# Leveling the Playing Field: Bringing Development of Biomarkers and Molecular Diagnostics up to the Standards for Drug Development

George Poste[1], David P. Carbone[2], David R. Parkinson[3], Jaap Verweij[4,5], Stephen M. Hewitt[6], and J. Milburn Jessup[7]

## Abstract

Molecular diagnostics are becoming increasingly important in clinical research to stratify or identify molecularly profiled patient cohorts for targeted therapies, to modify the dose of a therapeutic, and to assess early response to therapy or monitor patients. Molecular diagnostics can also be used to identify the pharmacogenetic risk of adverse drug reactions. The articles in this *CCR Focus* section on molecular diagnosis describe the development and use of markers to guide medical decisions regarding cancer patients. They define sources of preanalytic variability that need to be minimized, as well as the regulatory and financial challenges involved in developing diagnostics and integrating them into clinical practice. They also outline a National Cancer Institute program to assist diagnostic development. Molecular diagnostic clinical tests require rigor in their development and clinical validation, with sensitivity, specificity, and validity comparable to those required for the development of therapeutics. These diagnostics must be offered at a realistic cost that reflects both their clinical value and the costs associated with their development. When genome-sequencing technologies move into the clinic, they must be integrated with and traceable to current technology because they may identify more efficient and accurate approaches to drug development. In addition, regulators may define progressive drug approval for companion diagnostics that requires further evidence regarding efficacy and safety before full approval can be achieved. One way to accomplish this is to emphasize phase IV postmarketing, hypothesis-driven clinical trials with biological characterization that would permit an accurate definition of the association of low-prevalence gene alterations with toxicity or response in large cohorts. *Clin Cancer Res; 18(6); 1515–23. ©2012 AACR.*

## Introduction

This is an especially exciting time in the field of medicine, created by the application of the burgeoning knowledge about the molecular and cellular biology of cancer to improve diagnosis and therapy. Recent results obtained through The Cancer Genome Atlas [TCGA (1–3)] and the Therapeutically Applicable Research to Generate Effective Treatments [TARGET (4–7)] programs, as well as independent research programs (8–10), have identified a large number of candidate molecular alterations in different cancers that represent both potential markers for novel diagnostics and targets for a new generation of oncology drugs. This set of *Focus* articles concentrates on various aspects surrounding the identification, validation, and introduction into regular practice of clinically useful molecular diagnostics. Success will demand unprecedented levels of collaboration between clinicians and clinical laboratory scientists in the design and monitoring of clinical trials. Challenges include the need to validate both the clinical and analytical performance of the diagnostic (11); the need to standardize preanalytical variables during specimen collection, stabilization, and processing (12); the need to pay rigorous attention to the analytical performance and validation of the assay (13); and the need to meet regulatory requirements (14). The first five articles in this series are a testament to the free-market approach to molecular diagnostics that characterizes the U.S. system of medical innovation. However, Andre and colleagues (15) describe an alternative approach in which a government entity [in this case, the French National Cancer Institute [NCI]] contributes to the development of molecular markers and ensures the high quality of molecular diagnostics by establishing

**Authors' Affiliations:** [1]Complex Adaptive Systems Initiative, Arizona State University, Scottsdale, Arizona; [2]Thoracic/Head and Neck Program and Thoracic Oncology Center, Vanderbilt-Ingram Cancer Center, Nashville, Tennessee; [3]Nodality, Inc., South San Francisco, California; [4]Department of Medical Oncology and [5]Daniel den Hoed Cancer Center, Erasmus University Medical Center, Rotterdam, The Netherlands; [6]Tissue Array Research Program, Laboratory of Pathology, Advanced Technology Center, National Cancer Institute, Gaithersburg; [7]Cancer Diagnosis Program, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, Maryland

**Note:** The opinions expressed in this article are those of the authors and do not necessarily represent those of the National Cancer Institute, the National Institutes of Health, or the Department of Health and Human Services.

**Corresponding Author:** J. Milburn Jessup, Diagnostics Evaluation Branch, Cancer Diagnosis Program, Division of Cancer Treatment and Diagnosis, National Cancer Institute, EPN 6040, 6130 Executive Boulevard, Rockville, MD 20892-7420. Phone: 301-435-9010; Fax: 301-402-7819; E-mail: jessupj@mail.nih.gov

