



Clinical proteomics: Promises, challenges and limitations of affinity arrays

Christian Betzen^{1,2}, Mohamed Saiel Saeed Alhamdani¹, Smiths Lueong¹, Christoph Schröder^{1,3}, Axel Stang⁴ and Jörg D. Hoheisel¹

¹ Division of Functional Genome Analysis, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany

² Department of Pediatric Medicine I, University Children's Hospital Heidelberg, Heidelberg, Germany

³ Sciomics GmbH, Heidelberg, Germany

⁴ Oncology, Asklepios Klinik Barmbek, Hamburg, Germany

After the establishment of DNA/RNA sequencing as a means of clinical diagnosis, the analysis of the proteome is next in line. As a matter of fact, proteome-based diagnostics is bound to be even more informative, since proteins are directly involved in the actual cellular processes that are responsible for disease. However, the structural variation and the biochemical differences between proteins, the much wider range in concentration and their spatial distribution as well as the fact that protein activity frequently relies on interaction increase the methodological complexity enormously, particularly if an accuracy and robustness is required that is sufficient for clinical utility. Here, we discuss the contribution that protein microarray formats could play towards proteome-based diagnostics.

Received: October 13, 2014

Revised: December 16, 2014

Accepted: January 13, 2015

Keywords:

Affinity profiling / Antibodies / Immunoassay / Interaction / Protein microarrays

Currently, we are getting a glimpse of how high-throughput analysis formats, such as next-generation sequencing, will have an enormous impact on the understanding of molecular aspects that underlie human physiology and pathology. The ability of reading and deciphering the genomic sequence of individual patients [1] will change the way diagnostics is done and thus translate quickly into diagnostic reality. Diagnostic sequencing, however, will have consequences more for diseases, which are caused by variations at the genetic and epigenetic level, such as cancer, although patient stratification could well be possible even if a causative link between markers and disease may not be established.

Analyses at the protein level hold even higher promise for identifying changes that go along with illness and for defining a fitting treatment. It is proteins that are responsible for

most known cellular activities. For this reason, the large majority of current drugs is affecting proteins. Knowledge about protein variations is therefore more likely to identify a link between changes in their abundance, structure or location and health consequences, and is better suited to delineate the effect of a drug. In particular in view of companion diagnostics, protein analyses will thus be more informative than the assessment of nucleic acids. What is missing in the field of proteomics, however, are assays of the kind represented by DNA sequencing, which combines high throughput for comprehensive coverage with high accuracy, low cost and good sensitivity, the last by means of DNA amplification processes.

One reason for proteomics lagging behind is the complexity of the proteome. An estimated one to several million protein derivatives are encoded by about 22 000 genes. This number is augmented by an unknown number of disease-relevant variations. In addition, the large number of structural and biochemical differences of proteins and a very wide range in protein concentration make things even more complicated. Recently, initial draft maps of the human proteome were published [2–4]. Although they represent the result of distinct scientific efforts, the discrepancy between the data shown and the anticipated complexity of the proteome in a

Correspondence: Jörg D. Hoheisel, Division of Functional Genome Analysis, Deutsches Krebsforschungszentrum (DKFZ), Im Neuenheimer Feld 580, 69120 Heidelberg, Germany

E-mail: j.hoheisel@dkfz.de

Fax: +49-6221-424687

Abbreviation: IHC, immunohistochemistry



Correspondence concerning this and other Viewpoint articles can be accessed on the journals' home page at: <http://viewpoint.clinical.proteomics-journal.com>

Correspondence for posting on these pages is welcome and can also be submitted at this site.

human being documents the still relatively infant status of knowledge.

