Protein Expression and Glycosylation Level Change Associated Pancreatic Cancer

By

Chen Li

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in The University of Michigan
2010

Doctoral Committee:

Professor David M. Lubman, Chair
Professor Michael D. Morris
Associate Professor Kerby A. Shedden
Associate Professor Kristina Hakansson

To my parents, and my wife

ACKNOWLEDGEMENT

I would like to sincerely thank my advisor, Dr. David M. Lubman for his support and encouragement for my research projects through my five years in the University of Michigan.

I also want to thank the rest of my committee members, Dr. Michael Morris, Dr. Kristina Hakansson and Dr. Kerby Shedden for being in my committee and spending time and effort on my dissertation and meetings.

My thanks also go to former and current members of the Lubman lab for their help in my research and life in Ann Arbor. Special thanks are giving to Dr. Jia Zhao, Dr. Yoo Chul and Dr. Tasneem Patwa for the trainings and advice during the first several years in the group.

I also owe my gratitude to my research collaborators, Dr. Diane Simeone, Dr. Mack Ruffin, Dr. Dean Brenner, Dr. Michelle Anderson and Dr. Eugene Zolotarevsky for providing me the samples that are essential for my projects.

Finally I give my deep appreciation to my parents and my wife for their support and understanding.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Chapter 1

Introduction

Pancreatic cancer

Pancreatic cancer is the 4th leading cause of cancer death in the United States[1]. According to the statistics of National Cancer Institute, there are 42 thousand new cases in 2009 and the estimated death is 35 thousand. Pancreatic cancer is the known for its extremely low 5year survival rate, which is only 5%. Surgical resection is only possible if the tumor is detected early. However, 85% of the patients were diagnosed with unresectable pancreatic cancer because symptoms usually do not develop until tumor extends to the retro peritoneum, adjacent organs and vasculature. Unresectable pancreatic cancer is associated with mean survival of 6.5 months[2]. A small minority of patients (10-20%) is found to have early resectable pancreatic cancer at presentation. While majority of these patients do not achieve curability, pancreatic resection is associated with substantially better outcomes including a five-fold higher 5-year survival rate. Unfortunately, overwhelming majority of patients do not present with early-stage disease and currently there are no clinically utilized strategies for detection of early pancreatic cancer. Current medical approaches for unresectable pancreatic cancer involve palliative chemotherapy and radiation[3]. Neither of these therapies is associated with long-term survival. Pancreatic caner is also sometimes difficult to distinguish from chronic pancreatitis. False diagnosis may lead to unnecessary surgery which could be life-threatening for some patients.

Existing biomarkers

Early-stage detection of pancreatic cancer can clearly affect the outcome of treatment against the disease and provide a prospect of long-term survival and even curability. Many strategies utilizing both invasive and noninvasive techniques have been employed to detect early pancreatic cancer among patients at high risk of developing this malignancy. Endoscopic ultrasound, magnetic resonance cholangiopancreatography and computed tomography scans have been studied for screening pancreatic cancer[4]. However, due to the unpleasantness and cost of these exams, their usage has been limited to high risk population which is determined by hereditary factors and the presence of genetic syndromes. Compared to the invasive exams, noninvasive method, such as serum biomarker test, is much easier to be widely accepted by the general population. CA-19-9 is the most famous serum biomarker for pancreatic cancer[5]. The abundance of this biomarker in serum elevates in 80% of the patients with pancreatic cancer, however it also elevates in pancreatitis patients. Researchers have found other potential biomarkers up-regulated in the serum for pancreatic cancer and combined them together to improve the sensitivity and specificity. Faca et al identified the proteins with increase levels in the sera of tumor-bearing mice and from these proteins five of them were chosen to be combined with/without CA-19-9 as a panel to distinguish humans with pancreatic cancer from pancreatitis/healthy controls[6]. In another study, a larger panel of 92 distinctive proteins is assembled, where the protein level in the sera is probed with antibodies microarray[7]. Panels of multiple biomarkers have shown much better performance than the single CA-19-9 test, but the sensitivity and specificity of these panels have not been very satisfactory since the low occurrence rate of this tumor requires an extreme low false positive rate.

Glycoprotein and glycomic profiling

Glycosylation is the most commonly occurring post translational modification on proteins involved in numerous biological processes, such as protein-protein interactions, protein folding, immune recognition, cell adhesion, and inter-cellular signaling. The function of glycoproteins is highly dependent on their carbohydrate structure. The alteration on the glycans is associated with multiple biological events and has been reported in a variety of diseases, especially cancer[8-11]. In the search for effective glycosylated biomarkers for targeted diseases, there has been great deal of effort invested in profiling and characterizing glycoproteins in complex samples. Many clinical biomarkers and therapeutic targets in cancer are glycoproteins, such as CA125 in ovarian cancer, Her2/neu in breast cancer and prostate-specific antigen (PSA) in prostate cancer [12-14]. In addition, the alteration in protein glycosylation which occurs through varying the heterogeneity of glycosylation sites or changing glycan structure of proteins on the cell surface and in body fluids have been shown to correlate with the development and/or progression of cancer and other disease states[15-17]. Identification of glycoprotein isoforms is becoming increasingly important to the diagnosis and management of human diseases as more diseases are found to result from glycan structural alterations such as I-cell disease and congenital disorders of glycosylation leukocyte adhesion deficiency type II[18-19].

Due to the complexity of the glycosylation structure, analysis of glycoprotein still presents many unconquered challenges against current technologies. Glycoproteins are often isolated from the proteome by affinity chromatography utilizing the selective binding between the glycans to lectins. Lectins are proteins that possess motifs that have

a strong affinity for glycans (the sugar component of glycoproteins). The isolated glycoproteins can be analyzed in their intact forms using RPLC and ESI-MS to obtain the molecular weight and relative abundance[20]. However, the resolved glycoforms, especially the ones with the glycan structure of interest, are often present at low copy number that requires highly sensitive detection technique. Alternatively, after enzymatic-cleavage of the intact proteins, the glycosylated peptides are analyzed with liquid chromatography and mass spectrometry[21-24]. Due to the additional molecular weight and heterogeneity of the glycans, the glycosylated peptides often show lower ionization efficiency than unglycosylated peptides thus purification is necessary. Graphitized carbon, cellulose, sepharose and lectins have been used as SPE materials for glycopeptide enrichment[25]. In order to identify as many glycopeptide peaks as possible, in addition to high-accuracy molecular weight, MS/MS is often required to elucidate the sequence of the peptides then consequently the structure of the glycans. However, due to complexity of the glycan structure, the intensity of many low-abundance glycopeptides is not high enough to generate high-quality MS/MS spectra. Glycosylated peptides can be further treated with glycan-releasing enzyme to fragment the hybrid molecules to glycans and peptides for structural determination. Normal phase chromatography is usually employed for the separation of glycans for the polar nature of the carbohydrate molecules[26]. This method provides an overall profile of the structural changes of glycans as the onset or progression of a disease occurs. However, compared to the glycopeptide profiling, the information about the protein origin of the glycan is lost.

Microarray approaches

Using LC-MS in glycomic analysis has proved to be effective in decoding the glycan structures and also generating quantitative information of the glycans or glycoproteins. Although very comprehensive, profiling the glycome with LC-MS method is time-consuming when a large number of samples are analyzed. The prolonged analyzing time might also affect the reproducibility of the experiment and consequently compromise the accuracy of the quantitative result.

A technique that is becoming increasingly popular involves protein microarrays. Microarrays have proved to be a high-throughput technique because they allow multiple analytes to be investigated simultaneously due to their physical attachment to unique regions on a single microarray slide that can be analyzed with a reagent of choice[27-29]. By taking advantage of the high-throughput nature of the microarray platform and exploiting the unique binding properties of glycoproteins, carbohydrates and glycans, comprehensive studies of glycoproteomes are made possible. A variety of microarray-based platforms for the study of glycoproteins have been pursued thus far.

Lectin array is based on binding of an intact glycoprotein or glycopeptides to the arrayed lectins, to provide in a characteristic fingerprint pattern that is highly sensitive to changes in a protein's glycan composition. Lectin arrays determine the most appropriate lectins for glycoprotein enrichment as well as for the removal of undesirable glycoproteins. Lectin microarrays could also be used for rapid and simple analysis of protein glycosylation[30-32]. Protein microarrays are a useful screening method for whole cell lysates, fractionated proteomes, intact glycoproteins, and antigen-antibody reactions. One very promising strategy that has been explored is the use of multidimensional-fractionation techniques to simplify cell lysates into less-complex

fractions to produce natural protein microarrays. Patwa et al. have developed a fractionated glycoprotein array technique to analyze the unique glycosylation pattern of serum from pancreatic cancer patients[33]. The glycoproteins in the serum were enriched with two broad-spectrum lectins then separated with RPLC into fractions that each contained only a few glycoproteins which were subsequently arrayed on nitrocellulose slides before being probed with five lectins with different glycan-selectivity. Due to the complexity of some samples and limited resolution of RP-HPLC fractionation, the accuracy of the biomarker candidates identified with the glycoprotein microarray might be compromised when the differentiated fraction contains several glycoproteins and none of them is dominant. In this case the protein that contributes the most to the observed lectin binding can not be determined.

In Chapter Two, the development of an antibody/lectin glycoprotein microarray technique is described. The use of antibodies is a means to capture and purify specific proteins in a natural state. The antibody microarray allows the efficient, multiplexed study of multiple individual proteins from complex mixtures. The variable carbohydrate structures attached to the captured proteins can be characterized by studying the interaction between the glycans and different kinds of biotinylated lectins or glycan-binding antibodies[34-35]. Antibody-lectin sandwich microarray provides a means to measure glycosylation change on specific proteins captured from complex samples with lectin probes in a high-throughput array format. Compared to HPLC-MS, it is not only a simple and reproducible way of investigating large number of samples but also a type of assay that analyzes proteins in their native form which can be crucial for understanding the functional role of glycoproteins in the biological environment.

Bead-based glycoprotein assay

Antibody microarray assay sometimes suffers low reproducibility when the microarray printed on slides show significant printing variation. To avoid the problem, in Chapter Three we developed a bead-based protein assay which is very similar to the antibody/lectin glycoprotein microarray assay on fluorescence/size labeled beads. The assay involves the capturing target proteins on the immobilized antibodies followed by characterizing and quantifying the captured proteins with different reporter molecules[36-37]. Multiple analytes can be measured in a single assay simultaneously by using a multiplexed bead set in which the identity of each antibody-immobilized bead is defined by size or a unique mixture of two fluorescence dyes (ex. purple/red). Beads sets with different color codes capture target proteins which react with the reporter molecule with different affinity. The unique color code for each kind of bead allows the discrete quantification of each analytes in the system. The glycoprotein-antibody array assay can be performed on the bead platform by using lectins to detect the glycan structure.

Cancer stem cell

Recently a hierarchy model of carcinogenesis may be able to explain the recurrence of tumor and failure of the treatments. The model was proposed based on the concept that tumor consists of a heterogeneous population and only a rare population called cancer stem cells is responsible for initiation and development of the disease[38-40]. The cancer stem cells share some ability of normal stem cells such as self-renewal and producing differentiated progenitor cells and self protection against apoptosis. One impact about identifying the cancer stem cells is that we have to revise the mechanism of drug resistance in cancer according to the self protection property of cancer stem cells and

their ability of regenerating the tumor which may cause recurrence after therapy. We might also be able design treatment specially targeting cancer stem cells or some key pathway in the cancer stem cells to effectively kill the tumor.

Simeone and coworkers in University of Michigan have found a type of pancreatic stem cells which could be isolated from normal cells by three surface antigens CD24/CD44/ESA[41-43]. They performed functional study by planting 500 cells with/without the three surface antigens to an immunodeficient mouse. The result shows that a tumor was produced on the side where the cells with the surface antigens were planted, which proved that the cancer stem cells were responsible for regenerating the tumor. The material for our proteomic study was extracted from such a xenografted tumor in mice. The reasons for using xenografted tumor in mice are we have difficulty obtaining primary human tumor, it has been proven that mouse model accurately represent many features in primary tumor and many tests can be done on mice. As I mentioned, there are on 0.2-0.8% of cancer stem cells in pancreatic tumor. From a mouse tumor, we can typical obtain 10k antibody labeled cancer stem cells using flow cytometry, which corresponds to around 1ug of total proteins.

Shot-gun protein profiling

Profiling the proteome of cancer stem cells is one of the best ways to understand the molecular mechanism underlying the cancer stem cells properties. Shot-gun protein profiling is the most sensitive protein profiling approach for whole proteome profiling[44]. A typical shotgun proteomic strategy includes four major steps: 1) cell lysis, 2) protein digestion, 3) protein prefractionation, 4) follow-up LC-MS analysis. There are two prefractionation strategies for shot-gun proteomics. One includes one or two-

dimensional gel or liquid-based separation of the proteins from cell lysate with electrophoresis or liquid chromatography followed by trypsin digestion and nanoLC-ESI-MS. This strategy is able to resolve proteins with PTMs and sometimes provide information on the protein mass and pI[45]. Alternatively, the cell lysate is directly digested and prefractionated with electrophoresis or chromatography separation before NanoLC-ESI-MS[46]. For small amounts of sample (under 1ug), the concern for the first strategy is that each protein fraction contains even less protein than the original amount, so that the efficiency of trypsin digestion of these protein fractions may be significantly lower than direct digestion of the cell lysate[47]. In addition, separating intact proteins at the micro scale level still presents a challenging technical issue due to the size of intact proteins[48]. Capillary electrophoresis is the only capillary-based method compatible with mass spectrometry for prefractionation of intact proteins with high molecular weight[49]. However, precipitation of proteins in the CE environment may still occur, especially in capillary isoelectric focusing. Direct digestion of cell lysate followed by prefractionation of the resulting peptides is thus a preferable method for analysis of small amounts of sample. The disadvantage of peptide level prefractionation is the distribution of tryptic peptides from high-abundance proteins into multiple fractions, where ionization of peptides from low-abundance protein may be suppressed.

CIEF

A preparative or first dimension liquid separation is frequently used before reverse-phase HPLC-MS to maximize the resolution of the enormous number of peptides from the digestion of a cell lysate, thus the overlapping of the peptides can be reduced to optimize the performance of the identification. The available preparative liquid separations include

ion-exchange, reverse-phase, normal-phase, size-exclusive chromatography and multiple types of electrophoretic separations[50]. The chromatography-based separations often present undesired retention of peptides with strong interaction with the chromatographic resin and, resulting poor recovery rate. In contrast, electrophoresis is usually performed in an open capillary, where peptides are separated based on migration under the influence of an electric field rather than based on retention. The inner wall of the capillaries used for electrophoresis is often coated with inert material to reduce the absorption of the analytes. Therefore electrophoresis provides much higher sample recovery rate than chromatographic methods.

