

Developing Precision Medicine in a Global World

Eric H. Rubin¹, Jeffrey D. Allen², Jan A. Nowak³, and Susan E. Bates⁴

Abstract

Advances in understanding the biology of cancer, as well as advances in diagnostic technologies, such as the advent of affordable high-resolution DNA sequencing, have had a major impact on the approach to identification of specific alterations in a given patient's cancer that could be used as a basis for treatment selection, and hence the development of companion diagnostics. Although there are now several examples of successful development of companion diagnostics that allow identification of patients who will achieve the greatest benefit from a new therapeutic, the path to coapproval of a diagnostic test along with a new therapeutic is complex and often inefficient. This review and the accompanying articles examine the current state of companion diagnostic development in the United States and Europe from academic, industry, regulatory, and economic perspectives.

See all articles in this CCR Focus section, "The Precision Medicine Conundrum: Approaches to Companion Diagnostic Co-development."

Clin Cancer Res; 20(6); 1419–27. ©2014 AACR.

Introduction

Concept and need for companion diagnostics

The concept of personalized medicine is based on the fundamental assumption that one or more molecular aberrations can be discovered as either etiologic or sustaining a given malignancy. Identifying the aberration in a given tumor will lead to therapies that are more selective (and therefore hopefully less toxic) than classical chemotherapeutics. The tests that identify molecular aberrations that are themselves targets of a specific therapeutic are among the tests considered companion diagnostics. As such, the tests have predictive value—in a clinical trial, the test ensures selection of a population of patients more likely to have a response to treatment, and in any individual patient increases the likelihood of success. Companion diagnostics are to be differentiated from prognostic tests—that infer outcome for a patient. For example, lactate dehydrogenase levels are prognostic in lymphoma—placing patients in a higher risk group. But they do not predict activity of any individual agent. Predictive and prognostic markers must also be distinguished from pharmacodynamic markers, tests that can show an impact on a particular oncologic pathway by a drug, and from pharmacologic markers, usually polymorphic variant testing, to determine susceptibility to drug toxicity. In general, pharmacodynamic mar-

kers have not yet joined the armamentarium of clinically useful tests in oncology.

Although the term companion diagnostics has been newly applied to a test "co-developed" to identify populations of patients who may benefit from a particular drug, the concept is not new. Nor is individualized selection of therapy for a patient based on a pathology report. What distinguishes the two in U.S. Food and Drug Administration (FDA) parlance is that the companion diagnostic provides a result that is "essential for the safe and effective use of a corresponding therapeutic product" (*italics added*; Text Box 1). In addition to identifying patients who are likely to benefit, the FDA definition of a common diagnostic also encompasses those tests that would identify patients who are at risk of adverse reactions, as well as tests that monitor response to treatment for the purpose of adjusting that treatment. FDA has taken the approach that the essential diagnostic should be "developed contemporaneously"; there is recognition that this may not always be possible.

The development of HER2 testing to accompany trastuzumab is considered the first example of a companion diagnostic from the standpoint of co-development. However, lessons learned in the development of tests for estrogen receptor (ER) positivity in breast cancer certainly informed development of HER2 testing (1, 2). Radioisotope-based receptor-binding assays were first described in 1973 for ER and were later superseded by immunohistochemical assays (3). Fourteen tests for ER assays can be found in the FDA Medical Devices listings. The first test for ER was approved in 1981, and the most recent approval was in 2013, based on data suggesting that the test was "substantially equivalent."

The FDA listing of approved companion diagnostics (reproduced in Table 1) includes 19 approvals, of which ten cover testing for HER2, three for EGF receptor (EGFR), two for B-RAF, and one each for ALK, KIT, RAS, and iron concentration. Examining the list of approved diagnostics

Authors' Affiliations: ¹Merck Research Laboratories, North Wales, Pennsylvania; ²Friends of Cancer Research, Arlington, Virginia; ³NorthShore University Health System Pathology and Laboratory Medicine, Evanston, Illinois; and ⁴Developmental Therapeutics Branch, National Cancer Institute, Bethesda, Maryland

Corresponding Author: Eric H. Rubin, Merck Research Laboratories, 351 North Sumneytown Pike, North Wales, PA 19454. Phone: 267-305-1717; Fax: 267-305-6402; E-mail: eric_rubin@merck.com

doi: 10.1158/1078-0432.CCR-14-0091

©2014 American Association for Cancer Research.

Text Box 1. FDA definition of companion diagnostic from 2011 draft guidance on *in vitro* companion diagnostic devices
<http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm262292.htm>

Definition and use of an IVD companion diagnostic device

An IVD companion diagnostic device is an *in vitro* diagnostic device that provides information that is essential for the safe and effective use of a corresponding therapeutic product.^a

The use of an IVD companion diagnostic device with a particular therapeutic product is stipulated in the instructions for use in the labeling of both the diagnostic device and the corresponding therapeutic product, as well as in the labeling of any generic equivalents of the therapeutic product.

