MASS SPECTROMETRY-BASED

CANCER BIOMARKER DISCOVERY

Olena V. Masui

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN CHEMISTRY
YORK UNIVERSITY
NORTH YORK, ONTARIO
CANADA

October 2012

Abstract

The aim of the projects in this thesis was to identify biomarkers for clear cell renal cell carcinomas (ccRCC) and head and neck/oral squamous cell carcinoma (HNOSCC), using quantitative or qualitative proteomics. Comparative analysis of cancerous and normal tissue homogenates, or secretome analysis of cancer cell cultures using liquid chromatography - mass spectrometry (LC-MS) and immunoassay techniques, allowed the identification of different types of biomarkers: diagnostic or prognostic, biofluid- or tissue-based.

Chapter 1 of this thesis provides general information on cancer and cancer biomarker discovery. Chapter 2 gives a brief introduction to the techniques used in this work and theories behind them. Chapters 3 – 5 are papers that resulted from the cancer biomarker discovery research performed here, and Chapter 6 contains the conclusions, the author's comments and the final remarks.

The papers on the identification of biomarkers for different diseases, to which the author of this thesis contributed, are listed in the Appendix.

Table of Contents

List of Figures

List of Abbreviations

14-3-3ζ	14-3-3 protein zeta/delta
14-3-3 σ	Stratifin
AHNAK	neuroblast differentiation-associated protein AHNAK
ccRCC	clear cell renal cell carcinomas
CE	collision energy
CEM	channel electron multiplier
CID	collision-induced dissociation
DC	direct current
EF	error factor
ENO1	alpha-enolase
ESI	electrospray ionization
FDR	false discovery rate
FTICR	Fourier transform-ion cyclotron resonance
Gal-1	galectin-1
GO	Gene Ontology
HNOSCC	head and neck/oral squamous cell carcinoma
hnRNPK	heterogeneous nuclear ribonucleoprotein K
HPLC	high-performance liquid chromatography
HRP	horseradish peroxidase
HSPB1	heat shock protein beta-1
HSPE1	10 kDa heat shock protein, mitochondrial

IDA	information-dependent acquisition
ICP	inductively-coupled plasma
IHC	Immunohistochemistry
IPA	Ingenuity pathway analysis
iTRAQ	isobaric tags for relative and absolute quantitation
LC-MS	liquid chromatography - mass spectrometry
LDHA	L-lactate dehydrogenase A chain
MALDI	matrix-assisted laser desorption ionization
MCP	microchannel plate
MS	mass spectrometry
MS1	precursor- ion scan or the first mass analyzer
MS2	product-ion scan of the first mass analyzer
OPLs	oral pre-malignant lesions
PBS	phosphate buffered saline
Pfn1	profilin-1
PIE	precursor ion exclusion
PPIA	peptidyl prolyl isomerase A/cyclophilin A
PSA	prostate specific antigen
PVDF	polyvinylidene difluoride
RCC	renal cell carcinoma
RF	radio frequency
RP	reverse phase
SDS-PAGE	sodium dodecyl sulfate - polyacrylamide gel electrophoresis

SCX	strong cation exchange
TMA	tissue microarray
TOF	time-of-flight
WB	western blotting

Acknowledgements

I would like to gratefully acknowledge the supervisory support and guidance of Professor K.W. Michael Siu, Director of the Centre for Research in Mass Spectrometry in the Department of Chemistry, York University. I would also like to thank Professor Imogen Coe, in the Department of Biology, and Professor Derek Wilson, in the Department of Chemistry, York University, for their support and encouragement as a members of my supervisory committee.

I would like to thank our collaborators, Professor George M. Yousef and Dr. Nicole M.A. White, in the Laboratory of Medicine and Pathobiology, and Professor Ranju Ralhan, in the Department of Otolaryngology - Head and Neck Surgery, at the University of Toronto, for greatly contributing to the projects in this thesis

Among my colleagues, I would like to extend my thanks to the past and present members of Professor Siu's research team, and especially acknowledge Dr. Leroi V. DeSouza, Dr. Udo Verkerk, Dr. Sebastien Voisin, Dr. Ajay Matta, and Ph.D candidate, John van Nostrand, for their support and direction.

I would also like to thank my husband, Hitoshi, for his love, support, and for proof-reading this thesis.

Chapter 1

Introduction to cancer biomarker discovery

1. 1. Cancer

Based on the GLOBOCAN 2008 estimates, about 12.7 million cancer cases and 7.6 million cancer deaths are estimated to have occurred in 2008 worldwide [1], which is, according to the Union for International Cancer Control [2], more than AIDS, tuberculosis, and malaria combined.

1.2. Causes

Global cancer rates have been increasing mainly due to an aging population and lifestyle changes in the developing world. The majority of cancers (90-95%) are due to environmental factors, while 5-10% are due to hereditary factors [3]. Environmental factors that contribute to cancer are tobacco (25-30%), diet (30-35%), infections (15-20%), ionizing and non-ionizing radiation (10%), stress, physical inactivity, physical injury, hypoxia and environmental pollutants [3, 4].

1.3. Pathophysiology

Cancer is a disease in which abnormal cells divide and grow uncontrollably, forming malignant tumors, and can invade adjacent tissues and organs [5]. This disease is characterized by dysregulation of cell growth caused by gene and protein mutation. For instance, in humans, the tumor-suppressor gene, p53, encodes protein 53 (p53), which

normally regulates cell growth by several mechanisms, one of which is the initiation of apoptosis, the programmed cell death that occurs when DNA damage appears to be irreparable [6]. Mutation of the p53 gene in prostate cancer cells inhibits their apoptosis [7], and mutation of p53 protein (e.g., acetylated lysine-120 replaced by a non-acetylated arginine) also inhibits apoptosis of cancer cells [8].

1.4. Type of cancers

The tumors that are at their original site where progression began and that are capable of invading adjacent tissues and organs are described as primary cancers. Cancerous cells can also spread (metastasize) to other organs through the blood vessels and lymph systems to form tumors that are described as metastatic or secondary cancers. Tumors that are not capable of invading surrounding tissues and adjacent organs and, in most cases, are harmless to human health, are called benign tumors [9].

There are many types of cancers: carcinomas are derived from epithelial cells of the skin or from tissues that line or cover internal organs; sarcomas arise from connective tissue in bone, cartilage, fat, muscle, blood vessels, or other connective or supportive tissue; leukemias originate in blood-forming tissue such as the bone marrow and then enter the blood; lymphomas and multiple myelomas begin in the cells of the immune system; central nervous system cancers begin in the tissues of the brain and spinal cord; germ cell cancers, derived from pluripotent cells, most often begin in the testes or the ovaries; and blastomas are derived from immature "precursor" cells or embryonic tissue.

The chances of surviving cancer vary greatly by the type and location of the tumors and the extent of the disease at the beginning of treatment.

1.5. Cancer staging and treatment

The stage of a cancer is a description of the extent to which the cancer has spread. The stage often takes into account the size of a tumor, how deeply it has penetrated, whether it has invaded adjacent organs, how many lymph nodes it has metastasized to (if any), and whether it has spread to distant organs. Staging of a tumor is often the most important predictor of survival, and cancer treatment is primarily determined by staging.

Cancer is treated by several methods: surgery, chemotherapy, radiation therapy, complementary medicine and/or palliative care. Surgery is most effective on localized, solid cancers. It is an important part of diagnosing and staging the tumor, as biopsies are typically required. Chemotherapy, in addition to surgery, has proven to be a useful therapeutic modality against breast, colorectal, ovarian, and other cancers [10]. The effectiveness of chemotherapy, unfortunately, is often limited, as the therapeutic agent is also toxic to healthy cells in the body. In addition to surgery and/or chemotherapy, radiation therapy is also often applied; it uses carefully targeted doses of high-energy radiation to kill cancer cells. Radiation therapy is prescribed in about 50% of all cancer cases and is effective in about 70% of those with bone metastasis. For certain types of cancers, such as early head and neck cancer, radiation may be used effectively alone [10]. As an alternative to conventional treatment, a "complementary medicine" such as a herbal remedy and naturopathic medicine, etc. is sometimes used; however, most of these remedies have not been scientifically proven [11]. Finally, palliative care is administered to reduce the physical and emotional distress of patients near end of life.

1.6. Diagnosis and prognosis

Most tumors are initially discovered either because of the presentation of symptoms or through screening. A patient suspected of harboring a tumor usually requires a thorough evaluation of her/his health record, a physical examination and diagnostic testing such as blood tests, X-rays, CT scans, endoscopy, and tumor biopsy. Histological examination of tumor cells provides information about the type, size, shape, appearance and stage of the tumor, and is the only unambiguous diagnostic tool (determining cancer staging). This diagnosis typically dictates how a tumor should be treated.

Benign tumors are harmless by definition and can be left untreated as long as there is continual monitoring. Stage I cancers (tumor localized to one part of the body) are, in many cases, treatable by surgical removal of the tumor alone. The more advanced stages of cancer, in which the tumor has invaded adjacent organs or has spread to distant organs, require chemo- and radiotherapies in addition to surgery. Prescribed treatments based solely on staging can cause more harm than good; for instance, Stage I cancers occasionally remain dormant for prolonged periods and treatment may be postponed, if it is too risky. In other cases, Stage I cancer can rapidly become more aggressive (invade adjacent and distant organs) [12] and become life-threatening; thus, surgery, chemo- and/or radiotherapies may be necessary. Selection of the most appropriate treatment requires prognostic information on how fast the cancer will progress.

The onset of the transformation of a less aggressive cancer into a more aggressive one may not immediately cause changes of cellular morphology and may be overlooked

by histological examination. However, such changes may be detectable at the molecular level using methods such as ELISA, immunohistochemistry (IHC), etc., but one must know what molecular changes to expect when the cancer becomes more aggressive. These molecular warnings could lead to more effective treatment, and hence, higher success rates as well as better patient care in the battle against cancer.

1.7. Cancer biomarkers

Molecular warnings of disease, such as those described in Section 1.6, are described as biomarkers. In general, a biomarker is a substance (DNA, RNA, protein, hormone, etc) present in tissues or body fluids that changes in concentration (commonly used), location, and/or structure, which signifies a change in the patient's condition, including the presence or progression of a disease or a response to therapy.

Various types of biomolecules can be used as biomarkers; for example, the increase in total cellular DNA indicates a cell entering S phase in preparation for mitotic division; increase in blood cholesterol indicates a risk of coronary and vascular disease; overexpression of the protein, carbonic anhydrase 9 (CAIX), indicates the progression of primary into metastatic ccRCC [13]; and expression of the neurofilament heavy chain (pNFH) proteins in cerebrospinal fluid helps to monitor drug efficacy in the treatment of amyotrophic lateral sclerosis (ALS) [14].

Cancer biomarkers can be classified into three categories: diagnostic, prognostic and predictive. A diagnostic biomarker indicates that an individual has a cancer. Prognostic biomarker predicts the likely course of a cancer in the near future, if left untreated. In contrast to prognostic biomarker, a predictive biomarker identifies

subpopulations of patients who are most likely to respond to a given therapy [15]. Both diagnostic and prognostic biomarkers can also be used to monitor disease recurrence after initial therapy.

Biomarkers harvested from biopsy samples or surgically removed tumors are described as tissue-based biomarkers. These biomarkers cannot be used to screen the general population because they require invasive methods to retrieve them. Biofluid-based biomarkers obtained from blood, urine, saliva, vaginal secretion, etc. are potentially more valuable because they can be sampled using less invasive means.

Samples employed in cancer biomarker discovery include tissues, biofluids and cell cultures. Although biofluid-based biomarkers are preferred, their discovery is typically challenging as their concentrations in a biofluid are typically low and the biofluid matrix is complex. Instead, it may be possible to first discover them in tissue samples, as long as they are eventually secreted. One strategy that has proven effective in integrating the two conditions is to examine the secretome of a cell culture of the tissue [16]. This has the advantage of maintaining high biomarker concentration while avoiding complexity of the biofluid matrix [17, 18].

The studies presented in this thesis focus on the discovery of potential diagnostic and prognostic tissue- and biofluid-based protein biomarkers for ccRCC and HNOSCC by the analyses of tissue and secretome samples.

1.8. Proteins

Proteins are an essential structural component of cells and are the molecular machinery in them. They play a multitude of regulatory, signaling, catalytic and other

roles in our body. Proteins are highly dynamic and their state (tertiary structure, chemical modification, expression and location, etc.) depend on the physiological state of the cell or the tissue where they are expressed.

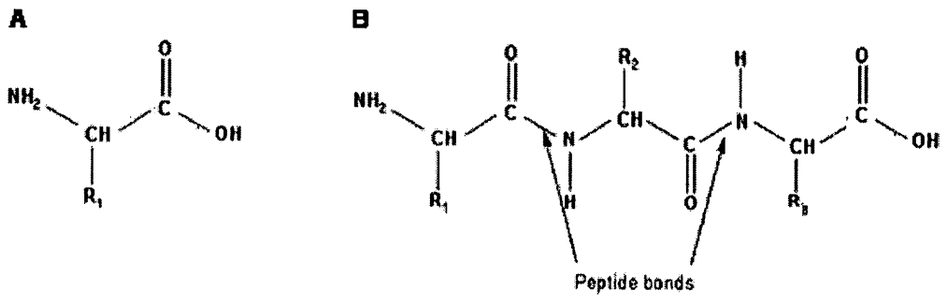


Figure 1.1. *Basic structure (**A**) of an amino acid and (**B**) of a tripeptide. Peptide bonds (CO-NH) link the amino-acid residues. R_1, R_2 and R_3 represent the side chains that characterize the residues.*

Proteins are polymers of amino acids covalently linked by peptide bonds (CO-NH). Proteins can be as small as human chaperonin 10, consisting of about 100 amino acids, and as large as human Titin, consisting of over 34,000 amino acids [20]. Short polymers made of less than 40 amino acids are generally classified as peptides. The basic structure of a generic amino acid and that of a generic tripeptide are given in Figure 1.1.

1.9. Challenges

In the field of biomarker detection, sensitivity and specificity are uniquely defined and are important characteristics of a given biomarker performance. Sensitivity is the fraction (or percentage) of cancer patients successfully identified. Specificity is the fraction of cancer-free individuals identified as such. An ideal biomarker would have

100% sensitivity and specificity, i.e., all individuals with cancer would have a positive test result, and all individuals without cancer would have a negative test result. Unfortunately, none of currently available protein biomarkers achieve this level of performance; in fact, many of the commonly given tests are far from ideal. There are many reasons for this inadequacy: tumor heterogeneity (multiple tumor subpopulations within a single type of cancer), matrix interference, individual variations, multiple protein functions, etc. [21]. For instance, the prostate specific antigen (PSA) test for prostate cancer, despite its wide use, has a relatively poor sensitivity and poor specificity, and is prescribed only for high-risk individuals who are being monitored for prostate cancer recurrence [22]. For the general population, the effectiveness of the PSA test, even used in conjunction with digital rectal examination, is subject of much discussion.

The sensitivity and specificity of a biomarker test can be increased by means of a panel of biomarkers. For example, a study showed that five antigens, galectin-3, PAK2, PHB2, RACK1, and RUVBL1, can discriminate individuals with early-stage cancers from those who are healthy [23]. Individually, these biomarkers have a sensitivity of 40-60% and a specificity of around 80%, but when combined, the sensitivity increases to 70%. It was suggested that this panel of biomarkers could be useful as a diagnostic tool to screen early-stage invasive breast cancer and pre-invasive breast cancer for woman at risk and could be particularly useful as complement to mammography for women with high breast density.

Another challenge is that most of cancer biomarkers are not cancer-specific, but are found in a number of different cancers. For example, alpha-enolase (ENO1) is upregulated in glioma [24] and in clear cell renal cell carcinoma (ccRCC) (Chapter 3).

Furthermore, not all protein biomarkers function in the same way in different cancers. For instance, in glioma, neuroblast differentiation-associated protein AHNAK (AHNAK), is downregulated [25], whereas in ccRCC it is upregulated (Chapter 3). Although neither ENO1 nor AHNAK is cancer-specific, as a panel, they can potentially discriminate between both glioma and ccRCC. Upregulation of ENO1 and downregulation of ANAK suggest the onset of glioma, but upregulation of both proteins suggests the onset of ccRCC.

From the foregoing, one may conclude that the current focus of protein biomarker discovery is to identify biomarkers or panels of biomarkers having (1) high sensitivity and specificity, (2) cancer-specificity, and (3) present in body fluids. Such biomarkers or panel of biomarkers in combination with other diagnostic modalities, e.g., imaging techniques, can ultimately improve the management of cancer and save lives.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics.

1. CA Cancer J Clin. 2011 Mar-Apr;61(2):69-90. Epub 2011 Feb 4. Erratum in: CA Cancer J Clin. 2011 Mar-Apr;61(2):134.

2. www.uicc.org

3. Anand P, Kunnumakkara AB, Sundaram C, Harikumar KB, Tharakan ST, Lai OS, Sung B, Aggarwal BB. Cancer is a preventable disease that requires major lifestyle changes. Pharm Res. 2008 Sep;25(9):2097-116. Epub 2008 Jul 15. Review. Erratum in: Pharm Res. 2008 Sep;25(9):2200.

4. Nelson DA, Tan TT, Rabson AB, Anderson D, Degenhardt K, White E. Hypoxia and defective apoptosis drive genomic instability and tumorigenesis. Genes Dev. 2004 Sep 1;18(17):2095-107. Epub 2004 Aug 16.

5. http://www.cancer.gov/dictionary

6. Read, A. P.; Strachan, T. Human molecular genetics 2. New York: Wiley; 1999. ISBN 0-471-33061-2. Chapter 18: Cancer Genetics.

7. Suh SO, Chen Y, Zaman MS, Hirata H, Yamamura S, Shahryari V, Liu J, Tabatabai ZL, Kakar S, Deng G, Tanaka Y, Dahiya R. MicroRNA-145 is regulated by DNA methylation and p53 gene mutation in prostate cancer. Carcinogenesis. 2011 May;32(5):772-8. Epub 2011 Feb 23.

8. Mellert H, Sykes SM, Murphy ME, McMahon SB. The ARF/oncogene pathway activates p53 acetylation within the DNA binding domain. Cell Cycle. 2007 Jun 1;6(11):1304-6. Epub 2007 Jun 24. Review.
9. http://medical-dictionary.thefreedictionary.com/benign
10. Holland, James F. (2009). Holland-Frei cancer medicine. (8th ed. ed.). New York: McGraw-Hill Medical.
11. Vickers A. Alternative cancer cures: "unproven" or "disproven"? CA Cancer J Clin. 2004 Mar-Apr;54(2):110-8. Review.
12. Rini BI, Campbell CS, Escudier B. Renal cell carcinoma. Lancet 2009; 373: 1119–32.
13. Tostain J, Li G, Gentil-Perret A, Gigante M. Carbonic anhydrase 9 in clear cell renal cell carcinoma: a marker for diagnosis, prognosis and treatment. Eur J Cancer. 2010 Dec;46(18):3141-8. Epub 2010 Aug 13. Review.
14. Levine TD, Bowser R, Hank NC, Gately S, Stephan D, Saperstein DS, Van Keuren-Jensen K. A Pilot Trial of Pioglitazone HCl and Tretinoin in ALS: Cerebrospinal Fluid Biomarkers to Monitor Drug Efficacy and Predict Rate of Disease Progression. Neurol Res Int. 2012;2012:582075. Epub 2012 Jun 28.
15. http://www.dako.com/index/knowledgecenter/kc_publications/kc_publications_connection/kc_publications_connection13-htm/28828_2009_conn13__difference_predictive_prognostic_biomarkers_brunner.pdf
16. Gunawardana, C. G., Kuk, C., Smith, C. R., Batruch, I. et al., Comprehensive analysis of conditioned media from ovarian cancer cell ines identifies novel

candidate markers of epithelial ovarian cancer. J. Proteome Res. 2009, 8, 4705–4713.

17. Villanueva, J., Philip, J., Chaparro, C. A., Li, Y. et al., Correcting common errors in identifying cancer-specific serum peptide signatures. J. Proteome Res. 2005, 4, 1060–1072.
18. Anderson, N. L., Anderson, N. G., The human plasma proteome: history, character, and diagnostic prospects. Mol. Cell. Proteomics 2002, 1, 845–867.
19. Grønborg M, Kristiansen TZ, Iwahori A, Chang R, Reddy R, Sato N, Molina H, Jensen ON, Hruban RH, Goggins MG, Maitra A, Pandey A. Biomarker discovery from pancreatic cancer secretome using a differential proteomic approach. Mol Cell Proteomics. 2006 Jan;5(1):157-71. Epub 2005 Oct 8.
20. http://web.expasy.org
21. Heppner GH. Tumor heterogeneity. Cancer Res. 1984 Jun;44(6):2259-65.
22. Thompson IM, Pauler DK, Goodman PJ, Tangen CM, Lucia MS, Parnes HL, Minasian LM, Ford LG, Lippman SM, Crawford ED, Crowley JJ, Coltman CA Jr. Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter. N Engl J Med. 2004 May 27;350(22):2239-46. Erratum in: N Engl J Med. 2004 Sep 30;351(14):1470.
23. Lacombe J, Mangé A, Jarlier M, Bascoul-Mollevi C, Rouanet P, Lamy PJ, Maudelonde T, Solassol J. Identification and validation of new autoantibodies for the diagnosis of DCIS and node negative early-stage breast cancers. Int J Cancer. 2012 Aug 7.

24. Beckner ME, Gobbel GT, Abounader R, Burovic F, Agostino NR, Laterra J, Pollack IF. Glycolytic glioma cells with active glycogen synthase are sensitive to PTEN and inhibitors of PI3K and gluconeogenesis. Lab Invest. 2005 Dec;85(12):1457-70.

25. Gentil BJ, Benaud C, Delphin C, Remy C, Berezowski V, Cecchelli R, Feraud O, Vittet D, Baudier J. Specific AHNAK expression in brain endothelial cells with barrier properties. J Cell Physiol. 2005 May;203(2):362-71.

Chapter 2

Techniques and methods

2.1. Mass spectrometry

2.1.1. Introduction

Mass spectrometry (MS) is an indispensable analytical tool in chemistry, biochemistry, pharmacy, medicine, and many related fields of science. MS is used to elucidate the structures of unknown substances, quantify environmental and forensic analytes, and perform quality control of drugs, foods, and polymers. MS can be used to deduce a compound's empirical formula from the atomic or molecular mass(es) of its constituents. The relative abundance of isotopes helps to determine the elements that contribute to the formula and estimate the number of atoms of each element. MS can be used to elucidate the connectivity of atoms within small molecules, identify functional groups, determine the sequence of constituents in macromolecules, and in some cases, even yield their three-dimensional structures. From its inception to the present, MS has undergone continual development and its applications are widening.

The basic principle of MS involves introducing a sample into the gaseous state; generating ions from it by a suitable method; separating these ions by a mass analyzer according to their mass-to-charge, m/z, ratios (the atomic mass per number of elementary charges); detecting the ions; and generating a mass spectrum. Analysis of the mass

spectrum results in qualitative and quantitative determinations of ions based on their respective m/z values and abundances.

MS is performed using a mass spectrometer, which consists of an ion source, a mass analyzer, and a detector (Figure 2.1). In advanced instrumentation, different types of mass analyzers can be arranged in tandem to exploit their respective strength in a single, so-called, "hybrid instrument." There are different techniques that are employable to generate, select, fragment, and detect ions.

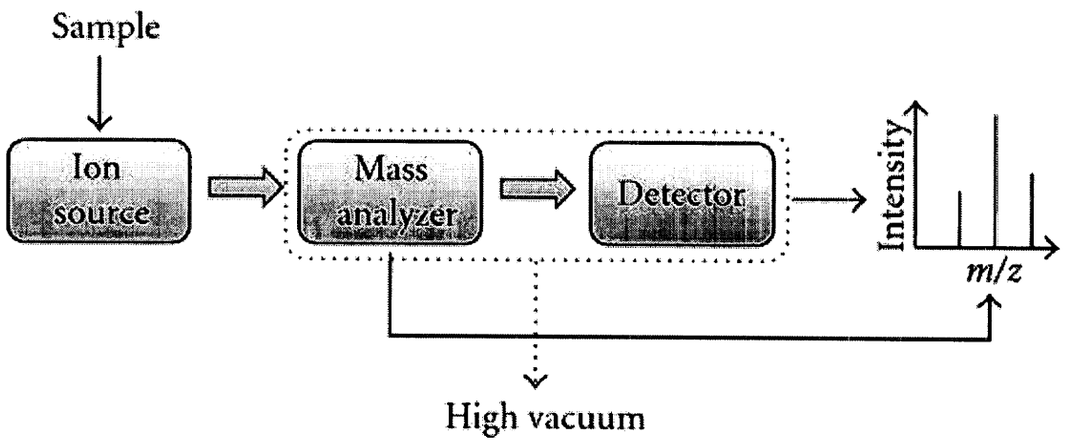


Figure 2.1. General scheme of a generic mass spectrometer. *Figure is reproduced from Banerjee et al. [1].*

2.1.2. Ion source

2.1.2.1. Creation of Ions

The mass analyzer can only separate ionized species. The analyte may be ionized by means of a plethora of techniques, including thermally (via an electrically heated thin metal ribbon), inductively-coupled plasma (ICP), glow discharge, electron ionization (EI), chemical ionization (CI), atmospheric pressure photoionization (APPI), fast atom

bombardment (FAB), electrospray ionization (ESI), and matrix-assisted laser desorption/ionization (MALDI).

Several of the ionization techniques can impart relatively high energies into the ionizing compound and cause its dissociation into atomic and fragment ions. These techniques are known as hard ionization techniques. Glow discharge and ICP create atomic ions and are useful for elemental analysis of metals, alloys, and metal oxides. EI is a widely used technique in organic analysis; the typically high (70 eV) energy of the ionization electrons results in fragmentation of the analyte, typically at weak bonds. The fragmentation pattern can be examined to reveal atom connectivity and structure. Gas chromatography (GC) is often used upstream from EIMS to separate the components in a sample. GC-MS has evolved into one of the most powerful analytical tools in the analysis of volatile and semi-volatile organic compounds.

By contrast, other techniques, such as MALDI and ESI [2], generate ions without extensive fragmentation; these are known as soft ionization techniques. MALDI and ESI are especially useful for the analysis of large biomolecules, including proteins, DNA and RNA. These techniques produce gas-phase ions from the solid state (MALDI) and solution (ESI), thus bypassing the roadblock of having to generate neutral analytes in the gas phase, a prerequisite in most other ionization techniques.

2.1.2.2. Electrospray ionization

ESI is the most widely used ionization technique for the study of soluble proteins and peptides among the group of atmospheric pressure ionization (API) methods, and is the primary ionization technique in liquid chromatography-mass spectrometry (LC-MS)

[3]. Since ESI is used extensively in the work presented in this thesis, a more detailed discussion of the technique is presented below.

ESI transfers the analyte from the solution phase to the gas phase as ions. The ions are produced either by the addition of a proton (H^+) to the analyte (M), the protonated analyte is typically denoted as $[M + H]^+$, or by the removal of a proton from the analyte, denoted as $[M - H]^-$. Ions may also be produced by the addition of cations such as sodium, which is ubiquitous when using glass vessels, giving the ion, $[M + Na]^+$. Multiply charged ions such as $[M + nH]^{n+}$ and $[M + nNa]^{n+}$ are often observed, especially when M is relatively large. In ESI, the analyte is dissolved in a mixture of water and a volatile organic solvent (e.g. methanol or acetonitrile), typically at a concentration of 10^{-6}–10^{-4} M. The dissolved analyte acquires protons from two sources [4]: the solution itself, which is typically at a pH of approximately 3.0 (samples are electrosprayed in dilute acid for positive ion detection), and from the electrolysis of water in the sample solvent, caused by the large potential difference between the capillary and the curtain plate.

The ESI is a process (shown in Figure 2.2) in which a high electric field causes the sample solution to emerge from the tip of a capillary as a jet in the shape of a 'Taylor cone' [5]. At the tip of the Taylor cone, charged droplets of the solution are ejected into the gas phase as a fine mist and begin to travel towards the curtain plate, guided by the electric field. As the droplets near the curtain plate entrance, they are met with a counter flow of dry N_2 that facilitates desolvation and minimizes sampling of solvent molecules into the orifice and entrance into the spectrometer.

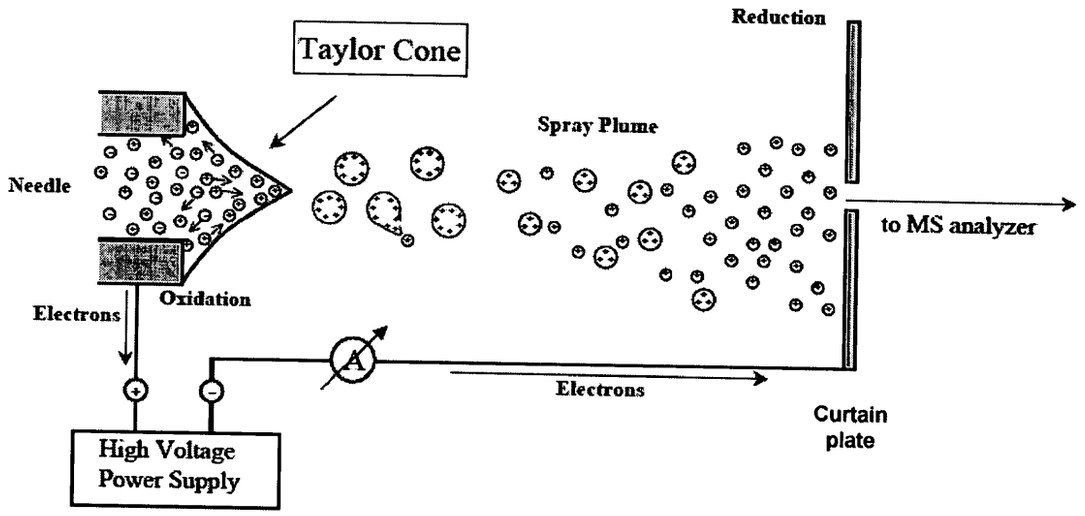


Figure 2.2. Schematic representation of the electrospray process in the positive ion detection mode *(Figure is adapted from Crotti et al. [5]).*

Two hypotheses have been forwarded on how ions are generated in ESI. The first, called "the ion evaporation mechanism" [6], suggests that once a droplet reaches a critical radius, Coulombic repulsions cause ions to be directly emitted from it. The second, called "the charged residue mechanism" [7], suggests that a series of fissions lead to the production of small droplets that eventually each carries one analyte molecule bearing one or more charges (shown in Figure 2.3). A modification of "the charged residue mechanism" [8] considers the effect of the external electric field of the ESI source on the positively and negatively charged ions contained within the droplets and suggests that they undergo axial charge separation, at least for the first-generation offspring droplets prior to further subdivisions. Generally, "the charged residue mechanism" is considered

more applicable to large molecules, while "the ion evaporation mechanism" is considered more relevant to small molecules.

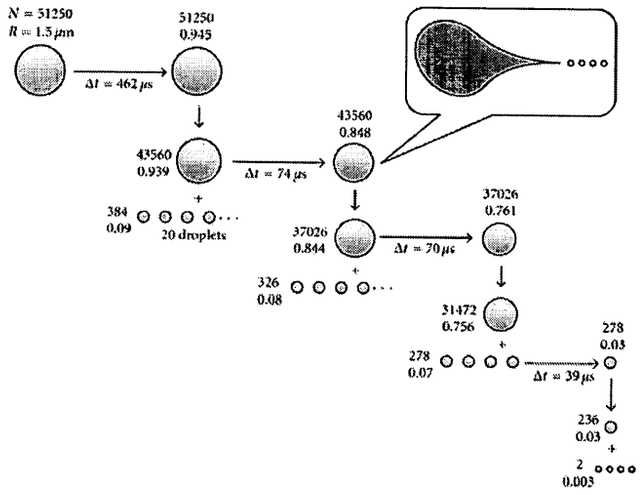


Figure 2.3. Time history of the charged methanol droplet produced by electrospray process. *The droplet at the top left is a typical parent droplet created at the ES capillary tip. The successive solvent evaporation and Coulomb fission leads to the charged nanodroplets that are the precursors of the gas-phase analyte ions. The numbers beside the droplets give radius R (μm) and number of elementary charges N on the ES droplet; Δt corresponds to the time required for evaporative droplet shrinkage to size where fission occurs. Only the first three successive fissions of a parent droplet are shown. At the bottom right, the fission of the offspring droplet to produce the charged nanodroplets is shown. The inset shows a drawing of droplet jet fission based on actual flash microphotograph. (Figure is reproduced from Banerjee et al. [9]).*

One of the advantages of using ESI to study proteins and other large molecules is that it can generate multiply charged ions, which lowers the m/z of high-mass analytes into a range accessible by most mass analyzers. The disadvantage of having multiply

charged ions is that it can lead to more complicated spectra, which may hinder protein identification.

2.1.3. Mass analyzer

Gas-phase analyte ions may vary in *m/z* and abundance by several orders of magnitude. These ions must be sorted by the mass analyzer, often in a timespan of a few milliseconds or less, and presented to the detector in ascending or descending *m/z* ratios with respect to time. Alternatively, the ions may be made to orbit or oscillate at frequencies unique to their *m/z* values, and these frequencies are then detected to determine the *m/z* of the ions present in the sample. The ions are manipulated, depended on the mass-analyzer type, by static or dynamic electric or magnetic fields to achieve spatial, and ultimately time, separation, or to coerce them into orbits or oscillations.

There are several types of mass analyzers: the magnetic sector (B), in which a continuous ion beam is deflected in a magnetic field with bending radius depending on the ion's momentum; the linear quadrupole (Q), in which a continuous ion beam moves in an oscillating electric field with separation achieved based on the stability of ion trajectories in the field; the linear ion trap quadrupole (LTQ), in which ions are trapped and separated in a linear, quadrupolar radio-frequency field by resonant excitation; the quadrupole ion trap (QIT), in which ions are trapped and separated in a three-dimensional radio-frequency quadrupole field by resonant excitation; the time-of-flight (TOF), in which ions of different m/z, having different velocities, are separated during their flight through a field-free drift tube; the Fourier transform-ion cyclotron resonance (FTICR), in which ions are trapped in an orbit in magnetic field and their m/z are deduced from their

cyclotron frequency; and the Orbitrap, in which ions oscillate in an inhomogeneous electric field, allowing their m/z to be deduced from their oscillation frequencies. Quadrupole and TOF were employed in the studies presented here; thus, these analyzers will be described in more detail.

2.1.3.1. Quadrupole

A quadrupole consists of four parallel, cylindrical or hyperbolic rod electrodes mounted in a square configuration as shown in Figure 2.4. The pairs of opposite rods are each held at the same potential which is composed of a direct current (DC) voltage and a superimposed radio-frequency (RF) voltage. Ions travel down the quadrupole in the axial space between the rods in the z-direction. An attractive force is exerted on them by one pair of rods having the opposite charge to the ions. If the voltage applied across the pairs of rods is periodic, attraction and repulsion in both the x- and y-directions will alternate in time. The resulting motion of the ions in the quadrupole field depends on their mass and charge, and is described by the Mathieu equation. The equation shows that the DC and RF fields can both stabilize, or destabilize, an ion's trajectory depending on the ion's *m/z* value and the ratio of DC to RF voltages. Only ions of a certain *m/z* window that have stable trajectories for the given ratio of voltages (resonance ions) will reach the detector; the other ions, having unstable trajectories (non-resonance ion), will collide with the rods or exit the quadrupole via the space between the rods (see Figure 2.5). Thus, the "window of transmission" can be tuned to transmit ions of a particular *m/z* value. By ramping the ratio of DC and RF voltages, a range of *m/z* values can be scanned to generate a mass spectrum. When the DC component of the applied voltages on the quadrupole rods is set to zero, the mass analyzer is said to be operating in "RF-only"

mode. This results in stable trajectories for all of the ions and they are all transmitted. The "RF only" setting is used for ion transmission and focusing. Thus, a quadrupole is an effective tool for transmitting an entire population of ions, as an ion lens, or for selectively transmitting ions of specific *m/z* values as a mass analyzer.

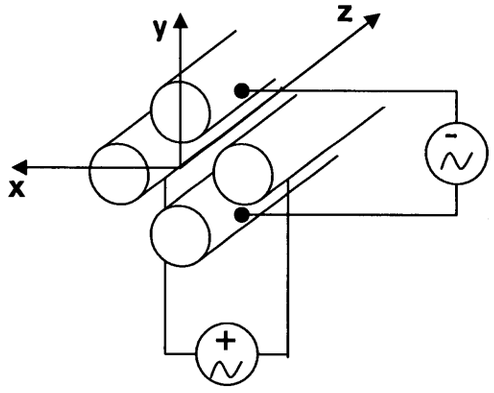


Figure 2.4. Schematic diagram of the quadrupole mass analyzer, conventionally the direction down the axis is defined as z and the quadrupolar field is applied across the xy plane.

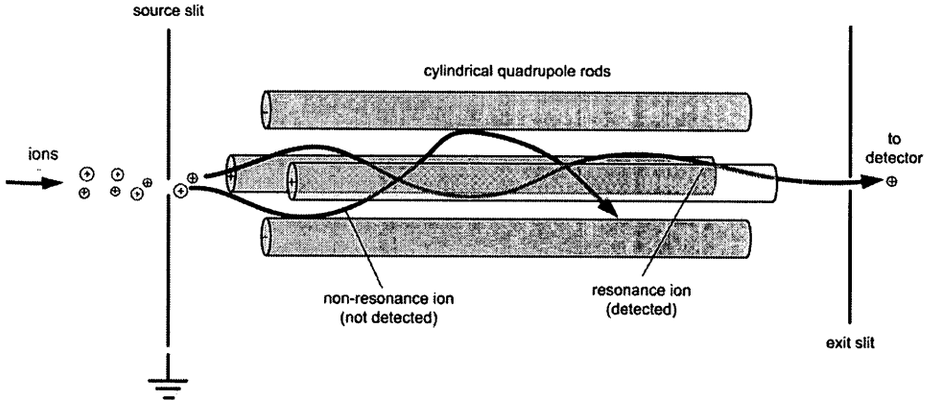


Figure 2.5. Schematic representation of ions with stable trajectory (resonance ions) and unstable trajectory (non-resonance ions) in the quadrupole filter. *Figure is adapted from Wittmann et al. [10].*

2.1.3.1.1. Collision cell

The quadrupole can also be used as a collision cell to induce ion fragmentation via collision-induced dissociation (CID). When used as a collision cell, the quadrupole is operated in "RF-only" mode: precursor ions are made to collide with a neutral gas (e.g. N_2 or He) with sufficient energy to cause the ions to fragment. Collision between the precursor ion and the inert gas results in conversion of the ion's kinetic energy into internal energy. As dissociation rates are typically much lower than the rates of intramolecular vibrational energy redistribution [11], the excess energy is equilibrated within the ion, leading to fragmentation of weaker bonds irrespective of collision site. Ionic fragments (product ions) are then transmitted to a second mass analyzer where they are separated for subsequent detection.

2.1.3.2. Time-of-Flight

The TOF mass analyzer determines the *m/z* value of an ion via a time measurement. The ions of different masses and charges are accelerated by an electric field of known strength before entering the TOF tube. All ions of a given charge acquire the same kinetic energy, E_k, which is proportional to the ion charge, ze, and the voltage applied across the accelerator grid, V, according to the equation:

$$E_k = zeV \qquad (1)$$

Because the kinetic energy is related to the mass, m, and the velocity, v, of the ions, according to the equation,

$$E_k = \frac{1}{2}mv^2 \qquad (2)$$

lighter ions travel faster than heavier ones, given the same kinetic energy, and will reach the detector at the end of the TOF tube earlier. The time, *t*, that it takes for the ion to travel the distance, *d*, to the detector is measured to deduce the velocity of the ion:

$$v = \frac{d}{t} \qquad (3)$$

By combining Equations 1 – 3 the time-of-flight can be used to deduct the *m/z* value of the ion, *viz.*:

$$t = \frac{d}{\sqrt{2eV}}\sqrt{\frac{m}{z}} \quad = \quad k\sqrt{\frac{m}{z}} \qquad (4)$$

2.1.4. Scan modes in MS

A modern tandem mass spectrometer is capable of a number of modes of operation, depending on how the mass analyzers are operated to achieve specific goals. Figure 2.6 shows a schematic representation of the main scanning modes. MS1 represents the first analyzer, which can transmit ions of a specific *m/z* or transmit all of the ions, CID occurs in the collision cell (RF-only quadrupole), and MS2 represents the second analyzer, which transmits or mass-analyzes precursor and product ions depending on the scan mode. In 'single-stage MS scanning,' all precursor ions are transmitted unhindered through MS1 and the collision cell into MS2, where they are scanned. In 'product ion scanning' a precursor ion is mass-selected in MS1, fragmented in the collision cell, and the product ions are then mass-analyzed in MS2. In 'precursor ion scanning,' a specific product ion *m/z* is monitored in MS2, while the precursor masses are scanned in MS1. Only precursor ions that fragment to give the specific product ions are detectable. In 'neutral loss scanning,' MS1 scans a range of *m/z* while MS2 also scans the same range of *m/z*, but at an offset from the MS1, which corresponds to a particular neutral loss. In

'selected reaction monitoring (SRM),' MS1 is set to transmit a particular precursor ion, while MS2 is set to monitor the presence of a particular fragment ion. SRM is typically used for detecting specific analytes that elute from LC with tandem MS being the means of detection and quantification. In the studies presented here, the 'single-stage MS scanning' and 'product ion scanning' were used for the identification and quantification of proteins.

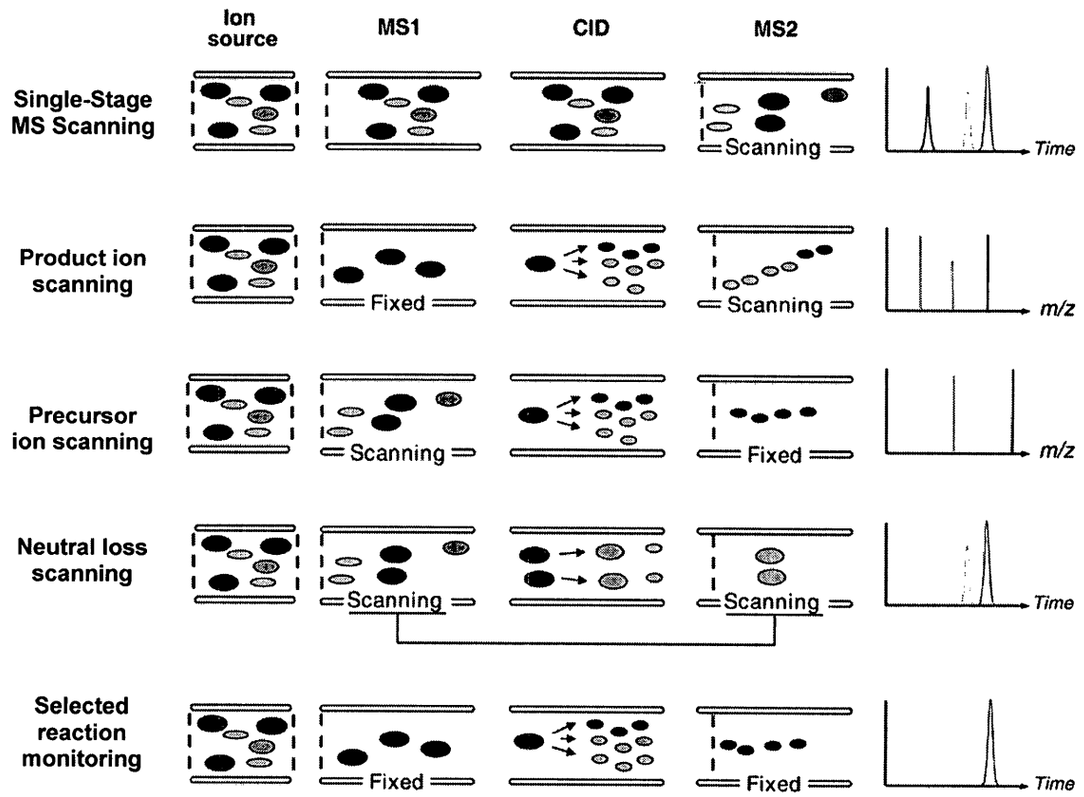


Figure 2.6. Schematic representation of different tandem MS scan modes. *(Figure is adapted from Domon and Aebersold [12]).*

2.1.5. Detector

There are essentially two types of ion detectors: one that generates a signal by sensing the oscillation or orbiting motion of the ions and another that generates a signal when ions collide with it. Only the FTICR and Orbitrap mass spectrometers use the first type of detector, while most others use the latter.

In the FTICR and Orbitrap, two electrodes are connected to an amplifier and sense the motion of the oscillating or orbiting ions. As a group of ions with a particular *m/z* value approaches one electrode while departing from the other, electrons flow through the amplifier, drawn to the first electrode by the ions' positive charges and sourced by the second. As the ions continue on their orbit or oscillation, they move away from the first electrode toward the second, causing the electric current to change direction. The resulting alternating current signal has the same frequency as the ions' oscillation or orbit. With many ions of different *m/z* values trapped, a complex periodic signal is generated, composed of superimposed alternating current signals. This complex signal can be amplified, digitized, and analyzed by Fourier transformation to ultimately yield a mass spectrum [2].

Among the second type of detectors, the simplest is the Faraday cup, which is an electrode with which the ions collide and deposit their charge [2]. The deposition of charge increases the electrode's potential, which is then measured by a sensitive electrometer. Faraday cups measure ion abundance with the highest accuracy and are, therefore, used in isotope-ratio mass spectrometry (IR-MS) [13].

Ion-to-photon detectors operate by converting the energy of an impacting ion to light, using a scintillation compound such as Rhodamine B or CsI, which is then detected by a photomultiplier tube. This detector is primarily used in MALDI mass spectrometers.

Ion traps, quadrupole and TOF analyzers usually use some type of electron multiplier detector. Because the number of ions reaching the detector at a particular instant is small, considerable amplification is necessary to obtain a signal. A channel electron multiplier (CEM) is used for this purpose. CEMs consist of a hollow tube with a semiconducting inner surface. When an ion strikes the inner surface near the entrance of the tube, it causes the emission of a number of electrons. As these electrons are accelerated down the tube under a potential gradient, they repeatedly strike the inner surface releasing secondary electrons, leading to a cascade of electrons. This multiplication process results in the generation of 10,000 electrons per incident ion (gain of 10^4). CEMs are unstable at gains exceeding 10^4. In TOF MS, the ion beam is usually much wider than in quadrupoles, and this necessitates the use of multiple, parallel, miniaturized CEMs in the form of the microchannel plate (MCP) detector to ensure a high efficiency of ion detection. This type of detector was used in this study and will be described in more detail below.

