Biochimica et Biophysica Acta xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

Review

Protein biomarker validation via proximity ligation assays ☆

A. Blokzijl, R. Nong, S. Darmanis, E. Hertz, U. Landegren, M. Kamali-Moghaddam *

Dept. of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, SE-751 08 Uppsala, Sweden

ARTICLE INFO

Article history:
Received 10 May 2013
Received in revised form 15 July 2013
Accepted 29 July 2013
Available online xxxx

Keywords:
Biomarkers
Protein detection
Proximity ligation assay
Diagnostic

ABSTRACT

The ability to detect minute amounts of specific proteins or protein modifications in blood as biomarkers for a plethora of human pathological conditions holds great promise for future medicine. Despite a large number of plausible candidate protein biomarkers published annually, the translation to clinical use is impeded by factors such as the required size of the initial studies, and limitations of the technologies used. The proximity ligation assay (PLA) is a versatile molecular tool that has the potential to address some obstacles, both in validation of biomarkers previously discovered using other techniques, and for future routine clinical diagnostic needs. The enhanced specificity of PLA extends the opportunities for large-scale, high-performance analyses of proteins. Besides advantages in the form of minimal sample consumption and an extended dynamic range, the PLA technique allows flexible assay reconfiguration. The technology can be adapted for detecting protein complexes, proximity between proteins in extracellular vesicles or in circulating tumor cells, and to address multiple post-translational modifications in the same protein molecule. We discuss herein requirements for biomarker validation, and how PLA may play an increasing role in this regard. We describe some recent developments of the technology, including proximity extension assays, the use of recombinant affinity reagents suitable for use in proximity assays, and the potential for single cell proteomics. This article is part of a Special Issue entitled: Biomarkers: A Proteomic Challenge.

1. Introduction

The discovery and application of blood biomarkers has a central and expanding role in clinical medicine. The ability to identify disease processes in internal organs by way of a venous blood sample promises to profoundly contribute to the accuracy and effectiveness of future medicine. Whereas DNA or RNA level analyses may require access to the tissue affected by disease, sufficiently sensitive protein analyses may permit detection of diagnostic molecules simply by sampling blood. While genetic approaches have increased in importance due to improved insights and new technologies, protein biomarker testing is slated for an equally or even more dramatic development. Current protein biomarkers are valuable for prognosis [1], prediction [2,3], stratification [4] and active surveillance [5] of indolent disease. There is, however, a vast need for more and better protein biomarkers for use in a range of pathological conditions. During the recent 15 years numerous potential biomarkers have been described in the medical literature. Nonetheless, the annual number of FDA-approved biomarkers over the same period of time is below an average of two new protein biomarkers per year [6,7]. This fact highlights the pressing need for efficient means to translate initial discoveries into clinical use.

Several factors contribute to the low translational success rate. Studies are often conducted using small numbers of samples, providing insufficient statistical power. This problem becomes even more severe when diagnostic algorithms involving multiple analytes are evaluated. Initial findings also often fail to be validated in independent sample cohorts, as required to exclude chance findings, differences in e.g. sample pre-processing, or cohort bias as confounding factors. Another shortcoming is the choice of samples and controls. It is usual practice to limit studies to comparisons between affected individuals and healthy subjects, while subjects with conditions that represent clinically relevant differential diagnoses are lacking. Claims of predictive importance obviously need to be validated in longitudinal follow-up studies [8].

Attempts have been made to streamline biomarker development by establishing a pipeline from initial discovery to clinical application [9,10]. Such a pipeline is intended to include the successive steps required to establish new biomarkers, along with a framework of analytical and statistical methods that can be used at each step to avoid the problems discussed above through proper study design.

Most protein biomarker studies have focused on the blood proteome, believed to be the most complete human proteome. The expectation is that the protein content of blood may include information about both normal and pathological processes taking place anywhere in the body. If we could properly capture this information we might be able to monitor a broad range of biological activities, anywhere in the

☆ This article is part of a Special Issue entitled: Biomarkers: A Proteomic Challenge.

* Corresponding author at: Dept. of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Box 815, SE-751 08 Uppsala, Sweden. Tel.: +46 18 471 4454; fax: +46 18 471 4808.
E-mail address: masood.kamali@igp.uu.se (M. Kamali-Moghaddam).

1570-9639/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.bbapap.2013.07.016

body, using minimally invasive procedures. A major limiting factor for such comprehensive and highly resolving analyses lies in the performance of current protein detection assays. The immense complexity of the blood proteome in terms of numbers of protein constituents, and the large differences in abundance among these, represent formidable challenges for protein detection methods [11]. Ideal methods need to meet the following requirements: 1) High sensitivity and specificity for distinguishing specific proteins in complex matrices; 2) Potential for multiplexing and high throughput in order to allow simultaneous analysis of panels of targets over broad dynamic ranges in large sample cohorts; 3) Capability of interrogating small sample volumes, in particular in studies such as for analyses of biobank samples, in pre- or neonatal care, cerebrospinal fluid analysis, bone marrow sampling, and in general where limited amounts of individual samples are available; 4) Flexibility of assay reconfiguration for investigating different protein isoforms and post-transitional modifications (PTMs) in different sample types; and 5) Versatility of readout format to suit diverse requirements for research, in large hospital settings, in primary health centers, and ultimately in patients' homes [12].

