GENETIC EPIDEMIOLOGY (C AMOS, SECTION EDITOR)

Considerations of Biomarker Application for Cancer Continuum in the Era of Precision Medicine

Rayjean J. Hung[1,2] · Elham Khodayari Moez[1] · Shana J. Kim[2] · Sanjeev Budhathoki[1] · Jennifer D. Brooks[2]

Abstract

Purpose of the Review The goal of this review is to highlight emerging biomarker research by the key phases of the cancer continuum and outline the methodological considerations for biomarker application.

Recent Findings While biomarkers have an established role in targeted therapy and, to some extent, disease monitoring, their role in early detection and survivorship remains to be elucidated. With the advent of omics technology, the discovery of biomarkers has been accelerated exponentially, therefore careful consideration to ensure an unbiased study design and robust validity is crucial.

Summary The rigor of biomarker research holds the key to the success of precision health care. The potential clinical utility and the feasibility of implementation should be central to future biomarker research study design.

Keywords Biomarker · Cancer · Precision medicine · Early detection · Prognosis · Survivorship

Introduction

With the era of multi-omics technologies blooming in the last two decades, precision medicine, a term that started to gain momentum with the US Institute of Medicine's National Research Council Report in 2011 [1], is now within sight. Initially narrowly defined as targeted treatments, the concept of precision medicine now includes tailored health care more broadly, based on an individual's risk profiles, including exposure history and/or molecular characterizations. There are international efforts underway to build genomic databases and infrastructure linkable with electronic medical records in the UK, France, and other countries to facilitate

This article is part of the Topical Collection on *Genetic Epidemiology*

✉ Rayjean J. Hung
  rayjean.hung@lunenfeld.ca

  Jennifer D. Brooks
  jennifer.brooks@utoronto.ca

1   Prosserman Centre for Population Health Research, Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, Canada

2   Dalla Lana School of Public Health, University of Toronto, Toronto, Canada

population-based precision health care [2, 3]. In particular, early detection and optimal management for cancer patients across the cancer continuum using the precision medicine approach have garnered substantial attention due to the continued high burden of disease and poor prognosis of most cancers either as years or quality of life lost after cancer diagnosis [4].

The promise of precision medicine heavily relies on the ability to accurately classify individuals into precise subgroups to be targeted for early detection or optimal disease management. While there are many different factors that can impact realizing the potential of precision medicine, without a doubt, reliable and validated biomarkers are the foundation needed to ensure its success. There has been a wide range of biomarker discovery and validation work, for various purposes across the cancer continuum, including exposure assessment, early detection, clinical management (e.g., treatment selection), and survivorship. With cancer as the disease model, in this review, we outline the application of emerging biomarkers with examples along the cancer continuum. Figure 1 illustrates the utility of biomarkers in each of the key phases in the cancer continuum. Examples of emerging biomarkers are described in the following sections, followed by the methodological challenges and considerations.
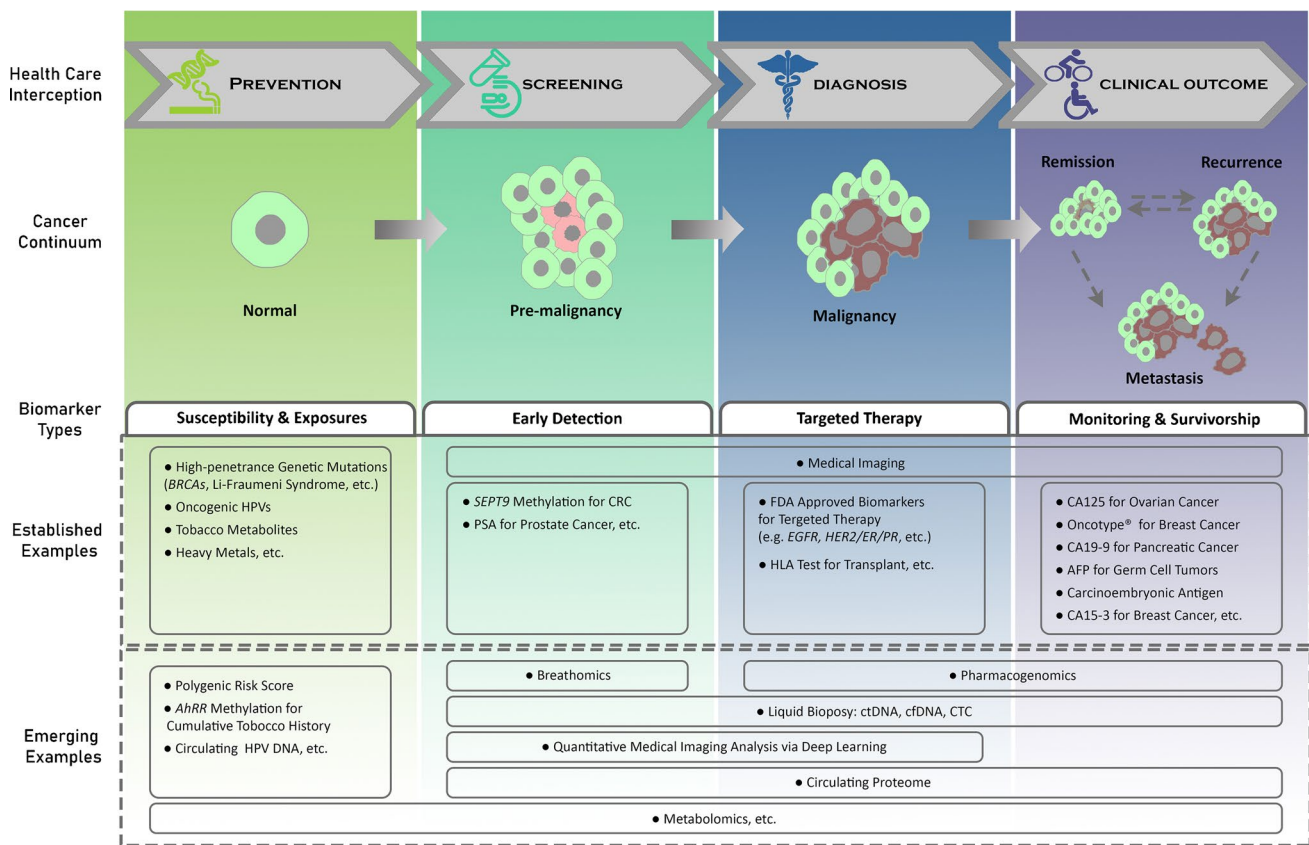
Fig. 1 Biomarker applications across the cancer continuum. Abbreviations: AFP, α-fetoprotein; *AhRR*, aryl-hydrocarbon receptor repressor; *BRCA*, breast cancer gene; CA125, cancer antigen 125; CA19-9, carbohydrate antigen 19–9; CRC, colorectal cancer; cfDNA, cell-free DNA; ctDNA, circulating tumor DNA; CTC, circulating tumor cell; *EGFR*, epidermal growth factor receptor; ER, estrogen receptor; HLA, human leukocyte antigen; HPV, human papillomavirus; PSA, prostate-specific antigen; *SEP9*, septin 9; PR, progesterone receptor

Biomarkers for Susceptibility and Exposures

It is widely accepted that carcinogenesis is a multifactorial process, which often involves both intrinsic susceptibility and extrinsic exposures, as well as accumulations of somatic mutations that ultimately initiate tumorigenesis.

Intrinsic susceptibility is determined by inherited genomes, the variations of which range from rare mutations that cause familial syndromes to common sequence variations. While rare mutations have higher penetrance and larger effect sizes, they only account for a limited fraction of heritability for most complex diseases such as cancer [5]. To capture the individual susceptibility represented by sequence variations, including both common and rare variants, polygenic risk scores (PRS), as the sum of the risk alleles weighted by their effect sizes, are the current mainstream approach [6]. The application of PRS following genome-wide discovery and rigorous validation has shown some successes for risk prediction of several cancer sites, such as breast, prostate, and lung cancer [7–9]. PRS are now being incorporated into risk prediction models used to assess absolute risk trajectory in a defined time window (for example, 5-year risk), which could potentially have clinical implications regarding screening eligibility [8, 10]. While the National Comprehensive Cancer Network currently does not recommend the routine use of PRS outside of clinical trials [11], in recognition of the rapid advances in PRS research for complex diseases, the International Common Disease Alliance (ICDA) outlined the considerations for responsible use of PRS in the clinic, potential benefits, harms, and possible mitigation strategy [12].