Capillary isoelectric focusing (CIEF) was the first electrophoresis method emerged as a useful tool for protein/peptide prefractionation because of its high resolution and orthogonal separation mechanism versus RP-HPLC[51]. This pI based separation provides an optimal resolution of 0.01 pH unit, which indicates a peak capacity of 700 in a pH range from 3 to 10. CIEF focusing is conducted through establishing a linear pH gradient in the capillary with a series of carrier ampholytes. The peptides are positioned along the pH gradient according to their isoelectric points and concentrated in narrow bands. Once the focusing is complete, the concentrated peptide bands, maintained by an electric field, are hydrodynamically transferred into samples loops or 96-well plates for LC-MS injection. The capillary used in CIEF is usually neutrally coated with hydropropyl cellulose to prevent electroosmotic flow (EOF) and absorption of analytes. Thus, sample loss can be reduced to a minimal level, which is essential in handling small quantities of peptides. The carrier ampholytes for the creation of the pH gradient present in each fraction need to be removed by a nano trap column

filled with C18 material before the RPLC column due to their competition with peptides for electrospray ionization. Another electrophoretic first dimensional separation for shot-gun proteomic study utilizes capillary isotachophoresis/capillary zone electrophoresis (CITP/CZE)[52], which separates and enriches the peptides to different level depending on their abundance. Low-abundance peptides are better enriched than high-abundance ones. CITP/CZE proves to be more robust and provides improved proteome coverage than CIEF. However, in comparative experiments, using CITP/CZE may introduce bias to the measurement of the peptides due to the abundance dependent enrichment; hence CIEF is preferred in our study.

In Chapter Four, our work of analyzing the whole proteome of pancreatic cancer stem cells is presented. An optimized protein extraction and peptide desalting procedure is used in our experiment. We employed CIEF as the first dimension separation and NanoLC as the second dimension. Peptides were identified with MS/MS spectra.

Autoimmune response

Autoimmune response was first known by autoimmune diseases, such as diabetes type 1 and rheumatoid arthritis, where the immune system fails to recognize and direct strong immune attack against self tissue[53]. Then autoimmune response was found to be happening not only in patients with autoimmune diseases but also in every normal person and it plays the role of eliminating the abnormal cells and maintaining the responsiveness of the immune system against foreign antigens. Unlike normal cells, tumor cells are abnormal cells and present unusual antigens such as mutated proteins, proteins with modified PTM, aberrantly expressed proteins and etc[54-58]. These antigens may trigger attacks from the immune system, in the forms of cytotoxic T cells or antibodies produced

by activated B cells. In this biomarker study, we are focused on studying the autoantibodies, which are sustained at a certain level in the serum for a long period of time after the autoimmune attack.

There are several facts about autoantibodies which make them excellent biomarker candidates. Autoantibodies can be directed against intracellular proteins as well as extracellular proteins including some important low-abundance proteins associated with the activation of cancer related genes[59-61]. The changes of autoantibodies in the patient serum have been found to precede the onset of clinical symptoms of multiple types of cancer. It can be used to help diagnosis of cancer, decide the stage of an existing tumor and determine the invasiveness of the disease. And as mentioned above the level of serum antibodies, such as immunoglobin G, is stable for a long period of time after the first immune attack and may even persist for a life time. When a few tumor associated antigens are released into the serum, cells in the immune system are able to bind the antigens at a very low concentration and start the autoimmune response, which provides us a biological amplification of the disease signal. Thus, autoantibodies against tumor associated antigens can be very sensitive biomarkers that provide signals about cancer detection.

Autoantibodies for cancer have another feature that usually only a subset of cancer patients have a differential level of autoantibody against a specific antigen when the level of the autoantibody for the other patients are just like normal people. This means using a single autoantibody as biomarker, it only distinguishes some of the cancer patients from the normal people[62-64]. However, the problem can be fixed by using

other autoantibodies to pick up the rest of the patients because the levels of different autoantibodies are usually not correlated[65-67].

In Chapter Five, our work of using a natural protein microarray to look for the different of autoimmune response in the sera of pancreatic cancer patients is described. A statistical analysis which catches the outlier pattern is applied to the dataset and the potential biomarker candidates were confirmed with another sample set. The combination of several different biomarkers increased the sensitivity of cancer detection as we expected.

Reference

1. Jemal A, Siegel R, Ward E, Hao YP, Xu JQ, Thun MJ, Ca-a Cancer Journal for Clinicians, 2009, 59, 225
2. Sohn TA, Lillemoe KD, Cameron JL, Huang JJ, Pitt HA, Yeo CJ, Journal of the American College of Surgeons, 1999, 188, 658
3. Cardenes HR, Chiorean EG, DeWitt J, Schmidt M, Loehrer P, Oncologist, 2006, 11, 612
4. D'Onofrio M, Gallotti A, Mucelli RP, Expert Review of Medical Devices, 2010, 7, 257
5. Rosty C, Goggins M, Hematology-Oncology Clinics of North America, 2002, 16, 37
6. Faca VM, Song KS, Wang H, Zhang Q, Krasnoselsky AL, Newcomb LF, Plentz RR, Gurumurthy S, Redston MS, Pitteri SJ, Pereira-Faca SR, Ireton RC, Katayama H, Glukhova V, Phanstiel D, Brenner DE, Anderson MA, Misek D, Scholler N, Urban ND, Barnett MJ, Edelstein C, Goodman GE, Thornquist MD, McIntosh MW, DePinho RA, Bardeesy N, Hanash SM, Plos Medicine, 2008, 5, 953
7. Orchekowski R, Hamelinck D, Li L, Gliwa E, VanBrocklin M, Marrero JA, Woude GFV, Feng ZD, Brand R, Haab BB, Cancer Research, 2005, 65, 11193
8. Zhao J, Patwa TH, Lubman DM, Simeone DM, Current Opinion in Molecular Therapeutics, 2008, 10, 602
9. Yuan HY, Li X, Wu JF, Li JP, Qu XJ, Xu WF, Tang W, Current Medicinal Chemistry, 2008, 15, 470
10. Tripodo C, Florena AM, Macor P, Di Bernardo A, Porcasi R, Guarnotta C, Ingrao S, Zerilli M, Secco E, Todaro M, Tedesco F, Franco V, Current Cancer Drug Targets, 2009, 9, 617
11. Dai Z, Zhou J, Qiu SJ, Liu YK, Fan J, Electrophoresis, 2009, 30, 2957
12. Levitt JM, Slawin KM, Curr Urol Rep, 2007, 8, 269
13. Hogdall E, Current Opinion in Obstetrics & Gynecology, 2008, 20, 4
14. Ferretti G, Felici A, Papaldo P, Fabi A, Cognetti F, Current Opinion in Obstetrics & Gynecology, 2007, 19, 56
15. Ohyama C, International Journal of Clinical Oncology, 2008, 13, 308
16. Meany DL, Zhang Z, Sokoll LJ, Zhang H, Chan DW, Journal of Proteome Research, 2009, 8, 613
17. Arnold JN, Saldova R, Hamid UMA, Rudd PM, Proteomics, 2008, 8, 3284
18. McDowell G, Gahl WA, Proceedings of the Society for Experimental Biology and Medicine, 1997, 215, 145
19. Leroy JG, Pediatric Research, 2006, 60, 643
20. Zhang YJ, Wang JL, Cai Y, Ying WT, Hao YW, Li L, Li XH, Dai SJ, Dai JQ, He FC, Qian XH, Wang J, Wei KH, Chinese Journal of Analytical Chemistry, 2004, 32, 415
21. Wu Y, Mechref Y, Klouckova I, Mayampurath A, Novotny MV, Tang H, Rapid Commun Mass Spectrom, 2010, 24, 965
22. Wada Y, Dell A, Haslam SM, Tissot B, Canis K, Azadi P, Backstrom M, Costello CE, Hansson GC, Hiki Y, Ishihara M, Ito H, Kakehi K, Karlsson N, Hayes CE, Kato K, Kawasaki N, Khoo KH, Kobayashi K, Kolarich D, Kondo A, Lebrilla C,

Nakano M, Narimatsu H, Novak J, Novotny MV, Ohno E, Packer NH, Palaima E, Renfrow MB, Tajiri M, Thomsson KA, Yagi H, Yu SY, Taniguchi N, Molecular & Cellular Proteomics, 2010, 9, 719
23. Ivancic MM, Gadgil HS, Halsall HB, Treuheit MJ, Analytical Biochemistry, 2010, 400, 25
24. Borisov OV, Field M, Ling VT, Harris RJ, Anal Chem, 2009, 81, 9744-54
25. Dalpathado DS, Desaire H, Analyst, 2008, 133, 731
26. Zhao J, Qiu WL, Simeone DM, Lubman DM, Journal of Proteome Research, 2007, 6, 1126
27. Voduc D, Kenney C, Nielsen TO, Seminars in Radiation Oncology, 2008, 18, 89
28. Uttamchandani M, Yao SQ, Current Pharmaceutical Design, 2008, 14, 2428-38
29. Tao SC, Chen CS, Zhu H, Combinatorial Chemistry & High Throughput Screening, 2007, 10, 706
30. Rosenfeld R, Bangio H, Gerwig GJ, Rosenberg R, Aloni R, Cohen Y, Amor Y, Plaschkes I, Kamerling JP, Maya RBY, Journal of Biochemical and Biophysical Methods, 2007, 70, 415
31. Koshi Y, Nakata E, Yamane H, Hamachi I, Journal of the American Chemical Society, 2006, 128, 10413
32. He Q, Zhang Y, Song C-y, Pan Z-c, Wang T-j, Zhang Y-k, Zaho Y-j, Journal of China Medical University, 2009, 38, 1
33. Patwa TH, Zhao J, Anderson MA, Simeone DM, Lubman DM, Analytical Chemistry, 2006, 78, 6411
34. Li C, Simeone DM, Brenner DE, Anderson MA, Shedden KA, Ruffin MT, Lubman DM, Journal of Proteome Research, 2009, 8, 483
35. Chen SM, LaRoche T, Hamelinck D, Bergsma D, Brenner D, Simeone D, Brand RE, Haab BB, Nature Methods, 2007, 4, 437
36. Minor LK, Expert Review of Molecular Diagnostics, 2005, 5, 561
37. Edwards BS, Oprea T, Prossnitz ER, Sklar LA, Current Opinion in Chemical Biology, 2004, 8, 392
38. Saikawa Y, Fukuda K, Takahashi T, Nakamura R, Takeuchi H, Kitagawa Y, Gastric Cancer, 2010, 13, 11
39. Powell AE, Shung CY, Saylor KW, Mullendorf KA, Weiss JB, Wong MH, Stem Cell Research, 2010, 4, 3
40. Jenq RR, van den Brink MRM, Nature Reviews Cancer, 2010, 10, 213
41. Li CW, Heidt DG, Dalerba P, Burant CF, Zhang LJ, Adsay V, Wicha M, Clarke MF, Simeone DM, Cancer Research, 2007, 67, 1030
42. Lee CJ, Li CW, Simeone DM, Translational Oncology, 2008, 1, 14
43. Dosch J, Lee CJ, Simeone DM, Stem Cells and Cancer, 2009, 185
44. Yates JR, Ruse CI, Nakorchevsky A, Annual Review of Biomedical Engineering, 2009, 11, 49
45. Yan F, Subramanian B, Nakeff A, Barder TJ, Parus SJ, Lubman DM, Analytical Chemistry, 2003, 75, 2299
46. Washburn MP, Wolters D, Yates JR, Nature Biotechnology, 2001, 19, 242
47. Strader MB, Tabb DL, Hervey WJ, Pan CL, Hurst GB, Analytical Chemistry, 2006, 78, 125

48. Lee SW, Berger SJ, Martinovic S, Pasa-Tolic L, Anderson GA, Shen YF, Zhao R, Smith RD, Proceedings of the National Academy of Sciences of the United States of America, 2002, 99, 5942
49. Zhou F, Johnston MV, Analytical Chemistry, 2004, 76, 2734
50. Sandra K, Moshir M, D'Hondt F, Tuytten R, Verleysen K, Kas K, Francois I, Sandra P, Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences, 2009, 877, 1019
51. Chen JZ, Balgley BM, DeVoe DL, Lee CS, Analytical Chemistry, 2003, 75, 3145
52. Guo T, Lee CS, Wang WJ, DeVoe DL, Balgley BM, Electrophoresis, 2006, 27, 3523
53. Invernizzi P, Gershwin ME, Journal of Autoimmunity, 2009, 33, 290
54. Pezzilli R, Corinaldesi R, Jop, 2010, 11, 89
55. Pentcheva-Hoang T, Corse E, Allison JP, Immunological Reviews, 2009, 229, 67
56. Holzer AM, Martiniuk F, Levis WR, J Drugs Dermatol, 2007, 6, 393
57. Boyman O, Surh CD, Sprent J, Expert Opinion on Biological Therapy, 2006, 6, 1323
58. Bei R, Masuelli L, Palumbo C, Modesti M, Modesti A, Cancer Letters, 2009, 281, 8
59. Disis ML, Cheever MA, Current Opinion in Immunology, 1996, 8, 637
60. Disis ML, Calenoff E, McLaughlin G, Murphy AE, Chen W, Groner B, Jeschke M, Lydon N, McGlynn E, Livingston RB, Moe R, Cheever MA, Cancer Research, 1994, 54, 16
61. Benmahrez K, Thierry D, Sorokine I, Dannamuller A, Kohiyama M, British Journal of Cancer, 1988, 57, 529
62. Soussi T, Cancer Research, 2000, 60, 1777
63. Gnjatic S, Wheeler C, Ebner M, Ritter E, Murray A, Altorki NK, Ferrara CA, Hepburne-Scott H, Joyce S, Koopman J, McAndrew MB, Workman N, Ritter G, Fallon R, Old LJ, Journal of Immunological Methods, 2009, 341, 50
64. Brichory FM, Misek DE, Yim AM, Krause MC, Giordano TJ, Beer DG, Hanash SM, Proceedings of the National Academy of Sciences of the United States of America, 2001, 98, 9824
65. Raedle J, Oremek G, Welker M, Roth WK, Caspary WF, Zeuzem S, Pancreas, 1996, 13, 241
66. Kotera Y, Fontenot JD, Pecher G, Metzgar RS, Finn OJ, Cancer Research, 1994, 54, 2856
67. Hamanaka Y, Suehiro Y, Fukui M, Shikichi K, Imai K, Hinoda Y, International Journal of Cancer, 2003, 103, 97

Chapter 2

Pancreatic cancer serum detection using a lectin/glyco-antibody array method

Abstract:

Pancreatic cancer is a formidable disease and early detection biomarkers are needed to make inroads into improving the outcomes in these patients. In this project lectin antibody microarrays were utilized to detect unique glycosylation patterns of proteins from serum. Antibodies to four potential glycoprotein markers that were found in previous studies were printed on nitrocellulose coated glass slides and these microarrays were hybridized against patient serum to extract the target glycoproteins. Lectins were then used to detect different glycan structural units on the captured glycoproteins in a sandwich assay format. The biotinylated lectins used to assess differential glycosylation patterns were Aleuria aurentia lectin (AAL), Sambucus nigra bark lectin (SNA), Maackia amurensis lectin II (MAL), Lens culinaris agglutinin (LCA), and Concanavalin A (ConA). Captured glycoproteins were evaluated on the microarray in situ by on-plate digestion and direct analysis using MALDI QIT-TOF mass spectroscopy. Analysis was performed using serum from 89 normal controls, 35 chronic pancreatitis samples, 37 diabetic samples and 22 pancreatic cancer samples. We found that this method had excellent reproducibility as measured by the signal deviation of control blocks as on-slide standard and 41 pairs of pure technical replicates. It was possible to discriminate cancer from the other disease groups and normal samples with high sensitivity and specificity where the response of Alpha-1-β glycoprotein to lectin SNA increased by 69% in the cancer sample compared to the other non-cancer groups (95% confidence interval 53% to 86%). These

data suggest that differential glycosylation patterns detected on high throughput lectin microarrays are a promising biomarker approach for the early detection of pancreatic cancer.