An IVD companion diagnostic device could be essential for the safe and effective use of a corresponding therapeutic product to:

- Identify patients who are most likely to benefit from a particular therapeutic product^b
- Identify patients likely to be at increased risk for serious adverse reactions as a result of treatment with a particular therapeutic product
- Monitor response to treatment for the purpose of adjusting treatment (e.g., schedule, dose, discontinuation) to achieve improved safety or effectiveness

FDA does not include in this definition clinical laboratory tests intended to provide information that is useful to the physician about the use of a therapeutic product, but that is not a determining factor in the safe and effective use of the product.^c

^aGenerally, this means that the use of the IVD companion diagnostic device with the therapeutic product allows the therapeutic product's benefits to exceed its risks.

^bThis may include identifying patients in a specific population for which the therapeutic is indicated because there is insufficient information about the safety and effectiveness of the therapeutic product in any other population. An example is a therapeutic that is indicated only for patients who by virtue of the presence of a marker in tumor cells are believed to be unlikely to respond to other therapies.

^cExamples of such tests are commonly used and well-understood biochemical assays (e.g., serum creatinine or transaminases) used to monitor organ function. Note, however, that circumstances may occur when use of such tests, in the context of the therapeutic product, rises to an IVD companion diagnostic device level and approval or clearance for such use will be necessary. Note also that a novel IVD device providing information that is useful in, but not a determining factor for the safe and effective use of a therapeutic product, would not be considered an IVD companion diagnostic device.

(Continued in the following column)

Text Box 1 (Cont'd)

Ideally, a therapeutic product and its corresponding IVD companion diagnostic device would be developed contemporaneously, with the clinical performance and clinical significance of the IVD companion diagnostic device established using data from the clinical development program of the corresponding therapeutic product—although FDA recognizes that there may be cases when contemporaneous development may not be possible. An IVD companion diagnostic device that supports the safe and effective use of a particular therapeutic may be a novel IVD device (i.e., a new test for a new analyte), a new version of an existing device developed by a different manufacturer, or an existing device that has already been approved or cleared for another purpose.

for HER2, the complexity of the regulatory process is immediately understood. Furthermore, this list of FDA-approved companion diagnostics is dwarfed by the list of tests commercially available to clinicians for understanding the biology of a given tumor. These tests are not approved by the FDA, but are developed by laboratories that are Clinical Laboratory Improvement Amendments (CLIA) certified, and thus subject to regulation with regard to analytical validation. These tests, termed laboratory-developed tests (LDT), in some cases are also companion diagnostics (although without FDA approval); in others, they are prognostic or diagnostic tests that point the clinician toward one treatment scheme or another. The Tufts Medical Center Evidence-based Practice Center identified 112 gene-based tests alone for solid and hematologic tumors, 50 introduced between 2006 and 2001 (Table 2), and approximately 79 tests in development (4). These tests have entered the clinical use in the United States using the same regulatory pathways followed for immunohistochemical assays performed in commercial and academic pathology laboratories in the course of cancer diagnosis. As described by Parkinson and colleagues, diagnostic tests in cancer need to have analytical validity, clinical validity, and clinical utility (5).

One of the commercially available LDTs in Table 2 is that for ERCC1, a DNA repair protein. ERCC1 is one of a family of proteins with a critical role in at least four different pathways of DNA repair, including nucleotide excision repair needed for the removal of platinum adducts. High levels of ERCC1 in patient tumors have been associated with resistance to cisplatin, and some studies have suggested that cisplatin should not be used in treating tumors that express high levels of ERCC1, to avoid toxicity without benefit (6, 7). Ongoing clinical trials are testing whether chemotherapy choice can be altered on the basis of ERCC1 levels, with cisplatin omitted in the ERCC1-high cohorts. One prospective trial reported improved response rates with genotyping, without a survival advantage (8). However, a recent study casting doubt on the accuracy of the antibody used for ERCC1 detection (9) calls the analytical validity into

Table 1. FDA-approved companion diagnostic devices for oncology

Drug trade name (generic name)	NDA/BLA	Device trade name	PMA	Device manufacturer
Eribitux (cetuximab)	BLA 125084	Therascreen KRAS RGQ PCR Kit	P110030	Qiagen Manchester, Ltd.
Eribitux (cetuximab); Vectibix (panitumumab)	BLA 125084; BLA 125147	Dako EGFR PharmDx Kit	P030044 S001-S002	Dako North America, Inc.
Gilotrif (afatinib)	NDA 201292	Therascreen EGFR RGQ PCR Kit	P120022	Qiagen Manchester, Ltd.
Gleevec/Glivec (imatinib mesylate)	NDA 021335; NDA 021588	DAKO C-KIT PharmDx	P040011 S001-S002	Dako North America, Inc.
Herceptin (trastuzumab)	BLA 103792	INFORM HER2 NEU	P940004 S001	Ventana Medical Systems, Inc.
Herceptin (trastuzumab)	BLA 103792	PATHVISION HER2 DNA Probe Kit	P980024 S001-S010	Abbott Molecular, Inc.
Herceptin (trastuzumab)	BLA 103792	PATHWAY ANTI-HER2 NEU (4B5) Rabbit monoclonal primary antibody	P990081 S001-S016	Ventana Medical Systems, Inc.
Herceptin (trastuzumab)	BLA 103792	INSITE HER2 NEU KIT	P040030	Biogenex Laboratories, Inc.
Herceptin (trastuzumab)	BLA 103792	SPOT-LIGHT HER2 CISH Kit	P050040 S001-S003	Life Technologies, Inc.
Herceptin (trastuzumab)	BLA 103792	Bond Oracle Her2 IHC System	P090015	Leica Biosystems
Herceptin (trastuzumab)	BLA 103792	HER2 CISH PharmDx Kit	P100024 S001-S004	Dako Denmark A/S
Herceptin (trastuzumab)	BLA 103792	INFORM HER2 DUAL ISH DNA Probe Cocktail Assay	P100027 S001-S006	Ventana Medical Systems, Inc.
Herceptin (trastuzumab); Perjeta (pertuzumab)	BLA 103792; BLA 125409	HercepTest	P980018 S001-S017	Dako Denmark A/S
Herceptin (trastuzumab); Perjeta (pertuzumab)	BLA 103792; BLA 125409	HER2 FISH PharmDx Kit	P040005 S001-S009	Dako Denmark A/S
Mekinist (trametinib); Tafinlar (dabrafenib)	NDA 204114; NDA 202806	THXID BRAF Kit	P120014	bioMérieux, Inc.
Tarceva (erlotinib)	NDA 021743	Cobas EGFR Mutation Test	P120019	Roche Molecular Systems, Inc.
Xalkori (crizotinib)	NDA 202570	VYSIS ALK Break Apart FISH Probe Kit	P110012 S001-S003	Abbott Molecular, Inc.
Zelboraf (vemurafenib)	NDA 202429	COBAS 4800 BRAF V600 Mutation Test	P110020 S001-S006	Roche Molecular Systems, Inc.

