

# An Inflection Point in Cancer Protein Biomarkers: What was and What's Next

## Authors

Anna D. Barker, Mario M. Alba, Parag Mallick, David B. Agus, and Jerry S. H. Lee

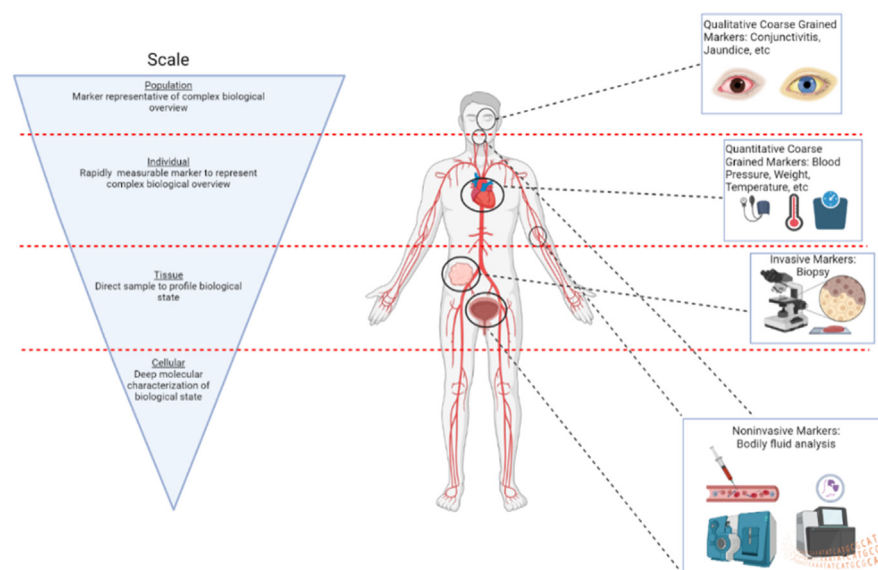
## Correspondence

[ABarker@eitm.org](mailto:ABarker@eitm.org)

## Graphical Abstract

### In Brief

Biomarkers are the highest value proposition in cancer medicine today. Despite progress in molecular sciences and evolving regulatory frameworks for biomarkers, the field has been more about promise than improving human health. Achieving precision medicine necessitates moving away from reductionist thinking to studying complex diseases as complex adaptive systems and defining biomarkers as representations of biological system states. This approach will enable redefining cancer as a disease of dysregulated communication, which could ultimately prove to be game changing for the field of biomarkers and the patients who will benefit.



### Highlights

- Protein biomarkers will be key to deconvoluting the complexity of cancer.
- Redefine biomarkers as representations of hierarchical biological system states.
- Big data and AI could be transformative for biomarker discovery in the future.
- Biomarkers discovery will require theoretical constructs to separate signal from noise.
- Build mechanistic frameworks to enable new data analytics in context of prior studies.



# An Inflection Point in Cancer Protein Biomarkers: What was and What's Next

Anna D. Barker<sup>1,2,\*</sup>, Mario M. Alba<sup>3</sup> , Parag Mallick<sup>4,5</sup> , David B. Agus<sup>1,6,7</sup> , and Jerry S. H. Lee<sup>1,6,7</sup> 

**Biomarkers remain the highest value proposition in cancer medicine today—especially protein biomarkers. Despite decades of evolving regulatory frameworks to facilitate the review of emerging technologies, biomarkers have been mostly about promise with very little to show for improvements in human health. Cancer is an emergent property of a complex system, and deconvoluting the integrative and dynamic nature of the overall system through biomarkers is a daunting proposition. The last 2 decades have seen an explosion of multiomics profiling and a range of advanced technologies for precision medicine, including the emergence of liquid biopsy, exciting advances in single-cell analysis, artificial intelligence (machine and deep learning) for data analysis, and many other advanced technologies that promise to transform biomarker discovery. Combining multiple omics modalities to acquire a more comprehensive landscape of the disease state, we are increasingly developing biomarkers to support therapy selection and patient monitoring. Furthering precision medicine, especially in oncology, necessitates moving away from the lens of reductionist thinking toward viewing and understanding that complex diseases are, in fact, complex adaptive systems. As such, we believe it is necessary to redefine biomarkers as representations of biological system states at different hierarchical levels of biological order. This definition could include traditional molecular, histologic, radiographic, or physiological characteristics, as well as emerging classes of digital markers and complex algorithms. To succeed in the future, we must move past purely observational individual studies and instead start building a mechanistic framework to enable integrative analysis of new studies within the context of prior studies. Identifying information in complex systems and applying theoretical constructs, such as information theory, to study cancer as a disease of dysregulated communication could prove to be “game changing” for the clinical outcome of cancer patients.**

Much has been written over the last two-plus decades about the inadequacies of both biomarker discovery and the distressing lack of the successful development and clinical adoption of cancer biomarkers (1–5). Literally, thousands of biomarkers are in the published literature, and if reports of biomarker discoveries mattered in terms of patient benefit, we should have countless biomarkers employed clinically today (3, 5). However, despite this “wealth” of reported discoveries, the number of Food and Drug Administration (FDA)-approved cancer biomarkers has changed little for several decades (6). Moreover, the problems leading to the dismal success of protein biomarkers has been reviewed, discussed, and rereviewed to the point that many of the sources of failure are well known for the biomarkers that do enter development (1, 2, 4, 5, 7–9). The authors that have studied, reported, and opined on why biomarkers (proteins and beyond) fail in end-to-end development (discovery to FDA approval and clinical application) cite various reasons for these dismal results ranging from the obvious example of poor quality samples to underpowering studies and bias to more complex regulatory questions of context of use and the fit for purpose of biomarker development that aims to become FDA-approved, clinically relevant assays (8–10).

However, nowhere in this dizzying array of studies, research reports, perspectives, or commentaries on biomarker failure is denial of their critical role in fundamentally understanding and ultimately controlling cancer. Biomarkers are crucial to achieving precision oncology.

Biomarkers are key to predicting cancer risk (screening), diagnosing the presence or absence of cancer, supporting disease prognosis, predicting therapy selection, subtyping patients for clinical trials, and monitoring both the safety and therapeutic signals in patients. In addition, biomarkers are key to developing pharmacodynamic biomarkers that can serve as endpoints for clinical trials and potentially become surrogate

From the <sup>1</sup>Lawrence J. Ellison Institute for Transformative Medicine, Los Angeles, California, USA; <sup>2</sup>Complex Adaptive Systems Initiative and School of Life Sciences, Arizona State University, Tempe, Arizona, USA; <sup>3</sup>Department of Pharmacology and Pharmaceutical Sciences, USC Alfred E. Mann School of Pharmacy, University of Southern California, Los Angeles, California, USA; <sup>4</sup>Canary Center at Stanford for Cancer Early Detection, and <sup>5</sup>Department of Radiology, Stanford University, Stanford, California, USA; <sup>6</sup>Keck School of Medicine, and <sup>7</sup>Viterbi School of Engineering, University of Southern California, Los Angeles, California, USA

\*For correspondence: Anna D. Barker, [ABarker@eitm.org](mailto:ABarker@eitm.org).

endpoints. All of these biomarker “classes” require assays that employ specific R&D strategies for establishing the metrics to determine the sensitivity (measure of positives) and specificity (measure of negatives) for specific context of uses (11). Positive and negative predictive values represent true positives and true negatives (12). These performance measures require careful consideration in the development of all biomarker assays and are especially relevant in cancer screening where even a seemingly small decrement in sensitivity or specificity can have a major impact on test outcomes and their ultimate value.

Interestingly, protein biomarkers remain the main target for biomarker discovery and development. In fact, most investigators agree that protein cancer biomarkers are key to realizing the promise of the decades of research that is currently driving the emerging concept of precision medicine. The Human Genome Project and subsequent large-scale projects such as The Cancer Genome Atlas, International Cancer Genome Consortium, Clinical Proteomics Tumor Analysis Consortium (CPTAC), and normal and tumor “atlases” at nearly all biologic scales now makes it possible to create a molecular profile of an individual patient’s cancer to predict therapy selection in the context of precision oncology (13–15). The combination of these “omics” profiles with clinical and real-world data is creating an unprecedented “big data” revolution that offers opportunities to apply machine and deep learning and other artificial intelligence (AI) approaches to potentially transform the discovery and development of new generations of cancer biomarkers. It is important to note that the initial generation, quality assessment, and interpretation of molecular data sets from these large-scale initiatives required iterative inputs from teams of disease experts, pathologists, technology platform specialists, and data scientists to separate true biological insights from misleading batch effects and other artifacts.

Given the history of biomarker discovery, it is possible that the impact of AI and other advanced technologies may not all be positive in improving the science of biomarkers but may further exacerbate problems that have historically plagued the field. Technologies such as liquid biopsy, digital pathology, advanced imaging, and broader availability of affordable genome sequencing may create unmanageable amounts of data of questionable quality that will add significant additional complexity to biomarker discovery and development (16–19). For example, liquid biopsy has evolved in a short period of time to gain the confidence of the clinical oncology world, setting the stage for collecting unprecedented quantities of patient data at unprecedented speeds. Further complicating this picture, liquid biopsy will also drive the emerging field of multicancer early detection (MCED) assays, which will provide complex data sets with limited existing knowledge to support meaningful analysis (<https://prevention.cancer.gov/news-and-events/blog/blood-samples-could-help>) (20–22).

Despite the dismal success rate of protein biomarkers, given their effector roles in nearly every aspect of biology, they remain the primary choice for the discovery and development of biomarkers of outcome and theragnosis in oncology. To achieve success in the discovery and development of protein biomarkers, the field must engage the interrogation of dysregulated cancer biology with new thinking that views and studies cancer as a complex adaptive system (CAS) (23, 24). The convergence of AI, “multiomics” profiling, liquid biopsies, MCED assays, imaging, and advanced technologies that capture complex spatial information from patients portend different paths for the discovery and development of protein biomarkers (alone or in combination with other interrogative technologies). For example, how will information-driven systems of proteins *versus* an expressed protein from a relatively simple alteration in a gene, or even a few genes, change what we know about the disease to inform and improve both biomarkers and prediction for therapeutic interventions.

In the following paragraphs, we will briefly trace some of the history of biomarker discovery and development as a backdrop for what we may expect in the future. It has taken decades to reach a point where we can reasonably expect to collect and integrate high quality data from multiple biological scales to identify and develop biomarkers that could be transformative. We are at an inflection point. It now requires rethinking what biomarkers are—and are not—and recognizing that biomarkers as a technology class are tools that must capture the complexity of the biology of cancer to achieve their promise. In that regard, we must move beyond single genomic alterations and/or small sets of altered genes in cancer to embrace systems proteomics, which may in the final analysis produce “coarse-grained” biomarkers that more accurately represent measurable changes reflected in disease phenotypes using a combination of invasive, minimally invasive, and noninvasive platforms (25–28) (Fig. 1). Future progress in the field will depend in large measure on understanding the origin, multiscale dynamics, and management of information (29–32). Given the need for application of mathematical constructs and information theories to make sense of “big data,” this may well represent the biggest challenge we have faced to date in cancer research.