**doi:** 10.1158/1078-0432.CCR-11-2206

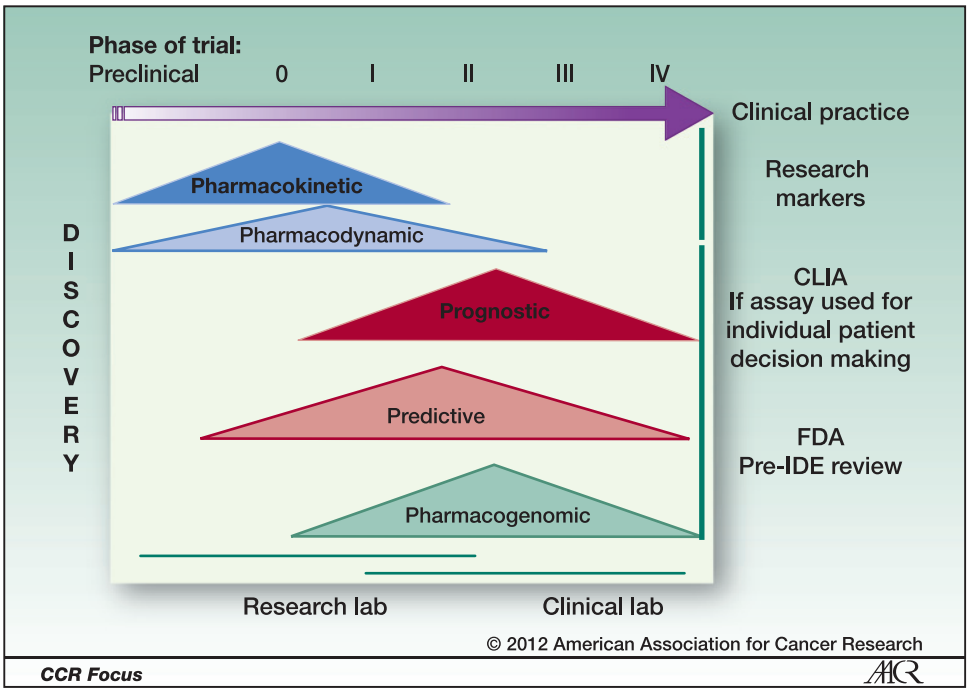


Figure 1. Types of markers and phases of clinical trials. Five types of markers are used in clinical research, as diagrammed within the context of the phase of clinical trial in which they are used. Pharmacokinetic and pharmacodynamic markers are used as research assays and are generally not used to guide medical decisions for patients in the trial. In contrast, the other 3 markers (prognostic, predictive, and pharmacogenetic) are often used for medical decision-making and therefore need to be performed in a CLIA-certified laboratory. Markers that are used for medical decision-making are also likely to require an FDA pre-IDE review. See text for further descriptions of the types of markers and phases of trials, and the CLIA and FDA requirements. IDE, Investigational Device Exemption.

standard operating procedures and distributing certified reagents throughout the country. In this work we define the different types of markers that comprise the contemporary universe of new molecular diagnostics, and emphasize the fundamental importance of standardized analytical parameters. We also suggest a reduction in phase III trials and propose metrics by which the quality of diagnostic development can be assessed by the regulatory agencies.

## Categories of Markers in Oncology

A biomarker is defined as a "characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" (16). Generally, basic and clinical investigators use the terms "molecular diagnostic," "marker," and "biomarker" interchangeably. There are 5 types of molecular diagnostics that have different functions and are usually associated with separate phases of clinical trials. Clinical research generally involves phase 0–III clinical trials, and phase IV trials are postmarket pharmacovigilance surveys that assess drug safety in large population sets and potential efficacy for new indications. In contrast, phase 0 trials are trials in which pharmacokinetic and pharmacodynamic effects are evaluated to enable comparisons with an investigational agent's behavior in preclinical models. The dosing in phase 0 trials is low enough that the biological effect of an agent on its intended target can be detected without risk of serious adverse events. Phase I trials are designed to determine the maximum tolerated dose of an agent with respect to its safety and toxicity profile without establishing efficacy in patients within a particular type of cancer. Phase II trials are larger studies in which an agent is dosed generally at or close to the maximum tolerated dose, and investigators can begin to ascertain the efficacy of the agent in different malignant diseases. Phase III trials are large, randomized trials that are intended to confirm the efficacy of an agent (alone or in combination) in a particular disease context compared with the standard-of-care therapy.

Five basic types of markers are used in clinical trials, often with phase-specific purposes (Fig. 1). Pharmacokinetic markers assess the absorption, distribution, metabolism, and excretion of a drug or agent, whereas pharmacodynamic markers measure how an agent affects its intended target. These markers are generally used in early-phase trials as drug development tools to correlate biological responses to agents with observable clinical effects in individual patients (Fig. 1). They are considered part of clinical research because, in general, the results are not provided to the patients or their attending physicians and are not used for medical decision-making. Although these markers do not need to be tested in a clinically certified laboratory, they play an important role in early drug development and should be analyzed with appropriately high standards of technical rigor, and should generally meet Good Laboratory Practice standards as defined by the U.S. Food and Drug Administration [FDA (17, 18)]. These markers can be useful for identifying or confirming the mechanism of action of an agent, and relating both mechanistic and off-target effects to drug dose and schedule. However, a recent trial sponsored by the National Cancer Institute (NCI) used a pharmacokinetic marker for medical decision-making. The trial investigators assessed the response of individual patients to a test dose of busulfan and then tailored the therapeutic dose of the drug in each patient to the kinetics of the production of

metabolites in a fixed period of time. Because this assay is now used for medical decision-making, it must be performed in a clinically certified laboratory.

Markers that are associated with survival or other clinical endpoints independently of any specific treatment are classified as prognostic markers. Such markers may also be useful for monitoring the response of patients to therapy (e.g., by enumerating circulating tumor cells in patients with colon, breast, or prostate carcinoma). Markers that are associated with a clinical endpoint (e.g., survival) but also assess the effectiveness of a particular treatment are designated predictive markers because they predict the response to that treatment (Fig. 1). Some predictive markers can identify individual patients who are more likely to respond to a particular drug, and thus can be used to select patients for therapy. For example, EML-ALK mutations in patients with non–small cell lung cancer (NSCLC) are considered a basis for the use of crizotinib. The final class of markers consists of pharmacogenomic markers that are used to identify the risk of organ-based toxicities or altered metabolism and/or responses to therapeutic agents (Fig. 1). These markers are usually inherited in the germ line and are typically nonsynonymous single nucleotide polymorphisms (SNPs) or genomic variants that involve more than one nucleotide.

### The need for clinical laboratory accreditation

Markers that are used for clinical decision-making within clinical trials are often termed integral markers because they usually are essential for the performance of the trial (see Schilsky and colleagues (11) elsewhere in this series). These markers may be prognostic, predictive, pharmacogenomic, or (occasionally) pharmacodynamic. Integral markers may be used to determine patients' eligibility for a trial and their assignment to therapy, to guide dose selection, and to stratify patients within a trial. Because in these cases the test result is known to the physician and patient, and influences clinical practice, the assay must be performed in a Clinical Laboratory Improvement Amendments (CLIA)–certified laboratory as stipulated in Federal Acquisition Regulation section 42 CFR 493 (19). In the United States, any patient-derived specimen that is analyzed for clinical decision-making must be performed in a CLIA-certified laboratory under an accreditation process managed by the Center for Medicaid and Medicare Services (CMS). These regulations apply to both clinical research and practice. The patients and/or their physicians will be informed of a test result or they may be informed indirectly by an event such as assignment to a particular treatment based on the result of the diagnostic test. The intent of these regulations is to ensure that the highest standards for reproducibility and reliability of assays performed on patients are being met, which will in turn facilitate optimal medical decision-making. Integral markers should be distinguished from integrated markers, which are used in all patients in a trial or in a predefined subset of patients but are not used for medical decision-making. Instead, such markers are intended for clinical research or development of a marker and its assay for use in a subsequent trial [see Williams and colleagues (13) in this issue].