Adapted from the FDA website on medical devices: <http://www.fda.gov/medicaldevices/productsandmedicalprocedures/invitrodiagnostics/ucm301431.htm>

Table 2. Commercially available genetic tests introduced between 2006 and 2011

		FDA approved
Breast	Breast profile	No
	deCODE Breast Cancer	No
	GeneSearch BLN Assay	Yes
	Her2 Neu overexpression	Yes
	Her2 Pro	No
	MammaPrint	Yes
	SPOT-Light HER2	Yes
	CISH Kit	
	Tamoxitest	No
	TOP2A FISH pharmDx Kit	Yes
Colorectal	BRAF mutation	No
	ColonSentry	No
	Colopath/ColorectAlert	No
	Cytokeratin 20 (CK20)	No
	KRAS mutation analysis	Yes ^a
	Oncotype DX colon cancer assay	Yes
	Septin-9 DNA methylation biomarker	No
	UGT1A1 Molecular Assay	Yes ^a
Genitourinary	ImmunoCyt/uCyt+	Yes
	NMP22BladderChek	Yes
Hematologic	G6PD	No
	Heme profile	No
	JAK2	No
	KIT Asp816Val mutation analysis	No
Lung	CellCorrect Kva-40 Lab Kit	No
	EGFR mutation analysis	Yes ^a
	ELSA-CYFRA 21-1	No
	ERCC1	No
	KRAS mutation analysis	No
MESOMARK	Yes	
Ovarian	OVA1	Yes
	OvaCheck	No
	OvaSure	No ^c
Prostate	Bayer Immuno 1Complexed PSA	Yes
	deCODE Prostate Cancer	No
	Hybritech free PSA test	Yes
	Progensa PCA3 Assay	No
	Prostate-63	No
	uPM3 test; PCA3Plus test	No

Table 2. Commercially available genetic tests introduced between 2006 and 2011 (Cont'd)

		FDA approved
Melanoma	Cobas 4800 BRAF V600 mutation test	Yes ^a
	THxID-BRAF Kit	Yes ^a
Other ^b	DakoCytomation's c-Kit (9.7) pharmDx	Yes
	LBA AFP-L3	Yes
	MGMT methylation testing	No
Multiple ^d	CellSearch	Yes
	CupPrint	No
	DPD deficiency	No
	EGFR assay	No
	miRview	No
	Pathwork Tissue of Origin test	Yes
	PI3K	No
	TheraGuide	No
Tumor profile	No	

^aTests that received FDA approval after 2011.

^b"Other" includes brain, liver, and upper gastrointestinal, respectively.

^cOvaSure sales halted after 2008 FDA review.

^dTests used for multiple cancers, including breast, colorectal, lung, ovarian, and prostate.

Adapted from Raman et al. (4).

question, despite its performance in a CLIA-approved laboratory. Both clinical validation (whether the test accurately predicts cisplatin resistance) and proof of clinical utility (whether knowing the result will allow selection of therapy) are as yet unproven, but remain under active investigation (10, 9).

Although many of the tests listed in Table 2 do not have the potential to become companion diagnostics, any discovery and validation of a new oncogenic target offers the potential for (i) a drug and (ii) a test. One example is the new discovery of activating mutations in MYD88, a Toll-like receptor, and interleukin-1 receptor adaptor protein, in Waldenstrom macroglobulinemia (11). Although therapies targeting MYD88 are not yet in the clinic, the identification of activating mutation has made the protein a target for drug development, and a companion diagnostic will be required. As shown in Table 3, data reported in 2012–2013 offer an unacceptably wide range of sensitivity, with a reported prevalence ranging from 70% to 100% of samples from patients with Waldenstrom macroglobulinemia (12). Any one of these tests could be commercialized and analytically validated by a CLIA-certified