2.1.5.1. Microchannel plates

A diagram of the MCP is shown in Figure 2.7A. The gain of an MCP is 10^3–10^4, i.e., lower than that of CEM. Two MCPs are often sandwiched together (see Figure 2.7B) with a slight angle between the channels in a chevron pattern to obtain gains of 10^6–10^7. Electrons exiting the first plate initiate the cascade in the second plate. Electrons that exit

the chevron on the opposite side are collected by an anode and further amplified and digitized.

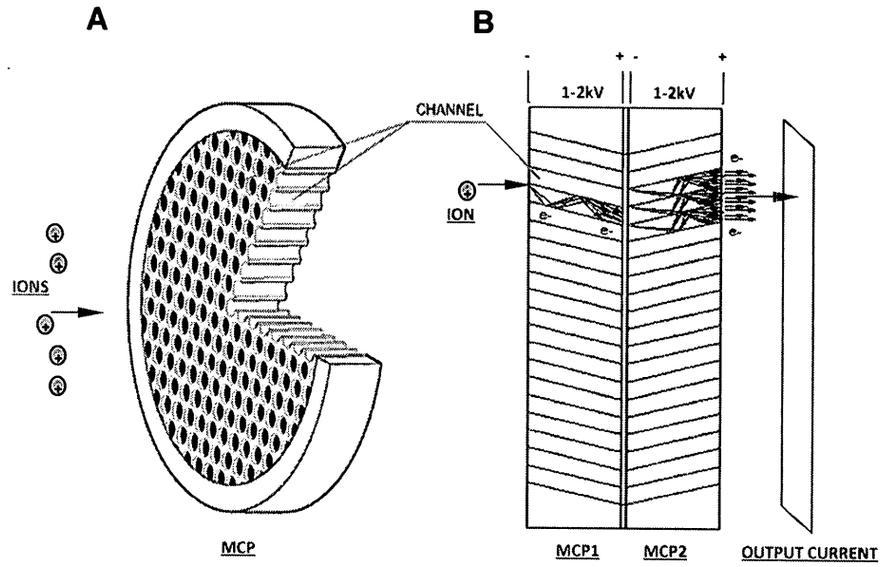


Figure 2.7. *(**A**) Microchannel plate (MCP) design. (**B**) Two MCP plates are arranged as a slightly spaced pair with their pores angling in opposite directions to enhance the electron cascading intensity. This figure combines two adapted images [14, 15].*

2.1.6. QSTAR Pulsar mass spectrometer

The QSTAR Pulsar mass spectrometer (Figure 2.8) from Applied Biosystems/MDS SCIEX (Foster City, CA) belongs to the QqTOF (Q = quadrupole mass selector, q = collision cell, TOF = mass analyzer) family of hybrid tandem mass spectrometers, and was configured as an HPLC-ESI-MS system and used for the proteomic analysis work presented in this thesis.

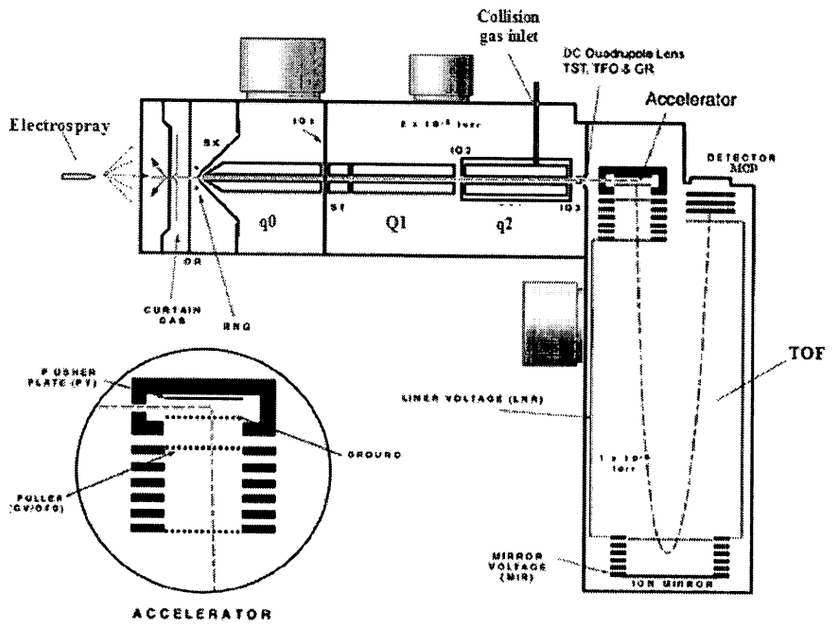


Figure 2.8. Schematic diagram of the QSTAR Pulsar. *The dashed line represents the ion path through the quadrupoles and TOF region. Figure is adapted [16].*

A nanobore high-performance liquid chromatography (HPLC) system (LC Packings, Amsterdam, Netherlands) was used upstream from the ESI source to separate the analytes (mostly tryptic peptides, see later), and both were controlled by a computer running Analyst QS 1.1 software. The HPLC system will be described further in a later section.

The mass spectrometer consists of an "RF only" quadrupole (q0), which functions as an ion lens, a quadrupole analyzer (Q1), which can be used to select ions for fragmentation, a second "RF only" quadrupole (q2), which functions as the collision cell, and a TOF mass analyzer. Low pressure is maintained within the instrument by three turbomolecular pumps, each of which evacuates gases at a rate of 250 – 500 L/min. The

q2 quadrupole is housed in a semi-isolated casing into which N_2 gas can be introduced for fragmentation. Semi-isolation is necessary for reaching a sufficiently high pressure (about $4x10^{-2}$ Torr) in the collision cell, yet maintaining a sufficiently low pressure elsewhere in the mass spectrometer.

The ions exiting q2 are injected into the TOF analyzer orthogonally to the TOF acceleration direction and are then reflected back towards the detector via an ion mirror. Orthogonal injection eliminates the problem of ion kinetic-energy spread along the z direction (the axis of the quadruples). Spread along the z direction becomes unimportant because the MCP detector has a large surface area and can capture all of the spread ions. The problem of ion kinetic-energy spread in the acceleration direction is solved by an ion mirror (reflectron) that enables the refocusing of ions having the same *m/z* value, but slightly different kinetic energy. Orthogonal injection and ion reflection via the reflectron were the technological advances that give the TOF mass analyzer its high-resolution.

The MCP detector is capable of differentiating ion-arrival events on a nanosecond timescale. The ion pulses are counted by the instrument's computer, which displays the output as a mass spectrum. The TOFMS resolution is about 10,000 - 12,000 as measured by dividing the *m/z* value of the ion being measured by the width of the ion peak at half maximum. The TOF mass accuracy typically exceeds 20 ppm and is determined and corrected using a procedure described as mass calibration using standard solutions containing analytes of accurately known *m/z* values. The QSTAR Pulsar mass spectrometer was operated in the positive ion mode in this study. An LC-MS run of a peptide sample, in the studies presented here, took 135 min. After loading the sample onto the chromatographic column, LC-MS data were acquired, using the Analyst QS 1.1

software, in series of 9-s cycles: 1-s survey scan (MS1), followed by four, 2-s product-ion scans (MS2) of the four most-abundant ion peaks in the survey scan (a schematic diagram of the process is shown in Figure 2.9).

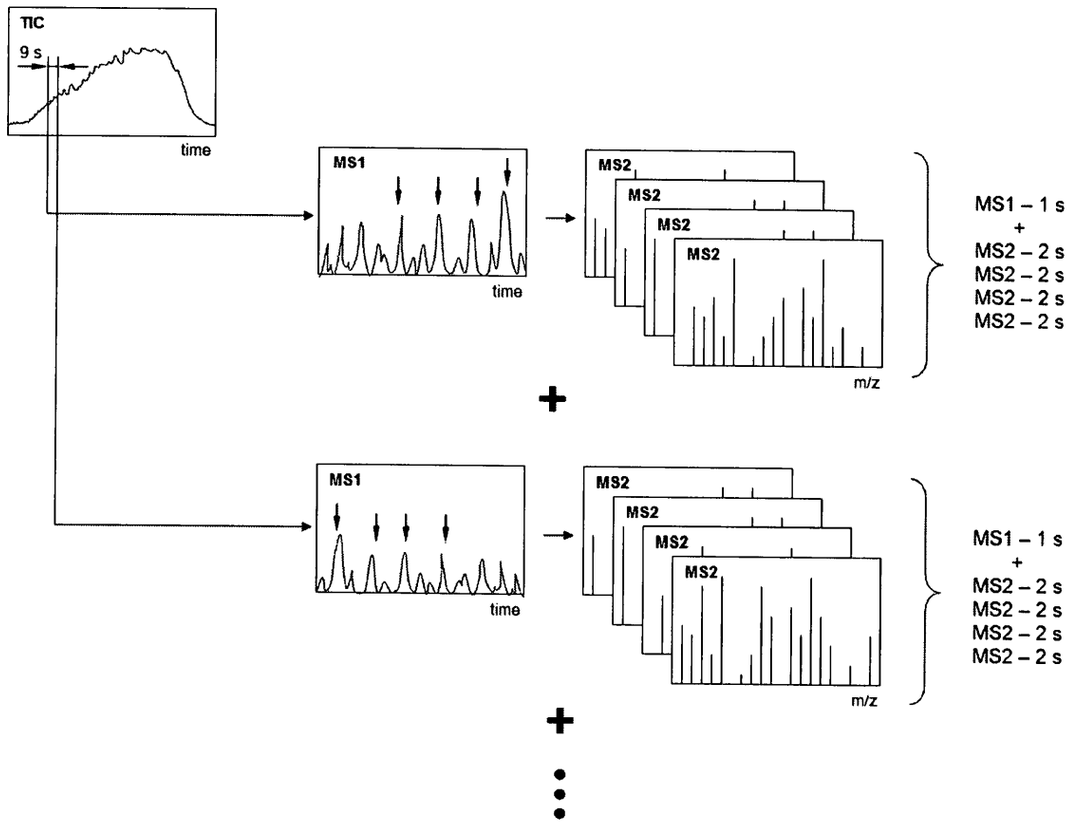


Figure 2.9. A schematic diagram represents the LC-MS data acquisition. *The TIC (total ion count) chromatogram shows all the ions eluted from RP LC column during 135-min LC-MS run. LC-MS data were acquired in the series of 9-s cycles: 1-s survey scan (MS1), followed by four, 2-s product-ion scans (MS2) of the four most-abundant ion peaks in the survey scan (as indicated by the blue arrows).*

2.1.7. Iterative runs with precursor ion exclusion (PIE)

Because of the overwhelming number of tryptic peptides in a digested protein sample (see later), only a small number of the peptides can be analyzed in a single LC-MS run. This is limited by the spectral acquisition speed of the mass spectrometer: only about 900 MS1 and 3600 MS2 spectra can be acquired in a 135-min LC-MS run (see Figure 2.9). As a result, only a limited number of proteins can be identified in one LC-MS run. The number of identified proteins can be increased by performing replicate LC-MS runs. Our experience is that about 17% of additional proteins can be identified in the second run and 8% in the third (i.e. a total increase of 25%). Additional runs beyond the third iteration identify few more proteins [17], the majority of the identifications being redundant.

PIE can increase the number of protein identifications in three replicate LC-MS runs by as much as 50% [18]. PIE works by excluding the precursor ions detected in the first LC-MS run (Figure 2.10A) from MS2 analysis in the second run (Figure 2.10B). Thus, the four peptides analyzed by MS2 in the second run, are next most abundant, and non-redundant. For the third run, the precursor ions detected in the first and second LC-MS runs are excluded from the MS2 analysis (Figure 2.10C). As a result, the total number of identified proteins in all LC-MS runs using PIE will be greater than just performing replicate runs under the identical conditions.

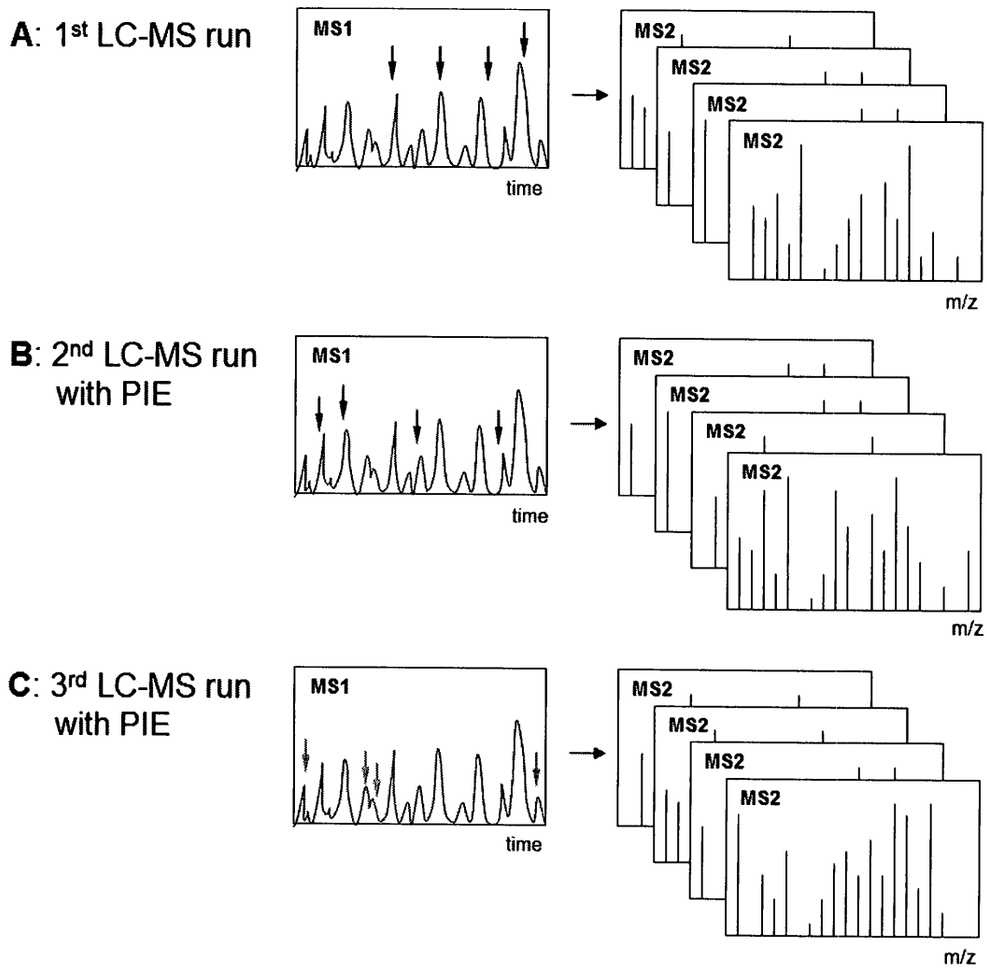


Figure 2.10. A schematic diagram illustrating the PIE strategy. *PIE minimizes the redundancy of identified peptides and increases the total number of identified proteins in three LC-MS runs.* ***A***, ***B***, *and* ***C*** *show the same group of peptides, eluted at a given time, in three different LC-MS runs. Precursor ions detected in the first LC-MS run* ***(A)*** *are excluded from the MS2 analysis in the second run* ***(B)***, *and the precursor ions detected in the first and second LC-MS runs are excluded from the MS2 analysis in the third run* ***(C)***.*The ions peaks selected for MS2 analysis are indicated by color-coded arrows.*

Wang's PIE method [18] was slightly modified and employed in this work. PIE lists were generated using an Excel template developed in-house. To generate the list for the second LC-MS run, the peptide summary of a sample, obtained after the first run, was imported into the Excel template where: (1) the *m/z* values and elution (retention) times of peptides identified with > 95% confidence were extracted; (2) alternative charge states (only +2, +3 and +4 were considered) of the peptides were calculated; and (3) the next three greater isotopic *m/z* values of extracted and calculated peptides were determined. The resulting m/z ratios from all three of these considerations constituted the PIE list. This list was saved as a text file and imported into the acquisition method for the second LC-MS run. The list, used for each LC-MS run, was cumulative of all the *m/z* values and elution times derived from all previous runs for the sample.

Wang et al. reported that the major time-consuming step was the cutting [18], pasting, and processing of data, which took about 30 min to generate a given PIE list for one sample. Our Excel template, developed in-house, reduced the processing time to less than 2 min [19].

2.1.8. Protein identification by MS

An MS2 or MS/MS analysis (see Figure 2.9) results in a product ion spectrum that typically comprises the precursor ion and a series of its fragment ions. While this fragmentation pattern can be studied manually and interpreted to give the primary structure (the sequence of amino acid residues) in a *de novo* manner, this task is time-consuming and demands much expertise. Peptide and protein identification in proteomics is typically accomplished by means of an automated search in which

"identification" is achieved by matching the *m/z* values of the ion peaks with those of theoretical ion peaks generated *in silico* on tryptic peptides deduced from all known proteins in a protein data bank. The best-matched tryptic peptide (and the protein from which the peptide is derived) constitutes the identification.

Dissociation of a peptide can potentially generate many types of fragment ions. This complexity is a function of the CID energy, the nature of the side chains and the residual sequence, and the three-dimensional structure of the peptide ions in the gas phase. A shorthand nomenclature has been developed to classify fragment ions and facilitate description. Figure 2.11 shows how fragment ions are named: a, b, and c ions are N-terminal-containing ions, while x, y, and z are C-terminal-containing ions; the subscripts denote the residues on which cleavage occurs.

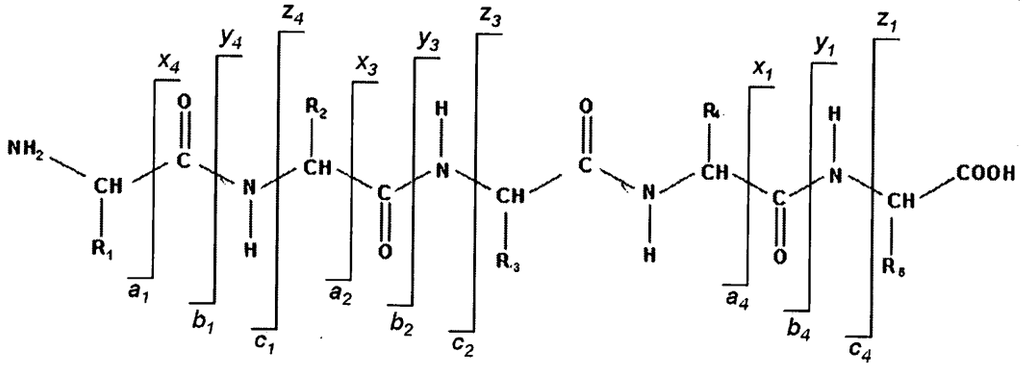


Figure 2.11. The fragmentation of backbone of a peptide. *The backbone of a peptide can fragment at three bonds CH-CO, CO-NH and NH-CH with each dissociation producing two fragments named according to the location of the charge and the amino acid position.*

Of the three types of bonds, CH-CO, CO-NH, and NH-CH, on the peptide backbone where fragmentation is possible, the peptide bond (CO-NH), which joins adjacent residues, is typically the weakest and, as a result, are the most easily dissociated in CID experiments. Thus, CID typically produces b- and y-type ions, which dominate the product ion spectrum as, for example, in Figure 2.12.

As pointed out earlier, by examining the mass difference between adjacent ions from the same peptide, it is possible to determine which amino acid is missing, thus allowing the sequence of the peptide to be determined *de novo* (see Figure 2.12). In reality, this is seldom unambiguous and often time-consuming, and *de novo* sequencing is really only practiced when database search is not a viable option, e.g. when the protein (or DNA) sequences of the organism are not available. Many computer-aided database search algorithms / programs are commercially available; the best of them include Mascot, Sequest, and ProteinPilot. Mass spectrometer manufacturers typically package one or several of these search programs with their instrument and operating software. The results of this dissertation research were obtained with ProteinPilot searches.

2.2. Trypsin digestion

Two fundamental approaches are currently used in proteomic analyses. In 'top-down' proteomics, intact proteins or large peptides are analyzed directly by MS. In 'bottom-up' proteomics, proteins are first cut via proteolytic digestion into peptides, which are then analyzed by MS. In cancer biomarker discovery, 'bottom-up' proteomics is the most commonly used approach, and was used in this study.

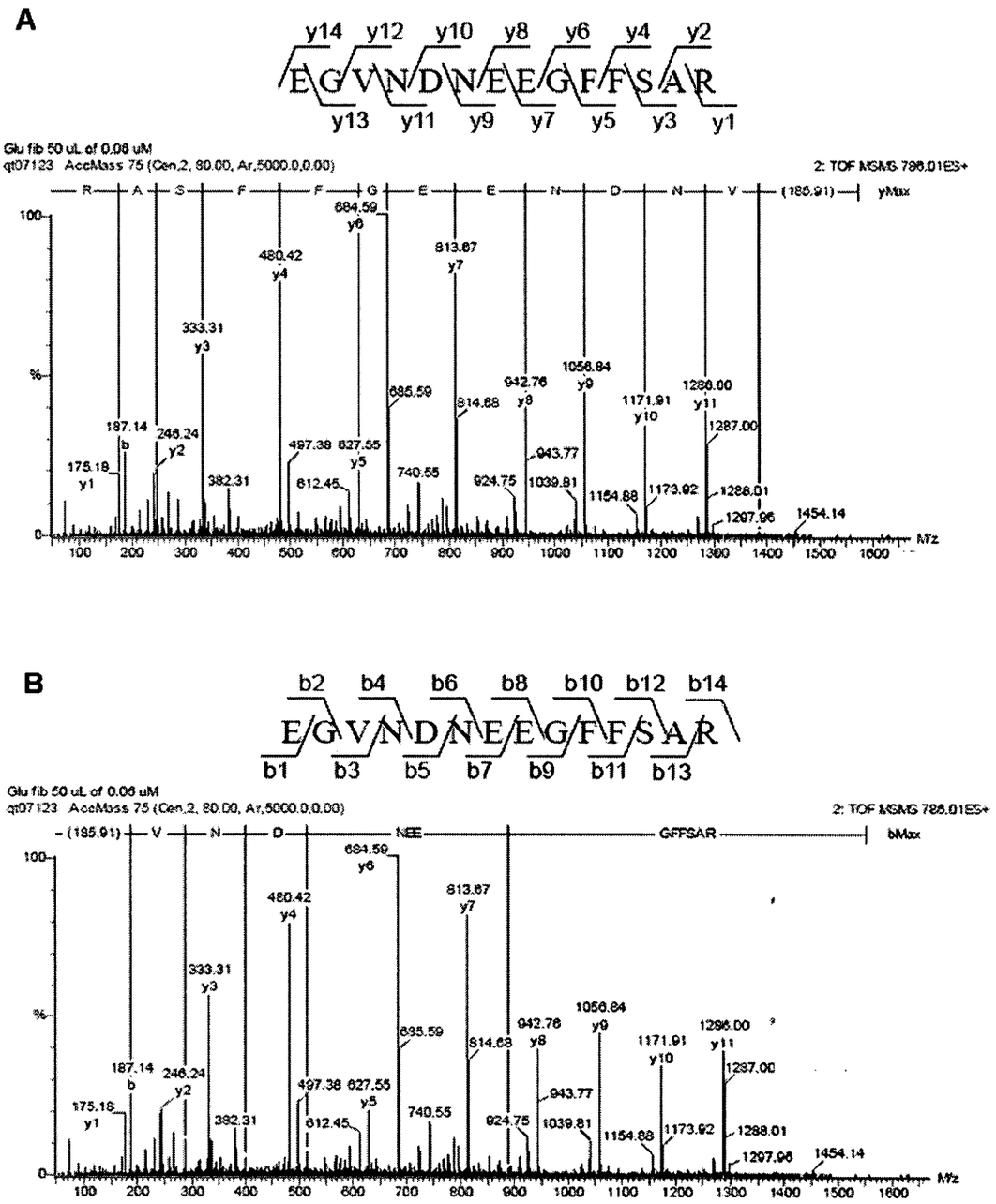


Figure 2.12. The CID QqTOF spectra of doubly charged Glu-fibrinopeptide. *(**A**) shows the y-series ions, (**B**) shows the b-series ions. Figures are reproduced [20].*

The enzymes typically used for proteolytic digestion of proteins are, Lys-C, Asp-N, Glu-C and trypsin, with the last being by far most commonly used and is the enzyme that was used exclusively in the research work described in this dissertation. Trypsin is a protease that cleaves proteins C-terminal to lysine (K) and arginine (R) residues. Lysines and arginines with a neighboring acidic amino acid - aspartic acid or glutamic acid – hydrolyze (cleaved by trypsin) more slowly; the presence of a neighboring proline C-terminal to lysine or arginine inhibits the hydrolysis completely [21]. Trypsin has high cleavage specificity, is aggressive, and is stable under a wide variety of conditions. Trypsin typically produces peptides within the preferred mass range for MS sequencing (due to the frequency of lysine and arginine residues) and more readily interpretable peptide fragmentation mass spectra relative to other enzymes. Trysin digestion is effective while the proteins are either in an SDS PAGE gel or in solution. All these properties help propel trypsin as the preferred proteolytic enzyme for proteomics.

Lys-C is active in harsher conditions and gives larger fragments than trypsin. Asp-N and Glu-C are also highly sequence-specific proteases, but less active than Lys-C. Other less sequence-specific proteases are generally avoided since they create complex mixtures of peptides, the spectra of which are more difficult to interpret [22].

2.3. Quantitative proteomics

Quantification in proteomics typically involves a comparison of the expression levels of proteins in two sample states. This is typically referred to as relative quantification. Relative quantification can be achieved by either label-free, or stable-isotope labeling approaches. In label-free quantification, each sample is prepared

separately, and subjected to individual LC-MS analyses. The quantification is then accomplished by either direct comparison of the mass spectral peak intensities of the precursor ion in the MS1 spectra of different biological samples (see Figure 2.13A), or counting the number of MS2 spectra identified for a given precursor (spectral counting) (see Figure 2.13B). Both approaches are relatively straightforward to implement, although both carry inherent weaknesses. In the first, differences in sample preparation and sample injection can result in peak intensity errors and retention time drifts, which can significantly complicate the comparison of multiple LC-MS runs [23]. In the second, normalization and statistical analysis of spectral counting datasets are necessary for reliable quantification. The acquisition and examination of many spectra of a given protein are required for reasonably accurate quantification.

Stable-isotope labeling is more costly and potentially more time-consuming; however, it is much less sensitive to experimental biases than label-free quantification. In addition, with standards of known concentrations, absolute quantification can be achieved. The 'spiked heavy peptides' method involves spiking known concentrations of heavier, isotopically labeled versions of the peptides to be quantified into an unlabeled sample, performing MS1 analysis, and quantifying the peptides using a heavy peptide standard curve (Figure 2.13C). One of the limitations of this approach is the purity of the spiked standard as there is likely interference from isobaric peptides present in the standard. However, this potential error can be corrected for provided the percents purity of the isotopically labeled and interfering peptides are known.

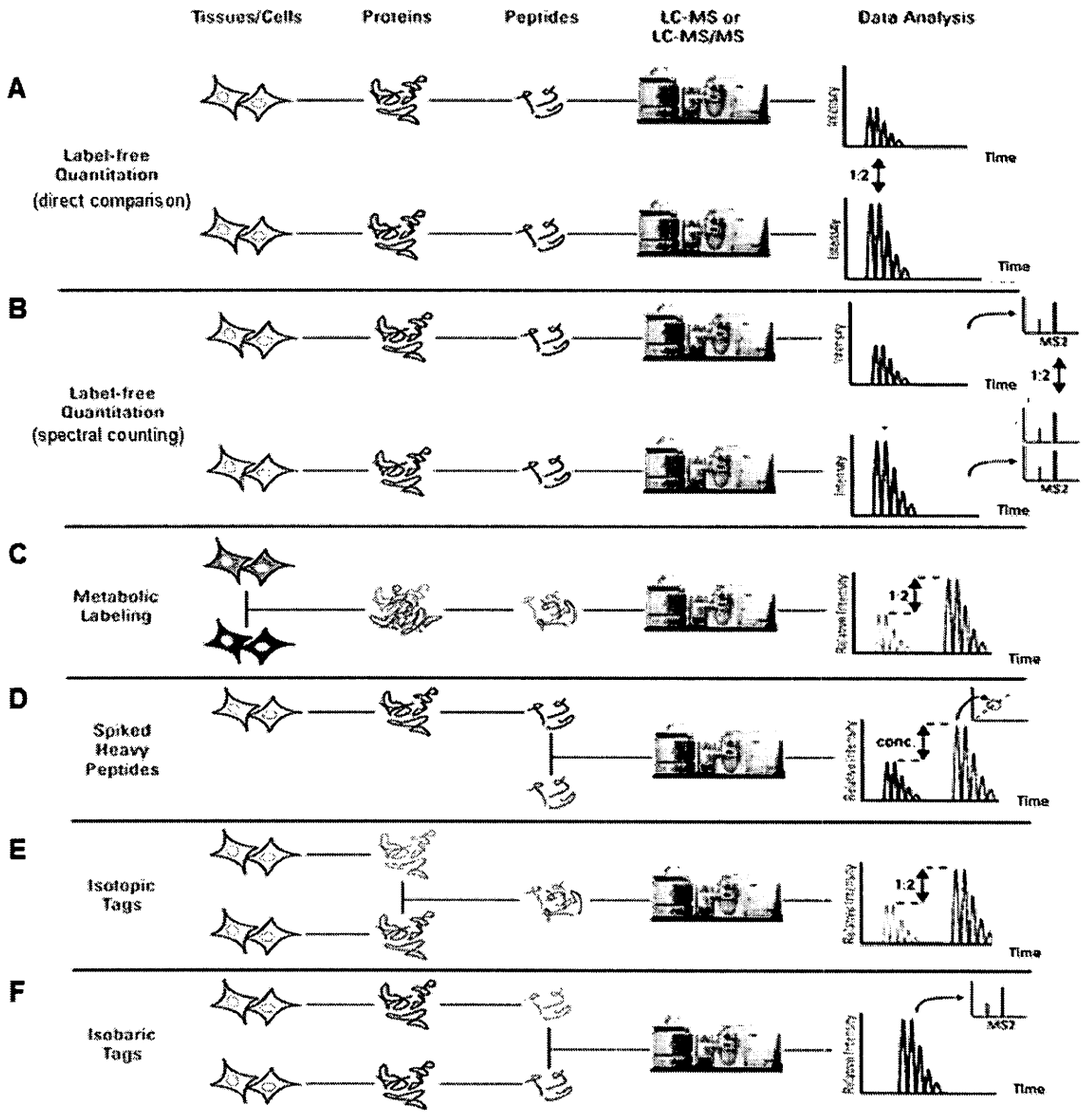


Figure 2.13. Overview of quantitative proteomics methods. *The drawings are color quoted to indicate the point in each workflow when samples are isotopically labeled (indicated by blue [light] and red [heavy]) for LC-MS analysis. Figure is adapted [24].*

The basic principle of the three remaining isotope-labeling methods, shown in Figure 2.13 (D-F), is that the peptides from the two samples being compared are labeled, one with light isotopes and the other with heavy isotopes. The two samples, comprising equal amounts of total proteins, are then combined and analyzed by LC-MS analysis. The individual peptides labeled with light and heavy isotopes are chemically identical, and therefore co-elute during LC separation and are detected simultaneously during MS analysis. The peak intensities of the heavy and light peptides are compared and the relative expression of the protein in the two samples can then be determined. The three methods differ in the way the isotopic labels are incorporated into the peptides.

'Metabolic labeling' (e.g. SILAC) uses living cell samples to metabolize isotopically labeled amino acids and incorporate them into their protein products. The proteins are then extracted from the cell samples and are combined and quantified via MS1 analysis (Figure 2.13D). The advantage of this method is that the heavy and light protein samples are combined before sample preparation for LC-MS analysis, thus the level of quantification bias from processing errors is low.

'Isotopic labeling' can be performed enzymatically (e.g. GIST, which will not be discussed here) or chemically (e.g. ICAT). Isotope-coded affinity tag (ICAT) labeling modifies a peptide's cysteine residues with a group containing zero or eight deuterium atoms. After labeling, the samples are combined, and the proteins are quantified via MS1 analysis (Figure 2.13E). The drawback is that ICAT will not work for the small fraction of proteins that lack a cysteine residue; in addition, cysteine is not a common residue and ICAT-labeling thus produces few peptides for quantification. This on one hand greatly

simplifies sample complexity, but on the other leads to much reduced redundancy which may lead to lower analytical reliability.

With 'isobaric labeling' (e.g. TMT and iTRAQ) peptides are tagged with various chemical groups that have the same mass (isobaric), thus in the MS1 spectrum the same (albeit differently labeled) peptide from different samples appears as a single peak, but after fragmentation, in the MS2 spectrum, labeled peptides from different samples yield reporter ions of different mass (Figure 2.13F). The main advantage of 'isobaric labeling' is that it allows the comparison of multiple (up to 8) samples at the same time. Isobaric tags for relative and absolute quantitation (iTRAQ) were exclusively used in this work and will be further discussed in the next section.

2.3.1. Isobaric Tags for Relative and Absolute Quantification

iTRAQ tags are isotopically variable, chemical reagents that are designed to bind to primary amine (NH_2) groups, such as those found at the N-terminus and the side chain of lysine residues in tryptic peptides. iTRAQ is available in sets of four or eight tags (4-plex and 8-plex, respectively). In this research work, 4-plex tags were used.

An iTRAQ tag consists of three parts: a 'reporter' group, N-methylpiperazine; a 'balancer' region; and a 'linker', N-hydroxy succinimide (NHS), which reacts with the primary amines of the peptide (Figure 2.14A). Thus, every peptide should have at least one label at the N-terminus and, for peptides containing lysine residues, a label on each lysine side-chain. The 4-plex kit can simultaneously compare up to four proteomic samples, and contains reagents with reporter ion masses of 114.1 Da, 115.1 Da, 116.1 Da and 117.1 Da (Figure 2.14A, C).

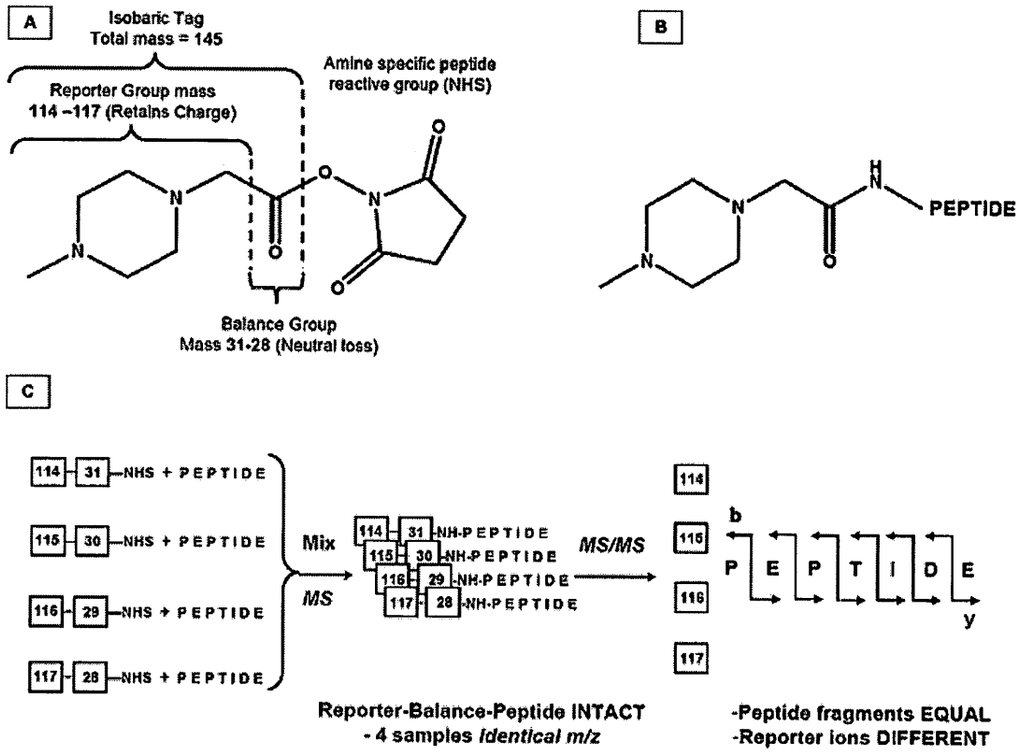


Figure 2.14. iTRAQ labeling of peptides for quantitative proteomics. *The chemical structure of iTRAQ reagents is shown in (**A**).The reporter group utilizes stable isotopes of carbon and nitrogen, whereas the balance group utilizes stable isotopes of carbon and oxygen, to form isobaric tags. Peptides are labeled on their amino groups (**B**). Four samples, each with a different isobaric tag, show identical m/z values in the MS1 mode of the mass spectrometer. Upon fragmentation in the MS2 mode, the reporter group is released and allows relative quantification in addition to the peptide sequence information (**C**). The figure is reproduced from Ross et al. [25].*

The 'balancer' region between the reporter and the linker has masses of 31, 30, 29 and 28 Da, respectively, to compensate for the difference in the mass of each reporter; this ensures that the mass of the tag, as a whole, is isobaric for all four tags.

Consequently, iTRAQ-labeled peptides of an identical sequence from different samples in a set are all isobaric and appear as a single precursor ion peak of identical *m/z* values. When the precursor ion peak is selected for fragmentation, the reporter and other sequence ions are produced (Figure 2.15). The former have four different m/z values, and the ratio among the four gives quantitative information, while the latter collectively give qualitative information and permits identification of the peptide in question.

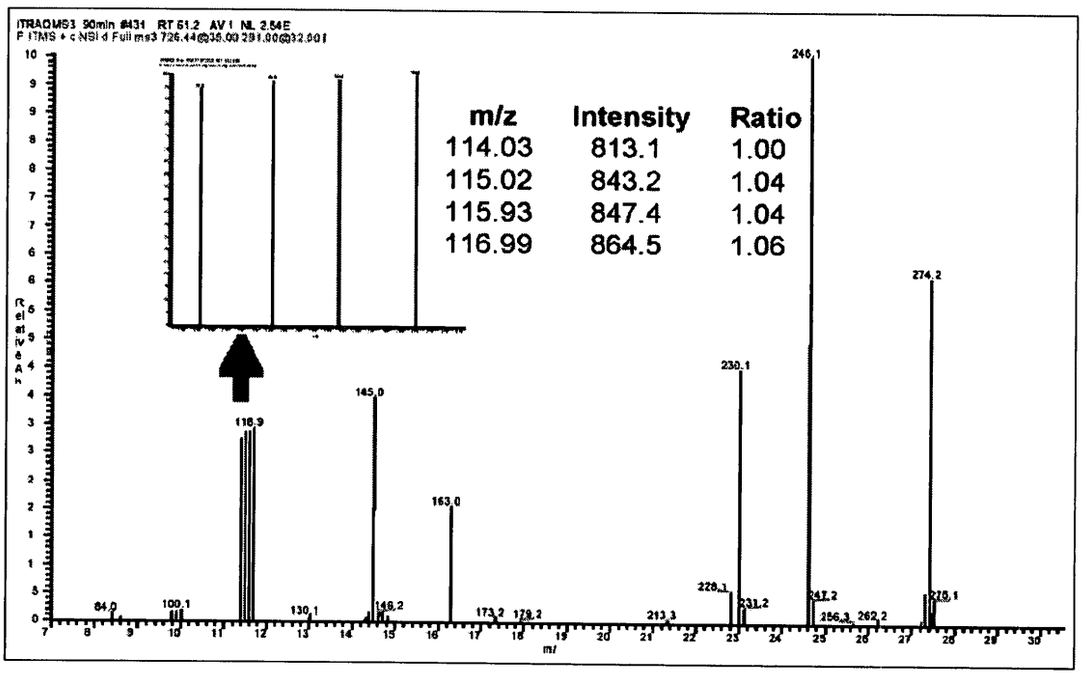


Figure 2.15. Quantification of the iTRAQ labels from the y-1 ion of L*VNELTEFAK* from BSA *(Figure is reproduced from Thermo Electron Corporation website). MS2 spectrum generated from a pool of four isobaric, iTRAQ labeled, tryptic peptides. A zoom image of spectrum shows the four 'reporter' groups (mass 114-117 Da) from which relative quantification of BSA from each sample can be calculated. The rest of the peaks can be used for BSA identification. Figure is reproduced [26].*

2.4. High-performance liquid chromatography

In proteomic analysis, a tissue homogenate may consist of thousands of proteins. After trypsin digestion, the number of peptides can increase to tens of thousands. This complex mixture must be resolved before MS analyses can be performed. A common analytical tool used for peptide separation is HPLC. To be compatible with the version of ESI employed commonly in proteomic analyses, which consumes nL / min of samples, nano-HPLC is typically implemented.

Nano-HPLC employs automated systems with precisely loaded samples, controlled flow rates in the nL / min range, high pressures of up to 450 bars, and on-line sample detection. The chromatographic column consists of a fused-silica capillary, packed with 3- to 5-μm diameter porous silica, usually coated with a hydrophobic material for reverse phase (RP) or antibodies for affinity chromatography; highly cross-linked styrene-divinylbenzene for size exclusion chromatography (SEC); or ion exchange resin for ion-exchange (IEX) chromatography.

As alluded to earlier, in 'bottom-up' proteomics, the analysis requires a proteolysis of all proteins into peptides that generates an even more complex mixture; thus, one-dimensional separation is typically not sufficient and additional separation dimensions must be introduced in order to reduce the sample complexity.

Size exclusion chromatography is a method in which large molecules, such as proteins or macromolecular complexes in solution, are separated by their size; thus, for tryptic peptides this technique is not the best choice. Also, affinity chromatography is not

applicable because it is based on specific interactions in a lock-and-key fashion between analytes (protein) and matrix-bound ligands (protein specific antibodies). IEX chromatography in both anion and cation forms is considered by many to be most suitable for the analysis of tryptic peptides, especially in tandem with RP chromatography. In this research work, strong cation exchange (SCX) chromatography was used in combination with RP chromatography.

2.4.1. Strong cation exchange chromatography

The separation of tryptic peptides using SCX chromatography is based on the charge of the peptides. The SCX column is packed with stationary phases consisting of cation-exchange resins with acidic sites such as the sulfonic acid group $-SO_3^-H^+$. When aqueous solutions of cationic peptides (M^{x+}) are introduced on the column, an exchange equilibrium ensues:

$$xRSO_3^-H^+ + M^{x+} \rightleftharpoons (RSO^{-3})_xM^{x+} + xH^+$$

When other cations that also have the affinity for $-SO_3^-H^+$, for example K^+, are introduced, they compete with the protonated peptides, *viz*,

$$(RSO^{-3})_xM^{x+} + K^+ \rightleftharpoons (RSO^{-3})_3K^+ + M^{x+}$$

causing some of the peptide ions on stationary phase to be transferred to the mobile phase. These peptide ions move down the column in a series of transfers between stationary and mobile phases. The peptide that has a higher affinity for the stationary

phase interacts with stationary phase longer and, therefore, has a longer retention time. In this research work, peptides were eluted using a gradient of potassium chloride.

2.4.2. Reverse Phase Chromatography

The RP chromatography is based on the partitioning of a sample between a hydrophobic stationary phase and a polar hydrophilic mobile phase. Here, the stationary phase is usually a siloxane with a C_8 chain (n-octyl) or C_{18} chain (n-octyldecyl). Peptides having greater hydrophobicity (often containing aromatic residues) partition more into the hydrophobic C_8 (or C_{18}) phase and, hence, elute later than those with lesser hydrophobicities. More effective competition with the stationary phase is rendered by increasing the hydrophobicity of the mobile phase (resulting in a shorter retention time). This is done by increasing the concentration of an organic solvent in the mobile phase, such as acetonitrile or methanol. As in SCX chromatography, the peptide ions move down the column in a series of transfers (partitioning steps) between stationary and mobile phases. The RP mobile phase is compatible with ESI; thus, the eluate from RP-LC is usually directly routed to the electrospray probe for on-line MS analysis.

2.5. Clustering analysis

After identifying and quantifying the proteins by LC-MS analysis, dysregulated proteins were selected that could potentially serve as cancer biomarkers. This selection was aided using clustering analysis to determine which of the dysregulated proteins could

consistently discriminate between cancer and normal samples. In clustering analysis, a set of objects are assigned to groups called clusters, so that the objects in the same cluster are more similar to each other, based on specific criteria, than to those in other clusters. In our case, the objects are the tissue samples and the criteria are the levels of protein expression.

Clustering is a common technique for data mining used in many fields, including bioinformatics. Clustering software are available, some of the most widely used include Unscrambler® X, MVCP, and Cluster 3. In bioinformatics, the most popular algorithm is the "shortest distance" approach, where objects are clustered such that those belonging to the same cluster have the shortest distance between them. Here, "distance" is used metaphorically to mean the difference in the protein expression of two samples being compared. Similar proteins expressions are indicated by a shorter distance between the two samples. These distances can be based on a single dimension (one protein) or multiple dimensions (many proteins). Tree clustering is a method to visualize the distances between the objects when forming clusters.

The distances can be computed using the 'Chebychev distance', 'power distance', 'Euclidean distance', 'squared Euclidean distance', 'city-block distance', and other methods; in biomarker studies, the last three are most commonly used. 'Euclidean distance' is simply the geometric distance in multidimensional space. It is computed as:

$$d(x,y) = \left(\sum (x_i - y_i)^2\right)^{\frac{1}{2}} \qquad (5)$$

The advantage of this method is that the distance between any two objects is not affected by the addition of new objects to the analysis, which may be outliers. However, the distances can be greatly affected by differences in scale among the dimensions from which the distances are computed.

If one wants to put a greater weight on the objects that are further apart, one may square the standard Euclidean distance:

$$d(x,y) = \sum (x_i - y_i)^2 \qquad (6)$$

Note that the Euclidean and squared Euclidean distances are usually computed from raw data, and not from standardized data.

The 'city-block distance' is also known as the 'rectilinear distance' or 'Manhattan length', in which the distance between two points is the sum of the absolute differences of their coordinates. In most cases, this distance yields results similar to the Euclidean distance; however, the effect of single large differences, for example outliers, is dampened, since they are not squared. The city-block distance is computed as:

$$d(x,y) = \sum |x_i - y_i| \qquad (7)$$

In the research work presented in this thesis, the data used for clustering analysis were standardized, therefore, it was appropriate to use the 'city-block distance'.

2.5.1. City-block distance

The city-block distance calculation used in the analysis can be explained and demonstrated on a simple example of finding the distance a taxi travels to its destination. Suppose a taxi drove to the North 7.0 kilometers and then to the West 2.4 kilometers; the distance it covered can be calculated using Eqn. 7:

$$d(x,y) = |0 - 7.0| + |0 - 2.4| = 9.4 \ \ km$$

The 'city-block distance' can be utilized to quantify differences or similarities in the protein expression between tissue samples and cluster the samples accordingly. Suppose there are five samples (S1-S5) and we know the expressions of proteins in them, the distances between each pair of samples are first calculated using Eqn. 7 (see Table 1).

Table 2.1. Distances between samples calculated base on 'city-block distance' method using protein expressions.