2.1 Introduction:

Pancreatic cancer continues to have a high mortality rate due to detection at a late stage of the disease[1]. In fact, 85% of patients initially present with advanced, non-resectable disease, highlighting the importance of identifying early detection biomarkers. In addition, in a subset of patients, it may be quite difficult to distinguish chronic pancreatitis and pancreatic cancer, necessitating unnecessary surgery in some patients that otherwise might not require it if an adequate biomarker to distinguish these two diseases was available. A serum biomarker test is expected to improve the efficiency of the diagnosis, where the blood contains the unique secretome of the tumor cells. Several serum markers have been investigated for pancreatic cancer. Elevated CA19-9 level has been cited as a potential marker of disease although it generally does not have the specificity or sensitivity for general screening[2-8]. It has been frequently utilized as a marker to monitor a patient's progress after surgery[9]. Other existing biomarkers relate to the inflammation that associates with the tumor and other pancreatic diseases that may be present[10-12]. It should be noted that no individual biomarker has been found to be conclusive at diagnosis to distinguish chronic pancreatitis and pancreatic cancer[13,14]. To our knowledge, there is no study comparing the serum of pancreatic cancer and diabetes which is a widely existing disease in patients at risk of pancreatic cancer. Discovery of new early detection biomarkers that are specific for pancreatic cancer remains a major challenge.

Post translational modification of the proteome in serum analysis has become an important area in biomarker research[15]. Of particular interest is the study of glycoproteins where unique protein glycosylation patterns are associated with cancer[16-25]. Glycans are involved in many biological processes including protein-protein interactions, protein folding, immune recognition, cell adhesion and inter-cellular signaling[26]. Alteration of glycan structure and coverage on several major glycoproteins in serum has been shown to contribute to the progression of cancer. In previous work, fucosylated haptoglobin was suggested as a biomarker for early detection of pancreatic cancer[27]. Also the glycoforms of alpha-1-acid glycoprotein have been found to vary in cancer patients compared to the healthy controls[28]. These biomarkers can be used to improve the confidence of the diagnosis through identification of disease-related glycan structures by various separation and mass spectrometry techniques[29-32]. In one such study using lectin extraction and mass spec analysis the glycosylated isoforms of alpha-antitrypsin were shown to change in cancer compared to normal samples or pancreatitis[33]. Other studies have removed the glycan groups from the glycoprotein content of the cell and used glycan profiling to show distinct differences between cancer and normal samples based on changes in carbohydrate structures in serum, although association with a particular protein is lost[34]. In other studies hydrazide columns have been used to extract glycoproteins from serum which were digested and analyzed by LC-MS/MS. In this report glycoproteins associated with cancer were found although the actual glycan structural information was not delineated[35].

Recently, various microarray formats have been utilized for studying glycosylation patterns. In one study examining sera samples from patients with colon and

pancreatic cancers, glycoproteins extracted from serum were printed on glass slides and hybridized against various lectins to study changes in the glycan patterns during cancer progression[36,37]. This method provides a means of studying subtle changes in glycan structure and is an excellent discovery platform but does not provide a high throughput mode for further validation. Other methods have included the use of glycan arrays where glycans are directly printed on glass slides[38] or alternatively lectin arrays where lectins are printed on a slide and glycoproteins or whole cells hybridized against them. The lectin array approach has been used to identify differences in glycoprotein surface markers for cancer cells compared to normal cells and between different types and stages of cancer in several studies[39,40]. Alternatively an antibody array approach has been used to capture proteins from serum and a lectin hybridized against the glycoprotein to study changes in glycan structure[41]. This method can screen large numbers of samples from serum for such changes but requires a discovery platform to choose the antibodies on the array for screening.

The antibody microarray is a favorable format for high throughput analysis, with a high level of specificity and reproducibility[42-44]. In the present study, we selected four glycoproteins as our target proteins for the antibody microarray. Three of them (Amyloid p component, Antithrombin and Kininogen) scored the first three highest Z values in previous study in which around ten potential biomarkers were found differentially expressed in the sera of pancreatic cancer and cancer-free (including pancreatitis) patients [36]. The other glycoprotein (A1BG) was reported in other work on urine sample analyzing bladder cancer patients [45]. Antibodies to these glycoprotein biomarkers were printed on nitrocellulose coated glass slides. The glycans on the printed

antibodies were first blocked to eliminate their interference in the hybridization with lectins. The target proteins in the serum were then captured on the antibody array and probed with several biotinylated lectins where streptavidinylated fluorescent dyes were used for detection. Eighty nine samples from normal controls, 35 chronic pancreatitis samples, 37 diabetics samples and 22 pancreatic cancer samples were processed using this method where non-cancer samples were randomly selected and all cancer samples available were used. Antibody specificity was verified by on-target digestion of the captured glycoproteins with subsequent on-slide MALDI-MS identification. The data was subjected to statistical analysis to display the variation for a single patient and the differentiation among the disease groups.

2.2 Experimental

2.2.1 Serum samples-

Inclusion criteria for the study included patients with a confirmed diagnosis of pancreatic cancer, chronic pancreatitis, long-term (for 10 or more years) Type II diabetes mellitus, or healthy adults with the ability to provide written, informed consent, and provide 40 ml of blood. Exclusion criteria included inability to provide informed consent, patients' actively undergoing chemotherapy or radiation therapy for pancreatic cancer, and patients with other malignancies diagnosed or treated within the last 5 years. The sera samples were obtained from patients with a confirmed diagnosis of pancreatic adenocarcinoma who were seen in the Multidisciplinary Pancreatic Tumor Clinic at the University of Michigan Comprehensive Cancer Center. All cancer sera samples used in this study were obtained from patients with stages III/IV pancreatic cancer. The mean age of the tumor group was 65.4 years (range 54-74 years). The sera from the normal, pancreatitis, and

diabetes groups was age and sex-matched to the tumor group. The chronic pancreatitis group was sampled when there were no symptoms of acute flare of their disease. All sera were processed using identical procedures. The samples were permitted to sit at room temperature for a minimum of 30 minutes (and a maximum of 60 minutes) to allow the clot to form in the red top tubes, and then centrifuged at 1,300 x g at 4^{o}C for 20 minutes. The serum was removed, transferred to polypropylene, capped tubes in 1 ml aliquots, and frozen. The frozen samples were stored at -70^{o}C until assayed. All serum samples were labeled with a unique identifier to protect the confidentiality of the patient. None of the samples were thawed more than twice before analysis. This study was approved by the Institutional Review Board for the University of Michigan Medical School.

2.2.2 Microarray preparation and serum hybridization

Alpha-1-β glycoprotein antibody was purchased from Novus, while Amyloid p component antibody and Antithrombin antibody were from Abcam. All the antibodies are monoclonal, raised in mouse, targeting human proteins. Antibodies were diluted to 50ug/mL in PBS and spotted on ultra-thin nitrocellulose coated slides (PATH slides, GenTel Bioscience) with a piezoelectric non-contact printer (Nano plotter; GESIM). Each spotting event that resulted in 500 pL of sample being deposited was programmed to occur 5 times/spot to ensure that 2.5 nL was being spotted per sample. The spots used by the MALDI-MS experiment were printed 50 times. Each antibody was printed in triplicate. The spot diameters were 280 um and 700 um for the spots that were printed 5 times and 50 times respectively. The spacing between the spots was 0.7mm. 14 blocks were printed on each slide in a 2X7 format and the block distance was 9.4mm.

Figure 1 presents an experimental flow chart of the microarray processing and on-target digestion for MALDI-MS. The antibody arrays on the slides were first chemically derivatized with a method similar to previous work [22] but modified for this work. The printed slides were dried in an oven at 30°C for 1h before gently being washed with PBST 0.1 (100% PBS with 0.1% tween 20) and coupling buffer (0.02M sodium acetate, pH 5.5), and then oxidized by 200 mM $NaIO_4$(Sigma) solution at 4C in the dark. After 3 hours the slides were removed from the oxidizing solution and rinsed with coupling buffer. The slides were immersed in 1 mM 4-(4-N-maleimidophenyl)butyric acid hydrazide hydrochloride (MPBH; Pierce Biotechnology) at room temperature for 2 hours to derivatize the carbonyl groups. 1 mM Cys-Gly dipeptide(Sigma) was incubated with the antibodies on the slides at 4°C overnight. The slides were blocked with 1% BSA for 1 hour and dried by centrifugation.

The slides were inserted into the SIMplex (Gentel) 16 Multi-Array device which separates the blocks and prevents cross contamination when different samples are applied on neighboring wells. Serum samples were diluted 10 times with PBST 0.1 containing 0.1% Brij. 100 uL of each sample was applied to the antibody array manually and left in a humidified chamber for 1 hour to prevent evaporation. Slides were rinsed with PBST 0.1 for 3 times to remove unbound proteins. The arrays were then treated with different detection biotinylated lectins (Vector Laboratory) to determine lectin response and streptavidinylated fluorescent dye (Alexa555; Invitrogen Biotechnology) was used for detection. After a final wash, the slides were dried and scanned with a microarray scanner (Genepix 4000A; Axon). The program Genepix Pro 6.0 was used to extract the numerical data. A threshold of signal to background ratio was set at 10 and less than 1%

of the spots were under this threshold and excluded. The mean of the intensity in each spot was taken as a single data point into analysis.

2.2.3 On-target digestion and MALDI-QIT-TOF

The microarray slides were incubated with 0.5 M lactose for 10 min and washed with PBST 0.1 to remove the captured lectin from the glycoprotein. After an additional wash with water the slides were dried with centrifugation. Trypsin was diluted to 50 ng/uL with 50 mM ammonium bicarbonate and printed on the microarray spots. The printed slides were moved into a humidity chamber and incubated at 37˚ C for 5min. 35 mg/mL 2,5-dihydroxybenzoic acid (DHB) (LaserBio Labs, France) in 50% acetonitrile was printed on the microarray by the microarray printer and allowed to dry.

Mass spectrometric analysis of the microarray slides was performed using the Axima quadrupole ion trap-TOF (MALDI-QIT) (Shimadzu Biotech, Manchester, UK). The microarray slide was analyzed directly by taping the slide onto the stainless steel MALDI plate and inserting it into the instrument for analysis. Acquisition and data processing were controlled by Launchpad software (Kratos, Manchester, UK). A pulsed N2 laser light (337 nm) with a pulse rate of 5 Hz was used for ionization. Each profile resulted from 2 laser shots. Argon was used as the collision gas for CID and helium was used for cooling the trapped ions. TOF was externally calibrated using 500 fmol/µL of bradykinin fragment 1-7 (757.40 m/z), angiotensin II (1046.54 m/z), P14R (1533.86 m/z), and ACTH (2465.20 m/z) (Sigma-Aldrich). The mass accuracy of the measurement under these conditions was 50 ppm.

2.3 Results and Discussion:

2.3.1 Optimization of serum concentration:

The antibodies were printed on ultrathin nitrocellulose slides and hybridized with serum in a 14 multi-array device, then visualized with biotinylated lectin and Alexafluor-555. In a reproducibility test, a common sample selected at random was applied to all 14 arrays. Lectin SNA was used for detection. The fluorescent image of the slide is shown in Figure 2a to illustrate the quality of the printed spots. There is a very low variation of the signal between different arrays over the slides. The intensity of the signal in each block was calculated as shown in Figure 2b. The standard deviations of the signal of all three antibodies within the slides were about 5% of the average.

In order to estimate the variation of the signal on different slides, 2 blocks on each slide were hybridized with the same two samples. The signals of these two blocks were compared across slides to calculate the variation. In multiple–slide experiments, the value of the slide to slide variation was about 10% of the average signal.

Different dilutions of serum were tested to determine the optimum concentration of the target glycoproteins. The same lectin SNA was used for detection in this test. There were seven dilutions of serum sample from 2 to 600 times dilution that were applied to the arrays. Figure 3 depicts how the intensity of the signal changes for the three antibodies with a decreasing dilution fold. A rising trend was noted from the 600X dilution to the 50X dilution for the three glycoproteins shown. In the 50X dilution to the 20X dilution the signal was relatively unchanged except for Antithrombin-III, where the signal increased 20% from the 50X dilution to the 20X. The signal remained the same from the 20X dilution until it reached the 5X dilution, where a saturation of the signal has occurred. A decrease of signal for all three glycoproteins from the 5X dilution to the 2X

dilution of serum sample can be seen in the figure 3, likely due to competing non-specific binding on the antibodies.

The result of the dilution test demonstrated that the antibodies were saturated by their target protein at 20X dilution or above in the process of hybridization (1hour, room temperature and gentle shaking). Below 50X dilution the antibodies were not completely occupied, so the signal decreased with additional dilution. The nonlinear relation between the concentration of the serum and the intensity of the signal could be attributed to various factors that may affect the antibody-antigen reaction, including accessibility of the antibodies, diffusion rate and solubility of the antigen in the hybridization buffer. Nonspecific binding on the antibodies was also considered as a possibility, but was further investigated and excluded by on-target digestion and MALDI-MS analysis.

To analyze the difference of the glycosylation on potential biomarker proteins, protein expression levels must be normalized. The protein level was estimated by antibody assay. In our experiment the three potential biomarkers were all relatively high abundance proteins in human serum (concentration >20mg/L) which could easily saturate the antibodies printed on the microarray. Under saturation conditions, the amount of target biomarkers captured on the antibody spots was equal to the capacity of the printed antibody which should be the same in all the replicate blocks. As a result, the need for protein assay was avoided and the intensity of the signal on the microarray directly represented the level of glycosylation.