Table 3. Variation in reported prevalence of L265P mutation in the MYD88 gene based on assay methodology (12)

Method	Source	WM	IgM MGUS	Reference
Sanger sequencing	BM CD19 ⁺	91%	10%	Treon and Hunter (12)
Sanger sequencing	BM	—	54%	Landgren and Staudt (25)
Sanger sequencing	—	70%	—	Ansell et al. (26)
PCR	BM	67%	—	Gachard et al. (27)
PCR	BM	79%	—	Poulain et al. (28)
Allele-specific PCR	BM/Lymph nodes	86%	87%	Jiménez et al. (29)
Allele-specific PCR	BM	72%	—	Mori et al. (30)
Allele-specific PCR	BM/PBMC	100%	47%	Varettoni et al. (31)
Allele-specific PCR	BM CD19 ⁺	93%	54%	Xu et al. (32)

Abbreviations: BM, bone marrow; IgM, immunoglobulin M; WM, Waldenstrom macroglobulinemia.

laboratory without proving the test accurately reflects the presence of the mutation in Waldenstrom macroglobulinemia (i.e., clinical validity) or that the test will affect outcome of treatment with a therapeutic (i.e., clinical utility). The development of a drug to target the L265P mutation of the *MYD88* gene will require a test that is not only proven to be analytically and clinically valid through the FDA approval process for companion diagnostics, but also demonstrate clinical utility to secure favorable reimbursement and optimal utilization.

One question that stands out in the discussion of diagnostics is the degree of regulation that is actually necessary. The papers in this *CCR Focus* section are aimed at understanding the current regulatory and reimbursement environment in the United States and in Europe, and the articles clearly show that companion diagnostics are under increasing scrutiny and regulatory control in both regions. Certainly, a breast cancer specimen incorrectly labeled HER2 negative could result in significant harm to a patient, but all tests are imperfect, and the availability of trastuzumab with an imperfect companion diagnostic has greatly benefited patients with breast cancer in general. The difficulty in settling on the most accurate test for HER2 points to the need for careful analytical validation and clinical validation (1, 13, 14). There exists a large body of evidence about laboratory errors. Some of the errors are "preanalytic," that is, due to the quality of sample obtained or how it was handled; some "postanalytic," due to reporting or interpretation. For pathology samples, where secondary review is considered some indicator of error, discrepancies have ranged from 1% to 15% (15). Quality assurance processes have been put into place to mitigate errors, but every new test introduces new opportunity for error. A study in 2007 reported that 20% of HER2 assays yielded incorrect results, with a major problem caused by variable tissue fixation (13).

Pathway to approval of a companion diagnostic

Companion diagnostics originate from preclinical hypotheses involving predictive biomarkers. One of the difficulties in development of companion diagnostics is that

a decision to invest in a companion diagnostic must typically be made before the predictive value of the biomarker, or the efficacy of the experimental therapeutic in a population defined by the diagnostic, is known. Initially, there must be conversion of an assay from a laboratory-based test (developed from cell lines or other preclinical models) to an assay that can be used on clinical specimens, such as surgical specimens or biopsies. The complexity of this transfer is often underappreciated by both the basic scientists who identified the potential predictive biomarker and the clinicians involved in designing the clinical studies that will use the predictive marker. Furthermore, the statistical rigor that would be applied to the evaluation of the biomarker in a clinical trial may not have been present in preclinical hypothesis testing, leading to a false-positive conclusion that a particular biomarker is predictive for drug efficacy.

As noted earlier, the FDA guidance states that "a companion diagnostic device can be *in vitro* diagnostic device or an imaging tool that provides information that is essential for the safe and effective use of a corresponding therapeutic product." Although interpretation of "essential for the safe and effective use of a corresponding therapeutic product" depends on the context of the patient population, analytical validity (the ability of the assay to accurately measure the biomarker) and clinical validity (the ability of the assay to accurately measure a clinical outcome of interest) do not depend on this context. Both analytical validity and clinical validity are typically assessed by 2 × 2 table analyses to identify sensitivity and specificity (and positive and negative predictive value).

Determination of whether an assay provides "information that is essential for the safe and effective use of a corresponding therapeutic product" relates to an assay's clinical utility. Clinical utility is discussed in detail in Parkinson and colleagues (5), and is often defined as an assessment of the use of the assay to improve patient outcomes (relative to the state of not using the assay). Because quantitative data are typically not available for this evaluation, clinical utility is often measured in a qualitative, subjective manner. For example, for a new drug that is much more effective than previous

treatments and with a safety profile that is no worse than previous treatments, it can be argued that a companion diagnostic that has a high negative predictive value (i.e., minimizes false negatives) could be considered to have acceptable clinical utility, even if the positive predictive value of the assay is not "high." To provide an example, based on original data from a registration trial that used an immunohistochemical assay to measure HER2 protein expression in breast cancer tissue, the negative and positive predictive value of the test for computed tomography-based response were 79% and 44%, respectively, for patients receiving the combination of trastuzumab and paclitaxel (from Herceptin label; http://www.accessdata.fda.gov/drugsatfda_docs/label/2000/trasgen020900lb.htm). Note that because only patients whose breast cancer tissue was scored as 2+ or 3+ were enrolled in this study (with +3 defined as "test positive"), 79% likely underestimates the true negative predictive value of the assay.