	Sample					Distance									
Protein	S1	S2	S3	S4	S5	S1/S2	S1/S3	S1/S4	S1/S5	S2/S3	S2/S4	S2/S5	S3/S4	S3/S5	S4/S5
A	1.1	2.4	2.4	0.8	0.8	9.1	9.2	1.0	1.1	5.3	9.3	9.4	9.4	9.5	0.9
B	0.9	5.6	4.2	1.0	0.9										
C	1.0	3.1	1.9	1.1	0.7										
D	1.2	1.3	2.6	0.9	1.1										
E	0.8	1.7	3.1	1.0	1.2										

For example, the distance between samples S1 and S2 (S1/S2) was calculated as:

$$d(x,y) = |1.1 - 2.4| + |0.9 - 5.6| + |1.0 - 3.1| + |1.2 - 1.3| + |0.8 - 1.7| = 9.1$$

Next, the samples were arranged in two clusters, where the samples from the same cluster have a small distance between them, while those from the different clusters have a large distance (see Figure 2.16A). The tree clustering representation, resulting from this analysis, is shown in Figure 2.16B.

Cluster analysis is not a typical statistical test; its purpose is to "put objects into clusters," using a combination of algorithms. Unlike many other statistical procedures, cluster analysis methods are mostly used when there are no a priori hypotheses, and research is still in the exploratory phase; in a sense, it finds the "most significant solution possible." Therefore, statistical significance testing is really not applicable nor appropriate here [27].

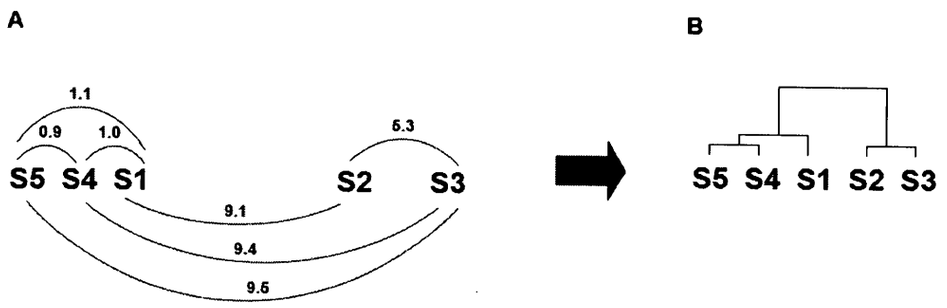


***Figure 2.16. Principle of clustering analysis. A**: the distances between each pair of samples were calculated based on the expressions of proteins in these samples using 'city-block distance' method; **B**: the tree clustering representation.*

In this research work, the clustering analysis was performed using Cluster 3.0 software and the result was visualized with TreeView software (Stanford University, Palo Alto, CA, http://www.dnachip.org), both of which were developed by Eisen et al. [28].

2.6. Verification of biomarkers

Differential expressions of the biomarker candidates, identified and quantified by MS, were verified by two independent techniques: Western blotting (WB) and immunohistochemistry (IHC). The verification ensures that the observed differential expressions are not caused by some unforeseen bias of the mass spectrometric method. Both techniques allow the quantification of a specific protein in a complex matrix by utilizing the very specific affinity of an antibody for the protein. The difference between the two techniques is that WB analyzes tissue homogenates or secretome solutions, allowing the sample to be concentrated for greater sensitivity, and the proteins are quantified instrumentally. IHC is performed on tissue sections typically suitable for histology, which reveals protein localization; however, protein quantification is more subjective and in according to a set of fixed criteria that typically include stain intensity and percent of cells stained.

2.6.1. Western blotting

2.6.1.1. Gel electrophoresis

In the first step of WB, the proteins are separated using gel electrophoresis either by their isoelectric point (pI), molecular weight, electric charge, or a combination of these. The nature of the separation depends on the treatment of the sample and the nature

of the gel. The proteins can be separated by their native 3-D structure or denatured and then separated by the length of the polypeptide. In the work presented in this thesis, denatured proteins were separated by their molecular weight using SDS-PAGE (sodium dodecyl sulfate - polyacrylamide gel electrophoresis). The proteins were denatured using 2-mercaptoethanol to reduce and break disulfide bonds [S-S] to give sulfhydryl groups [SH + SH], thereby disrupting the secondary and higher order structures. The denatured proteins were loaded into wells in the gel and a voltage was applied along the length of the gel to cause the proteins, covered in the negatively charged SDS, to migrate to the positively charged electrode. The proteins travel through the polyacrylamide gel matrix at different speeds depending on their size, the smaller ones migrating furthest.

2.6.1.2. Transfer and blocking

Once separated, the proteins were then transferred from the gel to a polyvinylidene difluoride (PVDF) membrane by sandwiching the gel and the membrane between two sheets of thick filter paper, and applying an electric current across the sandwich (positive on the PVDF side). The membrane was then treated with non-fat dry milk to prevent the non-specific binding of antibodies to the membrane in the subsequent step; this step is called "blocking." The milk proteins bind to the membrane where there are no transferred proteins; thus, when antibodies are added in the following step, they cannot bind by non-specific hydrophobic interactions, but must bind to their target proteins.

2.6.1.3. Detection

After blocking, the membrane was incubated with a primary antibody specific to the protein of interest (Figure 2.17). Such antibodies are generated by the immune system of a host such as a mouse or a rabbit, or by an immune cell culture, when exposed to the protein of interest. The membrane was then treated with a secondary antibody directed at a part of the primary antibody. The secondary antibody is covalently bound to an enzyme such as horseradish peroxidase (HRP) or alkaline phosphatase, which can catalyze a signaling chemical reaction in proportion to the amount of target protein present. The reaction can be chromogenic, resulting in the deposition of an insoluble colored precipitate; fluorogenic, resulting in the deposition of a fluorescent material; or chemiluminescent, resulting in the production of light.

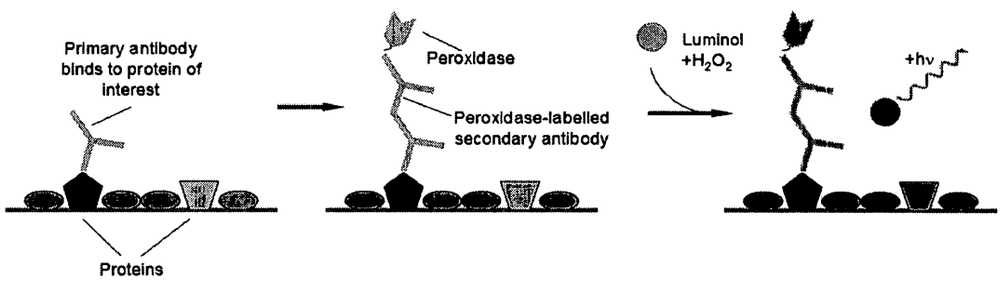


Figure 2.17. Protein detection in Western blot.

In this research work, a horseradish peroxidase-linked secondary antibody was used to catalyze the chemiluminescent peroxide oxidation of luminol. The luminol generates an excited state, the 3-aminophthalate intermediate. This intermediate relaxes to a lower energy state and releases photons. This is presented pictorially in Figure 2.18.

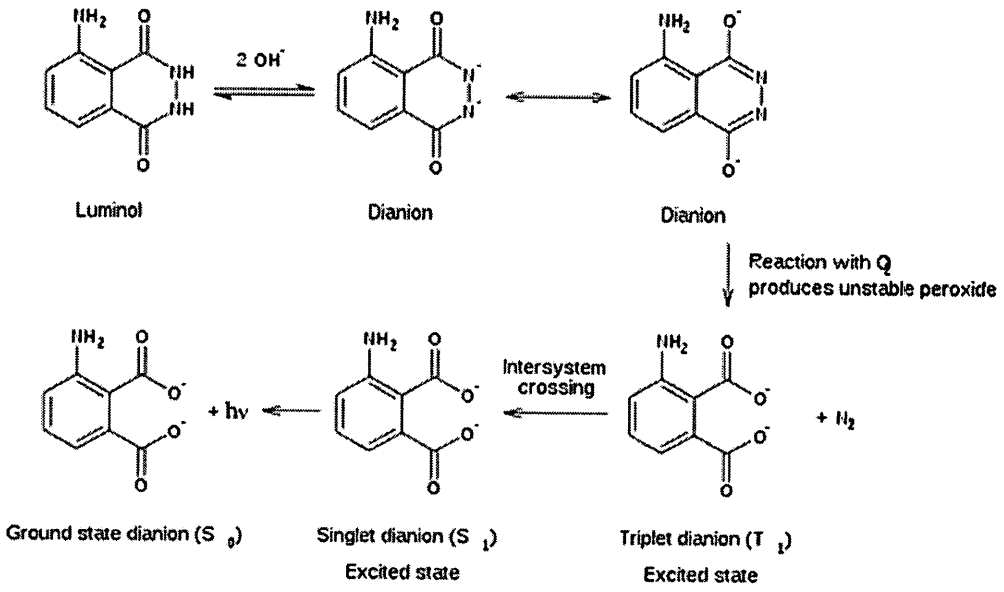


Figure 2.18. Chemiluminescent reaction of luminol. *Figure is reproduced from Thermo Scientific web page [29].*

The chemiluminescence was captured on photographic film, and the image was analyzed by densitometry using ImageJ, a publicly available Java-based image processing program [30].

2.6.2. Immunohistochemistry

IHC visualizes the distribution and localization of target proteins within cells and tissue sections by binding target-specific, dye-labeled antibodies to the proteins. Sectioned multiple miniature tissue samples were arranged on a slide for comparative analysis, forming a tissue microarray (TMA). Appropriate sample preparation and treatment is critical to maintaining cell morphology, tissue architecture, and the antigenicity of target proteins.

2.6.2.1. Tissue fixation and embedding

To preserve cell morphology and tissue architecture, the tissue was first fixed with formaldehyde. The formaldehyde cross-links the primary amino groups of proteins with other nearby nitrogen atoms in proteins or DNA through a $-CH_2-$ linkage. Fixed tissue samples were embedded in paraffin to maintain the architecture of the sample during storage and sectioning for IHC. Formalin-fixed paraffin-embedded tissues were used in this work. Samples too sensitive for chemical fixation or exposure to solvents (used to remove paraffin) could be encased in a cryogenic embedding medium and then snap-frozen in liquid nitrogen.

2.6.2.2. Sectioning and mounting

Formalin-fixed paraffin-embedded tissues were sectioned into 5 μm slices with a microtome. These sections were then mounted onto electrostatically charged glass slides, which leave amino groups on the surface of the glass to which the tissue can directly couple. Frozen sections could be cut using a pre-cooled cryostat, mount to adhesive glass slides, dry overnight at room temperature, and fix by immersion in pre-cooled (-20°C) acetone.

2.6.2.3. Epitope recovery

Paraffin and $-CH_2-$ bridges can mask the protein epitopes and prevent antibody binding. To unmask them, tissues were deparaffinized in xylene and treated by heat-induced epitope retrieval (HIER). Heat causes cross-linked protein epitopes to 'unfold', while a buffer solution maintains the conformation of the 'unfolded' protein.

2.6.2.4. Quenching/blocking endogenous target activity

For imaging approaches that depend on biotin, peroxidases or phosphatases for the amplification or enzymatic detection of target proteins, it is necessary to quench endogenous forms of these proteins to avoid false positive detection and high background signals. In this research work, sections were incubated with saturating amounts (0.3% v/v) H_2O_2 which resulted in irreversible inactivation of endogenous peroxidases.

2.6.2.5. Blocking nonspecific sites

Although antibodies show preferential affinity for specific epitopes, antibodies may weakly bind to non-target proteins, through "reactive sites" that are similar to the cognate binding sites on the target protein. This nonspecific binding causes background imaging that can mask the detection of the target antigen. To reduce background imaging, the tissues were incubated with a buffer that blocked the reactive sites to which the primary or secondary antibodies might otherwise bind. Common blocking buffers may contain normal serum, non-fat dry milk, BSA or gelatin. In this research work, tissues were blocked with 10% FBS (fetal bovine serum) in PBS.

2.6.2.6. Immunodetection

IHC target antigens are detected through chromogenic or fluorescent imaging, and the type of readout depends on the experimental design. In fluorescence detection, the primary or secondary antibody is conjugated to a fluorophore, and the location of the antigens can be seen by fluorescence microscopy. In chromogenic detection, the antibody

is conjugated to an enzyme such as HRP or alkaline phosphatase, which catalyzes a color forming reaction that deposits a dark precipitate at the antigen site. The sites can be seen by optical transmission microscopy. In this work, chromogenic detection was used. Tissues were first introduced to primary antibodies specific to the proteins of interest, and then to biotin-tagged secondary antibodies having an affinity for the primary antibodies (shown in Figure 2.19).

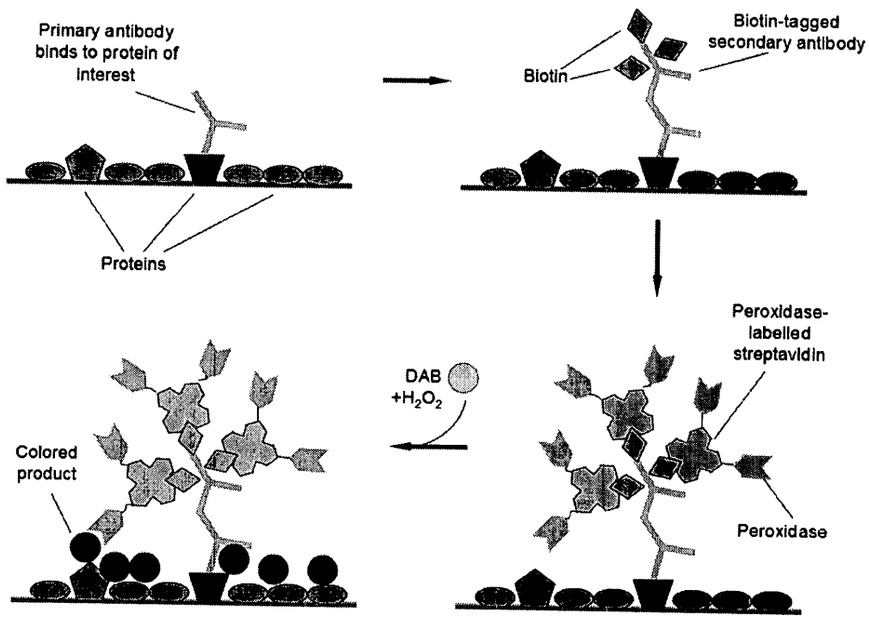


Figure 2.19. Protein detection in immunohistochemistry.

The biotin-tags were used to bind HRP-labeled streptavidin; biotin and streptavidin have an unusually high affinity and specificity for each other. When 3,3'-

diaminobenzidine (DAB) was added to the tissue section in the presence of H_2O_2, HRP catalyzed the oxidative polymerization of DAB, producing a colored, insoluble precipitate at the antigen sites. The signal was amplified because each secondary antibody contained several biotins and streptavidin was labeled with two HRP.

2.6.2.7. Counterstaining and mounting

After IHC staining of the target proteins, a second stain is often applied to provide contrast that helps the IHC stain to stand out. Many of these counterstains show specificity for organelles, while others stain the whole cell. Chromogenic and fluorescent counterstains are commercially available for IHC and include: hematoxylin, Hoechst stain, and 4',6-diamidino-2-phenylindole (DAPI). In this work, hematoxylin was used, which, when oxidized and combined with aluminum ions, formed a metal-dye complex that stained the nuclei of mammalian cells blue by binding to lysine residues on nuclear histones. After staining, the tissue samples were preserved for long-term usage and storage by mounting a coverslip with a DPX mountant (distyrene, plasticizer and xylene mixture).

References

1. Banerjee S, Mazumdar S. Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. Int J Anal Chem. 2012;2012:282574. Epub 2011 Dec 15.
2. Gross J. Mass Spectrometry. Springer-Verlag Berlin Heidelberg 2011, 2nd ed., DOI 10.1007/978-3-642-10711-5, 12.
3. Steenwyk RC, Hutzler JM, Sams J, Shen Z, Siuzdak G. Atmospheric pressure desorption/ionization on silicon ion trap mass spectrometry applied to the quantitation of midazolam in rat plasma and determination of midazolam 1'-hydroxylation kinetics in human liver microsomes. Rapid Commun Mass Spectrom. 2006;20(24):3717-22.
4. Crotti S, Seraglia R, Traldi P. Some thoughts on electrospray ionization mechanisms. Eur J Mass Spectrom (Chichester, Eng). 2011;17(2):85-99. Review.
5. Taylor G (1964) "Disintegration of Water Droplets in an Electric Field". Proc. Roy. Soc. London. Ser. A 280 (1382): 383.
6. Nguyen S, Fenn JB (January 2007). "Gas-phase ions of solute species from charged droplets of solutions". Proc. Natl. Acad. Sci. U.S.A. 104 (4): 1111–7.
7. Dole M, Mack LL, Hines RL, Mobley RC, Ferguson LD, Alice MB (1968). "Molecular Beams of Macroions". Journal of Chemical Physics 49 (5): 2240–2249.
8. Banerjee S, Prakash H, Mazumdar S. Evidence of molecular fragmentation inside the charged droplets produced by electrospray process. J Am Soc Mass Spectrom. 2011 Oct;22(10):1707-17. Epub 2011 Jul 7.

9. Zhou S, Cook KD. A mechanistic study of electrospray mass spectrometry: charge gradients within electrospray droplets and their influence on ion response.
10. Wittmann C. Fluxome analysis using GC-MS. Microb Cell Fact. 2007 Feb 7;6:6.
11. Laskin J, Liftshitz C. Intramolecular Vibrational Energy Redistribution and Ergodicity of Biomolecular Dissociation in Principles of Mass Spectrometry Applied to Biomoelcules., Wiley Interscience, Hoboken, New Jersey, 2006, pp 239-275.
12. Domon B. and Aebersold R. Mass Spectrometry and Protein Analysis. Science. 2006; 312, 212-217.
13. Goetz, A.; Platzner, I. T. (Itzhak Thomas); Habfast, K.; Walder, A. J. (1997). Modern isotope ratio mass spectrometry. London: J. Wiley.
14. http://www.amptek.com/a111
15. htmlhttp://what-when-how.com/remote-sensing-from-air-and-space/energy-in-electromagnetic-waves-remote-sensing/
16. AB/ Sciex QSTAR user manual.
17. Liu H, Sadygov R G, and Yates J R, III. A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. Anal. Chem. 2004 76 4193 4201
18. Wang N, Li L. Exploring the precursor ion exclusion feature of liquid chromatography-electrospray ionization quadrupole time-of-flight mass spectrometry for improving protein identification in shotgun proteome analysis. Anal Chem. 2008 Jun 15;80(12):4696-710. Epub 2008 May 15.

19. Masui O, White N M A, DeSouza LV, Krakovska O, Matta A, Metias S, Khalil B, Romaschin A D, Honey R J, Stewart R, Pace K, Bjarnason G A, Siu K W M, Yousef G M. Quantitative proteomic analysis in metastatic renal cell carcinoma reveals a unique set of proteins with potential prognostic significance. Paper was resubmitted, after minor corrections, on August 28, 2012 in "Molecular and Cellular Proteomics" (ID#: MCP/2012/020701).
20. http://biochem.ncsu.edu/ MassSpec/docs/Peptide SequencingMS.pdf.
21. Thiede B, Lamer S, Mattow J, Siejak F, Dimmler C, Rudel T, Jungblut PR. Analysis of missed cleavage sites, tryptophan oxidation and N-terminal pyroglutamylation after in-gel tryptic digestion. Rapid Commun Mass Spectrom. 2000;14(6):496-502.
22. Steen H, Mann M. The ABC's (and XYZ's) of peptide sequencing. Nat Rev Mol Cell Biol. 2004 Sep;5(9):699-711. Review.
23. Zhu W, Smith JW, Huang CM. Mass spectrometry-based label-free quantitative proteomics. J Biomed Biotechnol. 2010;2010:840518. Epub 2009 Nov 10. Review.
24. http://www.piercenet.com/browse.cfm?fldID=754E41C1-8B46-444E-AEE8-E33E5D37A1DE#icat
25. P.L. Ross, Y.N. Huang, J.N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson, D.J. Pappin. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol. Cell. Proteomics, 3 (2004), pp. 1154–1169.

26. http://www.thermo.com/eThermo/CMA/PDFs/Various/File_27402.pdf

27. http://www.obgyn.cam.ac.uk/cam-only/statsbook/stcluan.html

28. Eisen M B, Spellman P T, Brown P O, and Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. U. S. A 95, 14863-14868, 1998.

29. http://www.piercenet.com/browse.cfm?fldID=F95B91A9-3DC1-4B56-8E8D-59CA044A8BA7

30. http://rsbweb.nih.gov/ij/

Chapter 3

Identification of differentially expressed proteins in primary clear cell renal cell carcinoma (ccRCC) tissues as potential diagnostic biofluid-based biomarkers

In the research described in this chapter, diagnostic, fluid-based biomarkers for the onset of primary ccRCC were sought, using iTRAQ LC-MS analysis to determine differentially expressed proteins in primary ccRCC, relative to normal tissue homogenates. From the resulting pool of dysregulated proteins, those proteins having secretory function were selected based on bioinformatic analysis, due to their potential as biofluid-based biomarkers. The dysregulation of these proteins was verified in two independent sets of tissue samples by WB and IHC analyses and is currently undergoing verification in the serum and urine of ccRCC patients using WB analysis.

This study was a collaboration of research groups from the Department of Chemistry and the Centre for Research in Mass Spectrometry at York University, in Toronto, Ontario, Canada; the Keenan Research Center in the Li Ka Shing Knowledge Institute and the Department of Laboratory Medicine, St. Michael's Hospital, Toronto, Canada; the Department of Laboratory Medicine and Pathobiology, University of Toronto, Canada; the Department of Surgery, St. Michael's Hospital, Toronto, Canada; and the Division of Medical Oncology and Hematology, Sunnybrook Health Sciences, Toronto, Canada.

The group from Li Ka Shing Knowledge Institute was headed by Professor George M. Yousef and included Dr. Nicole M. A. White, Dr. Alexander D. Romaschin,

Shereen Metias, and Bishoy Khalil. They performed pathological analyses of kidney tissues, carried out tissue isolations and purifications, constructed TMAs, quantified immunoexpression of proteins, performed affinity chromatography of sera, and performed bioinformatic analyses.

The group from York University's Centre for Research in Mass Spectrometry was headed by Professor K. W. Michael Siu, and included Dr. Leroi V. DeSouza, Dr. Olga Krakovska and the author of this thesis. I performed the trypsin digestion of the proteins, iTRAQ labeling of the resulting peptides, SCX chromatography and RP LC-MS analysis, Protein Pilot database search, clustering analysis, SignalP, SecretomeP and other bioinformatics analyses, WB analysis and quantitation, IHC analysis, and developed the Excel template for the automated generation of PIE lists. Recalculation of iTRAQ ratios was carried out by Dr. Olga Krakovska. This chapter presents details of this study.

Abbreviations

AHNAK: neuroblast differentiation-associated protein AHNAK; ccRCC: clear cell renal cell carcinoma; CE: collision energy; EF: error factor; ENO1: alpha-enolase; FDR: false discovery rate; HSPB1: heat shock protein beta-1; HSPE1: 10 kDa heat shock protein, mitochondrial; IDA: information-dependent acquisition; IHC: immunohistochemistry; iTRAQ: isobaric tags for relative and absolute quantitation; LDHA: L-lactate dehydrogenase A chain; MS: mass spectrometry; PBS: phosphate buffered saline; RCC: renal cell carcinoma; PTM: post translational modification; RP: reverse phase; SCX: strong cation exchange; TMA: tissue microarray; UniProtKB: UniProt Knowledgebase; WB: Western blot.

Abstract

BACKGROUND: There are currently no biomarkers for the early and accurate detection of clear cell renal cell carcinoma (ccRCC). Diagnosis of cancer, and the decision to use nephrectomy, rely on imaging studies which are not always accurate.

METHODS: We employed high throughput quantitative proteomics, using isobaric tags for relative and absolute quantitation (iTRAQ) labeling and LC-MS/MS analysis, to identify proteins that are differentially expressed in ten paired ccRCC and normal kidney tissue samples. The involvement of dysregulated proteins in carcinogenesis and their "secretory" potential were investigated by in-silico analysis. Secretory proteins with potential diagnostic utility were verified by Western blot and immunohistochemistry analyses on two independent sets of tissue samples.

RESULTS: 55 proteins were identified to be significantly dysregulated in ccRCC, versus normal, kidney tissues. Of these, 54 have been reported to play a role in carcinogenesis, and 39 are secreted proteins. Alpha-enolase (ENO1), L-lactate dehydrogenase A chain (LDHA), heat shock protein beta-1 (HSPB1), and 10 kDa heat shock protein, mitochondrial (HSPE1) have been reported to be involved in metabolism, growth, proliferation, apoptosis, cell cycle, and hypoxia. By contrast, neuroblast differentiation-associated protein (AHNAK) has no records of involvement in any of these processes. The dysregulation of these 5 proteins was verified in two independent sets of patients.

CONCLUSIONS: Our study is the comprehensive quantitative proteomics analysis in ccRCC, which can lead to the development of an accurate test for the early detection of kidney cancer and the confirmation of the nature of kidney masses without

the need for invasive biopsies. Confirmation of the upregulation of AHNAK, ENO1, LDHA, and downregulation HSPB1 in the serum or urine of ccRCC patients would support their utility as biofluid-based biomarkers for the early diagnosis of ccRCC. Such biomarkers would greatly improve patient treatment and increase overall survival.

Introduction

Renal cell carcinoma (RCC) is the most common neoplasm in the adult kidney, with an increasing incidence over the past 20 years (1). Histopathologically, about 80% of RCCs are of the clear-cell type, 15% are papillary, and the remaining 5% are other types. Early diagnosis of RCC is associated with a favorable prognosis (5-year survival rate ~ 85%). Unfortunately, RCC is often asymptomatic, with about 30% of patients diagnosed at the metastatic stage when the prospects for cure are dismal (5-year survival rate ~9%) (2). Traditional methods of screening (history, physical examination and urine analysis) for detection of asymptomatic RCC are ineffective (3). The diagnosis of RCC, and the subsequent resection of the kidney are based on incidental imaging, which are not always accurate. There are currently no biomarkers available for the early diagnosis of RCC or for determining the nature of renal masses. A non-invasive test, for example, a biomarker that can be measured in serum or urine, will have a significant impact on patient management.

Few chromosomal abnormalities have been documented in RCC, including VHL mutation (3p-), 5q21+ (70%), and 14q- (41%) (4,5). The pathogenesis of RCC, however, is not yet fully understood. Understanding the tumor biology of RCC at the molecular level is essential to improve diagnosis, prognosis, and treatment options (6,7).

Proteomics, combined with mass spectrometry (MS), offers great promise for unveiling the complex molecular events of tumorigenesis and identifying cancer biomarkers. Proteomic technologies are being used in studies of dynamic protein expression, post-translational modifications, cellular and sub-cellular protein distribution,

and protein-protein interactions that have culminated in the identification of many cancer biomarkers.

Initial studies of the serum "spectral signatures" provided by surface enhanced laser desorption and ionization (SELDI) MS and MALDI-TOF MS (8.9) were promising, however, these studies were hampered by concerns regarding the reproducibility, artifacts of sample processing, and "black-box" approach where no proteins were identified or quantified, and no relevance of proteins to cancer biology was determined (10). Quantitative tissue proteomics is a promising alternative strategy for the discovery and identification of tumor biomarkers. One of the advantages of this strategy is that the relevant proteins are much more abundant in tissues. Additionally, it directly probes the protein profile of the diseased tissue rather than indirectly probing the profile of some systemic fluid that may or may not be changed by the disease.

In this study, we performed quantitative proteomic analysis using isobaric tags for relative and absolute quantitation (iTRAQ) labeling and LC-MS to identify proteins that were dysregulated in ccRCC versus normal, kidney tissues. We have identified a number of proteins that can distinguish between tumor samples and normal tissues with accuracy. We have also identified a number of secreted proteins that can serve as potential diagnostic markers. Finally, we have elucidated the potential involvement of these proteins in RCC pathogenesis. The most interesting proteins were verified by Western blot (WB) analysis on same tissue samples used for LC-MS analysis and by immunohistochemistry (IHC) on an independent set of tumor samples.

Materials and Methods

Specimen preparation and protein extraction

Clear cell RCC tissues and corresponding normal kidney tissues from the same patient were obtained from nephrectomy specimens at St. Michael's Hospital, Toronto, Canada. As ccRCC is known to arise from the proximal tubules (11), the kidney cortex is considered a suitable representation of normal kidney tissue (12). All specimens were histologically confirmed. The study was approved by the Research Ethics Boards of York University and St. Michael's Hospital.

Tissues were prepared as described previously (13-15). Briefly, tissues were homogenized in a protease-inhibitor cocktail (Roche, Laval, Canada). Cell debris was separated, and the clarified supernatant was used for analysis. A reference sample was prepared from a pool of 30 combined normal kidney tissues. Protein concentrations were determined using the Bradford assay (Sigma-Aldrich, St. Louis, USA) (13;15).

iTRAQ sample labeling

For iTRAQ LC-MS analysis, 100 μg of sample were denatured, disulfide bonds were reduced, and the cysteine residues were blocked as per the iTRAQ protocol (Applied Biosystems, Foster City, CA). Samples were then digested with trypsin and labeled with the iTRAQ tags (**Table 1**). Labeling of the reference sample was randomized for each set to eliminate any potential bias associated with a particular iTRAQ reporter tag. The iTRAQ-labeled samples were then dried using a vacuum centrifuge (Thermo Savant SC110 A, Holbrook, NY, USA).

Table 1. iTRAQ labeling of kidney tissue samples.

Set/iTRAQ tag	114	115	116	117
1	C1	C2	N1	RN
2	C3	N2	RN	C4
3	N3	RN	C5	C6
4	RN	N4	N5	C7
5	N6	N7	C8	RN
6	N8	C9	RN	N9
7	C10	RN	N10	

C: ccRCC tissue samples; N: normal kidney sample; RN: reference sample comprised of 30 normal kidney samples.

Strong cation exchange (SCX) chromatography

The iTRAQ sets were dissolved in 1.7 mL of Buffer A (15 mM KH_2PO_4 in 25% acetonitrile, pH 3.0) and filtered using a 0.45-μm syringe filter (Millipore, Cambridge, ON, Canada). Each set was then separated by off-line SCX chromatography using an HP1050 HPLC instrument (Agilent, Palo Alto, CA) with a 2.1-mm internal diameter x 100-mm-length PolyLC Polysulfoethyl A column packed with 5-μm beads with 300-Å pores (The Nest Group, Southborough, MA) as described previously (16). Separation was performed using a linear binary gradient over 1 h (**Table 2**). Buffer C was used to strip the column after the run. A total of 30 SCX fractions were collected per iTRAQ set. These fractions were dried using a vacuum centrifuge as before.

Table 2. LC gradient for strong cation exchange (SCX) chromatography

Time, min	0	2	58	60	65	75	80	90
Buffer A, %	100	100	0	0	0	0	100	100
Buffer B, %	0	0	100	100	0	0	0	0
Buffer C, %	0	0	0	0	100	100	0	0

Buffer A: 15 mM KH_2PO_4 in 25% acetonitrile, pH 3.0; buffer B: Buffer A containing 350 mM KCl; buffer C: Buffer A containing 1 M KCl.

Reverse phase (RP) LC-MS

The SCX fractions were analyzed in triplicate using precursor ion exclusion lists to minimize redundancy. Fractions were analyzed by a nanobore LC system (LC Packings, Amsterdam, Netherlands) and a QSTAR Pulsar mass spectrometer (Applied Biosystems/MDS SCIEX, Foster City, CA) in positive ion mode, externally calibrated with tryptic peptides from bovine serum albumin. The first five fractions were not analyzed because they consisted of the void volume which contained unreacted iTRAQ labels as well as byproducts that would compromise the reverse RP column. Fractions 6 – 17 were re-dissolved in 16 μL of eluant A [consisting of 94.9% deionized water, 5.0% methanol, and 0.1% formic acid (pH 3)]. For subsequent fractions, the amount of eluant A was incremented by 2 μL over the preceding fraction to accommodate the increase in the amount of KCl. A 1-μL aliquot of the sample (~1 μg of total peptides) was loaded onto a C18 RP pre-column (LC Packings: 300 μm x 5 mm) and desalted before separation on an RP analytical column (75-μm x 150-mm packed in-house with 3-μm Kromasil C18 beads with 100 Å pores, The Nest Group, Southborough, USA). Eluant A was used to load the sample onto the C18 pre-column at a flow rate of 25 μL min^{-1}. After 4 min, the C18 pre-column was switched in-line with the RP analytical column. Separation was performed at 100 nL min^{-1} using a nonlinear binary gradient (**Table 3**) starting with eluant A and transitioning to eluant B (5.0% deionized water, 94.9% methanol, and 0.1% formic acid).

Table 3. LC gradient for reverse phase LC

Time (min)	0.1	5	10	70	85	95	98	135
B (%)	5	5	15	35	80	80	5	Stop

Buffer A: 94.9% deionized water, 5.0% methanol, and 0.1% formic acid (pH 3); buffer B: 5.0% deionized water, 94.9% methanol, and 0.1% formic acid.

MS data were acquired in information-dependent acquisition (IDA) mode using the Analyst QS 1.1 software (Applied Biosystems/MDS SCIEX). The LC-MS analysis was performed using a 1-s TOF-MS survey scan from 400 to 1500 Da, followed by four, 2-s product-ion scans, from 80 to 2000 Da, of the four most-abundant ion peaks in the survey scan. The collision energy (CE) was automatically controlled by the IDA CE parameter script. Switching criteria were set for ions with $m/z \geq 400$ and <1500, charge states of +2 to +4, and abundances of ≥10 counts. Using Analyst QS 1.1 controlled dynamic exclusion, former target ions were excluded for 30 s, and ions within a 100-ppm window were ignored.

Bioinformatics Analysis

Protein identification by Protein Pilot

MS data of each fraction was used to identify proteins by searching a concatenated Swissprot/Panther database of 66082 distinct human protein entries (version June 2, 2010). The database was searched using Protein Pilot software, version 2.0.1 (AB SCIEX, Foster City, USA), which uses the Paragon algorithm (17). Protein identification

was performed with MMTS selected as cysteine modification, with the search option 'emphasis on biological modifications' checked, and with one of missed and/or non-specific cleavages permitted. Peptide and protein summaries and false discovery rate (FDR) reports were generated.

To minimize redundancy in subsequent iterations, a precursor ion exclusion list, generated in–house, was added to the acquisition method after each iteration as described in section 2.1.7 of this thesis. Tolerance windows for exclusion were set at 100 ppm for *m/z* and 360 s for elution time.

iTRAQ ratio re-calculation and identification of dysregulated proteins

To identify non-redundant proteins, data acquired for all 25 fractions from each iTRAQ set injected in triplicate were searched against a database that was created by concatenating the Swissprot human protein database and its reverse (as of June 2, 2010). Only proteins identified with local false discovery rate (FDR) $\leq$ 5% were considered for further analysis (18).

Proteins identified in seven iTRAQ sets were compiled and matched by accession numbers. Redundant proteins and peptides, and proteins identified in reverse sequence were removed from the list. To improve the confidence of protein quantitation, the mean expression iTRAQ ratios of the proteins were re-calculated, using a script written in Matlab (version 7.7.0.471), based on the criteria that the protein must be identified by a minimum of three peptides, with $\geq$95% confidence, and with an expression ratio error factor (EF) <11.1%. Proteins were considered to be dysregulated if iTRAQ ratios were $\geq$1.5 or $\leq$0.67 in $\geq$50% in ccRCC relative to normal samples.

Clustering analysis of ccRCC and normal samples based on dysregulated proteins

To determine if identified dysregulated proteins can discriminate between ccRCC and normal samples, a clustering analysis was performed. Proteins were included in the analysis if quantification was available in at least 50% of the samples. The average iTRAQ ratios were logarithmically transformed, and the city-block distance method was used for the hierarchical clustering of proteins and samples. As a control, the samples were hierarchically clustered based on quantified proteins without dysregulated proteins. Hierarchical clustering analysis was performed using Cluster 3.0 software and the result was visualized using TreeView software (19).

Selection of candidate ccRCC markers

Dysregulated proteins were selected for further verification if their involvement in tumorigenesis was documented (through UniProtKB and PubMed search) and if they could potentially be identified in serum. Such serum-based biomarkers must be secreted or shed into the extracellular space. Dysregulated proteins were considered to be "secreted" if they satisfied at least one of the following four criteria: (1) their subcellular location is extracellular or membrane-bound, according to Ingenuity Pathway Analysis; (2) they are classically secreted, according to SignalP 4.0 analysis; or (3) they are non-classically secreted, according to SecretomeP 2.0 analysis; (4) they are non-classically secreted by the exosome pathway, according to Ingenuity Pathway Analysis.

SignalP predicts the presence and the location of signal peptide cleavage sites in the amino-acid sequences by a combination of artificial neural networks and hidden Markov model algorithms to detect signal peptides from input protein sequences. The D

score indicates superior discrimination performance of secretory and non-secretory proteins (D > 0.450 was considered to be significant for secretory protein).

SecretomeP was used for non-classical and leaderless protein secretion. SecretomeP utilizes a neural network combining six protein features to predict whether a protein sequence undergoes non-classical secretion. A given protein is considered non-classically secreted if it contains a non-classical (non-signal) peptide-triggered protein secretion with NN-score $\geq$ 0.5; only proteins that did not contain a signal peptide as determined by SignalP were legitimate candidates for this analysis.

Western blot analysis

Twenty micrograms of total protein were electrophoretically separated on a 10% SDS-PAGE gel. Proteins were then transferred to a PVDF membrane and probed with rabbit polyclonal antibodies for L-lactate-dehydrogenase A (LDHA) and α-enolase (ENO1), and mouse monoclonal antibodies for 10kDA heat shock protein (HSPE1), neuroblast differentiation-associated protein (AHNAK), heat shock protein β1 (HSPB1), and β-actin (Abcam, Cambridge, USA). β-actin was used as a loading control. Membranes were incubated with primary antibodies overnight at 4°C. Protein expression was visualized after incubation with secondary anti-rabbit or anti-mouse antibodies conjugated to horseradish peroxidase and enhanced chemiluminescence reagent (Amersham Pharmacia Biotech, Piscataway, USA). The luminescence was captured on photographic film and the intensity of the resulting bands was determined by densitometry using ImageJ (http://rsbweb.nih.gov/ij/). Tumor samples were compared to

normal kidney samples using the paired sample two-tailed t-test. Value $p \leq 0.05$ was considered as significant.

Tissue microarray construction and IHC

Appropriate areas from ccRCC and normal kidney tissues were selected and circled from donor blocks by a pathologist. Tissue microarray (TMA) blocks containing duplicate 1.0-mm cores from each specimen were constructed with a manual tissue microarrayer (Beecher Instruments, Sun Prairie, USA). The TMAs contained 85 ccRCC and matched normal kidney tissues from the same patient. In addition, each block contained two marker cores for TMA orientation.

TMA sections were cut 5-μm thick and placed on charged slides. Slides were deparaffinized in xylene, hydrated in gradient ethanol, and pre-treated in a microwave oven for 20 min at 800 W in 1 L of citrate buffer (0.01 M, pH 6.0) for antigen retrieval. Sections were then incubated with hydrogen peroxide (0.3% v/v) in PBS for 15 min to quench the endogenous peroxidase activity, followed by blocking with 10% FBS (fetal bovine serum) in PBS (1X) to preclude non-specific binding. Thereafter, the slides were incubated overnight at 4°C with the desired primary antibody (AHNAK, HSPB1, ENO1, LDHA, and HSPE1) in 1 X PBS. Protein expression was detected using the streptavidin-biotin complex with the Dako LSAB+ kit (Dako Cytomation, Glostrup, Denmark) and diaminobenzidine as the chromogen. All procedures were carried out at room temperature unless otherwise specified. Slides were washed with 0.025% Triton X 100 in PBS (0.1 M, pH = 7.3) three times after each step. Finally, sections were counterstained with Mayer's

hematoxylin and mounted with DPX mountant. In the negative control tissue sections, the primary antibody was replaced by isotype specific non-immune mouse / rabbit IgG.

Immunoexpression was scored by assessing the cytoplasmic, nuclear, and membrane staining (in tumors only) intensity and frequency. Intensity was scored as 0 if there was no expression; 1 if there was weak intensity immunoexpression; 2 for moderate intensity; and 3 for strong immunoexpression. Frequency of immunoexpression was scored as 0 if there was no expression; 1 if 1-25% of the cells showed immunoexpression; 2 for 26-50% cells with expression; 3 if 51-75% cells showed immunoexpression; and 4 if 76-100% cells showed positive immunoexpression. In order to quantitatively determine the direction of dysregulation of ccRCC compared to matched normal kidney tissue, we calculated the sum of the intensity and frequency scores and compared the combined scores from the tumor tissue to normal tissue. Cancer scores that had a combined score that was ±1 from the normal score were not considered significant and were called "no change." Cancer scores that had a combined score of 2 or more above the normal combined score for that case were labeled "increased expression." Cancer scores that had a combined score of 2 or more below the matched normal tissue combined score were labeled "decreased expression."

Results

Identification of dysregulated proteins in ccRCC versus normal tissues

A schematic of the work flow is shown in **Figure 1**. Based on LC-MS analysis a total of 1591 non-redundant proteins with local FDR < 5% were identified; 345 of which were reliably quantified. Fifty five proteins fulfilled our criteria for a dysregulation (see Materials and Methods) between ccRCC and normal kidney: 15 were upregulated (iTRAQ ratios of $\geq$1.5) and 40 were downregulated (iTRAQ ratios of $\leq$0.67). **Table 4** shows a heat map of the 55 dysregulated proteins and **Table 5** shows a full name of proteins and their accession numbers.

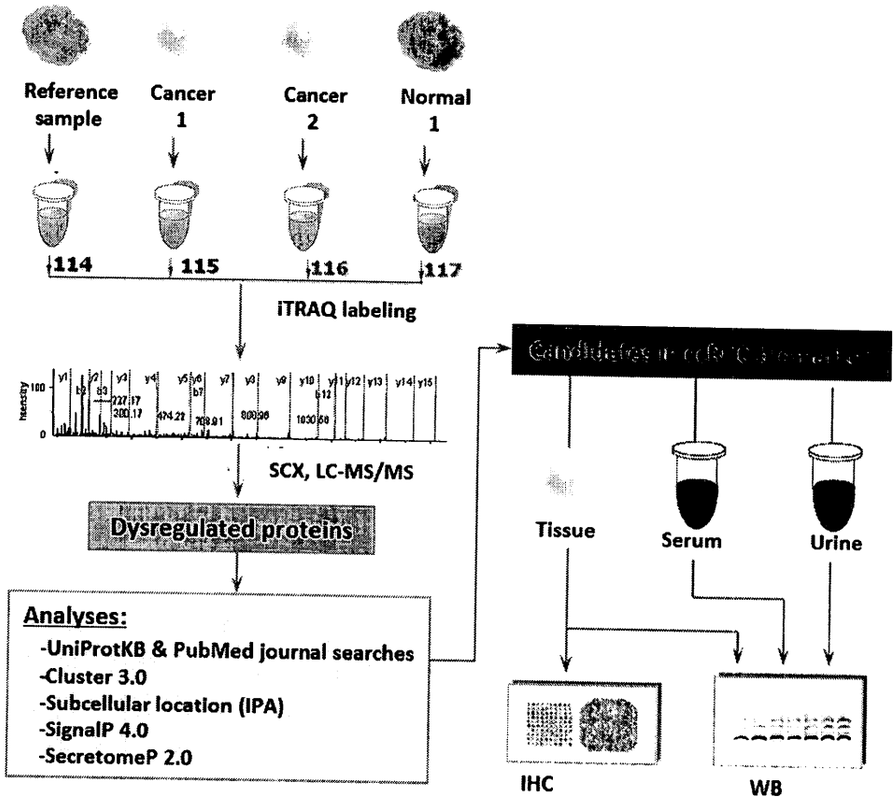
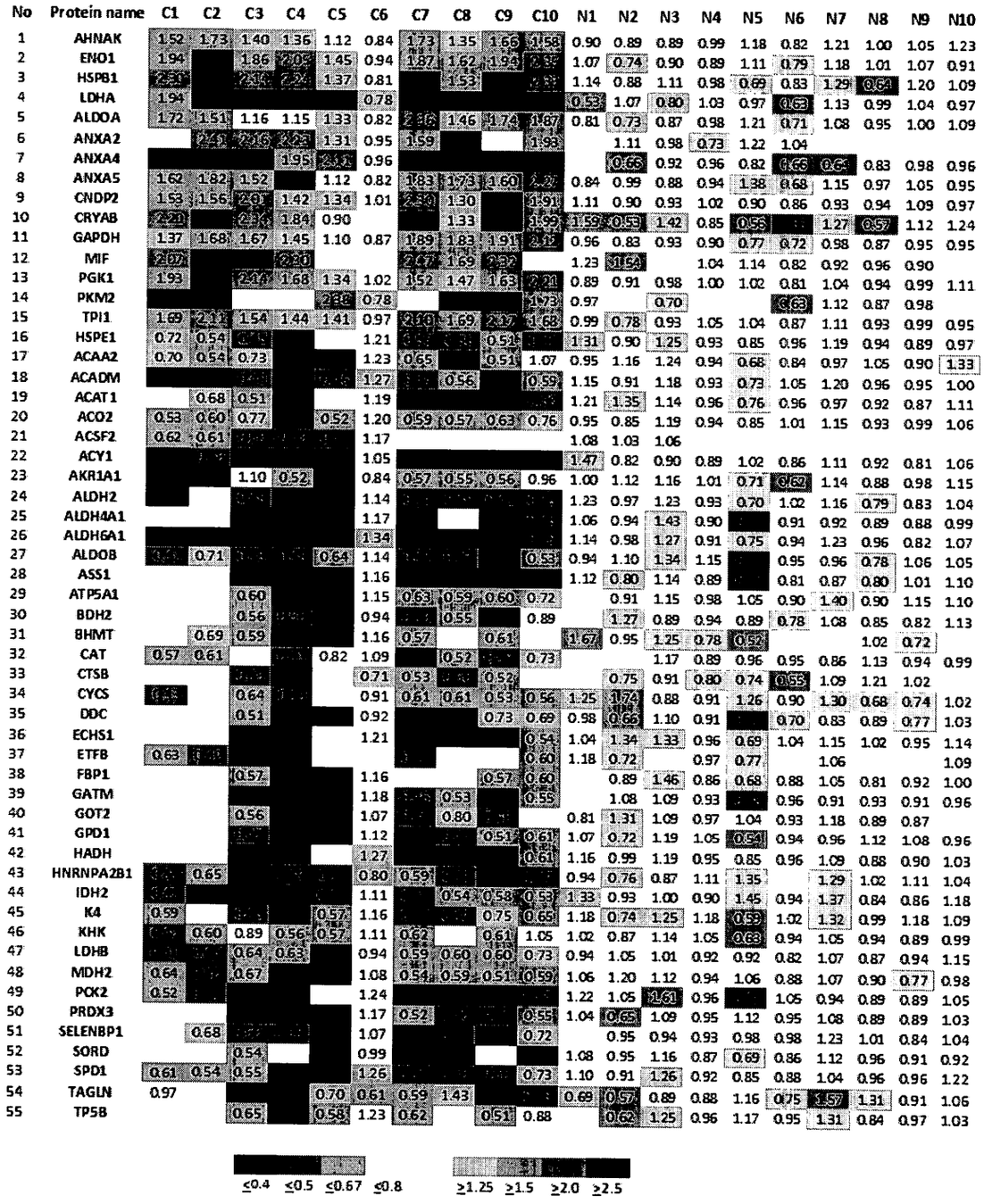


Figure 1. Work flow for quantitative proteomic analysis. *Ten pairs of ccRCC and normal matched kidney tissues samples from the same patient were analyzed. The reference sample consisted of a homogenate of thirty normal kidney tissues. Each sample was digested individually with trypsin and labeled with the appropriate iTRAQ tag. The labeled digests were then pooled and separated by offline SCX LC. Each fraction was analyzed in triplicate by on-line RP nano-LC-MS. Exclusion lists were used to minimize redundancy. MS data were analyzed by Protein Pilot to identify and quantify proteins using a cut-off of 5% local FDR. Dysregulated proteins were further analyzed by clustering, their involvement in tumorigenesis processes, and potential of being shed into serum. Their dysregulation of selected proteins was verified by WB and IHC.*

Table 4. A list of 55 proteins that were dysregulated in ccRCC compared to normal kidney tissue samples as identified by LC-MS analysis.



