A Recipe for Proteomics Diagnostic Test Development: The OVA1 Test, from Biomarker Discovery to FDA Clearance

Eric T. Fung

Although diagnostic test development remains challenging, novel technologies, including proteomics, genomics, and microRNA analysis, provide opportunities to identify biomarkers that in principle could accelerate the development of new diagnostic tests. Unfortunately, the literature is littered with initial biomarker discoveries that have failed to reach the clinic. We sought to identify a recipe that combines the basic ingredients of clinical research with novel analytical tools to create a new diagnostic test.

In the test kitchen, we understood that the most important ingredient is the unmet clinical need. Having made the decision to develop a test for ovarian cancer, we discussed with clinicians what they felt were the most pressing needs in this field. Our initial instinct was to pursue ovarian cancer screening, but in our conversations with key opinion leaders, including Ian Jacobs (University College London), Bob Bast (MD Anderson), and Daniel Chan (Johns Hopkins University), we came to understand that because of the low prevalence of ovarian cancer, development of a screening test would require large studies that would exceed our budget and time constraints. Additionally, because a positive initial result in a screening test would likely lead to pelvic surgery, the test would demand a level of clinical specificity that we were unlikely to achieve. Our colleagues, however, identified a critical unmet need in the area of ovarian tumor triage. Although ovarian tumors are relatively common, only a fraction of them are malignant. Being able to identify the malignant ones preoperatively would permit better preoperative management of women with ovarian tumors. In particular, women with a high likelihood of malignancy could benefit from referral to specialist surgeons (e.g., gynecologic oncologists), who would be able to perform debulking and staging surgeries that form the basis of optimal care for ovarian cancer (1).

Having identified the clinical question, we found it was time to turn our attention to the identification of biomarkers. Numerous technologies exist for biomarker discovery, each with inherent advantages and disadvantages. Our test kitchen focused on SELDI-TOF mass spectrometry, a technology with a throughput that allows the assessment of relatively large numbers (i.e., hundreds) of clinical samples. We knew that the choice of technology, although important, is secondary to the quality of the clinical samples. On the clinical side, the samples should be appropriately pedigreed, i.e., with adequate clinical history. On the analytical side, we needed to understand the handling history of the samples (e.g., how they were collected, number of freeze–thaw cycles, and so on). We needed to balance real life with the ideal scenario. Our initial studies were entirely retrospective, but we attempted to mitigate variability in analytical and clinical quality by running relatively large studies (500+ in one of our very early studies) from multiple institutions (2). These experiments were done with our colleagues from Johns Hopkins (Daniel Chan and Zhen Zhang). We reasoned that the use of samples from these various institutions would mitigate some of the biases caused by variation in demographics, sample acquisition, and sample handling across these institutions. A marker robust enough to reveal itself under these relatively poorly controlled conditions might be robust enough to withstand further validation, particularly as clinical and analytical parameters became more explicitly defined.

After this initial study, we had some biomarker candidates that, in conjunction with cancer antigen 125 (CA125), appeared to provide a modest improvement over CA125 alone for the detection of early-stage ovarian cancer. From these preliminary results, we proceeded to the analysis of additional samples and even began a prospective collection with colleagues at Rigshospitalet (Copenhagen), led by Claus and Estrid Hogdall. These studies culminated in our watershed "megastudy," which was a multi-institutional analysis encompassing more than 600 individuals. We then set about to define our best markers. None of our individual biomarker candidates [the final 7 SELDI marker candidates (the final 7 SELDI marker candidates)] were developed.

1 Vermillion, Inc., Fremont, CA.
2 Nonstandard abbreviations: CA125, cancer antigen 125; ITIH4, inter-α-trypsin inhibitor heavy chain 4; CTAP3, connective tissue–activating peptide III; FDA, US Food and Drug Administration; IDE, investigational device exemption.

327
candidates were ITIH4 (inter-α-trypsin inhibitor heavy chain 4), transthyretin, apolipoprotein A1, hepcidin, β2-microglobulin, transferrin, and CTAP3 (connective tissue–activating peptide III), in addition to CA125) were exciting, and indeed some tasters dismissed them as being nonspecific. The results, however, clearly demonstrated that a combination of these markers had the ability to discriminate between benign and malignant ovarian tumors, and, equally importantly, these markers could be validated with samples from all over the world. From then on, we would validate only candidate markers found in this study. This idea went against our creative instincts, which are to constantly look for better markers and to create the perfect test; however, the practical need to deliver a product intervened.

We now had defined 2 of the major ingredients: the clinical question and the marker set. What still remained was the regulatory strategy. A popular approach these days is to open up a CLIA laboratory and to offer the test there. Although we considered this approach, we decided that obtaining US Food and Drug Administration (FDA) clearance would be the better route for our test, particularly given the recent history of multimarker tests for ovarian cancer that had been proposed. We had several rounds of dialogue with the FDA to discuss both our intended use claim and the design of our clinical trial. We initiated our multicenter clinical trial in 2007 with Fred Ueland (University of Kentucky) as our principal investigator and were careful to include sites at which ovarian tumors are typically evaluated, including women’s health clinics, primary care centers, and a handful of academic medical centers.