For assays with nonbinary outputs, the assay cutoff is another variable that affects clinical utility, and may be decided before registration studies are initiated, thus with some risk that performance of the assay cutoff in large populations may be poorly understood. For example, a cutoff selected on the basis of response rates might yield very different negative and positive predictive value related to overall survival, with overall survival data often not available before initiation of registration studies. Similarly, from a clinical utility perspective, a cutoff that is selected on the basis of a single-arm study of a new drug may perform quite differently in a randomized setting, where potential prognostic effects of the assay for the control arm (i.e., a standard-of-care treatment) may become apparent.

FDA and the European Medicines Agency approval pathways for companion diagnostics are discussed in detail in articles in this *CCR Focus* section (16–18). Similar to therapeutics, companion diagnostics are also the subject of health technology assessments of cost effectiveness (19). In cases of low prevalence of test-positive patients, the cumulative cost of screen failures can be considerable, adding to the overall cost of use of the therapeutic. Indeed, it is interesting to note that "personalized" medicine was originally envisioned as an approach to reduce cost and toxicity for patients, allowing ineffective treatments to be avoided. However, the reality is that there is can be an increase in cost at both the industry level, related to development of the companion diagnostic, and at the health care delivery level, where diagnostic tests are used to determine the most appropriate therapy. Additional complexity for patients and physicians can exist in the case where multiple different assays, perhaps with different cutoffs, are approved for drugs that have the same target but were developed by different companies.

As noted in Dr. Mansfield's contribution, the concept of coapproval of therapeutic product and companion diagnostics has been an important part of FDA's effort to promote implementation of personalized or precision medicine (17). The message to be gleaned is the recognition that laboratory testing and molecular diagnostic

testing, in particular, is complex, and although a co-development model has merit, it cannot accommodate all clinical situations in the same way. The European experience as described by Byron and colleagues and Pignatti and colleagues acknowledges that a broad range of companion diagnostic tests for a specific drug is currently allowed, without the need for concurrent development with the drug, and without a requirement for a specific, regulatory authority-approved test in the drug label (18, 19). Although that route may present some challenges in terms of regulatory oversight, it utilizes all the laboratory testing capabilities that are available to bring precision medicine to the individual patient.

Recently, the classical paradigm of "one biomarker-one test" has been changing, in part due to advances in technology, as well as to obviate the need for multiple biopsies to obtain sufficient tissue for multiple tests. When multiple tests are needed on one specimen, there may not be sufficient tissue available to perform all the tests. This is an issue most pressing with carcinoma of the lung, where multiple genetic tests may ultimately need to be performed to detect different actionable oncogenic mutations. This problem may be solved, in part, by the advent of next-generation DNA sequencing (NGS) platforms. One example of use of this technology is the Lung Master Protocol effort, which is a collaboration involving Friends of Cancer Research, FDA, National Cancer Institute's National Clinical Trials Network/Southwest Oncology Group, patient advocacy organizations, several pharmaceutical companies, as well as a diagnostic company (17).

In this project, a master protocol will govern how multiple drugs will be tested as potential treatments for squamous cell carcinoma of the lung. Each arm of the study will test a different drug that has been determined to target a unique genetic alteration(s). The use of NGS will help identify which patient is a molecular match to each arm. Using a common screening platform will reduce the amount of tissue that would be needed, as compared with if each substudy were being conducted as completely independent trials. The testing platform has previously demonstrated analytic validity and the trial results will help evaluate the clinical outcome produced by each drug/biomarker (20). In this case, one test will be able to provide the determination of which drug may be a beneficial option for patients based on the genetic characteristics of their tumor. Successful drug/biomarker pairs will be eligible for FDA approval and sponsors could pursue market authorization with the NGS platform as the companion diagnostic for the specific biomarker signature or perform the necessary bridging studies to validate the use of an alternative diagnostic tool. This will create a rapidly evolving infrastructure that can simultaneously examine the safety and efficacy of new drugs. This approach will have the ability to improve enrollment, enhance consistency, increase efficiency, and reduce costs.

Furthermore, the FDA recently approved the first diagnostics using NGS, setting the stage for use of similar diagnostics to detect multiple gene mutations in a single

biopsy specimen. The approved devices include methods used to diagnose cystic fibrosis (by detecting DNA mutations in the cystic fibrosis transmembrane conductance regulator gene) as well as the Illumina MiSeqDx platform, which allows development of NGS tests for any part of a patient's genome (21). Notably, the variables that govern the accuracy of complex, multigene tests are numerous and will require the same or greater attention that many of the single gene tests required (examined in detail in ref. 22). The variables include overall quality of the tissue specimen, fraction of tumor cells in the sample, and depth of sequencing, among others.

Projects like the Lung Master Protocol and marketing approval of NGS instruments demonstrate the continued expansion of the role diagnostic tools play in guiding therapy. As the biology of patients and tumors continue to be further understood, identifiable subpopulations of patients that are best fit for some treatment regimens will likely become smaller as compared with treating broad types of cancers with conventional cytotoxic drugs. Collaborative development consortia, like the Lung Master Protocol, and novel technology will be essential to support new, precision drug development in the future. Although this approach will increase the likelihood that patients that match a targeted therapy will positively respond, developing a drug with a companion diagnostic continues to present several challenges, some of which are highlighted next.

