Hypothesis/Commentary

The Road from Discovery to Clinical Diagnostics: Lessons Learned from the First FDA-Cleared *In Vitro* Diagnostic Multivariate Index Assay of Proteomic Biomarkers

Zhen Zhang and Daniel W. Chan

Abstract

Background: After more than a decade of biomarker discovery research using advanced genomic and proteomic technologies, very few biomarkers have been translated into clinical diagnostics for patient care. This has become an urgent issue to be addressed because the continuing funding from both the public and private sources are called into question.

Methods: We use as an example, OVA1, the first *in vitro* diagnostic multivariate index assay (IVDMIA) of proteomic biomarkers recently cleared by the US FDA (Food and Drug Administration) to describe our experience through the long road from biomarker discovery, to validation, and finally to multi-institutional trial for regulatory approval by the FDA.

Results: We discuss 3 issues that are key bridges in the path of biomarker development to actual clinical diagnostics: 1) to generate sufficient and "portable" evidence in preliminary validation studies to support investment for large-scale validation trials; 2) to carefully and clearly define clinical utility that balances desire for broad applicability and feasibility for completing clinical trials for regulatory approval; and 3) to select/develop assays with analytical performance suitable for clinical deployment.

Conclusions: We learned that the road from biomarker discovery, validation, to clinical diagnostics could be long and winding, and often frustrating. However, we also know that, with the right approaches, at the end of the road, there is a rainbow waiting for us.

Impact: Provide insights and recommendations for the translation of proteomic biomarkers into clinical diagnostics. *Cancer Epidemiol Biomarkers Prev; 19(12); 2995–9. ©2010 AACR.*

Introduction: Bridges in the Paths from Biomarker Development to Clinical Applications

Advances in genomic and proteomic technologies for the analysis of clinical specimens have generated an ever increasing number of publications on the discovery of novel biomarkers and their potential clinical applications. A recent search of PubMed using the simple term of "biomarker" returned closed to half a million of entries. Search with the more specific phrase of "ovarian cancer biomarker" still returned more than 8,000 entries. However, with this proliferation of "biomarker discoveries," very few new biomarkers have been rigorously validated and become available for actual clinical use. In fact, for the past decade or so, few new tumor markers have been cleared or approved for clinical use by FDA. Successful

translation of biomarker discoveries has become an urgent issue not only for the many unmet needs in patient care but also to justify continued support for biomarker research from public and private funding sources (1).

The OVA1 test is an *In Vitro* Diagnostic Multivariate Index Assay (IVDMIA) of Proteomic Biomarkers that has been recently cleared by the FDA for assessing ovarian cancer risk in women diagnosed with ovarian tumor prior to a planned surgery. OVA1 analyzes 5 proteomic biomarkers in serum and the results are combined through an algorithm to yield a single-valued index within the range of 0–10. A menopausal status-dependent cutoff is used to classify a patient into high- or low-risk group. OVA1 provides additional information to assist in identifying patients for referral to a gynecologic oncologist. A number of studies have confirmed that ovarian cancer patients operated by oncologist specialist tend to have overall better outcomes (2). In a prospective multiple-center clinical study, the addition of OVA1 in pre-operative clinical assessment was found to improve sensitivity in the prediction of malignancy for ovarian tumor.

The road from development of biomarkers to clinical practices could take many possible paths. However, it is unequivocal that prior to clinical use, any biomarkers

Authors' Affiliation: Center for Biomarker Discovery, Department of Pathology, Johns Hopkins Medical Institutions, Baltimore, Maryland

Corresponding Author: Daniel W. Chan, Center for Biomarker Discovery, Department of Pathology, Johns Hopkins Medical Institutions, Baltimore, MD 21231. Phone: 410-955-2674; Fax: 1-443-287-6388. E-mail: dchan@jhmi.edu

have to prove their safety and efficacy in independent clinical trials using an appropriate study population for a clearly defined intended use. In this article, we select following 3 issues that are the key "bridges" in the path from discovered biomarkers to actual clinical diagnostics:

1. to generate sufficient and "portable" evidence in preliminary validation studies to support investment for large-scale validation trials;
2. to define clinical utility that balances desire for broad applicability and feasibility for completing clinical trials, and for regulatory approval; and
3. to select/develop assays with analytical performance suitable for clinical deployment.