### THE DISMAL HISTORY OF PROTEIN BIOMARKERS AND SOME HOPEFUL SIGNS

The field of biomarker science and the development paths for biomarkers discovery and development have struggled for years to determine what biomarkers are (and are not) and if protein biomarkers even existed at all. Some of this confusion comes from attempts to define what the term biomarker means, especially in terms of regulatory review. Currently most studies employ the definition of a biomarker as defined “officially” in 2001 by a task force of the National Institutes of Health and the FDA. This often-referenced definition, “A characteristic that is objectively measured and evaluated as

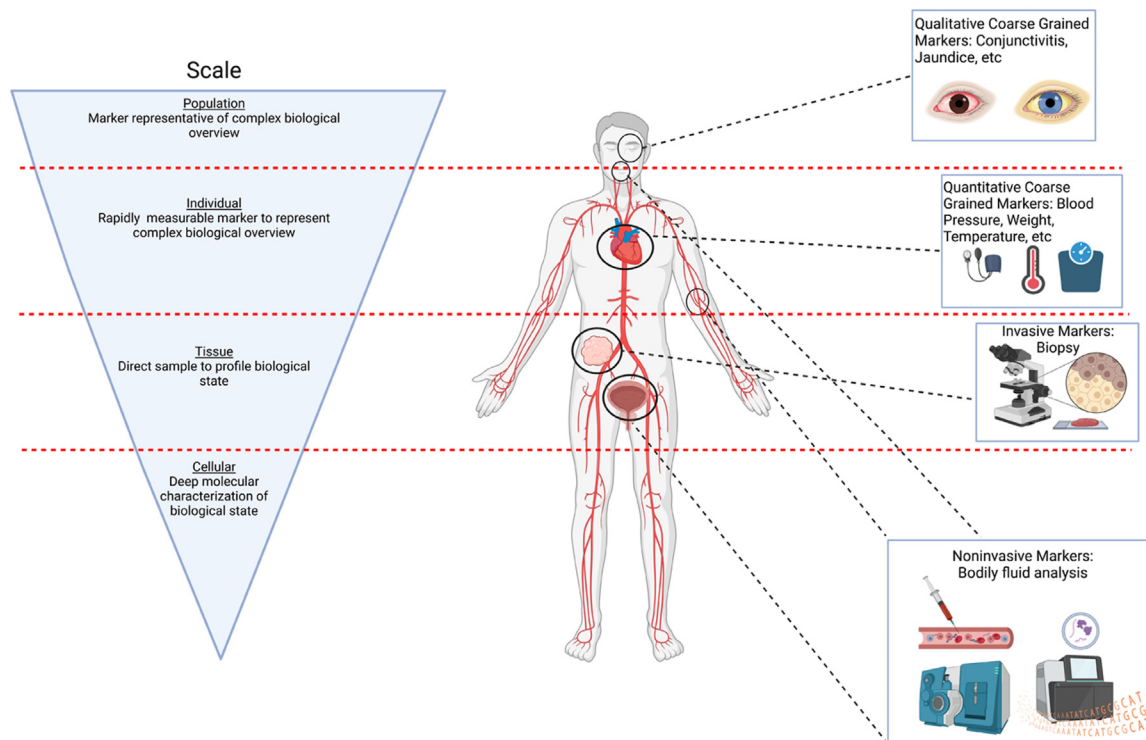


FIGURE 1. **An overview of biomarkers.** Cancer is an emergent property of a complex adaptive system (CAS) that requires we move beyond the single analyte–disease correlation model and instead look at biomarkers as multiscaled. Complex (sometimes termed “Coarse grain”) biomarkers will be required to capture sufficient multiscale information that represents an overall system state and allow us to build the mechanistic framework to describe the “diagnostic window” transitional state between healthy and diseased. This new generation of biomarkers can be simple representations of vastly more complex biological systems, that is, jaundice and the diagnosis of liver failure. (Created with [BioRender.com](#)).

an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention” (33) provides a reasonable construct for identifying a biomarker. Related, and as important, is the definition of a surrogate endpoint (33). Perhaps the best and most accepted definition of a surrogate endpoint was proposed in 1995 by Dr Robert Temple of the FDA, “a surrogate endpoint in a clinical trial is ‘a laboratory measurement’ or a physical sign used as substitute for a clinically meaningful endpoint that measures directly how a patient feels, functions, or survives.” (34) Interestingly, as biomarker discovery and development continue to advance, it is likely that these definitions will also need to evolve (35).

Biomarkers are likely the highest value proposition in biomedicine today, especially protein biomarkers for cancer, but frankly even with the rise of multiomics and wider use of companion diagnostics for therapeutics and significant improvements in mass spectrometry and the protein technologies, the discovery and development of biomarkers remains somewhat daunting. This should not be a surprise as the bar for clinically useful biomarkers is extremely high, especially for cancer biomarkers. Depending on the use envisioned for a cancer biomarker to be used in patients, some or all the following may apply. The authors propose that a biomarker

must be (1) integrative to reflect the complexity of cancer; (2) represent the dynamics of cancer; (3) deconvolute mixed subtypes of cancer; (4) ubiquitous; (5) simple to interpret; (6) sample friendly in terms of patients; (7) be derived from high quality fit for purpose samples; (8) amenable to develop using simple technologies; (9) able to seamlessly integrate into clinical practice, hospitals, and clinical care workflows; and (10) be cost effective. The bar is even higher for a biomarker surrogate endpoint that purports to be a direct measure of how a patient feels, functions, or survives (34). These criteria point to the fact that the discovery and development of a clinically useful (successful) cancer biomarker is a formidable challenge that cost in excess of 100 million dollars end-to-end (36).

A number of authors have reviewed in detail the difficulties that have plagued the field of protein biomarkers (3–5, 27, 36), and virtually all of these explorations make clear that the failure of protein biomarker begins in biomarker discovery and cascades through the entire biomarker development cycle. As shown in Table 1, biomarkers (particularly protein cancer biomarkers) have proven especially hard to discover, and false discovery has resulted in less than a single protein biomarker being approved by the FDA since the early 1990s (27). Clearly, the failure of protein biomarkers in discovery is inexorably linked to our lack of understanding of the biology of

TABLE 1  
Why protein biomarker fail in discovery and what has changed

Failure source	What goes wrong	Changes that mitigate the problem
An interesting but irrelevant clinical question	Biomarker does not address the clinical question or is not sufficiently different from an existing clinical biomarker (5).	“Multiomics,” MRM mass spectrometry, new generation proteomic technologies imaging, AI, and engagement of multidisciplinary teams are vastly improving opportunities to understand root causes of disrupted molecular pathways; and with the addition of real-world patient data, biomarker discovery will become more successful (27, 47, 48, 67, 70, 125–127).
Experimental design	Design is underpowered, does not recognize or control for bias anticipated analytics needs, or specifically address the question (10, 39, 128, 129).	Improved statistical models, Bayesian designs, and AI are proving new approaches to design robust proteomics biomarkers that minimize bias and other confounders (130).
Samples	Poor quality (often convenience) samples of insufficient numbers to power the study (8, 131).	The last decade has steadily improved the sources and quality of human samples—including approaches that maintain the fidelity of the biology of the sample, including preanalytics ( <a href="https://biospecimens.cancer.gov/bestpractices/index.asp">https://biospecimens.cancer.gov/bestpractices/index.asp</a> ) (132, 133).
Technology standards	Technology standards in areas such as mass spectrometry, sequencing technologies, etc. are not used to ensure quality and reproducibility (131).	Increased sophistication of protein measurement technologies, especially mass spectrometry, increasingly requires wider spread of use of standards at all levels of sample interrogation (27, 45).
Data quality and metadata	Poor samples equal to data that does not assess the biology targeted in the hypothesis; lack of metadata (especially clinical data) renders the biomarker useless (134, 135).	Data quality remains a major challenge, but the eras of “big data” is requiring investigators to consider data relative to sample quality and ensure appropriately consented metadata (27, 45, 66).
Flawed analytics	Biomarker studies require increasingly sophisticated analytics—well beyond “p values” and many researchers are not trained in ML and AI overall (136, 137).	The analytics space remains a work in progress. Data quantity, quality, and velocity from proteomics technologies require new thinking and approaches in the design phase of biomarker discovery inclusive of analytics that are AI based (138).
Reproducibility	All of the above and more render many biomarker studies irreproducible—significant source of overall failure (4, 131, 139).	Reproducibility of results across all of oncology, especially biomarker, research was (and remains) a major source of biomarker failure. FDA has emphasized the critical nature of sensitivity, specificity, and reproducibility of biomarkers in all of their guidance documents and reviews—which is helping (45, 128) ( <a href="https://www.gpo.gov/fdsys/pkg/BILLS-114hr34enr/pdf/BILLS-114hr34enr.pdf">https://www.gpo.gov/fdsys/pkg/BILLS-114hr34enr/pdf/BILLS-114hr34enr.pdf</a> ).

expressed proteins that derive from dysregulated genes that exist in cells which are themselves highly complex-regulated environments.

Deconvoluting the underlying complexity of cancer through biomarkers is a mandate for all aspects of discovery studies that aim to identify any type of cancer biomarkers, especially proteins. Unfortunately, as pointed out in numerous studies, just the opposite is true, with a plethora of reports of false discovery. Table 1 summarizes many of the reasons for the failure of biomarkers including beginning with an irrelevant clinical question; flawed experimental design; convenience samples of poor quality and insufficient numbers; lack of technology standards (even if they exist); poor data quality, especially lack of metadata; inadequate or just wrong analytics; and lack of attention to the reproducibility of results.

Table 1 also illustrates that some of the root causes for biomarker failure are being addressed, but the creation of robust protein cancer biomarker pipelines will require

solutions for these barriers and some that are less obvious. For example, addressing preanalytical variability of patient samples is a critical, but often overlooked, source of failure for protein biomarkers, and fortunately, standards are in various stages of development (4, 37–39). For example, application of the National Cancer Institute’s (NCI) best practices is designed to minimize preanalytical variability (<https://biospecimens.cancer.gov/bestpractices/index.asp>), and the goals of consortia such as BloodPAC aim to improve the reproducibility and value of blood-based cancer assays through the development of standards and frameworks (40). All of these efforts are reenforced by the need for biomarker studies to be clinically relevant in response to the FDA’s requirements that workflows warrant standard operating procedures ensure “fit for purpose” (41) with NCI’s biospecimen evidence-based practices serving as a potential guide. Despite these efforts there are, as of this writing, no national standards to control the impact of preanalytics on biomarker

discovery and derivative development pipelines. The reason for this lack of uniform standards is in the spirit of siloed science, as there is an abundance of standards. Unfortunately, to date, the focus has been more on standards development than harmonization—and the result is limited progress in the lack of reproducibility of biomarker research.

