OPINION

Precision diagnostics: moving towards protein biomarker signatures of clinical utility in cancer

Carl A. K. Borrebaeck

Abstract | Interest in precision diagnostics has been fuelled by the concept that early detection of cancer would benefit patients; that is, if detected early, more tumours should be resectable and treatment more efficacious. Serum contains massive amounts of potentially diagnostic information, and affinity proteomics has risen as an accurate approach to decipher this, to generate actionable information that should result in more precise and evidence-based options to manage cancer. To achieve this, we need to move from single to multiplex biomarkers, a so-called signature, that can provide significantly increased diagnostic accuracy. This Opinion article focuses on the progress being made in identifying protein biomarker signatures of clinical utility, using blood-based proteomics.

The World Health Organization (WHO) has proposed that millions of cancer patients could be saved from premature death if early detection and treatment were available[1]. Finding the tumour at an early stage when it is still localized and possibly even before clinical symptoms develop is one important application of specific biomarkers. This would certainly increase the number of radically resected tumours and therefore in many cases also increase overall survival[2]. Apart from early diagnosis, biomarkers could also provide physicians with actionable information leading to evidence-based selection of the optimal therapy (predictive biomarkers) and improved and more precise prognostication of disease progression (prognostic biomarkers). Ideally protein biomarkers should be found in a minimally invasive liquid biopsy, such as a simple blood sample. However, the question is whether blood contains enough information and whether we are even close to this scenario?

Tremendous efforts have been made over recent decades to find protein cancer biomarkers of clinical utility[3–6]. A striking discrepancy exists between these efforts and the number of US Food and Drug Administration (FDA) approved biomarkers, despite the fact that well over a thousand single candidate cancer biomarkers have been known for several years[7]. Currently, only a handful based on a liquid biopsy are FDA approved[8] and none of these is routinely used for early clinical diagnosis, although a few — for example, CA125 (also known as mucin 16) for ovarian cancer, prostate-specific antigen (PSA) for prostate cancer and CA19-9 for pancreatic cancer — have been proposed to be useful for longitudinal disease monitoring[8,9]. The reasons for this dismal progress are several: one major one is the lack of a clearly defined and relevant clinical question that the biomarkers should address, as merely discriminating between cancer patients and healthy individuals is certainly not enough. Furthermore, sample quality, technology platform, bioinformatic evaluation and reliance on single biomarkers are all contributing to the fact that cancer biomarkers are scarce[6,10].

The development of cancer biomarkers of clinical utility[11,12], that is, biomarkers that generate clinically useful information that could change the course of the disease for a patient, is a multiphase collaboration between various stakeholders, such as, academia, funding agencies, health-care providers, reimbursement organizations or authorities and commercial companies. Consequently, it is a lengthy and winding journey that needs to be carefully planned and to focus on which clinical questions the biomarkers should address. The clinical question being addressed will have direct implications for sample acquisition, including necessary clinical documentation, as well as pre-analytical variables that might act as confounding factors. However, biomarkers of clinical utility all relate to an ability to deliver accurate and improved diagnostic information to the clinicians, as discussed below. In this Opinion article I discuss the recent evidence now supporting the concept that protein biomarker signatures, as opposed to individual biomarkers, can provide the long-sought accuracy in cancer diagnostics. I will not discuss details of other biomarkers obtained from liquid biopsies, such as circulating tumour cells (CTCs) or circulating tumour DNA (ctDNA), although these could eventually be combined with protein biomarkers.

Multiparametric proteomics

Protein biomarker discovery has in many cases been technology driven instead of focusing on a specific clinical need. Proteomic technology platforms have developed rapidly during recent years, illustrated by substantially increased resolution, that is, depth of proteome coverage, and speed in, for example, mass spectrometry (MS) analysis and selected reaction monitoring or multiple reaction monitoring (SRM/MRM) for targeted proteomics[13]. Multiplexed enzyme-linked immunosorbent assay (ELISA) has also demonstrated clinical applicability (for review see REF. 14) and has paved the way for next-generation multiparametric diagnostics, that is, high-density antibody microarrays[15,16], which are discussed below. Such protein or antibody microarrays can theoretically display almost unlimited resolution of the most complex proteomes. However, in contrast to the more mature transcriptional profiling technologies, the proteome coverage of antibody microarrays

is limited by the number of available well-characterized antibodies[17]. For more in-depth details of antibody microarrays; see REFS 18–21 for reviews. Improved accuracy has been achieved through the use of antibody microarrays[22–24], reverse phase protein arrays[25,26] and bead-based arrays[27–29], demonstrating the feasibility of multiparametric proteomic analysis.