In the following, we will explain how these issues have played critical roles at the various decision points during the development of the OVA1 test. We believe that lessons we learned from this process could be generalized in many current efforts to bring biomarker discovery into clinical diagnostics.

Evidence of Portability of Biomarker Discriminatory Power across Multiple Sites

The final panel of biomarkers in OVA1 consists of CA125, transthyretin (prealbumin), apolipoprotein A1, beta 2 microglobulin, and transferrin. These biomarkers, other than CA125, were part of 7 biomarkers discovered through proteomic analysis of serum specimens from multiple centers (3; 4). Ideally, nested case-control study design using prospectively collected samples from a cohort (e.g., the prospective randomized open blinded end-point study design; ref. 5) could theoretically avoid many typical sources of biases and confounding factors in a pure case-control study of retrospective samples. However, such "pristine" samples are often scarce and not always available for discovery studies. As an alternative, we have proposed the use of samples from multiple sites each with its own cases and controls. We believe that under the assumption that many of the biases and confounding factors are more likely to be site-specific, the use of multisite samples at discovery will allow us to cross-compare and validate discoveries independently derived from each of the individual sites to identify biomarkers that are more likely to be "portable" from sites to sites, less sensitive to variations in sample collection, processing, and storage conditions (6), and have a better chance to ultimately survive the required multicenter clinical trial study to gain clearance for clinical use.

Even though the initial discovery studies of these proteomic biomarkers involved more than 500 patient specimens, for actual development of a clinical diagnostic test, additional evidence is often needed to prove their effectiveness for a more clearly defined clinical utility and to further test their portability over correspondingly defined target populations from diverse sites. In the case of OVA1, the desired clinical utility was determined to be the assessment of ovarian cancer risk among women with known pelvic masses. In Figure 1, we show the portability of the 7 proteomic biomarkers (without CA125) in separating ovarian cancer from patients with benign pelvic masses. In Figure 1A, ovarian cancer and benign tumor samples from 1 clinical site are plotted in the first 2 dimensions resulted from unsupervised cluster analysis [principal component analysis (PCA)] of measurement of the biomarkers. The coefficients of the first 2 PCA dimensions are fixed and then used to project samples from 5 independent clinical sites in Figure 1B. These plots, through unsupervised analysis, show the natural separation of patients with benign ovarian tumors from those with malignant tumors. Furthermore, the pattern of separation persists over samples from geographically extremely diverse sites. To some degree, the decision
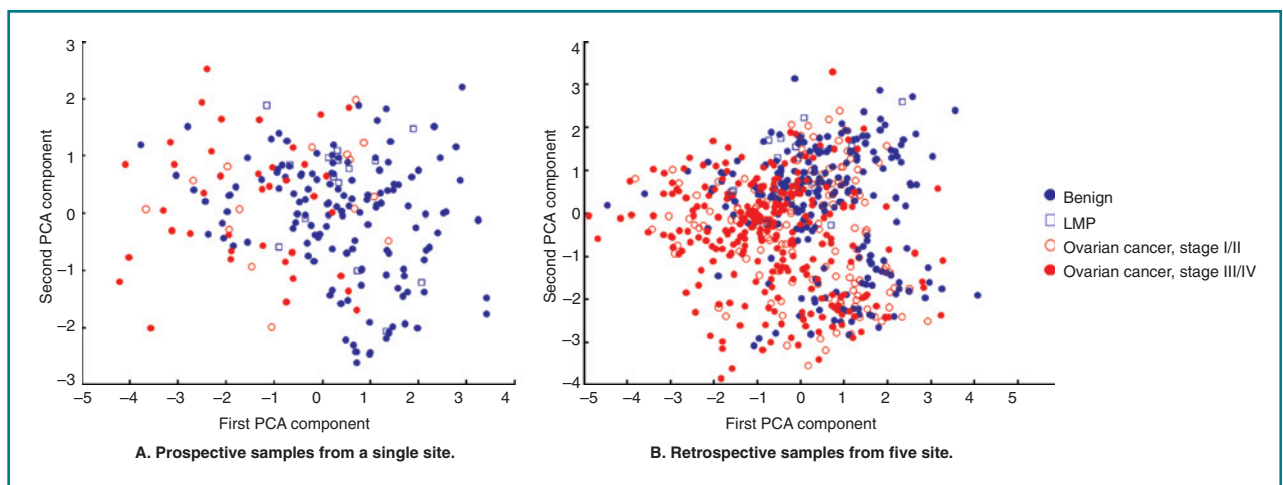

