Published in final edited form as:

Clin Cancer Res. 2012 March 15; 18(6): 1515–1523. doi:10.1158/1078-0432.CCR-11-2206.

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

George Poste¹, David P. Carbone², David R. Parkinson³, Jaap Verweij⁴, Stephen Hewitt⁵, and J. Milburn Jessup^{6,7}

¹Chief Scientist, Complex Adaptive Systems Initiative, Arizona State University, 1475 North Scottsdale Road, Suite 361 Scottsdale, Arizona 85257 ²Professor of Medicine and Cancer Biology, Director, Thoracic/Head and Neck Program and Thoracic Oncology Center, Vanderbilt-Ingram Cancer Center, 691 Preston Building, Nashville, TN 37232-6838 ³Chief Executive Officer, Nodality, Inc., 201 Gateway Boulevard, South San Francisco, CA 94080 ⁴Chairman, Dept. of Medical Oncology and Daniel den Hoed Cancer Center, Erasmus University Medical Center, PO Box 2040, 3000 CA ROTTERDAM, Netherlands ⁵Director, Tissue Array Research Program (TARP), Laboratory of Pathology, Advanced Technology Center, National Cancer Institute, Gaithersburg, MD ⁶Cancer Diagnosis Program, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD

Abstract

Molecular diagnostics are increasingly important in clinical research to stratify or identify molecularly profiled patient cohorts for targeted therapies, to modify the dose of a therapeutic, or to assess early response to therapy or monitor patients. Molecular diagnostics can also be used to identify pharmacogenetic risk of adverse drug reactions. The articles of this *CCR Focus* section on Molecular Diagnosis describe the development and use of markers for medical decision-making in the cancer patient. They define the sources of preanalytic variability to minimize as well as the regulatory and financial challenges in diagnostic development and integration into clinical practice. They also outline an NCI program to assist diagnostic development. Molecular diagnostic clinical tests require rigor in their development and clinical validation with sufficient sensitivity, specificity and validity that is comparable to that used for 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. A way to accomplish this is to emphasize Phase IV post-marketing hypothesis driven clinical trials with biological characterization that permits accurate definition of the association of low prevalence gene alterations with toxicity or response in large cohorts.

⁷To whom correspondence should be addressed: Chief, 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; jessupj@mail.nih.gov.

Introduction

This is an especially exciting time in medicine created by the application of the burgeoning knowledge of the molecular and cell biology of cancer to improve diagnosis and therapy. Recent results with 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 demands unprecedented levels of collaboration between clinicians and clinical laboratory scientists in the design and monitoring of clinical trials, including the need for validation of the clinical as well as analytical performance of the diagnostic (11), the need for standardization of preanalytical variables during specimen collection, stabilization and processing (12), the need also standards and rigorous attention to analytical performance and validation of the assay (13) and fulfillment of the regulatory requirements (14). All of these papers are testament to the free market approach to molecular diagnostics that characterizes the United States system of medical innovation. Andre et al (15) describe an alternative in which a government entity (the French NCI) contributes to the development of molecular markers and ensures that molecular diagnostics are performed with sufficient quality by distributing SOPs and certified reagents throughout the country. The present manuscript defines the different types of markers that comprise the contemporary universe of new molecular diagnostics and emphasizes the fundamental importance of standardized analytical parameters, suggesting a reduction in Phase III trials and proposing metrics by which the quality of diagnostic development may be assessed by the regulatory agencies.

Categories of Markers in Oncology

A biomarker has been 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 five types of molecular diagnostic that have different functions and are usually associated with separate phases of clinical trials. Clinical research generally involves Phase 0–III clinical trials while Phase IV trials are post-market pharmacovigilance surveys that assess drug safety in large population sets and potential efficacy in new indications. In contrast, Phase 0 trials are trials in which Pharmacokinetic (PK) and Pharmacodynamic (PD) effects are evaluated enabling comparisons with an investigational agent’s behavior in preclinical models. The dosing in Phase 0 is sufficiently low that the biological effect of an agent may be detectable on its intended target without risk for serious adverse events. Phase I trials determine the maximum tolerated dose (MTD) 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 that use an agent dosed generally at or close to the MTD that begin to ascertain efficacy in different malignancies while Phase III trials are large randomized trials intended to confirm the efficacy of an agent (alone or in combination) in a particular disease context compared to the “standard-of-care” therapy.