Extrinsic exposures can be measured in various sources of biospecimens, including peripheral blood, saliva, urine, and sputum, depending on the target molecules. For example, it has long been established that tobacco-specific carcinogens such as 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL) can be measured in urine samples with a half-life of around 10–45 days, while many other metabolites have shorter half-life and tend to represent only recent exposures [13]. A recent multi-ethnic

prospective study based on total nicotine equivalents (TNE), a sum of nicotine and its major metabolites, shows that the nicotine uptake may partially explain the difference in lung cancer risk across racial groups while consuming similar levels of cigarettes, although their rapid metabolism remains to be the main challenge [14]. On the other hand, there is growing literature that demonstrates that aryl-hydrocarbon receptor repressor (*AhRR*) hypomethylation measured in peripheral blood is a robust biomarker for long-term tobacco smoking history, exhibiting a strong dose–response relationship with intensity, duration, and time since quitting, which has been replicated in independent studies, including multiple racial groups [15–17]. *AhRR* hypomethylation has been shown to predict smoking-related morbidity and morality after adjusting for self-reported smoking history in multiple prospective cohorts [18, 19], although its clinical utility remains to be determined. Exposures to specific carcinogens (e.g., tobacco, aflatoxin, and ultraviolet light exposure) can also be identified based on the somatic mutation signatures. This information is informative for understanding carcinogenesis but not applicable for risk assessment due to the need for tumor tissues [20].

Another prominent example of a biomarker for extrinsic exposure to cancer risk is the detection of the oncogenic human papillomavirus (HPV), which is considered a necessary cause of cervical cancer and substantially increases the risk for oropharyngeal cancer (OPC) [21]. For cervical cancer, the standard test includes the HPV DNA test in conjunction with the Pap test. For OPC, it has been shown that HPV 16 E6 antibody can be detected in serum samples more than 10 years prior to cancer diagnosis [22]. While it is associated with an almost 100-fold increase in oropharyngeal cancer risk, the low incidence of OPC in the general population and the associated low positive predictive values pose challenges in using it as a population-wide screening tool [23]. However, a risk prediction model that specifically assesses absolute risks in the defined time window can help to determine the actionability depending on the individual's risk profile. In addition, the recent development of HPV-seq aiming to detect circulating HPV DNA may further enhance the possibility of early detection of HPV-related cancers [24], although its validation in pre-diagnostic samples will be required.

An alternative approach of determine the extrinsic exposures is metabolomics, an untargeted approach to analyzing hundreds or thousands of small molecular metabolites by their chemical properties or atomic weights. A comprehensive review of metabolomics in cancer research was recently published, thus not included here [25].

Biomarkers for Early Detection

With mortality rates remaining high for many cancer types, early detection holds the key to improving cancer survival. Cancer screening programs for lung, breast, cervical, and colorectal cancers have been shown to reduce mortality substantially [26]. Current guidelines are typically based on age and smoking history (lung) and modified by family history (breast). Development and validation of risk prediction models based on individual's medical history has been a very active area of research, and there is growing evidence that well-validated risk models can improve screening efficiency [27]. However, the role of biomarkers for early detection still has a long way to go. Biomarker tests that are clinically available, such as *SEPT9* methylation for colorectal cancer or prostate-specific-antigen (PSA), either lack sensitivity (former) or specificity (latter) [28].

Liquid biopsy aims to detect circulating tumor cells (CTC) or tumor-derived materials from blood or other bodily fluids, as a potential biomarker for the early detection of cancers [29, 30]. In particular, research on circulating cell-free DNA (cfDNA), short fragments of DNA of about 150 to 200 base pairs in length released into the circulation through tumor apoptosis or necrosis, has gained momentum in recent years [31, 32]. While the faction of tumor-derived DNA in the total cfDNA is low, the abundance of cfDNA was shown to correlate with the tumor burden [29, 30]. Since genetic and epigenetic alterations are some of the earliest events in the carcinogenic process, cfDNA provides the opportunity to identify these alterations, potentially even before clinical detection (e.g., symptoms, palpable lump, and visibility on imagery) with the minimally invasive procedure (i.e., blood draw).

Previous cfDNA research focusing on somatic mutations has shown low sensitivity in cancer detection because of their dependency on a limited set of recurrent mutations [31]. The sensitivity is particularly low for early-stage cancer with a low tumor burden. To improve the performance of cfDNA in cancer diagnosis, a combination of mutation markers with other markers has been used. A study combining mutation detection and protein markers reported sensitivity between 69 and 98% across eight tumor types and stated specificity of higher than 95% based on a multicancer test panel, although the performance for the early-stage cancers was relatively poor [33]. Since the study was largely based on symptomatic patients at cancer diagnosis, it is unclear how the test panel will perform in the early detection setting, where most patients will have early-stage disease.

Epigenetic changes, particularly DNA methylation, represent another promising biomarker for early detection using cfDNA. It is well established that DNA methylation plays an important role in carcinogenesis. Aberrant promoter hypermethylation associated with the silencing of tumor

suppression genes is known to affect tumor progression, while global hypomethylation can lead to genomic instability [34]. There are hundreds of differentially methylated regions across the genome, which when used in combination can help to improve sensitivity. Furthermore, its cell-type-specific nature can be used to infer tissues of origin [35]. Methylation based on cfDNA has moved from a few targeted regions to genome-wide interrogation in recent years. Based on 7 tumor types and healthy controls, methylation profiles in cfDNA were shown to be specific to the tissue of origin and achieved high discriminating accuracy (area under the receiver operating curve [AUC] > 0.90) for pancreatic cancer and lung cancer in the validation set, with comparable performance in early versus late-stage patients [36]. CfDNA fragmentomics is another recent topic of interest that focuses on fragmentation patterns of plasma cfDNA, which is known to be non-random and has the potential to enrich the selection of cfDNA targets that are of interest [37].

In general, these results from circulating cfDNA or circulating tumor DNA (ctDNA) have been encouraging, and several biotechnology companies have been established in the recent years with the goal to bring liquid biopsy tests to the masses [38]. However, large scale prospective studies evaluating the predictive performances in real-world setting will be necessary before cfDNA based test for early cancer detection is feasible and assessment of specifically early-stage cancers would be the key [39].

Another intriguing area of research is breathomics, which focuses on the analysis of volatile organic compounds (VOCs) in the exhaled breath for cancer detection. Initially started for respiratory diseases, an increasing number of studies have been conducted for other cancer sites, such as digestive tract cancers. Despite study heterogeneity, recent studies have shown promising predictive accuracy for differentiating cancer samples versus non-cancer samples [40–42]. However, its cancer specificity and utility for early detection before cancer diagnosis remain to be assessed in prospective studies.

Biomarkers for Clinical Management

At the time of cancer diagnosis, biomarkers can be used to inform optimal patient management based on predicted clinical outcomes and treatment response. For example, breast cancer patients are routinely tested for the expression of estrogen and progesterone receptor (ER/PR) and human epidermal growth factor receptor (HER2) to determine treatment options. In the last 10 years, oncotype-dx (a multi-gene panel test) has been increasingly used to predict the risk of recurrence to inform decisions related to the risks and benefits of chemotherapy in women with early-stage ER + / HER2 − disease [43]. Similarly, epidermal growth factor receptor (*EGFR*) mutations, anaplastic lymphoma kinase

(*ALK*) rearrangement, and programmed death-ligand 1 (PD-L1) are tested among lung cancer patients to identify the best-suited target therapy [43]. These are encapsulated in the field of pharmacogenetics, which is at the forefront of precision medicine today.