Analytical design considerations

Sample acquisition and documentation. Although novel technologies open up new avenues for clinical proteomics by introducing substantially improved proteome coverage, the quality of available samples may be problematic, as these have in most cases been collected for more traditional analytical set-ups, such as immunohistochemistry. The dream scenario for management of a complex disease such as cancer would be to obtain the necessary information to guide clinical decisions from a single blood sample. To achieve this, the quality of a blood sample is crucial, as it is quite easy to introduce confounding factors during sample collection[30]. Consequently, sample acquisition procedures should be strictly standardized. However, this has not been the case in many retrospective studies, in which standard operating procedures did not exist or were highly variable[31]. The introduction of such unknown pre-analytical variables can and will affect sample integrity and introduce bias, due to differences in parameters such as protein stability and concentration[31]. Furthermore, comprehensive information about patient demographics, such as gender, age, tumour stage and treatment schemes, as well as lifestyle factors, such as smoking and alcohol habits, is necessary to design the correct case–control studies, but these data are not always easily accessible.

The complexity of sample acquisition can be illustrated by the example of breast cancer, for which the patients may have received neoadjuvant therapy followed by various adjuvant treatments, such as chemotherapy, hormonal treatment, radiation, targeted therapy and even combinations of these. Consequently, each subgroup must be clearly defined and enough samples must be collected from each cohort, as differences in treatment modalities could have a major impact on the results of a proteomic analysis. Therefore, sample quality, as well as the available clinical documentation, has an essential role in clinical proteomics but can often be difficult to obtain in materials from biobanks, in which unrecorded variations can exist as confounding factors.

Finally, sample sizes usually increase as the biomarker moves from discovery to pre-validation to validation phase and finally into clinical testing. Therefore, power calculations should be carried out at all stages, unless there is no interventional intention, that is, no anticipated consequence for the patients, in a particular study[32,33].

Bioinformatic considerations. In the past, the bioinformatic analysis of a proteomic experiment has been an overlooked source of data overfitting[34], that is, the overinterpretation of the significance of differently expressed proteins in, for example, tumour versus normal tissue. This problem increases when handling multiplexed 'big' data, as a subset of the markers will always randomly correlate with whatever parameter is analysed, and this has to be adjusted for using statistical methods[35]. This problem frequently becomes an issue when subdividing the data set according to different features or subgroups, resulting in overinterpretation of the clinical significance and in the end, publication of potentially incorrect data. Diagnostic, prognostic and predictive problems are suitable for self-learning algorithms, such as support vector machines and artificial neural networks, which quickly generate unbiased results from large data sets (BOX 1). However, when defining larger biomarker signatures these approaches are particularly susceptible to overfitting, and, as a minimum, the leave-one-out cross-validation procedures should always be applied. A stricter and more desirable approach to minimize model instability for biomarker signatures is to first generate a classifier, using a training set of samples. This classifier is then tested in several truly independent test sets, which minimizes the risk of model instability and data overfitting.

As argued below, it is evident that biomarker signatures could be the future for precision diagnostics of complex diseases such as cancer, as it is the combination of biomarkers that contains valuable

Box 1 | **Bioinformatic principles and terms discussed in this article**

Leave-one-out cross-validation
Cross-validation is a validation technique for assessing how the results of a statistical analysis will apply to an independent data set. A model is given a data set of known data on which training is performed (training set) and a data set of unknown data against which the model is tested (test set). When performing leave-one-out, the training set consists of all data points except one, which is then used as a test set.

Support vector machine (SVM)
SVM is a supervised learning model in which machine-learning algorithms analyse data used for classification.

Kullback–Leibler divergence/error (K-L error)
K-L error is a measure of the difference between two probability distributions (for details, see REF. 37).

Summed squared error (SSE)
SSE is a measure of the discrepancy between the data and an estimation model, by which a small SSE indicates a good fit of the model to the data.

Backward elimination
This process produces a ranking of biomarkers with the purpose of assigning low ranks to randomly correlated biomarkers, consequently generating a signature where all biomarkers contribute to the accuracy of a classification (for details, see REFS 36,37 and FIG. 1).

Receiver operating characteristic (ROC)
ROC curves display the relationship between diagnostic sensitivity and specificity. The performance of a test classifier is measured as the area under the curve (AUC). A classifier performing no better than random will have an AUC of 0.5, whereas a perfect classifier will have an AUC of 1. Usually for diagnostic tests it is considered that AUC 0.6–0.7 is poor, AUC 0.7–0.8 is fair, AUC 0.8–0.9 is good and AUC 0.9–1.0 is excellent (see for example, Further information, The Area Under an ROC Curve).

Model instability
This refers to the situation in which the predictive equation changes considerably between training sets[74].

Artificial neural networks
These are systems that are self-learning and trained rather than explicitly programmed.

information. However, during biomarker analysis the challenge is to define the biomarker combination delivering optimal analytical accuracy. This cannot be based simply on *P* values, independently calculated for each marker, as this loses information about synergistic contributions between the markers that might be crucial for improving the classification accuracy. The challenge is to define a combination of 'orthogonal biomarkers' that do not depend on each other, such that the information from each is used optimally for the predictive power of the signature. To achieve this, we need, in an ordered approach, to eliminate the markers with the lowest impact on the accuracy, which can be achieved by combining the leave-one-out cross-validation procedure with a backward elimination algorithm[36]. Briefly, iteratively eliminating markers one by one and identifying those that contributed the least to a correct sample classification, using summed squared error or the 'Kullback–Leibler divergence' (REF. 37), would produce a ranking of the markers (FIG. 1). This enables the selection of a biomarker signature displaying optimal accuracy for each application.

