Proteomic Approaches to the Discovery of Cancer Biomarkers for Early Detection and Personalized Medicine

Kazufumi Honda*, Masaya Ono, Miki Shitashige, Mari Masuda, Masahiro Kamita, Nami Miura and Tesshi Yamada

Department of Chemotherapy and Clinical Research, National Cancer Center Research Institute, Tokyo, Japan

*For reprints and all correspondence: Kazufumi Honda, Department of Chemotherapy and Clinical Research, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan.
E-mail: khonda@ncc.go.jp

Received July 29, 2012; accepted October 31, 2012

Cancer biomarkers for the early detection of malignancies and selection of therapeutic strategies have been requested in the clinical field. Accurate and informative cancer biomarkers hold significant promise for improvements in the early detection of disease and in the selection of the most effective therapeutic strategies. Recently, significant progress in the comprehensive analysis of the human genome, epigenome, transcriptome, proteome and metabolome has led to revolutionary changes in the discovery of cancer biomarkers. The Human Proteome Organization has launched a global Human Proteome Project to map the entire human protein set. The Human Proteome Project research group has focused on three working proteomic pillars—mass spectrometry-based, antibody-based and knowledge-based proteomics—and each of these technologies is advancing rapidly. In this review, we introduce the proteomic platforms that are currently being used for cancer biomarker discovery, and describe examples of novel cancer biomarkers that were identified with each proteomic technology.

Key words: cancer biomarker – personalized medicine – early detection – proteomics

INTRODUCTION

Despite progress in four important areas (imaging technology, surgical management, therapeutic modalities and molecular-targeted therapies), cancer has remained a leading cause of mortality in Japan. Biomarkers, when used for the early detection of cancer and selection of therapeutic strategy, are powerful tools that can improve the outcomes of cancer treatments and reduce cancer-related mortalities. In the past decade, significant progress in the comprehensive analyses of the human genome, epigenome, transcriptome, proteome and metabolome has led to the discovery of a variety of biomarkers that can be used to detect early-stage cancers and predict tumor progression, drug response, clinical outcome or some combination of these three trajectories (1). Moreover, bioinformatics technology based on computational science has greatly facilitated the discovery of biomarkers that are strongly associated with the pathophysiology of particular cancers. For example, multiple dynamic alterations—including protein phosphorylation, protein trafficking and localization, and protein–protein interactions—that have secondary effects are frequently observed in cancer cells. Specific post-translation modifications (e.g. phosphorylation) and/or the status (e.g. nuclear localization) of particular proteins in cancer cells may be meaningful in clinical situations as potential cancer biomarkers. Recently, proteomic analyses have greatly facilitated the comprehensive cataloging of protein expression profiles in not only cell lines but also clinical samples, including serum/plasma, urine, spinal fluid, synovial fluid and tissues. In this review, we focus on the proteomic approaches used in the discovery of...
cancer biomarkers, and describe the current state of these proteomic studies.

**PLATFORM FOR PROTEOMIC ANALYSIS**

Recently, and following the successful completion of the Human Genome Project, the Human Proteome Organization (HUPO) (http://www.hupo.org/) officially launched a global Human Proteome Project (HPP) (http://www.hupo.org/research/hpp/), which aims to map the entire human protein set (2). A systematic global effort will be necessary to achieve this goal with respect to protein abundance, distribution and subcellular localization, as well as with respect to protein interactions with other biomolecules and protein functions at specific time points. As a general experimental strategy, the HPP research group has focused on three working pillars—mass spectrometry (MS)-based, antibody-based and knowledge-based proteomic strategies. In this section, we describe the main proteomic platforms used for cancer biomarker discovery (2).

**MS-BASED PROTEOMICS**

Recently, MS has evolved to a level such that it can be used to assess the complexity of the human proteome (3). MS has played a critical role in protein biomarker discovery in research on cancer and other diseases.