Pharmacogenetic biomarkers include somatic characterizations pharmacogenetic biomarkers directly measured in the tumor tissues, (e.g., *EGFR* and ER/PR/HER2 as described above), or inherited germline mutations or variations. Germline genetic variation can lead to variation in treatment response by influencing the pharmacokinetics of different therapeutic agents, leading to drug-induced adverse events and altered drug response. While a number of variants in candidate genes have been identified over the years (e.g., cytochrome P450 family 2 subfamily C member 9 (*CYP2C9)* and warfarin, cytochrome P450 family 2 subfamily D member 6 (*CYP2D6)* for tamoxifen), it is now possible to conduct genome-wide screens for genetic markers that are associated with absorption, distribution, metabolism, and excretion of the drug, and subsequently affect drug efficacy and risk of adverse effects [44]. As of June 2021, there are approximately 480 indications related to biomarkers approved by the US Food and Drug Administration (FDA) for 325 medications (among those, 110 are related to oncology), demonstrating the impact of precision medicine in current clinical practice in general [45]. While most drugs labeled by FDA are based on the variations in the somatic events, approximately half of the European Medicines Agency (EMA) labels related to biomarkers refer to pharmacokinetic germline variations [46].

While most of the tests to identify cancer treatment options performed at diagnosis are based on somatic characterizations, the tests to monitor cancer progression can be done using peripheral blood. For example, cancer antigen 125 (CA-125) is commonly used to monitor cancer burden after treatment for ovarian cancer [47]. The use of peripheral blood for the identification of circulating proteomic biomarkers for early detection and patient management is an emerging area of research that capitalizes on the advantages of using a non-invasive approach.

Proteomics, which characterizes protein networks and monitors post-translational modifications, structural changes, and protein–protein interactions, provides a rich source of information for outcome predictions to inform patient management. Proteomics has the potential to identify biomarkers for multiple cancer-related outcomes, including early detection, prognosis, metastasis, tumor growth, and aggressiveness [48–53]. For example, a recent proteomics study classified patients with hepatocellular carcinomas into subtypes with distinct clinical trajectories and underlying tumor biology that could be targeted for personalized therapies [54]. Furthermore, investigations into the tumor microenvironment, more specifically the cancer-associated

fibroblast (CAF), through proteome approaches, have identified an increased level of nicotinamide N-methyltransferase (NNMT) expression as a potential biomarker of metastasis [55]. Identification of kinases driving tumor growth and aggressiveness is one example with an important clinical impact [56].

Circulating proteomics focuses on biofluid-based protein markers measured noninvasively in serum, plasma, urine, or salvia. With the advent of high-throughput technology based on nucleic acid aptamers or proximity extension assays, researchers can now analyze thousands of protein markers with as little as 30 μl of plasma or serum [57, 58]. Proteogenomic combines information from proteomics, genomics, and transcriptomics and has been increasingly recognized as an important field in precision oncology [59–62]. Several recent large-scale investigations [63, 64] using this approach have found that incorporating information from the circulating proteome helps to fill in the gaps between genomics and observed phenotypes for a wide range of complex diseases, including cancer. Through the assessment of protein quantitative trait loci (pQTL) and their associated networks, this integrative approach could facilitate the prioritization of potential drug targets for future clinical trials [63, 64].

Biomarker for Survivorship

Cancer survivorship represents a distinct phase in the cancer continuum after treatment and before end-of-life. This period largely focuses on the long-term follow-up to manage late effects of cancer treatment, surveillance for cancer recurrence, and improve the quality of life of survivors to overall decrease the risk of mortality [65]. Biomarkers have the potential to play an important role in the cancer survivorship phase, particularly to detect and manage late effects of cancer treatment; however, this area of research is largely unstudied.

The benefits of cancer therapies come with a host of challenges related to the side-effects/toxicities that can persist into the survivorship phase (long-term effects) or present later as a consequence of the treatment (late effect) [66]. These effects can be far reaching impacting cardiovascular, musculoskeletal, respiratory and mental health, the endocrine system, reproductive organs, and immune system [67, 68].

Late effects of cardiotoxic chemotherapy (particularly anthracyclines) and chest radiation include congestive heart failure, coronary artery disease, and scarring and inflammation of the heart, which are often life-threatening. Troponin and natriuretic peptides are biomarkers of myocardial damage and heart failure, with the potential to detect acute cancer therapy-related cardiotoxicity during [69] and after treatment completion [70–75]. However, current evidence for the use of troponin or natriuretic peptides for clinical management of cancer survivors with long-term follow-up is limited and remains controversial. Some novel cardiac biomarkers (e.g., micro-RNAs) show promise in predicting cardiac dysfunction in cancer patients after treatment [76–79], although this remains largely uncharted territory.

Biomarkers of bone formation, turnover and resorption, circulating thyroid, sex hormones, and other routine blood measures (e.g., lipid panels, fasting glucose) have the potential to monitor musculoskeletal, endocrine, and metabolic effects of cancer treatments [80]. While these biomarkers are often utilized in the clinical setting for screening, there are no established or consistent guidelines for the use of these biomarkers for detecting or managing survivorship care. Little work has been done on predicting these outcomes to allow for early intervention for prevention and improved outcomes.

The application of biomarkers for use in follow-up, long-term surveillance, and management of late effects in cancer survivors has great potential. Given there are more cancer survivors in the population than ever before, there is a need for research in this space to ultimately improve outcomes and quality of life for survivors.

Methodological Issue

It is worth noting that the emerging examples provided in previous sections can often be applied to multiple phases in the cancer continuum. For example, in addition to early detection, the liquid biopsy approach can be applied to monitor treatment response or minimum residual disease [29–32]. Regardless of the type of biomarker and the phase of the cancer continuum, the processes involved in the discovery and validation of biomarkers broadly consist of five steps: biomarker discovery, biomarker validity, pre-clinical validation, clinical utility assessment, and regulatory approval. These move from small laboratory-based studies to large prospective population-based studies. The reality of this process is that it presents several methodological challenges at the level of the individual, biospecimen, assay, and analysis.

At the individual level, one of the most significant challenges is in minimizing the potential for selection bias in who participates in discovery and validation studies. The ultimate goal is to have biomarkers that are validated in population-based studies reflecting the diversity and variability (e.g., age, sex, and racial distribution) of the target population (i.e., to whom the biomarker is expected to apply), allowing for accurate clinical application. This requires reproducibility of results in multiple populations for external validation and ensuring transferability across ethnic groups, sex, age, and health systems (e.g., high versus low resource settings). This will be determined by the target population (e.g., average-risk individuals in the population for risk stratification or screening for high-risk individuals),

target outcome (e.g., any cancer, ER-positive breast cancer, metastasis, etc.), and the goal of the biomarker (e.g., risk assessment, early detection, and disease progression). In turn, this will inform the validation metric of interest.

At the level of the sample and experiment, it is necessary to determine the appropriate biospecimen for assessment (e.g., blood, saliva, and tumor), the timing of collection (e.g., fasting blood sample, time to diagnosis, pre- or posttreatment), if a single measurement is sufficiently representative, the impact of batch effects, the minimal detectable limits (i.e., assay sensitivity), as well as ease of clinical interpretation of results [81]. In practice, this involves the testing and re-testing of samples to ensure both inter- and intra-batch reproducibility. To minimize the introduction of bias, samples from individuals with and without disease should be randomized within the same batch, and assays conducted in a blinded fashion. Specifically for cfDNA, guidelines for sample pre-analytical conditions, such as plasma preparation, cfDNA extraction, and storage condition, have been proposed to optimize the assay accuracy and reproducibility within and between labs [82]. As we move to a multi-omics approach for biomarker discovery and validation, standardization of sample collection, processing, and analytic pipeline will be necessary to ensure reproducibility and transferability across populations and health care settings.