The power of biomarker signatures

Today, we have ample evidence that we need a multitude of biomarkers, a so-called biomarker signature, to mine the wealth of information contained in biological fluids such as serum or plasma. Combinations of biomarkers contain much more information[38,39] than a single biomarker, where the latter does not display sufficient discriminatory power to substantially affect clinical decisions.

Lessons learned from genomics. Genomics has been the forerunner in the use of the concept of biomarker signatures, owing to the fact that transcriptional profiling has been technically easy to perform. This is exemplified by a 70-gene prognostic signature in sporadic breast tumours that, with 83% accuracy, is predictive of poor prognosis (a short time period to distant metastases)[40,41]. This gene signature was shown in a prospective, European multi-centre trial to affect treatment decisions for adjuvant chemotherapy[42], and received 510(k) FDA clearance in 2007. Furthermore, in a prospective study on patients with oestrogen receptor-positive (ER^+) and $HER2^-$ (also known as $ERBB2^-$) breast cancer, a 21-gene signature could identify low-risk patients who had a good outcome despite not receiving adjuvant chemotherapy[43]. Recently, a novel approach focusing on monitoring four to six chromosomal rearrangements in cell-free ctDNA in plasma was, in a limited cohort of patients with breast cancer, demonstrated to be highly accurate (93%) for postsurgical discrimination of patients who eventually displayed metastatic disease[44]. If validated in an independent, larger cohort, the clinical utility clearly lies in the fact that detection of ctDNA preceded the clinical diagnosis in 86% of patients by an average of 11 months, potentially enabling initiation of an earlier interventional therapy[44]. These studies demonstrate that the information content in biomarker signatures could enable clinical decisions in cancer, a concept further supported by combining genomic and proteomic markers, as well as including standard clinical variables, as exemplified below.

Combining genomic and proteomic biomarkers. Prostate cancer has also attracted a lot of attention and here the clinical utility lies in the possibility of distinguishing prostate cancer from benign prostate conditions. PSA is associated with prostate cancer but also with benign prostate indications such as hyperplasia. Consequently, owing to poor test specificity, most patients with an elevated PSA level do not have cancer, which often results in over-treatment with no health benefit for the patient[45,46]. Improved test accuracy would not only minimize the number of negative biopsies but would also result in more cost-effective detection[47]. An equally important focus of prostate cancer biomarker research, which could guide treatment strategies, is to define markers for high-risk disease, most widely defined by the National Comprehensive Cancer Network (NCCN) (see Further information) as a Gleason score of >8 or PSA >20 ng ml^{-1} (REF. 46). To achieve this, various combinations of genomic or proteomic biomarkers have been investigated, resulting in both 510(k) FDA cleared tests and laboratory developed tests (LDTs); for review see REF. 46. In a recent attempt to further improve detection of prostate cancer, a combination of genetic tumour markers (232 single nucleotide polymorphisms (SNPs)), six plasma protein biomarkers from the literature and five standard clinical variables was tested. This approach demonstrated improved performance (receiver operating characteristic (ROC) area under the curve (AUC) of 0.74; see BOX 1), compared with PSA alone (ROC AUC of 0.56), for the detection of tumours with a Gleason score of at least 7 (REF. 48). This model was shown to also reduce the number of benign biopsies by 44%[48].

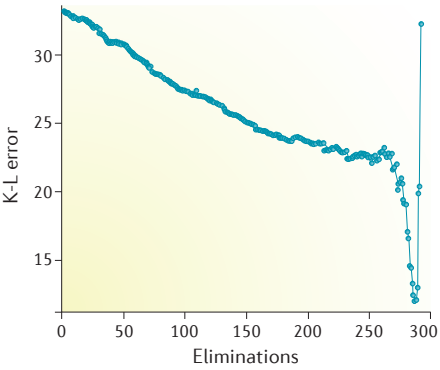

Figure 1 | **An example of the backward elimination principle.** Backward elimination enables the generation of a condensed biomarker signature, consisting of biomarkers all contributing orthogonal information. Kullback–Leibler divergence error (K-L error; BOX 1)[37] is displayed as a function of the number of biomarker eliminations, and enables selection of a signature displaying the lowest divergence error, which in this example occurred after approximately 275 markers had been eliminated. In this example, the remaining 25 biomarkers displayed the highest achievable accuracy, when the K-L error reached its minimum.