The “big data” revolution is proving to be both an asset and a liability for biomarker discovery and development. Siloed science drives data protectionism and sequestered data sets that are of insufficient power to drive quality biomarker studies. In addition, to establish predictive clinical cancer biomarkers will require agreement on common data elements, data harmonization standards and significant increase in open access to longitudinal and paired clinical outcomes data (5, 42, 43). Recent federal initiatives purport to address some of the problems associated with data availability, harmonization, and longitudinal data access through mandate that data from federally funded grants (e.g., NCI) be publicly available (<https://www.whitehouse.gov/ostp/news-updates/2022/08/31/what-they-are-saying-white-house-federally-funded-research-guidance-hailed-as-a-win-for-innovation-and-equity/>) (44, 45).

Looking into the future where cancer is viewed as a complex adaptive (and evolving) system *versus* a collection of “parts,” it will be necessary to develop systems level biomarkers. It could be argued that the best systems level biomarkers we have for cancer today are stage and grade, which offer “snapshots” and something of a surrogate for the extent of tumor development over time inclusive of evolution and the resulting heterogeneity that defines most cancers. Although these may be systems level representations of cancer, they tell you almost nothing about the cancer that is actionable. Cancer is an emergent property of a complex system, and deconvoluting the integrative and dynamic nature of the overall system through biomarkers is a daunting proposition. That said, perhaps these biomarkers will be more reflective of measures of system states at hierarchical levels more reflective of the phenotype. Such biomarkers will require multidisciplinary teams (cancer biologists and clinicians, mathematicians, physicists, and engineers) to iteratively optimize and monitor multiscale systems dynamics (46–48). Perhaps in the future, there will be more integrative, coarse-grained biomarkers such as a patient’s performance score, which may well offer a measure of the integrative and dynamics aspects of the overall system.

TECHNOLOGY ADVANCES PROMISE BIOMARKER SUCCESS—BUT TO DATE A “Slow SLOG”

As noted in the prior section, biomarker discovery and subsequent development has maintained an uncertain and often detached, relationship with advances in technologies—often delaying and sometimes slowing progress. For example, the success of imatinib for chronic myeloid leukemia (CML) is the best example of how advances in technology informed the

development of a highly clinically important biomarker discovery. The Philadelphia chromosome discovery was first reported by David Hungerford and Peter Nowell, based on application of new cytogenetics techniques that revealed the alterations in chromosome numbers in seven CML patients (49). A decade later, Janet Rowley’s use of quinacrine fluorescence karyotyping nine CML patients revealed the origin of the Philadelphia chromosome and paved the way for a new methodology to characterize chromosomal alterations across cancers (50). Both discoveries, at the time, were dismissed and mired by problems that continue to plague biomarker studies today: insufficient patient samples to determine whether the results were artifacts of new technologies or potential new clinical insights. Now we know these initial discoveries paved the way to the identification of the *BCR-ABL* fusion gene as the principal pathogenetic driver of CML and approval of the first successful tyrosine kinase inhibitor targeting the function of a fusion gene, imatinib, by the FDA in 2001 based on the dramatic results of pivotal clinical trials (51).

Imatinib initiated an era of biomarker discovery focused on finding “more” *BCR-ABL*-like targets and spurred 2 decades of evolution of molecular profiling technologies. We are now in an era of multiomics, with advances in genomics, transcriptomics, and epigenomics; are protein biomarkers still considered the “holy grail of biomarkers?” Although genomics and transcriptomics are often touted as comprehensive for creating patient molecular profiles, the fact is that these modifications do not necessarily manifest in the proteome (36). The proteome remains perhaps the most impacted biological measure of disease on a broadscale that reflects the state of the underlying biology (3). Protein posttranslational modifications alter signaling networks which create opportunities for discovery of new therapeutic targets (52). Although all aspects of the digital information that derives from the genomic alterations play some role in cancer, it is the systems of proteins (systems proteomics) that affect functional changes that provide a path to fundamentally understand and control cancer.

*PSA—A Learning Experience and Cautionary Tale*

The imatinib experience taught us that achieving an increased understanding of a disease at the molecular level does not necessarily lead to the discovery of clinically important protein biomarkers. Although protein cancer biomarkers (and cancer biomarkers overall) have historically suffered from significant failure, we have gained valuable lessons that cancer is not a singular event but rather a dynamic state change that evolves across space and time. Clearly, the early and somewhat naïve assumption that single protein biomarkers would capture cancer’s complexity led to the unfortunate reality that approximately 1% of published biomarkers achieved clinical utilization (4). For example, first approved by FDA in 1986, prostate-specific antigen (PSA), a single protein, was one of the few early biomarkers that advanced to clinical use, and it has taught us a great deal

about the challenges associated with its use in both the screening and diagnostics settings for a heterogeneous disease like prostate cancer.

The expansive deployment of PSA tests in the United States in the early 1990s led to significant overdiagnosis and unnecessary invasive and often harmful therapeutic interventions (53, 54). It was estimated during this time that only 25 to 30% of “positive” cases would ultimately develop cancer, and by 2009, PSA screening represented over \$750 million a year in unnecessary follow-up testing (36). Another learning experience from PSA testing was the realization that signals from other biological processes, such as inflammation and trauma (55), could produce positive PSA results and an abundance of false positives. It also became clear that a PSA test offered little, if any, guidance on the optimal point for disease intervention, and it became clear that deployment of a screening test like PSA in the population at large requires exquisitely validated sensitivity and specificity (56–58). Validation across all affected populations and the addition of familial history, multiomics profiles, and clinical data have continued to improve the value of PSA (59).

Ironically, although PSA does not capture the complexity of prostate cancer, it has recently demonstrated the value of an imperfect screening test for early detection of this disease. As early as 2014, and again in 2018, the U.S. Preventive Services Task Force recommended that men of 55 to 69 years should decide if they wanted to have PSA screening, or not, and further recommended that men of 70 years of age or older should not receive PSA screening (<https://www.uspreventiveservicestaskforce.org/uspstf/index.php/recommendation/prostate-cancer-screening>). In subsequent years, prostate cancer increased 3% per year from 2014, and the proportion of advanced prostate cancer has more than doubled over the past 10 years (<https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/2023-cancer-facts-figures.html>). This alarming increase in late stage prostate cancer prompted the American Cancer Society to make evidence-based recommendations in 2023 for screening that combines age, risk, and PSA results (<https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/2023-cancer-facts-figures.html>). Although PSA is not a perfect test and research is required to determine if PSA screening has differential effects in affected populations, it is a cautionary tale. PSA tests apparently measure an important disease signal, and its deployment as a screening assay plays a significant role in identifying disease at a point where intervention is warranted. The tradeoffs in terms of slowing or preventing advanced disease speak to the value of identifying early disease signals, even in the setting of accepting a percentage of false positive results. In addition to the potential role of PSA in identifying key disease signals, one of the major outcomes from the PSA experience is the extent to which efforts to improve the assay have enabled significant

advances in understanding the complex biology of prostate cancer. Imperfect tests, especially PSA, have made clear that cancer biomarkers will require a high degree of dimensionality to reflect the heterogeneity and complexity of most cancers and the need to understand how biomarkers can be leveraged to understand state changes in cancer. These findings and improvements in protein biomarker discovery, development, plus the availability of new FDA regulatory pathways have benefitted the field of biomarker science, especially the promise of protein biomarkers.

In recent years, multiple advances in technologies, like the *BCR-ABL* story, are supporting and accelerating the successful discovery and development of the protein biomarkers of the future. Sequencing technologies and multiomics approaches are evolving rapidly to enable sequencing workflows, reduce costs, and enable progress in precision medicine, especially oncology. Fortunately, the story of protein biomarker discovery and development is not all doom and gloom. There are several examples that demonstrate the power of multiomics technologies and protein-based biomarkers to support more effective discovery, development, and utilization of clinically important cancer biomarkers.

Launched in 2010, the I-SPY 2 consortium is an excellent example of a learning system that utilizes a suite of biomarkers to pair patients with the most rational therapeutic intervention for their disease. The I-SPY 2 consortium has continuously enrolled patients on a multiarmed, master protocol to evaluate new neoadjuvant therapies in combination with standard-of-care chemotherapy for breast cancer (60). Utilizing Bayesian adaptive randomization and clinical biomarkers to assign patients to specific breast cancer biomarker-based subtypes, a number of therapies have successfully “graduated” to move on to small phase 3 trials (61). Moreover, the multiple arms of the trial and biomarker-based subtype treatments support advancing successful therapies to patients and in parallel provide the power needed for biomarker regulatory approval. This innovative approach to trial design is reducing the costs of clinical trials, and as an enduring “platform” trial, it is streamlining the organization and set-up of further clinical trials (62). Master protocols have been employed to perform adaptive platform trials beyond breast cancer including the Beat AML trial (63) and the GBM AGILE Trial (64) for glioblastoma, and the model continues to be adopted across other diseases (<https://clinicaltrials.gov/ct2/show/NCT04488081>). As observed in I-SPY 2, these trials offer exciting opportunities to identify protein biomarkers that can serve to assign treatment based on molecular biomarkers, subtype disease, and advance precision oncology.

Whole exome sequencing of specimens in lung cancer cohorts have identified numerous genomic alterations, with somatic mutations as high as 8 to 10 mutation per mega base, reflecting the vast genetic heterogeneity in lung cancer (43). Next generation sequencing (NGS) mutations can be traced to single base-pair resolution levels, without sacrificing accuracy

or precision, and multiple cancer types are beginning to leverage these technologies for better multiomics resolution and associated biomarker development. In fact, genomics screening has quickly become a standard of care in non-small cell lung cancer as the National Comprehensive Cancer Network's 2023 principle of molecular and biomarker analysis guidelines have recommended nine genomic screening biomarkers along with PD-L1 expression for therapeutic decision making in non-small cell lung cancer (65). Multiomics changes also differentiate tumors from normal tissues, paving the way for improved targeted therapeutic intervention to extend patient survival (66). The NCI lung-MAP trial and the UK national lung matrix trial are just two examples of trials that are leveraging NGS to identify multiple gene panels as biomarkers for lung cancer.