## Differences between Discovery and Clinical Research, and their Impact on Treatment

Whereas integral markers need to be performed in CLIA-certified laboratories, studies for discovery or correlative research in which the patients or their physicians will not learn the test results are often performed in a research laboratory where the standards for reproducibility and reliability are not as closely regulated (Table 1). Apart from the absence of a need to inform patients about test results, several other factors may prevent the direct use of discovery research data in the clinic. Clinical research and studies for discovery often use different sources for specimens. For instance, TCGA uses fresh-frozen specimens that are collected in a way that minimizes damage to DNA, RNA, and proteins before the specimen and its analytes are assessed. In contrast, clinical specimens are collected in busy clinical settings such as interventional radiology or operating suites,

**Table 1.** Differences between discovery and clinical marker development

| | Discovery (TCGA) | Clinical (hospital) |
|---|---|---|
| Preanalytic | | |
| Samples | Fresh-frozen<br>Samples of convenience<br>(bias by site) | FFPE<br>Samples of convenience<br>(bias by trial or locale) |
| Analytic | | |
| Approach | Unbiased | Focused on analyte in tissue/matrix |
| Power | Detect variant | Reduce false negatives/ false positives |
| Runs | Batch mode | Each test is independent |
| Sensitivity/ specificity | May have 10%+ FDR | Requires high sensitivity/specificity |
| Risk to patient | Null to low | High |
| Postanalytic | | |
| | Not regulated | Regulated by FDA and CMS |

Discovery research is oriented toward the identification of novel markers that may be important drivers or other indicators of malignant transformation and/or neoplastic progression. In contrast, in clinical marker development, an assay is developed for a marker that may be an important therapeutic target or marker associated with clinical outcome. This assay must be reliable, reproducible, and performed in a clinical laboratory under federal regulations. This table summarizes many of the differences between discovery and clinical marker development in terms of preanalytic, analytic, and postanalytic factors.

or from biopsy procedures in which the time to transport specimens to a pathology laboratory for fixation and stabilization of the specimen is longer and subject to considerable variation between settings compared with a dedicated research setting. Clinical specimens are therefore likely to undergo more degradation than impeccably curated research specimens. More information on the importance of preanalytic variables is provided by Hewitt and colleagues (12) elsewhere in this series.

Differences in preanalytic factors may have a profound influence on the identification and reproducibility of putative markers (20, 21). Studies for discovery are just what the name implies: a search for gene or protein alterations or variants that are not otherwise predefined. In contrast, in clinical research involving integral markers, the markers are predefined and must use assays whose analytical performance is validated. There is minimal risk to the patient in discovery protocols because the specimens are deidentified and assay results are not provided to the specimen donor. In contrast, clinical studies in which an assay generates false-positive or -negative results can result in exposure to a potentially toxic treatment or inappropriate denial of therapy, respectively. Protocols for discovery research are often unbiased and use sufficient numbers of samples with the statistical power to detect a variant at a certain level of prevalence with a predefined false discovery rate. In contrast, integral markers are predefined, with assays that have been validated to have high sensitivity and specificity and concomitant low false-negative and -positive rates. An additional and important difference between assays for discovery research and integral assays is that discovery assays are typically performed in batches, whereas integral clinical assays are performed when samples are received by the clinical laboratory and typically are conducted individually or in small groups. This increases the interassay variability and requires greater controls to ensure that the inter-run variability is minimized. Clinical assays for medical decision-making thus face greater challenges in terms of reliability and reproducibility than discovery-stage assays. Nevertheless, it is critical to ensure that discovery research data are of sufficiently high quality to justify the time and cost required to establish and validate a clinical assay. A common cause of failure to confirm the clinical usefulness of a marker identified in discovery research is that insufficient rigor was applied in the original identification of the marker. The principles set forth by McShane and colleagues (21) [Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK)] are critical to ensure that discovery markers meet a high level of rigor before any translation into a clinical diagnostic is contemplated.

## Why Clinicians and Translational and Clinical Scientists Need to Understand the Principles of Analytical and Clinical Validity and How They Relate to Clinical Utility

Discovery research often involves the use of advanced technologies to search for genetic or protein alterations that may be useful targets for therapy or correlate with response to particular therapies. The requirement for sufficient statistical power to identify various levels of gene or protein variation should be addressed in the sample set. Once candidate variants have been identified, confirmation will require larger datasets, such as those provided by the TCGA and the TARGET programs, or similarly sized private-sector efforts. These larger-scale efforts are necessary not only to confirm the initial findings but also to establish the prevalence of the molecular marker in a sample set from the intended clinical-use population.

Using these approaches, investigators have made several important findings that are leading to the development of new therapeutics or their clinical assessment in cohorts of patients identified by new molecular diagnostic clinical assays, such as IDH1/2 mutations in glioblastoma multiforme (22) and Jak mutations in acute lymphoblastic leukemia (6). Unfortunately, it can take as long as 2 to 5 years to develop and validate clinical assays to measure these mutations in a sufficiently large number of clinical samples. Before a discovery candidate can be converted to a validated clinical assay, the analytical performance of the assay must meet certain standards, as defined by Westgard (23), Lee and colleagues (24), and Wagner (25). The FDA has provided guidance for satisfactory analytical performance of molecular diagnostics that includes suitable controls and analyses of inter- and intrareproducibility tests (26). The Cancer Diagnosis Program of the NCI provides downloadable templates that address the validation of analytical performance of FISH and immunohistochemistry assays (27). These issues are further discussed by Schilsky and colleagues (12) elsewhere in this section.