Initially, proteomic studies were based on two-dimensional gel electrophoresis and subsequent MS. This sequential approach greatly facilitated the identification of peptide sequences in proteins that were present in differential abundance on gels (4–6). Subsequently, proteomic pattern analysis was developed as one approach to biomarker discovery. Mass spectra and pattern recognition have enabled researchers and clinicians to use bioinformatics to distinguish cancer patients from healthy subjects (7). One such MS method is surface-enhanced laser-desorption ionization time-of-flight MS (SELDI-TOF-MS). This technique can be used to analyze the masses of protein directly without digestion using the enzyme (8,9).

MS can be coupled with liquid chromatography separation (LC); this combination is called LC/MS. Typically, in an LC/MS analysis, whole proteins that are present in complex biological samples are digested to peptide fragments by enzyme, and LC/MS is then used to identify the thousands of proteins in the biological samples; these samples can be tissues, serum, plasma, urine or protein elution of cell lines. The LC/MS-based methods, which can be used to comprehensively analyze digested peptides, are collectively called shotgun proteomics (10). There are two major approaches to shotgun proteomics, and each method is designed for accurate quantification of the peptide amount. One method is a labeling method that uses non-radio isotope labels and the other is a non-labeling method. We developed a third method that uses non-labeling shotgun proteomics, named two-dimensional image-converted analysis of LC and MS (LC/MS 2DICAL) (11,12).

**ANTIBODY-BASED PROTEOMICS**

The human proteome consists of ~20 500 non-redundant proteins that are defined as representative isoforms from every protein-coding gene locus. There are ongoing efforts in the field of proteomics to generate specific antibodies that recognize each component of the human proteome. Antibody-based proteomics plays a pivotal role in the cancer biomarker discovery and validation pipeline; specifically, it facilitates high-throughput evaluation of cancer biomarkers and provides a logical strategy for the systematic generation and use of specific antibodies to explore the proteome. The Human Protein Atlas project has been set up to systematically generate specific antibodies on a global scale and to utilize these antibodies for studies of the corresponding proteins and protein isoforms.

The use of antibodies for protein profiling on a global scale is an intuitive approach that should facilitate the system examination of the cancer proteome; antibody-based approaches can be used in conjunction with a wide range of high-throughput assays such as immunohistochemistry (IHC) on tissue microarrays (TMAs) and protein microarrays (13–15).

TMA is a method of assembling multiple tissue samples from an individual paraffin block to simultaneously evaluate multiple biomarkers using IHC; TMA can potentially become an accelerated molecular method for using a large-scale library of antibodies to examine the association between molecular biomarkers and clinical outcomes (16).

Protein microarray formats can be divided into two major classes: forward-phase protein arrays (FPPAs) and reverse-phase protein arrays (RPPAs) (17). In the FPPA formats, the analytes are captured from the solution phase by a capture molecule, usually an antibody that is immobilized on a substrate and acts as a bait molecule. Each spot on an array contains one type of immobilized antibody or bait protein. Each array is incubated with one test sample (e.g. a cellular lysate or serum sample) that represents a specific treatment condition, and multiple analytes from that sample are measured simultaneously. In contrast, RPPA immobilizes an aliquot of an individual complex test sample (e.g. serum, plasma, protein elution from a tissue sample or a cell lysate) in each array spot. In RPPA formats, each array is incubated with one detection protein, which is usually an antibody, and a single analyte is measured and directly compared across multiple samples (18).

RPPAs are a direct descendant of the miniaturized immunoassay, and the RPPA technology platform is designed for quantitative, multiplexed analysis of specific protein modifications (phosphorylation, glycosylation, processing or some combination thereof), or total forms of cellular proteins from a limited amount of samples. In addition, RPPAs
can be used to interrogate samples of tissues, cells, plasma, serum or body fluids (19,20). Moreover, we have developed a quantitative assay that is based on RPPAs and fluorescent immunoassays (21,22).