As sequencing methods continue to advance in depth of coverage, so does the ability to potentially develop high sensitivity and specificity protein biomarkers to detect early-stage tumors. While multiomics markers are providing much needed support for targeted therapies, especially in precision oncology, markers for immunotherapy remain a major challenge. Complex (black box) biomarkers such as tumor mutational burden (43) are of limited use in predicting patient response and monitoring patients that do respond. Patient molecular profiles based on classic genomics without consideration of epigenomics are insufficient to support optimal decision making by physicians. Epigenetic changes are common across all human cancers, as posttranslational and histone modifications play significant roles in altering the expression of various oncogenes and tumor suppressors (67). It has been shown that mRNA levels tend to correlate poorly with protein abundance, but by leveraging genomic level data, prediction of protein expression will improve (67–69). Although transcriptomics has allowed us to have a better understanding of the genetically driven RNA transcript alterations, genomic, epigenetic, and transcriptomic alterations will manifest in ways that will impact protein expression, making proteomics the strongest candidate for clinically relevant biomarkers. In recognition of these findings, proteogenomics as an emerging field has increased significantly in recent years as the science has evolved. By combining multiple omics modalities to acquire a more comprehensive landscape of the disease state, we are increasingly developing biomarkers to support therapy selection and patient monitoring.

#### *Proteogenomics*

Proteogenomics represents an application of multiomics to biomarker discovery and development. Proteogenomics, the combination of genomics and proteomics, has emerged rapidly over the past decade to define genomics-driven proteomic workflows. The NCI CPTAC flagship programs have been largely successful in developing assays and potential markers across multiple cancers, leveraging the proteogenomic approach (38, 70–72). As of 2022, over 3500

assays for 3224 unique peptides and 1742 unique proteins have been validated by the CPTAC initiative to develop reproducible targeted mass spectrometry-based assays for biomarkers in cancer (73–75). Advances in NGS will allow for ever-better identification of proteins to create pipelines of interest to both CPTAC and the industry. Selected companies maintain that next generation biomarker research should encompass genomic sequencing, transcriptomics, neoantigen prediction, immune profiling, and single-cell sequencing (<https://www.science.org/content/resource/genomic-biomarkers-multiomics-next-generation-sequencing-cancer-research>). Proteogenomics continues to advance, with recent CPTAC reports demonstrating the ability of these pipelines to generate biomarkers of patient responses to chemotherapy and therapeutics (71). Progress in proteogenomic-based approaches are powerful and promising technologies that look to improve and advance the discovery and development of protein biomarkers to support precision oncology in the future (76).

The emergence of proteogenomics highlights the need for integration across the research discovery, development, and delivery continuum. The focus on individual investigator studies has historically focused on single genes and/or proteins and resulted in what has become “siloed science.” Nowhere is that truer than in oncology where thousands of “biomarkers” are reported in the literature (77) and few ever even make it into development. To achieve progress in proteogenomics, and deconvolute the complexity of cancer, it will require that silos come down and the “omics” research communities unite through seamless partnerships that will result in high value biomarkers that will move the field forward (<https://www.cancer.gov/news-events/cancer-currents-blog/2021/cancer-moonshot-midpoint-progress-update>). As discussed later in this perspective, these partnerships are essential to fundamentally understand that cancer is an emergent multiscale complex system driven by dysregulated communication where each scale has its own integrative structure and emergent properties.

#### *Liquid Biopsy*

Proteogenomics is not alone in what portends an improved future for protein biomarkers, and one of these advances, liquid biopsy, may prove to be transformative for all aspects of cancer diagnosis, treatment, and even prevention. Liquid biopsy for cancer generally refers to the analysis of circulating tumor DNA but may also include isolation and analysis of other tumor-derived material from bodily fluid matrices. Isolating these fractions is challenging, as the abundance of circulating tumor DNA to cell-free DNA in circulation can range from 0.1% to 90%, and although imperfect, it is generally reflective of the tumor burden in a given patient (20). Despite several challenges, the method has expanded exponentially recently as a popular orthogonal technique to analyze patients that is less invasive compared to traditional needle biopsy.

### *Early Multiple Cancer Detection Assays*

The rapid evolution and expansion of liquid biopsy methodologies is best personified in what is a new class of MCEDs. As of 2022, there are more than 27 companies developing MCED assays. Many MCED assays claim to detect over 50 cancers using CpG methylation analysis of cell-free DNA coupled to NGS or other methods. This assay is in a trial to evaluate results in 140,000 patients to validate its predictive value. Whether or not these assays will make a difference in cancer survival is yet to be determined, but the field of MCED is without question a highly promising development for biomarker research (78).

Although overall liquid biopsy continues to develop, and similar to biomarkers, reproducibility in this setting illustrates the critical need for larger data to power these technologies. For example, it has been posited that liquid biopsy-based MCEDs may be the next PSA, leaving little doubt that a great deal of more research is required for these assays to ever become screening tools for early detection of cancer in the population (<https://prevention.cancer.gov/news-and-events/blog/blood-samples-could-help>). Liquid biopsy in general, and MCEDs, in particular, offer hope that achieving scale and application of stringent quality control metrics in sampling will significantly improve biomarkers for all phases of the patient's journey—well beyond detecting early cancer “signals” (21, 22). Reflecting the promise of the MCEDs as diagnostics, federal agencies are attempting to develop infrastructure to determine if identifying potential cancers very early also increases survival (<https://prevention.cancer.gov/news-and-events/blog/blood-samples-could-help>). In hopes of answering this question, the NCI has expanded trials to evaluate 24,000 patients over a span of 4 years (<https://prevention.cancer.gov/major-programs/multi-cancer-detection-mcd-research>). One thing is clear, for these assays to be effective as screening tools in the general population will require near perfect sensitivity and specificity.

### *Single-Cell Analysis*

The science of biomarkers is also benefiting from single-cell analysis. For example, the NCI Cancer Moonshot is establishing a clinical, experimental, computational, and organizational framework to create publicly available tumor atlases that includes single-cell analysis (44, 45). In addition, single-cell analysis is enabling deeper profiling of the microenvironment, an increasingly important source of potential cancer biomarkers, as is mass spectrometry imaging that profiles single-cell populations within the tumor immune microenvironment. This approach allows investigators to identify cellular subtypes by combining both tandem mass tags with antibodies for a targeted proteomic approach that can be further multiplexed to provide biomarkers that characterize the overall cancer immune profile (79–81). Targeting single-cell interactions promises to transform therapeutic intervention as cellular subtypes

associated with resistance to hormone ablation therapy in prostate cancer and in melanoma patients have been identified using these technologies (82, 83). Although there are myriad challenges (e.g., reproducibility and scaling) in single-cell proteomic analysis, these technologies offer much needed opportunities to molecularly characterize the tumor immune microenvironment and increase the effectiveness of these therapies in cancer patients (84). Single-cell analysis continues to grow, and new guidelines have also been established by the community at large to help continue its maturity (85).

As briefly reviewed here, the last 2 decades have seen an explosion of multiomics profiling and a range of advanced technologies designed to support precision medicine. This review does not include advanced imaging technologies, nanotechnology devices, synthetic biomarkers, and several other advances that enable probing the proteome. The emergence of liquid biopsy, continued development of single-cell analysis, and these other technologies could have the potential to transform biomarker discovery. However, in all of these domains, it will ultimately be critical to understand how information is encoded into defined pathways and networks, and to map the perturbations in those networks in the context of a given disease, and importantly around a clinical decision. Clearly translating multiomics profiling into an understanding of biological signaling networks as they relate to disease highlights the challenge of understanding the overlap of these networks in complex systems. Diseases are often more than the sum of their parts, and while a singular focus on omics domains may yield a lot of data, it will not be sufficient to capture the information needed to define a disease phenotype or drive improved clinical decision making. Advancing precision medicine necessitates moving away from the lens of reductionist thinking toward the understanding that complex diseases are in fact CASs. In addition, the field needs to adopt communication models that maximizes information and reduces noise at all levels in a CAS (Fig. 2). Theoretical constructs such as Shannon information theory is one approach that may be able to decode the information needed for biomarkers at various biologic scales.

### DISCUSSION: EMBRACING COMPLEXITY, REALIZING THE DIAGNOSTIC PROMISE OF PROTEIN BIOMARKERS

#### *Embracing Complexity*

Clearly, for the biomarkers of the future, especially protein biomarkers, to be more predictive requires that they increasingly capture the complexity of cancer information and associated signals. In that regard, it will become necessary to redefine biomarkers as representations of a biological system state at different hierarchical levels of biological order (Fig. 1). This definition would include traditional molecular, histologic, radiographic, or physiological characteristics, as well as emerging classes of digital markers and complex algorithms. In the case of protein biomarkers, it will be necessary to move away from single or small numbers of proteins to effectively

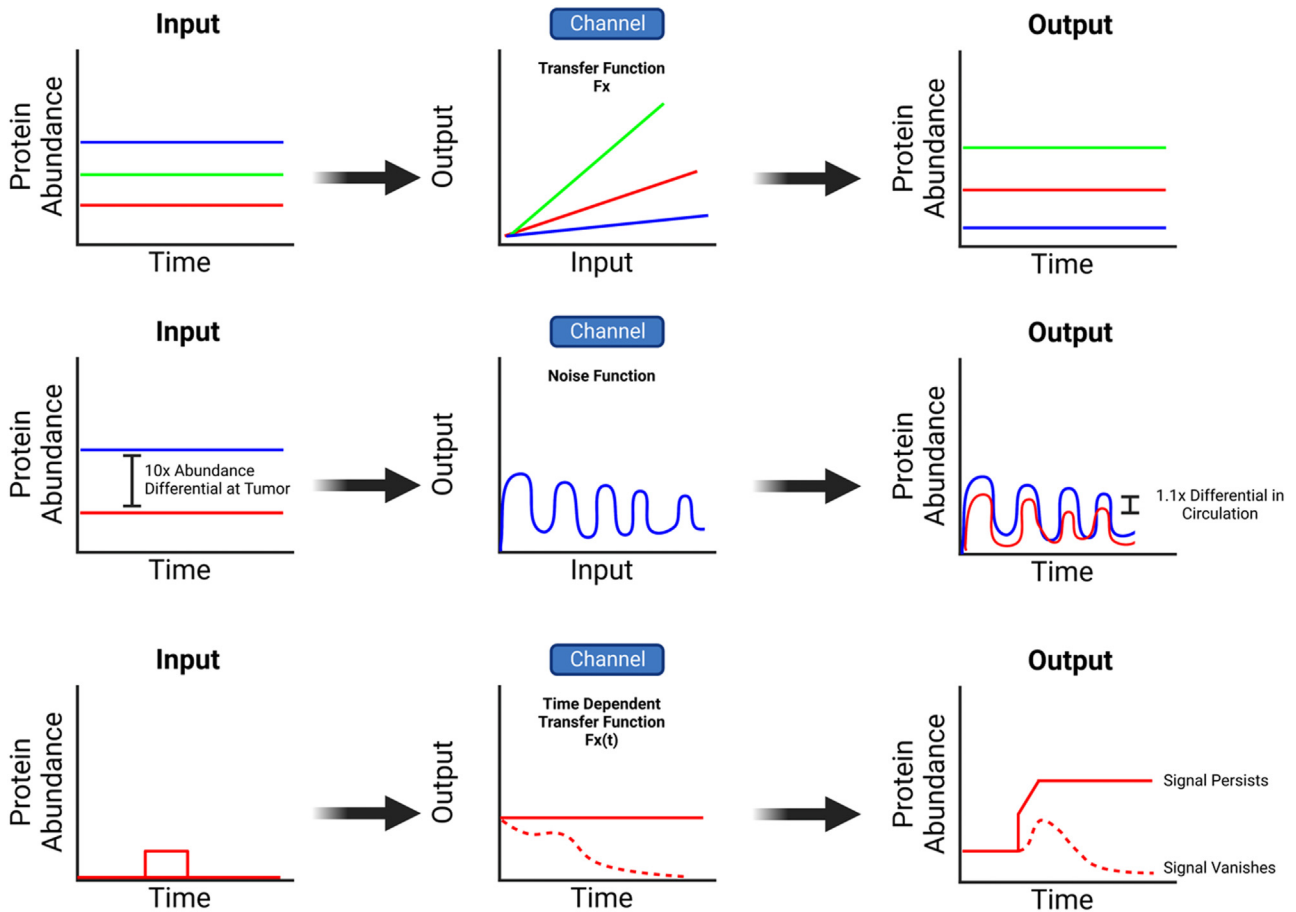


FIGURE 2. **Information transfer.** Understanding where and how biological information is encoded and transferred is essential to collect the right data to support accurate measurements. Uni-dimensional collection of a singular parameter is susceptible to bias and can lead to misrepresentation of the biological state in question (Created with [BioRender.com](https://BioRender.com)).

