The Long Journey of Cancer Biomarkers from the Bench to the Clinic

Maria P. Pavlou,1,2 Eleftherios P. Diamandis,1,2,3 and Ivan M. Blasutig2,3*

BACKGROUND: Protein cancer biomarkers serve multiple clinical purposes, both early and late, during disease progression. The search for new and better biomarkers has become an integral component of contemporary cancer research. However, the number of new biomarkers cleared by the US Food and Drug Administration has declined substantially over the last 10 years, raising concerns regarding the efficiency of the biomarker-development pipeline.

CONTENT: We describe different clinical uses of cancer biomarkers and their performance requirements. We also present examples of protein cancer biomarkers currently in clinical use and their limitations. The major barriers that candidate biomarkers need to overcome to reach the clinic are addressed. Finally, the long and arduous journey of a protein cancer biomarker from the bench to the clinic is outlined with an example.

SUMMARY: The journey of a protein biomarker from the bench to the clinic is long and challenging. Every step needs to be meticulously planned and executed to succeed. The history of clinically useful biomarkers suggests that at least a decade is required for the transition of a marker from the bench to the bedside. Therefore, it may be too early to expect that the new technological advances will catalyze the anticipated biomarker revolution any time soon.

© 2012 American Association for Clinical Chemistry

The search for cancer biomarkers has become an integral component of cancer research because of the potential of biomarkers to enable early detection of diseases and to provide diagnostic, prognostic, and predictive information. According to the NIH, a biomarker is defined as “a characteristic used to measure and evaluate objectively normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention.” During the last decade, the maturation of high-throughput -omics technologies (i.e., genomics, transcriptomics, proteomics, peptidomics, and metabolomics) has set the pace for biomarker discovery to a point whereby it has evolved from being an observational byproduct of clinical practice, toward being a large-scale systematic process, often referred to as a “pipeline.” Interestingly, despite the intensified interest and investment by major stakeholders, including academia, industry, and government, the number of biomarkers receiving US Food and Drug Administration (FDA)4 clearance has declined substantially over the last 10 years to less than one protein biomarker per year (2, 3).

A simplistic version of a biomarker development pipeline could be divided into 4 main phases (Fig. 1). First, preclinical exploratory studies are performed to identify promising biomarkers. Usually during this discovery phase, small numbers of samples from diseased and nondiseased groups are compared to identify molecules exhibiting discriminating potential. When properly performed, high-throughput technologies enable the simultaneous unbiased assessment of thousands of molecules from which potential biomarkers can be identified. The candidate biomarkers are prioritized on the basis of various selection criteria that depend on the discovery platform, the type of biomarker being sought, and the available bioinformatics tools. Judging by the numerous publications reporting novel candidate biomarkers, the discovery phase seems to be productive, and numerous reviews have summarized a generalized strategy regarding this first phase of the pipeline (4–7).

The second step includes the development and validation of a robust assay to measure the analyte of interest in the intended clinical sample. In practice, assay development is an iterative process that occurs at every step in the pipeline and may not end even after an
assay is marketed. The assay is used to quantify the biomarker, usually in a retrospective fashion, using patient samples stored in tumor banks, which is the third step of the pipeline. In contrast to the discovery phase, retrospective validation studies require large numbers of samples to ensure statistical rigor, as well as samples that reflect the biological variability of the targeted population. Only a very limited number of analytes that continue to show discriminatory potential will make it to the next phase, which is evaluation in prospective trials. Although the pipeline is most easily conceptualized as 4 distinct sequential steps, each phase can involve multiple studies performed at various time points throughout the biomarker’s developmental lifecycle.

Biomarkers come in different forms, such as DNA, RNA, proteins, peptides, and metabolites. Given that proteins are the biological endpoints that control most biological processes and can be quantified efficiently, economically, and with high analytical sensitivity in the circulation, it is not surprising that proteins have gained the most attention as potential circulating biomarkers. With the discovery of soft ionization methods by Fenn and Tanaka (who shared the 2002 Nobel Prize in Chemistry), the analysis of complex proteomic mixtures using mass spectrometry became feasible. Since this discovery, thousands of studies have been performed focusing on deciphering the proteomes of various biological materials, including blood, other fluids, tissues, and cell lines in the quest to identify novel biomarkers. As mass spectrometry–based proteomics turned quantitative, the excitement regarding the potential to discover novel biomarkers was heightened (8). However, the number of biomarkers entering the clinic has not increased, suggesting that the development pipeline is not efficient. A major problem is that the commonly used approach based on unbiased high-