**Knowledge-based Proteomics**

Molecular-level information has been accumulated on a large scale to understand the molecular biology of human cells and the causes of human diseases. The initial goals of these large-scale efforts focused on sequencing the entire human genome and mapping human transcriptomes. Recently, information about human proteomes has attracted increasing attention. The molecular and functional complexity of human proteomes poses challenges to researchers, and this complexity requires bioinformatic resources specifically aimed at capturing, integrating and maintaining up-to-date available knowledge. The HPP will deliver a comprehensive atlas of human proteins in their biological context. It will generate publicly accessible data and informational resources that will, in turn, support further exploration of the human proteome by basic and clinical scientists. The HPP will be built on a knowledge-based database to integrate information derived from the major technological pillars, which include shotgun and targeted MS and polyclonal and monoclonal antibodies (2). For knowledge-based proteomic approaches, the HPP working group has decided to draw upon the UniProtKB/Swiss-Prot (23), PRIDE (24), PeptodeAtras (25), GPMDB (26) and Human Protein Atlas (14) databases.

**Identification of Cancer Biomarkers via Proteomic Analysis**

**Plasma and Serum Biomarkers for Cancer**

Human plasma proteome analyses have the potential to significantly improve disease diagnosis and therapeutic monitoring. However, analysis of plasma and serum proteomes is a significant obstacle, as these samples contain extremely complex human-derived proteomes and other tissue proteomes as subsets. At the high abundance end of the protein-content spectrum, the concentration of serum albumin is 35–50 mg/ml; in contrast at the low abundance end, the normal range of interleukin 6 in plasma is 0–5 pg/ml. These two clinically useful proteins differ in plasma abundance by a factor of $10^{10}$. Given the low abundance in serum and plasma of known cancer biomarkers, proteomic technologies that provide sufficient depth of analysis for biomarker discovery are now urgently needed (27). Moreover, post-translational modifications of proteins are important in many biological processes, and post-translational changes have relevance to disease and cancer. We recently identified a novel and specific post-translational modification of alpha-fibrinogen in plasma samples from pancreatic cancers using 2DICAL; specifically, we found that a proline on alpha-fibrinogen was hydroxylated to form 4-hydroxylated alpha-fibrinogen. We generated a specific antibody that recognizes 4-hydroxylated alpha-fibrinogen, and validated the accuracy of assays based on this antibody for the diagnosis of pancreatic cancer. The levels of serum 4-hydroxylated alpha-fibrinogen were significantly higher in Stage-I patients with pancreatic cancer than in control subjects (11). We are currently developing a sandwich enzyme-linked immunosorbent assay (ELISA) system for prospective validation of 4-hydroxylated alpha-fibrinogen-positive cancer samples.

In addition, we used SELDI-TOF-MS to identify another plasma biomarker for the early detection of pancreatic cancer (28); the C-termini of the apolipoprotein A1 homodimer is specifically and identifiably modified in patients with pancreatic cancer and pancreatitis (29,30).

To validate the diagnostic accuracy of biomarker candidates, it is important that the verification studies are conducted with a very large number of samples. As the tool for the rapid validation of plasma and serum biomarkers, we developed an RPPA on which 380 plasma samples were individually printed. Those plasma samples included samples from healthy controls and patients with pancreatic cancer, benign pancreatic cysts and tumors, chronic pancreatitis, esophageal cancer, hepatocellular carcinoma, gastric cancer, colorectal cancer or cholangiocarcinoma (21). This tool can rapidly validate the usefulness of individual antibodies that recognize respective plasma components to decide whether an individual antibody (biomarker candidate) should be subjected to further development for use in clinical situations. We used a combination of 2DICAL and plasma-RPPA to successfully identify plasma and serum biomarkers that can be used for the early detection of colorectal cancers (Fig. 1). Plasma and serum protein profiling is also useful for identifying predictive biomarkers; predictive biomarkers are used to assess the efficacy and side effects of a particular therapeutic modality. Gemcitabine monotherapy is currently the standard of care for patients with advanced pancreatic cancer, but the occurrence of severe neutropenia and thrombocytopenia can sometimes be life-threatening. In order to predict the hematologic toxicities of gemcitabine, we used serum plasma profiling to assess the state of predictive biomarkers. We found that low levels of haptoglobin, a predictive biomarker, are associated with the gemcitabine-induced hematologic toxicities (31).