Current issues in the development of companion diagnostics

Key issues that often arise in companion diagnostic development include cutoff selection (for diagnostics with nonbinary measurements) and "prescreening" that can occur as a result of the availability of nonapproved tests used by patients and physicians to gain information about tumor genotype and phenotype.

About the cutoff selection, in many cases, there may be a monotonic relationship between a biomarker expression level and clinical outcome after treatment with the matching therapeutic, such as tumor shrinkage (Fig. 1). In these cases, setting a high cutoff for the test will favor positive predictive value at the expense of negative predictive value. Conversely, setting a low cutoff will favor negative predictive value at the expense of positive predictive value. From a payer perspective, a high cutoff might be favored because this would restrict the use of a new therapeutic to a patient population that would have a large clinical benefit. On the other hand, from an individual patient perspective, a low cutoff might be favored because this would increase the likelihood that the patient would be able to receive the new treatment, even if the chances of response were low (assuming that the treatment was associated with an acceptable safety profile).

Another issue with companion diagnostic development is that advances in technology have allowed increasing availability of genetic and phenotypic information about a patient's cancer at affordable prices (21). This creates tension between the patient and physician desire for infor-

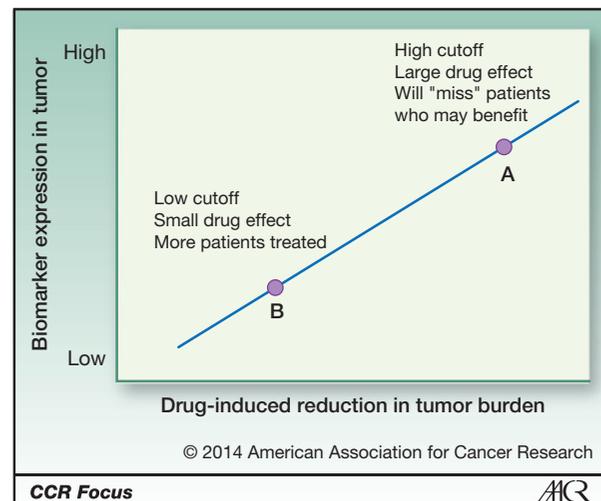


Figure 1. Example of the correlation between tumor biomarker expression and drug-induced reduction in tumor burden. Examples of two cutoff selections for test positive versus test negative are shown (A and B), with different implications for false-positive and false-negative rates, as well as differences in the percentage of patients who are deemed to be test positive.

mation that could be helpful in understanding unique aspects of a patient's cancer (even if that information is imperfect) versus regulatory concerns over harm that could result from decisions made from faulty tests. For example, a reporter from the *New York Times* had her DNA sequenced by three different companies, with discordant risk predictions identified for several diseases (23).

The availability of these kinds of tests can confound companion diagnostic development. For example, when patients are "prescreened" before being screened for enrollment into a study, using an assay that is different from the companion diagnostic, the study estimate of the effect of the new treatment can be biased (24). This bias is caused by discordance between the test used as a prescreen and the companion diagnostic test. For example, patients who would have scored positive on the companion diagnostic but not the prescreen test would not be entered into the study. FDA and industry alike discourage enrollment of prescreened patients in trials when possible, but with greater availability of direct-to-consumer genetic testing, this can be unavoidable.

Reimbursement for companion diagnostics

Molecular diagnostic tests are recognized for reimbursement purposes by specific current procedural terminology (CPT) codes. Through 2012, the CPT codes used for molecular tests were based on methodology (e.g., nucleic acid extraction, DNA amplification, separation by electrophoretic methods each had specific CPT codes) and, in that, differed from other laboratory tests for which codes are analyte specific. This system was very useful, in that, it could accommodate various technical approaches that could be applied to measure molecular analytes at a time when technology was rapidly evolving. Payers, however, balked at the lack of transparency of those codes and the need to

report multiple codes for single procedures. Consequently, a work group established by the American Medical Association (AMA) CPT Editorial Panel developed a replacement CPT coding system for molecular diagnostic tests that is exquisitely analyte specific and technology agnostic for more than one hundred of the most common molecular diagnostic tests [e.g., 81275 *KRAS* (*v-Ki-ras2* *Kirsten rat sarcoma viral oncogene*), gene analysis, variants in codons 12 and 13]. Other less common molecular tests, nearly 600, have been similarly defined and assigned codes in groups (tier 2), acknowledging to some extent the differences in technical and interpretive work needed to achieve clinically useful information. A similar AMA-sponsored work group is currently developing CPT codes to recognize various clinically defined tests that require NGS strategies. Although it is conceivable that such NGS tests can replace extant assays, the challenge will be in determining whether such an approach offers sufficient benefit in terms of cost efficiency and clinical utility to warrant recognition (and reimbursement) as CPT-defined procedures.

Establishing reimbursement for the refined molecular codes has been problematic. Rather than following the usual process for determining procedure values through the AMA Relative Value Update Committee-mediated process or via the cross-walk process for laboratory tests on the Clinical Laboratory Fee Schedule, CMS opted to leave the responsibility for pricing the new codes, to the local Medicare administrative contractors, using a process known as "gap fill." They were neither prepared to determine the pricing nor had the resources to make fair and informed pricing decisions or coverage policies. As a consequence, none of the tests were priced at the beginning of 2013 so claims were not paid. With completion of the gap fill process, some molecular tests are now being covered but many are not. None of the tier-2 codes have had pricing determined. Furthermore, the failure to price the codes has led many payers to misinterpret this as a reason for noncoverage. These problems with reimbursement rates and policies imperil the existence of many smaller molecular diagnostic laboratories.