Fig. 1. The biomarker development pipeline.
The 4 main phases of biomarker development are depicted as described in the text. For a biomarker to move from one phase to another it needs to overcome multiple challenges at different levels. Only biomarkers that will reach the last step successfully will be implemented in the clinic. Although depicted as 4 sequential steps, the phases are not always distinct from each other and the pipeline is not always linear. Each phase can involve multiple studies performed at various time points throughout the biomarker’s developmental lifecycle.
throughput discovery to biomarker identification suffers from relatively high false-positive rates. The large number of false positives, in turn, is slowing identification and validation of true biomarkers. Regrettably, even biomarkers with good performance never enter the clinic, because of lack of rigorous validation or a scarcity of data showing a clear or superior contribution to existing clinical practices. The key to tackling both these issues is to clearly define the clinical question that the biomarker should address before any biomarker study is undertaken. If investigators do not know the clinical question to be answered, it is impossible to design an appropriate study to identify and validate a true biomarker.

The journey of a protein biomarker from the bench to the clinic is long and challenging, and every step must be meticulously planned and executed to succeed. This is evident not only from the numerous articles addressing possible causes and potential solutions for the paucity of novel biomarkers, but also from other documents such as the NIH Roadmap initiative and the US FDA’s Critical Path Initiative for drugs and diagnostics (9). Our goal in this review was to focus on the clinical uses of protein cancer biomarkers, providing examples and discussing their strengths and weaknesses. We mention barriers that a potential biomarker has to overcome to reach the clinic, suggest possible solutions to these barriers, and summarize the regulatory procedures required for a biomarker to obtain FDA approval. Finally, we follow the uphill path of a protein cancer biomarker from the bench to the bedside.

**Clinical Uses of Cancer Biomarkers**

Ideally, a cancer biomarker test would be a blood test that is positive only in patients with cancer and is correlated with disease stage, provides prognostic information, predicts response to treatment, and is easily and reproducibly performed. To date, only 18 protein cancer biomarkers have been cleared by the FDA (Table 1). Although these are in clinical use, they are far from ideal. Below, we describe the main uses of cancer biomarkers, provide examples, and highlight their main shortcomings.