Figure 1. PCA of clinical samples from multiple sites using 7 proteomic biomarkers. A, scatterplot of samples from a prospective study in the first 2 PCA dimensions. B, Scatterplot of samples from 5 independent clinical sites in the first 2 PCA dimension.

to move forward a panel of biomarkers for further development should rely more on such "portability" of discriminatory power than how well they do over a set of samples from a single site.

Definition of Clinical Utility

Ideally, a clearly defined clinical utility should be integrated into the study design at the earliest phase of biomarker discovery that would allow the selection of the right samples and ask the right question to find the right biomarkers. However, the reality of biomarker research is that many biomarkers under study are the results of exploratory studies analyzing expression profiles of clinical samples that are heterogeneous within groups. It is expected that with progression in the stage/phase of the biomarker study and definitely as part of the decision to further develop them into clinical diagnostics, the intended clinical utility of the diagnostics and the corresponding target populations will have to be further refined.

The clinical applications of a biomarker should not be simply defined as biomarker for a particular cancer, or even by terms such as "detection," "screening," or "prognosis" alone. The clinical applications of a biomarker should be defined by their clinical utilities at a specific decision-making point along the disease progression path. In other words, we need to clearly define "Who" are the people the biomarkers are intended for (a point along disease progression, such as general population, high-risk population, or women with a pelvic mass, etc.) and take into consideration what would happen to the patient if the biomarker results are positive (elevated risk of cancer) or negative, and the "costs" of false positive or negative results.

There is always a tradeoff between having a broad applicability of a diagnostic test and the practical feasibility of completing the necessary validation studies. Such decisions require a good understanding of the disease epidemiology and clinical reality including available means of clinical interventions for positive test results and the consequence of false positives and negatives.

The defined clinical utility determines the desired performance characteristics of a diagnostic test under development and conversely is also constrained by the actual performance of the test. In Figure 2 we illustrate in a tabular form simplified examples of how the desired/ required performance of a test could be affected by 4 interwoven factors: 1) size of biomarker target population; 2) disease prevalence; 3) consequence of false negative; and 4) consequence of false positive. We use 2 examples of clinical applications: a) OVA1 is intended to assess preoperatively the risk of ovarian cancer in women scheduled for surgery due to suspected ovarian cancer. The test result aids in the decision to refer the woman to a gynecologic oncologist for surgery for better long-term outcome. b) OcaScr is a fictitious test to screen
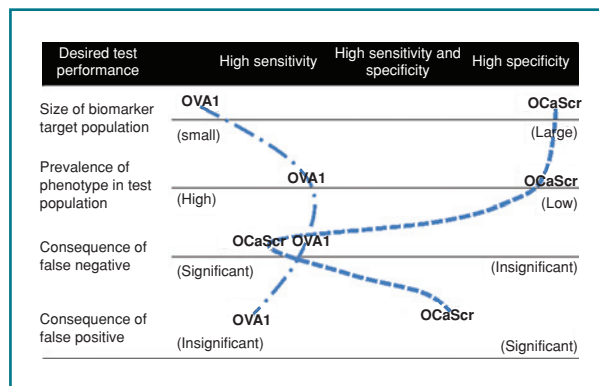

Figure 2. A simplified multidimensional space characterizing clinical needs for cancer detection and desired performance.

