers and patients, which is what actually happened. On the other hand, amyloid PET scans performed badly on all six questions so the model would predict difficulty in obtaining payer coverage. Of course, Dr. Quinn noted, people will still disagree on the answers to the questions. However, the framework provides structure and specificity to discussion of the issues.

In closing, Dr. Quinn closed by noting that business models for diagnostic tests can be divided across three axes (see Figure 6). The first axis is the functional category of the test. For example, does the test assess risk, screen, provide a diagnosis, or help select treatment, among other functions?



**Figure 8. Three Axes of Business Models for Diagnostic Tests**

The second axis categorizes the value proposition of the test. Dr. Quinn presented seven categories in the figure but noted that this is not intended to be an exhaustive list. Some of these value propositions may be seen as equally important, but measurement of them can be difficult and comparisons between the measures may be unintelligible or difficult to quantify. What does it mean if an R-value is larger or a p-value is smaller? What if you have a tumor of unknown origin test, where three different value propositions come into play? Which outcomes measures are best to use in this circumstance?

The third axis addresses outcome measures. These measures can be divided into three groups depending on whether they are more physical/objective or perceived by the patient.

These conceptual tools illustrate the types of specific information needed to answer Frueh’s questions in a way that can foster constructive communication. While payers and test developers may still disagree on the answers to the questions, the elements of these axes provide a vocabulary better geared to decisions, rather than merely to sorting of information.

## **ENDING THOUGHTS AND CONTINUING DIALOGUE**

Dr. Tunis closed the day by thanking the participants for engaging in this complicated and important discussion. Defining the concept of “adequate evidence” is difficult and different entities may reach different conclusions. The answer is not scientific, he reiterated, but a convergence of viewpoints about the right balance between acceptable levels of uncertainty and feasibility. In the absence of a national agency to establish the standards of evidence for clinical utility, a concrete statement that approximates

some convergence of viewpoints should have some impact. Dr. Tunis explained that CMTP would be forming a technology working group to review information, draft ideas about methodologic recommendations, and coordinate input and feedback, and expressed his hope that the participants would remain engaged in the dialogue going forward.

## APPENDICES

### Appendix A

#### Demonstrating Clinical Utility of NGS in Clinical Oncology

#### AGENDA

**JULY 7, 2014**

8:00 am—3:00 pm EDT

Maryland Room, 21<sup>st</sup> Floor

World Trade Center, Baltimore

<b>Topic</b>	<b>Time (EDT)</b>
<b>Breakfast and Check In</b>	8:00 am—8:30 am
<b>Welcome and Overview</b> <i>Donna Messner, PhD, Center for Medical Technology Policy</i> <b>Opening Remarks</b> <i>Sean Tunis, MD, MSc, Center for Medical Technology Policy</i>	8:30 am—9:00 am
<b>Opportunities and Challenges for NGS as an Emerging Clinical Tool</b> <ul style="list-style-type: none"> <li>• <i>What is the basic set of challenges presented by the emerging clinical use of NGS?</i></li> <li>• <i>What concerns do payers have? How do clinicians view these concerns?</i></li> <li>• <i>Can a shared understanding be gained of the problem set and approach needed to address it?</i></li> </ul> <b>Tamara Syrek Jensen, JD, Centers for Medicare and Medicaid</b> <b>Dane Dickson, MD, Palmetto GBA</b> <b>John Pfeifer, MD, PhD, Washington University</b>	9:00 am—10:30 am
<b>Break</b>	10:30 am—11:00 am
<b>Current Evidence and Uses of NGS</b> <ul style="list-style-type: none"> <li>• <i>How is NGS/genomic information being used clinically (in cancer centers) and commercially (by companies like Foundation Med)?</i></li> <li>• <i>What evidence is available to support decision-making? What is the threshold for “actionable” information?</i></li> <li>• <i>In what sense are gene panels “actionable”?</i></li> <li>• <i>What programs are being implemented to gain new information from patient care?</i></li> </ul> <b>Keyur Pravinchandra Patel, MD PhD, MD Anderson Cancer Center</b> <b>John Pfeifer, MD, PhD, Washington University</b> <b>Gary Palmer, MD, JD, MBA, MPH, Foundation Medicine</b>	11:00 am—12:30 pm
<b>Lunch</b>	12:30 pm—1:00 pm