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>Official gene namea</th>
<th>Clinical use</th>
<th>Cancer type</th>
<th>Source type</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-fetoprotein (AFP)</td>
<td>AFP</td>
<td>Staging</td>
<td>Nonseminomatous testicular</td>
<td>Serum</td>
</tr>
<tr>
<td>Human chorionic gonadotropin (hCG)</td>
<td>CGB</td>
<td>Staging</td>
<td>Testicular</td>
<td>Serum</td>
</tr>
<tr>
<td>Carbohydrate antigen 19:9 (CA19:9)</td>
<td>MUC16</td>
<td>Monitoring</td>
<td>Pancreatic</td>
<td>Serum</td>
</tr>
<tr>
<td>Carbohydrate antigen 125 (CA125)</td>
<td>MUC16</td>
<td>Monitoring</td>
<td>Ovarian</td>
<td>Serum</td>
</tr>
<tr>
<td>Carcinoembryonic antigen (CEA)</td>
<td>PSG2</td>
<td>Monitoring</td>
<td>Colorectal</td>
<td>Tissue</td>
</tr>
<tr>
<td>Epidermal growth factor receptor (EGFR)</td>
<td>EGFR</td>
<td>Prediction</td>
<td>Colorectal</td>
<td>Tissue</td>
</tr>
<tr>
<td>v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog (KIT)</td>
<td>KIT</td>
<td>Prediction</td>
<td>Gastrointestinal</td>
<td>Tissue</td>
</tr>
<tr>
<td>Thyroglobulin</td>
<td>TG</td>
<td>Monitoring</td>
<td>Thyroid</td>
<td>Serum</td>
</tr>
<tr>
<td>Prostate specific antigen (PSA)</td>
<td>KLK3</td>
<td>Screening and monitoring</td>
<td>Prostate</td>
<td>Serum</td>
</tr>
<tr>
<td>Carbohydrate antigen 15:3 (CA 15:3)</td>
<td>MUC1</td>
<td>Monitoring</td>
<td>Breast</td>
<td>Serum</td>
</tr>
<tr>
<td>Carbohydrate antigen 27:29 (CA27:29)</td>
<td>MUC1</td>
<td>Monitoring</td>
<td>Breast</td>
<td>Serum</td>
</tr>
<tr>
<td>Estrogen receptor (ER)</td>
<td>ESR1</td>
<td>Prognosis and prediction</td>
<td>Breast</td>
<td>Tissue</td>
</tr>
<tr>
<td>Progesterone receptor (PR)</td>
<td>PGR</td>
<td>Prognosis and prediction</td>
<td>Breast</td>
<td>Tissue</td>
</tr>
<tr>
<td>v-erb-b2 erythroblastic leukemia viral oncogene homolog 2 (HER2-neu)</td>
<td>ERBB2</td>
<td>Prognosis and prediction</td>
<td>Breast</td>
<td>Tissue</td>
</tr>
<tr>
<td>Nuclear matrix protein 22 (NMP-22)</td>
<td></td>
<td>Screening and monitoring</td>
<td>Bladder</td>
<td>Urine</td>
</tr>
<tr>
<td>Fibrin/fibrinogen degradation products (FDP)</td>
<td></td>
<td>Monitoring</td>
<td>Bladder</td>
<td>Urine</td>
</tr>
<tr>
<td>Bladder tumor antigen (BTA)</td>
<td></td>
<td>Monitoring</td>
<td>Bladder</td>
<td>Urine</td>
</tr>
<tr>
<td>High molecular CEA and mucin</td>
<td></td>
<td>Monitoring</td>
<td>Bladder</td>
<td>Urine</td>
</tr>
</tbody>
</table>

* Human genes: AFP, alpha-fetoprotein; CGB, chorionic gonadotropin, beta polypeptide; MUC16, mucin 16, cell surface associated; PSG2, pregnancy specific beta-1-glycoprotein 2; EGFR, epidermal growth factor receptor; KIT, Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog; TG, thyroglobulin; KLK3, kallikrein-related peptidase 3; MUC1, mucin 1, cell surface associated; ESR1, estrogen receptor 1; PGR, progesterone receptor; ERBB2, v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian).
“Early is the watchword for cancer control, early diagnosis, early treatment will save many lives” was the slogan on one of the posters from the Work Projects Administration Poster Collection (www.loc.gov/pictures/resource/cph.3b48900) designed to publicize health and safety programs in the mid-1930s. But it was not until 1971, when President Richard Nixon declared the war on cancer, that this concept was truly thrust into the public spotlight. In most cases, by the time cancer becomes symptomatic, the tumor has already spread and treatments become ineffective. The objective of a screening program is to detect cancer at a curable stage, before symptoms develop. For example, localized breast cancer has a 5-year survival rate of more than 90%, whereas in patients with metastasis the 5-year survival drops to 60% if regional and 30% if distant.

A useful screening biomarker must meet several criteria. It should be able to detect the disease at an early and asymptomatic stage and result in decreased morbidity or increased survival rates. Accurately determining the effect of early detection on survival rates can be difficult and must factor in lead-time bias, the perceived increase in survival rates that occurs solely by detecting the disease earlier. Moreover, a screening test must be highly specific to minimize false positives. In this context, the term specificity refers to the proportion of healthy people with a marker result below established cutoffs. Within a population of 1 million screened individuals a test with 99.9% specificity will still lead to 1000 false-positive results. Therefore, high specificity is necessary for a good screening biomarker given that even a small false-positive rate could trigger a large number of unnecessary diagnostic procedures with the associated psychological stress and cost. The optimal sensitivity and specificity requirements for any marker will depend on many factors and must take into account the consequences of producing either a false-positive or false-negative result. Ultimately, for a screening program to be considered successful, it must be cost-effective and noninvasive and lead to a measurable reduction of disease-specific morbidity and/or mortality. As a result, screening for early detection is best suited for diseases in which the prevalence is relatively high, medical care is accessible, and patients are willing to collaborate for further follow-up and treatment.