High-dose interleukin-2 (IL-2) induces durable therapeutic responses in a small subset of patients with metastatic melanoma or with renal cell carcinoma, but simple pretreatment predictors of response have not been identified. To identify predictive biomarkers of clinical response to IL-2 treatments, sera that were obtained from patients who were treated with high-dose IL-2 were collected and analyzed using a customized antibody array. Elevated levels of vascular endothelial growth factor (VEGF) or fibronectin in serum were identified as independent predictors of positive responses for high-dose IL-2 therapy (32).
IDENTIFICATION OF THERAPEUTIC TARGETS AND COMPANION BIOMARKERS

Results from genomic and proteomic studies are eagerly awaited for selecting patients, avoiding the use in non-targeted situation, and reducing the cost of development of the drug. One of the major contributions that proteomics has made to medical communities is the identification of potential targets for effective cancer treatments. Many cancers are characterized by particular alterations in certain signaling pathways, and the identification of a particular alteration in a particular patient could facilitate the selection of an effective therapy targeted at that specific pathway. For example, pathogenesis of non-small cell lung cancer often depends on the activation of the epithelial growth factor receptor (EGFR) signaling pathway, and gefitinib and erlotinib are used for patients who have clinical indicators of activated EGFR signaling (33,34). Profiles of protein expression and post-translational modifications are powerful tools for identifying specific therapeutic targets in cancers. Proteomic technology greatly facilitates the comprehensive analysis of protein expression and post-translational modifications.

One of the most important issues in oncogenesis with regard to post-translational modifications of proteins is aberrant phosphorylation; consequently, inhibition of aberrant

Figure 1. The pipeline for the discovery of a plasma biomarker. There are three phases in the discovery of a plasma biomarker: discovery, validation and assay development. In the discovery phase, MS/antibody-based proteomic approaches are used to screen candidates for plasma biomarkers of cancer. During the validation phase, reverse phase plasma arrays (RPPAs) containing 380 tightly printed plasma samples are used for rapid validation studies. Finally, during the assay development phase, biomarker based-assays are developed in concert with clinical trials for those biomarkers that were validated in the validation phase (assay development phase).
protein phosphorylation has been a focus of molecular-targeted therapies. It is important that the profile of protein phosphorylation within a tumor sample is determined in detail when selecting an appropriate therapeutic strategy. In addition, new molecular targets may be identified during analyses of the phosphorylation protein profiles of cancer cells.

Recently, proteomic analyses have given rise to new technologies for the analysis of phosphoprotein expression profiles. Knowledge of the players in signal transduction mechanisms has accumulated mainly through the study of individual molecules in specific pathways. The emergence of high-throughput transcriptome technologies has resulted in a more detailed and objective view of the downstream transcriptional changes that follow respective stimuli. However, many critical events involved in cellular response are mediated by changes in post-translational protein modification rather than changes in the pattern of transcription.

Development of global and quantitative methods for elucidating dynamic phosphorylation events is therefore essential for a systematic understanding of cancer biology. MS has matured to become a powerful technology for phosphoproteomic analysis. Olsen et al. (35) identified the site-specific phosphorylation dynamics in signal networks under EGFR signaling globally. They stimulated HeLa cells with EGF, and then quantified comprehensively the phosphorylation status of phosphopeptides that had been eluted from the cells. These assays involved a combination of the techniques including stable isotope labeling using amino acids in cell culture (SILAC), chromatography of strong-cation exchange (SCX) and titanium dioxide (TiO2) for phosphopeptide enrichment, and high-accuracy multistage MS. This technique allowed this group to detect 6600 phosphorylation sites on 2244 proteins and to determine the temporal dynamics of the peptide phosphorylation that followed stimulation with EGF (36). More recently, this group applied high-resolution MS-based proteomics to investigate the proteome and phosphoproteome of human cells over the course of the cell cycle, and quantified 6027 proteins, identified 20 443 unique phosphorylation sites and monitored dynamics of the phosphoproteins (35). Individual antibodies that recognize specific phosphopeptides have been generated by many individual scientists and companies to identify the phosphorylation status of particular proteins. RPPAs that employ site-specific antibodies for protein phosphorylations can directly measure the level of phosphorylated protein isoforms, and screens that combine RPPA and antibodies have been used to generate extensive phosphoproteomic profiles. As an alternative to conventional, target-oriented drug discovery, Sevecka et al. (37) reported that RPPA could be used to identify compounds on the basis of the state that they induce in a signaling network. Using a high-throughput strategy, this group screened 84 kinase and phosphatase inhibitors for the ability to induce different states under EGFR signaling.