In addition to the considerations for patient and sample selection, one of the most important aspects to ensure the robustness of the biomarker, particularly when the discovery was based on a large number of analytes (e.g., metabolomics and proteomics), is the issue of multiple comparisons. In recent omics analyses, the importance of minimizing false positives has been well recognized with a regular adaptation of variations of false discovery rates or the more conservative family-wise error rate, depending on the research hypothesis [83]. Alternative approaches, such as the Bayes factor, which are less susceptible to the impact of multiple comparisons, or other Bayesian approaches that can incorporate biological prior, have not been widely adapted into the analytical pipeline but would be a welcomed change. Recently, Ou et al. outline the statistical considerations to move biomarkers from bench to bedside [84]. Dimensionality reduction techniques, such as principal component analysis, *t*-distributed stochastic neighbor embedding, or uniform manifold approximation and projection, can also be used to mitigate the issues of multiple comparisons, in particular when the main interest is to identify a pattern rather than a few analytes [85].

It is now considered a state-of-the-art practice to build classification algorithms with a cross-validation approach, which helps to reduce (although not eliminate) model over-optimism. However, the importance of model calibration is often overlooked in most disciplines. Classification accuracy is only reliable when it is based on a well-calibrated model,

which is typically assessed by comparing the predictive probability against the observed probability [86]. To minimize model overfitting, both the discriminatory ability and the calibration of the model should be assessed in the hold-out testing set and, when possible, in the external validation set. Table 1 outlines the key methodological considerations for biomarker research and application.

Guidelines for Biomarker Development and Validation

The early detection research network (EDRN), established by the National Institute of Health, is perhaps one of the most prominent organizations providing data, funding, software, and guidelines (PRoBE, prospective sample collection–retrospective-blinded evaluation) for biomarker research and reporting [81]. Specifically for liquid biopsy, several initiatives such as the European Liquid Biopsy Society (ELBS), Blood Profiling Atlas in Cancer (BloodPAC), and CancerID were launched with the goal to define best practices for liquid biopsy assay development [31]. In addition, several guidelines related to risk prediction tools have been developed to provide a common standard for reporting results from risk models, which are relevant for biomarker applications of all purposes. For example, the TRIPOD (transparent reporting of a multivariable prediction model of individual prognosis and diagnosis) statement has been widely disseminated [87] and was recently updated to incorporate models developed based on an artificial intelligence approach (TRIPOD-AI) [88]. Specifically for genetics, the polygenic risk score reporting standard (PRS-RS) was recently reported as an update of the genetic risk prediction studies (GRIPS) statement [89].

Implications and Future Perspectives

As medical imaging modality advances, it is increasingly recognized that medical images (e.g., computed tomography (CT), magnetic resonance imaging) contain a wealth of information that can be analyzed using deep learning approaches, such as convolution neural networks (CNN) [90]. For example, using CNNs to differentiate benign versus malignant pulmonary nodules detected on low-dose CT scans is an active area of research [90]. Furthermore, imaging features and different layers of molecular characterizations (genetics, epigenetics, transcription, protein expression, etc.) are interconnected within a complex biological network. Multi-omics studies can provide a holistic picture of the disease pathways and help to uncover the key elements for precision medicine. While gaining an increasing level of interest in recent years, further development of robust analytic pipelines for the integration of data from multiple

Table 1 Challenges and possible mitigation strategy in biomarker application

Considerations	Challenges	Possible mitigation
Individual		
	■ Validity of the target population	■ Ensure representativeness of the study population ■ Minimize selection bias ■ Appropriate comparison group ■ Independent validation set
	■ Generalizability and transferability of the results	■ Include diverse populations in the study sample, including groups with different ancestral backgrounds
Biospecimen		
	■ Tissue specificity	■ Select the optimal biospecimen for the intended biomarker taking into account of ease of collection that is fit for purpose for implementation
	■ Timing of sample collection	■ Define optimal sample collection time point out front, instead of convenience sampling
	■ Pre-analytic variability	■ Establish and share best practices and standard operating procedures for sample collection, preparation, and storage
Target molecule		
	■ Detectability	■ Characterize the baseline distribution and the distribution in the target population ■ Assess the abundance in the target biospecimen ■ Evaluate the dynamic nature of the molecule (clearance, half-life, etc.) ■ Consideration of differences by biological sex, age, race/ethnicity
	■ Dynamic nature	■ Consider sequential and repeat measurements to assess intra- and inter-individual variability
	■ Interpretability	■ Biological role in carcinogenesis
Assay experiment		
	■ Reliability and accuracy	■ Assess reproducibility and concordance ■ Implement QC samples and procedures in each batch ■ Assess minimal detectable limit
	■ Batch effect	■ Randomize cases and controls in the experimental process in each batch ■ Blinded experiment
	■ Interpretability	■ Laboratory and clinical guidelines to ensure consistent and equitable outcomes (e.g., defined thresholds for action)
Data analytics		
	■ Multiple comparison, false positives	■ Address multiple comparisons in the analytical pipeline via stringent threshold, dimensionality reduction, or Bayesian approach ■ External validation/replication
	■ Statistical power	■ Ensure adequate sample size ■ Avoid ad-hoc analyses ■ Consortium collaboration
	■ Model over-fitting	■ Correct for over-optimism with bootstrapping or cross-validation ■ Performance to be evaluated in the hold-out test
	■ Model validity	■ Model calibration as a required component of the report ■ External validation
	■ Reproducibility	■ Provide open-source codes and parameter setting
	■ Model implementation	■ Assess added values of the biomarker ■ Cost-effectiveness analysis and benefit-harm ratio ■ Increase model interpretability to facilitate clinical translation

omics platforms is needed to ensure reproducibility and optimize discovery [91, 92].

Careful evaluation of the cost-effectiveness and benefit-to-harm ratio are needed before any biomarker can be brought into action. Take cancer screening, for example, biomarkers will need to be evaluated in the context of current screening guidelines and assess whether the addition of biomarkers can increase the screening efficiency, either through increasing detection rate, reducing over-diagnosis, unnecessary follow-up, or further reduction of morality rate. In addition, it is important to evaluate whether the introduction of a biomarker impacts the

screening uptake or modifies the health-related behaviors before mass introduction [93, 94]. The cost-effectiveness of the biomarker will depend on its performance and the necessary resources and infrastructure needed to support implementation. For example, a recent microsimulation analysis for lung cancer low-dose CT screening showed that a biomarker to differentiate benignly versus malignant pulmonary nodules with 90% specificity and 90% sensitivity can remain cost effective at the cost of $500. This cost-effectiveness threshold changes as the balance between specificity and sensitivity shifts [95].

Beyond these considerations, there are multiple factors that will affect the success of implementation. These include the turnaround time of the results, proper result interpretations and risk communications, resources needed to support follow-up and subsequent testing such as imaging or biopsy, and the training of healthcare professionals to support implementation. Perhaps one of the most pressing challenges are social–ethical–legal considerations related to equitable access and portability across ancestries [96].

Conclusions

While there are a healthy plethora of emerging cancer biomarkers, only a small fraction will survive the "death valley" of the translational process, either because the research laboratories do not have the resources to continue supporting the ever-growing need for large and prospective sample series, or lack of industry to develop the biomarker panel to meet the necessary clinical standard. This highlights the need to promote the partnership between academia and industry to accelerate the translation of new biomarkers into the clinic while properly safeguarding the independence of discovery and validation work prior to commercialization, ensuring rigor. Ideally, a government-supported framework would help to accelerate these processes and maximize resource efficiency, particularly research investment at the discovery and validation stage of the biomarker development. This is the field that is driving the progress of precision medicine, and its success can only be achieved with collaborative team science across disciplines, from biomedical discovery to data science and clinical implementation.

Declarations

Conflict of Interest The authors declare no competing interests.