One of the most well-known and studied cancer screening biomarkers is prostate-specific antigen (PSA). In 1986 PSA was approved for the monitoring of prostate cancer, and this monitoring resulted in a considerable improvement in prostate cancer treatment. Eight years later, PSA was cleared by the FDA for screening of prostate cancer, based on studies showing that increased PSA concentrations in asymptomatic individuals were associated with increased risk of prostate cancer. Widespread implementation of the PSA blood test resulted in early detection of approximately 90% of prostate cancer cases and therefore contributed to an apparently longer survival period after diagnosis.

Despite its widespread use, PSA suffers from several major limitations. In addition to being found in patients with prostate cancer, increased concentrations of PSA can be found in individuals with benign conditions such as benign prostate enlargement and prostate inflammation and infection. Additionally, PSA values do not correlate well with tumor aggressiveness. High-grade prostate tumors display the greatest risk for spreading and lead to prostate cancer-related death, whereas low-grade tumors may remain localized and not pose a threat to the patient’s life; treatment of the latter type of tumors may be more harmful to patient’s quality of life than the actual disease. As a result, the contribution of PSA screening in decreasing disease-specific mortality is still being contested on the basis of controversial results from 2 prospective prostate cancer–screening trials, the European Randomized Study of Screening for Prostate Cancer and the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. Data from the European Randomized Study of Screening for Prostate Cancer trial showed that PSA screening resulted in a 20% decrease in the prostate cancer–specific mortality rate, whereas the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial demonstrated no significant reduction in disease-specific mortality. Both trials showed that PSA screening led to overdiagnosis and overtreatment of prostate cancer, which were accompanied by potentially harmful results, including unnecessary biopsies, side effects from treatment, and increased psychological stress. Therefore, even PSA, the most widely known cancer biomarker, is far from ideal. Thus a major unmet clinical need remains: the ability to screen for prostate cancer by using a method that has greater diagnostic specificity than PSA for prostate cancer and can discriminate between indolent and aggressive (potentially life-threatening) disease.

A diagnostic test, contrary to screening, would be prescribed to an individual who has presented with symptoms. However, the characteristics of an ideal diagnostic biomarker are similar to the characteristics for screening. Notably, there is no protein biomarker recommended in practice guidelines for cancer diagnosis, but many of the well-known markers are widely used as diagnostic aids.
PREDICTION AND PROGNOSIS
A prognostic factor is a patient or disease characteristic at the time of diagnosis that provides information about the natural history of the disease, independent of therapy. On the other hand, a predictive biomarker predicts response to a therapeutic intervention compared to the effect of a different therapy or no treatment. Predictive biomarkers form the foundations for personalized medicine.

A prognostic test is used to classify patients into different risk groups. If the prognostic value of certain biomarkers is high, these markers may eventually be incorporated into the tumor-node-metastasis (TNM) staging system, as is the case for 3 serum biomarkers, α fetoprotein, human chorionic gonadotropin, and lactate dehydrogenase in testicular cancer, for which there is a TNMS system in place with a site-specific prognostic factor (S is for site-specific prognostic factors) (15). Prognostic biomarkers are important because patients with a good prognosis could be spared from receiving adjuvant therapy, thus avoiding side effects and reducing treatment cost. Because most prognostic biomarkers lack the required diagnostic accuracy, clinicians prefer overtreatment to undertreatment.

Evaluating prognostic and predictive biomarkers, especially in a prospective manner, is challenging given that endpoints of interest such as disease-free and overall survival take years to reach and often require complicated statistical analyses (16).

The most widely used prognostic and predictive biomarker is the estrogen receptor (ER) tissue marker in breast cancer. For prognosis, patients with ER-positive tumors have better outcomes than those with ER-negative tumors. However, the prognostic value of ER is time dependent (5 years or earlier after diagnosis), and the prognostic impact of ER in lymph-node-negative patients is limited (17, 18). The need for more accurate prognostic breast cancer biomarkers has led to intensified research and resulted in a series of multigene classifiers, proposed as potentially useful adjuncts for breast cancer patient management. However, the clinical value of these classifiers remains to be established (19).