2. Discovery of new biomarkers

Identification of new biomarkers most often commences with a discovery phase, during which an initial, large set of proteins is measured in relevant clinical material. Any interesting protein biomarker candidates that are identified can then be further validated in subsequent steps. Two general approaches are adopted at this stage, namely targeted and untargeted biomarker discovery. In the untargeted approach, the goal is to investigate the entirety of accessible proteins using methods that do not require any prior selection of targets, such as those based on shotgun mass spectrometry (MS) or 2D gel electrophoresis. Identified targets with promising characteristics as biomarkers are selected for further validation in independent and larger sample collections, using assays developed for that purpose. Untargeted approaches are also referred to as unbiased since discovery of new putative biomarkers is performed without any prior assumptions as to what constitutes promising markers. However, methods used for untargeted analyses tend to be relatively insensitive and cannot handle the immense complexity of the blood proteome, in practice limiting analyses to more abundant proteins that represent only a small fraction of the proteome [10]. Accordingly, untargeted analyses may be more meaningful using material other than blood such as fluids near affected tissues (proximal fluids) or tissue homogenates, in order to establish a first set of protein targets that may be further evaluated in blood using more sensitive methods. By analyzing proximal fluids, proteins released from the tumor may be efficiently investigated before they are diluted or degraded while in transit to the blood stream.

Targeted approaches make use of prior knowledge to generate a set of proteins to be evaluated as biomarkers. Since this approach focuses on selected proteins, ranging in numbers up to a hundred or so, the choice of targets becomes crucial. Selection can be based on factors such as differential expression of proteins based on existing genomic, transcriptomic or proteomic data, or knowledge about properties of the proteins. In the following part, we will discuss a few criteria that might be valuable to consider when generating lists of specific proteins to target.

Since carcinomas arise because of malignant transformation of epithelia from organs whose cells have as one of their functions to secrete molecules to the lumen, secreted proteins may end up interstitially upon invasive growth of malignant cells. Proteins that are normally subject to exocrine secretion can thus be considered as possible markers since increased concentrations in blood may signal malignant transformation of the corresponding organ [13] The prostate cancer marker prostate specific antigen (PSA), normally exhibiting a million-fold concentration difference between seminal fluid and blood, belongs to this category [14,15]. There is also a speculation that it may be of help to identify genes selectively expressed during embryonal development but not in adult tissues, as these may encode oncofetal or carcinoembryonal proteins that become expressed anew in poorly differentiated tumors. Such proteins could therefore represent promising biomarkers. Results of RNA sequencing experiments from embryonic tissues could help identify such genes.

Proteins with exclusive expression in a single or just a few tissues would be expected to have a better chance to identify disease processes in those tissues than more generally expressed proteins. Lists of proteins with highly restricted tissue expression are of interest also because such gene products have an increased risk of being targets for autoimmunity. It is not uncommon for proteins with restricted expression to be highly expressed in the relevant tissue, further increasing their utility as markers.

Another strategy to find valuable clinical protein markers is to investigate variants of already known markers in order to more precisely capture clinically relevant variation. Such protein variants might arise as a consequence of variable start of transcription or translation, differential splicing, protein processing, or PTMs, or the biomarker variants may represent covalent or noncovalent complexes with interacting proteins [16].

Yet another example of protein variants that may be of interest diagnostically are ones subject to activating mutations in malignancy. For example, a small number of Ras mutations are frequently observed in a broad range of tumors, and BCR-ABL fusion proteins are a regular occurrence in chronic myelogenous leukemia, recommending these aberrant proteins as biomarkers.

Finally, the mere act of improving the limit of detection of some known biomarker protein has the potential to greatly extend their utility, e.g. by permitting earlier diagnosis or more clearly distinguishing cases from controls.

3. Technologies for validation of biomarkers

Proteins identified as biomarkers can be detected in body fluids such as blood, cerebrospinal fluid (CSF), or urine. Irrespective of the approach used during the discovery phase, promising biomarker candidate proteins are selected and subjected to subsequent validation in larger sample collections. In this section we will briefly discuss technologies used during this stage such as MS and sandwich immunoassays. In addition we will discuss the role of proximity-based assays, a novel class of protein detection assays, in biomarker validation.

4. MS

MS is by far the most powerful technology for investigating thousands of factors from various chemical classes, such as lipids, sugars or proteins, in a single run. However, as discussed limitations of sensitivity and dynamic range render untargeted MS unsuitable to trace low abundance proteins in e.g. plasma samples. Instead, targeted MS approaches such as selected/multiple reaction monitoring SRM/MRM are attracting increasing interest. In these methods signature peptides, representing proteins of interest, are selectively targeted. By comparing these peptides to externally added versions that include stable isotopes, precise quantitation is possible. The reactions can be multiplexed to some degree, and the sensitivity of the assays may be further augmented, by using antibodies to enrich experimental and control peptides from the samples. Recent advances and implementations of SRM/MRM are reviewed by Picotti and Aebersold [17]. Table 1 presents a comparison between protein marker detection by MS, sandwich immunoassay, and PLA.

5. Sandwich immune assays

Immunoassays are the most commonly used protein detection assays, and among these the sandwich ELISA has achieved great popularity in clinical practice. These assays are most often performed using an immobilized capture antibody, while the captured target protein is revealed via binding of second antibodies with coupled enzymes, giving

Please cite this article as: A. Blokzijl, et al., Protein biomarker validation via proximity ligation assays, Biochim. Biophys. Acta (2013), http://dx.doi.org/10.1016/j.bbapap.2013.07.016