Global Analysis of Proteomics for Discovery of Biomarkers in Hepatocellular Carcinoma

Hong Li¹, Jiebingi Yuan¹, Fang Zhai¹, Jianqiang Zhang¹, Hua He² and Xuefeng Xia¹,³*

¹College of Pharmacy, Chongqing University, Chongqing 401331, China
²Institute for Viral Hepatitis, Ministry of Education, Chongqing Medical University, Chongqing 400010, China
³Houston Methodist Research Institute, Houston, Texas 77030, USA

Abstract

Hepatocellular Carcinoma (HCC) is a most lethal malignancy. At first presentation, patients often already have advanced disease, and their treatment options for cure are very limited. Measurable biomarkers for early diagnosis of HCC are urgently needed to prolong the median over-all survival rate and reduce therapeutic expenditures. Proteomics is a powerful analytical technique that has been widely applied to search for cancer biomarkers, including in HCC. High-sensitivity, high-throughput, and non-invasive technologies should be capable of contributing to early diagnosis of HCC. In this review, we provide a flow scheme for proteomic study in HCC, and an overview of technical approaches for protein quantification. In addition, we present a broad summary of the HCC biomarkers proposed by studies in recent years, which have used various quantitative proteomic approaches.

Keywords: HCC; Proteomics; Protein quantification; Biomarker

Introduction

Hepatocellular Carcinoma (HCC) is a most aggressive and devastating cancer, the third leading cause of cancer-related mortality worldwide, with 0.6 million new cases annually [1,2]. In developing countries, HCC accounts for 84% of the world total incidence and 83% of the world total death rate [3]. The rate of HCC incidence is increasing, partly due to the maturation of persons infected with the hepatitis C and B viruses, but also changes in lifestyle that lead to chronic alcohol abuse [4], non-alcoholic steatohepatitis, diabetes, and obesity [5,6]. Furthermore, more than 60% of patients have advanced stage disease, with metastasis, at the time of diagnosis [7], resulting in a very low overall 5-year survival rate (<16%) [8]. This is in contrast to the high 5-year survival rate (>93% with surgical intervention) when diagnosis is at early-stage, such as Barcelona Clinic Liver Cancer stage 0 and A [9].

Current methods to diagnose HCC include ultrasonography, computed tomography (CT), magnetic resonance imaging, and biopsy. While highly accurate, the biopsy procedure is painful. Diagnosis achieved through ultrasonography can be objective, as it depends on the size and character of the focal liver changes, but accuracy is also related to the experience of the operator and the quality of the equipment [10]. Focal liver lesions suspected on ultrasonography should be further confirmed with CT, magnetic resonance imaging, or both. Together, these methods can achieve a diagnosis of HCC with sensitivity and specificity of 89% and 99%, respectively [11]. Unfortunately, biopsy and imaging methods only detect disease that is already advanced, when nodes are obvious, and this means little to curative treatment. What is needed are non-invasive effective biomarkers to diagnosis HCC at the early stage.

The most widely used biomarker currently used in HCC is the protein alpha-fetoprotein (AFP) in blood. However, while elevated levels may indicate a potential liver disease or cancer, screening tests have such low sensitivity (at best, 60%) and specificity (being confused with intrahepatic cholangiocarcinomas or colon cancer metastases) that they cannot be relied upon for diagnosis of HCC, even in high-risk groups. At best, AFP may only indicate response to treatment. Therefore, AFP is not recommended in the current guidelines of the American Association for the Study of Liver Diseases [12].

Research efforts in finding suitable biomarkers for early diagnosis of HCC has turned toward the complex interactions of biological molecules, aided by increasingly refined maps of networks such as protein-protein interactions and signal pathways. In particular, approaches based on high-throughput proteomics offer a versatile platform to assay the disturbance of global proteins. Through associated differentiations, we may be able to find dysfunctional compounds that could lead to the discovery of biomarkers useful for the early diagnosis of HCC, as well as gain insight into the mechanisms driving oncogenesis. In the present review, we describe current strategies in proteomics for discovering useful biomarkers.