References

1. National Research Council (US) Committee on A Framework for Developing a New Taxonomy of Disease. Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease. Washington (DC): National Academies Press (US); 2011.
2. Lethimonnier F, Levy Y. Genomic medicine France 2025. Ann Oncol. 2018;29:783–4.
3. Turnbull C, Scott RH, Thomas E, Jones L, Murugaesu N, Pretty FB, Halai D, Baple E, Craig C, Hamblin A, Henderson S, Patch C, O'Neill A, Devereau A, Smith K, Martin AR, Sosinsky A, McDonagh EM, Sultana R, Mueller M, Smedley D, Toms A, Dinh L, Fowler T, Bale M, Hubbard T, Rendon A, Hill S, Caulfield MJ, Genomes P. The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. BMJ. 2018;361:k1687.
4. Beane J, Campbell JD, Lel J, Vick J, Spira A. Genomic approaches to accelerate cancer interception. Lancet Oncol. 2017;18:e494–502.
5. Penney, KL., Michailidou, K., Carere, DA., Zhang, C., Pierce, B., Lindstrom, S. Kraft, P. in *Schottenfeld and Fraumeni cancer epidemiology and prevention, 4th Edition*. (eds M. J. Thun *et al.*) Ch. 5, 53–76 (Oxford University Press, 2018).
6. Chatterjee N, Shi J, García-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nat Rev Genet. 2016;17(7):392–406
7. Plym, A., Penney, KL., Kalia, S., Kraft, P., Conti, DV., Haiman, C., Mucci, LA. Kibel, AS. Evaluation of a multiethnic polygenic risk score model for prostate cancer. *J Natl Cancer Inst* (2021).
8. Hung, RJ., Warkentin, MT., Brhane, Y., Chatterjee, N., Christiani, DC., Landi, MT., Caporaso, NE., Liu, G., Johansson, M., Albanes, D., Marchand, LL., Tardon, A., Rennert, G., Bojesen, SE., Chen, C., Field, JK., Kiemeney, LA., Lazarus, P., Zienolddiny, S., Lam, S., Andrew, AS., Arnold, SM., Aldrich, MC., Bickeboller, H., Risch, A., Schabath, MB., McKay, JD., Brennan, P. Amos, CI. Assessing lung cancer absolute risk trajectory based on a polygenic risk model. *Cancer Res* **81**, 1607–1615 (2021).
9. Mars, N., Widen, E., Kerminen, S., Meretoja, T., Pirinen, M., Della Briotta Parolo, P., Palta, P., FinnGen, Palotie, A., Kaprio, J., Joensuu, H., Daly, M. Ripatti, S. The role of polygenic risk and susceptibility genes in breast cancer over the course of life. *Nat Commun* **11**, 6383 (2020).
10. Pal Choudhury, P., Brook, MN., Hurson, AN., Lee, A., Mulder, CV., Coulson, P., Schoemaker, MJ., Jones, ME., Swerdlow, AJ., Chatterjee, N., Antoniou, AC. Garcia-Closas, M. Comparative validation of the BOADICEA and Tyrer-Cuzick breast cancer risk models incorporating classical risk factors and polygenic risk in a population-based prospective cohort of women of European ancestry. *Breast Cancer Res* **23**, 22 (2021).
11. NCCN guidelines for genetics/familial high-risk assessment: breast, ovarian and pancreatic V.1.2022. National Comprehensive Cancer Network. (2021).
12. Polygenic Risk Score Task Force of the International Common Disease, A. Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. *Nature medicine* **27**, 1876–1884 (2021).
13. Chang CM, Edwards SH, Arab A, Del Valle-Pinero AY, Yang L, Hatsukami DK. Biomarkers of tobacco exposure: summary of an FDA-sponsored public workshop. Cancer Epidemiol Biomarkers Prev. 2017;26:291–302.
14. Stram DO, Park SL, Haiman CA, Murphy SE, Patel Y, Hecht SSL, Marchand L. Racial/ethnic differences in lung cancer incidence in the multiethnic cohort study: an update. J Natl Cancer Inst. 2019;111:811–9.

15. Kaur G, Begum R, Thota S, Batra S. A systematic review of smoking-related epigenetic alterations. Arch Toxicol. 2019;93:2715–40.

16. Park SL, Patel YM, Loo LWM, Mullen DJ, Offringa IA, Maunakea A, Stram DO, Siegmund K, Murphy SE, Tiirikainen M. Le, Marchand L. Association of internal smoking dose with blood DNA methylation in three racial/ethnic populations. Clin Epigenetics. 2018;10:110.

17. Prince C, Hammerton G, Taylor AE, Anderson EL, Timpson NJ, Davey Smith G, Munafo G, Relton MR, Richmond CL. Investigating the impact of cigarette smoking behaviours on DNA methylation patterns in adolescence. Hum Mol Genet. 2019;28:155–65.

18. Fasanelli F, Baglietto L, Ponzi E, Guida F, Campanella G, Johansson M, Grankvist K, Johansson M, Assumma MB, Naccarati A, Chadeau-Hyam M, Ala U, Faltus C, Kaaks R, Risch A, De Stavola B, Hodge A, Giles GG, Southey MC, Relton CL, Haycock PC, Lund E, Polidoro S, Sandanger TM, Severi G, Vineis P. Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. Nat Commun. 2015;6:10192.

19. Bojesen SE, Timpson N, Relton C, Davey Smith G, Nordestgaard BG. AHRR (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. Thorax. 2017;72:646–53.

20. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Huang MN, Tiang NG, Wu Y, Boot A, Covington KR, Gordenin DA, Bergstrom EN, Islam SMA, Lopez-Bigas N, Klimczak LJ, McPherson JR, Morganella S, Sabarinathan R, Wheeler DA, Mustonen V, Group, PMSW. Getz G Rozen, SG. Stratton, MR. Consortium, P. The repertoire of mutational signatures in human cancer. Nature. 2020;578:94–101.

21. Franceschi, S., El-serag, HB., Forman, D., Newton, R. Plummer, M. in *Schottenfeld and Fraumeni cancer epidemiology and prevention* (eds M. Thun *et al.*) Ch. 24, 433–460 (Oxford University Press, 2018).

22. Kreimer AR, Ferreiro-Iglesias A, Nygard M, Bender N, Schroeder L, Hildesheim A, Robbins HA, Pawlita M, Langseth H, Schlecht NF, Tinker LF, Agalliu I, Smoller SW, Ness-Jensen E, Hveem K, D'Souza G, Visvanathan K, May B, Ursin G, Weiderpass E, Giles GG, Milne RL, Cai Q, Blot WJ, Zheng W, Weinstein SJ, Albanes D, Brenner N, Hoffman-Bolton J, Kaaks R, Barricarte A, Tjonneland A, Sacerdote C, Trichopoulou A, Vermeulen RCH, Huang WY, Freedman ND, Brennan P, Waterboer T, Johansson M. Timing of HPV16-E6 antibody seroconversion before OPSCC: findings from the HPVC3 consortium. Ann Oncol. 2019;30:1335–43.

23. Kreimer AR, Shiels MS, Fakhry C, Johansson M, Pawlita M, Brennan P, Hildesheim A, Waterboer T. Screening for human papillomavirus-driven oropharyngeal cancer: Considerations for feasibility and strategies for research. Cancer. 2018;124:1859–66.

24. Leung E, Han K, Zou J, Zhao Z, Zheng Y, Wang TT, Rostami A, Siu LL, Pugh TJ, Bratman SV. HPV sequencing facilitates ultrasensitive detection of HPV circulating tumor DNA. Clin Cancer Res. 2021;27:5857–68.

25. Schmidt DR, Patel R, Kirsch DG, Lewis CA, Vander Heiden MG, Locasale JW. Metabolomics in cancer research and emerging applications in clinical oncology. CA Cancer J Clin. 2021;71:333–58.

26. Smith RA, Andrews KS, Brooks D, Fedewa SA, Manassaram-Baptiste D, Saslow D, Wender RC. Cancer screening in the United States, 2019: a review of current American Cancer Society guidelines and current issues in cancer screening. CA Cancer J Clin. 2019;69:184–210.