2.3.2 Antibody specificity test with MALDI-QIT-TOF

In order to validate the specificity of the antibodies we performed on-target digestion and MALDI-QIT-TOF of the spots after elution of biotinylated lectins captured on the

glycoproteins with a concentrated sugar solution. A trypsin solution with 50mM ammonium bicarbonate was printed with the microarray printer using the same spot lay-out as in the antibody printing. The volume of the trypsin solution was 4nL which in a humidity chamber lasts about 5 minutes before drying out. Ammonium bicarbonate usually decomposes at the same time. 2,5-dihydroxybenzoic acid was then dissolved in 50% acetonitrile and printed on the digested spots. The matrix solution itself is very acidic and stops the digestion to prevent further digestion of antibodies and trypsin autolysis. Acetonitrile also partially dissolved the nitrocellulose film and the digested peptides on the film were extracted and mixed with matrix. Nitrocellulose film has been reported as an excellent substrate for MALDI-MS[46]. The presence of nitrocellulose in the mixture did not affect the crystallization of DHB.

The specificity (specific binding vs. non-specific binding) of the antibody as a function of the dilution times of the serum can be determined by comparing the spectrum from the arrays processed with different conditions. In the experiment one control array (incubated with blocking buffer) and two sample arrays which were hybridized with 2X and 10X dilution of the same serum were tested. The presented figures are the spectra of Amyloid p component antibody spot. Figure 4a shows the spectrum of the Amyloid p component spot in the control array which only contained the antibody(anti human Amyloid p component). All the peaks in the spectrum are the peptides digested from the antibody and the enzyme itself. The top 3 peaks are attributed to the antibody digest. The intensity of the other peaks was too low to be identified. The spectra in figures 4b and 4c are generated from the Amyloid p component spots in the sample arrays. In the mass spectrum of 10X dilution, 3 new peaks appeared which were all identified by

MS/MS to be tryptic peptides of Amyloid p component. This result indicated that no other protein was captured on the antibody or the amount was too low to be detected. In the case of the 2X dilution, 2 additional peaks emerged in the spectrum where one of them was identified as human serum albumin while the other one was not identified. The extra peaks are a sign of nonspecific binding on the antibody. Thus, only when the concentration of the sample was increased to 2X the dilution of the serum does non-specific binding begin to affect the specificity of the antibody.

2.3.3 Detecting glycosylation on captured protein by blocked antibody arrays

The chemical derivatization method was employed to block the glycans on the antibodies to eliminate their binding with the lectins used for detection of glycoproteins[41]. The cis-diol groups on the glycans were gently oxidized and converted to aldehyde groups which then reacted with hydrazide-maleimide bifunctional cross-linking reagent and capped with a Cys-Gly dipeptide. After the derivatization reaction the lectins could not recognize the modified oligosaccharide group.

All the antibodies were tested against several samples and lectins to evaluate the effectiveness of the protocol. The underivatized antibodies responded to some of the lectins, but after derivatization the binding greatly decreased or disappeared. The serum solution was incubated against the derivatized antibody array where the spots showed lectin binding on proteins captured by the antibodies, indicating that the antibodies maintained their function after derivatization.

2.3.4 Characterizing glycan structure of potential biomarkers with different lectins

A previous study described ten potential biomarkers in the sera of normal and cancer patients that significantly changed their response to several lectins[18]. We chose four of

these target proteins (Antithrombin-III, Amyloid p component, alpha-1-β glycoprotein and kininogen) as a proof of concept to determine the proteins which provided the best discrimination of samples from patients in different groups based on lectin response. The biotinylated lectins used were Aleuria aurentia lectin (AAL), Sambucus nigra bark lectin (SNA), Maackia amurensis lectin II (MAL), Lens culinaris agglutinin (LCA), and Concanavalin A (ConA). AAL and LCA bind fucose linked (α-1,6) to N-acetylglucosamine or (α -1,3) to N-acetyllactosamine related structures. Both MAL and SNA recognize sialic acid on the terminal branches. MAL detects glycans containing NeuAc-Gal-GlcNAc with sialic acid at the 3 position of galactose while SNA binds preferentially to sialic acid attached to terminal galactose in an (α-2,6) and an (α-2,3) linkage to a lesser degree. ConA recognizes α-linked mannose including high-mannose-type and hybrid-type structures. These lectins were selected since fucosylation and sialylation have been shown to be related to cancer development[27,33] and ConA binds to almost all the N-linked glycoproteins where its signal translates into a general level of glycosylation.

Figure 5 shows the result of an initial test of the four antibodies with five lectins. The contrast and brightness in the images were optimized to differentiate the three groups. The borders were drawn by hydrophobic marker pens to prevent the cross contamination between the blocks. The lectins used were AAL, LCA, SNA, MAL and ConA. In each of the images, three normal samples were hybrized in the first column of arrays, three pancreatitis samples were in the second column and three cancer samples were incubated in the third column of arrays. In each array, antibodies were printed in different columns in the order of A1BG, Kininogen, Amyloid p component and Antithrombin-III. For LCA,

AAL, SNA and MAL the three cancer samples all showed a stronger response than the pancreatitis and normal samples, whereas the blocks probed with ConA showed equal signal in the three groups. A binding pattern was shared between LCA and AAL, which agreed with their same specificity on fucosylated N-linked glycans, though the signal of LCA was lower in intensity. These lectins were found to preferentially distinguish normal and chronic pancreatitis samples from cancer samples (data not shown). MAL was not used for subsequent analysis due to its low sensitivity with these antibodies. Of the 4 antibodies, 3 of them (A1BG, Amyloid p component and Antithrombin-III) displayed a signal-to-background ratio of higher than 20, and were chosen for large set analysis.

2.3.5 High throughput analysis and data quality test

183 samples from patients with various genders, fasting status and disease classes were processed in 4 batches on 16 slides. Since the signal to background ratio for all the valid spots were higher than 10, the signals were directly used for analysis without taking into account the background. 41 of the patients in the groups of normal, chronic pancreatitis and diabetics contributed three samples with two samples collected twice under fasting conditions and the other sample was collected under non-fasting conditions. The lectin used in this study is SNA. Two patients provided only double fasting samples which are used for the data quality test. For the other samples including some of the normal, pancreatitis and all the cancer patients, the information of the gender and fasting status is not available After adjusting for fasting status, gender, and disease category, the data points were compared to a normal reference distribution (figure not shown). Based on

this comparison, two outlying data points from the antibody of Antithrombin-III were excluded from all subsequent analysis.

The accuracy of the antibody microarray analysis is heavily dependent on the reproducibility of the technique which is also used as a means to filter out unreliable antibodies in distinguishing cancer from other disease classes. Reproducibility is assessed by fitting a linear mixed effects model to log2 scale expression data, separately for each antibody. This is a type of ANOVA model in which fixed effects for fasting status, gender, and disease category are included along with random effects for patients, and batches within patients. Thus the expression variation for every antibody around the mean for its fasting/gender/disease group is described in terms of three variance components (residual, patient and batch within patient), which characterize the reproducibility at different levels of the experiment. . Residual variance represents variation for technical replicates (same person, batch, and fasting status). Batch variance represents technical variation for the same person and fasting status across batches. Patient variance represents stable biological variation across people. Table 1 shows the three variance components on the standard deviation scale, for the three antibodies. For example, the residual SD for A1BG is 0.21, which means that two thirds of the replicates will lie within (2^0.21-1) x 100% = 16% of the true values and 95% of the replicates will lie within (2^0.42-1) x 100% = 34% of the true value. Alternatively, the reproducibility could be exhibited by the correlation of the replicate spots in log2 scale which is presented in Figure 5. The scatter plots suggest that the technical error is not limited to a handful of outliers, consistent with our finding of an approximately normal distribution

of residual variance, as discussed above. Figure 5 shows data for all non-cancer patients and antibodies pooled.

2.3.6 Examination of potential bias

Sex, fasting status and other related diseases could all possibly become the source of bias in biomarker validation.[47] As discussed above, linear mixed effects models were built separately for each of the three antibodies, with these potentially biasing factors modeled as fixed effects. As always, one level of each factor variable is omitted, so the implicit fixed effect for a normal, non-fasting female is zero, and all other factor settings are interpreted as deviations from this arbitrary baseline setting. The results are listed in Table 2. Fixed effect estimates are shown along with the standard error. The T- Since log2 scale data are analyzed, we also converted the effects to the raw scale. For A1BG the factors have small and non-significant effects. For Amyloid there is a significant effect for fasting, and for Antithrombin-III there is a significant effect for sex and disease (mainly due to pancreatitis). These effects are statistically significant but are small in magnitude relative to the residual and patient variation, and to the response in cancer.

2.3.7 Antibody performance in distinguishing cancer and non-cancers

Table 3 provides information concerning the discrimination between cancer and non-cancer samples. These estimates are obtained through the same linear mixed effects model discussed above. A fixed effect estimate b on the log2 scale implies a percent change of (2^b-1)x100% on the raw scale, where negative and positive percent changes correspond to decreased, and increased expression, respectively. The A1BG signal increases by 69% %=(2^0.76-1)*100% in cancer samples compared to normal, chronic pancreatitis, and diabetic samples. The Amyloid signal increases 33%, and Antithrombin-

III is essentially unchanged. The total standard deviation from technical and biological variation (within disease classes) is around 0.32= sqrt(0.216^2 + 0.207^2 + 0.127^2) for A1BG, summing the variance components from table 1.. Thus approximately 95% of the time, the experimental result for a particular non-cancer sample is expected to fall between 35% below the mean level or 56% above the mean level for all non-cancer samples ((2^-0.64-1)*100=-35 and (2^0.64-1)*100=56, where 0.64=2*0.32 is two standard deviations. Thus the 69% average increase in A1BG associated with pancreatic cancer falls well outside the range of normal variation... ROC curves in Figure 6 were also constructed for each of the three markers, based on their ability to distinguish pancreatic cancer from non-cancer samples (a pool of normals, pancreatitis, and diabetes). All three markers show some discrimination where only A1BG is potentially useful on its own. A1BG distinguished cancer and non-cancer samples with a 100% sensitivity and a 98% specificity. The AUC value measuring the area under the ROC curve for A1BG is 0.998. For Amyloid p component the cancer samples were distinguished from non-cancer samples with 88% sensitivity and 68% specificity and its AUC value is 0.875. The discrimination for Antithrombin-III is due to the differences between cancer and pancreatitis and it would be unable to distinguish cancer from diabetes based on these data. According to the scatter plot in Figure 7 where the signals of A1BG and Amyloid were used as X and Y axes, the overlap of the cancer samples with the non-cancer groups is around 20%. The extent of the difference in 4 patient groups is also shown in Figure 8 which depicts the distribution of the measurement for the antibody A1BG. The boxplot provides the upper and lower quartiles of the measurements with respect to the median

value (red line in the middle of each box). The lines provide the ranges of the measurements, excluding outliers (+).

The use of the antibody microarray to capture potential biomarkers available in cancer serum provides a means for high throughput and analysis of glycosylation patterns. Because of the specific goal of quantifying the glycans in this study, antibodies were saturated with the analytes by optimizing the dilution times of the sera according to the saturation curve. Thus the response of the lectin from the microarray directly represented the level of the particular glycosylation without concern about the various concentrations of the proteins in different samples. This strategy also defined the sensitive steps in the experiment where the serum was aliquoted, diluted and hybridized with the microarray, while in other applications of antibody microarrays, factors such as precipitation, heterogeneity of the serum and conditions in hybridization may vary and lead to bias in the method.

In this study antibody specificity was confirmed by direct MALDI-MS of the microarray spots. Traditional immunoblotting is based on the same interaction as in the antibody microarray and does not exclude undesirable binding. MALDI-MS can identify the tryptic peptides of any captured abundant protein on the target. The microarray printer was essential in precisely depositing the extremely small amount of enzyme and matrix on top of the antibody spots [48]. In this experiment, the nitrocellulose surface generated high quality mass spectra. In spite of peaks from the antibody that dominated the mass spectra, target proteins were readily identified and non-specific binding was also found when the serum was not sufficiently diluted.

To access the technical error of the assay, a comprehensive reproducibility test was applied by using two fasting samples from the same patients (drawn at two times) as pure technical replicates. The samples were disordered before being incubated on the antibody arrays. In most other duplicate studies[42-45], variations from an entire slide or batch was more likely to be detected while the individual variability of the single blocks within the slide and batch were ignored. In this work, by statistically comparing the pairs of technical replicates that were distributed across the slides, we were able to evaluate the divergence of the signal from the ideal value that resulted from the technical error.

The pancreatic cancer samples could be clearly distinguished from other disease states and normal samples. The ROC curves showed that Alpha-1-β glycoprotein response to SNA resulted in specific detection of pancreatic cancer with high sensitivity and specificity. A combined ROC curve of Alpha-1-β glycoprotein and amyloid did not provide any improvement in discrimination. However, a limitation of the analysis was that the cancer samples studies were all from late stage pancreatic cancer, thus the study served as a proof of principle experiment. To have an impact on the morbidity and mortality associated with pancreatic cancer, we will need to examine an adequately sized test and training set of early stage pancreatic cancer samples and compare results with appropriate age and sex matched normals and disease controls (i.e chronic pancreatitis and diabetics). Such a serum set is currently being collected as part of a multi-institution collaborative effort by the Early Detection Research Network (EDRN).

2.4 Conclusion:

In this work, an antibody/glycoprotein/lectin sandwich assay was developed for screening potential markers of pancreatic cancer. These markers were chosen for study based upon our previous work using a lectin glycoarray approach. Three potential markers were chosen and their corresponding antibodies were printed on coated glass slides. They were exposed to sera from 89 normal samples, 35 chronic pancreatitis samples, 37 diabetic samples and 22 pancreatic cancer samples. The captured glycoproteins were analyzed against four different lectins where SNA was found to provide the best results.

Further, MALDI QIT-TOF MS was used for direct analysis of the captured glycoproteins to optimize dilution conditions of the serum and for minimizing nonspecific binding. It was shown that the pancreatic cancer samples could be clearly distinguished from other disease states and normal samples. The ROC curves showed that Alpha-1-β glycoprotein response to SNA resulted in specific detection of pancreatic cancer with high sensitivity and specificity. The resulting scatterplots also showed the ability to clearly distinguish pancreatic cancer from chronic pancreatitis, diabetics or normal samples. The protein Amyloid also showed an acceptable ability to discriminate pancreatic cancer according to the ROC curve whereas Antithrombin-III could not provide such discrimination. A combined ROC curve of Alpha-1-β glycoprotein and Amyloid did not provide any improvement in discrimination due to correlation between the two markers.