<b>Topic</b>	<b>Time (EDT)</b>
<p><b>Future Evidence Frameworks</b></p> <ul style="list-style-type: none"> <li>• <i>What will future evidence requirements look like?</i></li> <li>• <i>What evidence is needed for clinical use decisions? What are the most expedient ways to get it?</i></li> <li>• <i>What's the right mix? Bucket/umbrella trials? Registries?</i></li> <li>• <i>Who will pay for studies?</i></li> </ul> <p><b>David Litwack, PhD, FDA</b>  <b>Jeff Allen, PhD, Friends of Cancer Research</b>  <b>Bruce Quinn, MD, PhD, Foley Hoag</b></p>	<p>1:00 pm—2:30 pm</p>
<p><b>Concluding Question</b></p> <ul style="list-style-type: none"> <li>• <i>Are we on the right track?</i></li> <li>• <i>Where do we go from here?</i></li> </ul> <p><b>Sean Tunis, MD, MSc, Center for Medical Technology Policy</b></p>	<p>2:30 pm—3:00 pm</p>

## Appendix B

### Demonstrating Clinical Utility of NGS in Clinical Oncology Workshop Participants July 7, 2014

<p><b>Jeff Allen, PhD</b> Executive Director Friends of Cancer Research</p>	<p><b>David Bernstein, PhD</b> Assistant Director of Research Policy American Society of Clinical Oncology</p>
<p><b>Paul R. Billings, MD, PhD</b> CMO (consulting) Thermo Fisher Scientific</p>	<p><b>Mitchell I. Burken, MD</b> Medicare Medical Director Novitas Solutions, Inc.</p>
<p><b>Deborah Collyar, BS</b> Director of Patient Advocate Research Team (PART) Program Patients Advocates in Research</p>	<p><b>Scott Devine, PhD</b> Outcomes Research Scientist Merck &amp; Co., Inc.</p>
<p><b>Dane J. Dickson, MD</b> Director of Clinical Science MoDx, Palmetto GBA</p>	<p><b>Dana Dilbeck</b> Vice President Market Access and Strategy MolecularHealth</p>
<p><b>Rob Dumanois</b> Manager – reimbursement strategy Life Sciences Solutions Thermo Fisher Scientific</p>	<p><b>Marcia Eisenberg, PhD</b> Senior Vice President Research and Development/Science and Technology Laboratory Corporation of America Holdings</p>
<p><b>Lindee Goh, PhD</b> Partner Tapestry Networks</p>	<p><b>Stanley R Hamilton, MD</b> Division Head, Pathology/lab Medicine The University of Texas MD Anderson Cancer Center</p>
<p><b>Ruslan Horblyuk, PhD, MBA</b> Director, Outcomes &amp; Evidence Oncology Global Health &amp; Value Pfizer, Inc.</p>	<p><b>Laura Housman, MPH, MBA</b> SVP, Chief Commercial Officer MolecularHealth</p>
<p><b>Greg Jones</b> Director of Business Development and Technical Affairs MDxHealth</p>	<p><b>Chris Karlovich, PhD</b> Principal Scientist, Molecular Diagnostics Clovis Oncology</p>

