From genome to bedside: Are we lost in translation?

Daniel F. Hayes*

University of Michigan Comprehensive Cancer Center, 1500 E. Medical Center Drive, Ann Arbor, MI 48109, USA

Abstract

The preceding decade has seen a remarkable technical explosion resulting in an entirely new field designated as “omics”. A committee convened by the United States Institute of Medicine (IOM) has defined omics as “characterization of global sets of biological molecules such as DNAs, RNAs, proteins, and metabolites”. The IOM report has established a roadmap for translating a newly discovered signature that emerges from an omics-based exploratory study to a true, analytically valid test that has both clinical validity and clinical utility for a specific intended use. This roadmap requires a multi-disciplinary team with expertise in the technical aspects of high-throughput assay development, bioinformatics, clinical test development, and clinical research and statistics. The investigative team should follow one of the pathways laid out by the IOM committee to establish clinical utility of an analytically validated omics-based test, and therefore acceptance by regulatory and guideline bodies: Prospective retrospective studies, or prospective studies in which the omics-based test is the primary objective of the trial itself. Although developed for omics-based tests, these concepts are applicable to any diagnostic test used to direct care of patients with cancer. These pathways are rigorous, and therefore not easily accomplished. However, if we are to apply these tests to direct management of our patients, we must approach the science of biomarker development with the same rigor that is used for therapeutic agent assessment. A “Bad Tumor Marker Is as Bad as a Bad Drug”.

Keywords: Tumor biomarker tests Omics-based tests

Introduction

The goal of “personalized medicine” is to get the right therapy to the right patient at the right time, using the right dose and schedule. Too often in oncology, clinicians have taken a one size fits all approach. Indeed, widespread application of adjuvant systemic therapy has led to a remarkable reduction in breast cancer mortality in the Western World over the last several decades [1,2]. These successes make one wonder if all patients should be treated with all adjuvant systemic therapies, in order to ensure maximum treatment of this potentially life-threatening disease. However, of course, anti-neoplastic therapies are associated with annoying side effects and occasional life-threatening events, as well as substantial inconvenience and cost. Therefore, as in all medicine, clinicians must weigh risks and benefits carefully.

How much benefit justifies the toxicity and cost of adjuvant systemic therapy? The answer to this question lies in the perspectives of the patient, her doctor, and her society, which must bear the burden of the cost. Sophisticated programs, such as Adjuvant! Online [3,4] permit the clinician and patient to make reasonable estimates of the risk of recurrence and mortality in the absence of therapy. Then, by applying the relative proportional effect of a given therapy, they can calculate the odds of that individual patient’s benefit in absolute percentage terms. For example, a 50 year old woman with a 2 cm, node negative, estrogen receptor (ER) positive, HER2 negative breast cancer is exposed to a 15–20% risk of distant recurrence over the succeeding 10 years if she is treated with local therapy (surgery, radiation) only. Adjuvant endocrine therapy (tamoxifen, aromatase inhibitors) reduces this risk by 40–50%. Therefore, approximately 6–10% of such women will benefit from application of adjuvant endocrine therapy, and 10–15% remain at “residual risk” in spite of the endocrine therapy. In other words, 80–85% of such women did not need adjuvant therapy, and of those who did, approximately 5–10% did not benefit. Given the relatively low toxicity profile of adjuvant endocrine therapy (especially the very low risk of life-threatening toxicities), most physicians would recommend, and most patients accept, at least 5 if not longer years of treatment, to gain this small potential benefit.

One might assume that adjuvant chemotherapy might further reduce this patient’s odds of recurrence by approximately one-third [5]. In this case, one-third of the 10–15% of patients with residual risk will be spared a recurrence, or in other words, 3–5% of such women will benefit. Nearly 100% of women will suffer the annoying
side effects of adjuvant chemotherapy, and sadly the life-threatening toxicities range from 1 to 2% [6]. Does this 3–5% absolute benefit outweigh the 1–2% side effect and life-threatening toxicity profile? This dilemma is very difficult, and different patients will make very different decisions [7].

