Biochimica et Biophysica Acta xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

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

Review

Toward an integrated pipeline for protein biomarker development ☆

Andrei P. Drabovich [a,b,*,1], Eduardo Martínez-Morillo [a,1], Eleftherios P. Diamandis [a,b,c,d,1]

[a] Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON, Canada
[b] Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada
[c] Department of Clinical Biochemistry, University Health Network, Toronto, ON, Canada
[d] Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, ON, Canada

ARTICLE INFO

Article history:
Received 29 May 2014
Received in revised form 8 August 2014
Accepted 4 September 2014
Available online xxxx

Keywords:
Protein biomarker
Biomarker discovery
Biomarker verification
Biomarker validation
Proteomics
Mass spectrometry

ABSTRACT

Protein biomarker development is a multidisciplinary task involving basic, translational and clinical research. Integration of multidisciplinary efforts in a single pipeline is challenging, but crucial to facilitate rational discovery of protein biomarkers and alleviate existing disappointments in the field. In this review, we discuss in detail individual phases of biomarker development pipeline, such as biomarker candidate identification, verification and validation. We focus on mass spectrometry as a principal technique for protein identification and quantification, and discuss complementary -omics approaches for selection of biomarker candidates. Proteomic samples, protein-based clinical laboratory tests and limitations of biomarker development are reviewed in detail, and critical assessment of all phases of biomarker development pipeline is provided. This article is part of a Special Issue entitled: Medical Proteomics.

1. Introduction

Biomedical and translational science literature widely claims that molecular markers will revolutionize diagnosis and prognosis of almost every disease, including cancer, neurodegeneration and cardiovascular diseases. Such expectations arise mainly from recent exciting developments in the high-throughput -omics technologies which are set to analyze expression of every human gene. Increased availability of -omics technologies makes them very attractive to search for biomarkers and, as a result, leads to a steadily increasing number of biomarker discovery studies. The number of publications which report on putative disease biomarkers is continuously increasing, while the number of novel Food and Drug Administration (FDA)-approved protein biomarkers remains very low [1–3]. Indeed, no major cancer biomarker for screening or early diagnosis has been approved in the last 25 years [4]. In fact, the lately approved ovarian cancer biomarker, human epididymis protein 4 (HE4), is intended for either monitoring cancer recurrence [5] or prediction of malignancy along with CA-125 [6], but not for early diagnosis. With the underestimated difficulties of biomarker discovery and development, many overstated expectations are followed by later disappointments in the actual progress in the field [4]. Rational design and implementation of individual phases as parts of an integrated pipeline should facilitate systematic development of protein biomarkers and may soon bring new successful stories to the field and alleviate existing disappointments.

2. Proteins as biomarkers

Various classes of molecules may be considered as potential disease biomarkers. Advantages of proteins as a class include their enormous diversity, dynamic turnover and secretion into blood and bodily fluids. There is an estimated number of 20,300 genes [7], 40,000 unique metabolites [8], ~100,000 mRNA transcripts, and up to 1.8 million of different proteoforms, if post-translational modifications are considered [9]. Such enormous diversity of proteoforms increases chances to identify a marker, or a panel of markers, for each disease state. Since protein sequences may also reflect some genomic variations, a single instrumentation platform of mass spectrometry can measure not only changes in protein abundance, but also genomic and transcriptomic variations, such as mutant proteins

Abbreviations: AQUA, absolute quantification; ELISA, enzyme-linked immunosorbent assay; FDA, (the US) Food and Drug Administration; iTRAQ, isobaric tags for relative and absolute quantification; LC–MS/MS, liquid chromatography–tandem mass spectrometry; LOD, limit of detection; LOQ, limit of quantification; m/z, mass-to-charge-ratio; PSA, prostate-specific antigen; PSAQ, Protein Standard Absolute Quantification; PTM, post-translational modification; QconCAT, quantification concatemer; SCX, strong cation exchange; SILAC, stable isotope labeling by amino acids in cell culture; SNP, single nucleotide polymorphism; SRM, selected reaction monitoring

☆ This article is part of a Special Issue entitled: Medical Proteomics.

* Corresponding author at: Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Joseph & Wolf Lebovic Ctr., 60 Murray St, Box 32, Flr 6, Rm L6-201, Toronto, ON M5T 3L9, Canada. Tel.: +1 416 586 4800x8805.
E-mail address: andrei.drabovich@utoronto.ca (A.P. Drabovich).

[1] A.P.D. and E. M-M. wrote the manuscript, and all authors contributed to the revision of the manuscript.

http://dx.doi.org/10.1016/j.bbapap.2014.09.006
1570-9639/

or alternatively spliced proteoforms. Finally, proteins secreted into blood and body fluids can be measured with the minimally invasive tests.

Enormous protein diversity, however, poses the analytical challenge of detecting a specific protein in complex biological matrices. For example, detection of a particular nucleotide in the genome of a human cell should meet the analytical challenge of searching through 3.2×10^9 nucleotides, while detection of a certain amino acid in interleukin 6 sequence in blood plasma has the challenge of searching through 10^{13} amino acids [10]. Use of post-translational modifications as biomarkers would be an even more challenging undertaking due to the even higher complexity and dynamic turnover. For those reasons, typical protein biomarker pipelines are still focused on discovery of consensus protein sequences with differential abundance in disease rather than on discovery of differences in proteoforms or post-translational modifications.

3. Addressing unmet clinical needs with protein biomarkers

Unmet clinical needs, the intended use of biomarkers and their potential to facilitate the medical decision making in combination with concurrent diagnostic procedures should be considered well in advance. Specific applications of biomarkers typically include diagnosis, screening, prognosis, disease monitoring, or prediction of the response to therapy [11].