Figures

Figure 2.1 An outline of the experimental flow of microarray processing and on-target digestion.

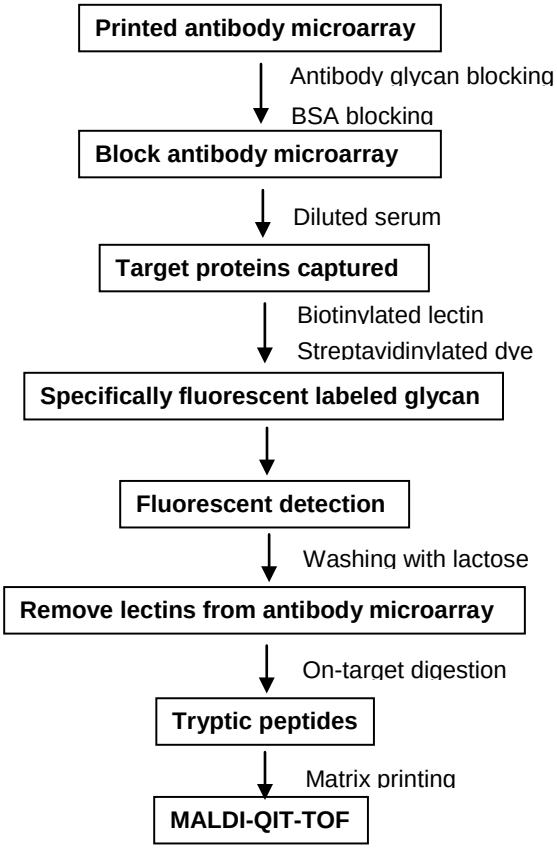
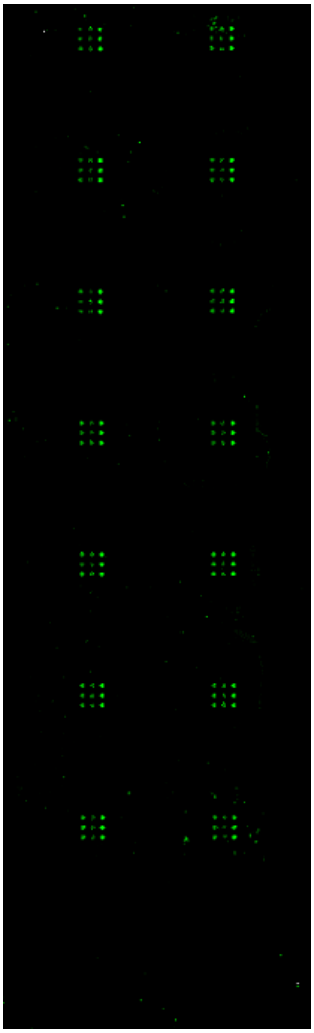
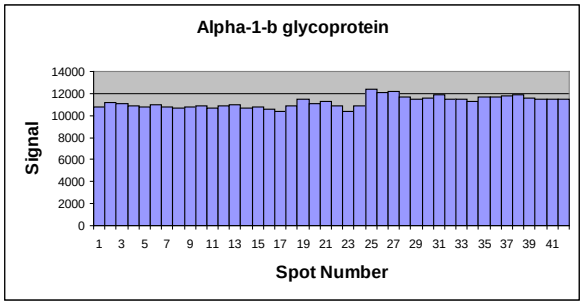
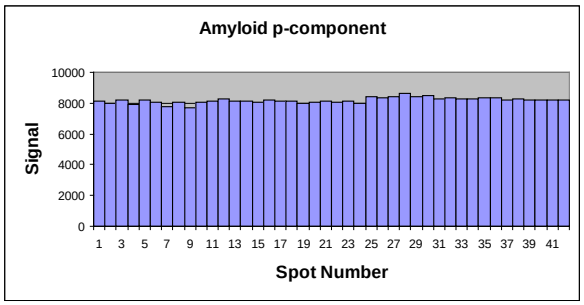


Figure 2.2. Block-block variation test. **a** The fluorescent image of a microarray slide with 14 identical antibody arrays, the antibodies in the three columns of each block are A1BG, SAMP, Antithrombin III. SNA lectin was used for detection; **b** The intensity of the fluorescent signals for each of the spots of the same antibodies in the slide.







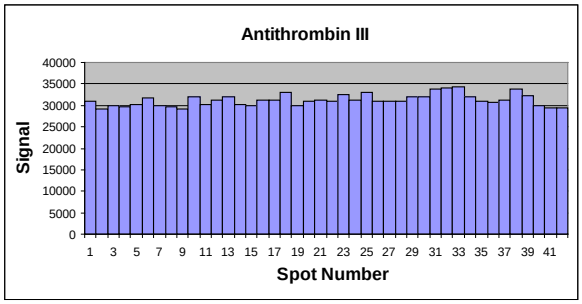


Figure 2.3. Saturation curve which shows the fluorescence intensity of the three antibodies when they are exposed to different dilutions of serum with SNA lectin detection. The numbers on X-axis are the folds that the serum has been diluted before hybridized with the antibody array. The y-axis is the intensity of the signal.

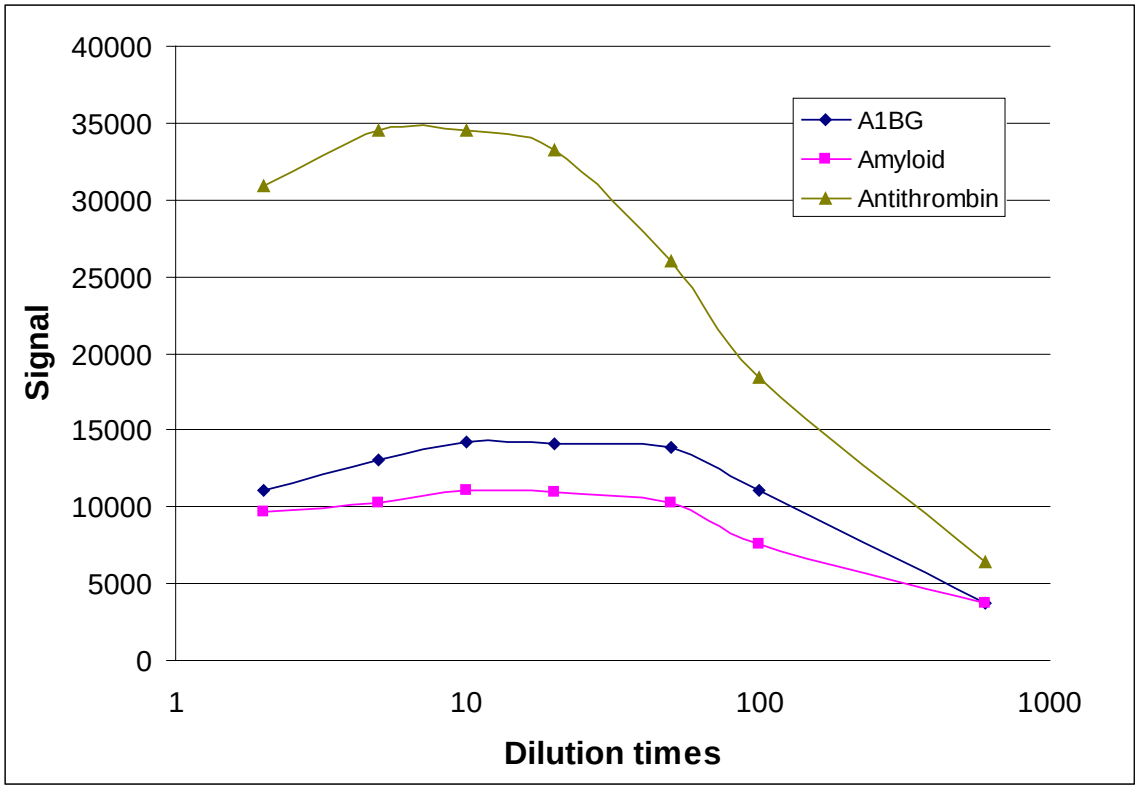


Figure 2.4. The MALDI-MS spectra generated on the microarray spots of SAMP antibody after on-target digestion. The peaks assigned as Amyloid p component were marked with green arrows where the unassigned peaks appearing in **c** were marked with black arrows. **a** control spot, without incubation of serum**; b** incubated with10X diluted serum; **c** incubated with 2X diluted serum.

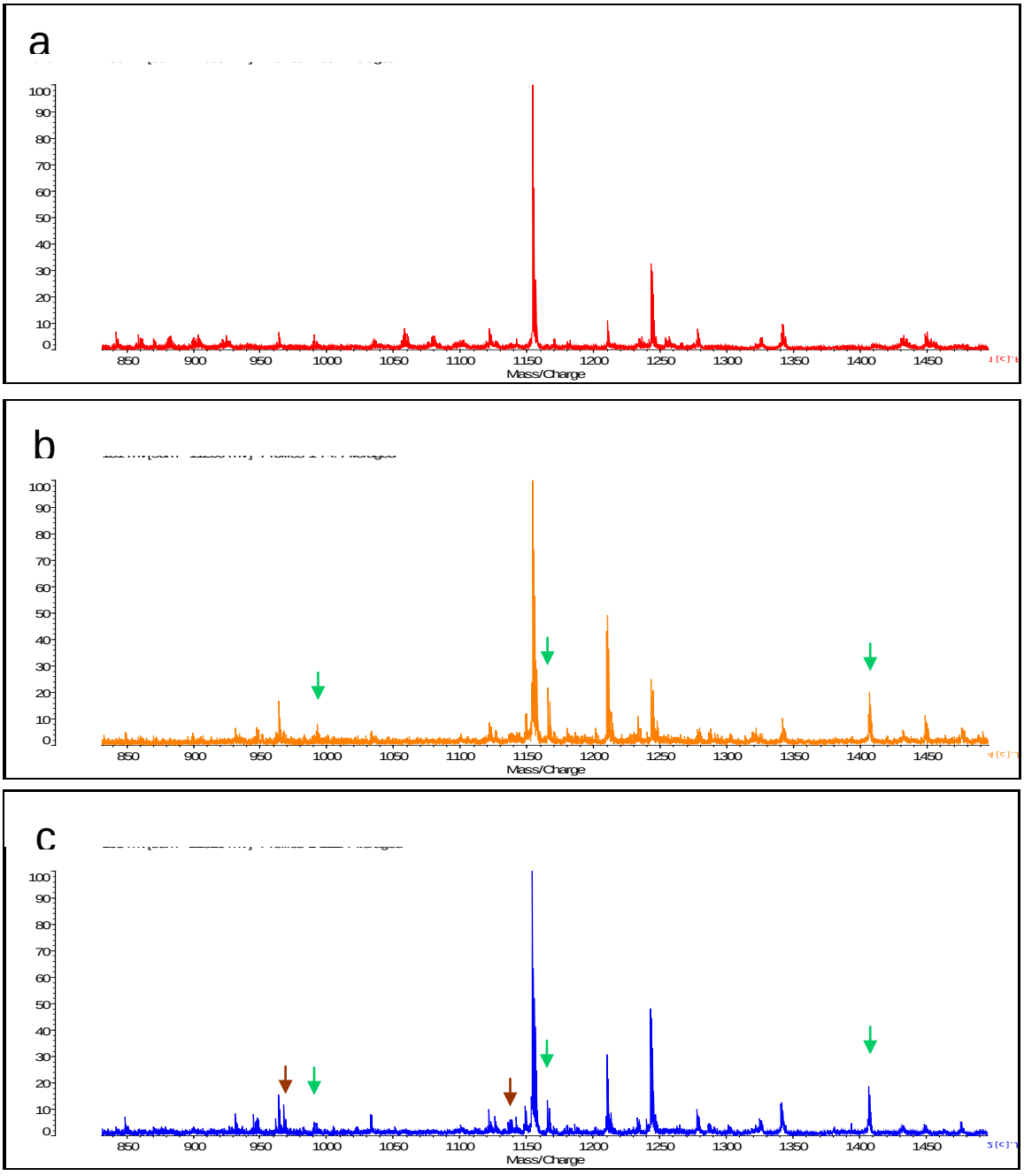


Figure 2.5 Scatter plot in log2 scale between every pair of technical replicates (a replicate is two distinct points same patient, same antibody, same fasting status and same batch) i.e. for any spot in the figure, its values on the aisles are the log2 intensities of any pair of replicate spots. The lectin SNA is used.

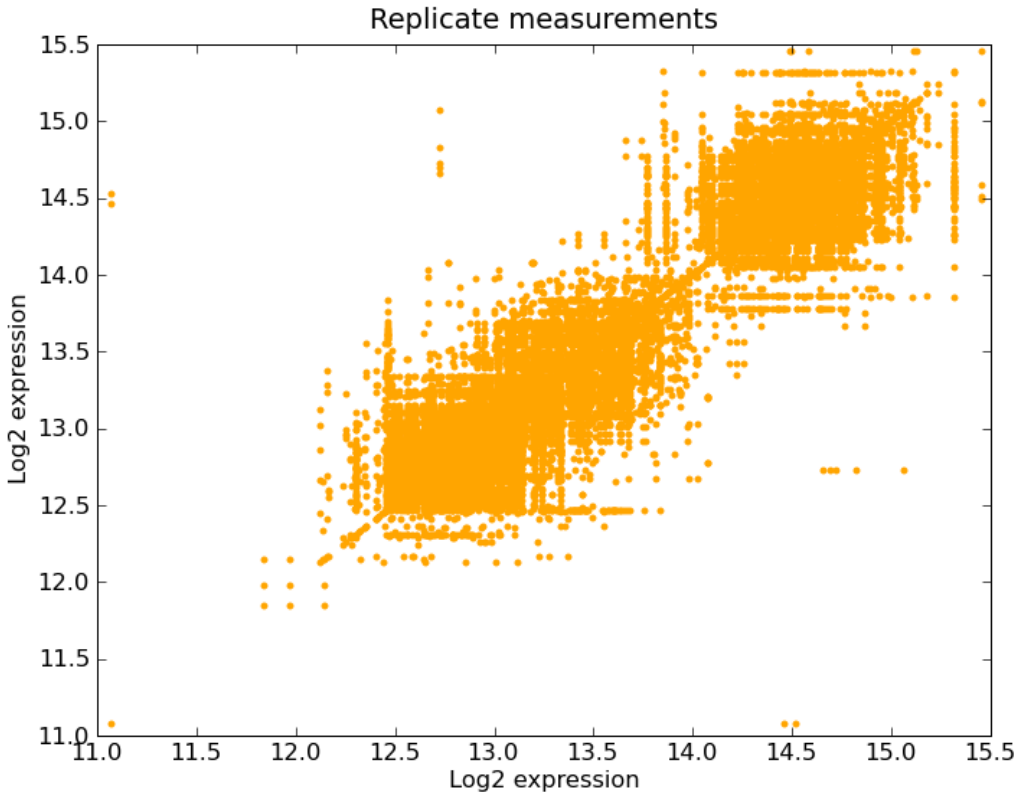


Figure 2.6 ROC curves for the three antibodies alone and A1BG and Amyloid combined.