27. Tammemagi MC, Ruparel M, Tremblay A, Myers R, Mayo J, Yee J, Atkar-Khattra S, Yuan R, Cressman S, English J, Bedard E, MacEachern P, Burrowes P, Quaife SL, Marshall H, Yang I, Bowman R, Passmore L, McWilliams A, Brims F, Lim KP, Mo L, Melsom S, Saffar B, Teh M, Sheehan R, Kuok Y, Manser R, Irving L, Steinfort D, McCusker M, Pascoe D, Fogarty P, Stone E, Lam DCL, Ng MY, Vardhanabhuti V, Berg CD, Hung RJ, Janes SM, Fong K, Lam S. USPSTF2013 versus PLCOm2012 lung cancer screening eligibility criteria (International Lung Screening Trial): interim analysis of a prospective cohort study. Lancet Oncol. 2022;23:138–48.

28. Church TR, Wandell M, Lofton-Day C, Mongin SJ, Burger M, Payne SR, Castanos-Velez E, Blumenstein BA, Rosch T, Osborn N, Snover D, Day RW, Ransohoff DF, Presept Clinical Study Steering Committee, I. & StudyT. Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. Gut. 2014;63:317–25.

29. Heitzer E, Haque IS, Roberts CES, Speicher MR. Nat Rev Genet. 2019;20:71–88.

30. Corcoran RB, Chabner BA. Application of cell-free DNA analysis to cancer treatment. N Engl J Med. 2018;379:1754–65.

31. Ignatiadis M, Sledge GW, Jeffrey SS. Liquid biopsy enters the clinic - implementation issues and future challenges. Nat Rev Clin Oncol. 2021;18:297–312.

32. Keller L, Pantel K. Unravelling tumour heterogeneity by single-cell profiling of circulating tumour cells. Nat Rev Cancer. 2019;19:553–67.

33. Cohen JD, Li L, Wang Y, Thoburn C, Afsari B, Danilova L, Douville C, Javed AA, Wong F, Mattox A, Hruban RH, Wolfgang CL, Goggins MG, Dal Molin M, Wang TL, Roden R, Klein AP, Ptak J, Dobbyn L, Schaefer J, Silliman N, Popoli M, Vogelstein JT, Browne JD, Schoen RE, Brand RE, Tie J, Gibbs P, Wong HL, Mansfield AS, Jen J, Hanash SM, Falconi M, Allen PJ, Zhou S, Bettegowda C, Diaz LA Jr, Tomasetti C, Kinzler KW, Vogelstein B, Lennon AM, Papadopoulos N. Detection and localization of surgically resectable cancers with a multi-analyte blood test. Science. 2018;359:926–30.

34. Bates SE. Epigenetic therapies for cancer. N Engl J Med. 2020;383:650–63.

35. Moss J, Magenheim J, Neiman D, Zemmour H, Loyfer N, Korach A, Samet Y, Maoz M, Druid H, Arner P, Fu KY, Kiss E, Spalding KL, Landesberg G, Zick A, Grinshpun A, Shapiro AMJ, Grompe M, Wittenberg AD, Glaser B, Shemer R, Kaplan T, Dor Y. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. Nat Commun. 2018;9:5068.

36. Shen SY, Singhania R, Fehringer G, Chakravarthy A, Roehrl MHA, Chadwick D, Zuzarte PC, Borgida A, Wang TT, Li T, Kis O, Zhao Z, Spreafico A, Medina TDS, Wang Y, Roulois D, Ettayebi I, Chen Z, Chow S, Murphy T, Arruda A, O'Kane GM, Liu J, Mansour M, McPherson JD, O'Brien C, Leighl N, Bedard PL, Fleshner N, Liu G, Minden MD, Gallinger S, Goldenberg A, Pugh TJ, Hoffman MM, Bratman SV, Hung RJ, De Carvalho DD. Sensitive tumour detection and classification using plasma cell-free DNA methylomes. Nature. 2018;563:579–83.

37. Lo, Y. M. D., Han, D. S. C., Jiang, P. & Chiu, R. W. K. Epigenetics, fragmentomics, and topology of cell-free DNA in liquid biopsies. *Science* 372 (2021).

38. Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV, Consortium C. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. Ann Oncol. 2020;31:745–59.

39. Klein EA, Richards D, Cohn A, Tummala M, Lapham R, Cosgrove D, Chung G, Clement J, Gao J, Hunkapiller N, Jamshidi A, Kurtzman KN, Seiden MV, Swanton C, Liu MC. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. Ann Oncol. 2021;32:1167–77.

40. Hanna GB, Boshier PR, Markar SR, Romano A. Accuracy and methodologic challenges of volatile organic compound-based

exhaled breath tests for cancer diagnosis: a systematic review and meta-analysis. JAMA Oncol. 2019;5:e182815.

41. Ibrahim W, Carr L, Cordell R, Wilde MJ, Salman D, Monks PS, Thomas P, Brightling CE, Siddiqui S, Greening NJ. Breathomics for the clinician: the use of volatile organic compounds in respiratory diseases. Thorax. 2021;76:514–21.

42. Wang L, Li J, Xiong X, Hao T, Zhang C, Gao Z, Zhong L, Zhao Y. Volatile organic compounds as a potential screening tool for neoplasm of the digestive system: a meta-analysis. Sci Rep. 2021;11:23716.

43. Zhong L, Li Y, Xiong L, Wang W, Wu M, Yuan T, Yang W, Tian C, Miao Z, Wang T, Yang S. Small molecules in targeted cancer therapy: advances, challenges, and future perspectives. Signal Transduct Target Ther. 2021;6:201.

44. Mehta D, Uber R, Ingle T, Li C, Liu Z, Thakkar S, Ning B, Wu L, Yang J, Harris S, Zhou G, Xu J, Tong W, Lesko L, Fang H. Study of pharmacogenomic information in FDA-approved drug labeling to facilitate application of precision medicine. Drug Discov Today. 2020;25:813–20.

45. Administration, U. F. a. D. Vol. 2022 (US Food and Drug Administration, 2021).

46. Lauschke VM, Zhou Y, Ingelman-Sundberg M. Novel genetic and epigenetic factors of importance for inter-individual differences in drug disposition, response and toxicity. Pharmacol Ther. 2019;197:122–52.

47. Charkhchi, P., Cybulski, C., Gronwald, J., Wong, F. O., Narod, S. A. & Akbari, M. R. CA125 and ovarian cancer: a comprehensive review. *Cancers (Basel)* **12** (2020).

48. Dagnino S, Bodinier B, Guida F, Smith-Byrne K, Petrovic D, Whitaker MD, Haugdahl Nost T, Agnoli C, Palli D, Sacerdote C, Panico S, Tumino R, Schulze MB, Johansson M, Keski-Rahkonen P, Scalbert A, Vineis P, Johansson M, Sandanger TM, Vermeulen RCH, Chadeau-Hyam M. Prospective Identification of elevated circulating CDCP1 in patients years before onset of lung cancer. Cancer Res. 2021;81:3738–48.

49. Nedjadi T, Benabdelkamal H, Albarakati N, Masood A, Al-Sayyad A, Alfadda AA, Alanazi IO, Al-Ammari A, Al-Maghrabi J. Circulating proteomic signature for detection of biomarkers in bladder cancer patients. Sci Rep. 2020;10:10999.

50. Loriot Y, Marabelle A, Guegan JP, Danlos FX, Besse B, Chaput N, Massard C, Planchard D, Robert C, Even C, Khettab M, Tselikas L, Friboulet L, Andre F, Nafia I, Le Loarer F, Soria JC, Bessede A, Italiano A. Plasma proteomics identifies leukemia inhibitory factor (LIF) as a novel predictive biomarker of immune-checkpoint blockade resistance. Ann Oncol. 2021;32:1381–90.

51. Rittmann MC, Hussung S, Braun LM, Klar RFU, Biesel EA, Fichtner-Feigl S, Fritsch R, Wittel UA, Ruess DA. Plasma biomarkers for prediction of early tumor recurrence after resection of pancreatic ductal adenocarcinoma. Sci Rep. 2021;11:7499.

52. Lam MPY, Ge Y. Harnessing the power of proteomics to assess drug safety and guide clinical trials. Circulation. 2018;137:1011–4.