The analytical validity of an assay (i.e., its ability to detect an analyte when it is present and not when it is absent) is the primary focus of diagnostic laboratory accreditation. As clinical molecular diagnostic tests evolve to play a greater role in guiding treatment decisions, establishing the clinical validity of the assay will become as important as determining its analytical validity. Regulatory agencies have indicated their intent to focus more on significant-risk tests, i.e., those associated with medical clinical decisions for individual patients. To establish clinical validity, investigators much show that a positive test result is associated with a particular clinical endpoint or event, and a negative test result is not. In this context, the clinical endpoint may be related to either survival (i.e., overall survival or a progression-free interval, or some other evaluation of the kinetics of disease progression) or response to therapy [e.g., meeting a prespecified reduction in tumor burden according to Response Evaluation Criteria in Solid Tumors (RECIST)]. To show clinical validity, the assay must characterize the analyte(s) in specimens collected within the clinical context of intended use, which may range widely. As the assay progresses from discovery to development, the samples must be obtained and processed in a manner similar to what will occur in routine clinical practice. For example, if the marker predicts a positive response to a therapy [e.g., mutation in the epidermal growth factor receptor (EGFR) in

NSCLC for treatment with erlotinib], a positive assay must associate consistently with a beneficial clinical response. These results can then be translated into clinical sensitivity and specificity, and may be used to define cutoffs by means of receiver operating characteristic curves (28) or other analyses (29, 30).

Finally, before payors will reimburse for a diagnostic marker and clinical practitioners can incorporate it into routine practice, investigators must show that the marker has high clinical utility. Measures of clinical utility abound (31–34), but clinical utility is not clearly defined other than that the benefits of the diagnostic must outweigh its risks (35). Various formulas for estimating clinical utility have been proposed (32–35) and used in technology assessment programs by organizations such as the National Institute for Health and Clinical Excellence [NICE (36)] in the UK, and the Center for Medical Technology Policy [CMPT (37)], the Agency for Healthcare Research and Quality [AHRQ (38)], the Centers for Disease Control and Prevention's Evaluation of Genomic Applications in Practice and Prevention [eGAPP (39)], and the Blue Cross Blue Shield Technology Evaluation Center [BCBS TEC (40)] in the United States. These organizations follow predefined criteria in an effort to assess the utility of a test along with its associated companion therapeutic(s) in an unbiased way. The BCBS TEC provides clear criteria on its website and assesses the utility of treatments and diagnostics in light of the following parameters: (i) the technology must be approved by the appropriate governmental regulatory bodies; (ii) the scientific evidence must enable conclusions about the effect of the technology on health outcomes; (iii) the technology must improve the net health outcome; (iv) the technology must be as beneficial as established alternatives; and (v) the improvement must be expected to be attainable outside clinical trial research investigations. These criteria appear to be similar to those used by the NICE and CMPT; however, the BCBS TEC also will consider reviewing topics that do not fulfill all of the criteria. For example, in the case of new diagnostic tests for oncology, the BCBS TEC supports the use of KRAS testing in patients with metastatic colorectal carcinoma when treatment with anti-EGFR antibodies is considered. The BCBS TEC originally ruled against the use of erlotinib in NSCLC patients with mutations in EGFR, but reversed its decision after analyzing nonrandomized clinical data. Although other organizations (e.g., the American Society of Clinical Oncology, and the National Comprehensive Cancer Network) publish guidelines for clinical practice, practitioners may need to track the reports provided by evaluation committees from the BCBS TEC, NICE, AHRQ, and CMPT because their reviews include the concerns of the payors who will be providing critical reimbursements for both the diagnostic test and the companion treatment. These assessments also often identify shortcomings in the development of a diagnostic or treatment in terms of the science involved, so that practitioners can understand the boundaries of the scientific evidence. For example, a BCBS TEC review (41) of the use of erlotinib in EGFR-mutated advanced NSCLC raised questions about whether the diagnostic test would serve solely as a predictor of therapeutic response or as a prognostic factor, and about the sensitivity and specificity of the test for predicting a response to an agent.

## The Need for International Biobanks and Diagnostic Databases

Next-generation genome sequencing is uncovering extensive genetic variants of uncertain significance in diverse cancers. As noted by Andre and colleagues (42), there is an essential need to identify clinically useful molecular and imaging markers in a registry or database so that these diagnostics can be identified by both clinicians and patients. Currently, several websites [e.g., the NCI site (43) and Vanderbilt University's "My Cancer Genome" (44)] track a number of diagnostics that are used to assess patients for eligibility for targeted therapies. As the assessment of genomic alterations in a given disease state becomes the standard of care, it will be important to integrate the reported findings, and their implications, into the electronic medical record systems used by attending clinicians. In addition, investigators are now conducting many clinical trials that require patients with relatively rare genetic abnormalities (often with a $\leq$1% prevalence), so the ability to match patients with these specific lesions with relevant clinical trials (institutionally, nationally, and internationally) will be of great value. This approach is already part of patient-oriented sites such as My Cancer Genome.

Although it is relatively straightforward to use standard gene and protein names such as those provided by the Human Genome Organization (45), this only allows indexing of the root gene or protein name. The creation of precise and standardized ontologies and semantics for validated alterations or variants of genes or proteins is more complex. The databases for these alterations should format content in consistent ways, perhaps by following the guidance of the Human Genome Variation Society (46). Institutions are still struggling with the poor state of nomenclature standardization. An important consideration for the future will be to develop standardized vocabularies and ontologies for the molecular targets, modules, and pathways that are altered in cancers, and for the patterns of variation in neoplasms that arise in specific cell linkages. Hanahan and Weinberg (47) recently noted that at least 12 molecular pathways have various nodes that permit cross-talk but also have net vulnerabilities that may be exploitable by appropriate combinations of therapeutics. If this type of rational, network-based pharmacology is to achieve its full potential, it will be essential to integrate knowledge about these pathways into simplified, actionable decision templates that can be adopted by physicians. To that end, gene and protein variation nomenclatures must be consistent with incorporation into such pathway analysis programs or languages as BIOPAX (48) or SMBL (49), respectively. Such pathway analyses are currently within the purview of the research investigator, but will eventually transition into tools that can be used by practicing physicians and will also provide

the logical foundation for reimbursement coding for specific molecular pathologies.