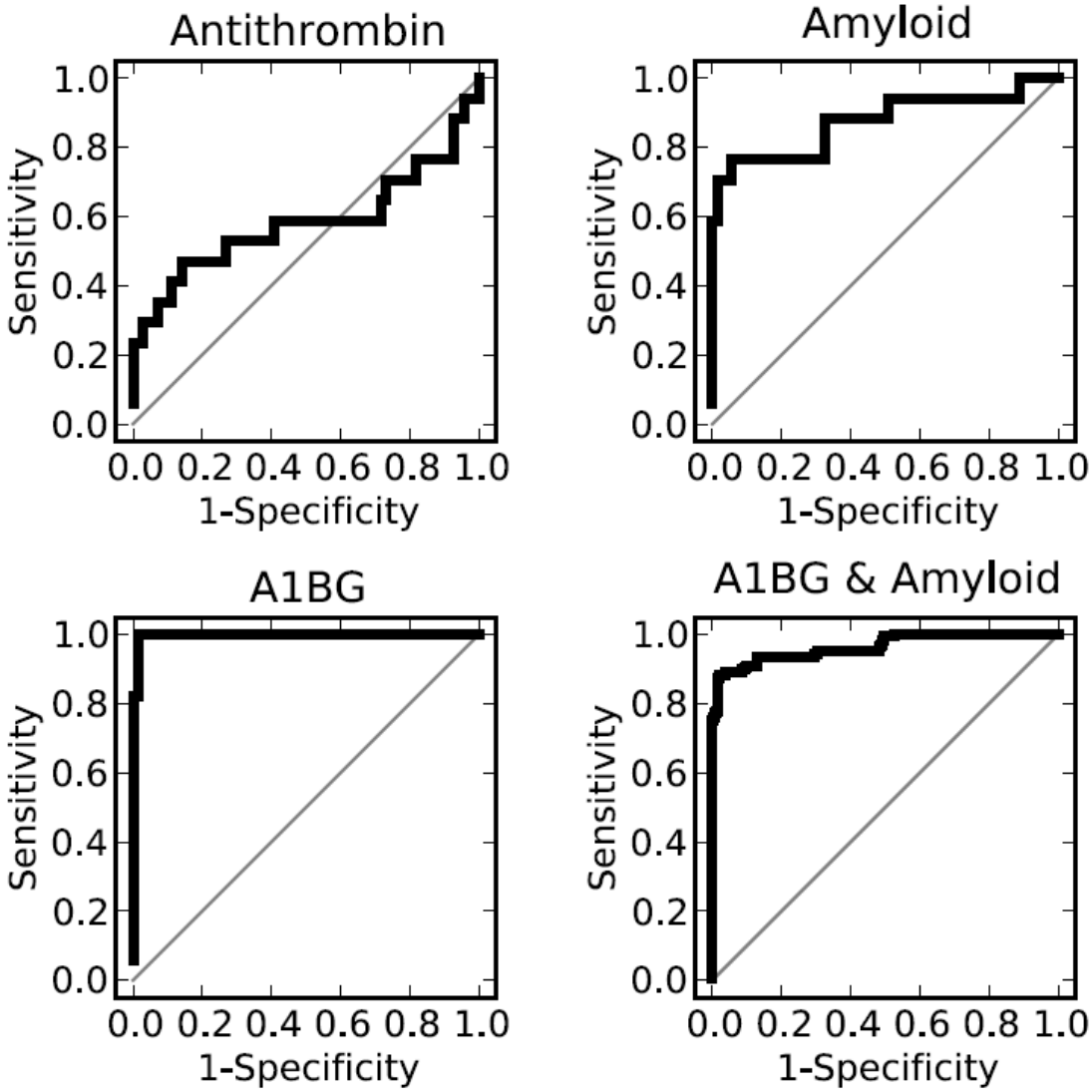


Figure 2.7 Scatter plot of sialylation level detected by lectin SNA on A1BG and Amyloid p component. Samples from different disease classes are differently colored. Normal spots are orange, pancreatitis spots are blue, diabetes spots are pink and cancer spots are dark blue.

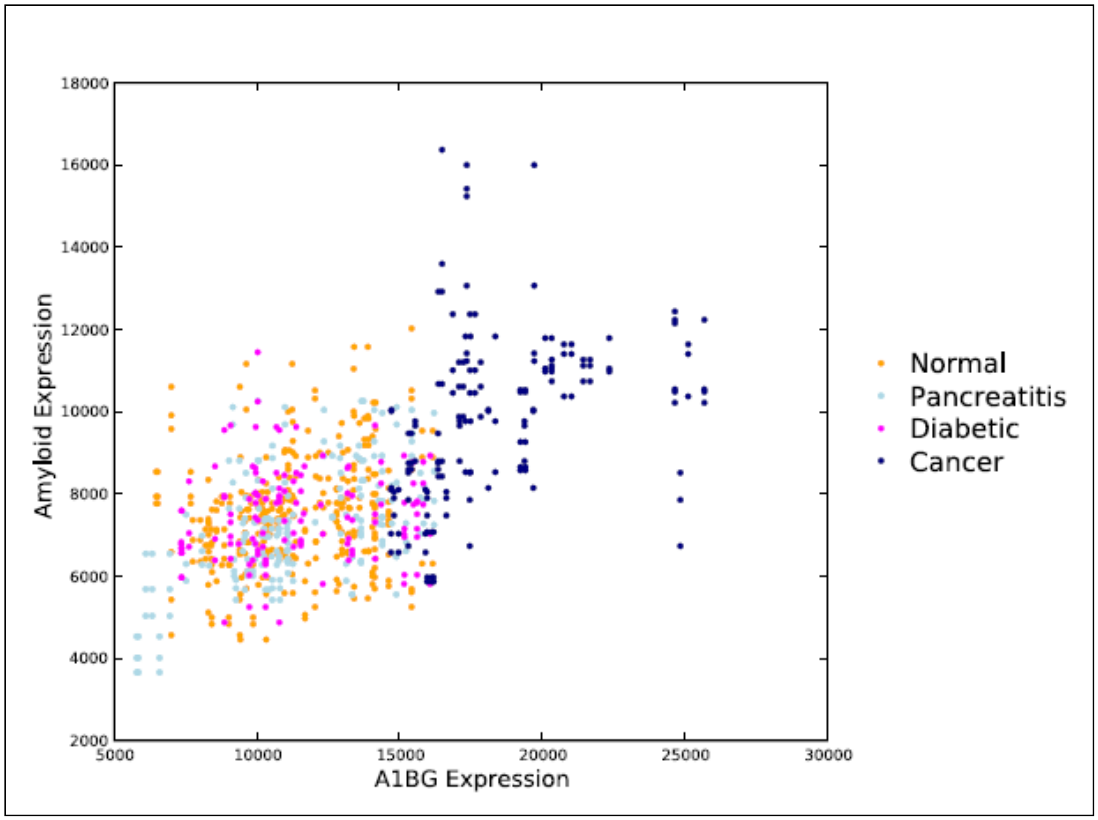


Figure 2.8 The boxplot depicts the distribution of the measurements for antibody A1BG. The boxes provide the upper and lower quartiles of the measurements with respect to the median value (red line in each box). The lines provide the ranges of the measurements, excluding outliers (+).

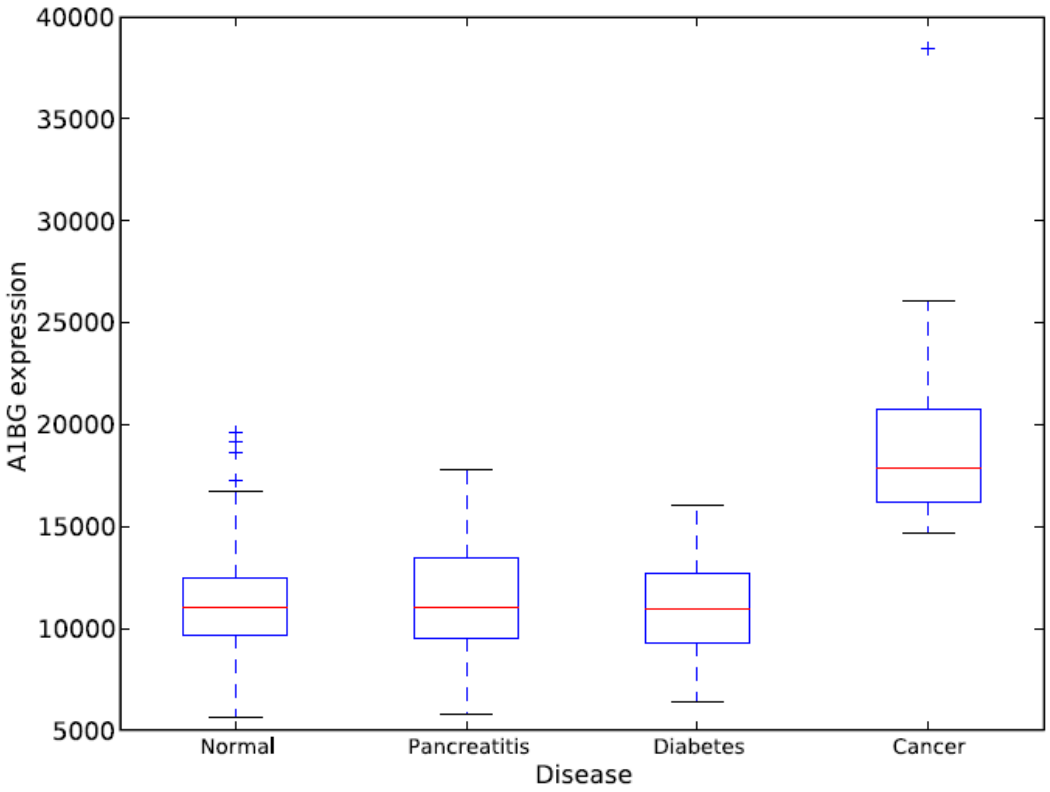


Table 2.1 Residual, batch and patient standard deviations of individual antibodies. These are the estimated standard deviations for variance components in the linear mixed effects model, as described in the text. Only samples with fasting replicates were analyzed.

Std. Dev. (log2)	**A1BG**	**Amyloid**	**Antithrombin-III**
Residual	**0.21559**	**0.19667**	**0.22877**
Batch	**0. 20701**	**0.16376**	**0.15382**
Patient	**0.12681**	**0.05973**	**0.19776**

Table 2.2 Fixed effects for sex, disease classes (excluding cancer) and fasting status. Point estimates, standard error and T value (ratio of point estimate to standard error) are given. The effect estimate (column 2) is on the log2 scale, and the percent effect on the raw scale is given in column 3. T-values greater than 2 in magnitude correspond to two-sided p-values smaller than 0.05. Only samples with fasting replicates were analyzed. The reference category is a normal (free of pancreatitis and diabetes), non-fasting female. Fixed effects are relative to this category.

A1BG	Estimate	Percent	Std. Error	T value
Male	**0.2948**	**2%**	**0.05324**	**0.55**
Female	**0**	**0%**	**---**	**---**
Pancreatitis	**-0.03688**	**-3%**	**0.65551**	**-0.56**
Diabetic	**-0.3033**	**-2%**	**0.06616**	**-0.46**
Normal	**0**	**0%**	**---**	**---**
Fasting	**0.06761**	**5%**	**0.03157**	**2.14**
Non-fasting	**0**	**0%**	**---**	**---**

SAMP	Estimate	Percent	Std. Error	T value
Male	**0.04110**	**3%**	**0.03808**	**1.1**
Female	**0**	**0%**		
Pancreatitis	**-0.03545**	**-2%**	**0.04829**	**-0.7**
Diabetic	**-0.00562**	**0%**	**0.04719**	**-0.1**
Normal	**0**	**0%**	**---**	**---**
Fasting	**0.06383**	**4%**	**0.02758**	**2.3**
Non-fasting	**0**	**0%**	**---**	**---**

Antithrombin-III	Estimate	Percent	Std. Error	T value
Male	**0.1134**	**8%**	**0.04489**	**2.5**
Female	**0**	**0%**	**---**	**---**
Pancreatitis	**-0.16112**	**-12%**	**0.0564**	**-2.9**
Diabetic	**0.03839**	**3%**	**0.05640**	**0.7**
Normal	**0**	**0%**	**---**	**---**
Fasting	**0.01663**	**1%**	**0.03088**	**0.5**
Non-fasting	**0**	**0%**	**---**	**---**

Table 2.3 Fixed effects of disease classes including cancer. Point estimates, standard error and T value (ratio of point estimate to standard error) are given. The effect estimate (column 2) is on the log2 scale, and the percent effect on the raw scale is given in column 3. T-values greater than 2 in magnitude correspond to two-sided p-values smaller than 0.05. The reference category is an individual free of diabetes and pancreatitis. Fixed effects are relative to this category.