No	Protein name	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10
1	AHNAK	1.52	1.73	1.40	1.36	1.12	0.84	1.73	1.35	1.66	1.58	0.90	0.89	0.89	0.99	1.18	0.82	1.21	1.00	1.05	1.23
2	ENO1	1.94		1.86	2.05	1.45	0.94	1.87	1.62	1.94	2.37	1.07	0.74	0.90	0.89	1.11	0.79	1.18	1.01	1.07	0.91
3	HSPB1	2.51		2.14	2.24	1.37	0.81		1.53		2.33	1.14	0.88	1.11	0.98	0.69	0.83	1.29	0.64	1.20	1.09
4	LDHA	1.94					0.78					0.53	1.07	0.80	1.03	0.97	0.63	1.13	0.99	1.04	0.97
5	ALDOA	1.72	1.51	1.16	1.15	1.33	0.82	2.36	1.46	1.74	1.87	0.81	0.73	0.87	0.98	1.21	0.71	1.08	0.95	1.00	1.09
6	ANXA2		2.41	2.16	2.23	1.31	0.95	1.69			1.93		1.11	0.98	0.73	1.22	1.04				
7	ANXA4				1.95	2.31	0.96						0.66	0.92	0.96	0.82	0.66	0.64	0.83	0.98	0.96
8	ANXA5	1.62	1.82	1.52		1.12	0.82	1.83	1.73	1.60	2.47	0.84	0.99	0.88	0.94	1.38	0.68	1.15	0.97	1.05	0.95
9	CNDP2	1.53	1.56	2.01	1.42	1.34	1.01	2.30	1.30		1.91	1.11	0.90	0.93	1.02	0.90	0.86	0.93	0.94	1.09	0.97
10	CRYAB	2.20		2.34	1.84	0.90			1.33		1.99	1.59	0.59	1.42	0.85	0.56		1.27	0.57	1.12	1.24
11	GAPDH	1.37	1.68	1.67	1.45	1.10	0.87	1.89	1.83	1.91	2.12	0.96	0.83	0.93	0.90	0.77	0.72	0.98	0.87	0.95	0.95
12	MIF	2.67			2.80			2.67	1.69	2.32		1.23	1.64		1.04	1.14	0.82	0.92	0.96	0.90	
13	PGK1	1.93		2.14	1.68	1.34	1.02	1.52	1.47	1.63	2.21	0.89	0.91	0.98	1.00	1.02	0.81	1.04	0.94	0.99	1.11
14	PKM2					2.88	0.78				1.73	0.97		0.70			0.63	1.12	0.87	0.98	
15	TPI1	1.69	2.11	1.54	1.44	1.41	0.97	2.80	1.69	2.17	1.68	0.99	0.78	0.93	1.05	1.04	0.87	1.11	0.93	0.99	0.95
16	HSPE1	0.72	0.54	0.48			1.21			0.51		1.31	0.90	1.25	0.93	0.85	0.96	1.19	0.94	0.89	0.97
17	ACAA2	0.70	0.54	0.73			1.23	0.65		0.51	1.07	0.95	1.16	1.24	0.94	0.68	0.84	0.97	1.05	0.90	1.33
18	ACADM						1.27		0.56		0.59	1.15	0.91	1.18	0.93	0.73	1.05	1.20	0.96	0.95	1.00
19	ACAT1		0.68	0.51			1.19					1.21	1.35	1.14	0.96	0.76	0.96	0.97	0.92	0.87	1.11
20	ACO2	0.53	0.60	0.77		0.52	1.20	0.59	0.57	0.63	0.76	0.95	0.85	1.19	0.94	0.85	1.01	1.15	0.93	0.99	1.06
21	ACSF2	0.62	0.61				1.17					1.08	1.03	1.06							
22	ACY1						1.05					1.47	0.82	0.90	0.89	1.02	0.86	1.11	0.92	0.81	1.06
23	AKR1A1			1.10	0.52		0.84	0.57	0.55	0.56	0.96	1.00	1.12	1.16	1.01	0.71	0.62	1.14	0.88	0.98	1.15
24	ALDH2						1.14					1.23	0.97	1.23	0.93	0.70	1.02	1.16	0.79	0.83	1.04
25	ALDH4A1						1.17					1.06	0.94	1.43	0.90		0.91	0.92	0.89	0.88	0.99
26	ALDH6A1						1.34					1.14	0.98	1.27	0.91	0.75	0.94	1.23	0.96	0.82	1.07
27	ALDOB		0.71			0.64	1.14				0.53	0.94	1.10	1.34	1.15		0.95	0.96	0.78	1.06	1.05
28	ASS1						1.16					1.12	0.80	1.14	0.89		0.81	0.87	0.80	1.01	1.10
29	ATP5A1			0.60			1.15	0.63	0.59	0.60	0.72		0.91	1.15	0.98	1.05	0.90	1.40	0.90	1.15	1.10
30	BDH2			0.56			0.94		0.55		0.89		1.27	0.89	0.94	0.89	0.78	1.08	0.85	0.82	1.13
31	BHMT		0.69	0.59			1.16	0.57		0.61		1.67	0.95	1.25	0.78	0.52			1.02	0.72	
32	CAT	0.57	0.61			0.82	1.09		0.52		0.73			1.17	0.89	0.96	0.95	0.86	1.13	0.94	0.99
33	CTSB						0.71	0.53		0.52			0.75	0.91	0.80	0.74	0.65	1.09	1.21	1.02	
34	CYCS	0.48		0.64			0.91	0.61	0.61	0.53	0.56	1.25	1.74	0.88	0.91	1.26	0.90	1.30	0.68	0.74	1.02
35	DDC			0.51			0.92			0.73	0.69	0.98	0.66	1.10	0.91		0.70	0.83	0.89	0.77	1.03
36	ECHS1						1.21				0.54	1.04	1.34	1.33	0.96	0.69	1.04	1.15	1.02	0.95	1.14
37	ETFB	0.63	0.48								0.60	1.18	0.72		0.97	0.77		1.06			1.09
38	FBP1			0.57			1.16			0.57	0.60		0.89	1.46	0.86	0.68	0.88	1.05	0.81	0.92	1.00
39	GATM						1.18		0.53		0.55		1.08	1.09	0.93		0.96	0.91	0.93	0.91	0.96
40	GOT2			0.56			1.07		0.80			0.81	1.31	1.09	0.97	1.04	0.93	1.18	0.89	0.87	
41	GPD1						1.12			0.51	0.61	1.07	0.72	1.19	1.05	0.54	0.94	0.96	1.12	1.08	0.96
42	HADH						1.27				0.61	1.16	0.99	1.19	0.95	0.85	0.96	1.09	0.88	0.90	1.03
43	HNRNPA2B1		0.65				0.80	0.59				0.94	0.76	0.87	1.11	1.35		1.29	1.02	1.11	1.04
44	IDH2						1.11		0.54	0.58	0.53	1.33	0.93	1.00	0.90	1.45	0.94	1.37	0.84	0.86	1.18
45	K4	0.59				0.57	1.16			0.75	0.65	1.18	0.74	1.25	1.18	0.59	1.02	1.32	0.99	1.18	1.09
46	KHK		0.60	0.89	0.56	0.57	1.11	0.62		0.61	1.05	1.02	0.87	1.14	1.05	0.68	0.94	1.05	0.94	0.89	0.99
47	LDHB			0.64	0.63		0.94	0.59	0.60	0.60	0.73	0.94	1.05	1.01	0.92	0.92	0.82	1.07	0.87	0.94	1.15
48	MDH2	0.64		0.67			1.08	0.54	0.59	0.51	0.59	1.06	1.20	1.12	0.94	1.06	0.88	1.07	0.90	0.77	0.98
49	PCK2	0.52					1.24					1.22	1.05	1.61	0.96		1.05	0.94	0.89	0.89	1.05
50	PRDX3						1.17	0.52			0.55	1.04	0.65	1.09	0.95	1.12	0.95	1.08	0.89	0.89	1.03
51	SELENBP1		0.68				1.07				0.72		0.95	0.94	0.93	0.98	0.98	1.23	1.01	0.84	1.04
52	SORD			0.54			0.99					1.08	0.95	1.16	0.87	0.69	0.86	1.12	0.96	0.91	0.92
53	SPD1	0.61	0.54	0.55			1.26				0.73	1.10	0.91	1.26	0.92	0.85	0.88	1.04	0.96	0.96	1.22
54	TAGLN	0.97				0.70	0.61	0.59	1.43			0.69	0.57	0.89	0.88	1.16	0.75	1.57	1.31	0.91	1.06
55	TP5B			0.65		0.58	1.23	0.62		0.51	0.88		0.62	1.25	0.96	1.17	0.95	1.31	0.84	0.97	1.03

≤0.4 ≤0.5 ≤0.67 ≤0.8 ≥1.25 ≥1.5 ≥2.0 ≥2.5

'Protein name' shows "gene name" according to UniProtKB. C: ccRCC tissue samples;

N: normal kidney tissue sample.

Table 5. The secretory capability of the 55 dysregulated proteins in ccRCC

No	Accession Number	Protein Name	Gene name	Dysre-gulation	Subcellular location	Exo-somes	SignalP, D	SecretomP, NN
1	sp|Q09666	Neuroblast differentiation-associated protein AHNAK	AHNAK	up	C, PM		0.106	NA
2	sp|P06733	Alpha-enolase	ENO1	up	C, PM	*	0.114	0.536
3	sp|P04792	Heat shock protein beta-1	HSPB1	up	C, PM	*	0.111	0.74
4	sp|P00338	L-lactate dehydrogenase A chain	LDHA	up	C, PM	*	0	0.549
5	sp|P04075	Fructose-bisphosphate aldolase A	ALDOA	up	C	*	0.1	0.356
6	sp|P07355-2	Isoform 2 of Annexin A2	ANXA2	up	E		0.127	0.746
7	sp|P09525	Annexin A4	ANXA4	up	PM	*	0.103	0.439
8	sp|P08758	Annexin A5	ANXA5	up	PM		0.108	0.55
9	sp|Q96KP4	Cytosolic non-specific dipeptidase	CNDP2	up	C		0.111	0.443
10	sp|P02511	Alpha-crystallin B chain	CRYAB	up	N, PM		0	0.864
11	sp|P04406	Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	up	C	*	0.154	0.467
12	sp|P14174	Macrophage migration inhibitory factor	MIF	up	E		0.112	0.776
13	sp|P00558	Phosphoglycerate kinase 1	PGK1	up	C		0.099	0.389
14	sp|P14618	Pyruvate kinase isozymes M1/M2	PKM2	up	C	*	0.148	0.42
15	sp|P60174	Triosephosphate isomerase	TPI1	up	C	*	0.143	0.51
16	sp|P61604	10 kDa heat shock protein, mitochondrial	HSPE1	down	M		0.213	0.57
17	sp|P42765	3-ketoacyl-CoA thiolase, mitochondrial	ACAA2	down	M		0.115	0.382
18	sp|P11310	Medium-chain specific acyl-CoA dehydrogenase	ACADM	down	MM		0.178	0.657
19	sp|P24752	Acetyl-CoA acetyltransferase, mitochondrial	ACAT1	down	M		0.188	0.593
20	sp|Q99798	Aconitate hydratase, mitochondrial	ACO2	down	M		0.165	0.413
21	sp|Q96CM8	Acyl-CoA synthetase family member 2, mitochondrial	ACSF2	down	M		0.302	0.547
22	sp|Q03154	Aminoacylase-1	ACY1	down	C		0.128	0.369
23	sp|P14550	Alcohol dehydrogenase [NADP+]	AKR1A1	down	Cytosol		0.163	0.49
24	sp|P05091	Aldehyde dehydrogenase, mitochondrial	ALDH2	down	MM		0.381	0.629
25	sp|P30038	Delta-1-pyrroline-5-carboxylate dehydrogenase	ALDH4A1	down	MM		0.27	0.584
26	sp|Q02252	Methylmalonate-semialdehyde dehydrogenase	ALDH6A1	down	M		0.108	0.789
27	sp|P05062	Fructose-bisphosphate aldolase B	ALDOB	down	C		0.126	0.342
28	sp|P00966	Argininosuccinate synthase	ASS1	down	Cytosol		0.105	0.571

Table 5. Continued

No	Accession Number	Protein Name	Gene name	Dysre-gulation	Subcellular location	Exo-somes	SignalP, D	SecretomP, NN
29	sp|P25705	ATP synthase subunit alpha, mitochondrial	ATP5A1	down	M, PM		0.298	0.574
30	sp|Q9BUT1	3-hydroxybutyrate dehydrogenase type 2	BDH2	down	C		0.264	0.818
31	sp|Q93088	Betaine--homocysteine S-methyltransferase 1	BHMT	down	C		0.114	0.346
32	sp|P04040	Catalase	CAT	down	Peroxisome		0.108	0.406
33	sp|P07858	Cathepsin B	CTSB	down	E, Lysosome		0.841	
34	sp|P99999	Cytochrome c	CYCS	down	MM		0.124	0.465
35	sp|P20711	Aromatic-L-amino-acid decarboxylase	DDC	down	Cytosol		0.102	0.646
36	sp|P30084	Enoyl-CoA hydratase, mitochondrial	ECHS1	down	M		0.284	0.855
37	sp|P38117	Electron transfer flavoprotein subunit beta	ETFB	down	MM		0.126	0.362
38	sp|P09467	Fructose-1,6-bisphosphatase 1	FBP1	down	Cytosol		0.115	0.515
39	sp|P50440	Glycine amidinotransferase, mitochondrial	GATM	down	C		0.141	0.699
40	sp|P00505	Aspartate aminotransferase, mitochondrial	GOT2	down	PM, M		0.317	0.545
41	sp|P21695	Glycerol-3-phosphate dehydrogenase [NAD+]	GPD1	down	C		0.162	0.533
42	sp|Q16836-2	Isoform 2 of Hydroxyacyl-coenzyme A dehydrogenase	HADH	down	MM		0.149	0.492
43	sp|P22626	Heterogeneous nuclear ribonucleoproteins A2/B1	HNRNPA2B1	down	N		0.104	0.081
44	sp|P48735	Isocitrate dehydrogenase [NADP], mitochondrial	IDH2	down	M		0.147	0.593
45	sp|P27144	Adenylate kinase isoenzyme 4, mitochondrial	K4	down	MM		0.208	0.738
46	sp|P50053-2	Isoform Peripheral of Ketohexokinase	KHK	down	Cytosol		0.101	0.445
47	sp|P07195	L-lactate dehydrogenase B chain	LDHB	down	C		0.137	0.569
48	sp|P40926	Malate dehydrogenase, mitochondrial	MDH2	down	PM, MM		0.174	0.644
49	sp|Q16822	Phosphoenolpyruvate carboxykinase [GTP]	PCK2	down	M		0.243	0.489
50	sp|P30048	Thioredoxin-dependent peroxide reductase	PRDX3	down	M		0.291	0.756
51	sp|Q13228	Selenium-binding protein 1	SELENBP1	down	N, C		0.115	0.496
52	sp|Q00796	Sorbitol dehydrogenase	SORD	down	E, M		0.119	0.515
53	sp|P10809	60 kDa heat shock protein, mitochondrial	SPD1	down	E, M		0.151	0.282
54	sp|Q01995	Transgelin	TAGLN	down	C		0.102	0.787
55	sp|P06576	ATP synthase subunit beta, mitochondrial	TP5B	down	PM, M		0.21	0.591

C: cytoplasm; E: extracellular; M: mitochondrial; MM: mitochondrion matrix; N: nucleus; PM: plasma membrane.

Clustering analysis indicated that differential protein expressions can discriminate between ccRCC and normal kidney tissue samples. Clustering analysis was performed on 345 proteins for which quantitative information was available. The samples clustered into two main groups: one that contained the 9 of 10 ccRCC samples (C1-C5, C7-C10); and a second that contained all normal samples (N1-N10) and one cancer sample (C6) (**Figure 2A**). To confirm the results of our clustering analysis, we performed control clustering analysis based on the 345 quantified minus the 55 dysregulated proteins. As expected, the cancer and normal samples failed to cluster into the two groups (**Figure 2B**).

Elucidating RCC pathogenesis through quantitative proteomics

We performed UniProtKB and literature searches on the 55 dysregulated proteins to investigate their biological significance in carcinogenesis. All proteins, except AHNAK, were found to be involved in at least one of the tumorigenesis-related processes (**Table 6**).

Figure 2A

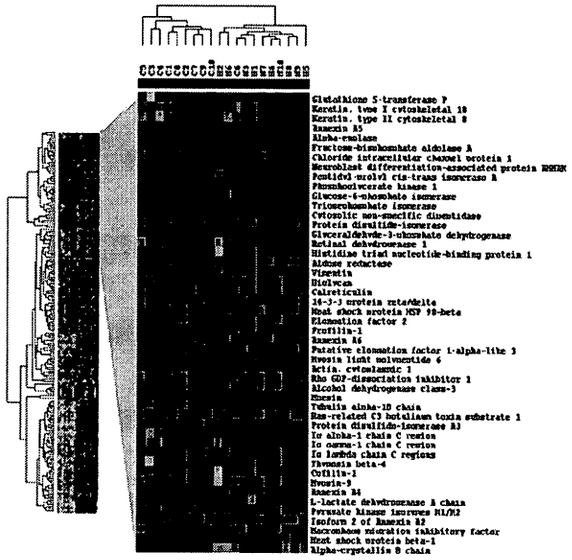


Figure 2B

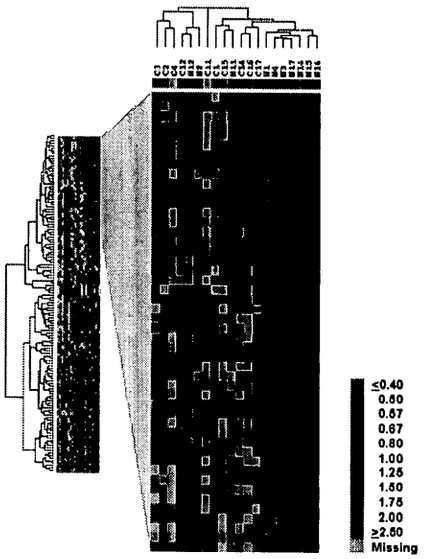


Figure 2. Hierarchical clustering analysis of ccRCC and normal kidney samples based on dysregulated proteins.

A: *To determine if differential protein expression can discriminate between ccRCC and normal kidney samples clustering analysis of these samples was performed based on 345 proteins for which quantitative information was available. The samples clustered into two main groups: one that contained the 9 of 10 ccRCC samples (C1-C5, C7-C10); and second contained all normal samples (N1-N10) plus cancer sample C6. The difference in expressions of dysregulated proteins between ccRCC and normal kidney samples was statistically significant ($p<0.001$).* ***B:*** *To verify that clustering of samples (in **A**) was not accidental we performed control clustering analysis based on the 345 quantified minus 55 dysregulated proteins. Primary and normal samples were mixed together. The difference in expressions of non-dysregulated proteins between ccRCC and normal kidney samples in this case was statistically insignificant ($p>0.5$) which means that expressions are similar. That conforms that clustering based on quantified proteins including dysregulated (in **A**) was not accidental. Clustering analysis suggests that dysregulated proteins can discriminate between cancer and normal samples.*

Table 6. Involvement of dysregulated proteins in tumorigenesis-related processes.

Tumorigenesis-related processes	**Protein name**
Carbohydrate and lipid metabolism	ACAA2, ACADM, ACAT1, ACO2, ACSF2, ACY1, AKR1A1, ALDH2, ALDH4A1, ALDH6A1, ALDOA, ALDOB, ASS1, ATP5A1, BDH2, BHMT, CAT, CNDP2, DDC, ECHS1, ENO1, ETFB, FBP1, GAPDH, GATM, GOT2, GPD1, HADH, HSPB1, IDH2, K4, KHK, LDHA, LDHB, MDH2, PCK2, PGK1, PKM2, SORD, TP5B, TPI1
Apoptosis	ACAA2, ANXA2, ANXA4, ANXA5, CRYAB, CTSB, CYCS, ENO1, GAPDH, HSPB1, HSPE1, LDHA, MIF, SELENBP1, SPD1
Growth and proliferation	CAT, ENO1, FBP1, HNRNPA2B1, HSPB1, HSPE1, LDHA, MIF, PRDX3, SELENBP1, SPD1
Cell cycle	ENO1, HSPB1, HSPE1, MIF
Hypoxia	ENO1, HSPB1, LDHA

For 'Protein name' was used gene name according to UniProtKB.

Interestingly, 41 of the dysregulated proteins were found to be involved in carbohydrate and lipid metabolism, in agreement with recent reports (20-22) (**Table 6**). Twenty eight proteins are involved in glycolysis, citric cycle, and acetyl-CoA metabolism, as shown in **Figure 3**. Fructose-bisphosphate aldolases A and B (ALDOA, ALDOB), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), ENO1, pyruvate kinase isozymes M1/M2 (PKM2) catalyze the reactions of glycolysis; and aconitate hydratase (ACO2), malate dehydrogenase (MDH2) and others catalyze citric acid cycle reactions. Also, LDHA and LDHB catalyze the conversion of pyruvate to lactate.

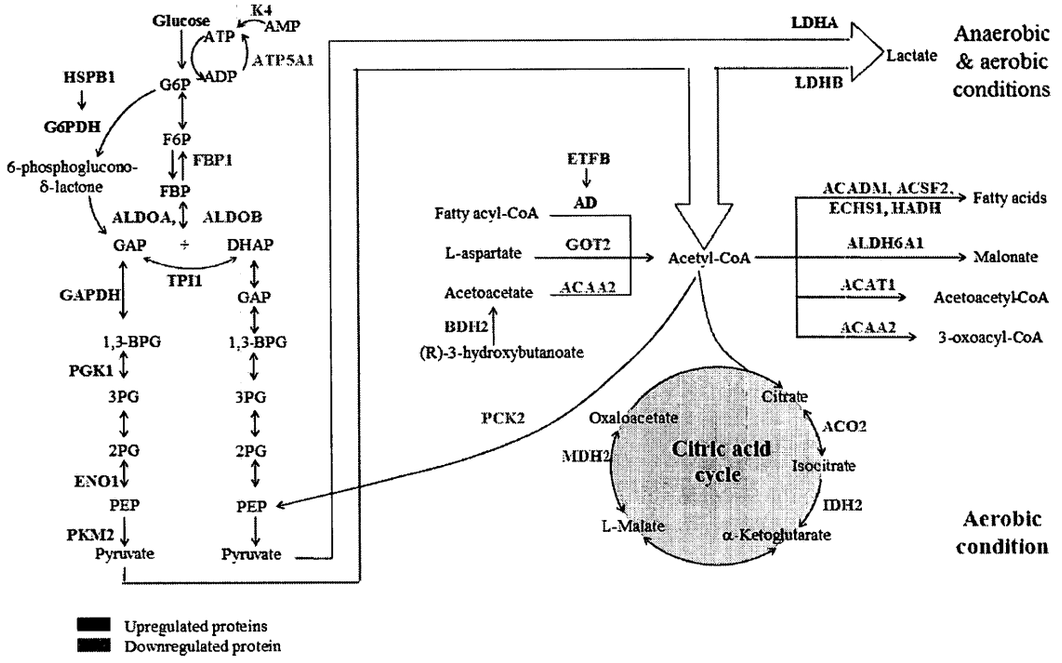


Figure 3. The involvement of a subgroup of dysregulated proteins in glycolysis, citric cycle, metabolism and catabolism of Acetyl-CoA.

A: *Most of the upregulated proteins are enzymes which catalyze the reactions of glycolysis, citric acid cycle, metabolism and catabolism of Acetyl-CoA. ETFB serves as a specific electron acceptor for several dehydrogenases, including five acytyl-CoA dehydrogenases (AD), glutaryl-CoA and sarcosine dehydrogenase; and HSPB1 forms a complex with G6PDH that increased its activity. In red lettering are shown upregulated proteins, in green dowdregulated proteins, in black documented metabolic reactants, products, and enzymes. Lines with arrowheads represent documented metabolic reactions.*

B: *Visualization of protein-protein interactions for dysregulated proteins in ccRCC using STRING analysis. Dysregulated proteins were used as input for STRING and are represented as spheres of distinct colors. Blue lines represent interactions between proteins and the thickness of the lines display the level of confidence associated with each interaction.*

Other metabolic processes were found to be affected as well. For instance, electron transfer flavoprotein subunit beta (ETFB) serves as a specific electron acceptor for several dehydrogenases, including five acyl-CoA dehydrogenases (AD), glutaryl-CoA and sarcosine dehydrogenase (UniProtKB), and HSPB1 forms a complex with glucose 6-phosphate dehydrogenase (G6PDH), increasing its activity, augmenting NADP(+) to NADPH reduction, and stimulating nucleotide synthesis (23).

Fifteen proteins are involved in apoptosis (**Table 6**). For example, alpha-crystallin B chain, annexin A4, and macrophage migration inhibitory factor (MIF) negatively regulate the apoptotic process; ENO1 induces ganglion cell death through an apoptotic process (24); LDHA mediates the unique apoptotic effect of c-Myc when glycolysis is blocked in lung carcinoma cells (25); gain- and loss-of-function studies indicate that HSPB1 mediates mitochondrial apoptosis in hepatocellular carcinoma cells (26); and HSPE1 regulates apoptosis of mouse ovarian granulosa cells (27).

Eleven proteins are involved in growth and proliferation (**Table 6**). For instance, catalase promotes growth of cells including T-cells, B-cells, myeloid leukemia cells, and melanoma cells (UniProtKB); LDHA directs pyruvic acid into the Krebs cycle rather than into glycolysis in AGS gastric cancer cells, which inhibits cell growth (28); silencing of ENO1 results in growth inhibition of gastric cancer cells (29); upregulation of HSPB1 in nonangiogenic cells results in expansive growth of breast cancer, xenograft model, MDA-MB-436 cells (30); and HSPE1 increases both proliferation and death in mouse P19 teratocarcinoma cells (31).

Four proteins are involved in the cell cycle (**Table 6**): MIF negatively regulates cell cycle arrest (UniProtKB); silencing of ENO1 results in cell cycle arrest of gastric

cancer cells (29); silencing of HSPB1 enhances G(1)-S arrest and cell death of human adenocarcinoma, MCF-7 cells (32); and HSPE1 interacts with guanine triphosphate–binding protein of the Ras superfamily involved in termination of M-phase (33).

Three proteins are involved in, and respond to, hypoxia (**Table 6**): ENO1 is upregulated in breast cancer cells MCF-7 during hypoxia (34); LDHA is transcriptionally controlled in human HepG2 hepatoma cells by the hypoxia inducible factors HIF1α and HIF2α (35); and tumor-induced angiogenesis in vascular multicellular prostate tumor spheroids results in elevated levels of HSPB1 (36).

Identification of dysregulated proteins that can serve as serum biomarkers

The aim herein was to identify which dysregulated proteins could potentially be secreted and, hence, serve as serum-based ccRCC biomarkers. Of the 55 dysregulated proteins, 39 (70.9%) satisfied one or more of the four criteria for being considered a "secretory" protein (see Materials and Methods for details) (**Table 5**): 5 proteins (9.1%) were found to be extracellular; 10 (18.2%) were membrane-bound according to subcellular location; 8 (14.5%) could be released from cells via the exosome pathway, according to Ingenuity Pathway Analysis; 1 (1.8%) could be classically secreted, according to SignalP analysis; and 32 (58.2%) proteins were likely to be non-classically secreted, according to SecretomeP analysis.

Interestingly, at least 5 of the dysregulated proteins, namely ENO1, LDHA, HSPB1, AHNAK and HSPE1, have been reported to be upregulated in the serum of cancer patients compared to healthy controls: serum ENO1 was elevated in patients with small-cell lung carcinoma (37); high serum LDH was linked to significantly poor survival

rates in colorectal cancer patients (38); high levels of HSPB1 were present in the serum of patients with breast cancer (39); serum AHNAK was elevated in patients with ovarian cancer (40), and HSPE1 was upregulated in the sera of ovarian cancer patients (41). Thus, it is indeed possible to detect the upregulation of these proteins in sera and/or urine of the ccRCC patients, which makes them especially promising as biofluid-based biomarkers for ccRCC.

Verification of the differential expression of potential biomarkers by WB analysis

The differential expressions of the five promising biofluid-based biomarker candidates were verified by WB analysis in samples of the discovery cohort (**Figure 4A**). As in the LC-MS results, the average expressions of AHNAK, ENO1, and HSPB1 were found to be significantly elevated in ccRCC compared to the matched normal tissues (1.68-fold, $p < 0.002$; 1.62-fold, $p < 0.01$; and 1.47-fold, $p < 0.01$, respectively). LDHA was also elevated in ccRCC, although less significantly (1.12-fold, $p < 0.4$) (**Figure 4B**). HSPE1 was significantly downregulated in ccRCC (0.47-fold, $p < 0.002$), which is also in agreement with our LC-MS findings.

We also assessed the consistency of the WB and the LC-MS results with respect to protein dysregulation, using the same set of samples. For ENO1, HSPB1, and HSPE1, 9 results out of 10 were consistent; for AHNAK, 8 out of 10 were consistent; and for LDHA, 6 out of 10 were consistent. Thus, overall, the WB results agree with the LC-MS results.

Figure 4A

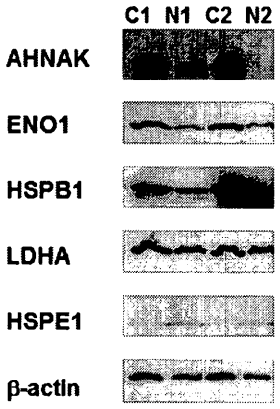


Figure 4B

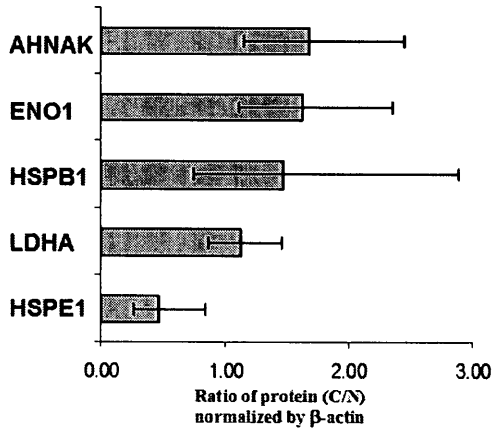


Figure 4. Verification of ENO1, LDHA, HSPB1, and HSPE1 dysregulation in ccRCC by Western blot analysis.

***A**: Representative blots showing the expression of proteins in normal kidney tissues (N1, N2) and cancerous ccRCC (C1, C2). For AHNAK, ENO1, HSPB1, LDHA, expression was increased and for HSPE1 was decreased in cancer tissues when compared to normal kidney tissue. b-actin was used as a loading control. **B**: Graphical representation of the average fold change in expression of the proteins between ten ccRCCs and matched normal specimens (C/N). Expressions of protein in normal samples were normalized (average =1).*

Verification of the differential expression of potential biomarkers by IHC analysis

The dysregulation of ENO1, HSPB1, HSPE1, LDHA, and AHNAK were also verified by IHC in an independent cohort of 85 patients, using TMAs consisting of paired ccRCC and normal kidney tissue obtained from the same patient as a reference **(Figure 5)**.

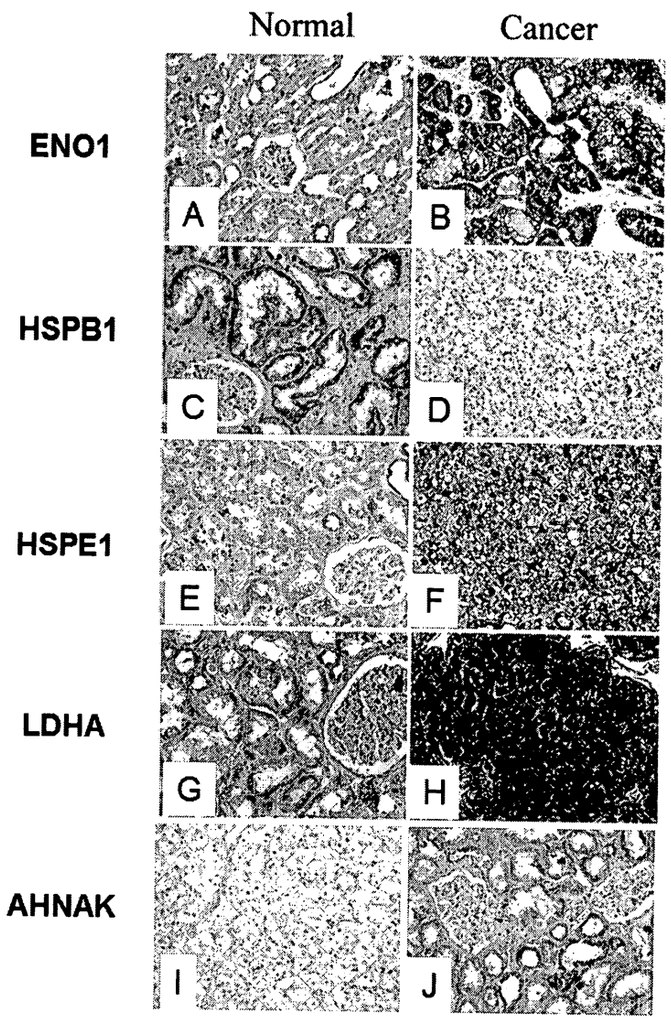


Figure 5. Verification of ENO1, HSPB1, HSPE1, LDHA, and AHNAK dysregulation in ccRCC by immunohistochemical analysis. A-J: *Representative photomicrographs showing differential expression of ENO1, HSPB1, HSPE1 and LDHA in ccRCC compared to normal kidney tissue by IHC. Original magnification × 200.*

The immunoexpression of ENO1 was elevated in 56 (70%) ccRCC tissues, which is in agreement with both the LC-MS and WB results. Eight tissues (10%) showed no change in expression and 16 (20%) showed decreased expression. Five cases were omitted from the analysis because the samples washed off the slide.

For HSPB1, 53 (69%) ccRCC tissues showed elevated expression, again, correlating well with our LC-MS and WB data. Thirteen (17%) showed decreased

expression and 11 (14%) showed no change in expression. Eight cases were omitted from the analysis because the samples washed off the slide.

The downregulation of HSPE1 that was observed using LC-MS and WB was also seen in 71 (92%) ccRCC tissues by IHC. There were only 6 (8%) cases that showed no change in expression and none showed increased expression. Eight cases were omitted from the analysis because the samples washed off the slide.

The immunoexpression of LDHA was elevated in 76 (96%) ccRCC tissues, while 2 (3%) showed no change in expression and only one (1%) showed decreased expression. This result correlates with both LC-MS and WB data. Six cases were omitted from the analysis because the samples washed off the slide.

Surprisingly, the upregulation of AHNAK that was observed using LC-MS and WB was not seen in any of the ccRCC tissues. One case was omitted from the analysis because the tissue washed off the slide.

Discussion

Early diagnosis of RCC is associated with a favorable prognosis. Unfortunately, a significant proportion of RCC patients are diagnosed at the metastatic stage, when the prospects for cure are dismal (5-year survival rate ~9%). Traditional method for detection of asymptomatic RCC, which includes medical history, physical examination, and urine analysis, is ineffective (42). More accurate methods such as non-invasive serum- or urine-based biomarkers are urgently required.

Using LC-MS based proteomic methods, 55 proteins were identified that are significantly dysregulated in ccRCC relative to normal kidney tissues (**Table 4**). Hierarchical cluster analysis showed that these proteins can clearly distinguish between ccRCC and normal kidney tissues (**Figure 2A**). Of the 55 dysregulated proteins, 54 (98%) had been reported to be involved in at least one tumorigenic process, and 39 (70.9%) were classified as "secretory" proteins according to our criteria; thus, they can potentially serve as biofluid-based biomarkers. Interestingly, five of the dysregulated "secretory" proteins, namely ENO1, LDHA, HSPB1, AHNAK and HSPE1, have also been reported to be upregulated in the serum of patients having various malignancies (37-41), and this supports the likelihood that these proteins (HSPE1 may be exception) will also be secreted at quantifiable levels in ccRCC, making them even more promising as candidates of biofluid-based biomarkers for primary ccRCC.

The dysregulation of these proteins was verified on two independent sets of tissues by WB and IHC. Both analyses confirmed the upregulation of ENO1, HSPB1, LDHA, and the downregulation of HSPE1. Unexpectedly, the upregulation of AHNAK was confirmed by WB but not by IHC, as will be discussed later. The verification of the

expression of these proteins in the sera and urine of ccRCC patients is currently underway.

The hierarchical cluster analysis divided the samples into two main groups: one that contained 9 out of 10 ccRCC samples and another that contained all of the normal samples plus one cancer sample (C6) (**Figure 2A**). The difference between cancer sample, C6, and the rest of the cancer samples may be due to the heterogeneity of ccRCC; the sample may represent a subpopulation of ccRCC that has a different set of biomarker proteins. For instance, this sample exhibited upregulation of disulfide-isomerase A3 relative to its matched normal kidney tissue sample, whereas the others did not, except for C1.

According to UniProtKB and PubMed searches, 40 of the 55 dysregulated proteins are involved in "metabolic" processes (**Figure 3** and **Table 6**). Metabolism is adapted by cancer cells to meet the requirements of rapid cell proliferation, growth, negative regulation of apoptosis, survival under hypoxia etc. (43,44). Our findings are similar to earlier reports on functional analyses in RCC (45;46). The link between carbohydrate metabolism and RCC is not surprising. In contrast to normal proliferating cells, tumor cells have to survive in environments with varying oxygen and nutrient supplies (47). The increase in lactate dehydrogenase and the activation of the pyruvate kinase pathway indicate active anaerobic glycolysis which is a reflection of the hypoxic conditions known to be an integral component of the pathogenesis of RCC (48;49). Also, the "clear cell" morphology of RCC is known to result from the accumulation of glycogen as a result of disturbed carbohydrate metabolism.

Although the upregulation of LDHA was statistically insignificant by WB analysis (1.12-fold, $p < 0.4$), it was quite significant by LC-MS analysis (2.15-fold, $p < 0.05$). A reason for the difference could be that in the LC-MS analysis, two members of the LDH family were separately quantified; LDHA was upregulated and LDHB was downregulated. For WB analysis, we used polyclonal antibody, which may not specific to LDHA and can also bind LDHB, since the two proteins share 76% of the same sequence. Furthermore, they have similar molecular weights of about 36.6 kDA, causing them to migrate at similar rates using SDS-PAGE. Thus, the two proteins are very likely to be detected together by WB, resulting in the apparently suppressed upregulation of LDHA.

The upregulation of AHNAK, found by LC-MS, was confirmed by WB, but this was contrary to the findings by IHC. An explanation for the reduced immunoexpression shown by IHC may be that in cancer cells, AHNAK undergoes post-translational modification (PTM) at or near the proteins' epitopes, preventing the antibody from binding during IHC analysis, and, as a result, the protein appears to be at lower concentration in the cancer tissues than in normal tissues. This PTM did not affect the WB analysis because it may cleave off during the analysis.

In our study, AHNAK was upregulated in ccRCC tissues, but was reportedly downregulated in the permeable angiogenic endothelial cells of brain tumors (glioma) (50). This suggests that AHNAK may be involved in different tumorigenic mechanisms in different tissue types. Similarly, we found HSPE1 to be downregulated in ccRCC tissues, while others found it to be upregulated in endometrial carcinoma (13); again, this suggests that HSPE1 may play different roles in different tissues.

Through quantitative proteomic analysis, we have identified differential protein expressions that can distinguish between ccRCC and normal kidney tissues. Most of these proteins are involved in biological pathways pertinent to tumor progression, and about 70% can be potentially shed into serum. If the upregulation of the most promising biomarker candidates are confirmed in serum and/or urine of ccRCC patients, this would support further clinical studies of these candidates as serological biomarkers for the early diagnosis of ccRCC, which may lead to greatly improved patient management and increased overall survival.

Acknowledgements

This work was supported by grants from the Canadian Cancer Society (CCS grant # 20185), Ministry of Research and Innovation of the Government of Ontario, Kidney Foundation of Canada, Cancer Research Society and Canadian Institutes of Health Research.

Reference List

1. Lipworth L, Tarone RE, McLaughlin JK. The epidemiology of renal cell carcinoma. J Urol 2006;176:2353-8.

2. Weiss RH, Lin PY. Kidney cancer: identification of novel targets for therapy. Kidney Int. 2006 Jan;69(2):224-32. Review.

3. Feldstein MS, Rhodes DJ, Parker AS, Orford RR, Castle EP. The haphazard approach to the early detection of asymptomatic renal cancer: results from a contemporary executive health programme. BJU Int 2009;104:53-6.

4. Beroukhim R, Brunet JP, Di NA, Mertz KD, Seeley A, Pires MM et al. Patterns of gene expression and copy-number alterations in von-hippel lindau disease-associated and sporadic clear cell carcinoma of the kidney. Cancer Res 2009;69:4674-81.

5. Girgis, A. H., Iakovlev, V. V., Beheshti, B., Bayani, J., Squire, J. A., Bui, A., Mankaruos, M., Youssef, Y., Khalil, B., Khella, H. W., Pasic, M. D., and Yousef, G. M. Genomic Characterization of Clear Cell Renal Cell Carcinoma Reveals Harmonized Multi-Level Molecular Changes Influencing Expression of Candidate Biomarkers. Cancer Res. 2012. Ref Type: In Press

6. White NM, Yousef GM. Translating molecular signatures of renal cell carcinoma into clinical practice. J Urol 2011;186:9-11.

7. Metias SM, Lianidou E, Yousef GM. MicroRNAs in clinical oncology: at the crossroads between promises and problems. J Clin Pathol 2009;62:771-6.

8. Adam BL, Qu Y, Davis JW, Ward MD, Clements MA, Cazares LH, Semmes OJ, Schellhammer PF, Yasui Y, Feng Z, Wright GW. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. Cancer Res 2002; 62, 3609–3614.

9. Baggerly KA, Morris JS, Wang J, Gold D, Xiao LC, Coombes KR. A comprehensive approach to the analysis of MALDI-TOF proteomics spectra from serum samples. Proteomics 2003; 3, 1667–1682.

10. Baggerly KA, Morris JS, Coombes KR. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. Bioinformatics 2004;20:777-85.

11. Pavlovich CP, Schmidt LS. Searching for the hereditary causes of renal-cell carcinoma. Nat Rev Cancer 2004;4:381-93.

12. Sarto C, Marocchi A, Sanchez JC, Giannone D, Frutiger S, Golaz O et al. Renal cell carcinoma and normal kidney protein expression. Electrophoresis 1997;18:599-604.

13. DeSouza LV, Grigull J, Ghanny S, Dube V, Romaschin AD, Colgan TJ, Siu KW. Endometrial carcinoma biomarker discovery and verification using differentially tagged clinical samples with multidimensional liquid chromatography and tandem mass spectrometry. Mol Cell Proteomics 2007;6:1170-82.

14. DeSouza L, Diehl G, Rodrigues MJ, Guo J, Romaschin AD, Colgan TJ, Siu KW. Search for cancer markers from endometrial tissues using differentially labeled tags iTRAQ and cICAT with multidimensional liquid chromatography and tandem mass spectrometry. J Proteome Res 2005;4:377-86.

15. Ralhan R, DeSouza LV, Matta A, Chandra TS, Ghanny S, Datta GS et al. Discovery and verification of head-and-neck cancer biomarkers by differential protein expression analysis using iTRAQ labeling, multidimensional liquid chromatography, and tandem mass spectrometry. Mol Cell Proteomics 2008;7:1162-73.

16. Ralhan R, Masui O, DeSouza LV, Matta A, Macha M, Siu KW. Identification of proteins secreted by head and neck cancer cell lines using LC-MS/MS: Strategy for discovery of candidate serological biomarkers. Proteomics 2011;11:2363-76.

17. Shilov IV, Seymour SL, Patel AA, Loboda A, Tang WH, Keating SP et al. The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. Mol Cell Proteomics 2007;6:1638-55.

18. Tang WH, Shilov IV, Seymour SL. Nonlinear fitting method for determining local false discovery rates from decoy database searches. J Proteome Res 2008;7:3661-7.

19. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 1998;95:14863-8.

20. Romaschin AD, Youssef Y, Chow TF, Siu KW, DeSouza LV, Honey RJ et al. Exploring the pathogenesis of renal cell carcinoma: pathway and bioinformatics analysis of dysregulated genes and proteins. Biol Chem 2009;390:125-35.

21. Siu KW, DeSouza LV, Scorilas A, Romaschin AD, Honey RJ, Stewart R et al. Differential protein expressions in renal cell carcinoma: new biomarker discovery by mass spectrometry. J Proteome Res 2009;8:3797-807.

22. Unwin RD, Craven RA, Harnden P, Hanrahan S, Totty N, Knowles M et al. Proteomic changes in renal cancer and co-ordinate demonstration of both the glycolytic and mitochondrial aspects of the Warburg effect. Proteomics 2003;3:1620-32.

23. Krüger A, Ostadal P, Neuzil P. Mini-invasive mechanical cardiac support. Cas Lek Cesk. 2011;150(12):647-50. Review. Czech.

24. Ren G, Adamus G. Cellular targets of anti-alpha-enolase autoantibodies of patients with autoimmune retinopathy. J Autoimmun 2004;23:161-7.

25. Shim H, Chun YS, Lewis BC, Dang CV. A unique glucose-dependent apoptotic pathway induced by c-Myc1. Proc Natl Acad Sci U S A 1998;95:1511-6.