The need for well-curated biospecimens and preanalytical standards for the acquisition, handling, and storage of such specimens has already been emphasized (50, 51). As pointed out by Hewitt and colleagues (12), the methods employed to collect the pristine samples used by large academic programs such as TCGA and TARGET are far too expensive to be used in routine clinical practice. However, once a candidate biomarker is discovered, the requirements for measuring it should be individually assessed, because such measurements may not require the same degree of rigor as used in the discovery phase. Genetic mutations that were discovered in pristine samples by whole-genome sequencing can often be accurately assessed routine paraffin-embedded specimens with the use of PCR-based technologies. Although it is appropriate to establish rigorous standard operating procedures for the collection, stabilization, and storage of specimens, cost constraints will invariably impose process limitations that introduce variable levels of specimen and/or analyte degradation. Nonetheless, clinical investigators are beginning to transition next-generation sequencing technologies into the clinical laboratory by performing whole-exome sequencing on formalin-fixed paraffin-embedded (FFPE) tissues. Because these specimens come with rich clinical annotation and are a commonly used type of diagnostic material, the possibility of identifying genetic variants routinely in clinical practice to support targeted therapeutics may soon become a reality.

## Taking Advantage of More Precisely Defined Patient Populations: Opportunities for Later-Stage Clinical Trials

The unchecked rise in the costs of medical care is unsustainable (52, 53). Current late-stage cancer drug development is both expensive and characterized by a high failure rate. Ever larger phase III clinical trials are associated with high risk of clinical trial failure to meet registration thresholds of efficacy and safety. The ability to develop drugs initially in more precisely defined and biologically appropriate patient populations would facilitate new strategies for enrichment/adaptive clinical trial designs. One approach would be to restrict early development to particularly targeted groups of patients, which would allow accurate assessment of the relationships of dose and schedule with efficacy and toxicity in carefully defined patient groups. This strategy has been linked to the policy initiative related to progressive approval, a variant of accelerated approval by the FDA associated with use of a new drug or biological in a carefully defined patient population for a well-circumscribed indication and patient population. Expansion of the initial approval and wider use of the drug would be associated with progressive development of increasing levels of evidence related to clinical validity and safety in broader groups of patients. This approach is particularly applicable to the development of "first in class" single agents in areas where there is a clearly specified and high, unmet medical need. For example, such an approach would have worked well in the development of imatinib for BCR-ABL–related chronic myeloid leukemia (54) and KIT-positive gastrointestinal tumor (55), crizotinib (56) for *EML4-ALK* expressing NSCLC, and vemurafenib for *B-RAF*–mutated melanoma (57). These examples also suggest that the likelihood of overestimating activity and/or underestimating side effects is limited. It should be possible to establish criteria by international consensus that further refine progressive approval based on levels of objective response according to RECIST, the duration of the response, and/or the absence of progression at first tumor assessment. Other possibilities (58, 59) include initial regulatory approvals based on large, randomized phase II trials designed with sufficient power to suggest that an agent has sufficient safety and clinical efficacy in cancer patients to be approved for controlled marketing. Final approval would then be contingent upon the timely development of more definite evidence collected in larger postmarketing studies.

The initial biomarkers associated with benefit from targeted therapies in cancer are often low-prevalence somatic mutations in driver oncogenes, challenging the historical paradigm of progressive evaluation through classical trial phases. If a mutation with a 1% prevalence identifies patients with a 95% frequency of benefit, this can be confirmed with high confidence in a relatively small single-arm trial in that subset of patients. Although the practical barriers to identifying such patients in the general population are considerable, the advantages of this approach may outweigh the disadvantages, provided that the marker used to select the patient population is suitably robust. Much larger randomized trials would be needed to prove the lack of benefit of a targeted agent in biomarker-negative populations, or the relative benefit of other therapies in this small subset of patients (who may respond differently). Adaptive design trials such as those used in the I-SPY2 [Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging and Molecular Analysis 2 (60-62)] and Battle-1 [Biomarker-integrated Approaches of Targeted Therapy for Lung Cancer Elimination 1 (63)] trials suggest approaches that may minimize the number of marker-negative patients who are treated before predicting that a larger trial involving such patients would not benefit from a randomized phase III trial (64). In I-SPY2, an experimental arm either graduates or is dropped if there is a $\geq$85% or <10% predictive probability of success, respectively, in a phase III trial (64). Similar approaches using standard randomized phase II trial designs also may limit the number of patients who will not benefit from inclusion in a large phase III trial because they are negative for an integral marker. Whether such approaches can be used for development of drugs with targets other than those that drive the oncogenetic phenotype, or combinations of agents, remains a matter of debate.

The issue of low-incidence drug toxicity is a special case. Historically, this issue has been addressed by regulatory requirements for very large patient numbers (often far

beyond those required to determine efficacy) prior to drug approval. The opportunity to define particular low-frequency patient groups at particular risk of toxicity will enable new, more efficient approaches. Data suggest that low-prevalence SNPs and other interindividual variants may be critical for determining drug efficacy and toxicity. However, it has been shown that even large phase III trials may be too small to include enough patients with uncommon SNPs that might identify patients at particular risk of toxicity or, for that matter, those who are most likely to benefit from treatment. In a recent review, Grossman (65) suggested that when a therapeutic is approved, the first 250,000 recipients should undergo testing and close surveillance for both toxicity and benefit. This can be achieved by means of large-scale observational studies with high-throughput SNP analysis linked to electronic health records and national indices maintained by the CMS and other agencies. Andy Grove, the former CEO of Intel, made a similar request in a recent *Science* editorial (66). Genome-wide association studies performed in large postmarket surveys might be an unbiased way to identify gene polymorphisms and variants in patients who first receive novel therapeutics under an FDA-accelerated approval. Companies that currently do this type of analysis might assist by providing the genetic variant data for such postmarket surveys. Once these data are generated, they can be linked with electronic health records as well as various national indices, such as the National Death Index, at a cost that might be considerably cheaper than current large phase III trials.