Development of proteomics toward HCC

In the mid-1970s, the development of analysis technology such as mass spectrometry (MS) and high-resolution two-dimensional gel electrophoresis (2DE) [13-15] enabled the study of proteins on a mass scale, the study now known as proteomics. Subsequent research focused on automated procedures that let biologists focus on data analysis. These efforts, primarily driven by work at the Argonne National Laboratory (United States), culminated in 1980 with establishment of the Human Protein Index Task Force. The purpose of the task force was to create the Human Protein Index (HPI) database, to catalog all the proteins in every human cell type.

The HPI project at first failed to attract large-scale support. This was partly because automated science was considered inappropriate in...
biology, and also because the possibility of sequencing entire genomes was not yet recognized. In 1996, Wilkins et al. [16] were the first to publish the concept of the proteome, defined as all the proteins produced through the genome. In the same year, the Australian government funded the Australian Proteome Analysis Facility, equipped with state-of-the-art technology dedicated to proteomic research [17].

In 1998, Chinese scientists began studying the proteome of the liver, and in March 2002, the Asian and Oceanian Human Proteome Organization launched the Human Liver Proteome Project (HLPP). Since 2006, the stated objectives of the HLPP have been the identification, characterization, and integration of the human liver proteome. Accomplishing these objectives entails creating expression and modification profiles, and maps of protein-protein interactions and proteome localization. In addition, the HLPP seeks to define the ORFeome (protein-encoding open reading frames), physiome (normal physiological dynamics), and pathome (differentially expressed pathways in pathogenesis) of the liver [18].

Completion of the Human Genome Project in April 2003 made reasonable the goal of creating a map of the entire human proteome map with all its networks of interactions, and interest in finding a biomarker of HCC intensified. Currently, the HLPP summarizes its mission focus as understanding the molecular mechanisms underlying liver function and disease. Proteomic research in HCC has included the search for suitable biomarkers, with some success achieved.

**General aspects of HCC proteomic research**

Proteomics is the large-scale study of proteins, especially their structure and function. In particular, proteomic research seeks to characterize biological processes, including disease and drug effects, by understanding the regulation and quantification of gene expression [19]. Thus, proteomic research can contribute greatly to our understanding of pathogenesis, as disease dysfunction is reflected in the differentiable genetic expression of proteins. Identifying aberrant proteins in fluids or HCC cells with high throughput proteomics is a powerful approach to search for HCC biomarkers. In this review we illustrate the ability of the proteomic platform to find HCC biomarkers, by concentrating on 25 proteins, each of which is involved in various functions such as apoptosis, ion transport, differentiation, and death.

The typical experimental workflow of a proteomic experiment in HCC is depicted in Figure 1. The first step is protein preparation, in which proteins are isolated from tissue or blood samples by protein labeling, laser capture microdissection (LCM), or subcellular fractionation, depending on the scientific question. The result is a mixture protein, and an additional fractionation step (protein separation) is required.

The method of protein separation depends on the goal of the research (Figure 1). For differences in protein ligand specificity, affinity chromatography is appropriate. To separate according to differences in molecular weight, dialysis and ultrafiltration, or gel filtration chromatography (GFC) may be chosen. Charged proteins may be resolved by electrophoresis, or ion exchange chromatography (IEC). To separate by solubility, techniques include salting out, isoelectric point precipitation (IPP), and organic solvent precipitation (OSP).

Subsequently, to identify proteins, we can combine a variety of methods such as image analysis, microsequencing, MS with peptide mass fingerprinting (PMF) of peptide fragments, or amino acid composition analysis (Figure 1). After identification, potential biomarkers are finally determined through protein informatics, by database matching or protein-protein interaction (PPI) mapping. Proteome informatics provides a way to understand the underlying pathways and the interactions between individual signature markers and non-markers. With further analysis of this information, we may obtain potential biomarkers for early diagnosis of HCC.

**Quantification strategies in HCC research**

In addition to identifying as many proteins as possible in a given sample, in proteomic research the quantification of these biomolecules is crucial to finding HCC biomarkers (Figure 2) [20]. Several methodologies have been developed to monitor quantitatively relative or absolute changes in protein levels.