A1BG	Estimate	Percent	Std. Error	T value
Cancer	**0.75515**	**69%**	**0.06972**	**10.8**
Pancreatitis	**0.02404**	**-2%**	**0.06323**	**-0.4**
Diabetic	**-0.01072**	**-1%**	**0.07169**	**-0.1**
Normal	**0**	**0%**	**---**	**---**

SAMP	Estimate	Percent	Std. Error	T value
Cancer	**0.41395**	**33%**	**0.06065**	**6.8**
Pancreatitis	**-0.04436**	**-3%**	**0.05510**	**-0.8**
Diabetic	**0.00274**	**0%**	**0.06065**	**0.0439**
Normal	**0**	**0%**	**---**	**---**

Antithrombin-III	Estimate	Percent	Std. Error	T value
Cancer	**0.03193**	**2%**	**0.07523**	**0.4**
Pancreatitis	**-0.1629**	**-11%**	**0.06808**	**-2.4**
Diabetic	**0.02831**	**2%**	**0.07523**	**0.4**
Normal	**0**	**0%**	**---**	**---**

Reference

1. Jemal A, Siegel R, Ward E, Murray T, CA Cancer J Clin 2006, 56, 106
2. Mann D V, Edwards R, Ho S, Lau W Y, Glazer G Eur J Surg Oncol 2000, 26, 474
3. Ferrone C R, Finkelstein D M, Thayer S P, Muzikansky A, Fernandez-del Castillo C, Warshaw AL J Clin Oncol 2006, 24, 2897
4. Duffy M J Ann Clin Biochem 1998, 35 (Pt 3), 364
5. Boeck S, Stieber P, Holdenrieder S, Wilkowski R, Heinemann V Oncology 2006, 70, 255
6. Dalgleish AG Britsh Medical Journal 2000, 321, 380
7. Chang CY, Huang SP, Chiu HM, Lee YC, Chen MF, Lin JT Hepatogastroenterology 2006, 53, 1
8. Kim JE, Lee KT, Lee JK, Paik SW, Rhee JC, Choi KW J Gastroenterol Hepatol 2004,19,182
9. Riker A, Libutti SK, Bartlett DL Surg Oncol 1998, 6,157
10. Wigmore SJ, Fearon KC, Sangster K, Maingay JP, Garden OJ, Ross JA Int J Oncol 2002, 21, 881
11. Fearon KC, Barber MD, Falconer JS, McMillan DC, Ross JA, Preston T World J Surg 1999, 23, 584
12. Dube DH, Bertozzi CR Nat Rev Drug Discov 2005, 4, 477
13. Garcea G, Neal C P, Pattenden C J, StewardW P, Berry D P Eur J Cancer 2005, 41, 2213
14. Rustgi A K Gastroenterology 2005, 129, 1344
15. VanMeter A, Signore M, Pierobon M, Espina V, Liotta LA, Petricoin EF Expert Review of Molecular Diagnostics 2007, 5,625
16. An HJ, Peavy TR, Hedrick JL, Lebrilla CB Anal Chem 2003, 75: 5628
17. Block TM, Comunale MA, Lowman M, Steel LF, Romano PR, Fimmel C, Tennant BC, London WT, Evans AA, Blumberg BS, Dwek RA, Mattu TS, Mehta AS Proc Natl Acad Sci U S A 2005, 102, 779
18. Ressom HW, Varghese RS, Goldman L, An Y, Loffredo CA, Kyselova Z, Mechref Y, Novotny M, Drake SK, Goldman R, Journal of Proteome Research 2008, 7, 603
19. Gessner P, Riedl S, Quentmaier A, and Kemmner W, Cancer Lett 1993, 75, 143
20. Gorelik E, Galili U, Raz A Cancer Metastasis Rev 2001, 20, 245
21. Morelle W, Michalski JC Curr Pharm Des 2005, 11, 2615
22. Peracaula R, Tabares G, Royle L, Harvey DJ, Dwek RA, Rudd PM, de Llorens R Glycobiology 2003, 13, 457
23. Poland DC, Kratz E, Vermeiden JP, De Groot SM, Bruyneel B De Vries, T and Van Dijk, W Prostate 2002, 52, 34
24. Marrero JA, Romano PR, Nikolaeva O, Steel L, Mehta A, Fimmel CJ, Comunale MA, D'Amelio A, Lok AS, Block TM J Hepatol 2005, 43: 1007
25. Ahmed N, Oliva KT, Barker G, Hoffmann P, Reeve S, Smith IA, Quinn MA, Rice GE Proteomics 2005, 5, 4625
26. Bertozzi CR, Kiessling LL Science 2001, 291, 2357
27. Okuyama N Ide Y, Nakano M, Nakagawa T, Yamanaka K, Moriwaki K, Murata K, Ohigashi H, Yokoyama S, Eguchi H, Ishikawa O, Ito T, Kato M, Kasahara A, Kawano S, Gu JG, Miyoshi E Int J Cancer 2006,118, 2803

28. Lacunza I, Kremmer T, Diez-Masa JC, Sanz J, de Frutos M Eletrophoresis 2007, 23, 4447
29. Yang ZP, Hancock WS Journal of Chromatograhpy A 2004, 1-2, 79
30. Drake RR, Schwegler EE, Malik G, Diaz J, Block T, Mehta A, Semmes OJ Molecular & Cellular Proteomics 2006, 10, 1957
31. Cho W, Jung K, Regnier FE Anal Chem 2008, 14, 5286
32. Kyselova Z, Mechref Y, Kang P, Goetz JA, Dobrolecki LE, Sledge GW, Schnaper L, Hickey RJ, Malkas LH, Novotny MV Clinical Chemistry 2008, 7, 1166
33. Zhao J, Simeone DM, Heidt D, Anderson MA, Lubman DM Journal of Proteome Research 2006, 7, 1792
34. Zhao J, Qiu WL, Simeone DM, Lubman DM Journal of Proteome Research 2007, 3, 1126
35. Zhang H, Li XJ, Martin DB, Aebersold R Nature Biotechnology 2003, 6, 660
36. Zhao J, Patwa TH, QiuWL, Shedden K, Hinderer R, Misek DE, Anderson MA, Simeone DM, Lubman DM Journal of Proteome Research 2007, 5, 1864
37. Qiu Y, Patwa TH, Xu L, Shedden K, Misek DE, Tuck M, Jin G, Ruffin MT, Turgeon DK, Synal S, Bresalier R, Marcon N, Brenner DE, Lubman DM Journal of Proteome Research 2008, 7, 1693
38. Alvarez RA, Blixt O Glycobiology 2006, 415, 292
39. Kuno A, Uchiyama N, Koseki-Kuno S, Ebe Y, Takashima S, Yamada M, Hirabayashi J Nature Methods 2005, 11, 851
40. Chen P, Liu, YK, Kang, XN Sun, L Yang, PY Tang, ZY Journal of Cancer Research and Clinical Oncology 2008 8, 851
41. Chen SM, LaRoche T, Hamelinck D, Bergsma D, Brenner D, Simeone D, Brand RE, Haab BB Nature Methods 2007, 5, 437
42. Ingvarsson J, Wingren C, Carlsson A, Ellmark P, Wahren B, Engstrom G, Harmenberg U, Krogh M, Peterson C, Borrebaeck CAK Proteomics 2008, 11, 2211
43. Haab BB Current Opinion in Biotechnology 2006, 4, 415
44. Orchekowski R, Hamelinck D, Li L, Gliwa E, VanBrocklin M, Marrero JA, Woude GFV, Feng ZD, Brand R, Haab BB Cancer Research 2005, 23, 11193
45. Kreunin P, Zhao J, Rosser C, Urquidi V, Lubman DM, Goodison S, Journal of Proteome Research 2007, 7 2631
46. Liang XL, Lubman DM, Rossi DT, Nordblom GD, Barksdale CM Anal Chem 1998, 3, 498
47. Ransohoff DF Nat Rev Cancer 2005, 5,142
48. Evans-Nguyen KM, Tao SC, Zhu H, Cotter RJ Anal Chem 2008, 5,1448

Chapter 3

A multiplexed bead assay for profiling glycosylation patterns on serum proteins

Abstract:

A multiplexed bead-based immunoassay was developed to simultaneously profile glycosylation patterns of serum proteins to investigate their usefulness as biomarkers for pancreatic cancer. The multiplex assay utilized protein specific capture antibodies chemically coupled individually to beads labeled with specific amounts of fluorescent dye. Captured proteins were detected based on the extent and specific type of glycosylation as determined by successive binding of fluorescent lectin probes. For immunoassay format comparison purposes, the same capture antibodies were printed as discrete spots on nitrocellulose surface microarrays. Relative to the microarrays, antibodies coupled to the beads had less non-specific binding to the lectins, resulting in very low background signal. The bead assays were performed in a 96-well filter plate enabling high-throughput screening applications. The antibody-conjugated beads responded linearly to the antigens at low concentration, whereas the signal saturates at 20-fold dilution of serum. The assay was tested with ConA and SNA lectins to examine the glycosylation patterns of alpha-1-β glycoprotein and serum amyloid p component for use as biomarkers for the detection of pancreatic cancer. We found combining the results from the two glycoproteins enabled us to distinguish pancreatic cancer from normal controls and pancreatitis ($p < 0.05$). The microarray platform was performed with the same sample set and results were compared between the two platforms.

3.1 Introduction

Carbohydrate modifications on glycoproteins show a high structural diversity reflecting inherent functional roles for particular biological environments. Serum represents one of the most physiologically relevant glyco-proteomes in the human body. The alteration of glycan structure and coverage on several major glycoproteins in serum has been shown to contribute to the progression of cancer [1-5]. Studies comparing the carbohydrate chains of glycoproteins produced in serum from patients with developed malignancy to patients with corresponding chronic disease and normal controls, may provide useful information for the diagnosis, prognosis, and development of therapeutic strategies [6-10].

Pancreatic adenocarcinoma is the 4th leading cause of cancer-related death in the United States [11]. According to the SEER database pancreatic adenocarcinoma is associated with a 4% five-year survival rate. While a majority of these patients do not achieve curability, pancreatic resection is associated with substantially improved outcomes including a five-fold higher 5-year survival rate [12]. Unfortunately, the overwhelming majority of patients do not present with early-stage disease and currently there are no clinically useful strategies for detection of early pancreatic cancer. Many strategies utilizing both invasive and noninvasive techniques have been employed to detect early pancreatic cancer among patients at high risk of developing this malignancy. Endoscopic ultrasound has been studied for screening purposes but has been found limited in several aspects [13]. CA 19-9 is a serologic biomarker of pancreatic cancer but it is ineffective at detecting early pancreatic cancer [14, 15]. Recently, a novel approach in cancer-biomarker discovery has exploited the field of glycomics to discover unique differential glycosylation patterns associated with malignant processes [16]. Glycoproteins from patients' serum were extracted and separated by a combination of

multi-lectin affinity and reverse-phase chromatography, then spotted on glass slides. Five lectins were utilized to investigate the differential glycosylation pattern of the proteins in each fraction. The study identified 10 potential glycoprotein biomarkers with significant variation in pancreatic cancer samples.

Microarray immunoassays have been increasingly used for profiling and characterizing captured proteins and their modifications, with the appeal of multiplexing and low sample cost. A new type of antibody-lectin array has been recently developed to detect the glycosylation of specific proteins in a complex mixture [17-19]. A variety of glycan structures of interest can be probed with commercially available lectins. Among the methods for glycosylation research, this technique provides the highest throughput and reproducibility in measuring the abundance and glycosylation of multiple proteins, which is essential for disease biomarker discovery. The same advantages are shared by microsphere or bead-based immunoassay techniques [20, 21]. Bead-based multiplex protein profiling is derived from antibody microarrays, but uses polystyrene beads labeled with amounts of discrete fluorophore instead of a flat surface for antibody immobilization. The fluorophore intensity varies between different bead types and is easily distinguished using a flow cytometer. Additional multiplexing capability is achieved through the use of differently sized beads also distinguishable by flow cytometers.

The MultiBeads® immunoassay platform commercialized by Assay Designs provides beads internally labeled with discrete amounts of fluorescent dye producing 12 spectrally distinct bead types. Two sizes of beads are available allowing 24 different bead types to be combined. Similar to an antibody sandwich assay, each antibody-

coupled bead captures an antigen that can be detected by a fluorophore conjugated secondary antibody. A mixture of bead types, each hosting a separate immunoassay can be simultaneously read in a flow cytometer. Each reading contains the measurement of multiple antigens, each diffentiated based on the intrinsic fluorescence and size of the host bead. The beads used in a large set of a parallel assay are uniformly fabricated while each block of the microarray is printed individually providing the potential for bead-based immunoassays to outperform microarray-based assays in terms of reproducibility and accuracy. In addition, protein microarrays typically use non-covalent interactions and require drying during the printing process to bind the antibody to the solid phase. Bead-based immunoassays utilize covalent linkage chemistry and the immobilized antibodies are maintained in aqueous buffers. The result is that bead-based assays typically produce lower background due to reduced nonspecific absorption.

To take advantage of the benefit of bead-based immunoassays for glycosylation detection, we converted an antibody-lectin technique developed on the microarray platform for use on the bead platform (experimental scheme shown in Figure 1). Two types of antibody coupled beads targeting alpha-1-β glycoprotein and serum amyloid p component were prepared and tested to find the optimal conditions. As a proof-of-concept, they are employed to measure the glycosylation pattern of samples from different groups of patients including pancreatic cancer, chronic pancreatitis, and normal controls. The performance of the microsphere based assay is analyzed and compared with the results of a microarray experiment using the same sample set.

3.2 Experimental Section

3.2.1 Serum samples

Inclusion criteria for the study included patients with pancreatic cancer, chronic pancreatitis, or healthy adults with the ability to provide written, informed consent, and provide 40 ml of blood. Exclusion criteria included inability to provide informed consent, patients' actively undergoing chemotherapy or radiation therapy for pancreatic cancer, and patients with other malignancies diagnosed or treated within the last 5 years. The sera samples were obtained from patients with a confirmed diagnosis of pancreatic adenocarcinoma who were seen in the Multidisciplinary Pancreatic Tumor Clinic at the University of Michigan Comprehensive Cancer Center. All cancer sera samples used in this study were obtained from patients with stages III/IV pancreatic cancer. The mean age of the tumor group was 65.4 years (range 54-74 years). The sera from the normal and pancreatitis groups was age and gender-matched to the tumor group. The chronic pancreatitis group was sampled when there were no symptoms of acute flare of their disease. All sera were processed using identical procedures. The samples were permitted to sit at room temperature for a minimum of 30 minutes (and a maximum of 60 minutes) to allow the clot to form in the red top tubes, and then centrifuged at 1,300 x g at 4^{o}C for 20 minutes. The serum was removed, transferred to polypropylene tubes in 1 ml aliquots, and frozen. The frozen samples were stored at -80^{o}C until assayed. All serum samples were labeled with a unique identifier to protect the confidentiality of the patient. None of the samples were thawed more than twice before analysis. This study was approved by the Institutional Review Board for the University of Michigan Medical School.

3.2.2 Mouse mAbs.

Alpha-1-β glycoprotein mAb was acquired from Novus, while amyloid p component mAb antibody was from Abcam. The mAbs were convalently coupled to 5.4 um latex

beads by Assay Designs. The mAbs were diluted with water to 0.5ug/uL for microarray printing.

3.2.3 Antibody blocking

To prevent reaction between the glycans on the antibodies and some specific detection lectins, the antibodies coupled to beads were chemically modified following the glycan-blocking protocol described in our previous work [17]. Briefly, the beads were washed with coupling buffer (AB34) then incubated in 0.2M NaIO4 for 3hours. When the oxidation reaction was finished, precipitation was removed by washing the beads with coupling buffer with 0.1% Tween 20 three times. The oxidized antibody beads were incubated with 1mM MPBH and 1mM Cys-Gly dipeptide for 2hours. Finally the beads were kept in 1mM Cys-Gly in dark at 4C overnight. The blocked beads must be extensively washed before being stored in a 4C refrigerator to eliminate reagent in the solution.