## Conclusions

Trialists, translational science investigators, and clinical laboratory scientists must collaborate to move new research discoveries about molecular alterations in cancer into clinical practice.

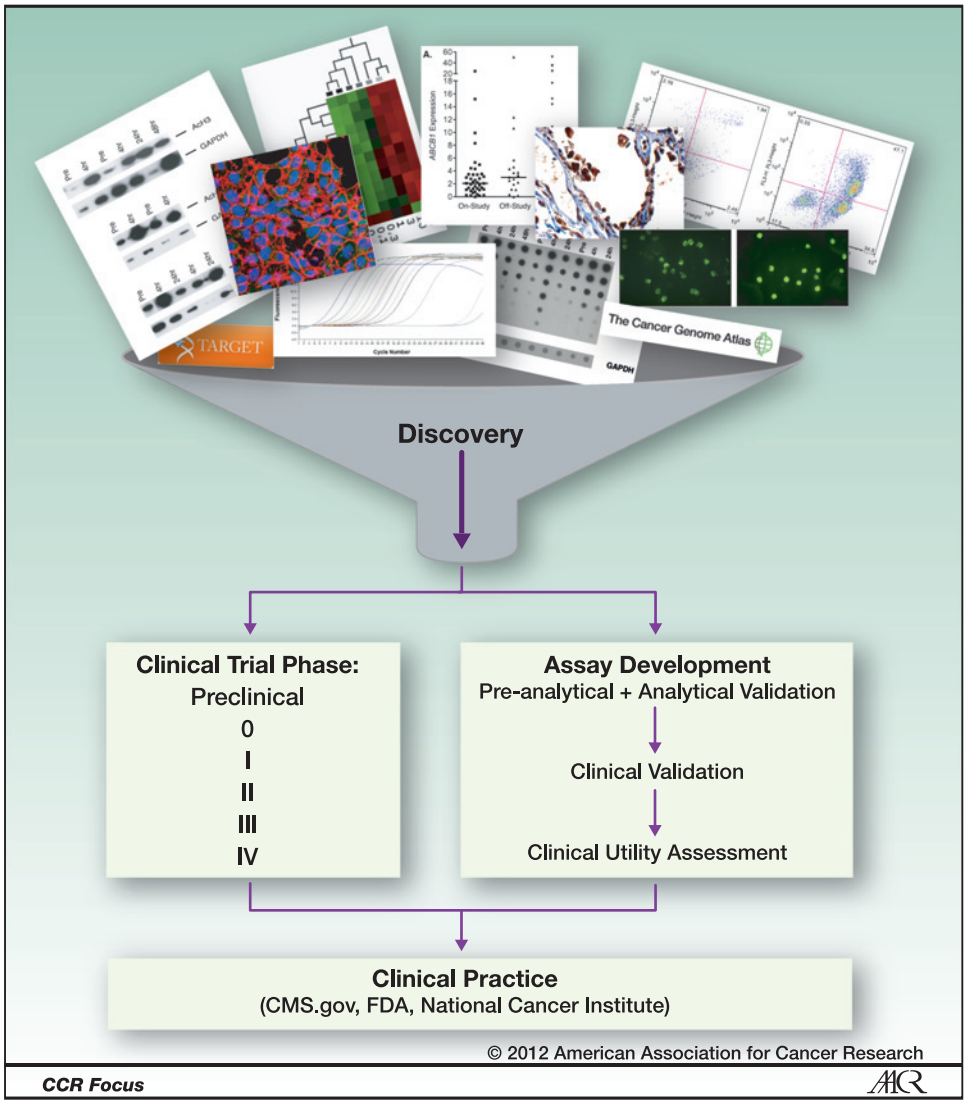


**Figure 2.** Flow of marker development. Markers that are candidates for molecular diagnostics come from multiple sources (e.g., TARGET and TCGA) as well as independent investigators who use various approaches to analyze aspects of biology. As these markers enter into the clinical research space, they must undergo a process of analytical validation as well as use in clinical trials. Once these diagnostics have been identified as successful prognostic, predictive, or pharmacogenomic markers, they may be used in clinical practice if reviewed positively by the FDA, CMS, or other payors. Assay images courtesy of the National Cancer Institute, laboratory of Susan E. Bates.

This is especially important for clinical validation studies of novel diagnostics, in which laboratory scientists must often work with samples that are collected under more variable conditions compared with discovery-phase protocols.

To move diagnostics forward in the most expeditious and cost-effective way, we may need a shift in the focus of clinical trials, with greater emphasis on randomized phase II trials, reduced emphasis on phase III trials, and greater reliance on phase IV trials that integrate electronic health records, and large databases for observational studies to identify adverse events and survival. Large-scale genotyping would enable investigators to correlate certain genotypes with phenotypes that confer therapeutic benefit, as well as those that confer a risk of serious toxicity. This would also be associated with a change in the FDA approval process, i.e., the agency could grant a provisional approval for larger, randomized phase II trials that would move to full approval after successful completion of a larger phase IV study.

Finally, as pointed out by Meshinchi and colleagues (14) elsewhere in this series, the FDA is taking a more proactive role in monitoring the use of integral markers in clinical trials. It is important to ensure that this monitoring, with its added regulatory burden of requiring data and reports from sponsors for an Investigational Device Exemption, is in turn evaluated to determine whether it results in an improved approval rate for molecular diagnostics and improved clinical outcomes for patients in terms of increased diagnostic accuracy and rational therapeutic selection based on molecular profiling and subtyping of individual patients' cancers (Fig. 2).

## Disclosure of Potential Conflicts of Interest

G. Poste is a consultant to and serves on the advisory board of Exelixis. The other authors disclosed no potential conflicts of interest.