Another factor that is limiting the ability of studying all aspects of the protein repertoire are the technical means currently at hand. MS in all its facets has led to a quantum leap in protein analysis and was instrumental for speeding up the process of analysing proteins with good accuracy and reproducibility. It has been the technique by which very many potential protein markers have been discovered. The fact, however, that only very few of them are actually used in clinical practice indicates shortcomings in making the results fit into a clinical setting [5]. Technical complexity of the analysis process, a still limited throughput and lacking sensitivity are aspects that have been restrictive to date. Sample preparation is still complex, and the capacity for quantitative measurements is not yet at a level required by routine diagnostics in a clinical setting [6, 7]. Another obstacle is the fact that proteins do not act on their own usually, but exhibit their activity together with other proteins or ligands.

Protein interaction is crucial for cellular activity. About 80% of the proteins assemble in multi-protein complexes. Yeast-two-hybrid and other procedures have contributed a lot to identifying interactions [8], but most results are qualitative and lack quantification of the contact strength, which is changing dependent of the conditions. Interaction is regulated by the intrinsic affinity of proteins to their ligands, which is affected and thus co-regulated by concentration, structure and environment. An analysis beyond the mere identification of interacting partners is therefore a multifaceted task and requires processes that take into account these features. In consequence, assays that utilise the affinity between molecular partners could add substantially to the portfolio of protein-analysing methods and complement analyses done by MS.

Protein microarrays represent such a methodology. They have become very powerful tools for research purposes, combining high-throughput with little sample consumption, good technical reproducibility and (semi-) quantitative accuracy with sensitivities that surpass that of ELISAs, down to the detection of individual binding events [9]. Nevertheless, also here, very little has yet spilled over into the more demanding area of clinical application. Protein arrays can be subdivided into three basic formats: antigen arrays present a large set of different proteins; antibody microarrays display many antibodies or other types of binder molecules that target proteins of interest; for reverse phase arrays, finally, protein extracts are isolated from individual (patient) specimens, printed in an array format and studied in comparison to each other.

Arrays that consist of individual proteins at defined positions are technically the most difficult to produce. While spotting pre-fabricated molecules works fine [10], it is more likely that in the long run *in situ* cell-free synthesis will prevail for applications, in which patient proteins are being studied individually, similarly to what happened in the field of DNA microarrays. Production of the latter format is far less labour-intensive and can be performed under standardised conditions [11–13], although progress of implementation is slow. While chemical *in situ* synthesis of peptides is well established, *in situ* protein synthesis requires enzymatic steps. It is thus more complicated to achieve high accuracy and reproducibility.

In terms of quality control, none of the processes is yet developed enough to come even close to meeting clinical requirements. Particularly a better definition of the structural composition of the arrayed proteins is needed. Once basic quality control measures for clinical utility will be established, an early application is likely to be the detection of infections with antigen microarrays that display a variety of prefabricated, pathogen-specific proteins. Incubation with the antibodies isolated from serological patient samples could allow a quick and accurate diagnosis by identifying proteins of the pathogenic organism, or a subtype thereof, against which antibodies were produced by the patient [14]. Such immunoprofiling could also identify allergies or other diseases, if disease-specific protein derivatives are known that can be placed onto the array (e.g. [15]).

Analysing patient proteins will require much more sophisticated arrays. For a start, the arrayed proteins need to

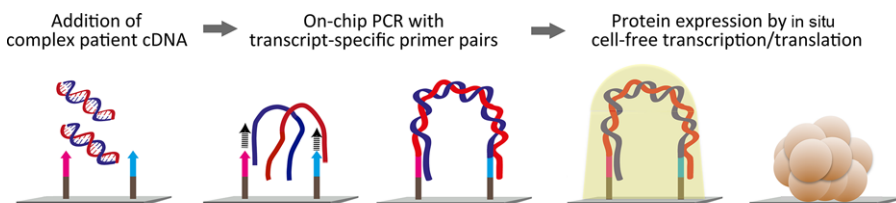


Figure 1. Schematic representation of the process of producing personalised protein microarrays. For each protein, a primer pair is present at a particular array position that permits copying by PCR onto the array the full-length gene transcript that was isolated from an individual patient. By *in situ* cell-free transcription and translation, a protein is being expressed as it was originally transcribed in the patient sample.