The ER was the first predictive biomarker recommended for routine use in breast cancer by the Tumor Marker Panel of the American Society of Clinical Oncology (20). It is used for selecting patients likely to respond to hormonal therapy. Endocrine treatment, such as selective ER modulators and aromatase inhibitors, are the most effective in patients with hormone receptor–positive tumors in early or advanced disease (21). However, the predictive efficacy of ER is far from ideal, because approximately one third of advanced ER-positive patients are intrinsically resistant to endocrine therapies, and the majority of ER-positive tumors that initially respond to endocrine therapy will eventually develop resistance (22).

MONITORING
Following therapy, cancer patients are monitored to ensure that they remain disease free or are treated promptly upon relapse. Additionally, monitoring biomarker concentrations could be indicative of a therapeutic response, with increasing concentrations associated with resistance and consideration for alternative therapies. A monitoring test needs to be both diagnostically sensitive and specific to ensure continuation of beneficial therapies and early discontinuation/replacement of ineffective therapies.

A typical example of a monitoring biomarker is the carbohydrate antigen 19-9 (CA19-9). CA19-9 was identified in 1979 as a useful marker for pancreatic and colorectal cancer (23) and in 2002 it was cleared by the FDA for monitoring pancreatic cancer patients. CA19-9 has relatively good diagnostic sensitivity (70%–90% in advanced disease, approximately 50% in early disease, and absent in approximately 5% of the general population) but poor diagnostic specificity (increased in numerous other cancer types as well as in benign pancreatic diseases and hepatobiliary inflammation) (24–27). Therefore, CA19-9 is recommended to assess response to therapy in patients with increased CA19-9 concentrations before treatment but is not recommended as a screening marker or as the sole method to identify recurrence of pancreatic cancer (28).

The Need for New Cancer Biomarkers

There are 18 FDA-cleared protein cancer biomarkers (Table 1), with several others being used clinically without FDA clearance. There is a clear need to identify additional biomarkers for optimal patient management. Numerous cancer researchers and organizations have strived to identify novel biomarkers that can fulfill these unmet clinical needs. So why are there so few new biomarkers entering the clinic, and can anything be done about it?

Challenges to Biomarker Development

Below, we outline some challenges of the biomarker-development pipeline after the discovery phase and suggest some ways to overcome these challenges. These are also summarized in Table 2. We also describe the path of a prognostic biomarker, which is an example that has yet to be cleared by the FDA despite over 20 years of testing and a wealth of supportive data.
A clinically useful biomarker should be measured reliably, so ease of adoption for use by routine clinical laboratories is key (29). Immunoassays such as ELISA remain the gold standard for validation and clinical use (30, 31). Mass spectrometry–based approaches for clinical applications remain an appealing prospect for numerous reasons including, but not limited to, high analytical specificity and sensitivity and multiplexing capabilities. However, assay complexity and expertise requirements are currently preventing the adoption of mass spectrometry into routine use in clinical laboratories (30).

Before assay development, numerous preanalytical variables must be assessed, such as choice of a biological matrix and sample collection, handling, and storage. Depending on the availability of critical reagents (antibodies, calibrators), assay development can take from weeks to years. In the case of immunoassays, the lack of high-quality antibodies has slowed the rate of development. To abbreviate that barrier, the Human Protein Atlas project took the initiative of developing multiple antibodies against all human proteins (32). As of November 2011, 15 598 antibodies corresponding to 12 238 protein-coding genes that cover >40% of the 19 559 human entries as defined by UniProt have been generated (33). The generated prototype assay needs careful validation. Basic analytical characteristics to be examined include assay dose–response curve, measuring range, limits of detection and quantification, accuracy, imprecision, and analytical specificity (34). Adhering to federal regulations that govern human diagnostic testing, known as the Clinical Laboratory Improvement Amendments, during analytical assay validation could increase the reliability and quality of the data (35).