53. Folkersen L, Gustafsson S, Wang Q, Hansen DH, Hedman AK, Schork A, Page K, Zhernakova DV, Wu Y, Peters J, Eriksson N, Bergen SE, Boutin TS, Bretherick AD, Enroth S, Kalnapenkis A, Gadin JR, Suur BE, Chen Y, Matic L, Gale JD, Lee J, Zhang W, Quazi A, Ala-Korpela M, Choi SH, Claringbould A, Danesh J, Davey Smith G, de Masi F, Elmstahl S, Engstrom G, Fauman E, Fernandez C, Franke L, Franks PW, Giedraitis V, Haley C, Hamsten A, Ingason A, Johansson A, Joshi PK, Lind L, Lindgren CM, Lubitz S, Palmer T, Macdonald-Dunlop E, Magnusson M, Melander O, Michaelsson K, Morris AP, Magi R, Nagle MW, Nilsson PM, Nilsson J, Orho-Melander M, Polasek O, Prins B, Palsson E, Qi T, Sjogren M, Sundstrom J, Surendran P, Vosa U, Werge T, Wernersson R, Westra HJ, Yang J, Zhernakova A, Arnlov J, Fu J, Smith JG, Esko T, Hayward C, Gyllensten U, Landen M, Siegbahn A, Wilson JF, Wallentin L, Butterworth AS, Holmes MV, Ingelsson E, Malarstig A. Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. Nat Metab. 2020;2:1135–48.

54. Jiang Y, Sun A, Zhao Y, Ying W, Sun H, Yang X, Xing B, Sun W, Ren L, Hu B, Li C, Zhang L, Qin G, Zhang M, Chen N, Zhang M, Huang Y, Zhou J, Zhao Y, Liu M, Zhu X, Qiu Y, Sun Y, Huang C, Yan M, Wang M, Liu W, Tian F, Xu H, Zhou J, Wu Z, Shi T, Zhu W, Qin J, Xie L, Fan J, Qian X, He F. Chinese Human Proteome Project, C. Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. Nature. 2019;567:257–61.

55. Eckert MA, Coscia F, Chryplewicz A, Chang JW, Hernandez KM, Pan S, Tienda SM, Nahotko DA, Li G, Blazenovic I, Lastra RR, Curtis M, Yamada SD, Perets R, McGregor SM, Andrade J, Fiehn O, Moellering RE, Mann M, Lengyel E. Proteomics reveals NNMT as a master metabolic regulator of cancer-associated fibroblasts. Nature. 2019;569:723–8.

56. Kwon YW, Jo HS, Bae S, Seo Y, Song P, Song M, Song P, Yoon JH. Application of proteomics in cancer: recent trends and approaches for biomarkers discovery. Front Med (Lausanne). 2021;8:747333.

57. Rohloff JC, Gelinas AD, Jarvis TC, Ochsner UA, Schneider DJ, Gold L, Janjic N. Nucleic acid ligands with protein-like side chains: modified aptamers and their use as diagnostic and therapeutic agents. Mol Ther Nucleic Acids. 2014;3:e201.

58. Assarsson E, Lundberg M, Holmquist G, Bjorkesten J, Thorsen SB, Ekman D, Eriksson A, Rennel Dickens E, Ohlsson S, Edfeldt G, Andersson AC, Lindstedt P, Stenvang J, Gullberg M, Fredriksson S. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. PLoS ONE. 2014;9:e95192.

59. Mertins P, Mani DR, Ruggles KV, Gillette MA, Clauser KR, Wang P, Wang X, Qiao JW, Cao S, Petralia F, Kawaler E, Mundt F, Krug K, Tu Z, Lei JT, Gatza ML, Wilkerson M, Perou CM, Yellapantula V, Huang KL, Lin C, McLellan MD, Yan P, Davies SR, Townsend RR, Skates SJ, Wang J, Zhang B, Kinsinger CR, Mesri M, Rodriguez H, Ding L, Paulovich AG, Fenyo D, Ellis MJ, Carr SA, Nci C. Proteogenomics connects somatic mutations to signalling in breast cancer. Nature. 2016;534:55–62.

60. Vasaikar S, Huang C, Wang X, Petyuk VA, Savage SR, Wen B, Dou Y, Zhang Y, Shi Z, Arshad OA, Gritsenko MA, Zimmerman LJ, McDermott JE, Clauss TR, Moore RJ, Zhao R, Monroe ME, Wang YT, Chambers MC, Slebos RJC, Lau KS, Mo Q, Ding L, Ellis M, Thiagarajan M, Kinsinger CR, Rodriguez H, Smith RD, Rodland KD, Liebler DC, Liu T, Zhang B. Clinical proteomic tumor analysis, C proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. Cell. 2019;177:1035-1049 e1019.

61. Huang KL, Li S, Mertins P, Cao S, Gunawardena HP, Ruggles KV, Mani DR, Clauser KR, Tanioka M, Usary J, Kavuri SM, Xie L, Yoon C, Qiao JW, Wrobel J, Wyczalkowski MA, Erdmann-Gilmore P, Snider JE, Hoog J, Singh P, Niu B, Guo Z, Sun SQ, Sanati S, Kawaler E, Wang X, Scott A, Ye K, McLellan MD, Wendl MC, Malovannaya A, Held JM, Gillette MA, Fenyo D, Kinsinger CR, Mesri M, Rodriguez H, Davies SR, Perou CM, Ma C, Reid Townsend R, Chen X, Carr SA, Ellis MJ, Ding L. Proteogenomic integration reveals therapeutic targets in breast cancer xenografts. Nat Commun. 2017;8:14864.

62. Matsunuma R, Chan DW, Kim BJ, Singh P, Han A, Saltzman AB, Cheng C, Lei JT, Wang J, Roberto da Silva L, Sahin E, Leng M, Fan C, Perou CM, Malovannaya A, Ellis MJ. DPYSL3 modulates mitosis, migration, and epithelial-to-mesenchymal transition in claudin-low breast cancer. Proc Natl Acad Sci U S A. 2018;115:E11978–87.

63. Ferkingstad E, Sulem P, Atlason BA, Sveinbjornsson G, Magnusson MI, Styrmisdottir EL, Gunnarsdottir K, Helgason A, Oddsson

A, Halldorsson BV, Jensson BO, Zink F, Halldorsson GH, Masson G, Arnadottir GA, Katrinardottir H, Juliusson K, Magnusson MK, Magnusson OT, Fridriksdottir R, Saevarsdottir S, Gudjonsson SA, Stacey SN, Rognvaldsson S, Eiriksdottir T, Olafsdottir TA, Steinthorsdottir V, Tragante V, Ulfarsson MO, Stefansson H, Jonsdottir I, Holm H, Rafnar T, Melsted P, Saemundsdottir J, Norddahl GL, Lund SH, Gudbjartsson DF, Thorsteinsdottir U, Stefansson K. Large-scale integration of the plasma proteome with genetics and disease. Nat Genet. 2021;53:1712–21.

64. Pietzner M, Wheeler E, Carrasco-Zanini J, Cortes A, Koprulu M, Worheide MA, Oerton E, Cook J, Stewart ID, Kerrison ND, Luan J, Raffler J, Arnold M, Arlt W, O'Rahilly S, Kastenmuller G, Gamazon ER, Hingorani AD, Scott RA, Wareham NJ, Langenberg C. Mapping the proteo-genomic convergence of human diseases. Science. 2021;374:1541.

65. Dulaney C, Wallace AS, Everett AS, Dover L, McDonald A, Kropp L. Defining health across the cancer continuum. Cureus. 2017;9:e1029.

66. Aziz NM, Rowland JH. Trends and advances in cancer survivorship research: challenge and opportunity. Semin Radiat Oncol. 2003;13:248–66.

67. Ganz PA. Late effects of cancer and its treatment. Semin Oncol Nurs. 2001;17:241–8.

68. Landier W, Armenian S, Bhatia S. Late effects of childhood cancer and its treatment. Pediatr Clin North Am. 2015;62:275–300.

69. Michel L, Rassaf T, Totzeck M. Biomarkers for the detection of apparent and subclinical cancer therapy-related cardiotoxicity. J Thorac Dis. 2018;10:S4282–95.