represent the molecules as present in each patient, with all the variations caused by mutations and differential splicing. This can be achieved by preparing on the array DNA copies of the transcripts of interest by an in situ PCR amplification; total RNA from the patient or the respective total cDNA would be the template. The arrayed DNA copies are then transcribed and translated by cell-free protein expression (Fig. 1). Although an initial personalised version of the proteome can be studied this way, it would still not reflect entirely the proteins as present in the patient sample. Many PTMs, such as phosphorylation, glycosylation or cleavage by one of the hundreds of human proteases, would be missing or be artificial in position and degree. Site-specific glycosylation, for example is frequently required for proper protein folding [16]. While processes may exist, by which also sample-specific PTMs could be copied onto the microarray, they are very far from being applicable in a clinical setting. Another problem is membrane-associated proteins, although several procedures have been established to produce membrane proteins in a functional format [17]. Last, as a means of quality control, processes need to be implemented, by which an appropriate structure of the arrayed proteins can be both achieved and confirmed, before many clinical applications could become feasible.

A particular segment of affinity-based analysis are assays that rely on the interaction of an antibody – or another binder of similar nature – and a target molecule [18]. Some immunoassays already meet the requirements of today's clinical routine diagnostics (Fig. 2). Immunohistochemistry (IHC) and ELISA are two assay formats, for instance, that are indispensable in today's analysis of protein biomarkers. IHC is very sensitive down to single molecule detection. It is performed on material that represents the diseased tissue and

provides information about protein quantity, location and spatial distribution, producing different types of information in a single assay. In addition, data evaluation is based on exquisite image analysis – mostly the visual functions of the human brain supported by sophisticated technical means – and an expert analysis system – again the human brain – that combines the accumulated knowledge of hundreds of pathology departments over many decades with personal experience. It represents a gold standard that will be difficult to match or surpass.

However, the situation is different for (screening) analyses that are performed on body fluids, a material predestined for diagnosis, since they can be obtained in relatively large quantities in a non-invasive or minimally invasive process. However, the analysis is done on the same sample material for all diseases; no particular tissue section is studied. Also, the assay may be performed on many, mostly even healthy individuals, and is not always supplemented by other information or prior knowledge. Finally, the assay produces no image with structural information, like IHC, but only a mere measurement value. Consequently, the requirements to assay accuracy could actually be higher. Therefore, it is unlikely that many – if any – single marker molecule of sufficient diagnostic or prognostic precision exist. Companion diagnostics increases this hurdle even further. Not only should the information be disease-specific, but should also be linked to the effect of the drug or the activity of the target protein. Consequently, signatures rather than individual biomarkers will be required.

Antibody microarrays (Fig. 3) present an analysis format that is well placed to perform such assays. If relevant binders exist, basically any number of proteins can be studied in parallel. Also, in comparison to ELISA, both the consumption of