capture the information in protein networks and systems. Over 90% of the human proteome has now been validated at high stringency to better enable the understanding of the proteome networks, thanks in large part to efforts from the Human Proteome Organization (86, 87). Systems proteomics will support the discovery and development of biomarkers that can capture complex systems information throughout the health-to-disease continuum of patients and thus enable the development of efficacious treatments and high value diagnostics and theragnostics. In this regard, using biomarkers to understand information architectures and information processing in complex biological systems will be of key importance for making precision oncology a reality.

*Reframing Cancer Diagnostics as an Information Transfer and Systems Management Problem*

As we look forward the next decade of cancer biomarker-related studies, one of our greatest opportunities to succeed will come from reframing the challenge of biomarker discovery and accordingly the approaches taken for data analysis and interpretation. To date, much of the focus of biomarker studies

has been through the lens of univariate binary classification. This framing essentially states the biomarker challenge as, “find a protein whose abundance differentiates state X (*i.e.*, healthy) from state Y (*i.e.*, diseased).” In the case of diagnostics for infectious disease, such a phrasing is reasonable. There is a clear mechanistic link, and appropriate reductionism, between the presence of a pathogen in the body and the presence of a pathogen’s shed material in locations that can be easily sampled. Furthermore, the wide molecular gulf between a pathogen and its host makes the differentiation of signals from the pathogen and background signals straightforward to interpret. Because there is such a significant difference between host and pathogen, it can often be straightforward to look at one or a few pathogen-derived biomarkers as differentiators in a relatively time-invariant manner. In therapeutics, the distinction between disease and healthy state is often referred to as the “therapeutic window” and enables thinking about how straightforward it is to therapeutically target a given disease. We adopt the term “diagnostic window” to describe the magnitude of difference between disease states and healthy states.

Unfortunately, when finding diagnostics for diseases like cancer, the binary classification framing of biomarker studies is unlikely to succeed. The simplest diagnostic windows, such as those in many infectious diseases, are ones in which a marker or set of markers are absent in one state and present in another. More complex diagnostic windows might operate with different effective set points, such as low in one state and high in another (such as with serum troponin before and after cardiac arrest (88)). However, as tumors are extremely like their host, particularly in their earliest stages, the signals from those cancers are less differentiated and thus the diagnostic window is likely incredibly small as within person, and across person, biological variance can be large and state-to-state difference can be small. Furthermore, cancers are multiscale, complex adaptive, and dynamic systems, so the signals they are producing are constantly changing, further exacerbating the challenge of finding a marker whose single-timepoint measurement is radically different between states (Fig. 2). Given these two factors, it becomes necessary to reframe the biomarker-based diagnostics challenge away from univariate binary classification and toward a more descriptive information-rich framework. These innovative approaches are already influencing the field of proteomics, with the Human Proteome Grand Project representing an excellent example of an increased understanding that proteomics functions as a network of networks (<https://hupo.org/page-1757429>).

We cast the challenge of describing the state and trajectory of a tumor as an information transfer and management problem as illustrated by the following example. Remote sensing dates back to mid-19th century when aerial photographs were taken from hot air balloons or by camera-bearing carrier pigeons. The term “remote sensing” was first used in the United States in the 1950s by the Office of Naval Research (89) and is now commonly used to describe the science of identifying, observing, and measuring an object without coming into direct contact with it. One can deconstruct such problems as containing two pieces: the first is the signal(s) coming from the monitored object(s) and the second is the communication of that signal from the source through a communication channel to a receiving device (Fig. 2). A communications framework (90) gives us both a language and set of techniques for interrogating our biological problem of biomarker discovery. It is quite natural to recast the early detection and therapeutic response prediction/monitoring problems as information transfer and management problems. Essentially, for cancer early detection and therapeutic response characterization, we would like to discern the presence and trajectory of a tumor without requiring direct measurement of that tumor. The “source” in our model would be the tumor or the host response to that tumor. The “signal” would be the protein(s) from that tumor, and the “channel” would be the poorly characterized and extremely noisy process of shedding by which proteins travel from the tumor into the circulation and persist therein (Fig. 2). Through this lens, proteins are “transmitted” from the tumor to the circulation as governed

by a series of information transfer functions. Alternately, signals influenced by the tumor, but not generated by the tumor, such as the immune response or changes in the microbiome could be considered in this framework. The sensor at the receiver may be one of a wide variety of techniques ranging from ELISA to broadscale proteomic profiling.

An additional benefit of a communications framework is that it enables the use of time-dependent, multiscale models describing signal origins and signal transfer. Despite thousands of case-control profiling studies attempting to discover biomarkers, the number of studies asking simple, fundamental questions about biomarker origins is extremely small (91–97). The existence of biomarkers is predicated on the assumption that processes occurring at the tumor are reflected at a distance, such as in the circulation, urine, or in breath exudate. Proteomics-based biomarker discovery approaches have asserted that many potential markers are tumor-derived proteins that may be detected in the circulation, *via* shedding (secreted, cleaved from the cell surface, or released from the cell by either lysis or exocytosis) from the tumor, and then transited into blood (77, 98). Consequently, the composition and dynamics of the tumor proteome may be reflected in the circulating proteome (99–103). Based upon the assumption of a relationship between the tumor proteome and the circulating proteome, logical arguments are presented regarding the need to give higher priority to abundant tumor proteins (94, 104–106). As shown in Figure 2, it is possible to examine these assumptions. In particular, we must examine how molecular changes propagate to the cell scale and are impacted by tissue and host scale effects. For example, what if the proteins produced by the tumor change over the course of its progression, such as through evolution of the cellular heterogeneity or changes in the microenvironment? Likewise, what if there is a nonmonotonic relationship between the abundance of a protein in the tumor and in the blood. Though attractive, the monotonicity assumption may not be universally true; there may be a wide range of functions relating a protein’s tumor levels to its circulating levels. This is especially true given the likely wide range of degradation and clearance rates (107–110); for reference, decades of research shows that intracellular protein turnover rates (111) vary from minutes, weeks, or even months (112–115). Consequently, it is possible that proteins with large case *versus* control differences in abundance at the tumor site may be either unobservable or uninformative when interrogated in the circulation. Furthermore, as both the patient and the cancer are constantly changing, it is highly likely that we will need diagnostics that anticipate time-dependent variation, such as those that arise through circadian rhythm or from changes in weight or as induced by hormonal therapy (116–118). To succeed going forward, we must get past purely observational studies and instead start building a communications framework in which to better interpret the vast array of observational data. The biomarkers must also look for the dynamic changes of the tumor, as many studies have shown static cells with tumor-

associated molecular changes that do not cause harm to the host (119, 120).

Building the communication frameworks needed to drive progress in the discovery of biomarkers that represent complex systems states will benefit from several current opportunities including the integrative analysis of new studies within the context of prior studies. For example, there are nearly 2000 publications citing TIMP1 as a putative biomarker in diseases as wide range as pancreatic cancer and rheumatoid arthritis. Unfortunately, with new studies being done in isolation, it is hard to contextualize any potential biomarker. Resources like MarkerDB and the GVK 13,000 biomarker database could potentially help in the future (121, 122). There is also an opportunity to use multidata-type analysis, such as the integration of transcriptome and proteome data, especially given that at the cell scale, there are key linkages between molecular scales (e.g., transcriptome to proteome) that can give a synergistic boost to data interpretation. To recognize this boost, it is necessary to treat the data, not simply as additional dimensions but instead as linked multivariate entities. Learning these relationships within any singular dataset is currently an intractable problem. However, appropriate use of prior information and prior mechanistic models can enable more grounded, integrative analysis. Finally, the field will benefit greatly from the development of controlled vocabularies and repositories for both biomarker-associated data and experimental metadata and results of these studies.

#### LOOKING FORWARD: A PORTENTOUS FUTURE

Dating to the mid 1990s, viewing cancer as a CAS is not a new concept (23). Albeit biologic CAS is increasingly defined in greater detail; simply stated, they consist of large numbers of interacting components with simple rules of operation that exhibit selforganization and information processing. CAS is also characterized by internal information flows and dynamic input from external information flows. These properties lead to subsystems (complex interacting networks) that exhibit properties that cannot be predicted based on understanding the individual components. Complex systems adapt and evolve giving rise to collective emergent properties, which cannot be predicted from studying the parts of the system. In fact, cancer is perhaps best described as an emergent property of a CAS.

The future of cancer biomarker discovery, especially protein biomarkers, will depend in large measure on identifying and understanding the dysregulated information architecture, transfer, and management processes of cancer in a CAS context. Information in biological systems is stored everywhere and flows everywhere in time scales that are difficult for most current technologies to capture, but coarse-grained elements of these systems can be measured and used as surrogates. Most of the problems encountered in biomarker discovery to date reflect this lack of understanding of what information is, where it is, and how information is related to data. Some of the

current AI approaches in development today promise to reduce the dimensionality of the data from a dysregulated system like cancer, but the translation to information will require identifying the signal based on a defined biological context. However, capturing the complexity and trajectory of the cancer CAS through systems of protein biomarkers will require the application of novel communication frameworks including theoretical constructs such as information theory (24, 123) that provide a basis for understanding how information is encoded and transmitted, which offers an opportunity to reduce uncertainty in the predictions based on converting data to information (90, 124). One such theoretical construct, mathematically based Shannon information theory (90), is especially promising for its applicability to biological information. The theory enables defining information based on context (including temporal and spatial information) and distinguishing signal (information) from noise (entropy) at escalating levels of complexity in cancer. Evolution suggests that signaling networks in biology are optimized to match information sources, which will open new areas of investigation based on dysregulated signaling in cancer (24).

