MINI-REVIEW

Implementation of Proteomics for Cancer Research: Past, Present, and Future

Parisa Karimi1*, Armin Shahrokni2, Mohammad R Nezami Ranjbar3

Abstract

Cancer is the leading cause of the death, accounts for about 13% of all annual deaths worldwide. Many different fields of science are collaborating together studying cancer to improve our knowledge of this lethal disease, and find better solutions for diagnosis and treatment. Proteomics is one of the most recent and rapidly growing areas in molecular biology that helps understanding cancer from an omics data analysis point of view. The human proteome project was officially initiated in 2008. Proteomics enables the scientists to interrogate a variety of biospecimens for their protein contents and measure the concentrations of these proteins. Current necessary equipment and technologies for cancer proteomics are mass spectrometry, protein microarrays, nanotechnology and bioinformatics. In this paper, we provide a brief review on proteomics and its application in cancer research. After a brief introduction including its definition, we summarize the history of major previous work conducted by researchers, followed by an overview on the role of proteomics in cancer studies. We also provide a list of different utilities in cancer proteomics and investigate their advantages and shortcomings from theoretical and practical angles. Finally, we explore some of the main challenges and conclude the paper with future directions in this field.

Keywords: Proteomics - cancer - biomarkers - mass spectrometry - bioinformatics

Asian Pac J Cancer Prev, 15 (6), 2433-2438

Introduction

Cancer is the leading cause of the death worldwide, which account for approximately 13% of all deaths each year (Ferlay et al., 2010). Only in 2013, a total of 1,660,290 new cancer cases and 580,350 cancer deaths are projected to occur in the United States (Siegel et al., 2013). Since 1990, the death for all cancers together has decreased by only 1% per year, a decline due to a combination of factors, including prevention specifically reductions in tobacco consumption, improvements in treatment of some cancers, and early detections of a few specific cancers (Edwards et al., 2010).

New treatments for several cancers decreased the death rates significantly; however, advances in cancer treatment and improvements outcomes have been modest during past decades, and current treatments represent limited efficacy against advanced cancers (Shahrokni and Karimi, 2012). Although drugs and surgery can be effective for cancers in early stages, they usually briefly extend survival of patients with metastatic cancers (Martin et al., 2010). In fact, advances in cancer treatment and improvements outcomes have been modest during past decades (Etzioni et al., 2003). A great deal of researches is invested to improve treatments for advanced cancers, since most people who develop cancer have advanced disease at the time of diagnosis. For instance, among patients with breast, colorectal, and lung cancers in the United States, 34%, 57%, and 72% have metastatic cancer at the time of diagnosis (Etzioni et al., 2003). Thus, early detection can significantly reduce cancer mortality (Etzioni et al., 2003).

The advantages of early detection of cancer are to identify cancer while still localized and curable, prevent the mortality, and reduce the morbidity and cost (Smith et al., 2002; Kamangar and Karimi, 2013; Karimi et al., 2013). In another word, early detection generally means more opportunities for intervention that ultimately lead to improvement in patient outcomes (Karimi et al., 2013). Although early detection of cancer is the goal of majority of researches in cancer field since decade ago, relatively few early detection approaches have been proven sufficiently effective and practical (Etzioni et al., 2003). Biomarkers are one of the key concepts of early detection of cancer (Pepe et al., 2001), and defined by National Cancer Institute (NCI) as “…the biological molecule found in blood, other body fluids, or tissues that is a sign of a normal or abnormal process, or of a condition or disease”. The ideal biomarker should be easily detectable, highly sensitive and specific for its

1Johns Hopkins Bloomberg School of Public Health, Baltimore, 2Memorial Sloan Kettering Cancer Center, Geriatric Oncology Service, New York, 3Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, VA, USA. *For correspondence: pkarimi@jhsph.edu
Protein microarrays are powerful tools to capture and measure proteins from biospecimen (Stoevesandt et al., 2009). A protein microarray typically consists of a small piece of plastic or glass coated with thousands of capture reagents (molecules that can grab specific proteins) (Chandra et al., 2011). By using this technology, scientists could isolate and study many potential biomarker proteins (Chandra et al., 2011). Protein microarrays can be miniaturized to contain tens of thousands of capture features arranged in a grid, each specific for a certain type of a protein (Mishra and Verma, 2010). Thus, they are considered a multiplexed device, and can test for multiple biomarkers simultaneously, which is essential for clinical use (Rusling et al., 2010).

Nanotechnology: Nanotechnology is the creation of manufacturing devices and components, ranged from 1 to 100 nanometers (Nicolini and Pechkova, 2010). Nanotechnology devices have the potential to significantly expand the capabilities of proteomics, e.g. addressing current limitations in selectively reaching a target protein in vivo through physical and biological barriers, detecting low abundance targets, and providing a “toolbox” to translate the discovery of protein biomarkers to novel therapeutic intervention (Hainaut and Plymoth, 2011).