Proteomic biomarker signatures. Advanced cancer diagnostics based merely on proteomics has only recently delivered biomarker signatures with the required clinical accuracy. This has been due to technological difficulties in decoding complex proteomes, as well as to lack of rigorous validation[49,50]. One consequence of these issues is that today only one multiparametric test has received 510(k) FDA clearance[8]. The test (OVA1 for ovarian cancer) uses five proteomic serum biomarkers (CA125, transthyretin, apolipoprotein A-I (APOA1), β2-microglobulin and transferrin), previously identified by surface-enhanced laser desorption ionization–time-of-flight (SELDI–TOF) MS[51] and immunoassay[52], to assist in identifying patients diagnosed with an ovarian tumour for direct referral to a gynaecological oncologist, which results in overall better outcomes[12]. The OVA1 test, displaying a ROC AUC of 0.90, correctly predicted ovarian malignancy in 91.4% of cases of early-stage disease, compared with 65.7% for CA125 alone[12,53]. A second-generation version of OVA1, also based on CA125, transferrin and APOA1, but with two additional biomarkers (follicle-stimulating hormone and human epididymis

protein 4 (also known as WFDC2)) was recently approved by the FDA. This test (OVA2) exhibited significantly improved specificity[54].

The impact of proteomics in cancer diagnostics could be advanced by the application of novel technologies that could significantly improve test accuracy and the use of samples from well-designed cohort studies[49]. There is plenty of information in serum, but owing to the vast dynamic range of protein concentrations, this information has been partly inaccessible by conventional MS unless pre-analytical procedures such as protein depletion and/or extensive fractionation are applied[55,56]. This somewhat limits the capacity of MS for biomarker discovery in large sample cohorts[57]. However, recent progress in multiplexing[58] and immunoaffinity MS using antipeptide antibodies[59] is paving the way for clinical diagnostics using MS. Furthermore, multiplexed MS has already demonstrated clinical utility in breast cancer molecular classification, displaying a ROC AUC of 0.83 (REF. 60).

Prognosis of cancer. Over the past couple of years, antibody microarrays have reached the point at which clinically relevant information related to risk classification and/or prognosis can be generated. For example, based on a microarray set-up with 162 antibody fragments, biomarker signatures associated with different risk groups of patients with prostate cancer were identified with ROC AUCs of up to 0.84 (REF. 61). Furthermore, patients with breast cancer could be classified with a ROC AUC of 0.85 into high- versus low-risk groups for developing metastatic disease, using a 21-protein signature[36]. This was based on a retrospective, longitudinal study on breast cancer samples collected between 0 and 36 months after the primary tumour resection, with a follow-up time of 5 years[36]. An antibody microarray containing 135 antibody fragments was used to identify the 21 proteins based on biomarker velocity. Clinical utility was demonstrated, as the protein signature provided an added value when combined with conventional clinical parameters (ROC AUC = 0.66). The combination increased the ROC AUC to 0.90 for prediction of recurrence[36]. In high-grade bladder cancer, an antibody microarray was also used to define a panel of two prognostic tissue biomarkers associated with recurrence-free and progression-free survival[62].

Early detection of cancer. The examples mentioned above demonstrate that combinations of biomarkers can increase diagnostic accuracy in several cancer indications. However, to achieve clinical impact in early diagnosis, performance must increase considerably, ideally reaching ROC AUC values of approximately 0.95 (BOX 1), as the consequence of false negatives will be substantial for the patient, especially when dealing with fast-progressing and deadly cancer indications[63]. Furthermore, false positives will also have a negative impact for patients, resulting in several undesirable side-effects due to overtreatment, as mentioned above for prostate cancer.

One extensively studied indication is pancreatic ductal adenocarcinoma (PDAC), as it is one of the most deadly cancers, with a 5-year survival of 5–6%, and a mortality exceeding that of breast cancer[64]. As early detection of pancreatic cancer could significantly improve survival, as well as reduce societal costs[65], it has been the focus of several studies over the past few years. In a retrospective study, patients with PDAC (samples mostly from late or undefined stages) could be discriminated from healthy individuals with a ROC AUC of 0.93, using a signature of three serum biomarkers (CA19-9, intercellular adhesion molecule 1 (ICAM1) and osteoprotegerin (OPG; also known as TNFRSF11B))[27]. The biomarkers were identified using a bead-based antibody array, initially screening for 83 markers[27]. However, in a population-based follow-up study the same three-biomarker panel failed to perform adequately[29] and could not be used for pre-diagnostic risk assessment, demonstrating the challenges for early diagnosis of PDAC.

To circumvent the lack of available samples for identifying biomarkers for early diagnosis, for which preferably stage I or II is needed, another study interrogated the tissue proteome of a genetically engineered mouse model of PDAC, at preinvasive and invasive stages, using a high-density antibody microarray[66]. In a follow-up study, plasma samples from both the mouse model and pre-diagnostic plasma from individuals that later succumbed to PDAC were analysed, using an antibody microarray platform containing 130 antibodies relevant to PDAC[67]. This combined approach enabled the identification of a panel of three protein biomarkers giving a ROC AUC of 0.68 in pre-diagnostic samples, increasing to 0.86 in diagnostic samples, that is, those derived from patients at the time of clinical diagnosis. As noted above, these results were based on a three-biomarker signature (HER2, tenascin and ER); it seems likely, however, that a larger biomarker signature would increase the analytical accuracy.