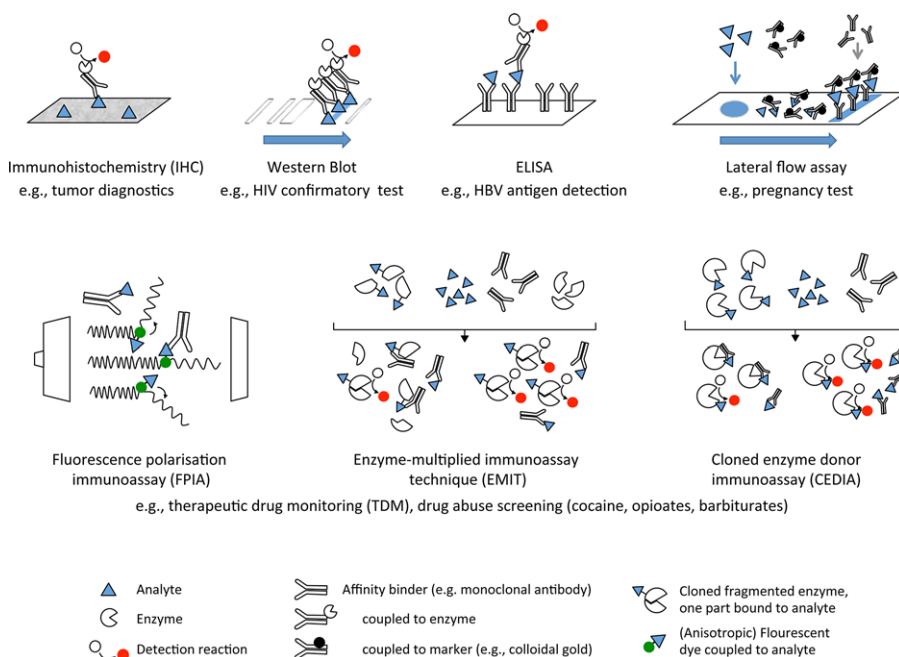


Figure 2. Immunoassays are shown that are currently used in clinical practice. The assay type is named and a typical application is mentioned.

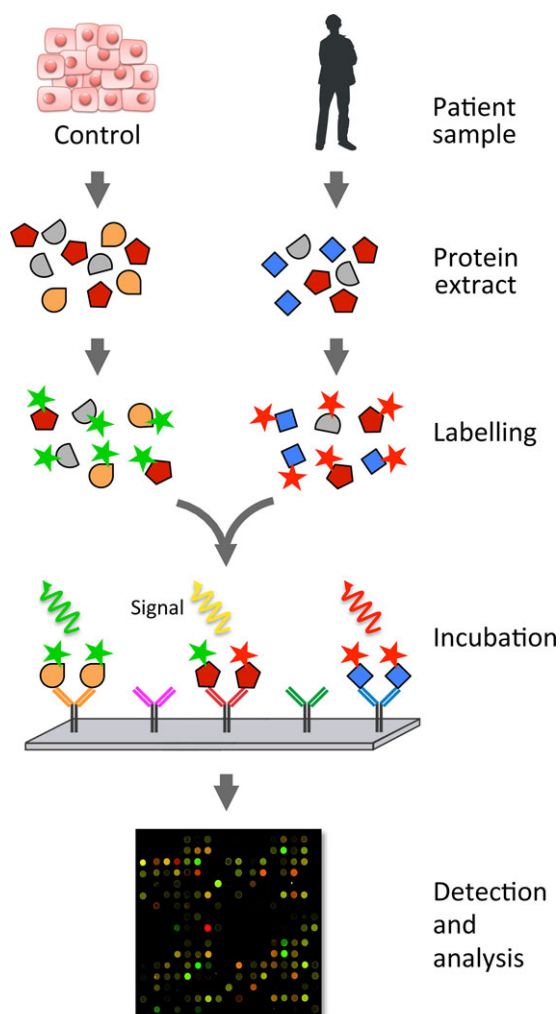


Figure 3. Workflow of an analysis by antibody microarray of the protein content of a patient sample. The protein extract is labelled with a fluorescent dye. Upon incubation with an appropriate control sample, which is labelled with another fluorophore, the relative signal intensities at the various antibody locations on the array are determined and can easily be compared to other analyses, if the same control is applied. Alternatively to the dual-colour approach shown here, analyses are performed that use a single dye only, which requires a different subsequent data processing. Also, a second antibody could be used for detection instead of direct antigen labelling, although the multiplex factor is limited by such an approach.

antibodies and of sample is very low, while better sensitivities can be achieved nevertheless. Most binder types exhibit an extraordinary structural robustness and are functional on an array surface. Technically, antibody microarrays already exhibit performance parameters that meet or exceed clinical requirements [19]. In addition, the format permits analysis of blood samples, for example without the need for sample fractionation or depletion of the most abundant serum proteins, such as albumin; depletion is bound to impair protein proportions artificially. Also, the presence of lipids in high

concentration can be dealt with. Finally, the format is not much different to established clinical immunoassays so that the transfer into routine diagnostics should not fail because of lacking acceptance in the clinical community. Corresponding to IHC, antibody microarrays can produce different forms of information in a single assay. Next to the abundances of the target proteins, structural differences or PTMs can be identified, for example. Particularly for companion diagnostics, the identification of protein isoforms could be crucial.