## Acknowledgments

The authors thank Professor Donald A. Berry (MD Anderson Cancer Center, Houston, TX) for valuable advice and discussion.

Received December 5, 2011; revised January 27, 2012; accepted January 27, 2012; published online March 15, 2012.

## References

1. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. Nature 2011;474:609–15.
2. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al; Cancer Genome Atlas Research Network. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell 2010; 17:98–110.
3. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 2008;455:1061–8.
4. Mullighan CG, Collins-Underwood JR, Phillips LA, Loudin MG, Liu W, Zhang J, et al. Rearrangement of CRLF2 in B-progenitor- and Down syndrome-associated acute lymphoblastic leukemia. Nat Genet 2009;41:1243–6.
5. Harvey RC, Mullighan CG, Wang X, Dobbin KK, Davidson GS, Bedrick EJ, et al. Identification of novel cluster groups in pediatric high-risk B-precursor acute lymphoblastic leukemia with gene expression profiling: correlation with genome-wide DNA copy number alterations, clinical characteristics, and outcome. Blood 2010;116:4874–84.
6. Mullighan CG, Zhang J, Harvey RC, Collins-Underwood JR, Schulman BA, Phillips LA, et al. JAK mutations in high-risk childhood acute lymphoblastic leukemia. Proc Natl Acad Sci U S A 2009;106:9414–8.
7. Zhang J, Mullighan CG, Harvey RC, Wu G, Chen X, Edmonson M, et al. Key pathways are frequently mutated in high-risk childhood acute lymphoblastic leukemia: a report from the Children's Oncology Group. Blood 2011;118:3080–7.
8. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, et al. An integrated genomic analysis of human glioblastoma multiforme. Science 2008;321:1807–12.
9. Wood LD, Parsons DW, Jones S, Lin J, Sjöblom T, Leary RJ, et al. The genomic landscapes of human breast and colorectal cancers. Science 2007;318:1108–13.
10. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 2008;321:1801–6.
11. Schilsky RL, Doroshow JH, LeBlanc M, Conley BA. Development and use of integral assays in clinical trials. Clin Cancer Res 2012;18:1540–6.
12. Hewitt SM, Badve SS, True LD. Impact of preanalytic factors on the design and application of integral biomarkers for directing patient therapy. Clin Cancer Res 2012;18:1524–30.
13. Williams PM, Lively TG, Jessup JM, Conley BA. Bridging the gap: moving predictive and prognostic assays from research to clinical use. Clin Cancer Res 2012;18:1531–9.
14. Meshinchi S, Hunger SP, Aplenc R, Adamson PC, Jessup JM. Lessons learned from the Investigational Device Exemption review of Children's Oncology Group Trial AAML1031. Clin Cancer Res 2012;18:1547–54.
15. Andre F, Nowak F, Arnedos M, Lacroix L, Viens P, Calvo F. Biomarker discovery, development, and implementation in France: a report from the French National Cancer Institute and cooperative groups. Clin Cancer Res 2012;18:1555–60.
16. Atkinson AJ Jr, Colburn WA, DeGruttola VG, DeMets DL, Downing GJ, Hoth DF, et al. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 2007; 69:89–95.
17. U.S. Food and Drug Administration [homepage on the Internet]. Draft guidance for industry, clinical investigators, and Food and Drug Administration staff—design considerations for pivotal clinical investigations for medical devices [cited 2011 Sept 20]. Available from: http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm081752.htm.
18. U.S. Food and Drug Administration [homepage on the Internet]. Draft guidance for industry and Food and Drug Administration staff—in vitro companion diagnostic devices [cited 2011 Jul 13]. Available from: http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm262292.htm.
19. U.S. Government Printing Office [homepage on the Internet]. Title 42–Public Health. Chapter IV—Centers for Medicare and Medicaid Services, Department of Health and Human Services (continued). Part 493–Laboratory requirements. Available from: http://www.access.gpo.gov/nara/cfr/waisidx_04/42cfr493_04.html.
20. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. J Natl Cancer Inst 2009;101: 1446–52.
21. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GMStatistics Subcommittee of the NCI-EORTC Working Group on Cancer Diagnostics. Reporting recommendations for tumor marker prognostic studies (REMARK). J Natl Cancer Inst 2005;97:1180–4.
22. Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, et al. IDH1 and IDH2 mutations in gliomas. N Engl J Med 2009;360:765–73.