The lauded transparency of the new molecular codes will be problematic for clinicians who try to use molecular diagnostic assays in situations outside those recognized for specific companion diagnostics, leading to denial of payment. For example, *KRAS* codon 12 and 13 mutation testing is indicated for consideration of treatment of stage 4 colorectal cancer with targeted therapies. Mutations in *KRAS* codons 61 and 146 are also thought to be relevant but testing for them would probably not be covered. Similarly, *KRAS* mutation status is not immediately relevant for non-small cell lung carcinoma in choosing a therapy, but it is very informative in that *KRAS* mutations are mutually exclusive of other significant driver mutations in *EGFR*, *ALK*, and *ROS-1*.

Conclusions

The companion diagnostic model has been adopted in the United States, the European Union, and elsewhere in the world as an approach to improve the efficacy of therapy

administered to patients by improving the accuracy of the tests used to select particular tumors for therapy. Although some of these efforts have been successful, and a model for future drug development, it is also important to understand the weaknesses of the co-development model. Although many quickly point to the Herceptin/HerceptTest combination as a model, the difficulties in establishing uniform *HER2* testing in the United States and elsewhere are notable, and although the notion that *HER2* protein "overexpression" seems a reasonable marker for tumors that might benefit from the drug, the eventual recognition that another "biomarker," *HER2* gene amplification might be a better test, should give us pause (14).

In 2004, a similar approach was applied to selecting patients whose tumors might respond EGFR monoclonal antibody therapy (erbitux). An immunohistochemical assay (DakoCytomation EGFR pharmDx; Table 1) was approved as the companion diagnostic to detect overexpression of EGFR. In 2008, reports emerged that *KRAS* codon 12 and 13 mutation predicted resistance to EGFR inhibitor, and in 2009, FDA issued a labeling change for the drug requiring a molecular determination of *KRAS* mutation status. An FDA-approved "companion diagnostic" test for *KRAS* was not available until July 2012, with the difference in time frame relative to the labeling change reinforcing the point that development works most efficiently if trials for both diagnostic and drug are carried out simultaneously. It is not always possible to know *a priori* what the best biomarker is. They are chosen, hopefully after extensive debate, based on the best knowledge and understanding, but always with an awareness that there will be new discoveries over the horizon.

The questions begging to be addressed and that are discussed in this *CCR Focus* section include the following:

How are companion diagnostic tests best developed (before being approved and commercialized) and how do they evolve?

Who performs clinical testing when there is no FDA-approved companion diagnostic for a biomarker?

How is clinical utility best determined?

How can regulatory processes be harmonized to prevent duplication of effort but still allow innovation?

How should reimbursement considerations impact the development of new companion diagnostics?

Disclosure of Potential Conflicts of Interest

E.H. Rubin is an employee of Merck & Co. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: E.H. Rubin, S.E. Bates

Writing, review, and/or revision of the manuscript: E.H. Rubin, J.D. Allen, J.A. Nowak, S.E. Bates

Acknowledgments

The authors thank Mark Raffeld for advice and review of the article and Robert Robey for editorial assistance.

Received January 13, 2014; accepted January 24, 2014; published online March 14, 2014.