26. Fu WM, Zhang JF, Wang H, Xi ZC, Wang WM, Zhuang P et al. Heat shock protein 27 mediates the effect of 1,3,5-trihydroxy-13,13-dimethyl-2H-pyran [7,6-b] xanthone on mitochondrial apoptosis in hepatocellular carcinoma. J Proteomics 2012;75:4833-43.

27. Ling J, Zhao K, Cui YG, Li Y, Wang X, Li M et al. Heat shock protein 10 regulated apoptosis of mouse ovarian granulosa cells. Gynecol Endocrinol 2011;27:63-71.

28. Cai Z, Zhao JS, Li JJ, Peng DN, Wang XY, Chen TL et al. A combined proteomics and metabolomics profiling of gastric cardia cancer reveals characteristic dysregulations in glucose metabolism. Mol Cell Proteomics 2010;9:2617-28.

29. Yan GR, Xu SH, Tan ZL, Yin XF, He QY. Proteomics characterization of gastrokine 1-induced growth inhibition of gastric cancer cells. Proteomics 2011;11:3657-64.

30. Straume O, Shimamura T, Lampa MJ, Carretero J, Oyan AM, Jia D et al. Suppression of heat shock protein 27 induces long-term dormancy in human breast cancer. Proc Natl Acad Sci U S A 2012;109:8699-704.

31. Galli G, Ghezzi P, Mascagni P, Marcucci F, Fratelli M. Mycobacterium tuberculosis heat shock protein 10 increases both proliferation and death in mouse P19 teratocarcinoma cells. In Vitro Cell Dev Biol Anim 1996;32:446-50.

32. Kanagasabai R, Karthikeyan K, Vedam K, Qien W, Zhu Q, Ilangovan G. Hsp27 protects adenocarcinoma cells from UV-induced apoptosis by Akt and p21-dependent pathways of survival. Mol Cancer Res 2010;8:1399-412.

33. Czarnecka AM, Campanella C, Zummo G, Cappello F. Mitochondrial chaperones in cancer: from molecular biology to clinical diagnostics. Cancer Biol Ther 2006;5:714-20.

34. Sedoris KC, Thomas SD, Miller DM. Hypoxia induces differential translation of enolase/MBP-1. BMC Cancer 2010; 10:157.

35. Semenza GL, Jiang BH, Leung SW, Passantino R, Concordet JP, Maire P, Giallongo A. Hypoxia response elements in the aldolase A, enolase 1, and lactate dehydrogenase A gene promoters contain essential binding sites for hypoxia-inducible factor 1. 1. J Biol Chem 1996;271:32529-37.

36. Wartenberg M, Donmez F, Ling FC, Acker H, Hescheler J, Sauer H. Tumor-induced angiogenesis studied in confrontation cultures of multicellular tumor spheroids and embryoid bodies grown from pluripotent embryonic stem cells 1. FASEB J 2001;15:995-1005.

37. Emin Erbaycu A, Gunduz A, Batum O, Zeren Ucar Z, Tuksavul F, Zeki Guclu S. Pre-treatment and treatment-induced neuron-specific enolase in patients with small-cell lung cancer: an open prospective study. Arch Bronconeumol. 2010 Jul;46(7):364-9. Epub 2010 May 27.

38. Koukourakis MI, Giatromanolaki A, Sivridis E, Gatter KC, Trarbach T, Folprecht G et al. Prognostic and predictive role of lactate dehydrogenase 5 expression in colorectal cancer patients treated with PTK787/ZK 222584 (vatalanib) antiangiogenic therapy. Clin Cancer Res 2011;17:4892-900.

39. Rui Z, Jian-Guo J, Yuan-Peng T, Hai P, Bing-Gen R. Use of serological proteomic methods to find biomarkers associated with breast cancer. Proteomics 2003;3:433-9.

40. Chatterjee M, Mohapatra S, Ionan A, Bawa G, Ali-Fehmi R, Wang X et al. Diagnostic markers of ovarian cancer by high-throughput antigen cloning and detection on arrays. Cancer Res 2006;66:1181-90.

41. Akyol S, Gercel-Taylor C, Reynolds LC, Taylor DD. HSP-10 in ovarian cancer: expression and suppression of T-cell signaling. Gynecol Oncol. 2006 Jun;101(3):481-6.

42. Feldstein MS, Rhodes DJ, Parker AS, Orford RR, Castle EP. The haphazard approach to the early detection of asymptomatic renal cancer: results from a contemporary executive health programme. BJU Int. 2009 Jul;104(1):53-6. Epub 2009 Jan 12.

43. Perroud B, Lee J, Valkova N, Dhirapong A, Lin PY, Fiehn O et al. Pathway analysis of kidney cancer using proteomics and metabolic profiling. Mol Cancer 2006;5:64.

44. Locasale JW, Cantley LC, Vander Heiden MG. Cancer's insatiable appetite. Nat Biotechnol 2009;27:916-7.

45. Boer JM, Huber WK, Sultmann H, Wilmer F, von HA, Haas S et al. Identification and classification of differentially expressed genes in renal cell carcinoma by

expression profiling on a global human 31,500-element cDNA array. Genome Res 2001;11:1861-70.

46. Gieseg MA, Cody T, Man MZ, Madore SJ, Rubin MA, Kaldjian EP. Expression profiling of human renal carcinomas with functional taxonomic analysis. BMC Bioinformatics 2002;3:26.

47. Mazurek S, Boschek CB, Hugo F, Eigenbrodt E. Pyruvate kinase type M2 and its role in tumor growth and spreading. Semin Cancer Biol 2005;15:300-8.

48. Kaelin WG, Jr. The von Hippel-Lindau gene, kidney cancer, and oxygen sensing. J Am Soc Nephrol 2003;14:2703-11.

49. Turner KJ, Moore JW, Jones A, Taylor CF, Cuthbert-Heavens D, Han C et al. Expression of hypoxia-inducible factors in human renal cancer: relationship to angiogenesis and to the von Hippel-Lindau gene mutation. Cancer Res 2002;62:2957-61.

50. Gentil BJ, Benaud C, Delphin C, Remy C, Berezowski V, Cecchelli R et al. Specific AHNAK expression in brain endothelial cells with barrier properties. J Cell Physiol 2005;203:362-71.

Chapter 4

Identification of differentially expressed proteins in metastatic, relative to primary, ccRCC tissues as potential prognostic biomarkers

This chapter describes research work in which prognostic, tissue-based biomarkers for the onset of metastatic ccRCC were sought, as in Chapter 3, using iTRAQ LC-MS analysis to determine differentially expressed proteins in metastatic, relative to primary, ccRCC kidney tissue homogenates. From the resulting pool of overexpressed proteins, those known to be involved in aggressive tumorigenesis, were selected for further verification. The proteins that were confirmed to be overexpressed in primary ccRCC tumors that *had metastasized* relative to those that *had not metastasized,* were proposed as potential prognostic biomarkers. If such biomarkers were to be found overexpressed in primary tumor biopsy samples, it could indicate a predisposition of the tumor to metastasize; therefore, the patient should seek more aggressive treatment.

In this project, my collaborators performed pathological analysis of kidney tissues, isolated and purified tissue proteins, constructed tissue microarrays, quantified the immunoexpression of the proteins, and performed portions of the bioinformatics analysis. I performed the trypsin digestion of the proteins and iTRAQ labeling of the resulting peptides, the SCX chromatography and RP LC-MS analysis, the Protein Pilot database search, the PIE list preparation, the clustering analysis, the bioinformatics analysis, the WB analysis and its quantitation, and the IHC analysis. Dr. Olga Krakovska recalculated the iTRAQ ratios.

The results from this project were resubmitted for publication, after minor corrections, on August 28, 2012 in "Molecular and Cellular Proteomics" (ID#: MCP/2012/020701).

Quantitative Proteomic Analysis in Metastatic Renal Cell Carcinoma Reveals a Unique Set of Proteins with Potential Prognostic Significance

Olena Masui,[1] Nicole M.A. White,[2,3] Leroi V. DeSouza,[1] Olga Krakovska,[1] Ajay Matta,[1] Shereen Metias,[2] Bishoy Khalil,[2] Alexander D. Romaschin[2,3], R. John Honey[4] , Robert Stewart[4], Kenneth Pace[4] , Georg A Bjarnason,[5] K.W. Michael Siu,[1] and George M. Yousef.[2,3*]

1. Department of Chemistry and Centre for Research in Mass Spectrometry, 4700 Keele Street York University, Toronto, Canada, M3J 1P3.
2. The Keenan Research Center in the Li Ka Shing Knowledge Institute and the Department of Laboratory Medicine, St. Michael's Hospital, Toronto, Canada, M5B 1W8.
3. Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada, M5S 1A8.
4. Department of Surgery, St. Michael's Hospital, Toronto, Canada, M5B 1W8
5. Division of Medical Oncology and Hematology, Sunnybrook Health Sciences, Toronto, Canada, M4N 3M5

* To whom correspondence should be addressed.

Running title: *Molecular profiling in metastatic kidney cancer*

Correspondence to:
George M Yousef, MD PhD FRCPC (Path) MSc MBBCh
Department of Laboratory Medicine, St. Michael's Hospital
30 Bond Street, Toronto, ON, M5B 1W8, Canada
Tel: 416-864-6060 Ext: 77605 **Fax**: 416-864-5648
e-mail: yousefg@smh.ca

Abbreviations

14-3-3ζ: 14-3-3 protein zeta/delta, ccRCC: clear cell renal cell carcinoma, CE: collision energy, FDR: false discovery rate, Gal-1: galectin-1, Gal-3: galectin-3, GO: Gene Ontology, IDA: information-dependent acquisition, IHC: immunohistochemistry, iTRAQ: isobaric tags for relative and absolute quantitation, PIE: precursor ion exclusion, Pfn1: profilin-1, RCC: renal cell carcinoma, RP: reverse phase, SCX: strong cation exchange, TMA: tissue microarray, WB: western blot.

Summary

Metastatic renal cell carcinoma (RCC) is one of the most treatment-resistant malignancies and patients have a dismal prognosis with <10% five-year survival rate. Identification of markers that can predict the potential of metastases will have a great impact in improving patient outcome. In this study, we used differential proteomics with isobaric tags for relative and absolute quantitation (iTRAQ) labeling and LC-MS/MS analysis to identify proteins that are differentially expressed in metastatic compared to primary RCC. We identified 1256 non-redundant proteins and 456 of these were quantified. Further analysis identified 29 proteins that were differentially expressed (12 overexpressed and 17 under expressed) in metastatic vs. primary RCC. Dysregulated protein expressions of profilin-1 (Pfn1), 14-3-3 zeta/delta (14-3-3ζ), and galectin-1 (Gal-1) were verified on two independent sets of tissues by western blot and immunohistochemical analysis. Hierarchical clustering analysis showed the protein expression profile specific for metastatic RCC can distinguish between aggressive and non-aggressive RCC. Pathway analysis showed that dysregulated proteins are involved in cellular processes related to tumor progression and metastasis. Furthermore, preliminary analysis using a small set of tumors showed that increased expression of Pfn1 is associated with poor outcome and is a potential prognostic marker in RCC. In addition, 14-3-3ζ and Gal-1 also showed higher expression in the tumors with poor prognosis compared to those with good prognosis. Dysregulated proteins in metastatic RCC represent potential prognostic markers for kidney cancer patients and a greater understanding of their involved biological pathways can serve as the foundation of the development of novel targeted therapies for metastatic RCC.

Introduction

Renal cell carcinoma (RCC) is the most common neoplasm of the adult kidney. Worldwide incidence and mortality rates of RCC are rising each decade (1). Seventy-five percent of kidney tumors are of the clear cell (ccRCC) subtype (2). Although modern imaging techniques for abdominal screening have led to increased incidental detection of renal tumors (3), unfortunately approximately 25-30% patients still have metastases at presentation.

The prognosis of RCC is quite variable. The greatest risk of recurrence following nephrectomy is within the first 3-5 years (4). The ability to predict which tumors will metastasize would have a significant impact on patient outcome since the likelihood of a favorable response to treatment is greater when the metastatic burden is limited and surgical resection of a single or limited number of metastases can result in longer survival (5). Furthermore, approximately 3% of patients will develop a second primary renal tumor, either synchronous or metachronous. Currently, patient prognosis is assessed by histological parameters and a multivariate analysis developed at Memorial Sloan Kettering (6), but neither is sufficiently accurate. A more accurate assessment of prognosis is urgently needed to better guide patient's management.

While surgery is curative for localized disease, many patients eventually relapse. Metastatic RCC is one of the most treatment-resistant malignancies with chemotherapy and radiotherapy having limited effect. The five-year survival rate for metastatic RCC is $\leq 10\%$ (7). Although there has been much progress in RCC treatment with the new era of antiangiogenic therapy, the majority of patients ultimately suffers a relapse and die from progression of their cancer. A more in-depth understanding of the pathogenesis of

metastasis will be a cornerstone towards developing new targeted therapies. A number of prognostic markers have previously been identified based on comparative analysis of primary and metastatic tumors, including C-reactive protein, tetraspanin 7 (TSPAN7), hypoxia-inducible factor 1 alpha (HIF-1α), phos-S6, U3 small nucleolar ribonucleoprotein protein (IMP3), carbonic anhydrase IX (CAIX) and microvascular density (8-14). However, no biomarker has yet had an established clinical role independent of stage (15). Differential protein expression between primary RCC and normal tissues was previously studied (16-18). Also, differential expression between primary and metastatic kidney has been investigated at the miRNA level (19, 20). Molecular analyses hold the promise of obtaining a better understanding of the pathogenesis of kidney cancer (21).

In this study, we aimed to elucidate the pathogenesis of RCC metastasis through proteomic analysis and to identify potential prognostic markers for kidney cancer. We performed quantitative proteomic analysis using isobaric tags for relative and absolute quantitation (iTRAQ) labeling and LC-MS/MS to identify proteins that were dysregulated in metastatic compared to primary RCC. Differential expressions of selected, biologically interesting proteins; profilin-1 (Pfn1), 14-3-3 zeta/delta (14-3-3ζ), and galectin-1 (Gal-1) were validated on two independent sets of tumors by Western blot (WB) analysis and immunohistochemistry (IHC). Hierarchical clustering analysis showed that differential protein expression can distinguish between aggressive and non-aggressive tumors. In order to explore the role of these dysregulated proteins in tumor progression, we performed Gene Ontology (GO) and pathway analyses. In addition, we

carried out preliminary analysis to assess the potential of Pfn1, 14-3-3ζ, and Gal-1 as prognostic markers in RCC.

Experimental Procedures

Patients and specimens

Primary ccRCC tissues and matched normal kidney tissues from the same patient were obtained from nephrectomy specimens at St. Michael's Hospital and the Ontario Tumor Bank, Toronto, Canada. We also collected unmatched metastatic RCC tissues. Specimens were collected immediately following nephrectomy and flash frozen in liquid nitrogen in 2mL cryogenic tubes. As RCC is known to arise from the proximal tubules (22), the kidney cortex is considered a suitable representation of normal kidney tissue (23). All specimens were histologically confirmed by a pathologist. The study was approved by the Research Ethics Boards of York University, St. Michael's Hospital and the Ontario Cancer Institute. Relevant clinical information on the patients is shown in **Supplementary Table 1**.

Tissue preparation and protein extraction

A schematic of the work flow is shown in **Figure 1**. Tissues were prepared as described previously (24-26). Briefly, tissues were washed three times in ice-cold phosphate buffered saline (PBS) and homogenized using a hand-held homogenizer in a protease-inhibitor cocktail (Roche, Laval, Canada). Cell debris was then removed by centrifugation at 4°C for 30 min at 14,000 rpm. The clarified supernatant was transferred to a fresh 1.5 mL tube. A reference sample was prepared from a pool of six combined normal kidney tissues. Protein concentrations were determined using the Bradford assay (Sigma-Aldrich, St. Louis, USA) (24, 26). Equal amounts of protein from each tissue type were digested with trypsin, labeled with iTRAQ and combined. Samples were then

separated by off-line SCX liquid chromatography and analyzed by RP LC-MS/MS. Proteins were identified and quantified using Protein Pilot and then subjected to additional characterization, including verification by WB, clustering, GO analysis, pathway analysis, and IHC.

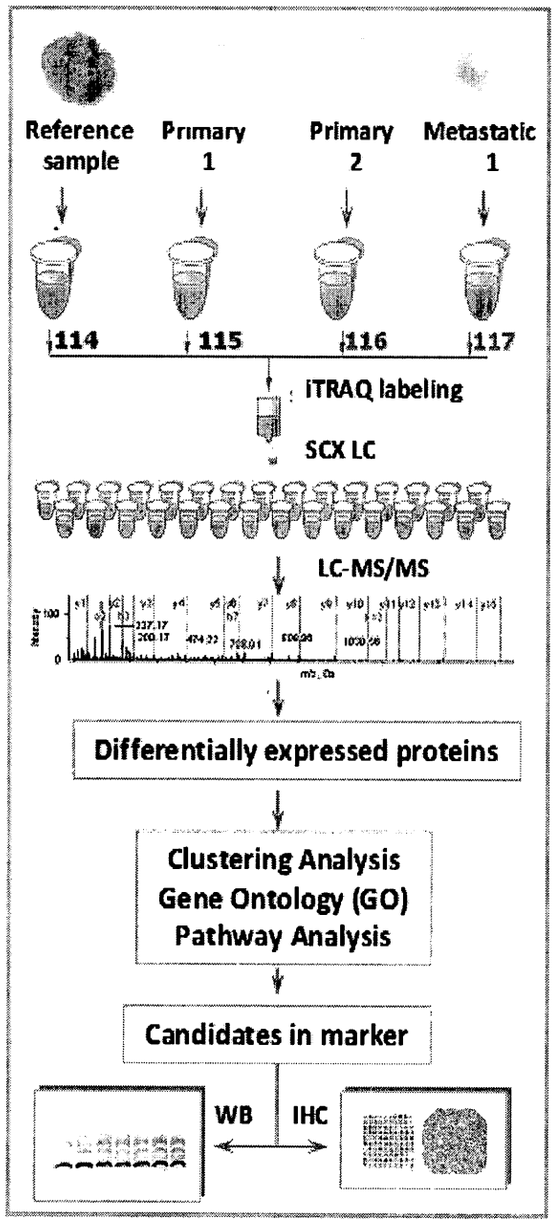


Figure 1 Work flow for quantitative proteomic analysis. *Six pairs of primary ccRCC and normal matched kidney tissues from the same patient as well as six metastatic RCC tissues were analyzed. The reference sample consisted of a homogenate of the six normal kidney tissues. Each sample was digested individually with trypsin and labeled with the appropriate iTRAQ tag. The labeled digests were then pooled and separated by offline SCX LC. Each fraction was analyzed in triplicate by on-line RP nano-LC-MS/MS. Exclusion lists were used to minimize redundancy. MS data were analyzed by Protein Pilot to identify and quantify proteins using a cut-off of 5% local FDR. Dysregulated proteins were validated by WB and IHC, and further analyzed by clustering, GO and pathway analyses.*

iTRAQ sample labeling

For iTRAQ LC-MS/MS analysis, 100 μg of clarified supernatants were denatured for 1 hour at 60°C, disulfide bonds were reduced, and the cysteine residues were blocked as described in the iTRAQ protocol (Applied Biosystems, Foster City, CA). Supernatants were then divided into sets of four, each containing one aliquot of the reference sample and three ccRCC malignant or individual non-malignant kidney samples. Each sample was then digested with trypsin and labeled with the iTRAQ tags (labeling details shown in **Supplementary Table 2**). Labeling of the reference sample was randomized for each set to eliminate any potential for bias that might be associated with a particular iTRAQ reporter tag. The iTRAQ-labeled samples were then combined according to the specified set and transferred into fresh 1.5 mL tubes. Each iTRAQ set was then dried using a vacuum centrifuge (Thermo Savant SC110 A, Holbrook, NY, USA).

Strong cation exchange (SCX) chromatography

The iTRAQ sets were dissolved in 1.7 mL of Buffer A (10mM H_3PO_4/KH_2PO_4, in an aqueous solution of 25% acetonitrile and acidified to a pH of 3.0 with phosphoric acid) and filtered using a 0.45-μm syringe filter (Millipore, Cambridge, ON, Canada). Each set was then separated by off-line SCX chromatography using an HP1050 HPLC instrument (Agilent, Palo Alto, CA) with a 2.1-mm internal diameter x 100-mm-length PolyLC Polysulfoethyl A column packed with 5-μm beads with 300-Å pores (The Nest Group, Southborough, MA) as described previously (27). Separation was performed using a linear binary gradient over 1 h (see details in **Supplementary Table 3**) of Buffer A and Buffer B, where Buffer B was composed of Buffer A and 350mM potassium

chloride. Buffer C was composed of Buffer A and 1M potassium chloride and was used to strip the column after the run. A total of 30 SCX fractions were collected per iTRAQ set. These fractions were dried using a vacuum centrifuge as before.

Reverse phase (RP) LC-MS/MS

The SCX fractions were analyzed in triplicate using a nanobore LC system (LC Packings, Amsterdam, Netherlands) and a QSTAR Pulsar mass spectrometer (Applied Biosystems/MDS SCIEX, Foster City, CA) in positive ion mode, externally calibrated with tryptic peptides from bovine serum albumin. The first five fractions were not analyzed because they consisted of the void volume which contained unreacted iTRAQ labels as well as byproducts that would compromise the RP column. Fractions 6 – 17 were redissolved in 16μL of eluant A [consisting of 94.9% deionized water, 5.0% methanol, and 0.1% formic acid (pH 3)]. For subsequent fractions, the amount of eluant A was incremented by 2μL over the preceding fraction to accommodate the increase in the amount of KCl. A 1μL aliquot of the sample (~1μg of total peptides) was loaded onto a C18 RP pre-column (LC Packings: 300 μm x 5 mm) and desalted before separation on an RP analytical column (75-μm x 150-mm packed in-house with 3-μm Kromasil C18 beads with 100 Å pores, The Nest Group, Southborough, USA). Eluant A, consisting of 94.9% deionized water, 5.0% methanol, and 0.1% formic acid (pH = 3), was used to load the sample onto the C18 precolumn at a flow rate of 25μL min^{-1}. After 4 min, the C18 precolumn was switched in-line with the RP analytical column. Separation was performed at 100nL min^{-1} using a nonlinear binary gradient (see gradient in table below)

starting with Eluant A and transitioning to Eluant B, which consisted of 5.0% deionized water, 94.9% methanol, and 0.1% formic acid.

Time (min)	0.1	5	10	70	85	95	98	135
B (%)	5	5	15	35	80	80	5	Stop

MS data were acquired in information-dependent acquisition (IDA) mode using the Analyst QS 1.1 software (Applied Biosystems/MDS SCIEX, Foster City, USA). The LC-MS/MS analysis was performed using a 1-s TOF-MS survey scan from 400 to 1500 Da, followed by four, 2-s product-ion scans, from 80 to 2000 Da, of the four most-abundant ion peaks in the survey scan. The collision energy (CE) was automatically controlled by the IDA CE parameter script. Switching criteria were set for ions with *m/z* $\geq$ 400 and $\leq$1500, charge states of +2 to +4, and abundances of $\geq$10 counts. Using Analyst QS 1.1 controlled dynamic exclusion, former target ions were excluded for 30 s, and ions within a 100-ppm window were ignored. Precursor ion exclusion (PIE) lists were used to minimize redundancy.

Bioinformatics Analysis

Protein identification by Protein Pilot

LC-MS/MS data of each fraction was used to identify proteins by searching a concatenated Swissprot/Panther database of 66082 distinct human protein entries (version June 2, 2010). The database was searched using Protein Pilot software, version 2.0.1 (AB SCIEX, Foster City, USA), which uses the Paragon algorithm (28). Protein identification was performed with MMTS selected as cysteine modification, with the search option 'emphasis on biological modifications,' and with 'PSPEP' (Proteomics System

Performance Evaluation Pipeline Software) analysis checked. Peptide and protein summaries, and false discovery rate (FDR) reports were generated. Only proteins identified with local FDR ≤ 5% were considered for further analysis (29).

Iterative runs with PIE

To minimize redundancy in subsequent iterations, a PIE list was added to the acquisition method after each iteration as described previously (30, 31). PIE lists were generated using an Excel template developed in-house. To generate the list for each iteration, the peptide summary of a fraction, obtained after the previous iteration, was imported into an Excel template where: (1) the m/z values and elution times of peptides identified with > 95% confidence were extracted; (2) alternative charge states (only +2, +3 and +4 were considered) of the peptides were calculated; and (3) the next three higher isotopic m/z values of extracted and calculated peptides were determined. The resulting m/z ratios from all three of these considerations constituted the PIE list. This list was saved and imported into the acquisition method for the next iteration. The list used for each iteration was cumulative of all the m/z values and elution times derived from all previous iterations for the fraction. Tolerance windows for exclusion were set at 100 ppm for *m/z* and 360 s for elution time. The template is available in **Supplementary Data, PIE Template.**

iTRAQ ratio re-calculation and identification of differentially expressed proteins

To identify non-redundant proteins, data acquired for all 25 fractions from each iTRAQ set injected in triplicate were searched using Protein Pilot software. Proteins

identified in five iTRAQ sets were compiled together and matched by accession numbers using a script written in Matlab (version 7.7.0.471). Redundant proteins and peptides, and proteins identified in reverse sequence were removed from the list. To improve the confidence of protein quantitation, the mean expression iTRAQ ratios of the proteins were re-calculated based on the criteria that the protein must be identified by a minimum of three peptides, with ≥95% confidence, and with an expression ratio error factor <11.1%. To enhance confidence in the protein quantitation even more, we included only 95% of all quantified proteins with the lowest computed error factor (which corresponds to a confidence > 0.05 in **Supplementary Table 4**) for further consideration. Proteins were considered to be differentially expressed if iTRAQ ratios were ≥1.5 or ≤0.67 in ≥50% in metastatic relative to primary ccRCC samples.

GO Analysis

Proteins were classified into groups according to biological processes (e.g. metabolic process), molecular function (e.g. protein binding), and subcellular compartmentalization (e.g. cytoplasm, organelle, etc.), using the GO Consortium databases.

Clustering analysis

To determine if differentially expressed proteins can discriminate between metastatic and primary RCC samples the samples were hierarchically clustered based on quantified proteins. Proteins were included in the analysis if quantification was available in at least 50% of the samples. The average iTRAQ ratios were logarithmically

transformed for hierarchical clustering via the City-block distance method. Hierarchical clustering analysis was performed using the Cluster 3.0 software and the result was visualized with the TreeView software (Stanford University, Palo Alto, CA, http://www.dnachip.org), both of which were developed by Eisen et al. (32).

Western blot analysis

Dysregulated protein expression in metastatic RCC samples was verified by WB analysis. Briefly, 30μg of total protein were electrophoretically separated on a 10% SDS-PAGE gel. Proteins were then transferred to a PVDF membrane and probed with the following polyclonal antibodies; anti-Gal-1 and anti-Pfn1 (both from Abcam, Cambridge, USA), and anti-14-3-3ζ (Santa Cruz Biotechnologies, Santa Cruz, USA). β-actin (Cell Signaling Technology, Danvers, USA) was used as a loading control. Membranes were incubated with primary antibodies overnight at 4°C. Protein expression was visualized after incubation with secondary anti-rabbit antibodies conjugated with horseradish peroxidase and enhanced chemoluminescence reagent (Amersham Pharmacia Biotech, Piscataway, USA). The intensity of protein staining was determined using ImageJ, a publicly available Java-based image processing program (http://rsbweb.nih.gov/ij/). Average protein expression was calculated based on two independent WB analyses. Primary ccRCC samples were compared to non-malignant kidney samples using the paired sample two-tailed t-test. Metastatic RCC samples were compared to primary ccRCC samples using the Mann-Whitney two-tailed test. $p \leq 0.05$ was considered as significant.

Tissue microarray construction and immunohistochemistry

Appropriate areas from normal kidney tissue, primary ccRCC and metastatic RCC were selected and circled from donor blocks by a pathologist. Tissue microarray (TMA) blocks containing duplicate 1.0-mm cores from each specimen were constructed with a manual tissue microarrayer (Beecher Instruments, Sun Prairie, USA). The TMAs contained 22 cases of primary ccRCC and matched normal kidney tissues from the same patient, 12 cases of primary ccRCC from patients who later developed metastasis, and 26 metastatic RCC tissues. In addition, each block contained two marker cores for TMA orientation.

TMA sections were cut 5μm thick and placed on charged slides. Slides were deparaffinized in xylene, hydrated in gradient ethanol, and pre-treated in a microwave oven for 20 min at 800 W in 1 L of citrate buffer (0.01 M, pH 6.0) for antigen retrieval. Sections were then incubated with hydrogen peroxide (0.3% v/v) in PBS for 15 min to quench the endogenous peroxidase activity, followed by blocking with 10% FBS (fetal bovine serum) in PBS to preclude non-specific binding. Thereafter, the slides were incubated with primary antibodies: Gal-1, Pfn1 or 14-3-3ζ overnight at 4°C. Protein expression was detected using the streptavidin-biotin complex with the Dako LSAB+ kit (Dako Cytomation, Glostrup, Denmark) and diaminobenzidine as the chromogen. All procedures were carried out at room temperature unless otherwise specified. Slides were washed with 0.025% Triton X 100 in PBS (0.1M, pH = 7.3) three times after each step. Finally, sections were counterstained with Mayer's hematoxylin and mounted with DPX mountant. In the negative control tissue sections, the primary antibody was replaced by isotype specific non-immune mouse / rabbit IgG.

Immunoexpression of each protein was evaluated by a pathologist. Quantification in tumor sections was classified into four categories: (A) moderate to strong membrane, cytoplasmic and nuclear staining in greater than 50% tumor cells; (B) moderate to strong cytoplasmic staining in >50% of either the cytoplasm or nuclei, but not both; (C) overall weak staining in the cytoplasm and/or nuclei; and (D) no staining.

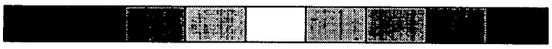
Results

Identification of differentially expressed proteins between primary and metastatic RCC

Using Protein Pilot, we identified a total of 1256 non-redundant proteins with local FDR $\leq$ 5% (**Supplementary Table 4**); 456 of these proteins were reliably quantified (**Supplementary Table 5**). Twenty-nine proteins met our definition for differential expression (see Experimental Procedures) in a comparison between metastatic and primary RCC: 12 were overexpressed (iTRAQ ratios of $\geq$1.5) and 17 were underexpressed (iTRAQ ratios of $\leq$0.67). **Table 1** and **Supplementary Table 6** shows a heat map of the 29 differentially expressed proteins. A literature search showed that all 29 proteins had previously been associated with other malignancies. For example, Gal-1 has previously been reported to be associated with cell migration and invasion in a metastatic murine lung cancer model (33). Gal-1 was also shown to have prognostic significance in epithelial ovarian cancer (34).

Table 1. Heat map showing expression of 29 proteins that are dysregulated in metastatic compared to primary renal cell carcinoma.

Gene symbol	M1	M2	M3	M4	M5	M6	P1	P2	P3	P4	P5	P6	Relation to Cancer	Ref.
PFN1 (Pfn1)	1.69	1.53	1.84	1.99	1.40	1.26	0.95	0.96	1.78	1.15	1.57	1.13	Marker for aggressive bladder ca	(48, 64)
LGALS1 (Gal-1)		2.45	2.18	1.98					1.46	1.25	1.10		Enhances tumorigenicity in lung ca	(33, 65)
YWHAZ (14-3-	1.47	1.59	1.44		1.81	1.58	0.96	0.98	1.40	1.29	2.30	0.86	Cervical ca mets	(66, 67)
ALDOA	1.67		1.34	1.67		2.17	1.14	0.94	1.82	1.32	1.64	1.01	Aggressive HCC and lung ca	(68, 69)
GAPDH	0.99		0.94	1.70	2.24	1.76	0.93	0.99	2.25	1.15	2.14	0.82	Up in colon ca and liver mets	(70)
GPI	0.98	2.42	0.98	1.10	2.43	1.74	1.07	0.98		0.99	1.63	0.85	Liver mets in colon ca	(71, 72)
LGALS3		1.72		1.67	2.14				1.66	1.11	2.22		Gastric ca, cell migration and	(52, 53)
P4HB	1.54	2.21	2.22	2.24	1.88	1.60	1.28	1.00	1.45	1.17	1.68	0.84	Up in melanoma and endometrial ca	(73, 74)
PKM2	1.77			2.04			1.33	0.91		2.10		0.84	Contributes to tumorigenesis	(75)
S100A11							1.21	0.62	2.14	1.93	2.31		Colon ca mets, cell growth in lung ca	(76, 77)
TPI1	0.98		1.03	1.65	2.29	1.16	0.98	1.00	1.93	1.34		0.88	Breast ca mets, gastric ca cell lines	(78, 79)
VIM	1.68	0.84	2.15	1.24		2.40	0.98	1.10	2.26	1.99	2.58	0.89	Bone mets, cell migration	(80, 81)
ACAA2		0.70	0.56	0.39		0.29	1.01	1.13	0.69	0.77		1.03	Apoptosis in HCC cells	(82)
AGMAT	0.27	0.19	0.10		0.20		1.09	1.19	0.19		0.31	1.17	Down in ccRCC	(49, 50)
AKR1A1	0.58	0.77	0.60		0.57	0.66	0.88	1.15	0.78	0.61	0.56	0.86	Dysregulation in gastric ca	(83)
ATP5B	0.94	0.59		0.34			1.60	0.94	0.58	0.56	0.39	0.90	Cell survival	(84)
CA2	0.61		0.38	0.40	0.92	1.02	0.90	0.92	0.69	0.62	0.40	1.06	Up in uterine ca and gastric ca mets	(85, 86)
CRYL1	0.57	0.96		0.33	0.71		0.99	1.09	1.01	0.86	0.69	0.99	Associated with HCC DFS	(87)
DCXR	0.38		0.37		0.29		0.96	1.23				1.38	Malignant progression of lesions	(88)
DDC					0.21	0.33	0.96	1.30	0.70			1.04	Marker for gastric ca	(89)
DLST			0.28				1.36	1.18			0.87	1.13	Mitochondrial respiratory complexes	(90)
FBP1	0.33		0.38	0.31			0.95	1.34		0.27	0.25	0.95	Down in HCC and colon ca	(91, 92)
GPD1	0.38			0.32		0.39	0.83	1.13		0.40		0.94	Down in breast ca	(93)
HADH	0.32			0.16	0.31		1.03	1.06		0.28	0.25	1.07	Metastatic melanoma	(94)
MDH1	0.67	0.87			0.33	0.82	0.83	0.91	0.90		0.84	0.81	Cell proliferation	(95)
MDH2	0.69	0.73		0.35	1.64		1.13	1.06	0.72	1.64	0.78	0.94	Metastatic muscle and liver tumors	(96)
PCK2				0.33	0.36		0.95	1.38		0.38	0.33	1.13	Down in pancreatic ca	(97)
PEBP1		1.23	0.31	0.30	0.53		1.01	0.96	0.80	0.57	0.34	0.83	Inhibits prostate ca invasion and	(51, 98)
PEPD				0.40			0.87	0.86				0.89	Metastatic potential in nude mice	(99)



≤0.4 ≤0.5 ≤0.67 ≤0.8 ≥1.25 ≥1.5 ≥2.0 ≥2.5

ca: cancer; DFS: disease-free survival; HCC: hepatocellular carcinoma; Mets: metastasis. For 'Protein name' we used the gene name according to UniProtKB. For a full name of the protein and its accession number, see ***Supplementary Table 6.***

Clustering analysis indicated differential protein expression can discriminate between metastatic and primary RCC

In order to examine the hypothesis that the metastatic potential is present at least in part of the primary tumor, cluster analysis was performed on 456 proteins for which quantitative information was available. The samples clustered into two main groups: one that contained the primary RCC samples, P1, P2 and P6; and a second that had the six metastatic cases plus the other three primary RCC samples, P3, P4 and P5 (data not shown). Clustering of the last three primary RCC samples with the six metastatic samples becomes less puzzling after data from clinical follow-up were examined: one of the primary RCC patients developed subsequent metastasis to the liver as well as having a history of colon cancer; a second patient also had an earlier cancer; while the third patient had no reported metastasis thus far. The other group of primary RCC patients, P1, P2 and P6 did not develop metastasis for five years. If validated on a larger tumor set, these data then strongly suggest that RCCs have a unique protein expression pattern that is required for metastasis and that differentially expressed proteins can discriminate between aggressive and non-aggressive RCCs.

Validation of dysregulated protein expression

To proceed with the first steps in validating our MS analysis, we confirmed the differential expressions of three dysregulated proteins by WB analysis using samples from the same RCC patient cohort. The proteins Gal-1, Pfn1 and 14-3-3ζ, were selected for validation based on their interesting biology and potential significance in RCC tumorigensis (35-40). The expressions of Gal-1, Pfn1, and 14-3-3ζ were found to be elevated in primary ccRCC when compared to normal kidney tissue (a typical set of results is shown in **Figure 2A**) when analyzed by WB analysis. Additionally, all three proteins were further upregulated in metastatic compared to primary ccRCC. Densitometry showed that 14-3-3ζ, Pfn1 and Gal-1 were upregulated 1.93 fold ($p < 0.05$), 2.28 fold ($p < 0.05$), and 2.50 fold ($p < 0.1$), respectively, in primary ccRCC when compared to normal kidney tissue (**Figure 2B**). When we compared protein expression in metastatic vs. primary tissues, 14-3-3ζ and Pfn1 showed significant increased expression (1.77-fold change, $p < 0.05$; and 1.92-fold change, $p < 0.01$, respectively), while Gal-1 showed 2.5 times increased in expression but this finding did not reach 95% statistical significance ($p < 0.1$).

We additionally verified the overexpressions of Pfn1, Gal-1, and 14-3-3ζ in metastatic vs. primary RCC in an independent cohort of patients by IHC using TMAs consisting of 22 cases of primary ccRCC and 26 metastatic RCC tissues (**Figure 3** shows a typical set of results). We quantified immunoexpression using the four-tier scale described in Experimental Procedures (ranging from A which is the highest to D which is no staining).

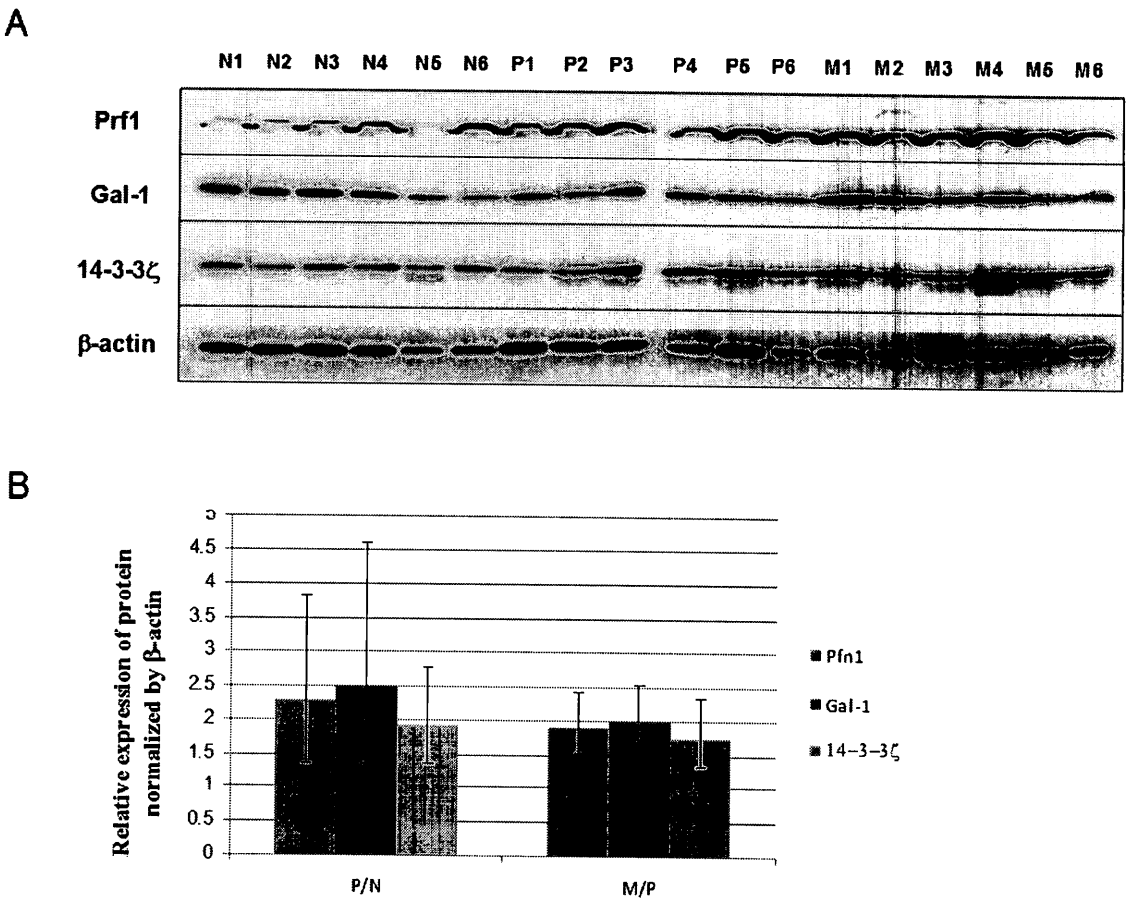


Figure 2. Verification of Gal-1, Pfn1, and 14-3-3 overexpression in metastasis by Western blot analysis. *(**A**) Representative blots showing the expression of Gal-1, Pfn1, and 14-3- in normal kidney tissues (N), primary (P), and metastatic ccRCC (M). For each of these proteins, expression was increased in primary tumor tissues when compared to normal kidney tissue; as well, the expression of all proteins was increased in metastatic tissues when compared to primary ccRCCs. -actin was used as a loading control. (**B**) Graphical representation of the average fold change in expression of the three proteins between six primary ccRCCs and matched normal specimens (P/N) and that between six primary ccRCCs and six unmatched metastatic cases.*

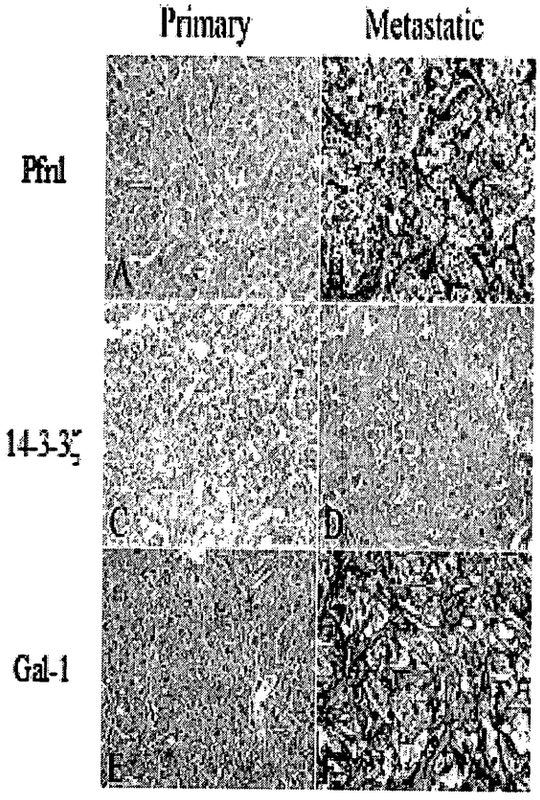


Figure 3. Verification of Gal-1, Pfn1, and 14-3-3ζ overexpression in metastatic ccRCCs by immuno-histochemical analysis. *Representative photomicrographs show the overexpression of Gal-1, Pfn1 and 14-3-3ζ in metastatic vs. primary RCC tissue by IHC. Pfn1 shows higher intensity and % positivity in metastatic (**B**) compared to primary tumors (**A**). 14-3-3ζ shows more intense cytoplasmic and nuclear staining in metastatic (**D**) than primary (**C**) RCCs. Similarly, Gal-1 staining is more intense in metastatic (F) than primary (E) tumor samples. Original magnification × 200.*

For Pfn1 expression, we found an increased expression in metastatic compared to primary ccRCC. Most of the primary tissues (19/22, 86%) showed the B expression level (**Figure 3A**). There were also two (9%) tumor samples that exhibited the A expression level, and one (5%) tumor sample that showed the C level. For metastasis, we found that the majority (18/23, 78%; 3 cases were omitted because they were washed off the slide) of samples showed the A level of expression (**Figure 3B**) while five (22%) showed the B level, indicating overall increased Pfn1 expression in metastatic compared to primary tissues. We then examined the expression of 14-3-3ζ in this independent tumor set. Similar to the MS results, 14-3-3ζ showed increased expression in metastatic when compared to primary ccRCC. 28% (6/22) primary ccRCC tumor samples had the A level of expression (**Figure 3C**), 36% (8/22) exhibited the B level, and 36% (8/22) the C level.

By contrast, in the 24 metastatic tissues that were available for the examination of 14-3-3ζ expression, we found that 67% (16/24, 2 cases were omitted because they were washed off the slide) showed the A level of expression (**Figure 3D**), while 29% (7/24) showed the B type level and 4% (1/24) showed the C level, again indicating an overall increase of 14-3-3ζ expression in metastatic compared to primary RCC. The differences in Gal-1 expression between primary and metastatic RCC was not statistically significant. (**Figures 3E-F**). This may be attributed to the high background staining that does not allow accurate quantification. A summary of our IHC analyses is shown in **Table 2**.

Table 2. Immunohistochemical expression of profilin-1, galectin-1 and 14-3-3ζ in metastatic and primary ccRCC.