Currently, antibody microarrays and similar formats with other binder molecules, such as single chain fragment variable antibodies, demonstrate in focussed studies great promise in addressing some unmet clinical needs with quite impressive accuracy (e.g. [20, 21]). What is missing towards a clinical applicability are studies beyond particular disease types, such as very recently performed on reverse-phase microarrays (see below). Even if a protein signature is very specific for one form of cancer, for example, it still needs to be confirmed that an analysis without prior knowledge would yield an accurate diagnosis. In case the signature overlaps too much with that of other cancer types or inflamed tissues, for instance, any patient stratification or prediction of outcome could be obscured substantially, rendering the assay useless in clinical practice. Another main obstacle currently is the availability of antibodies of sufficient quality, which target protein conformations that are informative. Although the *Antibodypedia* database (version 8, October 2014) reports the existence of 1 423 189 antibodies covering gene products encoded by 19 275 genes, many are of insufficient quality. Also, binders are missing that target specifically the many post-translational modified versions of a protein. Particularly, the identification of structural differences is difficult to achieve; many antibodies bind to linear epitopes. However, other binder formats, such as nano-bodies from camels or sharks, for instance, exhibit an overall better distinction of structural variations [22].

In a format reverse to that of antibody microarrays, protein extracts obtained from patients are printed as a whole onto reverse-phase microarrays and undergo an analysis with mostly individual antibodies [23]. The format has the advantage that all PTMs of the proteins are present, as long as they survive the protein extraction process. Protein phosphorylation, for example is prone to change even during relatively short periods of ischemia. For studying patient samples with a small number of well-defined antibodies, the reverse-phase arrays are probably superior to antibody arrays. A downside of the approach is the fact that many proteins of low abundance may not be detectable, particularly if serum or plasma are studied; only an overall small amount of protein is analysed and the enormous excess of albumin and globulins could mask rare proteins and obscure the measurement. The potential of reverse-phase arrays was demonstrated by two recent studies on several thousand patient tissue samples with 131 and 181 antibodies, respectively [24, 25]. They suggest that a protein-based analysis with samples from multiple diseases may allow to identify therapeutic modalities that are

not obvious from studying a single disease or from RNA and DNA analysis alone. They also indicate that different changes at the protein level may lead to the same functional consequence, again highlighting the importance of a broad protein analysis.

In conclusion, while few protein microarray formats are at the brink of becoming robust and accurate enough for clinical utility, much improvement still needs to be done. Particularly, availability of very good antibodies against the various protein forms is holding up the development. However, this obstacle may be overcome relatively soon, as the production of such binders is a focus of current activities at both the academic and commercial level. And not just array technology depends on binders, but basically all kinds of proteomics, including MS, require this resource. Concerning MS, a combination of affinity arrays and MS could actually be a methodology of choice for robust protein diagnosis, since they produce different types of information that are complementary to a large extent, but also sufficiently supplemental to act as an internal control. At the same time, they improve each other, the affinity array acting as a kind of sample preparation procedure required for MS, MS in turn functioning as a readout process that qualifies further the signals produced on the arrays [26].

Many aspects that are important for a clinical application of all kinds of analysis formats are obviously also relevant for an assessment of the clinical utility of affinity arrays. Sensitivity and specificity is such an issue, for example. However, the required accuracy very much depends on the assay type and the disease. Even markers of limited sensitivity and specificity could be of clinical utility, such as the serum marker CA19-9 [27]. Also, the actual purpose of the assay – screening, initial diagnosis, diagnosis of recurrence, prognosis, companion diagnostics, treatment monitoring – will require different levels of accuracy to be of clinical value. Therefore, no common threshold can be defined.