<p><b>Ira Klein, MD,MBA,FACP</b>                  Chief of Staff, Office of the Chief Medical Officer,                  National Accounts Clinical Sales &amp; Strategy                  Aetna</p>	<p><b>Melvin Limson, PhD</b>                  Director of Scientific Programs                  Association for Molecular Pathology</p>
<p><b>Erick Lin, MD, PhD</b>                  Manager, Medical Affairs                  Illumina, Inc.</p>	<p><b>David Litwack, PhD</b>                  Personalized Medicine Staff                  Center for Devices and Radiological Health                  U.S. Food and Drug Administration (FDA)</p>
<p><b>Bryan Loy, MD, MBA</b>                  Market Medical Officer-Kentucky                  Physician lead-Cancer                  Health Guidance Organization                  Humana</p>	<p><b>Gary H. Lyman, MD, MPH</b>                  Co-director                  Hutchinson Institute for Cancer Outcomes                  Research (HICOR)</p>
<p><b>Gary M. Martucci</b>                  Senior Director, Strategic &amp; National Accounts                  Foundation Medicine, Inc.</p>	<p><b>Robert McDonough, MD, JD, MPP</b>                  Head of Clinical Policy Research &amp; Development                  Aetna</p>
<p><b>Jennifer Morrissette, PhD</b>                  Clinical Director,                  Center for Personalized Diagnostics                  Hospital of the University of Pennsylvania</p>	<p><b>Jonathan Myles, MD, FCAP</b>                  Staff Pathologist                  Department of Anatomic Pathology                  Cleveland Clinic</p>
<p><b>Lee N. Newcomer, MD, MHA</b>                  Senior Vice President, Oncology, Genetics and                  Women's health                  UnitedHealth Group</p>	<p><b>Jan A. Nowak, PhD, MD</b>                  Northshore University Health System                  Economic Affairs Committee Co-Chair                  Association for Molecular Pathology</p>
<p><b>Frank S. Ong, MD</b>                  Associate Director                  Medical Affairs                  Illumina, Inc.</p>	<p><b>Gary Palmer, MD, JD, MBA, MPH</b>                  Senior Vice President                  Medical Affairs                  Foundation Medicine, Inc.</p>
<p><b>Keyur P Patel, MD, PhD</b>                  Assistant Professor, Hematopathology Adm                  The University of Texas MD Anderson Cancer Center</p>	<p><b>John D. Pfeifer, MD, PhD</b>                  Vice Chairman for Clinical Affairs,                  Department of Pathology and Immunology                  Professor of Pathology and Immunology                  Professor of Obstetrics and Gynecology                  Washington University School of Medicine</p>

<p><b>Bruce Quinn</b> Senior Health Policy Specialist Foley Hoag LLP</p>	<p><b>Mitch Raponi, PhD</b> Senior Director, Molecular Diagnostics Clovis Oncology</p>
<p><b>Alan B. Rosenberg, MD</b> Vice President Clinical Pharmacy and Medical Policy WellPoint, Inc.</p>	<p><b>Andrea Saltzman, RN, CIP</b> Broad Institute Assistant Director, Office of Research Subject Protection</p>
<p><b>Richard L. Schilsky, MD, FACP, FASCO</b> Chief Medical Officer American Society of Clinical Oncology</p>	<p><b>Mary Lou Smith, JD, MBA</b> Co-founder, Research Advocacy Network Co-chair, ECOG Patient Representative Committee and RTOG Patient Advocate Committee</p>
<p><b>Kendall Stevinson, MD, MPH</b> Director Global Health Outcomes, Oncology Merck &amp; Co., Inc.</p>	<p><b>Tamara Syrek Jensen, JD</b> Director, Coverage and Analysis Group Center for Clinical Standards and Quality Centers for Medicare &amp; Medicaid Services</p>
<p><b>Robyn Temple-Smolkin, PhD, HCLD</b> Project Manager, Clinical / Scientific Programs Association for Molecular Pathology</p>	<p><b>Tyler Whisman, PharmD, BCOP</b> Managing Pharmacist Specialty Pharmacy Strategy Humana Pharmacy Solutions</p>
<b>CENTER FOR MEDICAL TECHNOLOGY POLICY (CMTP)</b>	
<p><b>Sean Tunis, MD, MSc</b> President and Chief Executive Officer</p>	<p><b>Donna A. Messner, PhD</b> Vice President and Senior Research Director</p>
<p><b>Elisabeth J. Houtsmuller, PhD</b> Vice President and Senior Research Director</p>	<p><b>Srijana Rajbhandary, MPH</b> Research Associate</p>
<p><b>Scott Allocco (Consultant)</b> President SJA Healthcare Strategies</p>	<p><b>Julie Simmons, CMP</b> Manager, Marketing and Communications</p>