The importance of assay technical quality was underpinned in a recent study in which tissue samples collected in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial were used to evaluate potential ovarian cancer biomarkers (36). In this study prediagnostic samples were used to evaluate in a phase III setting the performance of 28 ovarian cancer biomarkers that have shown promise in phase II studies. Although the main conclusion was that CA125 outperformed all other biomarkers, one finding was that assay performance correlates strongly with biomarker performance. None of the markers for which the assays had

---

**Table 2. Reasons for biomarker failures.**

<table>
<thead>
<tr>
<th>Reason</th>
<th>Solutions</th>
<th>Frequency</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fraud</td>
<td></td>
<td>Low</td>
<td>Potti et al. (79)</td>
</tr>
<tr>
<td>Preanalytical factors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Patient selection bias</td>
<td>Clearly define the clinical question to be addressed</td>
<td>High</td>
<td>Xu et al. (80)</td>
</tr>
<tr>
<td>• Sample collection, handling and storage</td>
<td>Use samples collected under detailed SOPs</td>
<td></td>
<td>Villanueva et al. (81)</td>
</tr>
<tr>
<td></td>
<td>Use well annotated samples</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analytical factors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Methodological artefacts</td>
<td>Validate the analytical method</td>
<td>High</td>
<td>Leman et al. (82)</td>
</tr>
<tr>
<td>• Poor analytical method</td>
<td>Use appropriate quality controls with all analyses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistics/Bioinformatics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Inappropriate statistical analysis</td>
<td>Seek and follow the expertise of an experienced biostatistician</td>
<td>High</td>
<td>Mor et al. (83)</td>
</tr>
<tr>
<td>• Data overfitting</td>
<td></td>
<td></td>
<td>Petricoin et al. (84)</td>
</tr>
<tr>
<td>• Small sample size</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Multiple hypothesis testing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Overlapping training and validation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• patient cohorts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical validation</td>
<td></td>
<td>Medium</td>
<td>Esrig et al. (85)</td>
</tr>
<tr>
<td>• Nonreproducible validation</td>
<td>Clearly define the clinical question to be addressed prior to undertaking any study</td>
<td></td>
<td>Malats et al. (86)</td>
</tr>
<tr>
<td>• Poor study design</td>
<td>Collaborate with an experienced biostatistician</td>
<td></td>
<td>Kim et al. (87)</td>
</tr>
<tr>
<td>• No adequate clinical performance</td>
<td>Use appropriate specimens to avoid bias</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Use validated analytical methods</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commercialization</td>
<td></td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>• Intellectual property</td>
<td>Apply for patents to obtain intellectual property rights as early as possible</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FDA approval</td>
<td></td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Seek FDA guidance early in development phases</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CVs >30% had adequate diagnostic sensitivity in phase II or phase III studies.

SAMPLE REQUIREMENTS
Identifying the appropriate target population to address the clinical question to be answered and ensuring sample integrity from collection to analysis are essential steps toward biomarker development. The population used in the validation, including both cases and controls, should be selected to closely match the target population, and the selection process must have clearly defined inclusion/exclusion criteria. Sample size requirements must be calculated to ensure adequate statistical power for each study. Sample sizes usually increase as the biomarker moves toward the clinic. Patient-related factors, including fasting, posture, circadian rhythms, age, and sex should be documented for every sample and carefully investigated to explore their relation to the analyte of interest. Any factor found to affect analyte concentrations must be thoroughly controlled in all future studies. Additionally, biomarker stability should be studied under different conditions to identify the most efficient protocol for collection, handling, and storage. Finally, extra caution should be given to devices and reagents used during sample collection and storage. Collection and processing of all samples should be performed using predefined standard operating procedures to minimize preanalytical biases. Using samples that are poorly annotated or from an inappropriate population will introduce bias into studies and lead to false results.

Access to desired biological resources has been one of the limiting factors during validation studies, and the need for biological samples far surpasses the available supply. High-quality samples with associated clinical data are sparse and “should be reserved for the most promising research studies or late-stage biomarker discovery efforts” (38). Toward this direction, there are efforts to create a central registry that will include information on novel biomarkers undergoing validation studies, especially those that are using samples from randomized trials, so that the most promising candidates can be identified for further study (39). In addition, efforts are also underway to generate high-quality sample collections for validating promising biomarkers. A large NIH initiative, the Early Detection Research Network, promotes collaboration among academic and industrial researchers and organizations, both nationally and internationally, for the development and testing of promising biomarkers or technologies for early detection of cancer (http://edrn.nci.nih.gov).

STATISTICS
Robust statistical analysis of validation data of potential biomarkers is essential. Possible biases, including small sample size, inappropriate controls, nonindependent training and validation cohorts, and multiple hypothesis testing should be critically examined. Furthermore, statistical significance, frequently represented by P values, is not adequate to assess clinical utility and can even be misleading. Researchers should closely collaborate with biostatisticians that are experienced in the biomarker field to help plan and perform their study.