Achieving this transition will not be easy. However, it is increasingly a scientific imperative that the identification of clinically relevant biomarkers (especially proteins) be developed for a cancer patient's journey from diagnosis to survivorship. Ideally, the future will embrace the discovery, development, and delivery of complex information-based biomarkers gleaned through theoretical constructs that capture cancer as a disease of multiscale communication in the context of a CAS.

*Author contributions*—A. D. B. conceptualization; A. D. B., M. M. A., P. M., and J. S. H. L. writing—original draft; A. D. B., M. M. A., P. M., D. B. A., and J. S. H. L. writing—review and editing; D. B. A. supervision.

*Conflict of interest*—A. D. B. declares the following financial relationships: Ellison Institute for Transformative Medicine, Caris Life Sciences, Pancreatic Cancer Action Network, and Sage Bionetworks. A. D. B. declares the following non-compensated relationships: NCI National Cancer Advisory Board, American Association for Cancer Research, Arizona State University, Milken Institute, Quantum Leap Healthcare Collaborative, and Friends of Cancer Research. P. M. is a paid faculty member of Stanford University, Paid Founder and Chief Scientist of Nautilus Biotechnology, a compensated member of the Scientific Advisory Boards of ChromaCode, OncoImmune, and January AI, and an uncompensated President of The Dream Finder Foundation. J. S. H. L. is the paid Chief Science and Innovation Officer of EITM, paid faculty member of USC, advisory board member of AtlasXomics, Inc, and paid consultant of the Henry M. Jackson Foundation outside the submitted work. The other authors declare no competing interests.

*Abbreviations*—The abbreviations used are: AI, artificial intelligence; CAS, complex adaptive system; CML, chronic myeloid leukemia; CPTAC, Clinical Proteomics Tumor

Analysis Consortium; FDA, Food and Drug Administration; MCED, multiple cancer early detection; NCI, National Cancer Institute; NGS, next generation sequencing; PSA, prostate-specific antigen.

Received November 25, 2022, and in revised form, April 26, 2023  
 Published, MCPRO Papers in Press, May 16, 2023, <https://doi.org/10.1016/j.mcpro.2023.100569>

## REFERENCES

1. Pepe, M. S., Etzioni, R., Feng, Z. D., Potter, J. D., Thompson, M. L., Thornquist, M., et al. (2001) Phases of biomarker development for early detection of cancer. *J. Natl. Cancer Inst.* **93**, 1054–1061
2. Diamandis, E. P. (2010) Cancer biomarkers: can we turn recent failures into success? *J. Natl. Cancer Inst.* **102**, 1462–1467
3. Rifai, N., Gillette, M. A., and Carr, S. A. (2006) Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat. Biotechnol.* **24**, 971–983
4. Kern, S. E. (2012) Why your new cancer biomarker may never work: recurrent patterns and remarkable diversity in biomarker failures. *Cancer Res.* **72**, 6097–6101
5. Poste, G., Compton, C. C., and Barker, A. D. (2015) The national biomarker development alliance: confronting the poor productivity of biomarker research and development. *Expert Rev. Mol. Diagn.* **15**, 211–218
6. Selleck, M. J., Senthil, M., and Wall, N. R. (2017) Making meaningful clinical use of biomarkers. *Biomark Insights*. <https://doi.org/10.1177/1177271917715236>
7. Diamandis, E. P. (2004) Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. *J. Natl. Cancer* **96**, 353–356
8. Mischak, H., Ioannidis, J. P., Argiles, A., Attwood, T. K., Bongcam-Rudloff, E., Broenstrup, M., et al. (2012) Implementation of proteomic biomarkers: making it work. *Eur. J. Clin. Invest* **42**, 1027–1036
9. Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials, Board on Health Care Services, Board on Health Sciences Policy, Institute of Medicine, Micheel, C. M., Nass, S. J., and Omenn, G. S. (2012). In: *Evolution of Translational Omics: Lessons Learned and the Path Forward*, National Academies Press (US), Washington, DC
10. Ransohoff, D. F. (2005) Opinion - bias as a threat to the validity of cancer molecular-marker research. *Nat. Rev. Cancer* **5**, 142–149
11. Trevethan, R. (2017) Sensitivity, specificity, and predictive values: foundations, plibilities, and pitfalls in research and practice. *Front. Public Health* **5**, 307
12. Safari, S., Baratloo, A., Elfil, M., and Negida, A. (2015) Evidence based emergency medicine part 2: positive and negative predictive values of diagnostic tests. *Emerg. (Tehran)* **3**, 87–88
13. Edwards, N. J., Oberti, M., Thangudu, R. R., Cai, S., McGarvey, P. B., Jacob, S., et al. (2015) The CPTAC data portal: a resource for cancer proteomics research. *J. Proteome Res.* **14**, 2707–2713
14. Jensen, M. A., Ferretti, V., Grossman, R. L., and Staudt, L. M. (2017) The NCI genomic data commons as an engine for precision medicine. *Blood* **130**, 453–459
15. Milella, M., Lawlor, R. T., Luchini, C., Johns, A. L., Casolino, R., Yoshino, T., et al. (2022) ICGC-ARGO precision medicine: an update on familial matters in pancreatic cancer. *Lancet Oncol.* **23**, 991–992
16. Gentry-Maharaj, A., Blyuss, O., Ryan, A., Burnell, M., Karpinskyj, C., Gunu, R., et al. (2020) Multi-marker longitudinal algorithms incorporating HE4 and CA125 in ovarian cancer screening of postmenopausal women. *Cancers(Basel)* **12**, 1931
17. Kuk, C., Kulasingam, V., Gunawardana, C. G., Smith, C. R., Batruch, I., and Diamandis, E. P. (2009) Mining the ovarian cancer ascites proteome for potential ovarian cancer biomarkers. *Mol. Cell. Proteomics* **8**, 661–669
18. Theil, G., and Fornara, P. (2018) [Innovations in personalized medicine: molecular characterization of liquid biopsy-fake or fact?]. *Urologe A* **57**, 1069–1074
19. Goldstein, N. S., Hewitt, S. M., Taylor, C. R., Yaziji, H., Hicks, D. G., and Comm, A.-H. (2007) Recommendations for improved standardization of immunohistochemistry. *Appl. Immunohistochem. Mol. Morphol.* **15**, 124–133
20. Corcoran, R. B., and Chabner, B. A. (2018) Application of cell-free DNA analysis to cancer treatment. *New Engl. J. Med.* **379**, 1754–1765
21. Walker, N. J., Rashid, M., Yu, S. R., Bignell, H., Lumby, C. K., Livi, C. M., et al. (2022) Hydroxymethylation profile of cell-free DNA is a biomarker for early colorectal cancer. *Sci. Rep.* **12**, 16566
22. Khoo, B. L., Grenici, G., Jing, T. Y., Lim, Y. B., Lee, S. C., Thiery, J. P., et al. (2016) Liquid biopsy and therapeutic response: circulating tumor cell cultures for evaluation of anticancer treatment. *Sci. Adv.* **2**, e1600274
23. Schwab, E. D., and Pienta, K. J. (1996) Cancer as a complex adaptive system. *Med. Hypotheses* **47**, 235–241
24. Brennan, M. D., Cheong, R., and Levchenko, A. (2012) Systems biology. How information theory handles cell signaling and uncertainty. *Science* **338**, 334–335
25. Olivier, M., Asmis, R., Hawkins, G. A., Howard, T. D., and Cox, L. A. (2019) The need for multi-omics biomarker signatures in precision medicine. *Int. J. Mol. Sci.* **20**, 4781
26. Yasui, Y., Pepe, M., Thompson, M. L., Adam, B. L., Wright, G. L., Qu, Y. S., et al. (2003) A data-analytic strategy for protein biomarker discovery: profiling of high-dimensional proteomic data for cancer detection. *Biostatistics* **4**, 449–463
27. Boys, E. L., Liu, J., Robinson, P. J., and Reddel, R. R. (2022) Clinical applications of mass spectrometry-based proteomics in cancer: where are we? *Proteomics* **23**, e2200238
28. Kim, J., and Schoeberl, B. (2015) Beyond static biomarkers-the dynamic response potential of signaling networks as an alternate biomarker? *Sci. Signal.* **8**, fs21
29. Alber, M., Buganza Tepole, A., Cannon, W. R., De, S., Dura-Bernal, S., Garikipati, K., et al. (2019) Integrating machine learning and multiscale modeling-perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. *NPJ Digit. Med.* **2**, 115
30. Deisboeck, T. S., Wang, Z., Macklin, P., and Cristini, V. (2011) Multiscale cancer modeling. *Annu. Rev. Biomed. Eng.* **13**, 127–155
31. Gentles, A. J., and Gallahan, D. (2011) Systems biology: confronting the complexity of cancer. *Cancer Res.* **71**, 5961–5964
32. Boehm, K. M., Khosravi, P., Vanguri, R., Gao, J., and Shah, S. P. (2022) Harnessing multimodal data integration to advance precision oncology. *Nat. Rev. Cancer* **22**, 114–126
33. FDA-NIH Biomarker Working Group. (2016) *BEST (Biomarkers, EndpointS, and Other Tools) Resource [Internet]*. Food and Drug Administration (US), Silver Spring, MD
34. Temple, R. J. (1995) A regulatory authority's opinion about surrogate endpoints. In: Nimmo, W. S., Tucker, G. T., eds. *Clinical Measurement in Drug Evaluation*, John Wiley & Sons, New York: 1–22
35. Delgado, A., and Guddati, A. K. (2021) Clinical endpoints in oncology - a primer. *Am. J. Cancer Res.* **11**, 1121–1131
36. Paulovich, A. G., Whiteaker, J. R., Hoofnagle, A. N., and Wang, P. (2008) The interface between biomarker discovery and clinical validation: the tar pit of the protein biomarker pipeline. *Proteomics Clin. Appl.* **2**, 1386–1402
37. Godsey, J. H., Silvestro, A., Barrett, J. C., Bramlett, K., Chudova, D., Deras, I., et al. (2020) Generic protocols for the analytical validation of next-generation sequencing-based ctDNA assays: a joint consensus recommendation of the BloodPAC's analytical variables working group. *Clin. Chem.* **66**, 1156–1166
38. Addona, T. A., Abbatiello, S. E., Schilling, B., Skates, S. J., Mani, D. R., Bunk, D. M., et al. (2009) Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat. Biotechnol.* **27**, 633–641
39. Klont, F., Horvatovich, P., Govorukhina, N., and Bischoff, R. (2019) Pre- and post-analytical factors in biomarker discovery. *Methods Mol. Biol.* **1959**, 1–22
40. Clarke, C. A., Lang, K., Putcha, G., Beer, J. P., Champagne, M., Ferris, A., et al. (2022) BLOODPAC: collaborating to chart a path towards blood-based screening for early cancer detection. *Clin. Transl. Sci.* **16**, 5–9
41. Moore, H. M. (2014) Moving toward biospecimen harmonization with evidence-based practices. *Biopreserv. Biobank.* **12**, 79–80
42. Lawler, M., Haussler, D., Siu, L. L., Haendel, M. A., McMurry, J. A., Knoppers, B. M., et al. (2017) Sharing clinical and genomic data on cancer - the need for global solutions. *New Engl. J. Med.* **376**, 2006–2009
43. Swanton, C., and Govindan, R. (2016) Clinical implications of genomic discoveries in lung cancer. *New Engl. J. Med.* **374**, 1864–1873