**Figure 1:** Workflow of proteomics research investigating biomarkers in HCC.

---

**Figure 1:** Workflow of proteomics research investigating biomarkers in HCC.
important insights into the molecular mechanisms of particular diseases by allowing comparison of the types and amounts of proteins between diseased and normal cells and tissues. Furthermore, proteomic analysis via MS technologies and MS-based quantitative strategies can provide a more global and accurate view of dynamic biological processes. Normally, this method can be divided into two broad categories: labeling quantitation (in vivo labeling or in vitro labeling) or label-free quantitation.

Because of its simplicity, affordability, and accuracy, the most popular technique used in the lab for in vivo labeling quantitation is stable isotope labeling with amino acids in cell culture (SILAC) [21]. Chen et al. [22] utilized SILAC to study HCC metastasis mechanisms and potential predictive biomarkers of HCC metastasis. As SILAC cannot quantitatively analyze samples that cannot be cultured, Ishihama et al. [23] invented an alternative approach, based on SILAC, known as culture-derived isotope tags (CDITs). CDITs are now more commonly applied than SILAC in the search for HCC biomarkers. For example, Li et al. combined CDITs with 2D liquid chromatography-tandem MS (2D-LC-MS/MS) and concluded that APEX1 (apurinic/apyrimidinic endodeoxyribonuclease 1) and ANP32A (acidic [leucine-rich] nuclear phosphoprotein 32 kDa family member A) have potential as biomarkers of HCC [24].

According to labeling different parts of proteins, the current lab usually use techniques include isotope-coded affinity tag (ICAT). Kang et al. [25], used cleavable stable isotope labeling (cICAT) combined with LC-electrospray ionization tandem mass spectrometry (LC-ESI-MS/MS) to compare the serum proteome between liver cirrhosis and HCC patients. They reported that alpha-2-macroglobulin was downregulated and AGP (alpha-1-acid glycoprotein) was upregulated in serum. The results of this study showed the power of this method to find potential HCC biomarkers.

Although widely used, the prototypical ICAT technique has a number of limitations. These include missed identification of proteins with few or no cysteine residues, lost information for post-translational modifications, differential reversed-phase elution of identical peptides labeled with hydrogen/deuterium isotope pairs, and the complicated interpretation of tandem MS due to addition of the biotin group [26,27]. To a certain extent, isobaric tags for relative and absolute quantitation (iTRAQ) solve this problem [28]. The work of Zhou et al. [29] verified the iTRAQ method for investigation of HCC biomarkers, by finding that HSP90A (heat shock protein 90) levels were elevated in HCC cells, serum, and tissues.

Although SILAC is the most reliable method for quantitative MS, the preparation of isotope-labeled compounds is time-consuming and expensive. In recent years, label-free quantitative technology based on liquid chromatography tandem MS (i.e., label-free LC-MS/MS) and has been recognized as a viable alternative. Reis et al. [30] utilized LC MS/MS to find that elevated levels of 14-3-3 sigma were diagnostically accurate for HCC in hepatocytes, with rates of 73.2%, and 72.7% for specificity and sensitivity, respectively. Naboulsi et al. [31] applied the label-free platform to find that Versican was significantly associated with well differentiated and early-stage HCC; the area under the receiver operating characteristic curve (AUROC) was 0.85. However, label-free quantitative techniques are still in infancy, and there are no consistent international standards.

Full utilization of MS has provided us with many potential biomarkers for diagnosing HCC (Table 1). For example, a1AT, B2M, ERBB3, Fu-HPX, Prx-II, CC3-a, PHB2, G2890, G3560, and vimentin are upregulated in blood or tissues [32-40], and AFM, and CLU are downregulated in the plasma of HCC patients [41,42]. These biomarkers also indicate novel molecular mechanisms that may be targeted for treatment.

Protein quantification based on two-dimensional gel electrophoresis and staining: Two-dimensional gel electrophoresis (2DE) is a powerful method for protein abundance studies, and the only one available for simultaneous resolution of thousands of proteins. O’Farrell [14] first described the technique in 1975. In 2DE, proteins are separated according to their charge by isoelectric focusing or immobilized pH gradient (in the first dimension), and then by size via sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE; the second dimension). Sun et al. [43] utilized 2DE to find that APO A1 was downregulated in HCC patients’ sera and the fold change was 3.59 compared with normal healthy controls. Other researchers have concluded that AACT, GFAP, hCE1, LMNB1, ConA-pCD, HSP90, OPN, and Hs-AFP-L3 are upregulated in HCC patients [43-50], indicating their potential power for diagnosis.