Protein name	Expression level	22 primary paired ccRCC	26 metastatic ccRCC
Pfn1	A B C	2 (9%) 19 (86%) 1 (5%)	*18 (78%) 5 (22%)
Gal-1	A B C D	12 (55%) 8 (36%) 2 (9%)	*13 (57%) 8 (35%) 1 (4%) 1 (4%)
14-3-3ζ	A B C	6 (28%) 8 (36%) 8 (36%)	**16 (67%) 7 (29%) 1 (4%)

** 3 cases were omitted because they were washed off the slide; ** 2 cases were omitted because they were washed off the slide; **A**: moderate to strong membrane, cytoplasmic and nuclear staining in > 50% tumor cells; **B**: moderate to strong cytoplasmic staining in >50% either the cytoplasm or nuclei but not both; **C**: overall weak staining in the cytoplasm and/or nuclei; **D**: no staining.*

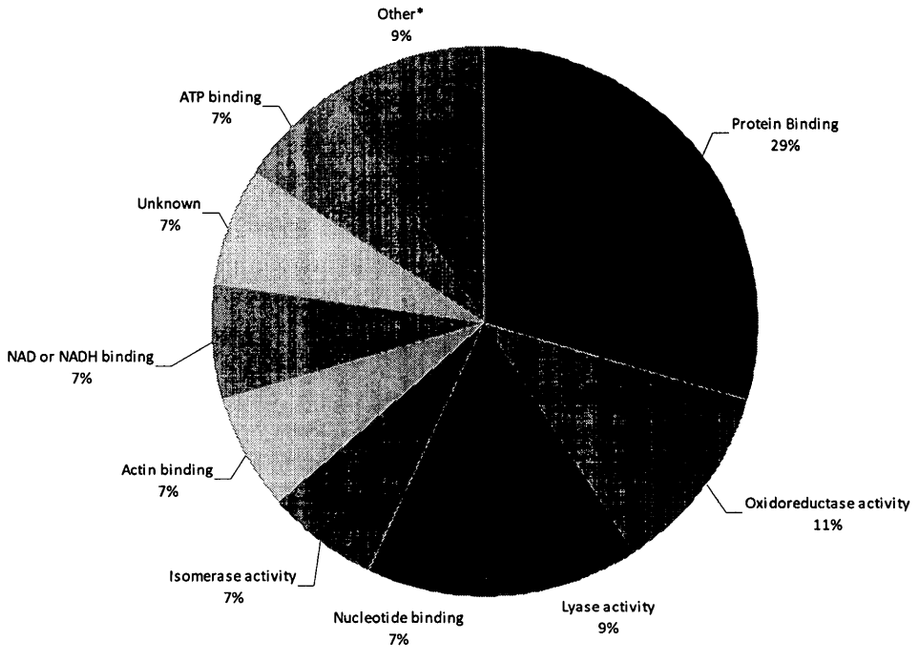
GO Analysis

We subjected the 29 identified proteins that were differentially expressed in metastatic vs. primary ccRCC to GO analysis and categorized them according to molecular function, biological processes and cellular component. When we analyzed dysregulated proteins for molecular function, we found that over one third of proteins were grouped under the GO term catalytic activity (GO:0003824, p=0.0016, **Figure 4A**), which included proteins involved in isomerase activity such as glucose-6-phosphate isomerase, triosephosphate isomerase 1 and protein disulfide isomerase, among others. In addition, we found 69% proteins analyzed were grouped under the GO term protein binding (GO:0005515, p<0.001). This category includes actin binding (GO:0003779) in which both Pfn1 and alcohol dehydrogenase A fall under.

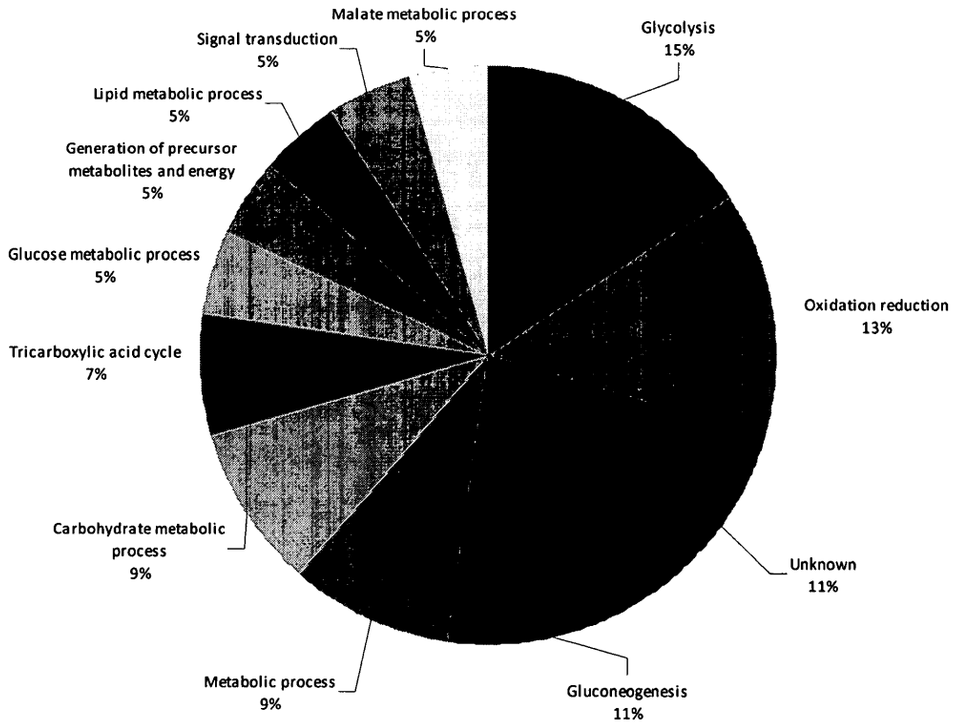
We also grouped dysregulated proteins based on biological processes. We found that a significant number of proteins were grouped under the multicellular organismal process (GO:0032501, p=0.00441) and biological regulation (GO:0065007; p<0.001, **Figure 4B**) which includes the proteins Gal-1 and Pfn1. In addition, when the 29 identified dysregulated proteins were analyzed for their cellular localization, we found 76% proteins were located intracellularly (p=0.03016), including proteins associated with the both plasma and organelle membranes (p=0.01703). There was also a significant association with the extracellular region (p<0.001, details not shown).

Figure 4. Gene Ontology (GO) Analysis. *Pie charts showing the results of GO analysis. The 29 dysregulated proteins were analyzed for* ***(A)*** *molecular function and* ***(B)*** *biological process. Significance values for each function and process are shown in the figure.*

Molecular Function



Biological Process



Differentially expressed proteins in metastatic RCC are involved in a number of pathways related to tumor progression and metastasis

We performed pathway analysis on the 29 dysregulated proteins. A subset of proteins, identified in our study, is involved in cell migration and invasion, cell-cell adhesion, and epithelial to mesenchymal transition (**Figure 5**).

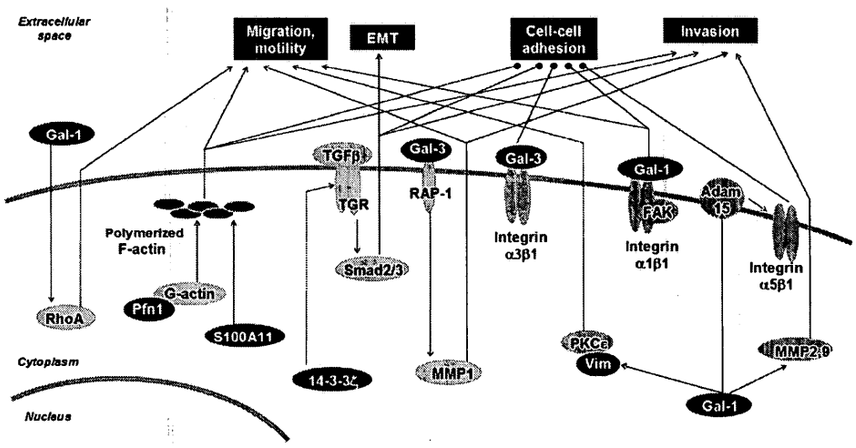


Figure 5. The involvement of a subgroup of differentially expressed proteins in cell-cell adhesion, migration and invasion. *Gal-1 is involved in cell migration through interaction with the α1β1 integrin, modulates cell adhesion and cell motility via Gal-1-induced expression of RhoA and alteration of the polimerization of the actin cytoskeleton, and promotes tumor invasion by reorganizing actin cytoskeleton and upregulating matrix metalloproteinase (MMP)-9 and MMP-2. 14-3-3ζ cooperates with ErbB2 to promote cell motility and migration via the activation of Src, and induce EMT by activating the TGFβ pathway to reduce cell adhesion. Pfn1 plays a role in cell adhesion and motility through interaction with G-actin. Pathways involved in tumor progression and metastasis are shown in white lettering on a black rectangle; differentially expressed proteins: white lettering on a red oval; other proteins involved in pathways: black letters on a grey oval. Lines with arrowheads represent documented interactions.*

Gal-1 has been reported to be involved in a number of pathways that can contribute to tumor progression and metastasis. It can promote tumor invasion by upregulating matrix metalloproteinase (MMP)-9 and MMP-2 and by reorganizing the actin cytoskeleton in lung adenocarcinoma (40). In addition, Gal-1 enhances the activation of Cdc42, thus increasing the number and length of filopodia on tumor cells (40). Dysregulated Pfn1 has been previously reported to be involved in the restoration of adherent junctions in breast cancer cells (41) and galectin-3 (Gal-3) has been shown to be involved in breast cancer cell adhesion (42). The identified proteins also have an effect on cellular migration, as S100A11 has been shown to mediate hypoxia-induced mitogenic factor-induced smooth muscle cell migration (43) and both 14-3-3ζ (44) and vimentin (45) have been shown to be involved in cellular migration.

The potential clinical significance of proteins dysregulated in metastatic RCC

In order to determine whether the dysregulated proteins identified herein had potential as prognostic markers or not, we performed additional preliminary analyses using IHC on an expanded set of primary RCC samples. We analyzed the expressions of Pfn1, Gal-1 and 14-3-3ζ in a total of 34 primary ccRCC: 12 cases with poor prognosis (developed metastasis within 3 years) and 22 cases with good prognosis (no metastasis within 5 years; same set used in the comparison with metastatic ccRCC, Table 2). We quantified immunoexpression in terms of the four categories described above. A summary of the expression levels of the three proteins in patients with poor vs. good prognosis is shown in **Figure 6.**

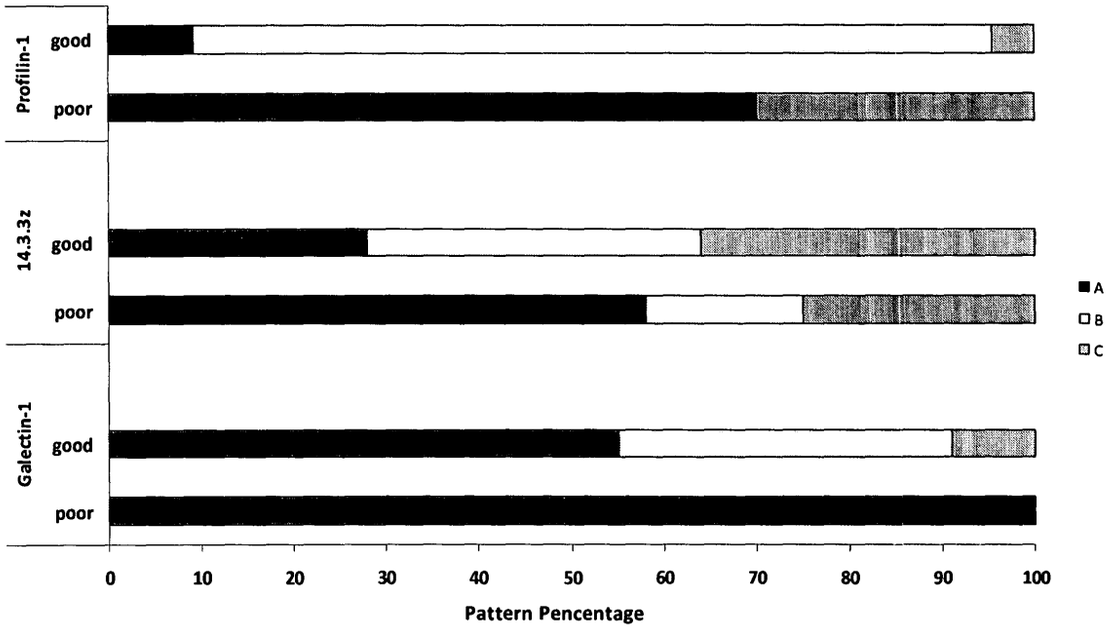


Figure 6. Stacked bar graphs showing differential staining patterns of Pfn1, Gal-1 and 14-3-3ζ between patients with good prognosis and those with poor prognosis. *Higher expression levels were associated with tumors with poor prognosis. Expression was quantified in a four-tier scale (from A with the highest expression to D with no expression, as described in the text).*

For the expression of Pfn1 in the 10 primary ccRCCs with poor prognosis (two cases were washed off during IHC handling), seven (70%) samples showed the A level of expression and three (30%) showed the C level. By contrast, only 2/22 cases (9%) in the good prognostic group showed the A expression level, while 19/22 (86%) showed the B level and 1/22 (5%) the C level of expression. Significantly, the percent of tumor cases displaying the A level of expression (70% vs. 5%, respectively) was drastically different

between the poor and the good prognosis cases, and resembled that in metastatic RCCs, suggesting that Pfn1 may have prognostic value for RCC patients.

We also investigated the potential of 14-3-3ζ as a prognostic marker. In the tumor cases with poor prognosis, 58% (7/12) exhibited the A level of expression. The B level was seen in 17% (2/12) and the C level in 25% (3/12) of cases. This is to be contrasted with the expression in 22 tumor cases with good prognosis, we found that only 6/22 (28%) of the samples showed the A expression level. The B and C levels were each seen in 36%. Again, the expression of 14-3-3ζ in ccRCC of poor prognosis resembled more closely to that of metastatic ccRCC and strongly suggests that 14-3-3ζ may have clinical significance for prognosticating kidney cancer patients.

Finally, for Gal-1 we found that 12/12 (100%) tumors from patients with poor prognosis exhibited the A level of expression. When we examined the patients with good prognosis, we found that only 12/22 (55%) exhibited the A level, while 8/22 (36%) the B level and 2/22 (9%) the C level of expression.

Discussion

Despite the many recent advances in metastatic RCC treatment through targeted therapies, the survival rate for metastatic RCC is extremely low at a five-year survival rate of <10%. Regrettably, there are no prognostic molecular markers that would predict whether a tumor will behave aggressively or remain indolent. It is abundantly clear that tumor biology plays a significant role in resultant tumor behavior (15). Unfortunately, RCC primary tumors that are placed in the same prognostic category based on currently used parameters, may behave differently. It is our hypothesis that the underlying biology of these tumors and differences in its detail will determine a particular tumor's potential for metastasis. In addition, we can use these biological differences to identify novel molecular markers that may be useful for diagnostic, prognostic, or predictive purposes, the success of which would pave the road to a new era of personalized medicine in kidney cancer (46, 47).

In this study, we performed quantitative proteomic profiling to identify differential protein expression between metastatic and primary RCC and to identify potential prognostic markers for RCC patients. We identified 29 proteins that were significantly differentially expressed in metastatic compared to primary RCC. Interestingly, all 29 proteins had previously been reported to be involved in tumor progression and metastasis (**Table 1 and Supplementary Table 6**). For example, two proteins that we identified to be dysregulated in our study, Pfn1 and agmatinase, were also reported to have increased (48) or decreased (49, 50) expression in RCC. In addition, many of the proteins we reported had been shown to be involved in cellular migration and invasion. For example, phosphatidylethanolamine-binding protein 1/Raf kinase

inhibitor protein (RKIP) which was downregulated in our study, was shown to inhibit the migration and invasive ability of prostate cancer cells through the extracellular matrix (51). Gal-3, which had increased expression in our study, was reported to facilitate cell migration and invasion *in vitro* and induce metastasis *in vivo* (52, 53). Furthermore, we identified a subgroup (5/12) of overexpressed enzymes that are involved in glycolysis. These include fructose-bisphosphate aldolase A, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate dehydrogenase, pyruvate kinase M1/M2, and triosephosphate isomerase. Interestingly, proteins involved in the glycolytic pathway have also been reported to be involved in the metastasis of other cancers such as melanoma (54, 55).

Hierarchical cluster analysis based on the quantified proteins resulted in patients clustered into two distinct groups. One group contained all the patients with metastatic RCC and three of the six primary RCCs, while the other group had the remaining three primaries. A review of the clinical conditions for the three primaries in the metastatic group showed that two of these patients had history of cancers and one had a subsequent RCC recurrence, suggesting that the protein expression profile of primary tumors may correlate with tumor aggressiveness. These data agree with those of Ramaswamy et al. (56) who hypothesized that the metastatic potential of tumors is encoded in the bulk of the primary tumor. This implies that there exists a biological difference between primary tumors that will metastasize and those that will remain indolent. The clinical implication is that exploitation of the protein expression profile may allow prediction of which patients will likely develop metastasis. This will encourage more intense follow up of these patients and likely leading to earlier detection of any new tumors, the result of

which is a more aggressive course of treatment and an overall increase of survival. Dissimilar tumor biology between more and less aggressive tumors may be the underlying factor for different response to treatment. This can be advantageous, as clinicians will have the ability to administer the optimal treatment for patients rather than alternative treatments that will not be effective and may cause negative side effects. A better understanding of the tumor biology will facilitate the realization of personalized medicine for RCC patients.

Pathway analysis predicted that a number of the identified dysregulated proteins are involved in similar biological signaling pathways **(Figure 6)**, suggesting that their dysregulation may be a cooperative effect imposed by the malignancy. For example, Gal-1 interacts with the α1β1 integrin subunit inducing the phosphorylation of focal adhesion kinase, which modulates cell migration (57). Binding of Gal-1 to integrin is involved in cell adhesion (37, 58). Gal-1 has also been found to be involved in cell motility (35, 37) and cellular invasion (40). 14-3-3ζ is an isoform in a family of evolutionally highly conserved acidic proteins expressed in all eukaryotic organisms (59) and has been shown to be involved in tumor progression and metastasis. 14-3-3ζ, in cooperation with ErbB2, was found to drive breast cancer metastasis (36). In addition, Pfn1, an actin-binding protein, plays a critical role in cell migration by regulating the actin-cytoskeleton pathway (38, 39). Pfn1 remodels the actin cytoskeleton by regulating actin polymerization via regeneration of actin monomers from disassembling filament networks through G-actin. Interestingly, our study showed that G-actin was overexpressed in all metastatic samples, which is in agreement with this mechanism (details shown in **Supplementary Table 5**).

An interesting point with respect to Pfn1 expression is similar to what Minamida et al (48) reported, we found Pfn1 overexpression in RCC tissues and cell lines. By contrast, Pfn1 has been found to be underexpressed in most, if not all, aggressive adenocarcinomas. For example, Pfn1 was reported to be downregulated in human breast cancer tissue and cell lines (60), pancreatic (60) and hepatic (61) carcinoma cells, squamous cell carcinoma (62), and nasopharyngeal cancer cell lines (63). These differences suggest that Pfn1 may be involved in different tumorigenic mechanisms in different tissue types.

In order to assess the potential prognostic value of our dysregulated proteins, we examined the expressions of Pfn1, Gal-1 and 14-3-3ζ on a small set of RCC patients who had good and poor prognosis. These proteins were chosen based on their increased expression in metastatic vs. primary tumors and their previous reported involvement as cancer markers (62-64). Our analysis indicated that of the three proteins Pfn1 has the most promise as a prognostic marker. Patients who had poor prognosis had high Pfn1 expression ("A" level of expression in 70% of the tumors) while only 1% of patients who showed good prognosis had this level of expression. Success in this preliminary study is limited due to the small number of cases examined, but the encouraging results certainly warrant follow up and validation on a larger cohort of patients.

In short, through quantitative proteomic analysis, we identified differential protein expressions that can distinguish between aggressive and non-aggressive RCC tumors. Many of these proteins are involved in biological pathways pertinent to tumor progression and metastasis. In addition, our preliminary analysis showed that some of

these dysregulated proteins may be useful clinical markers. Validation of these markers would greatly improve RCC patient treatment and increase overall survival.

Acknowledgements

This work was supported by grants from the Canadian Cancer Society (CCS grant # 20185), the Ministry of Research and Innovation of the Government of Ontario, the Kidney Foundation of Canada, and the Cancer Research Society.

Reference List

1. Hollingsworth, J. M., Miller, D. C., Daignault, S., and Hollenbeck, B. K. (2006) Rising incidence of small renal masses: a need to reassess treatment effect. *J. Natl. Cancer Inst.* 98, 1331-1334

2. Cheville, J. C., Lohse, C. M., Zincke, H., Weaver, A. L., and Blute, M. L. (2003) Comparisons of outcome and prognostic features among histologic subtypes of renal cell carcinoma. *Am. J. Surg. Pathol.* 27, 612-624

3. Russo, P. (2001) Renal cell carcinoma : clinical features and management. *Methods Mol. Med.* 53, 3-33

4. Lam, J. S., Shvarts, O., Leppert, J. T., Pantuck, A. J., Figlin, R. A., and Belldegrun, A. S. (2005) Postoperative surveillance protocol for patients with localized and locally advanced renal cell carcinoma based on a validated prognostic nomogram and risk group stratification system. *J. Urol.* 174, 466-472

5. Hofmann, H. S., Neef, H., Krohe, K., Andreev, P., and Silber, R. E. (2005) Prognostic factors and survival after pulmonary resection of metastatic renal cell carcinoma. *Eur. Urol.* 48, 77-81

6. Motzer, R. J., Mazumdar, M., Bacik, J., Berg, W., Amsterdam, A., and Ferrara, J. (1999) Survival and prognostic stratification of 670 patients with advanced renal cell carcinoma. *J. Clin. Oncol.* 17, 2530-2540

7. Tsui, K. H., Shvarts, O., Smith, R. B., Figlin, R. A., deKernion, J. B., and Belldegrun, A. (2000) Prognostic indicators for renal cell carcinoma: a multivariate analysis of 643 patients using the revised 1997 TNM staging criteria. *J. Urol.* 163, 1090-1095

8. Johnson, T. V., Young, A. N., Force, S., and Master, V. A. (2011) C-reactive protein may represent sensitive measure of renal cell carcinoma metastasis. *Urol. Nurs.* 31, 181-2, 194

9. Dubinski, W., Gabril, M., Iakovlev, V. V., Scorilas, A., Youssef, Y. M., Faragalla, H., Kovacs, K., Rotondo, F., Metias, S., Arsanious, A., Plotkin, A., Girgis, A. H., Streutker, C. J., and Yousef, G. M. (2011) Assessment of the prognostic significance of endoglin (CD105) in clear cell renal cell carcinoma using automated image analysis. *Hum. Pathol.*

10. Iakovlev, V. V., Gabril, M., Dubinski, W., Scorilas, A., Youssef, Y. M., Faragalla, H., Kovacs, K., Rotondo, F., Metias, S., Arsanious, A., Plotkin, A., Girgis, A. H., Streutker, C. J., and Yousef, G. M. (2012) Microvascular density as an independent predictor of clinical outcome in renal cell carcinoma: an automated image analysis study. *Lab Invest* 92, 46-56

11. Wuttig, D., Zastrow, S., Fussel, S., Toma, M. I., Meinhardt, M., Kalman, K., Junker, K., Sanjmyatav, J., Boll, K., Hackermuller, J., Rolle, A., Grimm, M. O., and Wirth, M. P. (2011) CD31, EDNRB and TSPAN7 are promising prognostic markers in clear-cell renal cell carcinoma revealed

by genome-wide expression analyses of primary tumors and metastases. *Int. J. Cancer*

12. Hoffmann, N. E., Sheinin, Y., Lohse, C. M., Parker, A. S., Leibovich, B. C., Jiang, Z., and Kwon, E. D. (2008) External validation of IMP3 expression as an independent prognostic marker for metastatic progression and death for patients with clear cell renal cell carcinoma. *Cancer* 112, 1471-1479

13. Tostain, J., Li, G., Gentil-Perret, A., and Gigante, M. (2010) Carbonic anhydrase 9 in clear cell renal cell carcinoma: a marker for diagnosis, prognosis and treatment. *Eur. J. Cancer* 46, 3141-3148

14. Schultz, L., Chaux, A., Albadine, R., Hicks, J., Kim, J. J., De Marzo, A. M., Allaf, M. E., Carducci, M. A., Rodriguez, R., Hammers, H. J., Argani, P., Reuter, V. E., and Netto, G. J. (2011) Immunoexpression status and prognostic value of mTOR and hypoxia-induced pathway members in primary and metastatic clear cell renal cell carcinomas. *Am. J. Surg. Pathol.* 35, 1549-1556

15. Arsanious, A., Bjarnason, G. A., and Yousef, G. M. (2009) From bench to bedside: current and future applications of molecular profiling in renal cell carcinoma. *Mol. Cancer* 8, 20

16. Romaschin, A. D., Youssef, Y., Chow, T. F., Siu, K. W., DeSouza, L. V., Honey, R. J., Stewart, R., Pace, K. T., and Yousef, G. M. (2009) Exploring the

pathogenesis of renal cell carcinoma: pathway and bioinformatics analysis of dysregulated genes and proteins. *Biol. Chem.* 390, 125-135

17. Shi, T., Dong, F., Liou, L. S., Duan, Z. H., Novick, A. C., and DiDonato, J. A. (2004) Differential protein profiling in renal-cell carcinoma. *Mol. Carcinog.* 40, 47-61

18. Siu, K. W., DeSouza, L. V., Scorilas, A., Romaschin, A. D., Honey, R. J., Stewart, R., Pace, K., Youssef, Y., Chow, T. F., and Yousef, G. M. (2009) Differential protein expressions in renal cell carcinoma: new biomarker discovery by mass spectrometry. *J. Proteome. Res.* 8, 3797-3807

19. Khella, H. W., White, N. M., Faragalla, H., Gabril, M., Boazak, M., Dorian, D., Khalil, B., Antonios, H., Bao, T. T., Pasic, M. D., Honey, R. J., Stewart, R., Pace, K. T., Bjarnason, G. A., Jewett, M. A., and Yousef, G. M. (2012) Exploring the role of miRNAs in renal cell carcinoma progression and metastasis through bioinformatic and experimental analyses. *Tumour. Biol.* 33, 131-140

20. White, N. M., Khella, H. W., Grigull, J., Adzovic, S., Youssef, Y. M., Honey, R. J., Stewart, R., Pace, K. T., Bjarnason, G. A., Jewett, M. A., Evans, A. J., Gabril, M., and Yousef, G. M. (2011) miRNA profiling in metastatic renal cell carcinoma reveals a tumour-suppressor effect for miR-215. *Br. J. Cancer* 105, 1741-1749

21. White, N. M., and Yousef, G. M. (2011) Translating molecular signatures of renal cell carcinoma into clinical practice. *J. Urol.* 186, 9-11

22. Pavlovich, C. P., and Schmidt, L. S. (2004) Searching for the hereditary causes of renal-cell carcinoma. *Nat. Rev. Cancer* 4, 381-393

23. Sarto, C., Marocchi, A., Sanchez, J. C., Giannone, D., Frutiger, S., Golaz, O., Wilkins, M. R., Doro, G., Cappellano, F., Hughes, G., Hochstrasser, D. F., and Mocarelli, P. (1997) Renal cell carcinoma and normal kidney protein expression. *Electrophoresis* 18, 599-604

24. DeSouza, L. V., Grigull, J., Ghanny, S., Dube, V., Romaschin, A. D., Colgan, T. J., and Siu, K. W. (2007) Endometrial carcinoma biomarker discovery and verification using differentially tagged clinical samples with multidimensional liquid chromatography and tandem mass spectrometry. *Mol. Cell Proteomics.* 6, 1170-1182

25. DeSouza, L., Diehl, G., Rodrigues, M. J., Guo, J., Romaschin, A. D., Colgan, T. J., and Siu, K. W. (2005) Search for cancer markers from endometrial tissues using differentially labeled tags iTRAQ and cICAT with multidimensional liquid chromatography and tandem mass spectrometry. *J. Proteome. Res.* 4, 377-386

26. Ralhan, R., DeSouza, L. V., Matta, A., Chandra, T. S., Ghanny, S., Datta, G. S., Bahadur, S., and Siu, K. W. (2008) Discovery and verification of head-and-neck cancer biomarkers by differential protein expression analysis

using iTRAQ labeling, multidimensional liquid chromatography, and tandem mass spectrometry. *Mol. Cell Proteomics.* 7, 1162-1173

27. Ralhan, R., Masui, O., DeSouza, L. V., Matta, A., Macha, M., and Siu, K. W. (2011) Identification of proteins secreted by head and neck cancer cell lines using LC-MS/MS: Strategy for discovery of candidate serological biomarkers. *Proteomics.* 11, 2363-2376

28. Shilov, I. V., Seymour, S. L., Patel, A. A., Loboda, A., Tang, W. H., Keating, S. P., Hunter, C. L., Nuwaysir, L. M., and Schaeffer, D. A. (2007) The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol. Cell Proteomics.* 6, 1638-1655

29. Tang, W. H., Shilov, I. V., and Seymour, S. L. (2008) Nonlinear fitting method for determining local false discovery rates from decoy database searches. *J. Proteome. Res.* 7, 3661-3667

30. Voisin, S. N., Krakovska, O., Matta, A., DeSouza, L. V., Romaschin, A. D., Colgan, T. J., and Siu, K. W. (2011) Identification of novel molecular targets for endometrial cancer using a drill-down LC-MS/MS approach with iTRAQ. *PLoS. One.* 6, e16352

31. Wang, N., and Li, L. (2008) Exploring the precursor ion exclusion feature of liquid chromatography-electrospray ionization quadrupole time-of-flight

mass spectrometry for improving protein identification in shotgun proteome analysis. *Anal. Chem.* 80, 4696-4710

32. Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A* 95, 14863-14868

33. Ito, K., and Ralph, S. J. (2012) Inhibiting galectin-1 reduces murine lung metastasis with increased CD4(+) and CD8 (+) T cells and reduced cancer cell adherence. *Clin. Exp. Metastasis*

34. Kim, H. J., Jeon, H. K., Cho, Y. J., Park, Y. A., Choi, J. J., Do, I. G., Song, S. Y., Lee, Y. Y., Choi, C. H., Kim, T. J., Bae, D. S., Lee, J. W., and Kim, B. G. (2012) High galectin-1 expression correlates with poor prognosis and is involved in epithelial ovarian cancer proliferation and invasion. *Eur. J. Cancer*

35. Camby, I., Belot, N., Lefranc, F., Sadeghi, N., de, L. Y., Kaltner, H., Musette, S., Darro, F., Danguy, A., Salmon, I., Gabius, H. J., and Kiss, R. (2002) Galectin-1 modulates human glioblastoma cell migration into the brain through modifications to the actin cytoskeleton and levels of expression of small GTPases. *J. Neuropathol. Exp. Neurol.* 61, 585-596

36. Lu, J., Guo, H., Treekitkarnmongkol, W., Li, P., Zhang, J., Shi, B., Ling, C., Zhou, X., Chen, T., Chiao, P. J., Feng, X., Seewaldt, V. L., Muller, W. J., Sahin, A., Hung, M. C., and Yu, D. (2009) 14-3-3zeta Cooperates with

ErbB2 to promote ductal carcinoma in situ progression to invasive breast cancer by inducing epithelial-mesenchymal transition. *Cancer Cell* 16, 195-207

37. Moiseeva, E. P., Spring, E. L., Baron, J. H., and de Bono, D. P. (1999) Galectin 1 modulates attachment, spreading and migration of cultured vascular smooth muscle cells via interactions with cellular receptors and components of extracellular matrix. *J. Vasc. Res.* 36, 47-58

38. Pollard, T. D., and Borisy, G. G. (2003) Cellular motility driven by assembly and disassembly of actin filaments. *Cell* 112, 453-465

39. Witke, W. (2004) The role of profilin complexes in cell motility and other cellular processes. *Trends Cell Biol.* 14, 461-469

40. Wu, M. H., Hong, T. M., Cheng, H. W., Pan, S. H., Liang, Y. R., Hong, H. C., Chiang, W. F., Wong, T. Y., Shieh, D. B., Shiau, A. L., Jin, Y. T., and Chen, Y. L. (2009) Galectin-1-mediated tumor invasion and metastasis, up-regulated matrix metalloproteinase expression, and reorganized actin cytoskeletons. *Mol. Cancer Res.* 7, 311-318

41. Zou, L., Hazan, R., and Roy, P. (2009) Profilin-1 overexpression restores adherens junctions in MDA-MB-231 breast cancer cells in R-cadherin-dependent manner. *Cell Motil. Cytoskeleton* 66, 1048-1056

42. Noma, N., Simizu, S., Kambayashi, Y., Kabe, Y., Suematsu, M., and Umezawa, K. (2012) Involvement of NF-kappaB-mediated expression of galectin-3-

binding protein in TNF-alpha-induced breast cancer cell adhesion. *Oncol. Rep.* 27, 2080-2084

43. Fan, C., Fu, Z., Su, Q., Angelini, D. J., Van, E. J., and Johns, R. A. (2011) S100A11 mediates hypoxia-induced mitogenic factor (HIMF)-induced smooth muscle cell migration, vesicular exocytosis, and nuclear activation. *Mol. Cell Proteomics.* 10, M110

44. Keshamouni, V. G., Michailidis, G., Grasso, C. S., Anthwal, S., Strahler, J. R., Walker, A., Arenberg, D. A., Reddy, R. C., Akulapalli, S., Thannickal, V. J., Standiford, T. J., Andrews, P. C., and Omenn, G. S. (2006) Differential protein expression profiling by iTRAQ-2DLC-MS/MS of lung cancer cells undergoing epithelial-mesenchymal transition reveals a migratory/invasive phenotype. *J. Proteome. Res.* 5, 1143-1154

45. Cheng, C. W., Wang, H. W., Chang, C. W., Chu, H. W., Chen, C. Y., Yu, J. C., Chao, J. I., Liu, H. F., Ding, S. L., and Shen, C. Y. (2012) MicroRNA-30a inhibits cell migration and invasion by downregulating vimentin expression and is a potential prognostic marker in breast cancer. *Breast Cancer Res. Treat.*

46. Diamandis, M., White, N. M., and Yousef, G. M. (2010) Personalized medicine: marking a new epoch in cancer patient management. *Mol. Cancer Res.* 8, 1175-1187

47. Yousef, G. M. (2012) Personalized cancer genomics: the road map to clinical implementation. *Clin. Chem.* 58, 661-663

48. Minamida, S., Iwamura, M., Kodera, Y., Kawashima, Y., Ikeda, M., Okusa, H., Fujita, T., Maeda, T., and Baba, S. (2011) Profilin 1 overexpression in renal cell carcinoma. *Int. J. Urol.* 18, 63-71

49. Dallmann, K., Junker, H., Balabanov, S., Zimmermann, U., Giebel, J., and Walther, R. (2004) Human agmatinase is diminished in the clear cell type of renal cell carcinoma. *Int. J. Cancer* 108, 342-347

50. Hwa, J. S., Park, H. J., Jung, J. H., Kam, S. C., Park, H. C., Kim, C. W., Kang, K. R., Hyun, J. S., and Chung, K. H. (2005) Identification of proteins differentially expressed in the conventional renal cell carcinoma by proteomic analysis. *J. Korean Med. Sci.* 20, 450-455

51. Xinzhou, H., Ning, Y., Ou, W., Xiaodan, L., Fumin, Y., Huitu, L., and Wei, Z. (2011) RKIp inhibits the migration and invasion of human prostate cancer PC-3M cells through regulation of extracellular matrix. *Mol. Biol. (Mosk)* 45, 1004-1011

52. Yang, L. P., Jiang, S., Liu, J. Q., Miao, X. Y., and Yang, Z. L. (2012) Association of Immunostaining of Galectin-3 and Sambucus nigra Agglutinin with Invasion, Metastasis and Poor Progression of Gallbladder Adenocarcinoma. *Hepatogastroenterology* 59

53. Wang, Y. G., Kim, S. J., Baek, J. H., Lee, H. W., Jeong, S. Y., and Chun, K. H. (2012) Galectin-3 Increases the Motility of Mouse Melanoma Cells by Regulating MMP-1 Expression. *Exp. Mol. Med.*

54. Gillies, R. J., Robey, I., and Gatenby, R. A. (2008) Causes and consequences of increased glucose metabolism of cancers. *J. Nucl. Med.* 49 Suppl 2, 24S-42S

55. Huang, S. K., Darfler, M. M., Nicholl, M. B., You, J., Bemis, K. G., Tegeler, T. J., Wang, M., Wery, J. P., Chong, K. K., Nguyen, L., Scolyer, R. A., and Hoon, D. S. (2009) LC/MS-based quantitative proteomic analysis of paraffin-embedded archival melanomas reveals potential proteomic biomarkers associated with metastasis. *PLoS. One.* 4, e4430

56. Ramaswamy, S., Ross, K. N., Lander, E. S., and Golub, T. R. (2003) A molecular signature of metastasis in primary solid tumors. *Nat. Genet.* 33, 49-54

57. Moiseeva, E. P., Williams, B., Goodall, A. H., and Samani, N. J. (2003) Galectin-1 interacts with beta-1 subunit of integrin. *Biochem. Biophys. Res. Commun.* 310, 1010-1016

58. Camby, I., Decaestecker, C., Lefranc, F., Kaltner, H., Gabius, H. J., and Kiss, R. (2005) Galectin-1 knocking down in human U87 glioblastoma cells alters their gene expression pattern. *Biochem. Biophys. Res. Commun.* 335, 27-35

59. Aitken, A. (2006) 14-3-3 proteins: a historic overview. *Semin. Cancer Biol.* 16, 162-172

60. Gronborg, M., Kristiansen, T. Z., Iwahori, A., Chang, R., Reddy, R., Sato, N., Molina, H., Jensen, O. N., Hruban, R. H., Goggins, M. G., Maitra, A., and Pandey, A. (2006) Biomarker discovery from pancreatic cancer secretome using a differential proteomic approach. *Mol. Cell Proteomics.* 5, 157-171

61. Wu, N., Zhang, W., Yang, Y., Liang, Y. L., Wang, L. Y., Jin, J. W., Cai, X. M., and Zha, X. L. (2006) Profilin 1 obtained by proteomic analysis in all-trans retinoic acid-treated hepatocarcinoma cell lines is involved in inhibition of cell proliferation and migration. *Proteomics.* 6, 6095-6106

62. Ma, C. Y., Zhang, C. P., Zhong, L. P., Pan, H. Y., Chen, W. T., Wang, L. Z., Andrew, O. W., Ji, T., and Han, W. (2011) Decreased expression of profilin 2 in oral squamous cell carcinoma and its clinicopathological implications. *Oncol. Rep.* 26, 813-823

63. Chan, C. M., Wong, S. C., Lam, M. Y., Hui, E. P., Chan, J. K., Lo, E. S., Cheuk, W., Wong, M. C., Tsao, S. W., and Chan, A. T. (2008) Proteomic comparison of nasopharyngeal cancer cell lines C666-1 and NP69 identifies down-regulation of annexin II and beta2-tubulin for nasopharyngeal carcinoma. *Arch. Pathol. Lab Med.* 132, 675-683

64. Zoidakis, J., Makridakis, M., Zerefos, P. G., Bitsika, V., Esteban, S., Frantzi, M., Stravodimos, K., Anagnou, N. P., Roubelakis, M. G., Sanchez-Carbayo,

M., and Vlahou, A. (2012) Profilin 1 is a Potential Biomarker for Bladder Cancer Aggressiveness. *Mol. Cell Proteomics.* 11, M111

65. Kuo, P. L., Huang, M. S., Cheng, D. E., Hung, J. Y., Yang, C. J., and Chou, S. H. (2012) Lung Cancer-derived Galectin-1 Enhances Tumorigenic Potentiation of Tumor-associated Dendritic Cells by Expressing Heparin-binding EGF-like Growth Factor. *J. Biol. Chem.* 287, 9753-9764

66. Huang, L., Zheng, M., Zhou, Q. M., Zhang, M. Y., Jia, W. H., Yun, J. P., and Wang, H. Y. (2011) Identification of a gene-expression signature for predicting lymph node metastasis in patients with early stage cervical carcinoma. *Cancer* 117, 3363-3373

67. Li, Y., Zou, L., Li, Q., Haibe-Kains, B., Tian, R., Li, Y., Desmedt, C., Sotiriou, C., Szallasi, Z., Iglehart, J. D., Richardson, A. L., and Wang, Z. C. (2010) Amplification of LAPTM4B and YWHAZ contributes to chemotherapy resistance and recurrence of breast cancer. *Nat. Med.* 16, 214-218

68. Lin, C. C., Chen, L. C., Tseng, V. S., Yan, J. J., Lai, W. W., Su, W. P., Lin, C. H., Huang, C. Y., and Su, W. C. (2011) Malignant pleural effusion cells show aberrant glucose metabolism gene expression. *Eur. Respir. J.* 37, 1453-1465

69. Hamaguchi, T., Iizuka, N., Tsunedomi, R., Hamamoto, Y., Miyamoto, T., Iida, M., Tokuhisa, Y., Sakamoto, K., Takashima, M., Tamesa, T., and Oka, M. (2008) Glycolysis module activated by hypoxia-inducible factor 1alpha is

related to the aggressive phenotype of hepatocellular carcinoma. *Int. J. Oncol.* 33, 725-731

70. Tang, Z., Yuan, S., Hu, Y., Zhang, H., Wu, W., Zeng, Z., Yang, J., Yun, J., Xu, R., and Huang, P. (2012) Over-expression of GAPDH in human colorectal carcinoma as a preferred target of 3-Bromopyruvate Propyl Ester. *J. Bioenerg. Biomembr.* 44, 117-125

71. Tsutsumi, S., Fukasawa, T., Yamauchi, H., Kato, T., Kigure, W., Morita, H., Asao, T., and Kuwano, H. (2009) Phosphoglucose isomerase enhances colorectal cancer metastasis. *Int. J. Oncol.* 35, 1117-1121

72. Funasaka, T., Hogan, V., and Raz, A. (2009) Phosphoglucose isomerase/autocrine motility factor mediates epithelial and mesenchymal phenotype conversions in breast cancer. *Cancer Res.* 69, 5349-5356

73. Colas, E., Perez, C., Cabrera, S., Pedrola, N., Monge, M., Castellvi, J., Eyzaguirre, F., Gregorio, J., Ruiz, A., Llaurado, M., Rigau, M., Garcia, M., Ertekin, T., Montes, M., Lopez-Lopez, R., Carreras, R., Xercavins, J., Ortega, A., Maes, T., Rosell, E., Doll, A., Abal, M., Reventos, J., and Gil-Moreno, A. (2011) Molecular markers of endometrial carcinoma detected in uterine aspirates. *Int. J. Cancer* 129, 2435-2444

74. Lovat, P. E., Corazzari, M., Armstrong, J. L., Martin, S., Pagliarini, V., Hill, D., Brown, A. M., Piacentini, M., Birch-Machin, M. A., and Redfern, C. P. (2008) Increasing melanoma cell death using inhibitors of protein

disulfide isomerases to abrogate survival responses to endoplasmic reticulum stress. *Cancer Res.* 68, 5363-5369

75. Ashrafian, H., O'Flaherty, L., Adam, J., Steeples, V., Chung, Y. L., East, P., Vanharanta, S., Lehtonen, H., Nye, E., Hatipoglu, E., Miranda, M., Howarth, K., Shukla, D., Troy, H., Griffiths, J., Spencer-Dene, B., Yusuf, M., Volpi, E., Maxwell, P. H., Stamp, G., Poulsom, R., Pugh, C. W., Costa, B., Bardella, C., Di Renzo, M. F., Kotlikoff, M. I., Launonen, V., Aaltonen, L., El-Bahrawy, M., Tomlinson, I., and Pollard, P. J. (2010) Expression profiling in progressive stages of fumarate-hydratase deficiency: the contribution of metabolic changes to tumorigenesis. *Cancer Res.* 70, 9153-9165

76. Hao, J., Wang, K., Yue, Y., Tian, T., Xu, A., Hao, J., Xiao, X., and He, D. (2012) Selective expression of S100A11 in lung cancer and its role in regulating proliferation of adenocarcinomas cells. *Mol. Cell Biochem.* 359, 323-332

77. Meding, S., Balluff, B., Elsner, M., Schone, C., Rauser, S., Nitsche, U., Maak, M., Schafer, A., Hauck, S. M., Ueffing, M., Langer, R., Hofler, H., Friess, H., Rosenberg, R., and Walch, A. (2012) Tissue Based Proteomics Reveals FXYD3, S100A11 and GSTM3 as Novel Markers for Regional Lymph Node Metastasis in Colon Cancer. *J. Pathol.*

78. Thongwatchara, P., Promwikorn, W., Srisomsap, C., Chokchaichamnankit, D., Boonyaphiphat, P., and Thongsuksai, P. (2011) Differential protein

expression in primary breast cancer and matched axillary node metastasis. *Oncol. Rep.* 26, 185-191

79. Wang, J. W., Peng, S. Y., Li, J. T., Wang, Y., Zhang, Z. P., Cheng, Y., Cheng, D. Q., Weng, W. H., Wu, X. S., Fei, X. Z., Quan, Z. W., Li, J. Y., Li, S. G., and Liu, Y. B. (2009) Identification of metastasis-associated proteins involved in gallbladder carcinoma metastasis by proteomic analysis and functional exploration of chloride intracellular channel 1. *Cancer Lett.* 281, 71-81

80. Hu, L., Lau, S. H., Tzang, C. H., Wen, J. M., Wang, W., Xie, D., Huang, M., Wang, Y., Wu, M. C., Huang, J. F., Zeng, W. F., Sham, J. S., Yang, M., and Guan, X. Y. (2004) Association of Vimentin overexpression and hepatocellular carcinoma metastasis. *Oncogene* 23, 298-302

81. Lang, S. H., Hyde, C., Reid, I. N., Hitchcock, I. S., Hart, C. A., Bryden, A. A., Villette, J. M., Stower, M. J., and Maitland, N. J. (2002) Enhanced expression of vimentin in motile prostate cell lines and in poorly differentiated and metastatic prostate carcinoma. *Prostate* 52, 253-263

82. Cao, W., Liu, N., Tang, S., Bao, L., Shen, L., Yuan, H., Zhao, X., and Lu, H. (2008) Acetyl-Coenzyme A acyltransferase 2 attenuates the apoptotic effects of BNIP3 in two human cell lines. *Biochim. Biophys. Acta* 1780, 873-880

83. Chen, Y. R., Juan, H. F., Huang, H. C., Huang, H. H., Lee, Y. J., Liao, M. Y., Tseng, C. W., Lin, L. L., Chen, J. Y., Wang, M. J., Chen, J. H., and Chen, Y. J. (2006) Quantitative proteomic and genomic profiling reveals metastasis-related protein expression patterns in gastric cancer cells. *J. Proteome. Res.* 5, 2727-2742

84. Ma, Z., Cao, M., Liu, Y., He, Y., Wang, Y., Yang, C., Wang, W., Du, Y., Zhou, M., and Gao, F. (2010) Mitochondrial F1Fo-ATP synthase translocates to cell surface in hepatocytes and has high activity in tumor-like acidic and hypoxic environment. *Acta Biochim. Biophys. Sin. (Shanghai)* 42, 530-537

85. Xie, H. L., Li, Z. Y., Gan, R. L., Li, X. J., Zhang, Q. L., Hui, M., and Zhou, X. T. (2010) Differential gene and protein expression in primary gastric carcinomas and their lymph node metastases as revealed by combined cDNA microarray and tissue microarray analysis. *J. Dig. Dis.* 11, 167-175

86. Hynninen, P., Parkkila, S., Huhtala, H., Pastorekova, S., Pastorek, J., Waheed, A., Sly, W. S., and Tomas, E. (2012) Carbonic anhydrase isozymes II, IX, and XII in uterine tumors. *APMIS* 120, 117-129

87. Cheng, I. K., Ching, A. K., Chan, T. C., Chan, A. W., Wong, C. K., Choy, K. W., Kwan, M., Lai, P. B., and Wong, N. (2010) Reduced CRYL1 expression in hepatocellular carcinoma confers cell growth advantages and correlates with adverse patient prognosis. *J. Pathol.* 220, 348-360

88. Cho-Vega, J. H., Vega, F., Schwartz, M. R., and Prieto, V. G. (2007) Expression of dicarbonyl/L-xylulose reductase (DCXR) in human skin and melanocytic lesions: morphological studies supporting cell adhesion function of DCXR. *J. Cutan. Pathol.* 34, 535-542

89. Sakakura, C., Takemura, M., Hagiwara, A., Shimomura, K., Miyagawa, K., Nakashima, S., Yoshikawa, T., Takagi, T., Kin, S., Nakase, Y., Fujiyama, J., Hayasizaki, Y., Okazaki, Y., and Yamagishi, H. (2004) Overexpression of dopa decarboxylase in peritoneal dissemination of gastric cancer and its potential as a novel marker for the detection of peritoneal micrometastases with real-time RT-PCR. *Br. J. Cancer* 90, 665-671

90. Kanamori, T., Nishimaki, K., Asoh, S., Ishibashi, Y., Takata, I., Kuwabara, T., Taira, K., Yamaguchi, H., Sugihara, S., Yamazaki, T., Ihara, Y., Nakano, K., Matuda, S., and Ohta, S. (2003) Truncated product of the bifunctional DLST gene involved in biogenesis of the respiratory chain. *EMBO J.* 22, 2913-2923

91. Wang, B., Hsu, S. H., Frankel, W., Ghoshal, K., and Jacob, S. T. (2012) Stat3-mediated activation of miR-23a suppresses gluconeogenesis in hepatocellular carcinoma by downregulating G6PC and PGC-1alpha. *Hepatology*

92. Chen, M., Zhang, J., Li, N., Qian, Z., Zhu, M., Li, Q., Zheng, J., Wang, X., and Shi, G. (2011) Promoter hypermethylation mediated downregulation of

FBP1 in human hepatocellular carcinoma and colon cancer. *PLoS. One.* 6, e25564

93. Hawthorn, L., Luce, J., Stein, L., and Rothschild, J. (2010) Integration of transcript expression, copy number and LOH analysis of infiltrating ductal carcinoma of the breast. *BMC. Cancer* 10, 460

94. Han, M. J., Wang, H., Beer, L. A., Tang, H. Y., Herlyn, M., and Speicher, D. W. (2010) A systems biology analysis of metastatic melanoma using in-depth three-dimensional protein profiling. *Proteomics.* 10, 4450-4462

95. Ross, C. D., Gomaa, M. A., Gillies, E., Juengel, R., and Medina, J. E. (2000) Tumor grade, microvessel density, and activities of malate dehydrogenase, lactate dehydrogenase, and hexokinase in squamous cell carcinoma. *Otolaryngol. Head Neck Surg.* 122, 195-200

96. Chaika, N. V., Yu, F., Purohit, V., Mehla, K., Lazenby, A. J., DiMaio, D., Anderson, J. M., Yeh, J. J., Johnson, K. R., Hollingsworth, M. A., and Singh, P. K. (2012) Differential expression of metabolic genes in tumor and stromal components of primary and metastatic loci in pancreatic adenocarcinoma. *PLoS. One.* 7, e32996

97. Wang, L., Liu, H. L., Li, Y., and Yuan, P. (2011) Proteomic analysis of pancreatic intraepithelial neoplasia and pancreatic carcinoma in rat models. *World J. Gastroenterol.* 17, 1434-1441

98. Yun, J., Frankenberger, C. A., Kuo, W. L., Boelens, M. C., Eves, E. M., Cheng, N., Liang, H., Li, W. H., Ishwaran, H., Minn, A. J., and Rosner, M. R. (2011) Signalling pathway for RKIP and Let-7 regulates and predicts metastatic breast cancer. *EMBO J.* 30, 4500-4514

99. Fan, H. Z., Liu, H., Zhang, C., Gao, D. M., Xue, Q., Chen, J., Sun, R. X., Liu, Y. K., and Yang, P. Y. (2009) Comparative proteomics and molecular mechanical analysis in CDA-II induced therapy of LCI-D20 hepatocellular carcinoma model. *J. Cancer Res. Clin. Oncol.* 135, 591-602

Chapter 5

Identification of proteins secreted by the cells of head and neck/oral squamous cell carcinoma (HNOSCC) as a strategy to find candidates for diagnostic serological biomarkers

In this project, we attempted to identify serological biomarkers in cultured cell lines for the diagnosis of HNOSCC. Cultured cell lines were used primarily because cancer tissue and body fluid samples were difficult to obtain and sample sizes were limited. However, there are also several advantages to using cultured cells, besides having an endless supply of sample; for example, the differentially expressed proteins identified in tissue samples may not be secretory, and therefore, not useful as serological biomarkers, whereas the growth medium of cell cultures mainly contains proteins that are secreted from the cells, often having a protein profile similar to the serum. In addition, body fluids are complex, requiring extra purification steps, and can dilute potential biomarkers, making them difficult to detect, whereas the cell culture growth medium is simpler and potential biomarkers can become relatively concentrated and easily detected. The disadvantage of cultured cell lines is that they are not grown in their native environment, and this may affect their protein profile.