Five basic types of markers are used in clinical trials often with phase-specific purposes (Figure 1). **Pharmacokinetic** markers assess the absorption, distribution, metabolism and excretion of a drug or agent while **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 (Figure 1). They are considered part of clinical research because in general the results are not provided to the patient or their attending physician and are not used for medical decision-making. While 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 (GLP) as defined by the Food and Drug Administration (FDA) (17,18). These markers can be useful for identification or confirmation of mechanism of action of an agent and relate both mechanistic and off target effects to drug dose and schedule. However, a recent NCI-sponsored trial uses a Pharmacokinetic marker for medical decision-making since the trial assesses the response of individual patients to a test dose of busulfan and then tailors the therapeutic dose of the drug in each patient to the kinetics of the production of metabolites in a fixed period of time. Since this assay now is used for medical decision-making, this pharmacokinetic is now performed in a clinically certified laboratory.

Markers that are associated with survival or other clinical endpoints independent of any specific treatment are classified as **Prognostic** markers. Such markers may also be useful in monitoring the response of patients to therapy e.g. Circulating Tumor Cell enumeration in patients with colon, breast, or prostate carcinoma. Markers that are associated with a clinical endpoint such as survival but also assess the effectiveness of a particular treatment are designated **Predictive** markers since they predict the response to that treatment (Figure 1). Some predictive markers can identify individual patients that are more likely to respond to particular drug and thereby can be used to select patients for therapy, e.g. EML-ALK mutations in patients with Non Small Cell Lung Cancer (NSCLC) as a basis for utilization of crizotinib. The final class of markers is **Pharmacogenomic** markers that are used to identify risk of organ-based toxicities or altered metabolism and/or responses to therapeutic agents (Figure 1). These markers are usually inherited in the germ line and most examples to date are typically non-synonymous 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 et al (11) elsewhere in this series). These markers may include prognostic, predictive, pharmacogenomic or occasionally pharmacodynamic markers. Integral markers may be used to determine patient eligibility for a trial, their assignment to therapy, dose selection, or patient stratification within a trial. Since in these cases the test result is known to physician and patient and influences clinical practice, the assay must be performed in a Clinical Laboratory Improvement Act of 1988 (CLIA)-certified laboratory as stipulated in FAR section 42 CFR 493 (19). In the United States, any specimen from a patient 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 clinical research as well as practice where the patient and/or their physician are informed of a test result or may learn indirectly of the test result, via such events as assignment to a particular treatment based on the result of a diagnostic test. The intent of these regulations is to assure that the highest standards of reproducibility and reliability of assays performed on patients in order to allow optimum medical decision-making. Integral markers should be distinguished from integrated markers that are performed in all patients within a trial or in a pre-defined subset but are not used for medical decision-making and instead are intended for clinical research or development of a marker and its assay for use in a subsequent trial (see Williams et al (13)).

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 where the patient or their physician will not learn the test result 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 of test results, several other factors may prevent direct use of discovery research data in the clinic. The source of specimens is often different between discovery and clinical research. 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 measured. In contrast, clinical specimens are collected in busy clinical settings such as interventional radiology, operating suites or 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 than in a dedicated research setting. Clinical specimens are therefore likely to have more degradation than impeccably curated research specimens. More information on the importance of pre-analytic variables is described in the manuscript by Hewitt et al (12) elsewhere in this series.

Differences in preanalytic factors may have a profound influence on the identification and reproducibility of putative markers (20, 21). Discovery is just that – a search for gene or protein alterations or variants that are not otherwise predefined whereas 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 since 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 respectively in exposure to a potentially toxic treatment or inappropriate denial of therapy. 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 validated to have a high sensitivity and specificity and a concomitant low false negative and positive rate. 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 assayed individually or in small groups. This increases inter-assay variability and requires greater controls to insure that inter-run variability is minimized. Clinical assays for medical decision-making thus face greater challenges for reliability and reproducibility than discovery stage assays. Despite these differences it is critical that discovery research data be of high quality to justify the time and cost needed to establish and validate a clinical assay. A common cause for 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 in REMARK by McShane et al. (21) are critical to assure that discovery markers meet a high level of rigor before any translation into a clinical diagnostic is contemplated.