The benefit of a larger panel of protein biomarkers was demonstrated using a cohort of 148 diagnostic samples derived from patients with either PDAC or pancreatitis[68]. Using an antibody microarray, a ROC AUC of up to 0.99 was achieved in discriminating cancer from inflammatory states of the pancreas[69]. This analytical accuracy was based on a serum biomarker signature of 25 different immunoregulatory proteins[68] and was optimized by the backward elimination principle[36] (FIG. 1). Of particular interest is

Glossary

Bead-based arrays
Similar to antibody microarrays but the antibodies are deposited on micro-beads instead of on a planar surface.

Biomarker velocity
The change in signal of a biomarker over time.

Enzyme-linked immunosorbent assay
(ELISA). A solid-phase immunoassay that measures the interaction between proteins and specific antibodies.

510(k)
A premarketing submission made to the US Food and Drug Administration (FDA) to demonstrate that the test is safe and effective. If cleared by the FDA, the test can be marketed in the United States.

Gleason score
A score given to a prostate cancer based on its microscopic appearance, whereby a higher Gleason score indicates a more aggressive tumour.

Antibody microarrays
Miniaturized enzyme-linked immunosorbent assay format.

Laboratory developed tests
(LDTs). *In vitro* diagnostic tests that are designed, manufactured and used in a single laboratory and not approved by the US Food and Drug Administration.

Reverse phase protein arrays
Arrays in which protein samples are deposited in micro-scale on a planar surface and probed with specific antibodies.

Selected reaction monitoring or multiple reaction monitoring
(SRM/MRM). Two names for a method used in tandem mass spectrometry to quantitatively target individual proteins or peptides.

the fact that PDAC could be distinguished with high accuracy from benign conditions[68], as well as from chronic or autoimmune pancreatitis[69]. Consequently, a signature based on 25 biomarkers confers increased robustness to the analysis, as the accuracy is less dependent on variation in single biomarkers. This was also recently demonstrated in a multi-centre trial for early diagnosis of PDAC[70], in which 338 case and control serum samples were collected from five different sites in Spain, and the discrimination of PDAC case versus healthy individual still achieved an analytical accuracy of a ROC AUC of 0.98 (REF. 70). Furthermore, an interesting finding in this study was the first indication that a protein marker panel could identify the location of a tumour within the pancreas, that is, head versus body or tail. As the tumour location could result in different treatment efficacy[71], biomarkers that can determine tumour localization could be of clinical relevance. Finally, a recent study also demonstrated the important discrimination for early detection between PDAC stage I/II and controls in a Chinese cohort, displaying a ROC AUC of 0.80 (REF. 72), based on a 25-biomarker signature. This was again carried out using a microarray set-up now based on 350 antibodies[72].

Conclusions

Consequently, we can today derive novel information from a liquid biopsy based on protein biomarker signatures in various cancer indications, information that is actionable and not available using conventional approaches. The clinical utility is directly apparent in early detection of PDAC, risk classification of prostate cancer and diagnosis of recurrence in breast cancer, although prediction of both therapy efficacy and disease progression is emerging as highly relevant in several cancer indications. Early detection strategies are also a focus of the 2016 Cancer Moonshot initiative, as well as of the Recalcitrant Cancer Research Act of 2012 (see Further information). In the latter, the emphasis is in particular on improving prevention, detection and diagnosis of PDAC and lung cancer. The general consensus today leans towards the view that prevention is preferable to treatment, and this is one grand challenge for the emerging field of precision diagnostics, in which we will see valuable clinical impacts, such as, increased overall survival — not for future generations but for today's. However, to finally benefit the patient, health authority regulatory models have to rapidly adapt to these developing multiparametric diagnostic technologies to facilitate their implementation in the clinical setting[73].

Carl A. K. Borrebaeck is at the Department of Immunotechnology, CREATE Health Translational Cancer Center, Medicon Village (Bldg 406), Lund University, 223 81 Lund, Sweden.