44. Page, D. B. (2022) The human tumor atlas network's beginning steps toward the future of collaborative multi-omic discovery. *Cell Rep. Med.* **3**, 100532
45. Rozenblatt-Rosen, O., Regev, A., Oberdoerffer, P., Nawy, T., Hupalowska, A., Rood, J. E., et al. (2020) The human tumor atlas network: charting tumor transitions across space and time at single-cell resolution. *Cell* **181**, 236–249
46. Kwong, G. A., Ghosh, S., Gamboa, L., Patriotis, C., Srivastava, S., and Bhatia, S. N. (2021) Synthetic biomarkers: a twenty-first century path to early cancer detection. *Nat. Rev. Cancer* **21**, 655–668
47. Nakayasu, E. S., Gritsenko, M., Piehowski, P. D., Gao, Y., Orton, D. J., Schepmoes, A. A., et al. (2021) Tutorial: best practices and considerations for mass-spectrometry-based protein biomarker discovery and validation. *Nat. Protoc.* **16**, 3737–3760
48. Mateo, J., Steuten, L., Aftimos, P., Andre, F., Davies, M., Garralda, E., et al. (2022) Delivering precision oncology to patients with cancer. *Nat. Med.* **28**, 658–665
49. Nowell, P. C., and Hungerford, D. A. (1960) A minute chromosome in human chronic granulocytic-leukemia. *Science* **142**, 1497
50. Rowley, J. D. (1973) Letter: a new consistent chromosomal abnormality in chronic myelogenous leukemia identified by quinacrine fluorescence and giemsa staining. *Nature* **243**, 290–293
51. Druker, B. J., Talpaz, M., Resta, D. J., Peng, B., Buchdunger, E., Ford, J. M., et al. (2001) Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *New Engl. J. Med.* **344**, 1031–1037
52. Jimenez, C. R., and Verheul, H. M. (2014) Mass spectrometry-based proteomics: from cancer biology to protein biomarkers, drug targets, and clinical applications. *Am. Soc. Clin. Oncol. Educ. Book*, e504–e510
53. Brawley, O. W., Thompson, I. M., Jr., and Gronberg, H. (2016) Evolving recommendations on prostate cancer screening. *Am. Soc. Clin. Oncol. Educ. Book* **35**, e80–e87
54. Lu-Yao, G. L., McLerran, D., Wasson, J., and Wennberg, J. E. (1993) An assessment of radical prostatectomy. Time trends, geographic variation, and outcomes. The prostate patient outcomes research team. *JAMA* **269**, 2633–2636
55. Prensner, J. R., Rubin, M. A., Wei, J. T., and Chinnaiyan, A. M. (2012) Beyond PSA: the next generation of prostate cancer biomarkers. *Sci. Transl. Med.* **4**, 127rv3
56. Baron, J. A. (2012) Screening for cancer with molecular markers: progress comes with potential problems. *Nat. Rev. Cancer* **12**, 368–371
57. Maxim, L. D., Niebo, R., and Utell, M. J. (2014) Screening tests: a review with examples. *Inhal. Toxicol.* **26**, 811–828
58. Tikkinen, K. A. O., Dahm, P., Lytvyn, L., Heen, A. F., Vernooij, R. W. M., Siemieniuk, R. A. C., et al. (2018) Prostate cancer screening with prostate-specific antigen (PSA) test: a clinical practice guideline. *BMJ* **362**, k3581
59. Crosby, D., Bhatia, S., Brindle, K. M., Coussens, L. M., Dive, C., Ember-ton, M., et al. (2022) Early detection of cancer. *Science* **375**, eaay9040
60. Barker, A. D., Sigman, C. C., Kelloff, G. J., Hylton, N. M., Berry, D. A., and Esserman, L. J. (2009) I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin. Pharmacol. Ther.* **86**, 97–100
61. Wang, H. Y., and Yee, D. (2019) I-SPY 2: a neoadjuvant adaptive clinical trial designed to improve outcomes in high-risk breast cancer. *Curr. Breast Cancer Rep.* **11**, 303–310
62. DiMasi, J. A., Hansen, R. W., and Grabowski, H. G. (2003) The price of innovation: new estimates of drug development costs. *J. Health Econ.* **22**, 151–185
63. Burd, A., Levine, R. L., Ruppert, A. S., Mims, A. S., Borate, U., Stein, E. M., et al. (2020) Precision medicine treatment in acute myeloid leukemia using prospective genomic profiling: feasibility and preliminary efficacy of the beat AML master trial. *Nat. Med.* **26**, 1852–1858
64. Alexander, B. M., Ba, S., Berger, M. S., Berry, D. A., Cavenee, W. K., Chang, S. M., et al. (2018) Adaptive global innovative learning environment for glioblastoma: Gbm AGILE. *Clin. Cancer Res.* **24**, 737–743
65. Ettinger, D. S., Wood, D. E., Aisner, D. L., Akerley, W., Bauman, J. R., Bharat, A., et al. (2023) NCCN Guidelines® insights: non-small cell lung cancer, version 2.2023. *J. Natl. Compr. Canc. Netw.* **21**, 340–350
66. Tannock, I. F., and Hickman, J. A. (2016) Limits to personalized cancer medicine. *New Engl. J. Med.* **375**, 1289–1294
67. Zhang, B., Whiteaker, J. R., Hoofnagle, A. N., Baird, G. S., Rodland, K. D., and Paulovich, A. G. (2019) Clinical potential of mass spectrometry-based proteogenomics. *Nat. Rev. Clin. Oncol.* **16**, 256–268
68. Meissner, F., Scheltema, R. A., Mollenkopf, H. J., and Mann, M. (2013) Direct proteomic quantification of the secretome of activated immune cells. *Science* **340**, 475–478
69. Koussounadis, A., Langdon, S. P., Um, I. H., Harrison, D. J., and Smith, V. A. (2015) Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. *Sci. Rep.* **5**, 10775
70. Kennedy, J. J., Abbatiello, S. E., Kim, K., Yan, P., Whiteaker, J. R., Lin, C. W., et al. (2014) Demonstrating the feasibility of large-scale development of standardized assays to quantify human proteins. *Nat. Methods* **11**, 149–155
71. Anurag, M., Jaehnig, E. J., Krug, K., Lei, J. T., Bergstrom, E. J., Kim, B. J., et al. (2022) Proteogenomic markers of chemotherapy resistance and response in triple-negative breast cancer. *Cancer Discov.* **12**, 2586–2605
72. Hoofnagle, A. N., Becker, J. O., Wener, M. H., and Heinecke, J. W. (2008) Quantification of thyroglobulin, a low-abundance serum protein, by immunoaffinity peptide enrichment and tandem mass spectrometry. *Clin. Chem.* **54**, 1796–1804
73. Whiteaker, J. R., Halusa, G. N., Hoofnagle, A. N., Sharma, V., MacLean, B., Yan, P., et al. (2014) CPTAC assay portal: a repository of targeted proteomic assays. *Nat. Methods* **11**, 703–704
74. Son, M., Kim, H., Yeo, I., Kim, Y., Sohn, A., and Kim, Y. (2019) Method validation by CPTAC guidelines for multi-protein marker assays using multiple reaction monitoring-mass spectrometry. *Biotechnol. Bio-process Eng.* **24**, 343–358
75. Rudnick, P. A., Markey, S. P., Roth, J., Mirokhin, Y., Yan, X. J., Tchekhovskoi, D. V., et al. (2016) A description of the clinical proteomic tumor analysis consortium (CPTAC) common data analysis pipeline. *J. Proteome Res.* **15**, 1023–1032
76. Soltis, A. R., Bateman, N. W., Liu, J., Nguyen, T., Franks, T. J., Zhang, X., et al. (2022) Proteogenomic analysis of lung adenocarcinoma reveals tumor heterogeneity, survival determinants, and therapeutically relevant pathways. *Cell Rep. Med.* **3**, 100819
77. Poste, G. (2011) Bring on the biomarkers. *Nature* **469**, 156–157
78. Cancer moonshot aims at liquid biopsies. *Nat. Biotechnol.* **40**, (2022), 1422
79. Jackson, H. W., Fischer, J. R., Zanotelli, V. R. T., Ali, H. R., Mechera, R., Soysal, S. D., et al. (2020) The single-cell pathology landscape of breast cancer. *Nature* **578**, 615–620
80. Chevrier, S., Levine, J. H., Zanotelli, V. R. T., Silina, K., Schulz, D., Bacac, M., et al. (2017) An immune atlas of clear cell renal cell carcinoma. *Cell* **169**, 736–749
81. Giesen, C., Wang, H. A. O., Schapiro, D., Zivanovic, N., Jacobs, A., Hattendorf, B., et al. (2014) Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Methods* **11**, 417–422
82. De Vargas Roditi, L., Jacobs, A., Rueschoff, J. H., Bankhead, P., Chevrier, S., Jackson, H. W., et al. (2022) Single-cell proteomics defines the cellular heterogeneity of localized prostate cancer. *Cell Rep. Med.* **3**, 100604
83. Subrahmanyam, P. B., Dong, Z. W., Gusenleitner, D., Giobbie-Hurder, A., Severgnini, M., Zhou, J., et al. (2018) Distinct predictive biomarker candidates for response to anti-CTLA-4 and anti-PD-1 immunotherapy in melanoma patients. *J. Immunother. Cancer* **6**, 18
84. Keating, S. M., Taylor, D. L., Plant, A. L., Litwack, E. D., Kuhn, P., Greenspan, E. J., et al. (2018) Opportunities and challenges in implementation of multiparameter single cell analysis platforms for clinical translation. *Clin. Transl. Sci.* **11**, 267–276
85. Gatto, L., Aebersold, R., Cox, J., Demichev, V., Derks, J., Emmott, E., et al. (2023) Initial recommendations for performing, benchmarking and reporting single-cell proteomics experiments. *Nat. Methods* **20**, 375–386
86. Adhikari, S., Nice, E. C., Deutsch, E. W., Lane, L., Omenn, G. S., Pennington, S. R., et al. (2020) A high-stringency blueprint of the human proteome. *Nat. Commun.* **11**, 5301
87. Omenn, G. S., Lane, L., Overall, C. M., Pineau, C., Packer, N. H., Cristea, I. M., et al. (2023) The 2022 report on the human proteome from the HUPO human proteome project. *J. Proteome Res.* **22**, 1024–1042