The need for multimarker algorithms to be developed on one set of samples and validated on a completely independent set of samples has been well established (3). The multicenter clinical trial was the opportunity to independently validate a specific marker algorithm in a prospective, real-life clinical setting. Although we were confident in the performance of our markers and had a number of preliminary algorithms, we needed to have a fixed algorithm. Our first instinct was to develop the final algorithm and assay on the SELDI platform because that platform had been used to discover the markers. For our FDA submission, we knew that we would need to present both clinical and analytical data. When we assessed the analytical performance of the SELDI assays, we realized that the reproducibility of the platform was not adequate for routine clinical use. Even in our expert hands, we still obtained CVs exceeding 10% for individual analytes, and we knew that CVs would be even worse in routine clinical laboratories. Fortunately, immunoassays existed for several but not all components of our assay.

We therefore needed to run the available immunoassays (CA125, β2-microglobulin, transferrin, apolipoprotein A1, transthyretin) on a training set and collect new prospective samples for validation of the immunoassay-based OVA1 assay. This investigation extended our clinical trial beyond our initial forecast, but in the end it was the right decision because the immunoassays were an improvement over the SELDI-based assay, both in clinical performance and in analytical-performance metrics.

We submitted our data in June 2008, and there was substantial dialogue and interaction with the FDA as we explained our data, discussed and refined the intended use claim, and answered clinical and analytical questions. We found this process to be a collaborative one; for example, the FDA asked us to consider setting different cutoffs for premenopausal and postmenopausal women. We obtained clearance on September 11, 2009 (4).

The end product is an ovarian tumor triage test that, when combined with clinical assessment such as imaging and physical examination, has a >90% sensitivity and a 90% negative predictive value in women with an ovarian tumor and for whom surgery is planned. The OVA1 score is between 0 and 10. For premenopausal women, a cutoff of 5.0 is used, and a cutoff of 4.4 is used for postmenopausal women. Scores above these cutoffs indicate a higher likelihood of malignancy. The test is not a screening test, which we did not set out to create. Instead, the finished product is one that can assist physicians in determining whether a patient would benefit from referral to a gynecologic oncologist because of a high likelihood of malignancy. The labeled claim has nuances that even we were not aware of at the outset. For example, we explicitly state in the intended use claim that the OVA1 test should not be used as a stand-alone diagnostic test. Our assumption had always been that physicians would use OVA1 in the context of physical and radiologic findings, but we were instructed to state this qualification outright. After 7 years, we had achieved the test that we had sought to create—an ovarian tumor triage test.

FDA-Cleared Assay

Mise en place is the concept of having all the ingredients in place before starting. This is highly recommended, although, understandably, some flexibility is required. The bouquet garnis can be a combination of any number of members, but it should have at least one of each expert listed below. The final dish is best served with champagne.

Biomarker panel

1 Clinical question
1 Bouquet garni comprising clinician, analytical chemist, statistician
1 (or more) Analytical methods
Several hundred to thousands of clinical samples, divided
Clinical trial
1 Pre-IDE meeting
1 Protocol
1 Bouquet garni comprising clinician, statistician, clinical operations staff
Multiple interactions with FDA

Marinate the bouquet garni with the clinical question. Apply the first batch of clinical samples to the analytical method(s). This should generate a set of candidate biomarkers. If not, reassess the clinical samples and the clinical questions. You may need to marinate longer to determine whether a new clinical question is necessary.

Apply the second batch of clinical samples to an analytical method, preferably the one that will be used for FDA clearance. Evaluate the performance of the candidate biomarkers. This step can be repeated as often as desired, each time with a new batch of clinical samples, until a satisfactory level of confidence in the performance of the biomarkers is reached.

Preheat oven by scheduling a pre-IDE meeting with the FDA.

Mix second and third clinical trial ingredients and, if necessary, some of the fourth ingredient. Season generously with patience and flexibility.

When ready, put the clinical trial in the oven. Periodically check oven for doneness, and, if necessary, add additional interactions with FDA. You may be asked to add additional ingredients requested by FDA.

When fully baked, frame clearance letter, and serve with champagne.³

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors’ Disclosures of Potential Conflicts of Interest: Upon manuscript submission, all authors completed the Disclosures of Potential Conflict of Interest form. Potential conflicts of interest:

Employment or Leadership: E.T. Fung, Vermillion, Inc.
Consultant or Advisory Role: E.T. Fung, Vermillion, Inc.
Stock Ownership: E.T. Fung, Vermillion, Inc.
Honoraria: None declared.
Research Funding: None declared.
Expert Testimony: None declared.

Role of Sponsor: The funding organizations played a direct role in the design and choice of the study, the choice of enrolled patients, the review and interpretation of data, and the preparation and final approval of the manuscript.

References

DOI: 10.1373/clinchem.2009.140855

3 Author’s note: Cooking is one of the author’s hobbies. This perspective is modeled (with apologies) after the recipes found in Cook’s Illustrated magazine.

Congratulations to Dr. Diann Weddle of Abbott Diagnostics for submitting “Amanita phalloides” as the poison used in the December 2009 “Lily Robinson and Up in Arms.” Her name was randomly drawn from dozens of entries as the winner of a $50.00 gift certificate to the AACC Bookstore!