carl.borrebaeck@immun.lth.se

doi:10.1038/nrc.2016.153
Published online 3 Feb 2017

1. World Health Organization. Cancer control: early detection: WHO guide for effective programmes. *WHO* http://www.who.int/cancer/publications/cancer_control_detection/en/ (2007).
2. Shimizu, Y., Yasui, K., Matsueda, K., Yanagisawa, A. & Yamao, K. Small carcinoma of the pancreas is curable: new computed tomography finding, pathological study and postoperative results from a single institute. *J. Gastroenterol. Hepatol.* **20**, 1591–1594 (2005).
3. Brennan, D. J. *et al.* Antibody-based proteomics: fast-tracking molecular diagnostics in oncology. *Nat. Rev. Cancer* **10**, 605–617 (2010).
4. Neagu, M., Constantin, C., Tanase, C. & Boda, D. Patented biomarker panels in early detection of cancer. *Recent Pat. Biomark.* **1**, 10–24 (2011).
5. Vlahou, A. Network views for personalized medicine. *Proteomics Clin. Appl.* **7**, 384–387 (2013).
6. Franzi, M., Bhat, A. & Latosinska, A. Clinical proteomic biomarkers: relevant issues on study design and technical considerations in biomarker development. *Clin. Transl Med.* **3**, 7–22 (2014).
7. Polanski, M. & Anderson, N. L. A list of candidate cancer biomarkers for targeted proteomics. *Biomark. Insights* **2**, 1–48 (2007).
8. Füzey, A. K., Levin, J., Chan, M. M. & Chan, D. W. Translation of proteomic biomarkers into FDA approved cancer diagnostics: issues and challenges. *Clin. Proteomics* **10**, 13–27 (2013).
9. Menon, U. *et al.* Risk algorithm using serial biomarker measurements doubles the number of screen-detected cancers compared with a single-threshold rule in the United Kingdom Collaborative Trial of Ovarian cancer screening. *J. Clin. Oncol.* **33**, 2062–2075 (2015).
10. Pavlou, M. P., Diamandis, E. P. & Blasutig, I. M. The long journey of cancer biomarkers from the bench to the clinic. *Clin. Chem.* **59**, 147–157 (2013).
11. Smart, A. A multi-dimensional model of clinical utility. *Int. J. Qual. Health Care* **18**, 377–382 (2006).
12. Zhang, Z. & Chan, D. W. The road from discovery to clinical diagnostics: lessons learned from the first FDA-cleared *in vitro* diagnostic multivariate index assay of proteomic biomarkers. *Cancer Epidemiol. Biomarkers Prev.* **19**, 2995–2299 (2010).
13. Kiyonami, R. *et al.* Increased selectivity, analytical precision, and throughput in targeted proteomics. *Mol. Cell. Proteomics* **10**, M110.002931 (2011).
14. Tighe, P. J., Ryder, R. R., Todd, I. & Fairclough, L. C. ELISA in the multiplexed era: potentials and pitfalls. *Proteomics Clin. Appl.* **9**, 406–422 (2015).
15. Haab, B. B. Applications of antibody array platforms. *Curr. Opin. Biotechnol.* **4**, 415–421 (2006).
16. Borrebaeck, C. A. K. & Wingren, C. Design of high-density antibody microarrays for disease proteomics: key technological issues. *J. Proteomics* **72**, 928–935 (2009).
17. Bradbury, A. & Pluckthun, A. Reproducibility: standardized antibodies used in research. *Nature* **518**, 27–29 (2015).
18. Alhamdani, M. S. S., Schröder, C. & Hoheisel, J. D. Oncoproteomic profiling with antibody microarrays. *Genome Med.* **1**, 1–7 (2009).
19. Liotta, L. A. *et al.* Protein microarrays: meeting analytical challenges for clinical applications. *Cancer Cell* **3**, 317–325 (2003).
20. Borrebaeck, C. A. K. & Wingren, C. Recombinant antibodies for the generation of antibody arrays. *Methods Mol. Biol.* **785**, 247–262 (2011).
21. Yu, X., Schneiderhan-Marra, N. & Joos, T. O. Protein microarray for personalized medicine. *Clin. Chem.* **56**, 376–387 (2010).
22. Ellmark, P. *et al.* Identification of protein expression signatures associated with *Helicobacter pylori* infection and gastric adenocarcinoma using recombinant antibody microarrays. *Mol. Cell. Proteomics* **5**, 1638–1646 (2006).
23. Carlsson, A. *et al.* Serum protein profiling of metastatic breast cancer using recombinant antibody microarrays. *Eur. J. Cancer* **44**, 472–480 (2008).
24. Shao, C. *et al.* Antibody microarray analysis of serum glycans in esophageal squamous cell carcinoma cases and controls. *Proteomics Clin. Appl.* **3**, 923–931 (2009).
25. Chapman, C. J. *et al.* Immunobiomarkers in small cell lung cancer: potential early cancer signals. *Clin. Cancer Res.* **17**, 1474–1480 (2011).
26. Sonntag, J. *et al.* Reverse phase protein array based tumor profiling identifies a biomarker signature for risk classification of hormone receptor-positive breast cancer. *Transl Proteomics* **2**, 52–59 (2014).
27. Brand, R. E. *et al.* Serum biomarker panels for the detection of pancreatic cancer. *Clin. Cancer Res.* **17**, 805–816 (2011).
28. Sauer, G. *et al.* Molecular indicators of non-sentinel node status in breast cancer determined in preoperative biopsies by multiplexed sandwich immunoassays. *J. Cancer Res. Clin. Oncol.* **137**, 1175–1184 (2011).
29. Nolen, B. M. *et al.* Prediagnostic serum biomarkers as early detection tools for pancreatic cancer in a large prospective cohort study. *PLoS ONE* **9**, e94928 (2014).
30. Rosskopf, S. *et al.* The pre-analytical processing of blood samples for detecting niomarkers on protein microarrays. *J. Immunol. Methods* **418**, 39–51 (2015).
31. Rai, A. J. *et al.* HUPO plasma proteome project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. *Proteomics* **5**, 3262–3277 (2005).
32. Alonzo, T. A., Pepe, M. S.& Moskowitz, C. S. Sample size calculations for comparative studies of medical tests for detecting presence of disease. *Stat. Med.* **21**, 835–852 (2002).
33. Melo, S. A. *et al.* Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. *Nature* **523**, 177–182 (2015).
34. Borrebaeck, C. A. K. Viewpoints in clinical proteomics. *Proteomics Clin. Appl.* **6**, 343 (2012).
35. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **57**, 289–300 (1995).
36. Carlsson, A. *et al.* Molecular serum portraits in patients with primary breast cancer predict the development of distant metastases. *Proc. Natl Acad. Sci. USA* **108**, 14252–14257 (2011).
37. Kullback, S. & Leibler, R. A. On information and sufficiency. *Ann. Math. Stat.* **22**, 79–86 (1951).
38. Hansh, S. M., Pitteri, S. J. & Faca, V. M. Mining the plasma proteome for cancer biomarkers. *Nature* **452**, 571–579 (2008).
39. Zethelius, B. *et al.* Use of multiple biomarkers to improve the prediction of death from cardiovascular causes. *N. Engl. J. Med.* **358**, 2107–2116 (2008).
40. Van 't Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).
41. Van de Vijver, M. J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009 (2002).
42. Cusumano, P. G. *et al.* European inter-institutional impact study of MammaPrint. *Breast* **23**, 423–428 (2014).
43. Sparano, J. A. *et al.* Prospective validation of a 21-gene expression assay in breast cancer. *N. Engl. J. Med.* **373**, 2005–2014 (2015).
44. Olsson, E. *et al.* Serial monitoring of circulation tumor DNA in patients with primary breast cancer for detection of occult metastatic disease. *EMBO Mol. Med.* **7**, 1034–1047 (2015).
45. Moyer, V. A. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. *Ann. Int. Med.* **157**, 120–134 (2012).
46. Sartori, D. A. & Chan, D. W. Biomarkers in prostate cancer: what's new? *Curr. Opin. Oncol.* **26**, 259–264 (2014).
47. Heijnsdijk, E. A., Denham, D. & de Koning, H. J. The cost-effectivness of prostate cancer detection with the use of prostate health index. *Value Health* **19**, 153–157 (2016).
48. Grönberg, H. *et al.* Prostate cancer screening in men aged 50–69 years (STHLM3): a prospective population-based diagnostic study. *Lancet Oncol.* **16**, 1667–1676 (2015).