Despite being well-established as a technique for protein analysis,
with MS has enabled researcher’s greater detail and flexibility in the analysis [54]. The coupling of gel-based classic proteomic approaches and MS has enabled researcher’s greater detail and flexibility in the analysis of the human proteome and finding HCC biomarkers [12]. In Table 2, we summarize studies that use mixed methods. These studies show that associations between certain proteins and HCC disease etiology or progression can be demonstrated with high sensitivity and specificity, and that these proteins are potential HCC biomarkers. The heterogeneous nature of these studies notwithstanding, their relevance to the search for HCC biomarkers cannot be questioned.

Due to the complexities of HCC etiology [55] and differences in clinical behaviors, no single protein is likely to have sufficient sensitivity and specificity for the detection of HCC, Particularly early HCC [10]. Rather, potential biomarkers may be considered in combination, to improve their efficiency. In Table 3, we summarize the recent research regarding biomarker combinations in HCC, with specificities and sensitivities, as a useful guide for present research. Among the biomarkers listed, combinations with AFP or AFP-L3 are particularly important in the diagnosis of HCC [56-58]. For example, Choi et al. [56] found that AFP-L3 combined with PIVKA-II had a sensitivity and specificity of 94.4% and 75.6%, respectively. Sun et al. [40] concluded

<table>
<thead>
<tr>
<th>Protein</th>
<th>Uniprot*</th>
<th>Sample</th>
<th>Cl. HCC</th>
<th>FC</th>
<th>AUROC</th>
<th>Spec</th>
<th>Sens</th>
<th>Platform</th>
<th>First Author, Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1AT</td>
<td>P01009</td>
<td>Plasma</td>
<td>↑</td>
<td>-</td>
<td>0.84</td>
<td>-</td>
<td>-</td>
<td>QTOF–LC–MS, ELISA</td>
<td>Fye, 2013</td>
</tr>
<tr>
<td>A2M</td>
<td>P01023</td>
<td>Plasma</td>
<td>↓</td>
<td>0.26</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>ICAT–LC–ESI–MS/MS</td>
<td>Kang, 2010</td>
</tr>
<tr>
<td>AFM</td>
<td>P43652</td>
<td>Plasma</td>
<td>↓</td>
<td>0.72</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>nUPLC–ESI–QTOF–MS &amp; TQMS</td>
<td>Lee, 2011</td>
</tr>
<tr>
<td>B2M</td>
<td>P61769</td>
<td>Plasma</td>
<td>↑</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>SELDI–TOF–MS</td>
<td>Nakatsu, 2010</td>
</tr>
<tr>
<td>CLU</td>
<td>P10909</td>
<td>Pl/ser</td>
<td>↓</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>SIC–MRM–MS</td>
<td>Zhao, 2010</td>
</tr>
<tr>
<td>ERBB3</td>
<td>P21860</td>
<td>Serum</td>
<td>↑</td>
<td>-</td>
<td>0.93/0.71</td>
<td>97%</td>
<td>71%</td>
<td>MALDI–TOF–MS, WB, ELISA</td>
<td>Hsieh, 2011</td>
</tr>
<tr>
<td>Fu–HPX</td>
<td>P02790</td>
<td>Plasma</td>
<td>↑</td>
<td>1.40</td>
<td>0.95</td>
<td>92%</td>
<td>92%</td>
<td>Lectin LC–MS/MS</td>
<td>Comunale, 2009</td>
</tr>
<tr>
<td>Ptx-II</td>
<td>P32119</td>
<td>Plasma</td>
<td>↑</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>MALDI–TOF–MS</td>
<td>Lu, 2010</td>
</tr>
<tr>
<td>CC3-a</td>
<td>P01024</td>
<td>Serum</td>
<td>↑</td>
<td>-</td>
<td>0.70</td>
<td>72-98%</td>
<td>41-77%</td>
<td>SELDI–TOF–MS</td>
<td>Kamnura, 2010</td>
</tr>
<tr>
<td>PHB2</td>
<td>Q99623</td>
<td>Tissue</td>
<td>↑</td>
<td>-</td>
<td>0.69</td>
<td>88%</td>
<td>41%</td>
<td>MALDI–TOF/TOF–MS</td>
<td>Cheng, 2013</td>
</tr>
<tr>
<td>Vimentin</td>
<td>P08670</td>
<td>Serum</td>
<td>↑</td>
<td>-</td>
<td>0.69</td>
<td>88%</td>
<td>41%</td>
<td>MALDI–TOF/TOF–MS</td>
<td>Sun, 2010</td>
</tr>
<tr>
<td>Versican</td>
<td>P13611</td>
<td>Tissues</td>
<td>↑</td>
<td>-</td>
<td>0.85</td>
<td>-</td>
<td>-</td>
<td>Label-free LC–M–MS</td>
<td>Naboulsi, 2016</td>
</tr>
<tr>
<td>G2890</td>
<td>-</td>
<td>Serum</td>
<td>↓</td>
<td>0.91</td>
<td>92%</td>
<td>83%</td>
<td>83%</td>
<td>MALDI–TOF–MS</td>
<td>Kamiyama, 2013</td>
</tr>
<tr>
<td>G3560</td>
<td>-</td>
<td>Serum</td>
<td>↑</td>
<td>0.85</td>
<td>89%</td>
<td>71%</td>
<td>71%</td>
<td>MALDI–TOF–MS</td>
<td>Kamiyama, 2013</td>
</tr>
</tbody>
</table>