Taken together, these considerations highlight the importance of generating highly accurate and validated tumor biomarker tests to help guide such decisions. However, introducing such tests into the clinic should not be taken lightly. Use of inaccurate tests to withhold adjuvant systemic therapy, either because they do not have analytical validity or because their clinical utility has not been established, is as dangerous as using therapeutic agents that have not been prepared properly or vetted in well-designed, prospective trials.

A matter of semantics

Over the last several decades, investigators and guidelines bodies have made efforts to organize tumor biomarker research and reporting in a fashion similar to that for therapeutics [8–18]. Importantly, The National Office of Public Health Genomics at the Centers for Disease Control and Prevention convened the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group, which articulated three semantic terms that are important for translation of laboratory findings to useful diagnostic tools to manage patients [19]. Analytical performance (analytical validity) describes how accurately and reliably the test detects and measures the analyte(s) of interest. Clinical validity provides some insight into how well the test relates to the clinical outcome of interest, e.g., response to therapy, survival, etc. However, clinical validity does not imply that the test should be used to guide clinical care. Rather, patient care should be altered only by tests that have been shown to have clinical utility, defined as whether the results of the test provide information that can contribute to and improve current optimal management of the patient’s disease. Indeed, before a tumor biomarker test is introduced into standard clinical management, it must have both analytical validity and clinical utility.

How does one determine clinical utility? As with therapeutics, it is important to insist on high levels of evidence that the tumor biomarker results do, in fact, improve patient outcomes when compared to management of the patient without the marker results. In 1996, members of the American Society of Clinical Oncology (ASCO) Tumor Marker Guidelines Committee proposed a level of evidence (LOE) scale for tumor biomarker tests, ranging from 1 (best) to 5 (worst) [20]. A revision of this scale was published in 2009 presenting a hierarchy of the types of studies that might be used to generate level 1 evidence [21] to demonstrate clinical utility for a specific use. Perhaps more importantly, these authors suggested that one could use archived specimens to do so, but they stressed that these specimens should be collected within previously conducted prospective trials that addressed the specific intended use for the marker. Alternatively, if such specimens are not available, then a prospective clinical trial is probably required to test the test.

Recently, a committee convened by the Institute of Medicine (IOM) of the United States to evaluate development and evaluation of a new omics-based biomarker test [22]. This Committee endorsed the EGAPP terminology, and put forward a roadmap for development of an omics-based test from initial concept to ultimate clinical utility. While specifically generated for omics-based tests, this roadmap is applicable to any diagnostic, and in particular tumor biomarker tests, (Fig. 1) [22]. This process involves the following three steps: 1) discovery of a biomarker that might be of biological and perhaps clinical interest, 2) analytical development of a test for the biomarker that has clinical/biological validity, and 3) evaluation of clinical utility for an intended use. The three strategies proposed by the IOM to generate high levels of evidence necessary to demonstrate clinical utility (illustrated in the far right panel of Fig. 1) follow the recommendations of Simon, Paik, and Hayes quite closely [21]. In particular, if one uses the strategy of prospective retrospective studies to generate high levels of evidence for clinical utility using archived specimens, investigators should ideally conduct prospective retrospective studies using trial quality methodologies [21].

In particular, the IOM Committee strongly recommended that any diagnostic test be “locked down” before it is applied to specimens for studies regarding clinical utility. In other words, the assay should be completely stable analytically, and no changes should be made to the standard operating procedure and protocols developed during the test development phase [22]. All too often, unstable and analytically invalid assays are taken into the clinical utility stage of test analysis, requiring changes to the assay. In therapeutic trials, such an approach is strictly forbidden—for example, one would not change the formula, dose, or schedule of a therapeutic agent in the midst of a prospective trial without a formal, written amendment, probably involving re-initiation of the trial with new accrual goals and new analysis of the statistical power of the trial. Such rigor should be applied to studies of the clinical utility of a tumor biomarker assay, since the results are intended to guide therapeutic decision making with the biomarker test.