LC-MS was used to identify proteins present in the growth medium of cultured HNOSCC cell lines to determine if some proteins that had been previously reported to be overexpressed in HNOSCC tissues were secreted into the extracellular space. Such

proteins would likely be found in the blood, making them potential diagnostic serological biomarkers for HNOSCC.

The results from this project were published in the journal, "Proteomics," (issue 12, pages 2363-2376 (2011). In this collaborative effort, biologists Dr. Ranju Ralhan, Dr. Ajay Matta, and Muzafar Macha cultured the HNOSCC cell-lines, collected conditioned media, and performed WB analysis on the conditioned media. Dr. Leroi V. DeSouza and I performed the trypsin digestion of proteins, LC-MS analysis, and Protein Pilot database searches. I also carried out SignalP, SecretomeP and other bioinformatics analyses.

Identification of proteins secreted by head and neck cancer cell lines using LC-MS/MS: Strategy for discovery of candidate serological biomarkers

Authors: Ranju Ralhan[1,2,3,4,5*#], Olena Masui[1*], Leroi V. DeSouza[1], Ajay Matta[1], Muzafar Macha[6] and K.W. Michael Siu[1,#]

Affiliations: [1]Department of Chemistry and Centre for Research in Mass Spectrometry, York University, 4700 Keele Street, Toronto, Ontario, Canada M3J 1P3

[2]Joseph and Mildred Sonshine Family Centre for Head and Neck Diseases, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario, Canada M5G 1X5

[3]Department of Otolaryngology-Head & Neck Surgery, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada M5G 1X5

[4]Alex & Simona Shnaider Research Laboratory in Molecular Oncology, Department of Pathology & Laboratory Medicine, Mount Sinai Hospital, Joseph & Wolf Lebovic Health Complex, 600 University Avenue, Room 6-500, Toronto, Ontario, Canada M5G 1X5.

[5]Department of Otolaryngology- Head & Neck Surgery, University of Toronto, Toronto, Ontario, Canada M5G 2N2.

[6]Department of Biochemistry, All India Institute of Medical Sciences, Ansari Nagar, New Delhi 110029, India

*** RR and OM are equal contributors to this study. # RR and KWMS are corresponding authors in this study.**

To whom correspondence should be addressed at the Shnaider Laboratory of Molecular Oncology, Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Joseph and Wolf Lebovic Health Complex, 600 University Avenue, Room 6-500, Toronto, Ontario, Canada M5G 1X5. Tel: (416) 586-4800 x 6426; Fax: (416)-586-8628, and the Centre for Research in Mass Spectrometry, Department of Chemistry, York University, 4700 Keele Street, Toronto, Ontario, Canada M3J 1P3. Tel: (416)650-8426; Fax: (416)736-5936; E-mail: kwmsiu@yorku.ca ; rralhan@mtsinai.on.ca.

Keywords: Proteomics, One-dimensional liquid chromatography/tandem mass spectrometry, secretome, head and neck cancer, biomarkers, secretory proteins

Abbreviations: 14-3-3 σ, stratifin; **CE,** collision energy; **FDR,** false discovery rate; **HNOSCC,** head and neck/oral squamous cell carcinoma; **hnRNPK,** heterogeneous nuclear ribonucleoprotein K; **IDA,** information-dependent acquisition; **IPA,** Ingenuity Pathway Analysis; **OPLs,** oral pre-malignant lesions; **PPIA,** peptidyl prolyl isomerase A/cyclophilin A.

Abstract

In search of blood-based biomarkers that would enhance the ability to diagnose head and neck/oral squamous cell carcinoma (HNOSCC) in early stages or predict its prognosis, we analyzed the HNOSCC secretome (ensemble of proteins secreted and/or shed from the tumor cells) for potential biomarkers using proteomic technologies. LC-MS/MS was used to identify proteins in the conditioned media of four HNOSCC cell lines (SCC4, HSC2, SCC38, and AMOSIII); 140 unique proteins were identified on the basis of 5% global false discovery rate, 122 of which were secretory proteins, with 29 being previously reported to be overexpressed in HNOSCC in comparison to normal head and neck tissues. Of these, five proteins including α-enolase, peptidyl prolyl isomerase A/cyclophilin A, 14-3-3 ζ, heterogeneous ribonucleoprotein K, and 14-3-3 σ were detected in the sera of HNOSCC patients by Western blot analysis. Our study provides the evidence that analysis of head and neck cancer cells' secretome is a viable strategy for identifying candidate serological biomarkers for HNOSCC. In future, these biomarkers may be useful in predicting the likelihood of transformation of oral pre-malignant lesions, prognosis of HNOSCC patients and evaluate response to therapy using minimally invasive tests.

1. Introduction

The interactions between cancer cells and the host's dynamic microenvironment play vital roles in tumor growth, invasion, and metastasis [1]. The cancer cells and the host's microenvironment secrete and shed proteins or their fragments extracellularly and into bodily fluids, including blood. These proteins and their fragments have been described to constitute the "cancer secretome" [2]. As about a quarter of all cellular proteins are secreted, many proteins relevant to carcinogenesis may be detectable in the blood or other bodily fluids [3, 4]. Sampling of bodily fluids is relatively straightforward, is minimally invasive, and can be repeated if necessary, thus providing longitudinal data over the course of disease investigation and/or treatment. In this view, analysis of proteins in serum/plasma and saliva using MS-based proteomic technologies have been extensively examined [5–7]. However, the major challenge when dealing with serum/plasma in particular is the complexity of blood, the wide dynamic range of protein concentrations, and the low concentrations of proteins that are directly relevant to cancer [8, 9]. An alternate strategy for examining the secretome, which proof-of-principle studies have been demonstrated to be effective, is to analyze tissue-proximal fluids and conditioned media of cancer cell lines for proteins that are released extracellularly [10–16]. Identification of proteases and growth factors in the cancer secretome underscores their utility in monitoring critical aspects of cancer progression, including invasion and metastasis [17, 18].

Head and neck/oral squamous cell carcinoma (HNOSCC) is the sixth most common cause of cancer deaths in the U.S.; despite improvements in therapeutic strategies, prognosis for the majority of HNOSCC patients remain bleak [19]. The quality

of life is often poor for survivors, mainly because the disease is often diagnosed in advanced stages. Moreover, lack of biomarkers that can predict progression of the disease or response to therapy further reflects our limited knowledge about head and neck carcinogenesis. The discovery of novel molecular targets for HNOSCC diagnostic, prognostic, and therapeutic applications has the potential to improve clinical strategy and outcome for this disease, especially if these targets are also relevant to oral pre-malignant lesions (OPLs), which are histopathologically dysplastic. Our earlier work, performed with tissue homogenates from HNOSCCs and OPLs using mass-tagging reagents, iTRAQ, and multidimensional LC-MS, resulted in identification of a panel of candidate biomarkers for OPLs [20] and HNOSCCs [21]. Other groups have also reported overexpression of several proteins in head and neck cancers in comparison to normal tissues [3, 22–40], which potentially could serve as putative biomarkers. Development of blood-based HNOSCC markers will enable us to predict progression of OPLs, prognosis of HNOSCCs and evaluate response to therapy.

Herein, we report the use of LC-MS/MS for HNOSCC secretome analysis of four HNOSCC cell lines (SCC4, HSC2, SCC38, and AMOSIII). Five proteins were detected in sera of HNOSCC patients thereby supporting our hypothesis that secretome analysis of head and neck cancer cell lines holds a potential for discovery of serum-based biomarkers. Once verified, large-scale analysis of these biomarkers can be carried out using standard ELISA-based assays or the more recently highly specialized MS-based targeted approaches such as multiple reaction monitoring, wherein specific antibodies are not available or difficult to generate [41–44]. In future studies, validation of such biomarkers will lead to the development of minimally invasive assays for accurately

predicting the progression or transformation of OPLs, determining the prognosis of the HNOSCC patients, as well as determining the response to therapy.

2 Materials and methods

2.1 Cell culture and collection of conditioned media

Cell lines used in this study were established from human head and neck squamous cell carcinomas from different sites: SCC4 (tongue), SCC38 (glottic larynx), HSC2 (mouth), and AMOSIII (buccal mucosa). SCC4 and SCC38 were obtained from American Type Culture Collection; HSC2 cells were purchased from Japanese Collection of Research Bioresources (Tokyo, Japan), while AMOSIII was developed by our group from a chronic tobacco user of Indian origin and deposited in American Type Culture Collection [45]. Such a different panel of cell lines gave us an advantage to study the biomarker diversity due to ethnicity and different sites of origin. The use of cell lines from different sites of origin would also help to find common markers that are likely to be universally applied for head and neck cancers. The cells were grown in DMEM, supplemented with 10% FBS (Invitrogen, Gaithersburg, MD) and 1 unit/mL of penicillin-streptomycin (Invitrogen) in 100-mm culture dishes to about 60% confluence at 37°C in a humidified atmosphere of 5% CO2 and 95% air [46]. Thereafter, the culture media were aspirated, and the plates were rinsed four times with phosphate-buffered saline (Sigma-Aldrich, St. Louis, MO) and once with fresh serum-free culture medium. These serum-free culture media were collected as 0 h controls. The cells were then incubated in fresh serum-free culture media for 24–48 h. The cell viability was checked using trypan blue dye assay after every 24 h. The serum-free conditioned media was collected after 24 h for SCC4, SCC38 cells, after 24 and 48 h for HSC2 cells, and after 48 h for AMOSIII cells, considering >95% cell viability until these time points. These conditioned media were filtered through 0.2 μm nylon filters, snap-frozen in liquid nitrogen and stored at −80°C

until further processing. The conditioned media from a total of 30 plates per cell type were pooled for analysis.

2.2 Protein precipitation from conditioned media

The workflow and experimental design used in this study are shown in Figure 1. Proteins in the conditioned media were precipitated using 0.02% sodium deoxycholate (Sigma- Aldrich) and 10% trichloroacetic acid (Sigma-Aldrich) as described earlier [47]. Following 2 h precipitation on ice, the samples were centrifuged for 30 min at 11 000×g and washed twice with ice-cold acetone. The precipitated proteins were redissolved in 50mM ammonium bicarbonate and their concentrations determined using the Bradford assay (Bio-Rad, Hercules, CA). The protein samples were then heated at 60^0C for 1 h in the presence of 5mM DTT, cooled to room temperature, and then alkylated by incubation with 10mM iodoacetamide for 1 h in dark. Sequencing grade trypsin (Promega, Madison, WI) at 1:20 w/w in 50mM ammonium bicarbonate was then added, and the samples were incubated at 37^0C overnight [46]. The digested samples were then dried under vacuum and redissolved in 10 mL of 0.1% formic acid.

2.3 LC-MS/MS analysis

The redissolved digested samples were analyzed using nanobore LC system (LC Packings, Amsterdam, The Netherlands) and QSTAR Pulsar (Applied Biosystems/MDS SCIEX, Foster City, CA) as described earlier 21. One-microliter aliquot of the sample (~ 100 μg) was loaded onto a C18 RP precolumn (LC Packings: 300 μm×5 mm) and desalted before separation on an RP analytical column (75 μm×150 mm packed in-house

with 3-μm Kromasil C18 beads with 100 Å pores, The Nest Group). We used a non-linear binary gradient: eluant A consisting of 94.9% deionized water, 5.0% ACN, and 0.1% formic acid (pH 3); and eluant B consisting of 5.0% deionized water, 94.9% ACN, and 0.1% formic acid for the separation. Eluant A was used to load the sample onto the C18 precolumn at a flow rate of 25 μL/min. After 8 min, the C18 precolumn was switched inline with the RP analytical column; separation was performed at 200 nL/min using a 180-min binary gradient shown in Table 1.

MS data were acquired in information-dependent acquisition (IDA) mode with the Analyst QS 1.1 and Bioanalyst Extension 1.1 software (Applied Biosystems/MDS SCIEX). MS cycles comprised a TOF MS survey scan with a mass range of 400–1500 Da for 1 s, followed by five product-ion scans with a mass range of 80–2000 Da for 2 s each. The collision energy was automatically controlled by the IDA collision energy Parameters script. Switching criteria were set to ions with m/z ≥400 and ≤1500, charge states of 2–4, and abundances of ≥10 counts. Former target ions were excluded for 30 s, and ions within a 6-Da window were ignored. Additionally, the IDA Extensions II script was set to "no repetition" before dynamic exclusion and to select a precursor ion nearest to a threshold of 10 counts on every fourth cycle [20, 21]. The protein samples from SCC4, SCC38, and HSC2 cell lines were analyzed using online LC-MS/MS in duplicates while AMOSIII cells were analyzed once only.

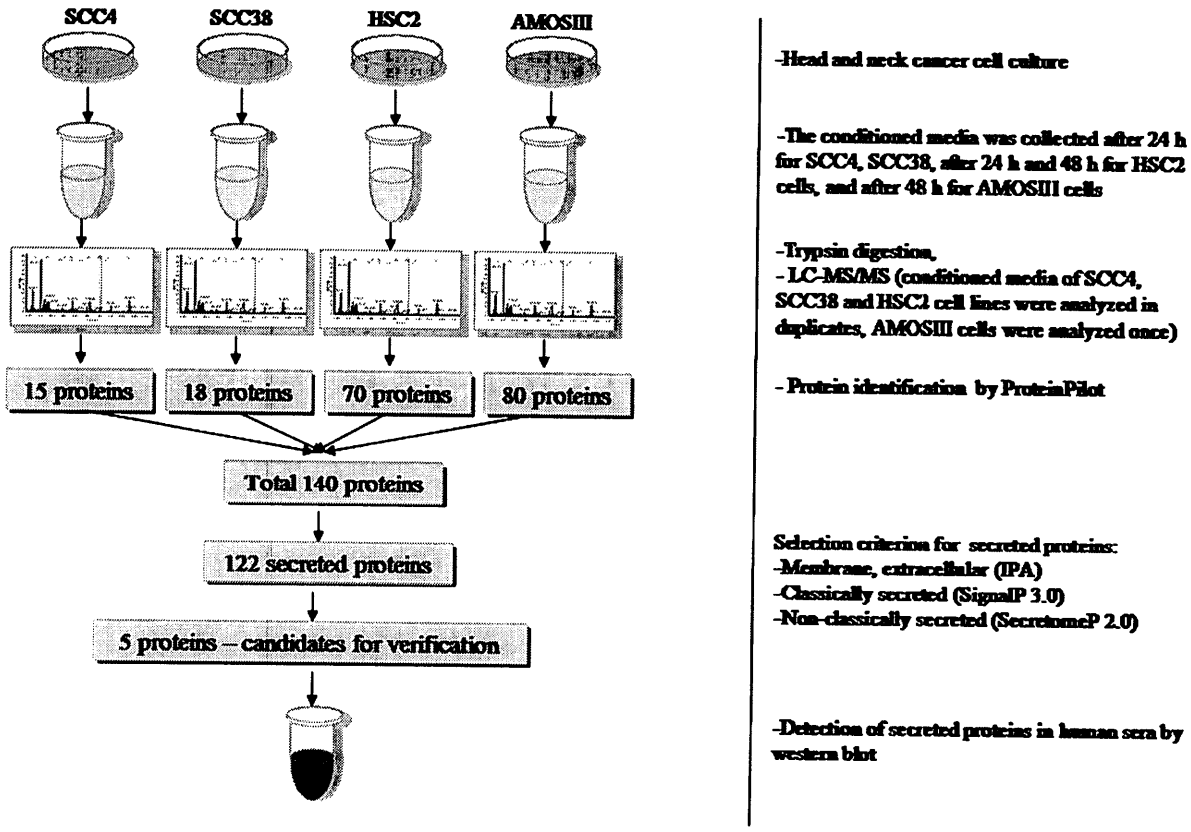


Figure 1 Outline of workflow for identification of proteins secreted by HNOSCC cell lines.

50 mM ammonium bicarbonate and their concentrations determined using the Bradford assay (Bio-Rad, Hercules, CA). The protein samples were then heated at 60°C for 1 h in the presence of 5 mM DTT, cooled to room temperature, and then alkylated by incubation with 10 mM iodoacetamide for 1 h in dark. Sequencing grade trypsin (Promega, Madison, WI) at 1:20 w/w in 50 mM ammonium bicarbonate was then added, and the samples were incubated at 37°C overnight 46. The digested samples were then dried under vacuum and redissolved in 10 µL of 0.1% formic acid.

Table 1. 180-min binary gradient used for separation

Time (min)	0	5	10	120	140	145	155	157	189
B (%)	5	5	15	35	60	80	80	5	Stop

2.4 Bioinformatics

LC-MS/MS data were searched using the ProteinPilot software 2.0.1. (Applied Biosystems) against an NCBI human protein database containing 207 920 protein sequences (version of the June 2, 2010) Protein identification was performed using a confidence threshold of 95% (ProteinPilot Unused score ≥1.31) with methyl methanethiosulfonate selected as cysteine modification, and with the search option "emphasis on biological modifications" checked. The same LC-MS/MS data was searched against a concatenated database plus decoy database created from the above NCBI human protein database containing 323 410 protein sequences. Identified proteins with global false discovery rate (FDR) less than 5% were considered for further analysis [48]. SignalP (http://www.cbs.dtu.dk/services/SignalP 3.0) was used to analyze identified proteins for classical protein secretion features [49]. SignalP predicts the presence and the location of signal peptide cleavage sites in the amino-acid sequences by a combination of artificial neural networks and hidden Markov model algorithms to detect signal peptides from input protein sequences. The presence of secretory signal peptide sequences was determined with a probability ≥0.9. SecretomeP (http://www.cbs.dtu.dk/services/ SecretomeP 2.0) was used for non-classical and leaderless protein secretion [50]. SecretomeP utilizes a neural network combining six protein features to predict whether a protein sequence undergoes non-classical secretion or not; these features include the number of atoms, number of positively charged residues, presence of transmembrane helices, presence of low-complexity regions, presence of pro-peptides, and subcellular localization. A given protein is considered non-classically secreted if it contains a non-classical (non-signal) peptide-triggered protein secretion with score ≥0.5; only proteins

that did not contain a signal peptide as determined by SignalP were legitimate candidates for this analysis. Ingenuity Pathway Analysis (IPA, Ingenuity Systems, www.ingenuity.com) was used to determine the subcellular localization and biological functions of the identified proteins (for detailed information on IPA, visit www.ingenuity.com). We classified a protein as "secreted" if it satisfied at least one of the following four criteria: (i) its subcellular location is extracellular or membrane-bound, according to IPA; (ii) it is classically secreted, according to SignalP; (iii) it is non-classically secreted, according to SecretomeP; or (iv) it is non-classically secreted by the exosome pathway, according to the literature.

2.5 Patient samples and Western blot Analysis

Serum samples were collected from patients diagnosed with HNOSCC (n=12) while control sera were obtained from healthy donors (n=12) undergoing annual checkup at the All India Institute of Medical Sciences, New Delhi, India, with approval of the All India Institute of Medical Sciences Ethics Committee after obtaining consent of the patients. Blood was collected and processed for serum isolation following the standard operating procedure recommended by the Early Detection Research Network [51]. The clinical details of the patients are given in Supporting Information Table IS). Sera samples were depleted of the 20 most-abundant proteins using the Proteoprep 20 Plasma Immunodepletion kit (Sigma-Aldrich) following the manufacturer's specifications before Western blotting. Briefly, 8 μL of sera samples were diluted in 100 μL of equilibration buffer and filtered through Corning Spin-X centrifuge tube filters. The diluted sera were loaded onto the packed bed medium and incubated for 15–20 min. The spin column was

centrifuged at 2000×g for 1 min and the flow through was collected in a fresh tube. The remaining serum proteins were washed from the spin column by adding 100 μL of equilibration buffer and centrifuged to collect the flow through in the same tube. The combined flow-through contained the serum proteins minus the targeted abundant proteins; this depleted serum sample was concentrated using speed-vac. Western blotting was carried out to detect the proteins: α-enolase, peptidyl prolyl isomerase A/cyclophilin A (PPIA), 14-3-3 ζ, heterogeneous ribonucleoprotein K (hnRNPK), and 14-3-3 σ in these sera samples as described earlier [20, 21, 52]. The criteria of selecting candidates for verification were (i) identification in our earlier iTRAQ studies; (ii) biological relevance; and (iii) prognostic relevance established in tissues in our earlier studies [20, 21, 52]. The details of the antibodies and dilutions used are given in Supporting Information Table IIS.

3 Results

3.1 Secretome analysis of HNOSCC cell lines

The LC-MS/MS data thus obtained was searched against NCBI human protein database plus a decoy database for each cell line (see Supporting Information Table IIIS a–f). The proteins with global FDR more than 5% were excluded from further analysis. Only unique proteins were considered for further analysis. Following these criteria, we identified a total of 183 unique proteins in the secretome analysis of HNOSCC cell lines (Table 2), of which 140 proteins were non-redundant among these cell lines (see Supporting Information Tables IVS). The number of unique peptides with a confidence score of ≥80% per identified protein in conditioned media of each of the four HNOSCC cell lines are shown in Table 2. The overlap of two independent replicates (Runs 1 and 2) of proteins identified in conditioned media of SCC4, SCC38, and HSC2 are shown in Supporting Information Fig. IS. On an average 60% overlap was observed among different runs for the same cell line (Supporting Information Fig. IS). The overlap of proteins identified in conditioned media of HSC2 collected twice at 24 h and once at 48 h are shown in Supporting Information Fig. IIS. Figure 2 shows the extent of overlaps between proteins identified in the conditioned media of the four cell lines. Supporting Information Fig. IIIS shows a Coomassie stained gel showing the reproducibility of the experimental procedures followed for secretome analysis.

Table 2. The total number of proteins identified per number of unique peptides in conditioned media of four HNOSCC cell lines

No. peptides identified	SCC4	HSC2	SCC38	AMOSIII
1	1	21	2	6
2	6	19	7	22
3	4	6	1	17
4	2	5	3	10
≥5	2	19	4	25
Total	15	70	18	80

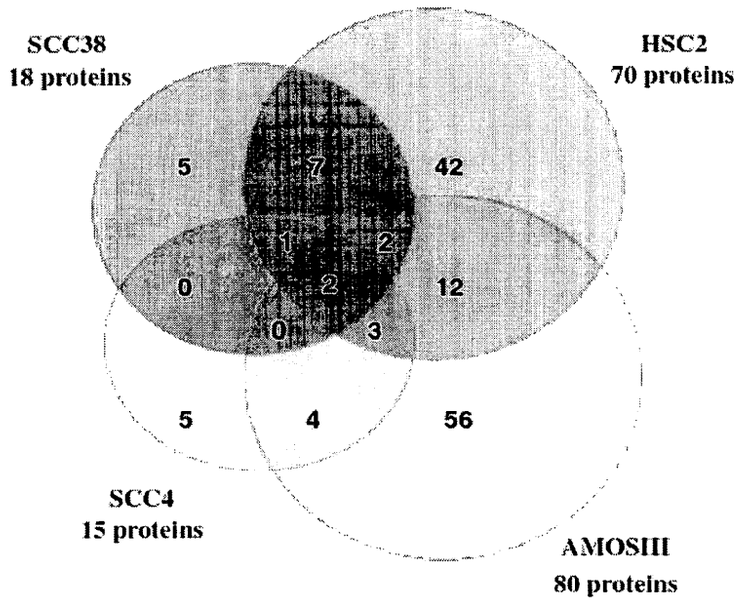


Figure 2. Overlaps of non-redundant proteins identified in the conditioned media of SCC4, HSC2, SCC38, and AMOSIII cells lines.

3.2 Bioinformatic analysis

Extracellular and membrane-bound proteins have higher probabilities of being secreted or shed, and being available in the conditioned media and probably in serum. Thus, using bioinformatics approach, we determined the subcellular localization of identified proteins using IPA software. About, 28.1% of the non-redundant proteins identified in this study were found to be extracellular, while 16.0% were membrane-bound (Fig. 3). Of the 140 proteins identified in this study, 122 proteins satisfied our criteria and were classified as secretory proteins while the remaining 18 proteins were excluded from further study (Table 3). According to the IPA anthology, 54 proteins in Table 3 were localized in the extracellular space and 21 were membrane proteins. Our analysis using SignalP bioinformatic program predicted that 75 are likely to be secreted. Another 29 proteins can be secreted according to SecretomeP analysis. Twenty-three proteins could be secreted from cells via the exosomal pathway [53–60]. Twenty-six proteins were reported to be present in either in blood, saliva, or secretome [22, 61] and therefore were included in Table 3. Cytoplasmic 14-3-3 σ was included in Table 3 because it can be found in extracellular space as well according to IPA.

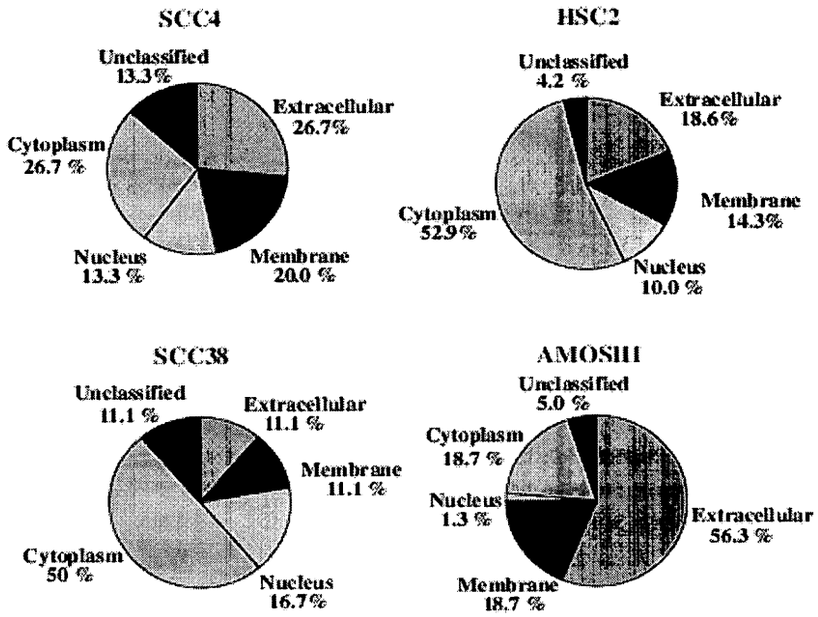


Figure 3. Subcellular locations of proteins identified in conditioned media of the SCC4, HSC2, SCC38, and AMOSIII cancer cell lines. Percentages of the proteins are given.

3.3 Western blot analysis

To determine the potential of 122 identified proteins as serum-based cancer biomarkers, we aimed to detect α-enolase, PPIA, hnRNPK, 14-3-3 ζ, and 14-3-3 σ in sera of HNOSCC patients and healthy controls using Western blot analysis. Our results of Western blotting revealed increased levels of α-enolase, PPIA, hnRNPK, 14-3-3 ζ, and 14-3-3 σ proteins in the majority of the serum samples obtained from HNOSCC patients, in comparison with the sera obtained form healthy volunteers as shown in Figs. 4A and B. A coomassie stained gel of sera samples is also shown in Supporting Information Fig. IVS demonstrating equal loading in each lane of SDS-PAGE.

Table 3. Secreted proteins identified by LC-MS/MS in HNOSCC conditioned media

#	Proteins	Accession no.	SCC4	SCC38	HSC2	AMOSIII	Protein ontology[a]	SignalP probability[b]	SecretomeP score[c]	Found in		Over-expressed in HNOSCC[e]
										Exosomes[d]	Blood/saliva/ secretome	
1	26kDa protein (aa 32-212) (*Homo sapiens*)	gil23835				*	M	0	0.759			
2	45kDa calcium-binding protein isoform 1 precursor	gil7706573				*	C	0.999				
3	Activated leukocyte cell adhesion molecule variant 1	gil94962177				*	M	0.996				
4	Activated T-cell marker CD109	gil37359236				*	M	0.998				
5	Agrin precursor	gil2988422			*	*	M	0.999	–			
6	Aldolase A (ALDOA)	gil49456715		*	*		C	0	0.356	[53]		
7	αEnolase	gil62896593			*	*	C	0	0.536	[54, 55]	[22]	[37]
8	Amyloid βA4 protein isoform e precursor	gil209915570	*		*	*	M	1				
9	Amyloid precursor protein homolog HSD-2	gil5702388			*	*	E	1				
10	B23 nucleophosmin (280 AA)	gil825671		*	*		N	0	0.811			[38]
11	β Actin variant	gil62897625	*	*	*		C	0	0.498	[56]	[61]	
12	β-2 Microglobulin	gil34616			*	*	M	1				
13	Calgizzarin (hCG2013819)	gil119603728		*		*	U	0	0.755			[37]
14	Calmodulin	gil825635		*	*	*	M	0	0.676			
15	Calreticulin precursor variant	gil62897681			*		C	1				
16	Calsyntenin 1, isoform CRA_b	gil119592036				*	M	0	0.132			
17	Caprin-1 isoform 1	gil42558250			*		M	0.112	0.099		[61]	
18	CD44 molecule (Indian blood group)	gil10432372			*		M	0.997	–			
19	Clusterin isoform 3	gil283806712			*		E	0.312	0.491			
20	Cofilin-1	gil5031635				*	N	0	0.628	[53]		[35]
21	Collagen α-1(I) chain	gil296439504			*		E	0.999				
22	Collagen, type XVII, α 1, isoform CRA_b	gil119569998			*		M	0	0.153			
23	Colony-stimulating factor 1 (macrophage), isoform CRA_b	gil119576821				*	E	0.997			[61]	
24	Complement component 1, q subcomponent-binding protein, isoform CRA_c	gil119610743			*		C	0.849	0.6			
25	Complement component C3	gil179665	*			*	E	1			[61]	
26	Connective tissue growth factor precursor	gil4503123				*	E	1				
27	C-X-C motif chemokine 2	gil4504155				*	E	1				
28	Cystatin-C precursor	gil4503107				*	E	1				
29	Dickkopf-related protein 3 precursor	gil40548389				*	E	1				
30	Dystroglycan	gil229462879				*	M	0.999				
31	EIF3S4	gil48146385			*		C	0	0.675			
32	Elongation factor 2	gil181969			*		C	0	0.38	[57]		
33	Eukaryotic translation elongation factor 1 α 1 variant	gil62897621			*	*	C	0.003	0.155	[53, 54]		[31, 38]
34	Family with sequence similarity 3, member C precursor	gil7661714				*	E	0.906				
35	Fatty acid-binding protein, epidermal	gil4557581			*		C	0	0.747			
36	Fibronectin 1, isoform CRA_j	gil119590945				*	M	0.997			[22]	[23]
37	Follistatin-related protein 1 precursor	gil5901956			*	*	E	1				
38	Glutathione *S*-transferase-P1c	gil726098			*		C	0.084	0.545	[58]	[22]	[21, 27, 32, 35, 37]
39	Glyceraldehyde-3-phosphate dehydrogenase	gil31645		*	*		C	0	0.467	[54, 55, 58]	[22, 61]	[37]
40	Granulin	gil183613				*	E	0.999			[61]	

Table 3. Secreted proteins identified by LC-MS/MS in HNOSCC conditioned media

#	Proteins	Accession no.	SCC4	SCC38	HSC2	AMOSIII	Protein ontology[a]	SignalP probability[b]	SecretomeP score[c]	Found in		Over-expressed in HNOSCC[a]
										Exosomes[d]	Blood/saliva/ secretome	
1	26kDa protein (aa 32-212) (*Homo sapiens*)	gil23835				*	M	0	0.759			
2	45kDa calcium-binding protein isoform 1 precursor	gil7706573				*	C	0.999				
3	Activated leukocyte cell adhesion molecule variant 1	gil94962177				*	M	0.996				
4	Activated T-cell marker CD109	gil37359236				*	M	0.998				
5	Agrin precursor	gil2988422			*	*	M	0.999	–			
6	Aldolase A (ALDOA)	gil49456715		*	*		C	0	0.356	[53]		
7	αEnolase	gil62896593			*	*	C	0	0.536	[54, 55]	[22]	[37]
8	Amyloid βA4 protein isoform e precursor	gil209915570	*		*	*	M	1				
9	Amyloid precursor protein homolog HSD-2	gil5702388			*	*	E	1				
10	B23 nucleophosmin (280 AA)	gil825671		*	*		N	0	0.811			[38]
11	β Actin variant	gil62897625	*	*	*		C	0	0.498	[56]	[61]	
12	β-2 Microglobulin	gil34616			*	*	M	1				
13	Calgizzarin (hCG2013819)	gil119603728		*		*	U	0	0.755			[37]
14	Calmodulin	gil825635		*	*	*	M	0	0.676			
15	Calreticulin precursor variant	gil62897681			*		C	1				
16	Calsyntenin 1, isoform CRA_b	gil119592036				*	M	0	0.132			
17	Caprin-1 isoform 1	gil42558250			*		M	0.112	0.099		[61]	
18	CD44 molecule (Indian blood group)	gil10432372			*		M	0.997	–			
19	Clusterin isoform 3	gil283806712			*		E	0.312	0.491			
20	Cofilin-1	gil5031635				*	N	0	0.628	[53]		[35]
21	Collagen α-1(I) chain	gil296439504			*		E	0.999				
22	Collagen, type XVII, α 1, isoform CRA_b	gil119569998			*		M	0	0.153			
23	Colony-stimulating factor 1 (macrophage), isoform CRA_b	gil119576821				*	E	0.997			[61]	
24	Complement component 1, q subcomponent-binding protein, isoform CRA_c	gil119610743			*		C	0.849	0.6			
25	Complement component C3	gil179665	*			*	E	1			[61]	
26	Connective tissue growth factor precursor	gil4503123				*	E	1				
27	C-X-C motif chemokine 2	gil4504155				*	E	1				
28	Cystatin-C precursor	gil4503107				*	E	1				
29	Dickkopf-related protein 3 precursor	gil40548389				*	E	1				
30	Dystroglycan	gil229462879				*	M	0.999				
31	EIF3S4	gil48146385			*		C	0	0.675			
32	Elongation factor 2	gil181969			*		C	0	0.38	[57]		
33	Eukaryotic translation elongation factor 1 α 1 variant	gil62897621			*	*	C	0.003	0.155	[53, 54]		[31, 38]
34	Family with sequence similarity 3, member C precursor	gil7661714				*	E	0.906				
35	Fatty acid-binding protein, epidermal	gil4557581			*		C	0	0.747			
36	Fibronectin 1, isoform CRA_j	gil119590945				*	M	0.997			[22]	[23]
37	Follistatin-related protein 1 precursor	gil5901956			*	*	E	1				
38	Glutathione *S*-transferase-P1c	gil726098			*		C	0.084	0.545	[58]	[22]	[21, 27, 32, 35, 37]
39	Glyceraldehyde-3-phosphate dehydrogenase	gil31645		*	*		C	0	0.467	[54, 55, 58]	[22, 61]	[37]
40	Granulin	gil183613				*	E	0.999			[61]	

Table 3. Continued

#	Proteins	Accession no.	SCC4	SCC38	HSC2	AMOSIII	Protein ontology[a]	SignalP probability[b]	SecretomeP score[c]	Found in		Over-expressed in HNOSCC[d]
										Exosomes[d]	Blood/saliva/ secretome	
41	Granulocyte colony-stimulating factor isoform b precursor	gil27437049				*	E	0.994				
42	GRP78 precursor	gil386758		*	*		C	1			[61]	[29]
43	H2AFX	gil48146275			*		N	0.116	0.535	[54]		
44	hCG1983058	gil119619436			*		U	0	0.642			
45	Heat shock protein 27	gil662841			*		C	0	0.74	[59]	[22]	[22, 31, 32, 34, 35, 37]
46	Heat shock protein HSP 90-α isoform 1	gil153792590			*	*	C	0	0.204	IPA		
47	Heat shock protein HSP 90-β	gil20149594			*		C	0	0.204	IPA		
48	Heparan sulfate proteoglycan	gil184427				*	M	1				
49	Histone H2A type 3	gil15617199	*				N	0.091	0.515			
50	Insulin-like growth factor binding protein 4	gil54696644				*	E	1				
51	Insulin-like growth factor binding protein 6	gil183894			*		E	1				
52	Insulin-like growth factor binding protein 7, isoform CRA_a	gil119625925	*		*	*	E	0.998			[61]	
53	Interleukin 8	gil33959				*	E	1				
54	Kunitz-type protease inhibitor 1 isoform 1 precursor	gil32313599				*	E	1				
55	Laminin α 3b chain	gil46020022				*	E	1				
56	Laminin B2 chain	gil186964				*	E	1				
57	Laminin S B3 chain	gil510703			*	*	E	1				[26]
58	Laminin subunit β-1 precursor	gil167614504				*	E	1				
59	Laminin, γ2	gil85566965				*	E	1				
60	Lectin, galactoside-binding, soluble, 3 binding protein, isoform CRA_a	gil119609949	*	*	*	*	M	1		[56]		
61	Lipocalin 2 (oncogene 24p3), isoform CRA_a	gil119608154	*			*	U	0	0.695			
62	Lactate dehydrogenase A chain isoform 3	gil260099723	*		*		C	0	0.549	[55]		[21, 38]
63	L-lactate dehydrogenase B chain	gil4557032			*		C	0	0.569	[58]		[38]
64	Lysyl oxidase-like 2, isoform CRA_b	gil119584030				*	E	0.999		IPA		[27]
65	Matrilysin preproprotein	gil4505219				*	E	1				
66	Matrix metallopeptidase 13 (collagenase 3)	gil45768662			*		E	1				
67	Matrix metalloproteinase 1 preproprotein variant	gil62898123	*			*	E	1				[27]
68	Melanoma-associated antigen MG50	gil6273399				*	U	0.987				
69	Mesothelin precursor	gil56406362				*	E	0.997				
70	MHC class I antigen	gil269965098				*	M	0.003	0.648			
71	Moesin, isoform CRA_b	gil119625804			*		M	0	0.53		[22]	
72	Myeloid leukemia-inhibitory factor (AA at 10)	gil187141				*	E	0.999				
73	Myristoylated alanine-rich C-kinase substrate	gil187387			*		M	0.001	0.366			[27]
74	Niemann-Pick disease, type C2 precursor variant	gil62896507				*	E	1				
75	Nuclease-sensitive element binding protein-1	gil457262			*		N	0.011	0.66			
76	Nucleobindin	gil1144316				*	C	1				
77	Pentaxin	gil35797				*	E	1				
78	Peptidylprolyl isomerase A-like, isoform CRA_c	gil119581485	*	*	*	*	C	0.001	0.339	[59]	[61]	[21]

Table 3. Continued

	Proteins	Accession no.	SCC4	SCC38	HSC2	AMOSIII	Protein ontology[a]	SignalP probability[b]	SecretomeP score[c]	Found in		Over-expressed in HNOSCC[e]
										Exosomes[d]	Blood/saliva/ secretome	
18	Triosephosphate isomerase isoform 2	gil226529917	*		*	*	C	0.013	0.39	[58]	[22]	[22, 31, 37]
19	Tripeptidyl-peptidase I precursor variant	gil62896635				*	C	0.998				
20	Ubiquitin C	gil67191208		*	*	*	C	0	0.693			
21	Vimentin	gil62414289				*	C	0.015	0.512			[32]
22	14-3-3 ζ Protein	gil30354619				*	C	0	0.252	[53]	[22]	[21]

Asterisk represents that the corresponding protein was identified in these cell lines.
a) The ontologies of the identified proteins were analyzed using IPA. C – cytoplasm, E – extracellular, M – plasma membrane, N – nucleus, U – unknown.
b) The classical secretion was predicted using the hidden Markov model of SignalP 3.0.
c) The nonclassical secretion of proteins was evaluated by the neural network output score of SecretomeP 2.0.
d) The nonclassical secretion of proteins by exosome pathway. Proteins were found in blood and saliva according to literature.
e) Proteins were found overexpressed in head and neck cancer according to literature.

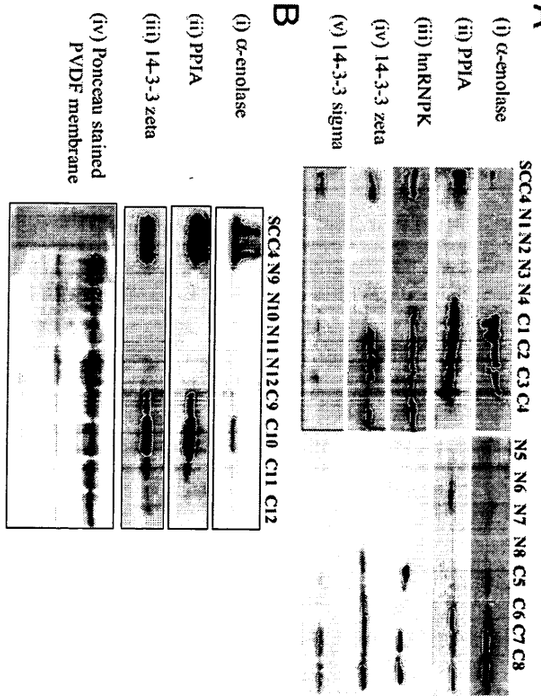


Figure 4. Detection of secretome proteins in HNOSCC patients' sera by Western blot analysis. (A) Sera from patients diagnosed with HNOSCCs (C1–C12) and normal controls (N1–N12) were depleted for abundant proteins using Proteoprep 20 Plasma Immunodepletion kit as described in Section 2. Equal amounts of depleted sera were electrophoresed on 12% SDS-PAGE. Western blot analysis was carried out using specific antibodies for the panel of proteins: α-enolase, PPIA, hnRNPK, 14-3-3 ζ and 14-3-3 σ. Western blotting showed increased expression of (i) α-enolase; (ii) PPIA; (iii) hnRNPK; (iv) 14-3-3 ζ and (v) 14-3-3 σ in sera samples of HNOSCCs (C1–C8) in comparison to sera of healthy controls (N1–N8). Whole cell lysates of SCC4 cells were used as a positive control; (B) Panel shows additional western blots of (i) α-enolase, (ii) PPIA, (iii) 14-3-3 ζ in sera from HNSCC patients (C8–C12) and normal controls (N8–N12) with whole cell lysates from SCC4 cells used as a positive control; (iv) Panel shows a ponceau-stained PVDF membrane showing equal loading of the protein.