The discovery of *diagnostic* biomarkers would benefit those diseases for which the correct diagnosis is clinically challenging. For example, there is no single diagnostic test for Alzheimer's disease, so its current diagnosis is based on several criteria including medical history, mental status testing and physical and neurological examinations [12]. Histological examination of post-mortem brain regions still remains the gold standard for Alzheimer's disease diagnosis [13]. Although two cerebrospinal fluid biomarkers, amyloid β-protein fragments 1–42 and tau protein, have been included in the diagnostic criteria for the symptomatic pre-dementia phase of Alzheimer's disease [14], new biomarkers are needed for diagnosis of asymptomatic preclinical phase [15] and for diagnosis through a minimally-invasive blood test [16]. Reduction of numbers of invasive biopsies and diagnostic costs are additional values of diagnostic biomarkers. Some diagnostic procedures, such as computed tomography (CT) scan, may not be readily available in small medical centers or remote areas. Blood-based biomarker tests will thus facilitate quick decision-making and ultimately reduce diagnostic costs.

Early detection of rapidly progressing fatal diseases, such as cancer, is needed to provide a sufficient time window for treatment. Identification of *screening* biomarkers for early detection of rare diseases with low prevalence in the population is very challenging and may not even be feasible for some cancers [17]. For example, potential screening biomarker for ovarian cancer should have specificity of at least 99.5% at 80% sensitivity, to provide a positive predictive value of 10% [18]. Biomarkers with *prognostic* value are needed to predict the disease outcome and prescribe relevant therapies. Likewise, *monitoring* of disease progression and *prediction* of therapy efficiency are other specific applications of biomarkers.

4. Biomarker development pipeline

Similar to drug discovery and development, biomarker development should be designed as a multiple step process. The ultimate purpose of the biomarker development pipeline (Fig. 1) is to assess as many candidates as possible and exclude ineffective markers as early as possible. Upon presentation of the unmet clinical needs, biomarker development includes identification of proteins in the relevant biological sample, qualification of biomarker candidates, verification of candidates in the independent set of samples, development of a pre-clinical assay, clinical validation, and, finally, assay approval by health agencies, such as the FDA. The cost and duration of the whole biomarker development pipeline may be as high and as long as the cost and duration of a drug discovery project. For instance, it took nearly 8 years since the discovery of human epididymis protein 4 (HE4), an ovarian cancer biomarker, to conduct all essential validation studies and receive an FDA approval [5,19]. Similar to HE4 story, heart failure biomarker interleukin 1 receptor-like 1 protein (also known as ST2) was discovered in 2003, while its clinical assay was cleared by the FDA in 2011 [20]. And it took almost two decades for tryptase, a serum-based biomarker of mastocytosis, to reach the clinic [21].

The first phase of a biomarker development project often involves *identification* of proteins in clinical samples. Even though mass spectrometry is the most powerful technique for protein identification, it still suffers from relatively poor capabilities for protein quantification. Various label-free and label-based approaches were introduced to equip global protein identification with quantification capabilities and thus facilitate selection of biomarker candidates. Following protein identification, biomarker *qualification* provides an evidence of association between protein abundance and the clinical outcome. Certain filtering criteria are usually applied to select a manageable number of candidates and proceed to the verification phase [22].

The aim of biomarker *verification* is to measure the most promising candidates in a large set of samples and exclude false candidates. ELISA and mass spectrometry-based selected reaction monitoring (SRM) [23] are commonly used assays for biomarker verification. Advantages of ELISA include low cost, high sensitivity and high-throughput measurement of proteins in biological fluids of high complexity, such as blood serum. SRM assays facilitate multiplex verification of medium- and high-abundance proteins for which immunoassays are not available and provide attractive multiplexing capabilities. Combination of immunoaffinity enrichment of proteins or peptides with mass spectrometry measurements resulted in SISCAPA [24], MSIA [25], iMALDI [26], and immuno-SILAC approaches [27] which advanced verification of novel protein biomarkers.

Biomarker verification is followed by development of a pre-clinical assay and biomarker *validation*. Proper validation includes measurement of each biomarker in hundreds of samples from multiple centers, blinded analysis, establishment of reference values and selection of clinically meaningful or surrogate endpoints [28,29].

Finally, a clinical-grade assay is developed and subjected for the approval by the FDA. The list of FDA-approved protein biomarker assays currently includes more than 200 proteins [30]. The majority of FDA-approved protein assays utilizes ELISA, and not a single mass spectrometry-based protein assay has been approved for clinical use yet [31]. In addition, there is no yet a single FDA-approved protein biomarker that has been discovered by mass spectrometry and proteomics. Recently approved cancer biomarkers, HE4 protein and PCA3 mRNA, were discovered by microarray-based differential transcriptomic approaches. Due to the long duration of biomarker development projects, we may soon witness the approval of protein biomarkers which were discovered by mass spectrometry and proteomics in the 2000s.

5. Biological samples for protein biomarker discovery

A variety of biological samples such as blood, proximal fluids, tissue samples, cell lines and laboratory animals (Fig. 2) are suitable for protein identification. Blood serum or blood plasma is routinely analyzed in the clinical laboratory due to their minimally invasive collection and systemic circulation. Being the fluid of choice from the clinician's and patient's perspectives, blood plasma, however, is the most challenging sample to analyze by proteomic techniques. Blood plasma proteins have a dynamic range of concentrations of more than ten orders of magnitude, with albumin and cytokines being the most and the least abundant proteins, respectively [3]. Such a wide dynamic range allows for identification of high- and medium-abundance proteins while low-

Please cite this article as: A.P. Drabovich, et al., Toward an integrated pipeline for protein biomarker development, Biochim. Biophys. Acta (2014), http://dx.doi.org/10.1016/j.bbapap.2014.09.006