Utilized Tools in Cancer Proteomics

Necessary equipment and technologies for cancer proteomics are as follow

Mass Spectrometry: Mass spectrometry (MS) is a novel and evolving technology that detect and quantify proteins in a complex biological matrix (Walther and Mann, 2010). It is usually coupled with liquid chromatography (LC) or gas chromatography (GC) to reach a better separation by including another dimension specifically for untargeted analysis (Maurer, 2010). Mass spectrometry methods are very precise, and could distinguish proteins that differ in composition by a single hydrogen atom that is the smallest atom (Gstaiger and Aebersold, 2009). Despite its potential, MS technologies are not yet capable to separate the complex protein mixtures from unprocessed human biospecimens (Wang et al., 2009). Other technologies such as organelle, protein fractionation, and affinity capture have been developed to reduce the complexity of proteins in biospecimens by enriching for a subset of proteins of interest (Boja and Rodriguez, 2012). In addition, these technologies improve the sensitivity of the instruments for detection and quantification of proteins (Baker et al., 2012).
therapeutic and diagnostic tests (Collins et al., 2009). Typical nano-devices are including nanoparticles used for the targeted delivery of anticancer drugs, energy-based therapeutics, and imaging contrast reagents (Sadat Tabatabaei Mirakabad et al., 2014). Moreover, nanowires arrays can be used in biosensors that measure minute quantities of biomarkers in biological fluids (Ray et al., 2011).

**Bioinformatics:** The role of bioinformatics in cancer proteomics include data modeling and database design, data interoperability and comparison, gene and protein expression analysis, structural predictions, vocabularies and ontologies, and modeling for systems biology (Li et al., 2002). Thus, the development of new bioinformatics tools for integrative analysis of genomic and proteomic data is necessary to drive the collaborative, multidisciplinary effort required to drive discovery from the laboratory to clinical practice (Strassberger et al., 2010). Bioinformatics can improve the quality and accuracy of the study if it is involved from the initiation phase by developing, examining, and refining statistically sound hypotheses (Ressom et al., 2012). Moreover, it helps to investigate appropriate experimental designs and conclude the minimum needed sample size and population specifications in each study (Varghese et al., 2012). Also, it is the key to implementation of practical algorithms and realization of accurate hypothesis testing methods (Ressom et al., 2012). Thus, it is very critical to infer true differences, e.g. real protein biomarkers, across samples by using proper statistical learning and artificial intelligence approaches, which have been modified and adapted for the specific problem at hand (Ressom et al., 2012). In addition, this utility enables researchers from different areas to understand, communicate, and interpret data by employing advanced information retrieval systems and data visualization techniques (Nezami Ranjbar et al., 2013). Finally, bioinformatics is required for integration of data from different studies and technologies to achieve better understanding of the underlying biological phenomena (Gortzak-Uzan et al., 2007). This can be done by utilization and development of network-based data analysis. For example, there are several tools to study the known pathways for different types of interactions, while protein-protein interactions (PPI) are not the only case of biomolecule interactions. There are still many other different types of interactions such as gene-protein or protein-metabolite interactions. However, at the present time, there are quite few tools that provide integration of multiple types of interaction networks (Rual et al., 2005).

**Biospecimens:** Current cancer research has come to rely heavily on the quality of biospecimens for the measurement of genetic and protein expression, and the linkage of that data with clinical findings (Lopez et al., 2011). Since cancer diagnosis and treatment usually begin with diagnostic biopsies followed by surgical resection of the tumor, there are many opportunities to collect valuable biospecimens that are useful in research (Ransohoff and Gourlay, 2010). Also, the use of less invasive approaches is increasing by collecting sera or other body fluids to use with more recent measurement technologies such as GC/LC-MS (Baker et al., 2012). In all cases above, one important key to get precise and reproducible results is performing sample collection, preservation, and preparation consistently for all biological replicates; otherwise, the introduced bias in or across experiments leads to unreal or meaningless results. In addition, by including quality control, samples are useful to monitor the analytical and technical variability in experiments (Boja and Rodriguez, 2012). For example, in GC/LC-MS studies, replicates of a pooled mixture of samples along with spiked in standard compounds with controlled concentrations are used to investigate the reproducibility of the measurements.