4 Discussion

Head and neck cancer is a heterogeneous disease including different sites such as oral cavity (includes tongue, buccal mucosa, alveolus, lip, retromolar trigone, floor of mouth), pharynx, larynx, nasopharynx, and salivary glands. In addition, different risk factors and ethnic variations add to the complexity of the molecular pathogenesis of this malignancy. Thus, our study is unique in giving a comprehensive overview of the secretome analysis of four HNOSCC cell lines obtained from different sites of origin and ethnicity. Our study resulted in identification of 140 non-redundant proteins among four cell lines on the basis of 5% global FDR. Following stringent criteria and bioinformatic analysis as described in materials and methods, we successfully identified 122 secretory proteins as compared to a recent report by Weng et al., [22] which identified 37 secreted proteins only. Notably, 15 of these 37 proteins including PPIA, α-enolase, 14-3-3 ζ, pyruvate kinase M2 (PKM2), glutathione-*S*-transferase P1, profilin 1, glyceraldehyde phosphate dehyrogenase, triosephosphate isomerase 1, fibronectin 1, HSP27, neutrophil gelatinase associated lipocalin, peroxiredoxin1, transforming growth factor-β induced protein IG-H3 precursor, thrombospodin 1, and plasminogen activator inhibitor 1 were common among these studies, thereby supporting our data. Among others, 12 proteins have also been reported in earlier investigations on the oral cancer secretome [22]. Interestingly, seven of these secretory proteins have been identified in our previous reports showing overexpression of these proteins in HNOSCC tissues using iTRAQ labeling followed by multidimensional LC-MS/MS suggesting their origin from head and neck cancer cells only [21]. These proteins included 14-3-3 ζ, PPIA or cyclophilin A, glutathione-*S*-transferase P1, 14-3-3 σ, calgizzarin, prothymosin α, and lactate

dehydrogenase A. In addition, overexpression of several other proteins identified in our secretome analysis has also been reported in head and neck cancer tissues [22–39].

A major limitation of our study is the use of only a single dimension of separation (nanoLC) for the analysis of the secretome that probably accounts for the fewer number of proteins identified. Several recent reports employing two-dimensional separation have shown that the secretome contains >1000 high confident proteins that can be reliably detected in replicate analysis [4, 13, 14, 62, 63]. These studies have demonstrated the efficacy of secretome-based strategies in a variety of cancer types [62]. In these studies, proteins secreted from cancer cells into serum-free media were resolved by one- or two-dimensional gels followed by in-gel tryptic digestion and analysis by MALDI-TOF MS or LC-MS/MS. Comparison of these studies showed in general that more proteins were detected in the secretome using the LC-MS/MS method than the MALDI-TOF [62]. Moreover, recent studies have focused on comparison and analysis of various first-dimensional separation techniques prior to nanoLC MS/MS and have provided evidence that the use of 1-D gel electrophoresis in combination with LC-MS/MS yield highest total number of identified proteins as well as highest reproducibility in biological replicates in comparison to other procedures such as RP and strong cation exchange chromatography [62, 63]. In light of these recent reports, the lower number of total proteins identified in the present study may be attributed to the use of one-dimensional LC MS/MS. Taken together, these reports suggest that cancer cell secretome is too complex for a direct nanoLC-MS/MS analysis. Nevertheless, the detection of some of the proteins identified in our secretome analysis in HNOSCC patients' sera underscores the great potential of

such cell-culture based secretome analysis for discovery of serum-based protein biomarkers.

α-Enolase, a glycolytic enzyme that converts 2-phospho-D-glycerate to phosphoenol pyruvate, is a multifunctional enzyme involved in several cellular processes, including growth control, tolerance to hypoxia, and allergic responses [64]. It is also a plasminogen-binding protein involved in the promotion of plasminogen activation by leukocytes. In our study, increased expression of α-enolase has been detected in five of eight HNOSCC serum samples in comparison to sera obtained from healthy volunteers. Similarly, α-enolase has been identified as a secretory marker in the serum of breast, kidney, prostate, ovarian, pancreatic, and non-small cell lung cancer patients. Interestingly, α-enolase has been proposed as a predictive maker in lung and prostate cancer for evaluating response to treatment in [65, 66]. PPIA contributes to the maintenance of proper conformation of nascent or denatured proteins and also provides protection against environmental insults [67, 68]. Upregulation of PPIA in small cell lung cancer, pancreatic cancer, breast cancer, colorectal cancer, squamous cell carcinoma, and melanoma has been reported [68–71]. In our study, we detected PPIA in HNOSCC serum samples. However, their relevance as head and neck serum-based biomarkers for diagnosis, prognosis, or predicting the response to therapy needs to be investigated.

In our recent studies using iTRAQ analysis on tissue homogenates, we observed increased expression of 14-3-3 ζ and 14-3-3 σ isoforms in OPLs and HNOSCCs in comparison to normal tissues [20, 21, 46, 72, 73]. In this study, we observed increased levels of both these isoforms in sera of HNOSCC patients in comparison to healthy

volunteers. Taken together, these studies suggest the relevance of both these isoforms in head and neck carcinogenesis and warrant a further investigation to determine their clinical impact on evaluation of prognosis and predicting response to therapy as serum-based biomarkers. Notably, 14-3-3 ζ has been shown to be secreted by tumor-associated monocytes/macrophages from the ascites of epithelial ovarian cancer patients, and speculated to play a role in regulating inflammatory pathways of the epithelial ovarian cancer microenvironment and serve as a biomarker of tumor-associated inflammation [74]. The expression of 14-3-3 ζ has also been associated with the degree of cancer peritoneal metastasis, the emergence of ascites, bilateral involvement, and the clinical stage of ovarian cancer patients [75]. In addition, 14-3-3 ζ has been reported in serum of lymphoma-bearing SJL mice and the tear fluid. 14-3-3 ζ is known to function as an adapter protein, interacting with over 100 cellular proteins, and is involved in blocking apoptosis and promoting cellular proliferation, adhesion, and cellular movement [76]. 14-3-3 ζ has also been reported to interact with β-catenin, enhance or inhibit β-catenin-dependent transcription, facilitate activation of β-catenin through Akt, and participate in stem-cell development [13]. Combined gene-expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma has identified 14-3-3 σ to be specifically overexpressed in the tumor epithelia and correlated with poor disease prognosis [77]. 14-3-3 σ was also identified in the proteomic analysis of laryngeal SCCs [78] and oral SCCs [79]. Most importantly, tissue proteomics showed the overexpression of 14-3-3 σ in OPLs and HNOSCCs, and correlated overexpression with disease prognosis [20, 21, 46].

Another important protein identified in our secretome analysis of cell lines and verified in serum of HNOSCC patients was hnRNPK. However, we failed to include hnRNPK in Table 3, as it did not meet our criteria to establish the secretory proteins, as they were probably overly stringent. The use of 5% of global FDR for confident protein identification is the accepted procedure to minimize false-positive identifications. However, this requirement can probably be relaxed when additional information is available, e.g. from a previous analysis, to reduce the exclusion of legitimate proteins. Tissue proteomics has identified hnRNPK in oral dysplastic lesions and head and neck cancers [20]. In addition, its overexpression in early preneoplastic stages has been independently verified and correlated with poor prognosis [20, 80]. It is noteworthy that hnRNPK has also been reported to be overexpressed in nasopharyngeal carcinoma [28, 29, 81], prostate cancer [82], and other human cancers [83].

To summarise, secretome analysis of HNOSCC cell lines using optimized cell culture and LC-MS/MS resulted in the identification of 122 secreted proteins. Further, we verified the expression of α-enolase, PPIA, 14-3-3 ζ, hnRNPK, and 14-3-3 σ in serum of HNOSCC patients establishing their potential as putative biomarkers. The detection of these proteins in sera underscores the analytical potential and viability for a blood-based assay for these putative HNOSCC biomarkers in the future. Individually, the proteins are not specific to HNOSCC and have been detected in other human cancers also. However, as a panel, they may provide enhanced selectivity to HNOSCC; this verification will require large-sample studies involving other cancer types. In short, the secretome analysis of head and neck cancer cell lines provided a rationale and tractable means to identify serological biomarkers and opportunities for MS-based clinical proteomics. Thus, results

of the current and previous investigations demonstrate that examining the secretome analysis of HNOSCC cells provided a viable and effective means to determine protein candidates that may serve as surrogate tumor markers.

Acknowledgements

R.R. gratefully acknowledges support from the Joseph and Mildred Sonshine Centre for Head and Neck Diseases, Alex and Simona Shnaider Research Laboratory in Molecular Oncology, Temmy Latner/Dynacare, and the Department of Otolaryngology-Head and Neck Surgery, Mount Sinai Hospital, University of Toronto. We thank AB SCIEX for reagent support and collaboration. K.W.M.S. acknowledges funding from the Canadian Institutes of Health Research (CIHR), and infrastructural support from the Ontario Research and Development Challenge Fund and AB SCIEX.

The authors have declared no conflict of interest.

References

[1] McAllister, S. S., Weinberg, R. A., Tumor-host interactions: a far-reaching relationship. J. Clin. Oncol. 2010, 28, 4022–4028.

[2] Liotta, L. A., Ferrari, M., Petricoin, E., Clinical proteomics: written in blood. Nature 2003, 425, 905.

[3] Schaaij-Visser, T. B., Brakenhoff, R. H., Leemans, C. R., Heck, A. J. et al., Protein biomarker discovery for head and neck cancer. J. Proteomics 2010, 73, 1790–1803.

[4] Pavlou, M. P., Diamandis, E. P., The cancer cell secretome: a good source for discovering biomarkers? J. Proteomics 2010, 73, 1896–1906.

[5] Hu, S., Arellano, M., Boontheung, P., Wang, J. et al., Salivary proteomics for oral cancer biomarker discovery. Clin. Cancer Res. 2008, 14, 6246–6252.

[6] Rifai, N., Gillette, M. A., Carr, S. A., Protein biomarker discovery and validation: the long and uncertain path to clinical utility. Nat. Biotechnol. 2006, 24, 971–983.

[7] Karagiannis, G. S., Pavlou, M. P., Diamandis, E. P., Cancer secretomics reveal pathophysiological pathways in cancer molecular oncology. Mol. Oncol. 2010, 4, 496–510.

[8] Villanueva, J., Philip, J., Chaparro, C. A., Li, Y. et al., Correcting common errors in identifying cancer-specific serum peptide signatures. J. Proteome Res. 2005, 4, 1060–1072.

[9] Anderson, N. L., Anderson, N. G., The human plasma proteome: history,

character, and diagnostic prospects. Mol. Cell. Proteomics 2002, 1, 845–867.

[10] Gunawardana, C. G., Kuk, C., Smith, C. R., Batruch, I. et al., Comprehensive analysis of conditioned media from ovarian cancer cell lines identifies novel candidate markers of epithelial ovarian cancer. J. Proteome Res. 2009, 8, 4705–4713.

[11] Planque, C., Kulasingam, V., Smith, C. R., Reckamp, K. et al., Identification of five candidate lung cancer biomarkers by proteomics analysis of conditioned media of four lung cancer cell lines. Mol. Cell. Proteomics 2009, 8, 2746–2758.

[12] Sardana, G., Jung, K., Stephan, C., Diamandis, E. P., Proteomic analysis of conditioned media from the PC3, LNCaP, and 22Rv1 prostate cancer cell lines: discovery and validation of candidate prostate cancer biomarkers. J. Proteome Res. 2008, 7, 3329–3338.

[13] Kulasingam, V., Diamandis, E. P., Proteomics analysis of conditioned media from three breast cancer cell lines: a mine for biomarkers and therapeutic targets. Mol. Cell. Proteomics 2007, 6, 1997–2011.

[14] Lawlor, K., Nazarian, A., Lacomis, L., Tempst, P., Villanueva, J., Pathway-based biomarker search by high-throughput proteomics profiling of secretomes. J. Proteome Res. 2009, 8, 1489–1503.

[15] Gourin, C. G., Zhi, W., Adam, B. L., Proteomic identification of serum biomarkers for head and neck cancer surveillance. Laryngoscope 2009, 119, 1291–1302.

[16] Bijian, K., Mlynarek, A. M., Balys, R. L., Jie, S. et al., Serum proteomic approach for the identification of serum biomarkers contributed by oral squamous cell carcinoma and host tissue microenvironment (dagger). J. Proteome Res. 2009, 8, 2173–2185.

[17] Allinen, M., Beroukhim, R., Cai, L., Brennan, C. et al., Molecular characterization of the tumor microenvironment in breast cancer. Cancer Cell 2004, 6, 17–32.

[18] Chen, S. T., Pan, T. L., Juan, H. F., Chen, T. Y. et al., Breast tumor microenvironment: proteomics highlights the treatments targeting secretome. J. Proteome Res. 2008, 7, 1379–1387.

[19] Jemal, A., Siegel, R., Xu, J., Ward, E., Cancer statistics. 2010. CA Cancer J. Clin. 2010, 60, 277–300.

[20] Ralhan, R., Desouza, L. V., Matta, A., Chandra Tripathi, S. et al., iTRAQ-multidimensional liquid chromatography and tandem mass spectrometry-based identification of potential biomarkers of oral epithelial dysplasia and novel networks between inflammation and premalignancy. J. Proteome Res. 2009, 8, 300–309.

[21] Ralhan, R., Desouza, L. V., Matta, A., Chandra Tripathi, S. et al., Discovery and verification of head-and-neck cancer biomarkers by differential protein expression analysis using iTRAQ labeling, multidimensional liquid chromatography, and tandem mass spectrometry. Mol. Cell. Proteomics 2008, 7, 1162–1173.

[22] Weng, L. P., Wu, C. C., Hsu, B. L., Chi, L. M. et al., Secretome-based

identification of Mac-2 binding protein as a potential oral cancer marker involved in cell growth and motility. J. Proteome Res. 2008, 7, 3765–3775.

[23] Vasko, V., Espinosa, A. V., Scouten, W., He, H. et al., Gene expression and functional evidence of epithelial-tomesenchymal transition in papillary thyroid carcinomainvasion. Proc. Natl. Acad. Sci. USA 2007, 104, 2803–2808.

[24] Park, J. I., Strock, C. J., Ball, D. W., Nelkin, B. D., The Ras/ Raf/MEK/ extracellular signal-regulated kinase pathway induces autocrine-paracrine growth inhibition via the leukemia inhibitory factor/JAK/STAT pathway. Mol. Cell Biol. 2003, 23, 543–554.

[25] Grille, S. J., Bellacosa, A., Upson, J., Klein-Szanto, A. J. et al., The protein kinase Akt induces epithelial mesenchymal transition and promotes enhanced motility and invasiveness of squamous cell carcinoma lines. Cancer Res. 2003, 63, 2172–2178.

[26] Goldfinger, L. E., Stack, M. S., Jones, J. C., Processing of laminin-5 and its functional consequences: role of plasmin and tissue-type plasminogen activator. J. Cell Biol. 1998, 141, 255–265.

[27] Ziober, A. F., Patel, K. R., Alawi, F., Gimotty, P. et al., Identification of a gene signature for rapid screening of oral squamous cell carcinoma. Clin. Cancer Res. 2006, 12, 5960–5971.

[28] Chen, L. C., Hsueh, C., Tsang, N. M., Liang, Y. et al., Heterogeneous ribonucleoprotein k and thymidine phosphorylase are independent

prognostic and therapeutic markers for nasopharyngeal carcinoma. Clin. Cancer Res. 2008, 14, 3807–3813.

[29] Sun, Y., Yi, H., Zhang, P. F., Li, M. Y. et al., Identification of differential proteins in nasopharyngeal carcinoma cells with p53 silence by proteome analysis. FEBS Lett. 2007, 581, 131–139.

[30] Leivo, I., Jee, K. J., Heikinheimo, K., Laine, M. et al., Characterization of gene expression in major types of salivary gland carcinomas with epithelial differentiation. Cancer Genet. Cytogenet. 2005, 156, 104–113.

[31] Wang, Z., Jiang, L., Huang, C., Li, Z. et al., Comparative proteomics approach to screening of potential diagnostic and therapeutic targets for oral squamous cell carcinoma. Mol. Cell. Proteomics 2008, 7, 1639–1650.

[32] Lo, W. Y., Tsai, M. H., Tsai, Y., Hua, C. H. et al., Identification of over-expressed proteins in oral squamous cell carcinoma (OSCC) patients by clinical proteomic analysis. Clin. Chim. Acta 2007, 376, 101–107.

[33] Koike, H., Uzawa, K., Nakashima, D., Shimada, K. et al., Identification of differentially expressed proteins in oralsquamous cell carcinoma using a global proteomic approach. Int. J. Oncol. 2005, 27, 59–67.

[34] He, Q. Y., Chen, J., Kung, H. F., Yuen, A. P., Chiu, J. F., Identification of tumor-associated proteins in oral tongue squamous cell carcinoma by proteomics. Proteomics 2004, 4, 271–278.

[35] Turhani, D., Krapfenbauer, K., Thurnher, D., Langen, H., Fountoulakis, M., Identification of differentially expressed, tumor-associated proteins in oral squamous cell carcinoma by proteomic analysis. Electrophoresis

2006, 27, 1417–1423.

[36] Celetti, A., Testa, D., Staibano, S., Merolla, F. et al., Overexpression of the cytokine osteopontin identifies aggressive laryngeal squamous cell carcinomas and enhances carcinoma cell proliferation and invasiveness. Clin. Cancer Res. 2005, 11, 8019–8027.

[37] Chen, J., He, Q. Y., Yuen, A. P., Chiu, J. F., Proteomics of buccal squamous cell carcinoma: the involvement of multiple pathways in tumorigenesis. Proteomics 2004, 4, 2465–2475.

[38] Onda, M., Emi, M., Yoshida, A., Miyamoto, S. et al., Comprehensive gene expression profiling of anaplastic thyroid cancers with cDNA microarray of 25 344 genes. Endocr. Relat. Cancer 2004, 11, 843–854.

[39] Seiwert, T. Y., Jagadeeswaran, R., Faoro, L., Janamanchi, V. et al., The MET receptor tyrosine kinase is a potential novel therapeutic target for head and neck squamous cell carcinoma. Cancer Res. 2009, 69, 3021–3031.

[40] Matta, A., Ralhan, R., DeSouza, L. V., Siu, K. W., Mass spectrometry-based clinical proteomics: head-and-neck cancer biomarkers and drug-targets discovery. Mass Spectrom. Rev. 2010, 29, 945–961.

[41] Kuzyk, M. A., Smith, D., Yang, J., Cross, T. J. et al., Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. Mol. Cell. Proteomics 2009, 8, 1860–1877.

[42] Addona, T. A., Abbatiello, S. E., Schilling, B., Skates, S. J. et al., Multi-site assessment of the precision and reproducibility of multiple reaction

monitoring-based measurements of proteins in plasma. Nat. Biotechnol. 2009, 27, 633–641.

[43] DeSouza, L. V., Taylor, A. M., Li, W., Minkoff, M. S. et al., Multiple reaction monitoring of mTRAQ-labeled peptides enables absolute quantification of endogenous levels of a potential cancer marker in cancerous and normal endometrial tissues. J. Proteome Res. 2008, 7, 3525–3534.

[44] DeSouza, L. V., Romaschin, A. D., Colgan, T. J., Siu, K. W., Absolute quantification of potential cancer markers in clinical tissue homogenates using multiple reaction monitoring on a hybrid triple quadrupole/linear ion trap tandem mass spectrometer. Anal. Chem. 2009, 81, 3462–3470.

[45] Kaur, J., Ralhan, R., Establishment and characterization of a cell line from smokeless tobacco associated oral squamous cell carcinoma. Oral Oncol. 2003, 39, 806–820.

[46] Matta, A., DeSouza, L. V., Shukla, N. K., Gupta, S. D. et al., Prognostic significance of head-and-neck cancer biomarkers previously discovered and identified using iTRAQ-labeling and multidimensional liquid chromatographytandem mass spectrometry. J. Proteome Res. 2008, 7, 2078–2087.

[47] Li, H., DeSouza, L. V., Ghanny, S., Li, W. et al., Identification of candidate biomarker proteins released by human endometrial and cervical cancer cells using two-dimensional liquid chromatography/tandem mass spectrometry. J. Proteome Res. 2007, 6, 2615–2622.

[48] Tang, W. H., Shilov, I. V., Seymour, S. L., Nonlinear fitting method for determining local false discovery rates from decoy database searches. J Proteome Res. 2008, 7, 3661–3667.

[49] Bendtsen, J. D., Nielsen, H., von Heijne, G., Brunak, S., Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 2004, 340, 783–795.

[50] Bendtsen, J. D., Jensen, L. J., Blom, N., Von Heijne, G., Brunak, S., Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng. Des. Sel. 2004, 17, 349–356.

[51] Tuck, M. K., Chan, D. W., Chia, D., Godwin, A. K. et al., Standard operating procedures for serum and plasma collection: early detection research network consensus statement standard operating procedure integration working group. J. Proteome Res. 2009, 8, 113–117.

[52] Kashat, L., So, A. K., Masui, O., Wang, X. S. et al., Secretome-based identification and characterization of potential biomarkers in thyroid cancer. J Proteome Res. 2010, 9, 5757–5769.

[53] Thery, C., Boussac, M., Veron, P., Ricciardi-Castagnoli, P. et al., Proteomic analysis of dendritic cell-derived exosomes: a secreted subcellular compartment distinct from apoptotic vesicles. J. Immunol. 2001, 166, 7309–7318.

[54] Wubbolts, R., Leckie, R. S., Veenhuizen, P. T., Schwarzmann, G. et al., Proteomic and biochemical analyses of human B cell-derived exosomes.

Potential implications for their function and multivesicular body formation. J. Biol. Chem. 2003, 278, 10963–10972.

[55] van Niel, G., Raposo, G., Candalh, C., Boussac, M. et al., Intestinal epithelial cells secrete exosome-like vesicles. Gastroenterology 2001, 121, 337–349.

[56] Skokos, D., Le Panse, S., Villa, I., Rousselle, J. C. et al., Mast cell-dependent B and T lymphocyte activation is mediated by the secretion of immunologically active exosomes. J. Immunol. 2001, 166, 868–876.

[57] Admyre, C., Johansson, S. M., Qazi, K. R., Filen, J. J. et al., Exosomes with immune modulatory features are present in human breast milk. J. Immunol. 2007, 179, 1969–1978.

[58] Mears, R., Craven, R. A., Hanrahan, S., Totty, N. et al., Proteomic analysis of melanoma-derived exosomes by twodimensional polyacrylamide gel electrophoresis and mass spectrometry. Proteomics 2004, 4, 4019–4031.

[59] Gonzales, P. A., Pisitkun, T., Hoffert, J. D., Tchapyjnikov, D. et al., Large-Scale Proteomics and Phosphoproteomics of Urinary Exosomes. J. Am. Soc. Nephrol. 2008, 20, 363–379.

[60] Pisitkun, T., Shen, R. F., Knepper, M. A., Identification and proteomic profiling of exosomes in human urine. Proc. Natl. Acad. Sci. USA 2004, 101, 13368–13373.

[61] States, D. J., Omenn, G. S., Blackwell, T. W., Fermin, D. et al., Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. Nat. Biotechnol. 2006, 24,

333–338.

[62] Wu, C. C., Hsu, C. W., Chen, C. D., Yu, C. J. et al., Candidate serological biomarkers for cancer identified from the secretomes of 23 cancer cell lines and the human protein atlas. Mol. Cell. Proteomics 2010, 9, 1100–1117.

[63] Piersma, S. R., Fiedler, U., Span, S., Lingnau, A. et al., Workflow comparison for label-free, quantitative secretome proteomics for cancer biomarker discovery: method evaluation, differential analysis, and verification in serum. J Proteome Res. 2010, 9, 1913–1922.

[64] Pancholi, V., Multifunctional alpha-enolase: its role in diseases. Cell Mol. Life Sci. 2001, 58, 902–920.

[65] van den Bemd, G. J., Krijgsveld, J., Luider, T. M., van Rijswijk, A. L. et al., Mass spectrometric identification of human prostate cancer-derived proteins in serum of xenograftbearing mice. Mol. Cell. Proteomics 2006, 5, 1830–1839.

[66] Karnak, D., Beder, S., Kayacan, O., Ibis, E. et al., Neuronspecific enolase and lung cancer. Am. J. Clin. Oncol. 2005, 28, 586–590.

[67] Tamesa, M. S., Kuramitsu, Y., Fujimoto, M.,Maeda, N. et al., Detection of autoantibodies against cyclophilin A and triosephosphate isomerase in sera from breast cancer patients by proteomic analysis. Electrophoresis 2009, 30, 2168–2181.

[68] Desmetz, C., Bascoul-Mollevi, C., Rochaix , P., Lamy, P.J. et al., Identification of a new panel of serum autoantibodies associated with the presence of in situ carcinoma of the breast in younger women. Clin. Cancer

Res. 2009, 15, 4733–4741.

[69] Obchoei, S., Wongkhan, S., Wongkham, C., Li, M. et al., Cyclophilin A: potential functions and therapeutic target for human cancer. Med. Sci. Monit. 2009, 15, RA221–RA232.

[70] Lee, J., Role of cyclophilin a during oncogenesis. Arch. Pharm. Res. 2010, 33, 181–187.

[71] Huang, C. F., Sun, Z. J., Zhao, Y. F., Chen, X. M. et al., Increased expression of peroxiredoxin 6 and cyclophilin A in squamous cell carcinoma of the tongue. Oral Dis. 2011, 17, 328–343.

[72] Arora, S., Matta, A., Shukla, N. K., Deo, S. V., Ralhan, R., Identification of differentially expressed genes in oral squamous cell carcinoma. Mol. Carcinog. 2005, 42, 97–108.

[73] Matta, A., Bahadur, S., Duggal, R., Gupta, S. D., Ralhan, R., Over-expression of 14-3-3zeta is an early event in oral cancer. BMC Cancer 2007, 7, 169.

[74] Kobayashi, R., Deavers, M., Patenia, R., Rice-Stitt, T. et al., 14-3-3 zeta protein secreted by tumor associated monocytes/macrophages from ascites of epithelial ovarian cancer patients. Cancer Immunol. Immunother. 2009, 58, 247–258.

[75] He, Y., Wu, X., Liu, X., Yan, G., Xu, C., LC-MS/MS analysis of ovarian cancer metastasis-related proteins using a nude mouse model: 14-3-3 zeta as a candidate biomarker. J. Proteome Res. 2010, 9, 6180–6190.

[76] Morrison, D. K., The 14-3-3 proteins: integrators of diverse signaling cues

that impact cell fate and cancer development. Trends Cell Biol. 2009, 19, 16–23.

[77] Badea, L., Herlea, V., Dima, S. O., Dumitrascu, T., Popescu, I., Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia. Hepatogastroenterology 2008, 55, 2016–2027.

[78] Sewell, D. A., Yuan, C. X., Robertson, E., Proteomic signatures in laryngeal squamous cell carcinoma. ORL J. Otorhinolaryngol. Relat. Spec. 2007, 69, 77–84.

[79] Roesch-Ely, M., Nees, M., Karsai, S., Ruess, A. et al., Proteomic analysis reveals successive aberrations in protein expression from healthy mucosa to invasive head and neck cancer. Oncogene 2007, 26, 54–64.

[80] Matta, A., Tripathi, S. C., DeSouza, L. V., Grigull, J. et al., Heterogeneous ribonucleoprotein K (hnRNP K) is a marker of oral leukoplakia and correlates with poor prognosis of squamous cell carcinoma. Int. J. Cancer 2009, 125, 1398–1406.

[81] Jiang, P. Z., Gan, M., Huang, H., Shen, X. M. et al., Proteomics-based identification of proteins with altered expression induced by 12-O-tetradecanoylphorbol 13-acetate in nasopharyngeal carcinoma CNE2 cells. Acta Biochim. Biophys. Sin. (Shanghai) 2005, 37, 97–106.

[82] Barboro, P., Repaci, E., Rubagotti, A., Salvi, S. et al., Heterogeneous nuclear ribonucleoprotein K: altered pattern of expression associated with diagnosis and prognosis of prostate cancer. Br. J. Cancer 2009, 100,

1608–1616.

[83] Roychoudhury, P., Chaudhuri, K., Evidence for heterogeneous nuclear ribonucleoprotein K overexpression in oral squamous cell carcinoma. Br. J. Cancer 2007, 97, 574–575.

[84] Xiaofang, Y., Yue, Z., Xialian, X., Zhibin, Y. et al., Serum tumour markers in patients with chronic kidney disease. Scand. J. Clin. Lab Invest. 2007, 67, 661–667.

[85] Culine, S., El Demery, M., Lamy, P. J., Iborra, F. et al., Docetaxel and cisplatin in patients with metastatic androgen independent prostate cancer and circulating neuroendocrine markers. J. Urol. 2007, 178, 844–848.

[86] Lenhard, M., Kuemper, C., Ditsch, N., Diebold, J. et al., Use of novel serum markers in clinical follow-up of Sertoli-Leydig cell tumours. Clin. Chem. Lab Med. 2007, 45, 657–661.

[87] Vos, M. J., Postma, T. J., Martens, F., Uitdehaag, B. M. et al., Serum levels of S-100B protein and neuron-specific enolase in glioma patients: a pilot study. Anticancer Res. 2004, 24, 2511–2514.

Chapter 6

General Discussion and Conclusions

6.1. Reproducibility of retention time

The reproducibility of the retention time of an analyte in RP-LC is one of the most important factors in the quality of protein identification using the methods as described in Chapters 3 and 4. For PIE to be effective, the retention times of the peptides, identified in the first LC-MS run, must be similar (within 3 min) to those in the second and third runs; otherwise, most of the same peptides will be identified again (see Figure 2.10).

The retention time is affected by variations in column flow rate, temperature, composition of the mobile phase, and surface composition of the stationary phase, which may be affected by contamination (by irreversibly adsorbed peptides) or degradation.

The LC-MS experiment was designed so that ten samples could be analyzed per day. Usually, the first LC-MS identification run of the ten samples was sequentially performed on the first day; then the second run of the same samples was performed on the next day with PIE; and finally, the third run of the samples was performed with PIE on the third day. It was noticed that if the same ten samples were run on two consecutive the days, the retention times would increase by 2 – 4 minutes. The shift was caused by the evaporation of methanol through an opening in the reservoir lid from which eluant B, consisting of 94.9% methanol, was withdrawn. The tolerance window for exclusion was set at 3 min; thus, the shift in retention times significantly affected the exclusion, causing redundant protein identifications. The problem of methanol evaporation was easily

solved by topping up the reservoir with fresh eluant B before each run. By doing so, the retention time shifted by no more than 0.5 – 1 min, allowing the PIE to be much more effective.

Occasionally, the temperature varied between consecutive days, causing a retention time shift. In this case, PIE could still to be used by compensating for the shift. The extent of the shift was first assessed by performing a survey run using one of the samples run the day before. Three peptides were selected, and their retention times from the two consecutive days (RT1 and RT2) were recorded along with their m/z values (Table 6.1). The ratio of the retention times for each peptide (RT2/RT1), and the average of all three ratios, *k*, were calculated. By multiplying RT1 of all of the peptides in the PIE list by *k*, the expected retention times were determined (re-calculated RT1). The re-calculated retention times were very similar to the actual times in the survey run (RT2 in Table 6.1), allowing the PIE method to work.

Table 6.1. Re-calculation of the retention times for PIE list.

						For PIE list	
m/z	**RT1, min**	**RT2, min**	**RT2/RT1**	**k=average of RT2/RT1**	**re-calculated RT1, min**	**m/z**	**re-calculated RT1, min**
893.101	35.41	40.61	1.15	1.1479	40.65	893.101	40.65
678.552	52.06	59.01	1.13		59.76	678.552	59.76
1123.874	69.88	81.304	1.16		80.22	1123.874	80.22

RT1: retention time of the peptide in 1stLC-MS run, RT2: retention time of the peptide in 2nd LC-MS run.

6.2. The role of Pfn1 in the tumors of different tissues

According to our findings (Chapter 5), and those of Minamida et al., [1] Pfn1 is overexpressed in ccRCC, but is underexpressed in most aggressive adenocarcinomas [2]. This suggests that Pfn1 is involved in different tumorigenic mechanisms in different tissue types, as was mentioned in Chapter 4.

In carcinogenesis, the ability of cells to invade and migrate past the basement membrane may be dependent on conformational changes within the cellular backbone. In most cells, Pfn1 remodels the actin cytoskeleton by regulating actin polymerization [3]. G-actin monomers are produced from the disassembling of F-actin filaments, where Pfn1 assists the exchange of ADP for ATP on G-actin (mechanism is shown in Figure 6.1A). The functions of Pfn1 in ccRCC have not been fully elucidated; however, the findings here suggest that Pfn1 may control migration of metastatic ccRCC cells via traditional actin-cytoskeleton pathways.

In contrast, Bae et al. reported that Pfn1 triggers cell migration in metastatic MDA-MB-231 breast cancer cells, not through a traditional actin cytoskeleton rearrangement, but through a regulation of PI(3,4)P(2) [2]. This mechanism is shown in Figure 6.1B. PI(3,4)P(2) recruits lamellipodin (Lpd) and Ena/VASP to the lamellipodial tip where the two proteins bind and trigger cell motility. If Pfn1 is expressed in the cell, it binds PI(3,4)P(2), thereby, limits the recruitment of Lpd and Ena/VASP which, in turn, inhibits cell motility. Therefore, Bae et al. proposed that Pfn1 inhibits the motility of MDA-MB-231 cells by negatively regulating PI(3,4)P(2).

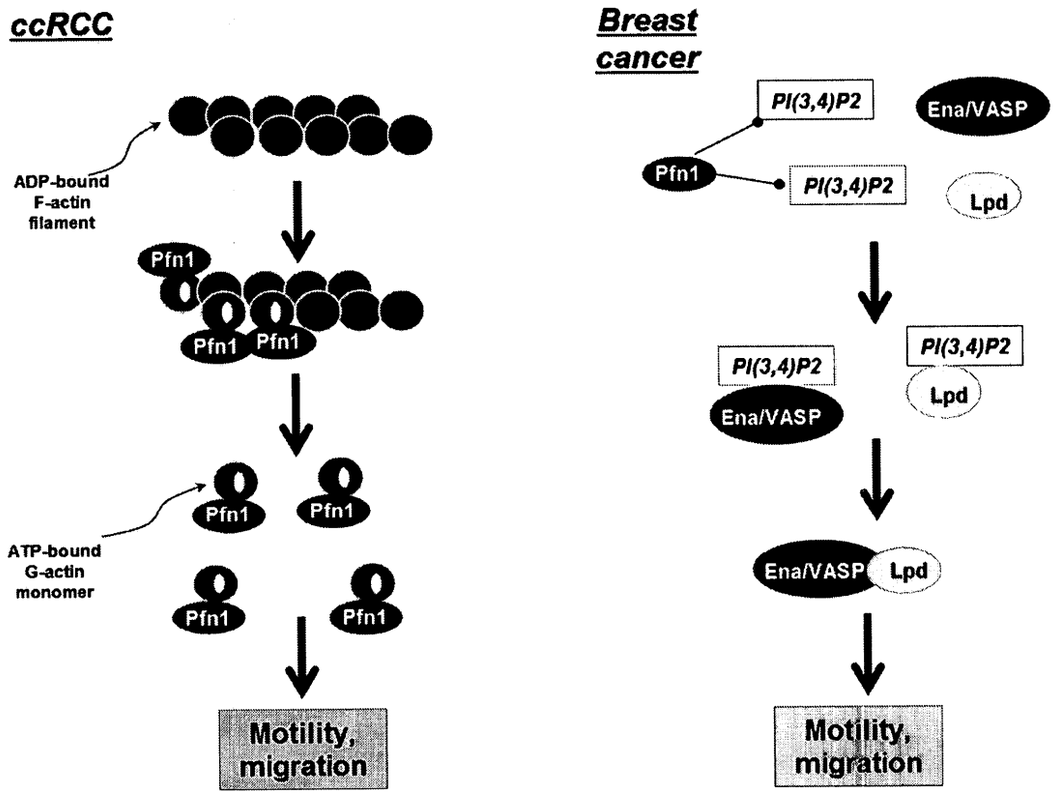


Figure 6.1. Different roles of Pfn1 in tumorigenic mechanisms in ccRCC and breast cancer. (A) *In ccRCC, Pfn1 may control migration of metastatic cells via traditional actin-cytoskeleton pathways.* ***(B)*** *In contrast, Pfn1 inhibits the migration of MDA-MB-231 cells by negatively regulating PI(3,4)P(2).*

6.3. The importance of using a panel of cancer biomarkers

As was discussed in Section 1.7.3, the sensitivity and specificity biomarkers markedly improve by assembling them into panels of biomarkers [4]. It was pointed out how cancer-specific, biomarker panels can be assembled from cancer- nonspecific biomarkers. Here, we discuss how the cancer-nonspecific biomarkers, identified in this work, can be assembled into panels for the diagnosis/prognosis of different types of cancers. The cancer-nonspecific biomarkers found in our investigation of ccRCC were

LDHA, ENO1, HSPB1, HSPE1, AHNAK, and Pfn1 (see chapter 3 and 4). These have also been reported as biomarkers for breast cancer [2, 5, 6], brain cancer [7-10], and endometrial carcinoma [11-13]. Figure 6.2 shows four panels assembled from the six biomarkers, which can be used for the diagnosis/prognosis of four different cancers. The overexpression of LDHA, ENO1, and HSPB1 indicates that one of the four cancers in present. The particular cancer is identified by the pattern of over- or underexpression of the remaining three biomarkers (HSPE1, AHNAK, or Pfn1).

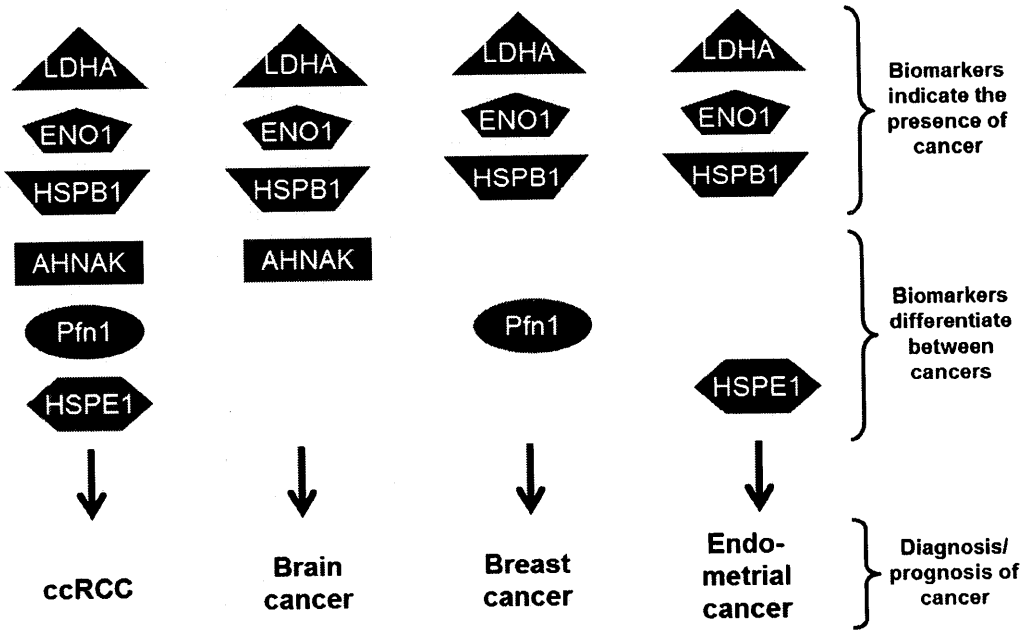


Figure 6.2. Panels of biomarkers for the diagnosis/prognosis of four different cancers constructed from the six biomarkers identified in this thesis. *Red symbolizes overexpression, whereas green symbolizes underexpression.*

6.4. Closing Remarks

Cancer biomarker discovery is a growing field whose findings will eventually improve the management of cancer. The availability and quality of biological samples, and the variability among those taken from different individuals, and even those obtained from the same person, remain as challenges to the researcher. Advancements in instrumentation and analytical methodologies have allowed progressively lower protein concentrations to be quantified at increasingly fast rates, and this has allowed access to proteins that were previously unattainable and allowed more thorough proteomic studies. The development of comprehensive panels of biomarkers with high sensitivity and specificity, which can account for different cancers and their subpopulations, will improve management of cancer patients.

The biomarker candidates for ccRCC and HNOSCC, identified in our discovery phase studies, must still be validated on a larger set of samples. Proteins Gal-1, Prf1, LDHA, and ENO1, are currently being validated in hundreds of primary and metastatic ccRCC patients using IHC analysis in combination with patient clinical histories.

After the biomarkers are validated, clinical assays based on ELISA or MRM-MS must be developed for clinical application. ELISA assays are considered to have the highest sensitivity, and zeptomoles of proteins have been detected [14], although its typical detection limit is on the attomole scale. The more recent MRM-MS-based targeted approach, using mass-spectrometers such as the AB SCIEX QTRAP 5500 System, can quantify of over 1000 proteins in a single LC-MS analysis and has a detection limit comparable to ELISA, but a superior, 4-orders of magnitude, dynamic range [15].

In summary, the focus of biomarker research today should be on the development of non-invasive biofluid-based biomarkers, and on the design of panels of diagnostic, prognostic, and predictive biomarkers which would make a significant impact on cancer patient management.

References

1. Minamida S, Iwamura M, Kodera Y, Kawashima Y, Ikeda M, Okusa H, Fujita T, Maeda T, Baba S. Profilin 1 overexpression in renal cell carcinoma. Int. J. Urol. 2011; 18, 63-71.
2. Bae YH, Ding Z, Das T, Wells A, Gertler F, Roy P. Profilin1 regulates PI(3,4)P2 and lamellipodin accumulation at the leading edge thus influencing motility of MDA-MB-231 cells. Proc Natl Acad Sci U S A. 2010 Dec 14;107(50):21547-52. Epub 2010 Nov 29.
3. Pollard TD, Borisy GG. Cellular motility driven by assembly and disassembly of actin filaments. Cell. 2003 Feb 21;112(4):453-65. Review. Erratum in: Cell. 2003 May 16;113(4):549.
4. Lacombe J, Mangé A, Jarlier M, Bascoul-Mollevi C, Rouanet P, Lamy PJ, Maudelonde T, Solassol J. Identification and validation of new autoantibodies for the diagnosis of DCIS and node negative early-stage breast cancers. Int J Cancer. 2012 Aug 7. doi: 10.1002/ijc.27766.
5. Hussien R, Brooks GA. Mitochondrial and plasma membrane lactate transporter and lactate dehydrogenase isoform expression in breast cancer cell lines. Physiol Genomics. 2011 Mar 16;43(5):255-64. Epub 2010 Dec 21.
6. Rambaruth ND, Greenwell P, Dwek MV. The lectin Helix pomatia agglutinin recognizes O-GlcNAc containing glycoproteins in human breast cancer. Glycobiology. 2012 Jun;22(6):839-48. Epub 2012 Feb 9.

7. Wolf A, Agnihotri S, Munoz D, Guha A. Developmental profile and regulation of the glycolytic enzyme hexokinase 2 in normal brain and glioblastoma multiforme. Neurobiol Dis. 2011 Oct;44(1):84-91. Epub 2011 Jun 25.
8. Beckner ME, Gobbel GT, Abounader R, Burovic F, Agostino NR, Laterra J, Pollack IF. Glycolytic glioma cells with active glycogen synthase are sensitive to PTEN and inhibitors of PI3K and gluconeogenesis. Lab Invest. 2005 Dec;85(12):1457-70.
9. Golembieski WA, Thomas SL, Schultz CR, Yunker CK, McClung HM, Lemke N, Cazacu S, Barker T, Sage EH, Brodie C, Rempel SA. HSP27 mediates SPARC-induced changes in glioma morphology, migration, and invasion. Glia. 2008 Aug 1;56(10):1061-75.
10. Gentil BJ, Benaud C, Delphin C, Remy C, Berezowski V, Cecchelli R, Feraud O, Vittet D, Baudier J. Specific AHNAK expression in brain endothelial cells with barrier properties. J Cell Physiol. 2005 May;203(2):362-71.
11. Voisin SN, Krakovska O, Matta A, DeSouza LV, Romaschin AD, Colgan TJ, Siu KW. Identification of novel molecular targets for endometrial cancer using a drill-down LC-MS/MS approach with iTRAQ. PLoS One. 2011 Jan 31;6(1):e16352.
12. Lomnytska MI, Becker S, Gemoll T, Lundgren C, Habermann J, Olsson A, Bodin I, Engström U, Hellman U, Hellman K, Hellström AC, Andersson S, Mints M, Auer G. Impact of genomic stability on protein expression in endometrioid endometrial cancer. Br J Cancer. 2012 Mar 27;106(7):1297-305.

13. Korneeva I, Caputo TA, Witkin SS. Cell-free 27 kDa heat shock protein (hsp27) and hsp27-cytochrome c complexes in the cervix of women with ovarian or endometrial cancer. Int J Cancer. 2002 Dec 10;102(5):483-6.

14. Pawlak M, Schick E, Bopp MA, Schneider MJ, Oroszlan P, Ehrat M. Zeptosens' protein microarrays: a novel high performance microarray platform for low abundance protein analysis. Proteomics. 2002; (2)383–393.

15. http://www.absciex.com/applications/drug-discovery-and-development-mass-spec/regulated-bioanalysis/dried-matrix-spot-analysis

Appendix: List of publications

1. **Masui O**, White NMA, DeSouza LV, Krakovska O, Matta A, Metias S, Khalil B, Romaschin AD, Honey RJ, Stewart R, Pace K, Bjarnason GA, Siu KW, Yousef GM. Quantitative proteomic analysis in metastatic renal cell carcinoma reveals a unique set of proteins with potential prognostic significance. Paper was resubmitted, after minor corrections, on August 28, 2012 in Mol Cell Proteomics (ID#: MCP/2012/020701).
2. Ralhan R, **Masui O**, Desouza LV, Matta A, Macha M, Siu KW. Identification of proteins secreted by head and neck cancer cell lines using LC-MS/MS: Strategy for discovery of candidate serological biomarkers. Proteomics. 2011 Jun;11(12):2363-76. doi: 10.1002/pmic.201000186. Epub 2011 May 20.
3. Chan CY, **Masui O**, Krakovska O, Belozerov VE, Voisin S, Ghanny S, Chen J, Moyez D, Zhu P, Evans KR, McDermott JC, Siu KW. Identification of differentially regulated secretome components during skeletal myogenesis. Mol Cell Proteomics. 2011 May;10(5):M110.004804. Epub 2011 Feb 22.
4. Kashat L, So AK, **Masui O**, Wang XS, Cao J, Meng X, Macmillan C, Ailles LE, Siu KW, Ralhan R, Walfish PG. Secretome-based identification and characterization of potential biomarkers in thyroid cancer. J Proteome Res. 2010 Nov 5;9(11):5757-69. Epub 2010 Oct 15.