Why Clinicians, 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 are identified, confirmation requires larger datasets such as provided by the TCGA or 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 several important findings have been generated 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 ALL (6). Unfortunately, it takes as long as 2–5 years to develop and validate clinical assays to measure these mutations in a sufficiently large number of clinical samples. The conversion from discovery candidates to a validated clinical assay requires that the analytical performance of the assay meets the principles defined by Westgard (23), Lee (24), others (25). The FDA has provided guidances for satisfactory analytical performance of molecular diagnostics that includes suitable controls, and analysis of inter-and intra-reproducibility testing (26). The Cancer Diagnosis Program of the NCI provides downloadable templates that address the validation of analytical performance of FISH and IHC assays (27). These issues are further expanded in Schilsky et al (12).

Analytical validity of an assay requires that it can detect the analyte when it is present and not when it is absent and is the primary focus of diagnostic laboratory accreditation. As clinical molecular diagnostic tests evolve to play a greater role in treatment decision-making, establishing the clinical validity of the assay assumes equal importance to analytical validity. Regulatory agencies have indicated their intent to focus more on significant “high risk” clinical tests, i.e. those associated with medical clinical decisions for individual patients. Clinical validity is the process by which a positive test result is associated with a particular clinical endpoint or event while a negative test is not. In this context, the clinical end point may be either related to survival (overall or a progression-free interval or other evaluation of the kinetics of disease progression) or response to therapy (e.g., meeting a prespecified reduction in tumor burden as reflected in the RECIST criteria). Clinical validity requires that the assay characterize the analyte(s) in specimens collected within the clinical context of intended use, which may range widely. As assays progress from discovery to development, 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), then 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 cut-offs through use of Receiver Operating Characteristic (ROC) curves (28) or other analyses (29, 30).

Finally, for a marker to progress from discovery research to be a clinically useful test, the diagnostic test must have high clinical utility before clinical guidelines adopt it, payers will reimburse it, and clinical practitioners assimilate it into routine practice. 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 utilized 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 (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 USA. These organizations follow predefined criteria which seek to provide an unbiased approach to

assessment of the utility of the test with any associated companion therapeutic(s). The BCBS TEC provides clear criteria on their website and assesses the utility of treatments and diagnostics in light of the following parameters: 1) the technology must be approved by the appropriate governmental regulatory bodies; 2) the scientific evidence must enable conclusions about the effect of the technology on health outcomes; 3) the technology must improve the net health outcome; 4) the technology must be as beneficial as established alternatives; and 5) the improvement must be expected to be attainable outside clinical trial research investigations. These criteria appear to be similar to those used by NICE and CMPT, although the BCBS TEC also will consider reviewing topics that do not fulfill all of the criteria. As is the specific case of new diagnostic tests for oncology, the BCBS TEC has supported the use of KRAS testing in metastatic colorectal carcinoma patients when treatment with anti-EGFR antibodies is considered. It also supports the use of erlotinib in patients with mutations in EGFR in Non Small Cell Lung Cancer after analysis of nonrandomized clinical data when it had originally ruled against the use of erlotinib. While other organizations release guidelines for clinical practice (e.g., American Society of Clinical Oncology, National Comprehensive Cancer Network, etc.), practitioners may need to track the reports provided by such evaluation committees as BCBS TEC, NICE, AHRQ, and CMPT because their reviews include the concerns of the payers who are critical for reimbursement of both the diagnostic test as well as the companion treatment. These assessments also often identify scientific short-comings in the development of a diagnostic or treatment in terms of the science 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 raises questions about whether the diagnostic test serves solely as a predictor of therapeutic response or as a prognostic factor and what is 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 Andre et al (42) have described, there is an essential need to identify clinically useful molecular and imaging markers in a registry or database so that these diagnostics may be identified by both clinicians and patients. Currently, there are several websites that track a number of diagnostics that are used for assessment of patients for eligibility for targeted therapies, such as the NCI's site (43) and Vanderbilt University's My Cancer Genome (44). As the assessment of genomic alterations in given disease states becomes the standard of care, it will be important to integrate the reporting of these findings, and their implications, into the electronic medical record systems used by attending clinicians. In addition, there are now many clinical trials for patients that require relatively rare genetic abnormalities (often 1% prevalence or less), so matching patients with these specific lesions with relevant clinical trials (institutionally, nationally, and internationally) will be of great value. This approach is already part of such patient-oriented sites like My Cancer Genome.