**Reagents:** There is a growing need in the field of proteomics for high-quality, well-characterized, and standard reagents that can improve the specificity and reproducibility of proteomic technologies (Paul et al., 2013). One widely used reagent in proteomic research is

### Table 2. Some of Cancer Biomarkers Discovered by Proteomics

<table>
<thead>
<tr>
<th>Cancer type</th>
<th>Biomarkers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast cancer</td>
<td>Fibrinogen A Fragment; S100A9; 21- protein signature; GCDFP-15 AAG; PARK7; S10A7; GD1R; DDAH1; DDAH2; Versican core protein precursor; AGR2; Ubiquitin; Ferritin light chain; CD13; OSF-2; RS/DJ-1 autoantibody</td>
</tr>
<tr>
<td>Esophageal cancer</td>
<td>Peroxiredoxin VI autoantibody; Alpha -actinin 4; 67 ku laminin receptor; Rho GDP dissociation inhibitor 2; alpha-enolase; Lamin A/C; nucleodis-chalcone kinase A</td>
</tr>
<tr>
<td>Gastric cancer</td>
<td>a1- antitrypsin precursor; Pepsinogen C; Cathepsin B; MABWP; Vimentin; galectin 1; DEAD-box protein 48 autoantibody; hnrNP A2/B1</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>TEF1 a; A 25-sign Proteinomic Signature; Autoantibodies against triosephosphate isomerase and superoxide dismutase (MsSOD); HSP27; Aminopeptidase-P; eLF-5A; 15 distinct MS peaks; PGP 9.5 autoantibody</td>
</tr>
<tr>
<td>Liver cancer</td>
<td>HOP, hnrNP C1/C2; eLF1A; Multiplex serum markers; Ferritin-light-unit; Adenylate kinase-3a-like 1; biliverdin reductase B; Tissue ferritin light chain; V10 fragment of vitronectin; Brain-derived neurotrophic factor</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>CCSA-2; CCSA-3; CCSA-4; SELENBP1; HSP-27; GST; Annexin II; L-FABP</td>
</tr>
</tbody>
</table>

DOI:http://dx.doi.org/10.7314/APJCP.2014.15.6.2433

Implementation of Proteomics for Cancer Research: Past, Present, and Future

Achievements of Proteomics in Cancer Research

Over the past few years, a large number of cancer biomarkers have been discovered by cancer proteomic studies (Tan et al., 2012). Table 1 summarizes some of cancer biomarkers discovered by proteomics.

Challenges and Future Directions

Proteomics has already delivered significant achievements to understand the cancer, and to identify proteins of potential interest for diagnosis and treatment (Srinivas et al., 2002). However, there are many challenges ahead that should not be underestimated. First, the proteome is highly complex, and current equipment and technologies cannot yet provide a definitive solution for its exploration (Wulfkuhle et al., 2003). Moreover, cancer is a multifactorial and diverse disease, so, a great deal of time and effort will be necessary to define its associated proteome modifications and to translate these into practical applications for the clinic (Wulfkuhle et al., 2003). However, there are some issues that have not been addressed completely. While running few samples is an easy task with many available technologies, it is still not convenient to perform experiments with tens or hundreds of samples to achieve a reliable data set with appropriate sample size from a statistical analysis point of view. The problem rises as preparing many samples at the same time for a long experiment may not lead to consistent results. Also, preparing samples batch-wise can introduce analytical variability to the experiment. This means that there is tradeoff between sample size and reproducibility of the measurements. One of the main challenges is verification and quantitation of proteins found by discovery step as potential biomarkers. Similarly, identification is still a challenging task for proteomic studies. Even for known compounds, many data base search approaches are not mature enough to come up with confident matches (Karimi et al., 2013). In addition, identification of unknown compounds, i.e. the biomolecules with no good match in the database, still remains as a rigorous procedure, especially when conducting untargeted experiments with lots of possible candidates. On the other hand, using isotope-labeled standards with known compound in targeted studies is not easy in practice. As mentioned before, one way of inferring biological functions is to take advantage of biomolecule interactions by building, developing, and searching biological interaction networks. However, to reveal true underlying interactions, it is required to remove partial correlations between pairs of biomolecules connected through different pathways. Moreover, there are always some missing or hidden intermediate data, which makes the inference more complicated and less accurate. Finally, if even all these issues are addressed for protein-protein interactions, there is still a more difficult and yet less explored challenge: how to integrate data from different types of omics studies (Ressom et al., 2012).

As a conclusion, proteomics remains a maturing field, but there are substantial reasons to be optimistic about its ability to deliver significant values to cancer research in the near future (Hanash et al., 2008). The fast rate of developments in proteomic technologies and their role in cancer research is increasing steadily (Ferrari, 2005).

References


Kamangar F, Karimi P (2013). The state of nutritional epidemiology: why we are still unsure of what we should eat?. Arch Iran Med, 16, 483-6.


Maurer HH (2010). Perspectives of liquid chromatography coupled to low-and high-resolution mass spectrometry for screening, identification, and quantification of drugs in clinical and forensic toxicology. Ther Drug Monit, 32, 324-7.