With the advent of -omics technologies such as RNA microarrays and the development of multigene signatures, statistical analysis of validation studies become more complex. Initially, the statistical methods used for analyzing complex outputs relied on tests developed and optimized for single-signal output. However, undertaken efforts resulted in newer and more appropriate methodologies, based on multisignal readings (40). For example, the output of multigene and multiprotein diagnostic tools, such as Oncotype Dx (41) and OVA1 (42), is a numeric score calculated from expression values of multiple genes or proteins, respectively. Although the algorithm used to calculate the diagnostic score is a vital component, it remains a “black box” for clinicians. To address these types of tests the FDA created a new diagnostic category entitled “in vitro diagnostic multivariate index assay” (43).

FUNDING
As a biomarker moves from one phase of development to another, the financial burden increases. Soon after discovery and early validation studies, usually taking place in an academic setting, industry may enter the scene and provide financial support for the remaining development phases. Being able to clearly demonstrate that a biomarker addresses an unmet clinical need by carefully planning and executing early validation studies is critical to acquiring an industry partner. A potential limitation at that stage may be issues related to intellectual property. A key element of biomarker commercialization is ownership, and industrial partners will not undertake the development of a marker that does not have potential for investment return, regardless of its clinical utility. Therefore investigators should apply for intellectual property rights as early as possible in the biomarker discovery process.

One of the reasons to seek FDA clearance of a biomarker is to obtain reimbursement by Medicare or private insurers, allowing for broader use of the test. On the other hand, inadequate reimbursement rates may discourage industrial partners from seeking FDA approval, opting for development of an in-house test which may receive less utilization but with higher re-
imbursement rates. The importance of inadequate reimbursement is highlighted by the existence of 2 government-commissioned reports recommending the reevaluation of reimbursement rates for diagnostics (44, 45). Because of their high financial risks, diagnostics have been regarded by industry as a less attractive investment opportunity than therapeutics, and consequently the development of diagnostic methods is not as well funded by industry as the development of therapeutic drugs (46). This reluctance to invest in diagnostics poses a challenge for biomarker development and further highlights the importance of carefully defining the clinical question to be addressed by the biomarker before initiating any biomarker discovery efforts.

REGULATORY ISSUES
Clinical validation of novel potential cancer biomarkers includes assessing their diagnostic utility in studies with different populations (37). After potential clinical usefulness is demonstrated, a company may pursue development of a product (in vitro diagnostic device) and conduct further validation studies to evaluate the test technically and clinically. When sufficient data are available on clinical utility for a specified clinical use, approval from regulatory agencies such as the FDA may be sought. Although obtaining regulatory approval is not necessary to market or obtain reimbursement for diagnostic tests in the US, obtaining such approval can increase the testing market and allow the marker to gain wider acceptance.

The FDA has established the Voluntary Exploratory Data Submission, which encourages sponsors to share their data with the agency, even at early stages of the development, without being considered as part of the regulatory decision-making process. The agency leans toward establishing a collaborative relationship with the sponsors rather than serving strictly as a regulatory body.

More details regarding regulatory processes for biomarker development in the US can be found on the FDA website (www.fda.gov) as well as in recent reviews (3, 47). Every country has its own regulatory body that governs the requirements for biomarker use within its own medical system, and approval by one country’s regulatory body does not translate into approval by another.

The Example of Urokinase-Type Plasminogen Activator and Its Inhibitor Plasminogen Activator Inhibitor Type 1

We will use a biomarker with proven clinical utility, but not widespread use, to outline some major implementation difficulties with cancer biomarkers. As mentioned earlier, one of the unmet clinical needs in breast cancer is the development of more diagnostically sensitive and specific prognostic markers (48). Current breast cancer prognosticators include lymph node status; number of involved nodes; tumor size, grade, and hormone receptor status; and HER2 (human epidermal growth factor receptor 2) amplification (49, 50). However, these factors can provide unequivocal prognostic information (favorable or poor) for only 30% of breast cancer patients (10).