While it is relatively straight forward to use standard gene and protein names such as provided by the Human Genome Organisation (HUGO, 45), this only allows indexing of the root gene or protein name. 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 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 altered in cancers and the patterns of variation for neoplasms arising in specific cell linkages. As Hanahan and Weinberg (47) have recently described, there are at least 12

molecular pathways with various nodes that permit cross-talk but whose net vulnerabilities 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 the knowledge into simplified, actionable decision templates that can be adopted by physicians. This requires that gene and protein variation nomenclature be consistent with incorporation into such pathway analysis programs or languages as BIOPAX (48) or SMOG (49), respectively. Such pathway analyses now are the purview of the research investigator but will eventually transition into tools that need to be usable by practicing physicians and also provide the logical foundation for reimbursement coding of specific molecular pathologies.

The need for well curated biospecimens and pre-analytical standards in their acquisition, handling and storage has already been emphasized (50, 51). As pointed out by Hewitt et al (12), the ability to collect the pristine samples that the large academic programs like TCGA and TARGET employ is 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 and may not require the same degree of rigor as used in the discovery phase. Genetic mutations discovered on pristine samples by whole genome sequencing can often be accurately assessed using PCR-based technologies from routine paraffin-embedded specimens. While it is appropriate to establish rigorous SOPs 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 tissues. Since these specimens come with rich clinical annotation and are the type of diagnostic material prevalent in the community, the prospect that genetic variants will be identified routinely in clinical practice to support targeted therapeutics is becoming a reality.

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

The unchecked rise in 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 prospect of developing drugs initially in more precisely defined, and biologically appropriate patient populations allows new strategies for enrichment/adaptive clinical trial designs. Possibilities include restricting early development to particularly targeted groups of patients, allowing 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 for development of “first in class” single agents in areas where there is a clearly specified and high unmet medical need. Examples where such an approach would have worked are for instance the development of agents such as Imatinib for BCR-ABL related Chronic Myeloid Leukemia (54) and KIT positive GIST (55), Crizotinib (56) in EML4-ALK expressing NSCLC, Vemurafenib in 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 by international consensus to build criteria that further refine progressive approval based on levels of objective response according to RECIST, the duration of these

responses, and/or absence of progression at first tumor assessment. Other possibilities (58, 59) include the potential for initial regulatory approvals to be based on large randomized Phase II trials designed with sufficient power to suggest an agent is 'safe enough' with enough clinical efficacy in cancer patients to be granted controlled marketing. Final approval would then be contingent upon the timely development of more definite evidence collected in larger post-marketing 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 this subset of patients. While the practical barriers to identifying these patients in the general population are considerable, the advantages of the 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 the I-SPY2 (Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging and Molecular Analysis 2) (60, 61) and Battle-1 (Biomarker-integrated Approaches of Targeted Therapy for Lung Cancer Elimination) trials (62) 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 (63). In I-SPY2 an experimental arm either graduates or is dropped if there is a $\geq 85\%$ or a $< 10\%$, respectively, predictive probability of success in a Phase III trial (63). 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 could 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 has been approached by regulatory requirements for very large patient numbers prior to drug approval, often far beyond those required to determine efficacy. The opportunity to define particular low frequency patient groups at particular risk of toxicity allows new, more efficient approaches. Data suggest that low prevalence SNPs and other inter-individual variants may be critical in determining drug efficacy and toxicity. However, even large Phase III trials have been shown to not be sufficiently sized to include enough patients with uncommon SNPs which might identify those at particular risk of toxicity or for that matter, benefit. In a recent review Grossman (64) suggested that as therapeutics are approved perhaps the first 250,000 recipients should undergo testing and close surveillance for both toxicity and benefit. This can be achieved using large scale observational studies with high throughput SNP analysis linked to electronic health records and National Indices maintained by CMS and other agencies. Grove, the previous CEO of Intel, made a similar request in a recent Science editorial (65). Genome Wide Association Studies (GWAS) might be an unbiased way to do this in large post market surveys if patients who receive the novel therapeutic also participate in GWAS studies. Companies that currently do this type of analysis might assist with providing the genetic variant data for such post-market surveys. Once these data are generated, they could be linked with electronic health records as well as various national indices such as the National Death Index, at a cost that may be considerably cheaper than current large Phase III trials.