70. Demissei BG, Hubbard RA, Zhang L, Smith AM, Sheline K, McDonald C, Narayan V, Domchek SM, DeMichele A, Shah P, Clark AS, Fox K, Matro J, Bradbury AR, Knollman H, Getz KD, Armenian SH, Januzzi JL, Tang WHW, Liu P, Ky B. Changes in cardiovascular biomarkers with breast cancer therapy and associations with cardiac dysfunction. J Am Heart Assoc. 2020;9:e014708.

71. Kitayama H, Kondo T, Sugiyama J, Kurimoto K, Nishino Y, Kawada M, Hirayama M, Tsuji Y. High-sensitive troponin T assay can predict anthracycline- and trastuzumab-induced cardiotoxicity in breast cancer patients. Breast Cancer. 2017;24:774–82.

72. Ky B, Putt M, Sawaya H, French B, Januzzi JL Jr, Sebag IA, Plana JC, Cohen V, Banchs J, Carver JR, Wiegers SE, Martin RP, Picard MH, Gerszten RE, Halpern EF, Passeri J, Kuter I, Scherrer-Crosbie M. Early increases in multiple biomarkers predict subsequent cardiotoxicity in patients with breast cancer treated with doxorubicin, taxanes, and trastuzumab. J Am Coll Cardiol. 2014;63:809–16.

73. Wang YD, Chen SX, Ren LQ. Serum B-type natriuretic peptide levels as a marker for anthracycline-induced cardiotoxicity. Oncol Lett. 2016;11:3483–92.

74. Lipshultz SE, Miller TL, Scully RE, Lipsitz SR, Rifai N, Silverman LB, Colan SD, Neuberg DS, Dahlberg SE, Henkel JM, Asselin BL, Athale UH, Clavell LA, Laverdiere C, Michon B, Schorin MA, Sallan SE. Changes in cardiac biomarkers during doxorubicin treatment of pediatric patients with high-risk acute lymphoblastic leukemia: associations with long-term echocardiographic outcomes. J Clin Oncol. 2012;30:1042–9.

75. Pavo N, Raderer M, Hulsmann M, Neuhold S, Adlbrecht C, Strunk G, Goliasch G, Gisslinger H, Steger GG, Hejna M, Kostler W, Zochbauer-Muller S, Marosi C, Kornek G, Auerbach L, Schneider S, Parschalk B, Scheithauer W, Pirker R, Drach J, Zielinski C, Pacher R. Cardiovascular biomarkers in patients with cancer and their association with all-cause mortality. Heart. 2015;101:1874–80.

76. Li Y, Zhong X, Cheng G, Zhao C, Zhang L, Hong Y, Wan Q, He R, Wang Z. Hs-CRP and all-cause, cardiovascular, and cancer mortality risk: a meta-analysis. Atherosclerosis. 2017;259:75–82.

77. Singh-Manoux A, Shipley MJ, Bell JA, Canonico M, Elbaz A, Kivimaki M. Association between inflammatory biomarkers and all-cause, cardiovascular and cancer-related mortality. CMAJ: Canadian Medical Assoc J =journal de l'Association medicale canadienne. 2017;189:E384–90.

78. Gai W, An J, Wang Z, Han X, Geng J, Liang Y, Guo Y. Research progress of biomarkers in early detection of chemotherapy-induced cardiotoxicity. Heart Fail Rev. 2021;26:1195–201.

79. Ananthan K, Lyon AR. The role of biomarkers in cardio-oncology. J Cardiovasc Transl Res. 2020;13:431–50.

80. Tonorezos ES, Hudson MM, Edgar AB, Kremer LC, Sklar CA, Wallace WH, Oeffinger KC. Screening and management of adverse endocrine outcomes in adult survivors of childhood and adolescent cancer. Lancet Diabetes Endocrinol. 2015;3:545–55.

81. Feng Z, Pepe MS. Adding rigor to biomarker evaluations-EDRN experience. Cancer Epidemiol Biomarkers Prev. 2020;29:2575–82.

82. Meddeb R, Pisareva E, Thierry AR. Guidelines for the preanalytical conditions for analyzing circulating cell-free DNA. Clin Chem. 2019;65:623–33.

83. Korthauer K, Kimes PK, Duvallet C, Reyes A, Subramanian A, Teng M, Shukla C, Alm EJ, Hicks SC. A practical guide to methods controlling false discoveries in computational biology. Genome Biol. 2019;20:118.

84. Ou FS, Michiels S, Shyr Y, Adjei AA, Oberg AL. Biomarker discovery and validation: statistical considerations. J Thorac Oncol. 2021;16:537–45.

85. Diaz-Papkovich A, Anderson-Trocme L, Gravel S. A review of UMAP in population genetics. J Hum Genet. 2021;66:85–91.

86. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Topic group "evaluating diagnostic, t. & prediction models" of the, S. i. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17:230.

87. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:g7594.

88. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, Logullo P, Beam AL, Peng L, Van Calster B, van Smeden M, Riley RD, Moons KG. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open. 2021;11:e048008.

89. Wand H, Lambert SA, Tamburro C, Iacocca MA, O'Sullivan JW, Sillari C, Kullo IJ, Rowley R, Dron JS, Brockman D, Venner E, McCarthy MI, Antoniou AC, Easton DF, Hegele RA, Khera AV, Chatterjee N, Kooperberg C, Edwards K, Vlessis K, Kinnear K, Danesh JN, Parkinson H, Ramos EM, Roberts MC, Ormond KE, Khoury MJ, Janssens A, Goddard KAB, Kraft P, MacArthur JAL, Inouye M, Wojcik GL. Improving reporting standards for polygenic scores in risk prediction studies. Nature. 2021;591:211–9.

90. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H. Artificial intelligence in radiology. Nat Rev Cancer. 2018;18:500–10.

91. Fang Z, Ma T, Tang G, Zhu L, Yan Q, Wang T, Celedon JC, Chen W, Tseng GC. Bayesian integrative model for multi-omics data with missingness. Bioinformatics. 2018;34:3801–8.

92. Chu J, Sun NA, Hu W, Chen X, Yi N, Shen Y. The application of Bayesian methods in cancer prognosis and prediction. Cancer Genomics Proteomics. 2022;19:1–11.

93. Brooks, J. D., Nabi, H. H., Andrulis, I. L., Antoniou, A. C., Chiquette, J., Despres, P., Devilee, P., Dorval, M., Droit, A., Easton, D. F., Eisen, A., Eloy, L., Fienberg, S., Goldgar, D., Hahnen, E., Joly, Y., Knoppers, B. M., Lofters, A., Masson, J. Y., Mittmann, N., Paquette, J. S., Pashayan, N., Schmutzler, R., Stockley, T., Tavtigian, S. V., Walker, M. J., Wolfson, M., Chiarelli, A. M. &

Simard, J. Personalized risk assessment for prevention and early detection of breast cancer: integration and implementation (PER-SPECTIVE I&I). *J Pers Med* **11** (2021).

94. Pashayan N, Antoniou AC, Ivanus U, Esserman LJ, Easton DF, French D, Sroczynski G, Hall P, Cuzick J, Evans DG, Simard J, Garcia-Closas M, Schmutzler R, Wegwarth O, Pharoah P, Moorthie S, De Montgolfier S, Baron C, Herceg Z, Turnbull C, Balleyguier C, Rossi PG, Wesseling J, Ritchie D, Tischkowitz M, Broeders M, Reisel D, Metspalu A, Callender T, de Koning H, Devilee P, Delaloge S, Schmidt MK, Widschwendter M. Personalized early detection and prevention of breast cancer: ENVISION consensus statement. Nat Rev Clin Oncol. 2020;17:687–705.

95. Toumazis I, Erdogan SA, Bastani M, Leung A, Plevritis SKA. Cost-effectiveness analysis of lung cancer screening with low-dose computed tomography and a diagnostic biomarker. JNCI Cancer Spectr. 2021;5:081.

96. Winkler EC, Knoppers BM. Ethical challenges of precision cancer medicine. Semin Cancer Biol. 2022;84:263–70.