49. Schully, S. D. *et al.* Leveraging biospecimen resources for discovery or validation of markers for early cancer detection. *J. Natl Cancer Inst.* **107,** djv012 (2015).
50. Ransohoff, D. F. Proteomics research to discover markers: what can we learn from Netflix? *Clin. Chem.* **56**, 172–176 (2010).
51. Zhang, Z. *et al.* Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res.* **64**, 5882–5890 (2004).
52. Bast, R. C. *et al.* Reactivity of a monoclonal antibody with human ovarian carcinoma. *J. Clin. Invest.* **68**, 1331–1337 (1981).
53. Bristow, R. E. *et al.* Ovarian malignancy risk classification of adnexal mass using a multivariate index assay. *Gynecol. Oncol.* **128**, 252–259 (2013).
54. Coleman, R. L. *et al.* Validation of a second-generation multivariate index assay for malignancy risk of adnexal masses. *Am. J. Obstet. Gynecol.* **82**, 1–11 (2016).
55. Anderson, N. L. & Anderson, N. G. The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* **11**, 845–867 (2002).
56. Makarov, A., Denisov, E., Lange, O. & Horning, S. Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer. *J. Am. Mass Spectrom.* **17**, 977–982 (2006).
57. Parker, C. E. & Borchers, C. H. Mass spectrometry based biomarker discovery, verification, and validation — quality assurance and control of protein biomarker assays. *Mol. Oncol.* **8**, 840–858 (2014).
58. Abbatiello, S. E. *et al.* Large-scale interlaboratory study to develop, analytically validate and apply highly multiplexed, quantitative peptide assays to measure cancer-relevant proteins in plasma. *Mol. Cell. Proteomics* **14**, 2357–2374 (2015).
59. Razavi, M., Anderson, N. L., Yip, R., Pope, M. E. & Pearson, T. W. Multiplexed longitudinal measurements of protein biomarkers in DBS using an automated SISCAPA workflow. *Bioanalysis* **8**, 1597–1160 (2016).
60. Waldemarson, S. *et al.* Proteomic analysis of breast tumors confirms the mRNA intrinstic molecular subtypes using different classifiers: a large-scale analysis using fresh frozen tissue samples. *Breast Cancer Res.* **18**, 69 (2016).
61. Nordström, M. *et al.* Identification of plasma protein profiles associated with risk groups of prostate cancer patients. *Proteomics Clin. Appl.* **8**, 951–962 (2014).
62. Lee, M.-S. *et al.* Prognostic significance of CREB-binding protein and CD81 expression in primary high grade non-muscular invasive bladder cancer: identification of novel biomarkers for bladder cancer using antibody microarray. *PLoS ONE* **10,** e0125405 (2015).
63. Hartwell, L., Mankoff, D., Paulovich, A., Ramsey, S. & Swisher, E. Cancer biomarkers: a systemic approach. *Nat. Biotechnol.* **24**, 905–908 (2006).
64. Kamisawa, T., Wood, L. D., Itoi, T. & Takaori, K. Pancreatic cancer. *Lancet* **388**, 73–85 (2016).
65. Ghatnekar, O. *et al.* Modelling the benefits of early diagnosis of pancreatic cancer using a biomarker signature. *Int. J. Cancer* **133**, 2392–2397 (2013).
66. Mirus, J. E. *et al.* Spatiotemporal proteomic analyses during pancreas cancer progression identifies serine/threonine stress kinase 4 (STK4) as a novel candidate biomarker for early stage disease. *Mol. Cell. Proteomics* **13**, 3484–3496 (2014).
67. Mirus, J. E. Cross-species antibody microarray interrogation identifies a 3-protein panel of plasma biomarkers for early diagnosis of pancreas cancer. *Clin. Cancer Res.* **21**, 1764–1771 (2015).
68. Wingren, C. *et al.* Identification of serum biomarker signatures associated with pancreatic cancer. *Cancer Res.* **72**, 2481–2490 (2012).
69. Sandström, A. *et al.* Serum proteome profiling of pancreatitis using recombinant antibody microarrays reveals disease-associated biomarker signatures. *Proteomics Clin. Appl.* **6**, 486–496 (2012).
70. Gerdtsson, A. S. *et al.* A multicenter trial defining a serum protein signature associated with pancreatic ductal adenocarcinoma. *Int. J. Proteomics* **2015**, 587250 (2015).
71. Wu, T.-C., Shao, Y.-F., Shan, Y., Wu, J.-X. & Zhao, P. Surgical effect of malignant tumor of body and tail of the pancreas: compare with pancreatic head cancer. *Chin. J. Surg.* **45**, 30–33 (2007).
72. Gerdsson, A. S. *et al.* Plasma protein profiling in a stage defined pancreatic cancer cohort - implications for early diagnosis. *Mol. Onc.* **10**, 1305–1316 (2016).
73. Mansfield, E. A. FDA perspective on companion diagnostics: an evolving paradigm. *Clin. Cancer Res.* **20**, 1453–1457 (2014).
74. Dinakarpandian, D. in *Big Data Analysis for Bioinformatics and Biomedical Discoveries* (ed. Ye, S. Q.) 249–264 (CRC Press, 2016).