*UniProt, Universal Protein Resource

Traditional 2DE is time-consuming and labor-intensive [51]. Two-dimensional fluorescence difference gel electrophoresis (2D-DIGE), which is based on traditional 2DE, overcomes these shortcomings well. The 2D-DIGE technique was first described by Jon Minden’s laboratory [52] and has subsequently been refined and marketed by Amersham Biosciences (GE Healthcare). Megger et al. [53] used 2D-DIGE to search for an HCC biomarker in HCC liver tissue, and concluded that betaine-homocysteine methyltransferase (BHMT) is upregulated in HCC liver tissue, and concluded that betaine-homocysteine methyltransferase (BHMT) is upregulated relative to non-tumorous liver tissue.

Quantitative measurement of differential protein levels, solely using MS, is not fully reliable due to the uneven ionization efficiency of peptides with different sequences—the signal intensity of various peptides in mass spectra is not usually proportional to their abundance. Therefore, 2DE is used commonly in conjunction with MS to measure relative differences in individual proteins, determined from the intensities of stained protein spots gained through image analysis [54]. The coupling of gel-based classic proteomic approaches with MS has enabled researcher’s greater detail and flexibility in the specificities and sensitivities of 94.4% and 75.6%, respectively. Sun et al. [40] concluded...
that the specificity of vimentin combined with AFP reached 98.2%. Kang et al. [25] showed an AUROC of 0.88 for the combination of AGP and AFP.

Other potential biomarkers have also shown impressive results. For example, Ahn et al. [59] reported a sensitivity of 100% for the combination of AACT and A1AT. Liu et al. [60] probed a combination of 5 proteins (CE, HRG, CD14, HGF, and C3), achieving a sensitivity of 79% and specificity of 72%, distinguishing early HCC from cirrhosis. The proteomic approach has thus proved useful in the search for biomarkers for early HCC diagnosis.

Conclusion

In this review, we have presented a flow chart of proteomics and the different quantitative methodologies applicable to identify protein alterations associated with HCC. We have also listed biomarker candidates gained by comparative 2DE and/or MS analysis of tissues or blood from HCC patients and various chronic liver diseases. Most researchers have proposed potential biomarkers without clinical verification. Therefore, putative biomarkers require clinical confirmation of sensitivity, specificity, reproducibility, and accuracy. However, because of the complex pathological mechanism of HCC, it is difficult to confirm a diagnosis of HCC with a single biomarker. Future studies should verify the insights from the existing literature and broaden them, defining the optimal mixture of surface markers that will identify HCC. Novel biomarkers can increase our understanding of oncogenesis, and may lead to better treatment strategies, with the ultimate goal of improving the prognosis of HCC patients.

References