23. Westgard JO. The need for a system of quality standards for modern quality management. Scand J Clin Lab Invest 1999;59:483–6.
24. Lee JW, Devanarayan V, Barrett YC, Weiner R, Allinson J, Fountain S, et al. Fit-for-purpose method development and validation for successful biomarker measurement. Pharm Res 2006;23:312–28.
25. Wagner JA. Strategic approach to fit-for-purpose biomarkers in drug development. Annu Rev Pharmacol Toxicol 2008;48:631–51.
26. U.S. Food and Drug Administration [homepage on the Internet] [cited 2011 Oct 30]. Available from: http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/Overview/default.htm.
27. National Cancer Institute [homepage on the Internet]. Templates for clinical assay development [cited 2011 Oct 30]. Available from: http://cdp.cancer.gov/diagnostics/templates.htm.
28. Grund B, Sabin C. Analysis of biomarker data: logs, odds ratios, and receiver operating characteristic curves. Curr Opin HIV AIDS 2010;5:473–9.
29. Jung K, Stephan C, Lein M, Henke W, Schnorr D, Brux B, et al. Analytical performance and clinical validity of two free prostate-specific antigen assays compared. Clin Chem 1996;42:1026–33.
30. Jakobsdottir J, Gorin MB, Conley YP, Ferrell RE, Weeks DE. Interpretation of genetic association studies: markers with replicated highly significant odds ratios may be poor classifiers. PLoS Genet 2009;5:e1000337.
31. Cicchetti DV, Koenig K, Klin A, Volkmar FR, Paul R, Sparrow S. From Bayes through marginal utility to effect sizes: a guide to understanding the clinical and statistical significance of the results of autism research findings. J Autism Dev Disord 2011;41:168–74.
32. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. BMJ 2004;329:168–9.
33. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM; Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. Ann Intern Med 2008;149:889–97.
34. Moreira J, Bisoffi Z, Narváez A, Van den Ende J. Bayesian clinical reasoning: does intuitive estimation of likelihood ratios on an ordinal scale outperform estimation of sensitivities and specificities? J Eval Clin Pract 2008;14:934–40.
35. National Human Genome Research Institute [homepage on the Internet]. Ensuring the safety and effectiveness of new genetic tests. Chapter 2. Promoting safe and effective genetic testing in the United States. A report of the Genetic Testing Task Force of the National Human Genome Research Institute [cited 2011 Oct 30]. Available from: http://www.genome.gov/10002404.
36. National Institute for Health and Clinical Excellence [homepage on the Internet] [cited 2011 Oct 30]. Available from: http://www.nice.org.uk/.
37. Center for Medical Technology Policy [homepage on the Internet] [cited 2011 Oct 30]. Available from: http://www.cmtpnet.org/.
38. Agency for Healthcare Research and Quality [homepage on the Internet] [cited 2011 Oct 30]. Available from: http://www.ahrq.gov/clinic/techix.htm.
39. Evaluation of Genomic Applications in Practice and Prevention [homepage on the Internet] [cited 2011 Oct 30]. Available from: http://www.egappreviews.org/.
40. Blue Cross Blue Shield Association [homepage on the Internet] [cited 2011 Oct 30]. Available from: http://www.bcbs.com/blueresources/tec/.
41. Blue Cross Blue Shield Association [homepage on the Internet] [cited 2011 Oct 30]. Available from: http://www.bcbs.com/blueresources/tec/vols/25/epidermal-growth-factor.html.
42. Andre F, McShane LM, Michiels S, Ransohoff DF, Altman DG, Reis-Filho JS, et al. Biomarker studies: a call for a comprehensive biomarker study registry. Nat Rev Clin Oncol 2011;8:171–6.
43. National Cancer Institute [homepage on the Internet] [cited 2011 Nov 26]. Available from: http://cancer.gov.
44. My Cancer Genome [homepage on the Internet] [cited 2011 Nov 26]. Available from: http://www.mycancergenome.org/.
45. HUGO Gene Nomenclature Committee [homepage on the Internet] [cited 2011 Nov 26]. Available from: http://genenames.org.
46. Human Genome Variation Society [homepage on the Internet] [cited 2011 Nov 26]. Available from: http://www.hgvs.org/.
47. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646–74.
48. Demir E, Cary MP, Paley S, Fukuda K, Lemer C, Vastrik I, et al. The BioPAX community standard for pathway data sharing. Nat Biotechnol 2010;28:935–42.
49. Ruebenacker O, Moraru II, Schaff JC, Blinov ML. Integrating BioPAX pathway knowledge with SBML models. IET Syst Biol 2009;3:317–28.
50. Poste G. Bring on the biomarkers. Nature 2011;469:156–7.
51. Moore HM, Compton CC, Alper J, Vaught JB. International approaches to advancing biospecimen science. Cancer Epidemiol Biomarkers Prev 2011;20:729–32.
52. Sleijfer S, Verweij J. The price of success: cost-effectiveness of molecularly targeted agents. Clin Pharmacol Ther 2009;85:136–8.
53. Sullivan R, Peppercorn J, Sikora K, Zalcberg J, Meropol NJ, Amir E, et al. Delivering affordable cancer care in high-income countries. Lancet Oncol 2011;12:933–80.
54. O'Brien SG, Guilhot F, Larson RA, Gathmann I, Baccarani M, Cervantes F, et al; IRIS Investigators. Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia. N Engl J Med 2003;348:994–1004.
55. Verweij J, Casali PG, Zalcberg J, LeCesne A, Reichardt P, Blay JY, et al. Progression-free survival in gastrointestinal stromal tumours with high-dose imatinib: randomised trial. Lancet 2004;364:1127–34.
56. Kwak EL, Bang YJ, Camidge DR, Shaw AT, Solomon B, Maki RG, et al. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. N Engl J Med 2010;363:1693–703.
57. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, et al; BRIM-3 Study Group. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N Engl J Med 2011;364:2507–16.
58. Stewart DJ, Kurzrock R. Cancer: the road to Amiens. J Clin Oncol 2009;27:328–33.
59. Sobrero A, Bruzzi P. Incremental advance or seismic shift? The need to raise the bar of efficacy for drug approval. J Clin Oncol 2009;27:5868–73.
60. I-SPY 2 TRIAL. Available from: http://ispy2.org/.
61. I-SPY 2 TRIAL: Neoadjuvant and personalized adaptive novel agents to treat breast cancer. Available from: http://clinicaltrials.gov/ct2/show/NCT01042379.
62. Barker AD, Sigman CC, Kelloff GJ, Hylton NM, Berry DA, Esserman LJ. I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. Clin Pharmacol Ther 2009;86:97–100.
63. Kim ES, Herbst RS, Wistuba II, Lee JJ, Blumenschein GR, Tsao A, et al. The BATTLE trial: personalizing therapy for lung cancer. Cancer Discov 2011;1:44–53.
64. Berry DA, Herbst RA, Rubin EH. Reports from the 2010 Clinical and Translational Cancer Research Think Tank Meeting: design strategies for personalized therapy trials. Clin Cancer Res 2012;18:638–44.
65. Grossman I. Routine pharmacogenetic testing in clinical practice: dream or reality? Pharmacogenomics 2007;8:1449–59.
66. Grove A. Rethinking clinical trials. Science 2011;333:1679.

Clinical Cancer Research

# Leveling the Playing Field: Bringing Development of Biomarkers and Molecular Diagnostics up to the Standards for Drug Development

George Poste, David P. Carbone, David R. Parkinson, et al.

*Clin Cancer Res* 2012;18:1515-1523.