Acknowledgements

The critical reading and comments suggested by J. Borrebaeck (University of California, Berkeley, USA) and C. Rose, C. Wingren, S. Ek, and expert bioinformatic guidance by C. Peterson and M. Ohlsson (CREATE Health Translational Cancer Center, Lund University, Sweden) are gratefully acknowledged.

Competing interests statement

The author declares competing interests: see Web version for details.

FURTHER INFORMATION

2016 Cancer Moonshot initiative: https://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative
National Comprehensive Cancer Network (NCCN): https://www.nccn.org/
Recalcitrant Cancer Research Act 2012: https://www.cancer.gov/about-nci/legislative/recent-public-laws#ui-id-8
The Area Under an ROC Curve: http://gim.unmc.edu/dxtests/roc3.htm

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

Author biography

Carl Borrebaeck received the first chair as Professor of Immunotechnology in Scandinavia in 1990. His main research interests are cancer proteomics, for early detection and prognosis, and antibody engineering, for the generation of human therapeutic antibodies. He founded CREATE Health Translational Cancer Center, Lund University, Sweden, in 2005. He received the AKZO Nobel Science Award in 2009, for his contributions to cancer proteomics and antibody-based therapy, the Research!Sweden Award in 2012 for his medical research and the Royal Academy of Engineering Sciences Gold Medal in 2012 for outstanding contributions to biomedical science. He is a permanent member of the Royal Swedish Academy of Engineering Sciences. He held the position of Vice-President of Lund University, 2009–2015, responsible for innovation and industrial collaboration. In 1996–97 he spent a sabbatical year at the Oklahoma Medical Research Foundation in the United States and did his postdoctoral training at the University of California in Davis, USA.

Competing interests statement

The author is one of the founders of a start-up diagnostic company in complex diseases.

Subject categories

Biological sciences / Cancer / Tumour biomarkers
[URI /631/67/1857]
Biological sciences / Drug discovery / Biomarkers
[URI /631/154/53]
Biological sciences / Biotechnology / Proteomics
[URI /631/61/475]

ToC blurb

000 **Precision diagnostics: moving towards protein biomarker signatures of clinical utility in cancer**

Carl A. K. Borrebaeck

This Opinion article focuses on the trends and progress being made in identifying protein biomarker signatures of clinical utility in cancer using, in particular, blood-based proteomics.