**Fig. 1.** Tumor biomarker test development process. The first stage encompasses discovery and test development and validation phases. In the second stage, an analytically-validated tumor biomarker test is evaluated for clinical utility, either in a prospective-retrospective study using archived specimens or in a prospective clinical trials designed to “test the test” for its intended clinical use. From Ref. [22] with permission.
What are the strategies to generate high levels of evidence of clinical utility?

As noted, if available, archived specimens collected from patients who have participated in prospective studies that address the intended use for the tumor biomarker test may be used in “prospective retrospective” studies to determine clinical utility [21]. It is strongly encouraged that such studies be prospectively planned, preferably within a written protocol precisely stating how patients to be included will be selected, the exact analytical methodology for the tumor biomarker assay, and what cut points will be used to distinguish “positive” from “negative”. Such a protocol must include a detailed statistical section stating power and the statistical modeling and testing to be incorporated in the analysis.

Nonetheless, even though performed “prospectively”, such studies contain biases that are not inherent in true prospective studies, such as specimen availability (it is almost always far less than 100%), pre-analytical concerns and assay failure, and under-powering (since the parent trial is almost always powered for the main therapeutic effect, not the tumor biomarker sub-analysis). Thus, it is essential that findings from such a trial be validated using specimens from a second data set, preferably from another prospective trial designed and conducted in a similar manner. Unless similar results, with similar magnitude of benefit for the patients, are demonstrated, the biomarker assay test should not be used to care for patients. If such specimens are not available, one must conduct prospective trials to “test the test”. There are several such trial designs, each with strengths and weakness, and the reader is referred to excellent reviews on this topic [23–25].

Recently an international registry has been established so that investigators can prospectively document their intention to perform a tumor biomarker study, in a manner similar to clinicaltrials.gov [26]. Thus, regardless of which strategy is pursued, an investigator can document that the study protocol, methods, and analytical techniques were prospectively considered.

What is the regulatory oversight of tumor biomarker tests?

A new therapeutic agent cannot be marketed in the Western world without approval from some regulatory body—in the case of the United States, the Food and Drug Administration (FDA). In this regard, the FDA often convenes an Oncology Drug Advisory Committee (ODAC) to review whether a new agent is sufficiently safe and effective to be administered routinely for its intended use. This is not the case for tumor biomarker tests. Indeed, there are two regulatory pathways in the U.S. for bringing a tumor biomarker test onto the market for clinical use: 1) application for FDA approval or clearance, under pre-market approval (PMA) or substantial equivalence (510K) mechanisms, respectively; and 2) development of a laboratory developed test (LDT) within an individual laboratory following practices described under the Clinical Laboratory Improvement Amendments (CLIA) of 1988. CLIA has no requirement that, or even a review process to determine if, individual tests have either clinical validity or utility, and LDT review processes are not universally required for all laboratories performing LDTs. Since FDA has historically exercised enforcement discretion towards most LDTs, they do not currently require premarket review by FDA.

For approval or clearance of a tumor biomarker test, FDA has not adopted the precise EGAPP terminology. FDA does require analytical validation of a new test and evidence that the test performance is aligned with the claims of the manufacturer. However, FDA does not require high levels of evidence that use of the test improves clinical outcome when compared to absence of the test (in other words, “clinical utility”). The uncertain regulatory environment has resulted in approval and marketing of tumor biomarker tests which, arguably, do not have clinical utility, use of LDTs which do not have analytical validity or clinical utility, and marketing of assays that have never been reviewed or approved/cleared by FDA. Thus, clinicians are left in a quandary about whether the results they receive from their clinical and pathology laboratories are accurate or worth using.