References

- Gown AM. Current issues in ER and HER2 testing by IHC in breast cancer. *Mod Pathol* 2008;21 Suppl 2:S8–S15.
- Fitzgibbons PL, Murphy DA, Hammond ME, Allred DC, Valenstein PN. Recommendations for validating estrogen and progesterone receptor immunohistochemistry assays. *Arch Pathol Lab Med* 2010; 134:930–5.
- Allred DC, Carlson RW, Berry DA, Burstein HJ, Edge SB, Goldstein LJ, et al. NCCN Task Force Report: Estrogen Receptor and Progesterone Receptor Testing in Breast Cancer by Immunohistochemistry. *J Natl Compr Canc Netw* 2009;7 Suppl 6:S1–S21.
- Raman G, Wallace B, Patel K, Lau J, Trikalinos TA. Update on Horizon Scans of Genetic Tests Currently Available for Clinical Use in Cancers. Technology Assessment Report. Tufts Evidence-based Practice Center 2011. Available from: www.cms.gov/Medicare/Coverage/DeterminationProcess/downloads/id81TA.pdf.
- Parkinson DR, McCormack RT, Keating SM, Gutman SI, Hamilton SR, Mansfield EA, et al. Evidence of clinical utility: an unmet need in molecular diagnostics for patients with cancer. *Clin Cancer Res* 2014;20: 1428–44.
- Reed E. ERCC1 measurements in clinical oncology. *N Engl J Med* 2006;355:1054–5.
- Altaha R, Liang X, Yu JJ, Reed E. Excision repair cross complementing-group 1: gene expression and platinum resistance. *Int J Mol Med* 2004;14:959–70.
- Cobo M, Isla D, Massuti B, Montes A, Sanchez JM, Provencio M, et al. Customizing cisplatin based on quantitative excision repair cross-complementing 1 mRNA expression: a phase III trial in non-small-cell lung cancer. *J Clin Oncol* 2007;25:2747–54.
- Friboulet L, Olaussen KA, Pignon JP, Shepherd FA, Tsao MS, Graziano S, et al. ERCC1 isoform expression and DNA repair in non-small-cell lung cancer. *N Engl J Med* 2013;368:1101–10.
- Besse B, Olaussen KA, Soria JC. ERCC1 and RRM1: ready for prime time? *J Clin Oncol* 2013;31:1050–60.
- Hunter Z, Xu L, Yang G, Zhou Y, Liu X, Cao Y, et al. The genomic landscape of Waldenström's Macroglobulinemia is characterized by highly recurring MYD88 and WHIM-like CXCR4 mutations, and small somatic deletions associated with B-cell lymphomagenesis. *Blood* 2013 Dec 23. [Epub ahead of print].
- Treon SP, Hunter ZR. A new era for Waldenström macroglobulinemia: MYD88 L265P. *Blood* 2013;121:4434–6.
- Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Arch Pathol Lab Med* 2007; 131:18–43.
- Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: american society of clinical oncology/college of American pathologists clinical practice guideline update. *Arch Pathol Lab Med* 2014;138: 241–56.
- Raab SS, Grzybicki DM. Quality in cancer diagnosis. *CA Cancer J Clin* 2010;60:139–65.
- Senderowicz AM, Pfaff O. Similarities and differences in the oncology drug approval process between FDA and European Union with emphasis on *in vitro* companion diagnostics. *Clin Cancer Res* 2014;20:1445–52.
- Mansfield EA. FDA perspective on companion diagnostics: an evolving paradigm. *Clin Cancer Res* 2014;20:1453–7.
- Pignatti F, Ehmann F, Hemmings R, Jonsson B, Nuebling M, Papaluca-Amati M, et al. Cancer drug development and the evolving regulatory framework for companion diagnostics in the European Union. *Clin Cancer Res* 2014;20:1458–68.
- Byron SK, Crabb N, George E, Marlow M, Newland A. The health technology assessment of companion diagnostics: experience of NICE. *Clin Cancer Res* 2014;20:1469–76.
- Frampton GM, Fichtenholtz A, Otto GA, Wang K, Downing SR, He J, et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat Biotechnol* 2013;31:1023–31.
- Collins FS, Hamburg MA. First FDA authorization for next-generation sequencer. *N Engl J Med* 2013;369:2369–71.
- Simon R, Roychowdhury S. Implementing personalized cancer genomics in clinical trials. *Nat Rev Drug Discov* 2013;12:358–69.
- Peikoff K. I had my DNA picture taken, with varying results. *New York Times* 2013. Available from: www.nytimes.com/2013/12/31/science/i-had-my-dna-picture-taken-with-varying-results.html?_r=0.
- Pennello GA. Analytical and clinical evaluation of biomarkers assays: when are biomarkers ready for prime time? *Clin Trials* 2013;10:666–76.
- Landgren O, Staudt L. MYD88 L265P somatic mutation in IgM MGUS. *N Engl J Med* 2012;367:2255–6.
- Ansell SM, Secreto FJ, Manske M, Braggio E, Hodge LS, Price-Troska T, et al. MYD88 Pathway Activation in Lymphoplasmacytic Lymphoma Drives Tumor Cell Growth and Cytokine Expression. *ASH Annual Meeting Abstracts* 2012;120:2699.
- Gachard N, Parrens M, Soubeyran I, Petit B, Marfak A, Rizzo D, et al. IGHV gene features and MYD88 L265P mutation separate the three marginal zone lymphoma entities and Waldenström macroglobulinemia/lymphoplasmacytic lymphomas. *Leukemia* 2013;27: 183–9.
- Poulain S, Roumier C, Decambron A, Renneville A, Herbaux C, Bertrand E, et al. MYD88 L265P mutation in Waldenström macroglobulinemia. *Blood* 2013;121:4504–11.
- Jiménez C, Sebastián E, Chillón MC, Giraldo P, Mariano Hernández J, Escalante F, et al. MYD88 L265P is a marker highly characteristic of, but not restricted to, Waldenström's macroglobulinemia. *Leukemia* 2013;27:1722–8.
- Mori N, Ohwashi M, Yoshinaga K, Mitsuhashi K, Tanaka N, Teramura M, et al. L265P mutation of the MYD88 gene is frequent in Waldenström's macroglobulinemia and its absence in myeloma. *PLoS ONE* 2013;8:e80088.
- Varettoni M, Arcaini L, Zibellini S, Boveri E, Rattotti S, Riboni R, et al. Prevalence and clinical significance of the MYD88 (L265P) somatic mutation in Waldenström's macroglobulinemia and related lymphoid neoplasms. *Blood* 2013;121:2522–8.
- Xu L, Hunter ZR, Yang G, Zhou Y, Cao Y, Liu X, et al. MYD88 L265P in Waldenström macroglobulinemia, immunoglobulin M monoclonal gammopathy, and other B-cell lymphoproliferative disorders using conventional and quantitative allele-specific polymerase chain reaction. *Blood* 2013;121:2051–8.