SUMMARY

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

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

To move diagnostics forward in the most expeditious and cost effective way may require a shift in the emphasis 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, large databases for observational studies to identify adverse events and survival as well as the possible support for large scale genotyping so that genotypes can be correlated with the phenotypes that confer therapeutic benefit of the risk of serious toxicity. This would also be associated with in a change in the FDA approval process that would entail a provisional approval in 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 et al(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 that this monitoring with its added regulatory burden of providing data and reports by sponsors for Investigational Drug Exemptions should, in turn, be evaluated to determine whether it results in an improved approval rate for molecular diagnostics and in improved clinical outcomes for patients in terms of increased diagnostic accuracy and rational therapeutic selection based on molecular profiling and subtyping of a patients cancer (Figure 2).

Acknowledgments

The opinions expressed in this manuscript 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. The authors gratefully acknowledge the advice and discussion provided by Professor Donald A. Berry, U. T. M. D. Anderson Cancer Center, Houston, TX.

References

1. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature*. 2011; 474:609–15. [PubMed: 21720365]
2. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. 2010; 17:98–110. [PubMed: 20129251]
3. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. 2008; 455:1061–1068. [PubMed: 18772890]
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*. 2010; 41:1243–1246. [PubMed: 19838194]
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; 116:4874–4884. [PubMed: 20699438]
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. [PubMed: 19470474]

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. [PubMed: 21680795]
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. [PubMed: 18772396]
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. [PubMed: 17932254]
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. [PubMed: 18772397]
11. Schilsky RL, Doroshow JH, LeBlanc M, Conley BA. Development and use of integral assays in clinical trials. *Clin Cancer Res*. 2012:18.
12. Hewitt SM, Badve SS, True LD. The impact of pre-analytic factors in the design and application of integral biomarkers for directing patient therapy. *Clin Cancer Res*. 2012:18.
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.
14. Meshinchi S, Hunger SP, Aplenc R, Adamson PC, Jessup JM. Lessons learned from the Investigational Device Exemption (IDE) review of Children's Oncology Group Trial AAML1031. *Clin Cancer Res*. 2012:18.
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 NCI and Cooperative Groups. *Clin Cancer Res*. 2012:18.
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. Draft Guidance for Industry, Clinical Investigators, and Food and Drug Administration Staff. Design Considerations for Pivotal Clinical Investigations for Medical Devices Downloaded from the FDA website.
<http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm081752.htm> on Sept. 20, 2011
18. Draft Guidance for Industry and Food and Drug Administration Staff. In Vitro Companion Diagnostic Devices Downloaded from the FDA website.
<http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm262327.htm> on July 13, 2011
19. 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. [PubMed: 19815849]
21. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Statistics 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. [PubMed: 16106022]
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. [PubMed: 19228619]
23. Westgard JO. The need for a system of quality standards for modern quality management. *Scand J Clin Lab Invest*. 1999; 59:483–6. [PubMed: 10667684]
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. [PubMed: 16397743]
25. Wagner JA. Strategic approach to fit-for-purpose biomarkers in drug development. *Annu Rev Pharmacol Toxicol*. 2008; 48:631–51. [PubMed: 17937595]
26. <http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/default.htm>
27. <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. [PubMed: 20978390]
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. [PubMed: 8674185]
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. [PubMed: 19197355]
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. [PubMed: 20490646]
32. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ*. 2004; 329:168–9. [PubMed: 15258077]
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. [PubMed: 19075208]
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. [PubMed: 19018928]
35. Ensuring the Safety and Effectiveness of New Genetic Tests. Chapter 2 of “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. Downloaded from the NHGRI website (<http://www.genome.gov/10002404>) on Sept. 29, 2011
36. <http://www.nice.org.uk/>
37. <http://www.cmtpn.net/>
38. <http://www.ahrq.gov/clinic/techix.htm>
39. <http://www.egapreviews.org/>
40. <http://www.bcbs.com/blueresources/tec/>
41. <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. [PubMed: 21364690]
43. <http://cancer.gov>
44. <http://www.mycancergenome.org/>
45. <http://genenames.org>
46. <http://www.hgvs.org/>
47. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011; 144:646–74. [PubMed: 21376230]
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. [PubMed: 20829833]
49. Ruebenacker O, Moraru II, Schaff JC, Blinov ML. Integrating BioPAX pathway knowledge with SBML models. *IET Syst Biol*. 2009; 3:317–28. [PubMed: 21028923]
50. Poste G. Bring on the biomarkers. *Nature*. 2011; 469:156–7. [PubMed: 21228852]
51. Moore HM, Compton CC, Alper J, Vaught JB. International approaches to advancing biospecimen science. *Cancer Epidemiol Biomarkers Prev*. 2011; 20:729–32. [PubMed: 21430299]
52. Sleijfer S, Verweij J. The price of success: cost-effectiveness of molecularly targeted agents. *Clin Pharmacol Ther*. 2009; 85:136–138. [PubMed: 19151639]
53. Sullivan R, Peppercorn J, Sikora K, Zalcborg J, Meropol NJ, Amir E, et al. Delivering affordable cancer care in high-income countries. *Lancet Oncology*. 2011; 12:933–98. [PubMed: 21958503]
54. O’Brien SG, Guilhot F, Larson RA, Gathmann I, Baccarani M, Cervantes F, et al. Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia. *N Engl J Med*. 2003; 348:994–1004. [PubMed: 12637609]