Reporting guidelines for tumor marker studies

Publication in peer-reviewed papers is one of the hallmarks of the scientific process. Peer-reviewed publication provides external perspectives of the data at hand, and presents the data to others who might use them—either for replication of the study or, preferably, application in the clinic. Unfortunately, the rigor of peer review for tumor biomarker tests has fallen far short of that for tumor treatment studies [17]. Whereas almost all therapeutic reports include descriptions of the protocols, including patient eligibility, trial conduct, and data analysis, such information has been rarely included in tumor biomarker papers. Over the last decade, two initiatives have been published, and endorsed by several journals, to improve reporting requirements for pre-analytical (BRISQ), and analytical (REMARK) considerations of tumor marker tests [17,27–37]. Recently, similar guidelines have been published for conduct and reporting of circulating tumor biomarkers (MONITOR) [38].

Case studies

Over the last two decades, several hundred putative tumor biomarker assays have been reported in the peer-reviewed literature. In spite of this plethora of publication, the ASCO Tumor Marker Guidelines Committee has only recommended a handful of tumor biomarker tests that they deemed had been shown, with high levels of evidence, to have sufficient analytical validity and clinical utility to guide treatment of patients with breast cancer. The recommended biomarkers include ER (and progesterone receptor), HER2, and urinary plasminogen activator/plasminogen activator inhibitor 1 (UAPA/PAI1) [18]. Subsequent collaborations between ASCO and the College of American Pathologists (CAP) have reviewed specific assays for ER, PgR, and HER2 [39–42]. In addition, the ASCO Tumor Marker Guidelines Committee has reviewed multi-parameter gene expression tests as well, and has recommended the use of the 21-gene recurrence score (RS) assay to determine prognosis in patients with node negative, ER positive breast cancer.

ER and adjuvant tamoxifen

Why are these guidelines so conservative? The Committee has attempted to only recommend assays for which high levels of evidence exist to document clinical utility for a specific, intended use. For example, adjuvant tamoxifen clearly reduces the odds of recurrence and mortality. Several of the original randomized trials included both women with ER positive and women with ER negative cancers, permitting the opportunity to pool these into a large, high level of evidence study that clearly documents that this effect is solely confined to patients with ER rich tumors [43]. Although many different assays were used in these trials, immunohistochemistry is now almost exclusively used in most clinical laboratories, and the recent efforts by the ASCO/CAP committee has resulted in better standardization and, therefore high analytical validity of this test [41,42]. Thus, tests for ER meet the criteria of analytical validity and clinical utility for determination of whether anti-estrogen therapy should be offered, and are used daily to
with-hold adjuvant tamoxifen from patients with ER negative (poor) disease.

21 gene recurrence score and prognosis

A major concern among many oncologist is the over-use of adjuvant chemotherapy. It is clear that many patients who either do not need it, or whose cancers may not respond, are treated so as to ensure that those who do need it and will respond are treated, as illustrated in the case reports described above. Several recent prospective retrospective studies conducted using archived specimens from previously conducted trials in which node negative, ER positive women only received tamoxifen has demonstrated that those with low RS, as determined by the 21-gene RS assay, have such a favorable prognosis that no more than 1–2% will benefit from adjuvant chemotherapy, even if their tumors are sensitive to the treatment [44,45]. These prognostic data are considered to represent level 1 evidence, and the assay is recommended by the ASCO Tumor Guidelines Committee for this use.

Prediction of response to chemotherapy

Several studies over the last 40 years have suggested that chemotherapy may not be equally active in all biological subgroups, regardless of prognosis [46]. In general, these studies suggest that patients with endocrine sensitive (hormone receptor positive), HER2 negative, and low proliferation may be relatively more resistant than patients with HER2 positive and/or ER/PgR/HER2 negative or high proliferative breast cancers. Intrinsic sub-typing has been used to classify these relative subgroups as luminal A and B (ER positive), HER2 like, and basal-like [47], and these terms have become useful shorthand to describe these biological subgroups [46]. It has been suggested that some measure of these, and other genes, might identify patients who are so unlikely to respond to chemotherapy that they should not be treated, in a manner analogous to with-holding tamoxifen from ER negative patients [48]. Indeed, at least two prospective, retrospective studies have indicated that patients with low RS do not benefit from adjuvant chemotherapy — not because they do not need it (prognosis) but because it may not work (prediction) [49,50]. However, these studies were retrospective, represented only a subgroup of enrolled patients in the trials, and one of the studies was used as the test set to develop the algorithm. Therefore, the available body of evidence falls short of level 1 for the predictive role of these tests in routine clinical care to with-hold chemotherapy based the assumption of resistance to it [46].