for ovarian cancer in postmenopausal women. In Figure 2, we can see that the target population for OVA1 is relatively small; yet the prevalence of ovarian cancer is quite high in this population (∼30% based on actual trial data); the consequence of false negative is only relatively significant because the woman will still have the surgery albeit by a nonspecialist; the consequence of false positive will mostly be the additional cost and effort to have surgery by a gynecologic oncologist. Most of OVA1 entries are on the left side that requires a test to have a high sensitivity. On the other hand, an ovarian cancer screening test will have a very large target population with an extremely low prevalence (∼1/2,500); the consequence of a false negative will be quite significant because it is a fast progressing and deadly disease; and the cost of workup procedures and confirmatory surgery from a false positive result is also relative high. Overall, entries of OCaSc are mostly in the right side and span across to the left too, indicating the need for a very high specificity and a reasonably high sensitivity.

Be Mindful of Assay Analytical Performance

For a diagnostic test to be cleared or approved for use in a clinical setting, a set of well-established criteria for assay analytical performance needs to be satisfied. One should not assume that research assays used for discovery, especially those that involve complex laboratory methods, can be directly translated to a robust clinical assay. The original development of the OVA1 test involved the use of 7 proteomic biomarkers to be measured by surface-enhanced laser desorption/ionization (SELDI) mass spectrometry analysis, the method for the original discovery work. However, after a prolonged effort and extensive cost, it became obvious that the SELDI assay was not able to have the required analytical performance as a clinical assay. The final choice of the 4 proteomic biomarkers and the addition of CA125, all measured by immunoassay, reflect the compromised choice due to the consideration of analytical performance (7).

Poor analytical performance not only hinders the deployability of biomarkers for clinical applications, but also increases the required sample size for clinical validation studies. Unfortunately, the required expertise and effort with associated cost for assay development are often underappreciated or even ignored in the development of biomarkers for clinical use. For example, a clinical assay needs to be robust and high throughput, that is, capable of analyzing a sufficiently large number of samples routinely over a sufficiently long period of time without the need for constant human intervention. This is often very different from the concept in a research setting for an assay to be considered as a "high-throughput" and "reproducible."

In addition to the analytical performance of individual assays, the development of an IVDMIA need also pay attention to the impact of the mathematical/computational algorithms in the derived multivariate models on the analytical performance of the single-valued IVDMIA results. It is possible that for certain combinations of biomarker values, the algorithm/formula could actually result in an amplified analytical variability in the IVDMIA results. It could be as simple as a mathematical division by an input biomarker that exhibits poor analytical behavior when its value is close to the limit of quantitation. However, for nonlinear multivariate algorithms, such "hot spots" of amplified variability at specific combinations of input values may not be easily identified through direct analysis of the mathematical models. A possible solution is to assess the analytical performance of the IVDMIA through statistical simulation based on available clinical data and individual assay precision data. Such assessment should be conducted during IVDMIA model development to eliminate or minimize these "hot spots" and be done again later as part of the IVDMIA precision study.

Final Thoughts and Comments

The biomarkers that were included in the OVA1 panel with CA125 were discovered during a time when clinical proteomics for biomarker discovery just started. Both the tools, such as the SELDI technology, and the discovery results reported in the literature have been the subjects of much debate and criticism (8, 9). From the beginning, we always believe that proteomic profiling by mass spectrometry may not be sufficient and reliable as diagnostics. However, we took advantage of the profiles that could differentiate disease from nondisease and were able to identify the actual proteins as potential biomarkers. Some of these promising biomarkers were used in OVA1. Using a multicenter study design and stringent statistical analysis approaches, we avoided or alleviated the effect of several commonly observed sources of biases and confounding variables. It is our belief that technologies are only to be used as tools. As scientists, what is important is to understand the strengths and weaknesses of a tool and to use it properly. The newer and better genomic and proteomic analysis technologies will allow us to see a much greater range of molecular changes in clinical samples. However, if we do not use the tools well and fail to pay attention to common sense yet critical issues, such as study design, randomization in experimental design, etc. (6), mistakes happened in the early time of clinical proteomics could repeat and possibly at a greater cost (10).