88. Katrukha, I. A., and Katrukha, A. G. (2021) Myocardial injury and the release of troponins I and T in the blood of patients. *Clin. Chem.* **67**, 124–130
89. Campbell, J. B. (2007) *Introduction to Remote Sensing*, 4th Ed, Guildford Press, New York, NY
90. Shannon, C. E., and Weaver, W. (1949) *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, IL
91. Avanzini, S., Kurtz, D. M., Chabon, J. J., Moding, E. J., Hori, S. S., Gambhir, S. S., et al. (2020) A mathematical model of ctDNA shedding predicts tumor detection size. *Sci. Adv.* **6**, eabc4308
92. Hori, S. S., and Gambhir, S. S. (2011) Mathematical model identifies blood biomarker-based early cancer detection strategies and limitations. *Sci. Transl. Med.* **3**, 109ra116
93. Hori, S. S., Lutz, A. M., Paulmurugan, R., and Gambhir, S. S. (2017) A model-based personalized cancer screening strategy for detecting early-stage tumors using blood-borne biomarkers. *Cancer Res.* **77**, 2570–2584
94. Lutz, A. M., Willmann, J. K., Cochran, F. V., Ray, P., and Gambhir, S. S. (2008) Cancer screening: a mathematical model relating secreted blood biomarker levels to tumor sizes. *PLoS Med.* **5**, e170
95. Machiraju, G. B., Mallick, P., and Frieboes, H. B. (2020) Multicompartment modeling of protein shedding kinetics during vascularized tumor growth. *Sci. Rep.* **10**, 16709
96. Frieboes, H. B., Curtis, L. T., Wu, M., Kani, K., and Mallick, P. (2015) Simulation of the protein-shedding kinetics of a fully vascularized tumor. *Cancer Inform.* **14**, 163–175
97. Fang, Q., Kani, K., Faca, V. M., Zhang, W., Zhang, Q., Jain, A., et al. (2011) Impact of protein stability, cellular localization, and abundance on proteomic detection of tumor-derived proteins in plasma. *PLoS One* **6**, e23090
98. Tainsky, M. A. (2009) Genomic and proteomic biomarkers for cancer: a multitude of opportunities. *Biochim. Biophys. Acta* **1796**, 176–193
99. Faca, V. M., Song, K. S., Wang, H., Zhang, Q., Krasnoselsky, A. L., Newcomb, L. F., et al. (2008) A mouse to human search for plasma proteome changes associated with pancreatic tumor development. *PLoS Med.* **5**, e123
100. Gast, M. C., Schellens, J. H., and Beijnen, J. H. (2009) Clinical proteomics in breast cancer: a review. *Breast Cancer Res. Treat.* **116**, 17–29
101. Hanash, S. M., Pitteri, S. J., and Faca, V. M. (2008) Mining the plasma proteome for cancer biomarkers. *Nature* **452**, 571–579
102. Pitteri, S. J., Faca, V. M., Kelly-Spratt, K. S., Kasarda, A. E., Wang, H., Zhang, Q., et al. (2008) Plasma proteome profiling of a mouse model of breast cancer identifies a set of up-regulated proteins in common with human breast cancer cells. *J. Proteome Res.* **7**, 1481–1489
103. Xue, H., Lu, B., and Lai, M. (2008) The cancer secretome: a reservoir of biomarkers. *J. Transl. Med.* **6**, 52
104. Anderson, N. L., and Anderson, N. G. (2002) The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* **1**, 845–867
105. Polanski, M., and Anderson, N. L. (2007) A list of candidate cancer biomarkers for targeted proteomics. *Biomark Insights* **1**, 1–48
106. Creighton, C., Kuick, R., Misek, D. E., Rickman, D. S., Brichory, F. M., Rouillard, J. M., et al. (2003) Profiling of pathway-specific changes in gene expression following growth of human cancer cell lines transplanted into mice. *Genome Biol.* **4**, R46
107. Balk, S., Ko, Y. J., and Bubley, G. J. (2003) Biology of prostate-specific antigen. *J. Clin. Oncol.* **21**, 383–391
108. Catalona, W. J., Smith, D. S., Ratliff, T. L., Dodds, K. M., Coplen, D. E., Yuan, J. J., et al. (1991) Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. *N. Engl. J. Med.* **324**, 1156–1161
109. Mastropalo, W., Fernandez, Z., and Miller, E. L. (1986) Pronounced increases in the concentration of an ovarian tumor marker, CA-125, in serum of a healthy subject during menstruation. *Clin. Chem.* **32**, 2110–2111
110. You, B., Perrin, P., Freyer, G., Ruffion, A., Tranchand, B., Hénin, E., et al. (2008) Advantages of prostate-specific antigen (PSA) clearance model over simple PSA half-life computation to describe PSA decrease after prostate adenectomy. *Clin. Biochem.* **41**, 785–795
111. Belle, A., Tanay, A., Bitincka, L., Shamir, R., and O'Shea, E. K. (2006) Quantification of protein half-lives in the budding yeast proteome. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 13004–13009
112. Bohley, P. (1995) The fates of proteins in cells. *Naturwissenschaften* **82**, 544–550
113. Carlson, N., Rogers, S., and Rechsteiner, M. (1987) Microinjection of ubiquitin: changes in protein degradation in HeLa cells subjected to heat-shock. *J. Cell Biol.* **104**, 547
114. Guruprasad, K., Reddy, B., and Pandit, M. W. (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence. *Protein Eng.* **4**, 155
115. Rogers, S., Wells, R., and Rechsteiner, M. (1986) Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* **234**, 364
116. Piening, B. D., Zhou, W., Contrepois, K., Rost, H., Gu Urban, G. J., Mishra, T., et al. (2018) Integrative personal omics profiles during periods of weight gain and loss. *Cell Syst.* **6**, 157–170.e8
117. Pitteri, S. J., and Hanash, S. M. (2011) Confounding effects of hormone replacement therapy in protein biomarker studies. *Cancer Epidemiol. Biomarkers Prev.* **20**, 134–139
118. Pitteri, S. J., Hanash, S. M., Aragaki, A., Amon, L. M., Chen, L., Busald Buson, T., et al. (2009) Postmenopausal estrogen and progestin effects on the serum proteome. *Genome Med.* **1**, 121
119. Meza-Espinoza, J. P., Vasquez-Jimenez, E. A., Barajas-Torres, R. L., Magana-Torres, M. T., and Gonzalez-Garcia, J. R. (2019) BCR/ABL1 transcripts in healthy individuals: a comparative analysis between first-degree relatives of patients with chronic myelogenous leukemia and subjects without antecedents of the disease. *Ann. Clin. Lab. Sci.* **49**, 703–709
120. Kuan, J. W., Su, A. T., Leong, C. F., Osato, M., and Sashida, G. (2020) Systematic review of normal subjects harbouring BCR-ABL1 fusion gene. *Acta Haematol.* **143**, 96–111
121. Wishart, D. S., Bartok, B., Oler, E., Liang, K. Y. H., Budinski, Z., Berjanskii, M., et al. (2021) MarkerDB: an online database of molecular biomarkers. *Nucleic Acids Res.* **49**, D1259–D1267
122. Jayaraman, K. (2011) 13,000 Biomarker deal. *Nat. Biotechnol.* **29**, 777
123. Shannon, C. E. (1948) A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423
124. Sturmburg, J. P. (2021) Health and disease are dynamic complex-adaptive states implications for practice and research. *Front. Psychiatry* **12**, 595124
125. Barker, A. D., and Lee, J. S. H. (2022) Translating "big data" in oncology for clinical benefit: progress or paralysis. *Cancer Res.* **82**, 2072–2075
126. Ho, D. (2020) Artificial intelligence in cancer therapy. *Science* **367**, 982–983
127. Lewis, J. R. R., Kerridge, I., and Lipworth, W. (2017) Use of real-world data for the research, development, and evaluation of oncology precision medicines. *JCO Precis. Oncol.* **1**, 1–11
128. Collins, F. S., and Tabak, L. A. (2014) NIH plans to enhance reproducibility. *Nature* **505**, 612–613
129. Prakash, A., Piening, B., Whiteaker, J., Zhang, H., Shaffer, S. A., Martin, D., et al. (2007) Assessing bias in experiment design for large scale mass spectrometry-based quantitative proteomics. *Mol. Cell. Proteomics* **6**, 1741–1748
130. Crook, O. M., Chung, C. W., and Deane, C. M. (2022) Challenges and opportunities for bayesian statistics in proteomics. *J. Proteome Res.* **21**, 849–864
131. Rodriguez, H., Rivers, R., Kinsinger, C., Mesri, M., Hiltke, T., Rahbar, A., et al. (2010) Reconstructing the pipeline by introducing multiplexed multiple reaction monitoring mass spectrometry for cancer biomarker verification: an NCI-CPTC initiative perspective. *Proteomics Clin. Appl.* **4**, 904–914
132. Engel, K. B., Vaught, J., and Moore, H. M. (2014) National cancer institute biospecimen evidence-based practices: a novel approach to pre-analytical standardization. *Biopreserv. Biobank.* **12**, 148–150
133. Hoofnagle, A. N., Whiteaker, J. R., Carr, S. A., Kuhn, E., Liu, T., Massoni, S. A., et al. (2016) Recommendations for the generation, quantification, storage, and handling of peptides used for mass spectrometry-based assays. *Clin. Chem.* **62**, 48–69

134. Kearney, P., Boniface, J. J., Price, N. D., and Hood, L. (2018) The building blocks of successful translation of proteomics to the clinic. *Curr. Opin. Biotech.* **51**, 123–129
135. Hayes, D. F., Allen, J., Compton, C., Gustavsen, G., Leonard, D. G. B., McCormack, R., et al. (2013) Breaking a vicious cycle. *Sci. Transl. Med.* **5**, 196cm6
136. Sardar, R., Sharma, A., and Gupta, D. (2021) Machine learning assisted prediction of prognostic biomarkers associated with COVID-19, using clinical and proteomics data. *Front. Genet.* **12**, 636441
137. Xu, L. L., Young, A., Zhou, A., and Rost, H. L. (2020) Machine learning in mass spectrometric analysis of DIA data. *Proteomics* **20**, e1900352
138. Havugimana, P. C., Hu, P. Z., and Emili, A. (2017) Protein complexes, big data, machine learning and integrative proteomics: lessons learned over a decade of systematic analysis of protein interaction networks. *Expert Rev. Proteomic* **14**, 845–855
139. Prakash, A., Sutton, J., Richmond, T., Piening, B., Whiteaker, J., Zhang, H., et al. (2006) Assessing reproducibility of mass spectrometry experiments for biomarker discovery in human plasma. *Mol. Cell. Proteomics* **5**, S280