Others have argued that the 21 Gene RS test is not needed to determine if a patient has luminal, chemo-resistant breast cancer [48]. Rather, they have proposed that measurement of ER, PgR, HER2, and Ki67 using immunohistochemistry (IHC) is sufficient to sub-classify patients into those likely or not to respond to chemotherapy. Dowsett et al. have demonstrated high concordance between the so-called “IHC4” test and the 21-gene RS [51,52]. However, this test is a highly analytically validated assay performed in one laboratory using a sophisticated mathematical algorithm to determine the IHC4 score. It is very different from IHC analysis performed in routine clinical pathology laboratories in various hospitals, with different methods and using different cutpoints for each marker.

Two pieces of data serve as cautionary notes to adoption of IHC for ER, PgR, HER2, and Ki67 to with-hold adjuvant chemotherapy. First, a recent update of the Oxford Early Breast Cancer Trials’ Collaborative Group (EBCTCG) overview has failed to demonstrate any biological subgroup that did not enjoy a reduction of risk of recurrence of approximately 30% due to adjuvant chemotherapy [5]. The studies included in the EBCTCG review only included ER, PgR, and tumor grade, and these assays were performed in many ways with variable quality control. They did not include HER2 or specific measures of proliferation. Therefore, authors have argued that the EBCTCG dataset does not represent a valid review of the hypothesis that luminal A cancers are resistant to chemotherapy [48]. In this regard, it has been suggested that one needs to include HER2 and especially Ki67, which these authors (and apparently the majority of panel members attending the St. Gallen Symposium) propose can be performed reliably using IHC. However, an effort to standardize Ki67 analysis by the International Ki67 Working Group has demonstrated an alarming lack of concordance among some of the world’s most accomplished pathologists, even when pre-analytical issues were removed from consideration [53,54].

In summary, it is appealing to reduce the amount of unnecessary adjuvant chemotherapy delivered to women who either do not need it or whose tumors are not responsive to it. In this regard, the 21-gene RS test has been shown, with high levels of evidence, to have clinical utility in node negative, ER positive women as a prognostic test — those with low RS are so unlikely to recur that even if chemotherapy works, it cannot help a sufficient number of women to outweigh its risks and costs. Use of other tests, such as IHC for ER, PgR, HER2 and Ki67, to make similar decisions is fraught with the uncertainty of analytical validity and therefore inaccuracy, and doing so should be viewed with considerable caution.

Use of any test to predict sensitivity or resistance to chemotherapy, especially in those who have a poor prognosis based on anatomic findings, such as positive axillary lymph nodes, has not yet been supported by high levels of evidence — especially for IHC analysis of ER, PgR, HER2, and Ki67. Indeed, a prospective randomized trial (the RxPonder Trial) is now being conducted by SWOG to address this issue. In the RxPonder trial, women with node positive, ER positive breast cancers have the 21 gene RS test performed, and if their RS is <25 they are randomly assigned to chemotherapy or not (they all receive appropriate endocrine therapy). Until this or other similar studies, such as the TailorRx and the MINDACT trial, are completed, it must be considered very risky to with-hold adjuvant chemotherapy from such patients, given the high stakes of being wrong.

Summary

In summary, a Bad Tumor Marker Is as Bad as a Bad Drug! If we are going to use tumor biomarker assays to direct patient care, then the clinician needs to be as certain about the analytical validity and clinical utility of the test as he/she is about the manufacture and clinical data regarding an anti-neoplastic. Only strict vigilance in applying the scientific method with rigor will result in generation of tumor biomarker tests in which the clinician will have confidence.

Conflict of interest statement

DH has received a stock grant from Oncimmune LLC, and research grants from Pfizer, Novartis, Veridex, and Janssen R&D. He has received an honoraria (educational grant) from Roche.

References