With the limitation of mass spectrometry, some of the discovered biomarkers that are used in OVA1 were of a relatively high level of abundance and often considered as being acute phase reactants and their specificity to cancer had been questioned (11). The decision to include them in the OVA1 panel were based on 1) that existing evidence linking inflammation and cancer initiation/progression made it plausible, as proven by our data, that such biomarkers could still provide complementary values to CA125 to improve sensitivity in detecting cancer, and 2) the targeted population of OVA1 (preoperative assessment of patients with confirmed ovarian tumor) and the inclusion of CA125 in the panel help to minimize the effect of possible nonspecificity of these biomarkers.

From the discovery of biomarkers to their use for a specific clinical indication, it requires the resolution of many interwoven issues and knowledge and expertise from very diverse areas. A basic understanding of these issues and appreciation of their complexity will help the collective effort of the biomarker research community to translate the large number of potential biomarkers into clinical diagnostics. In this article, we use the development of the OVA1 IVDMIA test as an example to discuss several critical issues that are key bridges in the path of biomarker development into clinical diagnostics and necessary steps prior to the commencement of clinical trials to show clinical utility and safety to obtain regulatory approval. First, we need to define carefully and clearly a specific clinical "intended use" in order to balance the desire for broad applicability and feasibility. Second, we need to generate sufficient and "portable" evidence in preliminary validation studies to support the investment for assay development and large-scale validation trials. Third, we need to select/develop assays with analytical performance suitable for clinical deployment. We learned that the road from biomarker discovery, validation, to clinical diagnostics could be long and winding, and often frustrating. However, we also know that, with the right approaches, at the end of the road, there is a rainbow waiting for us.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgment

References

1. Chan DW. Will cancer proteomics suffer from premature death?Clin Proteom 2010;6(1–2):1–3.
2. Giede KC, Kieser K, Dodge J, Rosen B. Who should operate on patients with ovarian cancer? An evidence-based review. Gynecol Oncol 2005;99(2):447–61.
3. Zhang Z, Bast RC Jr, Yu Y, Li J, Sokoll LJ, Rai AJ, et al. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. Cancer Res 2004;64(16):5882–90.
4. Rai AJ, Zhang Z, Rosenzweig J, et al. Proteomic approaches to tumor marker discovery – Identification of biomarkers for ovarian cancer. Arch Pathol Med 2002;126(12):1518–26.
5. Pepe MS, Feng Z, Janes H, Bossuyt PM, Potter JD. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: Standards for study design. J Natl Cancer Inst 2008; 100(20):1432–8.
6. Zhang Z, Chan DW. Cancer Proteomics: In Pursuit of "True" Biomarker Discovery. Cancer Epidemiol Biomarkers Prev 2005;14(10):2283–6.
7. Fung ET. A recipe for proteomics diagnostic test development: The OVA1 test, from biomarker discovery to FDA clearance. Clin Chem 2010;56(2):327–9.
8. Petricoin E, Liotta La. Counterpoint: The vision for a new diagnostic paradigm. Clin Chem 2003;49(8):1276–8.
9. Diamandis EP. Point: Proteomic patterns in biological fluids: Do they represent the future of cancer diagnostics?Clin Chem 2003;49 (8):1272–5.
10. Diamandis EP. Cancer Biomarkers: Can We Turn Recent Failures into Success?J Natl Cancer Inst 2010;1–6.
11. Schweigert FJ, Sehouli J, Zhang Z, Chan DW. Transthyretin, a biomarker for nutritional status and ovarian cancer. Cancer Res 2005;65 (3):1114.

The Road from Discovery to Clinical Diagnostics: Lessons Learned from the First FDA-Cleared *In Vitro* Diagnostic Multivariate Index Assay of Proteomic Biomarkers

Zhen Zhang and Daniel W. Chan

Updated version	Access the most recent version of this article at: doi: 10.1158/1055-9965.EPI-10-0580

Cited articles	This article cites 10 articles, 6 of which you can access for free at: http://cebp.aacrjournals.org/content/19/12/2995.full#ref-list-1
Citing articles	This article has been cited by 8 HighWire-hosted articles. Access the articles at: http://cebp.aacrjournals.org/content/19/12/2995.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://cebp.aacrjournals.org/content/19/12/2995. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.