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. [PubMed: 15451219]
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. [PubMed: 20979469]
57. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N Engl J Med*. 2011; 364:2507–16. [PubMed: 21639808]
58. Stewart DJ, Kurzrock R. Cancer: the road to Amiens. *J Clin Oncol*. 2009; 27:328–33. [PubMed: 19064964]
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. [PubMed: 19826122]
60. <http://ispy2.org/>, <http://clinicaltrials.gov/ct2/show/NCT01042379>
61. 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. [PubMed: 19440188]
62. Kim ES, Herbst RS, Wistuba II, Lee JJ, Blumenschein GR, Tsao A, et al. The BATTLE trial: Personalizing therapy for lung cancer. *Cancer Discovery*. 2011; 1:44–53.
63. Berry DA, Herbst RA, Rubin EH. Reports from 2010 Clinical and Translational Cancer Research Think Tank Meeting: Design Strategies for Personalized Therapy Trials. *Clin Cancer Res*. 2011 In Press.
64. Grossman I. Routine pharmacogenetic testing in clinical practice: dream or reality? *Pharmacogen*. 2007; 8:1449–59.
65. Grove A. Rethinking clinical trials. *Science*. 2011; 333:1679. [PubMed: 21940863]

Figure 1. The Types of Markers and Phases of Clinical Trials

There are five types of markers 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 generally not for medical decision-making in patients on the trial. In contrast, the other three markers (Prognostic, Predictive, and Pharmacogenetic) often are used for medical decision-making and therefore need to be performed in a CLIA-certified laboratory. In addition, these markers if used for medical decision-making also are likely to require a FDA pre-Investigational Drug Exemption (pre-IDE) review. Please see text for further descriptions of the types of markers and phases of trials as well as the CLIA and FDA requirements.

Figure 2. The Flow of Marker Development

As documented in this manuscript, markers that are candidates for molecular diagnostics come from multiple sources that include TARGET, 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 the 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.

Table 1
Differences between Discovery and Clinical Marker Development

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, Clinical Marker development develops an assay for a marker that may be an important therapeutic target or marker associated with clinical outcome. This assay needs to 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 Post-Analytic Factors.

	<u>Discovery (TCGA)</u>	<u>Clinical (Hospital)</u>
Preanalytic	Discovery	Clinical
Samples	Fresh Frozen Convenience Samples (bias by site)	FFPE Convenience Samples (bias by trial or locale)
Analytic		
Approach	Unbiased	Focused on Analyte in Tissue/Matrix
Power	Detect Variant	Reduce FP/FN
Runs	Batch Mode	Each test is independent
Sensitivity/Specificity	May have 10%+ FDR	Needs high Sensitivity/Specificity
Risk to Patient	Null -Low	High
Post-analytic		
	Not regulated	Regulated by FDA, CMS