A current view is that highly complex protein and particularly antibody microarrays are a very powerful tool for the identification of biomarkers and the generation of clinically relevant signatures, but that the actual clinical assay will subsequently be set up on a less complex platform that deals with fewer molecules (e.g. [28]). This argument for a diagnosis with relatively few proteins, however, is based more on current technical capabilities than scientific reasons. As for DNA sequencing, technology will develop in a way – with the availability of a renewable and comprehensive source of very well defined antibodies (or other binders) being an essential prerequisite – that could allow whole proteome analyses to be performed within an appropriate time window and at a reasonable cost. As a matter of fact, such data accumulation may actually be required for personalised diagnostics, since compensating effects occurring in the proteome elsewhere may offset in a person-specific manner the changes caused by defined marker molecules. A thorough diagnosis may therefore require a view at very many proteins. It could consequently be cost-effective to look always at the entire proteome rather than

studying a substantial, but varying portion that is needed for properly diagnosing a particular disease-type in an individual patient. In view of this, array-based affinity proteomics could well become a diagnostic process that will be the norm rather than the exception.

Our work on protein microarrays was financially supported by grants to J.D.H. of the European Commission, most recently as part of the Affinity Proteome and Affinomics consortia, and the German Federal Ministry of Education and Research (BMBF) within the NGFN PaCaNet project. S.L. was supported by a long-term fellowship of the Deutscher Akademischer Austausch Dienst (DAAD). C.B. was funded by a stipend of the Medical Faculty of Heidelberg University.

The authors have declared the following potential conflict of interest: C.S. and J.D.H. are co-founders of the DKFZ spin-off company Sciomics GmbH that provides services in the field of antibody microarrays. The remaining authors have declared no conflict.

References

- [1] The International Cancer Genome Consortium, International network of cancer genome projects. *Nature* 2010, 464, 993–998.
- [2] Kim, M.-S., Pinto, S. M., Getnet, D., Nirujogi, R. S. et al., A draft map of the human proteome. *Nature* 2014, 509, 575–581.
- [3] Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A. et al., Mass-spectrometry-based draft of the human proteome. *Nature* 2014, 509, 582–587.
- [4] Fagerberg, L., Hallström, B. M., Oksvold, P., Kampf, C. et al., Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Prot.* 2014, 13, 397–406.
- [5] Strathmann, F. G., Hoofnagle, A. N., Current and future applications of mass spectrometry to the clinical laboratory. *Am. J. Clin. Pathol.* 2011, 136, 609–616.
- [6] Meng, Q. H., Mass spectrometry applications in clinical diagnostics. *J. Clin. Exp. Pathol.* 2013, S6, e001.
- [7] Poste, G., Bring on the biomarkers. *Nature* 2011, 469, 156–157.
- [8] Syafrizayanti, Betzen, C., Hoheisel, J. D., Kastelic, D., Methods for analysing and quantifying protein-protein interactions. *Exp. Rev. Prot.* 2014, 11, 107–120.
- [9] Schmidt, R., Jacak, J., Schirwitz, C., Stadler, V. et al., Single-molecule detection on a protein-array assay platform for the exposure of a tuberculosis antigen. *J. Prot. Res.* 2011, 10, 1316–1322.
- [10] MacBeath, G., Schreiber, S. L., Printing proteins as microarrays for high-throughput function determination. *Science* 2000, 289, 1760–1763.
- [11] He, M., Taussig, M. J., Single step generation of protein arrays from DNA by cell-free expression and in situ immobilisation (PISA method). *Nucleic Acid. Res.* 2001, 29, e73.