The serine protease urokinase-type plasminogen activator (uPA) and its inhibitor, plasminogen activator inhibitor type 1 (PAI-1), is one example of potential prognostic breast cancer biomarkers. The concentrations of uPA and PAI-1 antigen correlate with recurrence risk; patients with node-negative breast cancer and low concentrations of uPA and PAI-1 can be classified as a low-risk subgroup and thus be spared systemic adjuvant therapy (51).

The prognostic potential of uPA and PAI-1 is not a recent finding. In 1988 Duffy et al. published a preliminary report demonstrating the prognostic potential of the enzymatic activity of serine protease uPA in patients with primary breast cancer (52). In less than a year, Janicke et al. reported that the tumor tissue antigen concentration of uPA (apart from its activity) also has prognostic relevance in breast cancer (53). In the following 3 years, PAI-1 also emerged as an additional prognostic breast cancer biomarker in both node-negative and node-positive breast cancer patients (54). Subsequently, between the early 1990s and mid-2000s, 15 individual studies (55–67) covering a variety of demographic areas supported the prognostic impact of uPA and PAI-1 in primary breast cancer.

The ELISAs used for determining antigen concentrations of uPA and PAI-1 were assessed by the Receptor and Biomarker Group of the European Organization for Research and Treatment of Cancer, and obtained international quality assurance (68, 69). Additionally, the optimum method for tissue preparation before analysis was determined, further standardizing the measurement of these markers (70).

uPA and PAI-1 are the first tumor markers to reach the highest level of evidence validation for their clinical utility in breast cancer management. The highest level of evidence, as defined by Hayes et al., can be achieved through “evidence from a single, high-powered, prospective, controlled study that is specifically designed to test the marker, or evidence from metaanalysis and/or overview of level II or III studies” (71). In the case of uPA/PAI-1, a metaanalysis of 18 datasets encompassing more than 8000 primary breast cancer patients with a median follow-up of 6.5 years was performed by the Receptor and Biomarker Group of the European Organization for Research and Treat-
The first detection of PSA in serum was reported as a potential ovarian cancer biomarker in the late 1990s. In 2010, PSA took approximately 6 years from monitoring recurrence and progression of ovarian cancer, and another 8 years for its clearance for screening. The intensified research for developing novel biomarkers is quite diverse. The clinical uses of cancer biomarkers are quite diverse. The intensified research for developing novel biomarkers stems from the fact that markers currently in clinical practice suffer from key limitations. Despite the high expectations regarding the availability of information from the completed human genome sequencing project and the advent of promising technological advances, the number of novel protein cancer biomarkers obtaining FDA clearance has decreased during the last decade. Many see this as a failure of the biomarker and -omics field to deliver results. However, we need to consider that biomarker development encompasses multiple contiguous phases, requires collaboration among different stakeholders, carries a major financial burden and faces many potential challenges. Even when all of these challenges are overcome, the history of clinically useful biomarkers suggests that at least a decade is required for the transition of a marker from bench to the bedside. Therefore, it may be too early to expect that the new technological advances will catalyze the anticipated biomarker revolution any time soon.

**Conclusion**

The clinical uses of cancer biomarkers are quite diverse. The intensified research for developing novel biomarkers stems from the fact that markers currently in clinical practice suffer from key limitations. Despite the high expectations regarding the availability of information from the completed human genome sequencing project and the advent of promising technological advances, the number of novel protein cancer biomarkers obtaining FDA clearance has decreased during the last decade. Many see this as a failure of the biomarker and -omics field to deliver results. However, we need to consider that biomarker development encompasses multiple contiguous phases, requires collaboration among different stakeholders, carries a major financial burden and faces many potential challenges. Even when all of these challenges are overcome, the history of clinically useful biomarkers suggests that at least a decade is required for the transition of a marker from bench to the bedside. Therefore, it may be too early to expect that the new technological advances will catalyze the anticipated biomarker revolution any time soon.

**References**


**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 or Potential Conflicts of Interest:** Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:

- **Employment or Leadership:** E.P. Diamandis, Clinical Chemistry, AACC.
- **Consultant or Advisory Role:** None declared.
- **Stock Ownership:** None declared.
- **Honoraria:** None declared.
- **Research Funding:** None declared.
- **Expert Testimony:** None declared.
- **Patents:** None declared.

**Role of Sponsor:** The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.