The assay developed by Sevecka et al. (19,38) allowed these researchers not only to identify interesting compounds that were overlooked in target-oriented screens, but also to uncover functional connections between proteins in signaling networks. RPPAs were originally developed to quantitatively measure numerous proteins extracted from a small number of cells obtained from tissue microdissection. RPPAs have been also used with the samples that were resected from patients with ovary (39), lung (40), prostate (41) or head and neck carcinomas (42). VanMeter et al. used RPPA and microdissected samples from lung cancers to quantify six phosphorylation sites on EGFR to evaluate whether EGFR mutation status in vivo was associated with the coordinated phosphorylation of multiple specific phosphorylation sites on EGFR. This RPPA-based analysis revealed simultaneous increases in the phosphorylation of two EGFR residues (1148Tyr and 1068Tyr) and the decreases in phosphorylation of EGFR1045Tyr, of human epidermal growth factor receptor type 2 1248Tyr and of insulin receptor substrate (IRS-1) 612Ser. These researchers concluded that a higher proportion of the EGFR mutant carcinoma cells exhibited activation of the phosphatidylinositol 3-kinase/protein kinase B/mammalian target of rapamycin pathway through 1148Tyr and 1068Tyr of EGFR.

In theory, protein-specific affinity reagents that specifically recognize each and every human protein could potentially be generated based on genome sequence information, and such validated affinity reagents could be used to explore protein expression patterns in various normal and cancer tissues. The Human Protein Atlas (HPA) project has started to generate the protein-specific affinity reagents toward all human proteins (15). A major effort in the HPA program was to use affinity-purified mono-specific antibodies for immunohistochemical analysis of a standardized set of TMAs that comprises 48 normal tissues and 216 different cancer specimens, which represent 20 different cancer types (43). The results are being made available as high-resolution images in an Internet-based Human Protein Atlas (http://www.proteinatlas.org/). With the release of the Human Protein Atlas version 7.0, an important objective was reached with the inclusion of 10 118 protein-coding genes corresponding to over 50% of the 19 627 human entries as defined by UniProt. With this high-throughput protein profiling effort, previously unrecognized associations between novel molecules and certain tumor types can be identified and analyzed. In order to search for proteins expressed in human melanocytes and melanomas, Stromberg et al. (44) used an antibody-based proteomics strategy to screen for proteins expressed in a TMA that contained normal tissue and cancer tissues. They reported that the expression level of the syntaxin-7 protein was inversely correlated with clinical stage of malignant melanoma.

CONCLUDING REMARKS
Over the past two decades, concerted scientific efforts have aimed at mapping the human genome. Recently, proteomic
technologies—including MS-based, antibody-based and knowledge-based proteomics—have matured. These proteomic approaches have the potential to greatly facilitate the discovery of new cancer biomarkers that can be applied to clinically applicable assays. It is hoped that this branch of proteomics will accelerate the development of biomarkers that can be used for the early detection of cancer and personalized therapeutic strategies in clinical situations.

**Funding**

This work was supported in part by the National Cancer Center Research and Development Fund (23-A-38 and 23-A-11).

**Conflict of interest statement**

None declared.

**References**