- [12] Ramachandran, N., Hainsworth, E., Bhullar, B., Eisenstein, S. et al., Self-assembling protein microarrays. *Science* 2004, *305*, 86–90.
- [13] Angenendt, P., Kreuzberger, J., Glokler, J., Hoheisel, J. D., Generation of high density protein microarrays by cell-free in situ expression of unpurified PCR products. *Mol. Cell. Proteomics* 2006, *5*, 1658–1666.
- [14] Li, B., Jiang, L., Song, Q., Yang, J. et al., Protein microarray for profiling antibody responses to *Yersinia pestis* live vaccine. *Infect. Immun.* 2005, *73*, 3734–3739.
- [15] Quintana, F. J., Fareza, M. F., Viglietta, V., Iglesias, A. H. et al., Antigen microarrays identify unique serum autoantibody signatures in clinical and pathologic subtypes of multiple sclerosis. *Proc. Natl. Acad. Sci. USA* 2008, *106*, 18889–18894.
- [16] Patwa, T., Li, C., Simeone, D. M., Lubman, D. M., Glycoprotein analysis using protein microarrays and mass spectrometry. *Mass. Spectrom. Rev.* 2010, *29*, 830–844.
- [17] Sachse, R., Dondapati, S. K., Fenz, S. F., Schmidt, T., Kubick, S., Membrane protein synthesis in cell-free systems: from bio-mimetic systems to bio-membranes. *FEBS Lett.* 2014, *588*, 2774–2781.
- [18] Brennan, D. J., O'Connor, D. P., Rexhepaj, E., Ponten, F., Gallagher, W. M., Antibody-based proteomics, fast-tracking molecular diagnostics in oncology. *Nat. Rev. Cancer* 2010, *10*, 605–617.
- [19] Schröder, C., Jacob, A., Tonack, S., Radon, T. et al., Dual-color proteomic profiling of complex samples with a microarray of 810 cancer-specific antibodies. *Mol. Cell. Proteomics* 2010, *9*, 1271–1280.
- [20] Wingren, C., Sandström, A., Segersvärd, R., Carlsson, A. et al., Identification of serum biomarker signatures associated with pancreatic cancer. *Cancer Res.* 2012, *72*, 2481–2490.
- [21] Mirus, J. E., Zhang, Y., Hollingsworth, M. A., Solan, J. L. et al., Spatiotemporal proteomic analyses during pancreas cancer progression identifies STK4 as a novel candidate biomarker for early stage disease. *Mol. Cell. Proteomics* 2014, *13*, 3484–3496.
- [22] Flajnik, M. F., Deschacht, N., Muyldermans, S., A case of convergence: why did a simple alternative to canonical antibodies arise in sharks and camels? *PLoS Biol.* 2011, *9*, e1001120.
- [23] Paweletz, C.P., Charboneau, L., Bichsel, V. E., Simone, N. L. et al., Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. *Oncogene* 2001, *20*, 1981–1989.
- [24] Akbani, R., Ng, P. K. S., Werner, H. M. J., Shahmoradgoli, M. et al., A pan-cancer proteomic perspective on The Cancer Genome Atlas. *Nat. Commun.* 2014, *5*, 3887.
- [25] Hoadley, K. A., Yau, C., Wolf, D. M., Cherniack, A. D. et al., Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* 2014, *158*, 929–944.
- [26] Olsson, N., James, P., Borrebaeck, C. A. K., Wingren, C., Quantitative proteomics targeting classes of motif-containing peptides using immunoaffinity-based mass spectrometry. *Mol. Cell. Proteomics* 2012, *11*, 342–354.
- [27] Bussom, S., Saif, M. W., Methods and rationale for the early detection of pancreatic cancer. *J. Pancreas* 2010, *11*, 128–130.
- [28] Schwenk, J. M., Igel, U., Kato, B. S., Nicholson, G. et al., Comparative protein profiling of serum and plasma using an antibody suspension bead array approach. *Proteomics* 2010, *10*, 532–540.