In mid-June, the payment contractor for Medicare in California agreed to reimburse oncologists for using a diagnostic test that relies on gene expression profiling to classify undifferentiated metastatic tumors with unknown sites of origin. Such tumors occur in about 30,000 new cancer cases each year.

The Pathwork Diagnostics test, which the U.S. Food and Drug Administration approved in 2010, analyzes the expression patterns of more than 2,000 genes to determine which of 15 common cancers the unidentified tumors mimic. The payment decision in California, where the Pathwork lab is based and samples are sent, enables the test’s wider use to aid in choosing treatment strategies in these situations.

A company-funded validation study that appeared in the January 2011 Journal of Molecular Diagnostics showed that the Pathwork® Tissue of Origin Test accurately identified tumors 88.5% of the time. It had only a 1% false-positive rate, the study said. A recent meta-analysis of immunohistochemistry (IHC) studies, which was also company-funded, showed that the traditional pathology method generated accurate diagnoses only two-thirds of the time.

And that’s how company officials are marketing the test—as a better, and potentially less costly diagnostic tool than repeated immunohistochemistry and imaging tests. “That’s the criteria for all diagnostic tests—that you accurately measure what you purport to measure,” said David Henner, M.D., Ph.D., chief medical officer for Pathwork Diagnostics. “We’re not a stand-alone diagnostic. We’re an aid to the physician getting to the correct diagnosis.”

**Uncertain Clinical Benefit**

But introducing a new test that is somewhat better than existing methods for identifying the site of origin of unidentified cancers doesn’t necessarily translate into real clinical benefit. And that assertion suggests that the test may not have meaningful clinical utility for practicing oncologists, according to
experts who attended a mid-June National Cancer Institute workshop, “Criteria for Use of Omics-Based Predictors in Clinical Trials.”

“The assumption is that those tumors will respond to the organ-specific chemo, and we don’t know that,” said Michael Birrer, M.D., a professor of medicine at Harvard Medical School and director of gynecologic medical oncology at Massachusetts General Hospital. “This test tells you it is gastric cancer. So you get gastric cancer chemo. What you’re obligated to show is that it has the same response rate as gastric cancer. I haven’t seen that data. It’s not a robust validation.”

That was the overriding message the NCI workshop sent to the 100 or so attendees, most of whom were academic researchers working on developing complex computer-driven tests for identifying and classifying tumors based on expressions of a single or a suite of proteomic or genomic biomarkers. In other words, efforts to develop new “omics” tests should be aimed at giving practicing oncologists useful tools that have proven clinical utility.

Developers’ tests could aim to identify specific targets for drug development, create subsets of common cancers that are most likely to respond to targeted therapies, or identify patients most likely to experience the toxic effects of a particular drug. But in each case, any test ought to be linked to clinical trials that show that using the test leads to superior patient outcomes. “Our goal has to be on clinical utility,” said James H. Doroshow, M.D., deputy director of NCI for clinical and translational research. “How do we bring new tests and information to the bedside that affects the treatments that patients get?”

Setting Research Standards High

Since the workshop took place in the aftermath of the Potti scandal at Duke University (See J. Natl. Cancer Inst. 2011; 103:916-7), the meeting focused heavily on what researchers need to do to produce valid, reproducible results that can meet the regulatory standards for using experimental tests in clinical trials, which requires an IDE (Investigational Device Exemption) from the FDA. Moving a test from the computer lab to a clinical trial setting requires attention to proper methods of handling tissue samples and assays, for instance, and creating repeatable algorithms and calculations that can be conducted at multiple locations with similar results.

However, IDE applicants often give FDA reviewers told the session that they often get very little information about how biospecimens are handled. “You would think you would have the protocols for sample collection, shipment, storage—everything written out,” said Reena Philip, Ph.D., associate director in the Office of In Vitro Diagnostic Device Evaluation and Safety at the FDA Center for Devices and Radiologic Health at FDA. “But that’s not what we see.”

“If you have some samples that are frozen and some fixed, you have to demonstrate that you got the same results from the frozen as from the fixed,” added Donna Roscoe, Ph.D., a reviewer in the division. “And if you’re doing a retrospective sampling, we’d ask you for your protocol for choosing the samples so we were assured that you weren’t cherry-picking the samples to show that your assay works.”

Jill Mesirov, Ph.D., associate director for computational biology and bioinformatics at the Broad Institute at MIT and Harvard, warned conference attendees that they must be willing to open the black box to reveal the decision algorithm and computer code so other researchers could reproduce the results. Although several labs are coming up with tools that allow for sharing, she touted a system developed in her lab called Gene Pattern, which uses software to embed everything from the original data to the algorithms to the test results. “When is a model ready for prime time?” she asked. “When there is independent verification.”

Journals and funding agencies, she said, could play a major in ensuring that data and source code were readily available to outside researchers. “They could require for publication that authors put their data in a publicly available database,” she said. “We need incentives to make the data and algorithms available.” Roger Peng, Ph.D., editor of the Journal of Computational Biology, said his journal adopted that policy 2 years ago.

Keith Baggerly, Ph.D., head of bioinformatics and computational biology at the University of Texas M.D. Anderson Cancer Center in Houston, said there was one simple reason for opening the data code to outside researchers: “Simple mistakes are very easy to make,” he said. His reanalysis of what turned out to be faulty Duke University data, which took him most of a year to obtain, led to the retraction of several papers. “What should NCI be looking for from us in clinical trials? It should be more than just a gene list,” he said. “It has to be a set of rules so that somebody else can check it before you start using it on patients.”

But few of the hundreds of papers published on predictive testing over the past decade have been aimed at clinical utility. Citing a recent review in the Journal of 16 studies claiming to establish proteomic biomarkers for variants of non-small-cell lung cancer (See J. Natl. Cancer Inst. 2010;102:1–11), Richard M. Simon, D.Sc., chief of the biometric research branch at NCI, said, “Most [studies] weren’t driven by people interested by intended use; they were driven by people interested in statistical modeling.”

Challenges Ahead

But academic researchers often lack the resources to carry out all the steps necessary to validate predictive genomic or proteomic analyses so they can be used in the clinic. “That takes a lot of work and a lot of time, and most academic investigators are not funded to do that. They have to look for a partnership with a really good diagnostics company,” Simon said.

And the ultimate test of a good predictive “omics” test—a clinical trial showing that stratification of patients for different treatment choices on the basis of test results led to better outcomes—is even more daunting and almost always requires teaming up with a deep-pocketed drug developer. Most oncology drug developments now are being developed with companion diagnostics. But the clinical trials seeking to prove the utility of such tests will require twice as many arms: one with a marker test and one without, where each arm gets the new drug or not. “This is going to take a lot more patients,” Simon said.

© Oxford University Press 2011. DOI: 10.1093/jnci/dj296