3.2.4 Sample incubation and flow cytometry detection

The beads coupled to A1BG and SAP antibodies were mixed with 20X diluted serum in eppendorf tubes and incubated on a shaker set at 300rpm for one hour at room temperature. Each tube contained six thousand beads of each type. The beads were then transferred to two identical 96-well filter plates where subsequent incubation and washing were done. The samples from different disease groups were randomized on the well plate to eliminate bias. The two duplicate plates were processed in parallel. Then the serum solution was drained and the beads were washed with PBST 3 times. A vacuum manifold was used to drain reagent and washing buffer. Biotinylated lectins (Vector Laboratory) were diluted to 1ug/mL and applied to each well. The lectin-glycoprotein

reaction was allowed to proceed for 45 minutes before being complete. The filter plates were rinsed to remove unbound lectins. The solution of 1ug/mL streptavidinylated Alexa 555 (Invitrogen Biotechnology) was added to each well for detection. Finally the beads were washed with water to remove detergent.

The fluorescent signal was read by a flow cytometer (FACSCalibur). The beads were sorted by flow cytometer based on size and inherent fluorescent intensities using the 670nm filter. Three hundred beads of each type were counted. The fluorescent signals of the analytes were measured at 575nm.

3.2.5 Statistical analysis

Data analysis was done in Weasel version 2.6. The signals were gated to exclude damaged or cross-linked beads. The medians of the select signal spots at each inherent fluorescent level were taken as a data point into analysis. All the samples were measured twice with duplicate wells. The reproducibility of the experiments was assessed by calculating the coefficient of variation (CV) and Pearson correlation for the pairs of duplicates. To compare the signal of cancer and non-cancer samples, a student T-test was applied to the data, where a difference is considered statistically significant when $P<0.05$. The agreement of the data from the two platforms, bead and microarray, was also estimated by Pearson correlation.

3.3 Results and discussion

3.3.1 Optimization of the experimental conditions

Necessity of Glycan blocking

The binding of the lectins to glycans on the antibodies printed on the arrays resulted in interference for the detection of captured glycoproteins. Previous work using antibody

microarrays for glycoprotein studies utilized a blocking procedure to prevent this interaction [17, 18]. As described in the experimental section, the cis-diol groups on the glycans were gently oxidized and converted to aldehyde groups which then reacted with hydrazide-maleimide bifunctional cross-linking reagent capped with a Cys-Gly dipeptide. The fluorescent signal of the lectin bound to the blocked antibodies decreased 5 to 10 fold depending on which lectin was used (data not shown). The derivatization procedure prevented the binding of the lectins to glycans on the antibodies but it resulted in reduced stability of the antibody on the antibody-coupled beads, where the signal of the blocked beads degraded over several days. To determine whether the glycan-blocking was necessary, the underivatised beads were treated with/without 20X diluted serum to assess their binding to 5 different lectins. The lectins utilized were ConA, SNA, AAL, MAL II and PNA.

In Table 1a, the signal of the lectin-treated beads (without serum) indicates the level of interaction between the lectin and the underivatised bead-coupled antibodies. The signal of ConA and SNA bound to underivatised antibodies is 20 fold lower than the signal of the lectins bound to the underivatised antibody spots exposed to serum, thus the underivatised antibodies do not bind to ConA and SNA strongly. AAL had strong interaction with the underivatised form. MAL II also shows a low level of signal. PNA barely binds to the antibodies or the captured proteins. The glycan on the Fc region of IgG in a native form is hidden and is not accessible for lectins. Glycosylation on the Fab region is dependent on the variable amino acid sequence which correlates with the selectivity and affinity of each of the specific antibodies. PNA recognizes galactose but usually shows a very low signal with IgG, because the specific interaction can be

prohibited by the presence of sialylation which is very common for serum glycoproteins. Although a majority of the oligosaccharides on the Fab region are fully sialylated, the glycan chains may associate with the positively charged amino acid side chains on the protein but not stay free in the solution. Therefore, the binding of SNA and ConA to their target glycan structure may be prevented, while the binding of AAL to fucosylation is less affected. Alternatively, the modification on lysine for covalent coupling may have hampered the accessibility of the glycan chain. On the microarray, the signal of the lectins bound to the unblocked antibodies is at a similar level to those bound to the antibody spots exposed to serum (Table 1b). Therefore the blocking effect of the antibody glycan is caused by coupling the antibodies to the beads. Additional experiments need to be done to clarify the mechanism. Since the underivatised antibodies on the beads do not bind to ConA and SNA, these two lectins were selected for large scale analysis.

Serum concentration

For experimentation, serum dilution for studying glycosylation was targeted to antibody saturation and minimization of background binding. At antibody saturation, glycosylation can be studied independently of target glycoprotein serum concentration since the amount of on-bead glycoprotein is standardized to the amount of antibody conjugated to the bead. Thus, differences in glycosylation can be measured between samples as lectin binding is directly related to level of its target glycan structures present on the proteins captured on the antibody array.

Different dilutions of serum (5X, 10X, 20X, 50X, and 200X) were tested to determine the optimum concentration of the target glycoproteins. The same lectin ConA was used for detection in this test. Figure 2 depicts how the intensity of the signal

changes for the two antibodies with the dilution fold. A rising trend was noted from the 200X dilution to the 20X dilution for both antibodies. From 20X dilution to 5X dilution, the signal of SAP only increased by 5%. For A1BG, the signal reached a peak at 20X dilution, where a saturation of the target protein on the antibody had occurred. The signal then decreased 10% when the dilution increased to 5X. The decrease of signal was observed along with an increase of outlier data points in the flow cytometry spectrum and a higher standard deviation, likely due to competing non-specific binding on the antibodies.

The result of the dilution test demonstrated that the antibodies were saturated by their target protein at 20X dilution in the process of hybridization (1hour, room temperature and gentle shaking). Below or above 20X dilution would not completely occupy all the antibodies or introduce strong non-specific binding. Thus, 20X was chosen as the optimal dilution fold.

3.3.2 Serum testing

Lectin ConA and SNA were used in two separate tests with antibody-coupled beads to analyze serum samples from normal control and patients with chronic pancreatitis and pancreatic cancer. In the experiment using ConA as the detection lectin, ten samples from each of the three groups were tested, where duplicate wells were made on the same plate. SNA was used in an experiment to test twenty samples from each group and duplicate wells were made on two identical plates.

Reproducibility

The reproducibility was assessed using duplicate sample pairs, on the same plate or two different plates. The average CV of the duplicate pairs for antibody A1BG and SAP were

9.8% and 6.9% respectively. On each of the two plates, 20 normal samples, 20 pancreatitis samples and 20 cancer samples were analyzed in parallel. The average CV of the duplicate pairs on two different plates for antibody A1BG and SAP were 7.5% and 5.9% respectively.

Biomarker performance

A1BG and SAP are pancreatic cancer related biomarker candidates. Their glycosylation levels have been found elevated in pancreatic cancer serum in a previous study using a RP HPLC-fraction microarray method [16]. Both of these two glycoproteins produced significant differences when analyzed against lectin SNA and ConA. Their potential to aid diagnosing pancreatic cancer was further tested utilizing the antibody microarray with lectin SNA in a previous study [17]. The results determined that A1BG was able to distinguish pancreatic cancer from chronic pancreatitis. On the bead based platform, a similar experiment was conducted with a different set of samples. The results (shown in Figure 3) are subjected to a t-test between each pair of sample classes. A P-value of 0.035 was obtained for SAP between samples from chronic pancreatitis patients and pancreatic cancer patients, while the p-value for normal control and pancreatic cancer was 0.096 which is above the significant level. For A1BG, the difference between normal control and pancreatic cancer patients is significant (p-value 0.026). The results demonstrate the importance of the multiplexed biomarker measurement, as neither of the two antibodies can significantly distinguish cancer samples from normal and pancreatitis samples by themselves, while together they create the capacity of differentiating cancer samples from the other two groups (Figure 3).

Mannosylation level of the glycoproteins were probed with ConA. The average signals and the variation for each group of samples are shown as a bar graph with error bars in compared to chronic pancreatitis was observed, but this did not reach statistical significance, $p < 0.11$.

3.3.3 Comparison to Microarray results

A microarray experiment using the same sample set and lectin SNA was conducted to compare the performance of these two platforms. We first calculated the Pearson correlations between the replicated well/block to estimate the reproducibility of the assays on these two platforms individually. As shown in table 2, the bead-based glycosylation assays using A1BG and SAP antibody show medium to high correlations (A1BG, r=0.85; SAP, r=0.72, total 57 samples) to their replicates on the other plate. For microarray-based assay, the correlations of the replicate blocks are slightly lower (A1BG, r=0.80; SAP, r=0.70, total 60 samples). To determine whether the two assays provide similar results on the same sample set, Pearson correlation was calculated between the microarray and bead results. The microarray assay yielded an overall medium agreement (A1BG, r=0.64; SAP, r=0.51; total 57 samples) with the bead-based assay. Within subgroups, higher agreement between the result of the microarray and bead assays was observed in the cancer cases (A1BG, r=0.84; SAP, r=0.63; 20 cases), probably due to the higher in-group variation of the cancer samples (data not shown). In terms of the specific antibody, the agreement is poorer with SAP, as it showed lower variation than A1BG. It is notable that though both assays have been able to distinguish the samples from the 2 disease classes and healthy controls, the ranking of the average signal varies in the results of these two assays. The average signal of the healthy controls ranked after cancer as the

2nd highest in the microarray results, while in the bead-based assay the healthy control group was the highest followed by cancer and pancreatitis.

3.4 Discussion

The bead-based glycoprotein detection method has been developed for multiplexed analysis of different kinds of glycosylation on biomarkers captured by antibodies coupled to the beads. The assay is ideally suited for large scale biomarker/sample screening. The analyses presented herein were conducted to measure the two types of glycosylation of two glycoprotein biomarker candidates that could be used to distinguish serum samples from pancreatic cancer patients and chronic pancreatitis and healthy controls, and evaluated the performance of this assay with respect to reproducibility and its agreement with the microarray-based assay.

The bead-based assay has several notable advantages over the microarray-based assay. Some key steps, such as array printing and slide scanning, that always introduce spatial variation are eliminated when antibody coupled beads are used to perform the assay in a 96-well plate. Improved reproducibility was observed for both of the two antibodies tested. Additionally, the number of microarray slides that can be printed in one batch is limited and different batches introduce batch-to-batch variation; whereas antibody-coupled beads prepared in one batch are enough to perform thousands of assays. In terms of procedure, unblocked bead-coupled antibodies bind to ConA and SNA at a very low level, even lower than blocked ones on microarray (data not shown), thus the whole experiment can be simplified by removing the chemical blocking procedure. We are currently testing additional antibodies to find out whether they would behave similarly after coupling to the beads. Furthermore, unlike the traditional antibody

sandwich assay using commercially available antibody pairs for capture and detection, glycosylation detection is performed by a single lectin. Therefore, the number of multiplexed assays that may be done in a single well is only limited by the types of internally labeled beads available.

3.5 Conclusion

We have developed a method for high-throughput glycoprotein biomarker screening using a novel bead-based antibody-lectin glycoprotein assay. Compared to the microarray platform, the new technique showed improved sensitivity and reproducibility in glycan detection. Additionally, the glycans on the bead-conjugated antibodies were not reactive to lectin ConA and SNA, so the glycan blocking step can be avoided. Using the bead-based assay, we discriminated glycosylation patterns among normal and other disease states with two glycoprotein biomarker candidates. This platform will be further tested by multiplexing more antibodies onto the beads and comparing the result with the microarray-based assay. Examination of a larger set of serum samples is planned to verify the validity of the biomarkers.

Figure 3.1. The diagram of bead based antibody-lectin multiplex assay. 1) production of 2 sizes of unique labeled beads; 2) chemical coupling different antibodies to different types of beads; 3) hybridization of the antibody-conjugated beads with diluted serum in a 96-well filter plate; 4) sequential reaction of antigen capture, glycan-lectin binding, and fluorescence detection; 5) detection of signal for each type of bead with a flow cytometer; 6) gating signal points to extract data for each analyte.

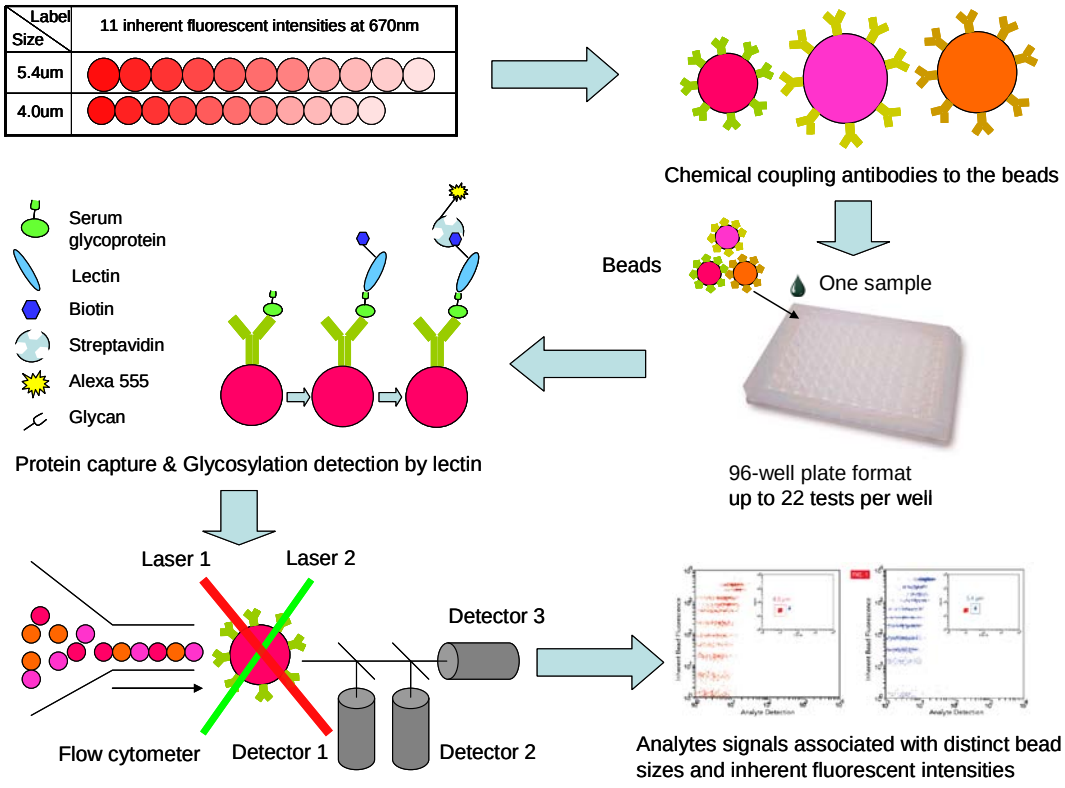


Figure 3.2. Signal-dilution curve made to obtain the optimum concentration of serum for bead hybridization using lectin SNA. The x-axis is the dilution fold of serum incubated with the beads in each assay ranging from 5-200 X, the y-axis is the resulting fluorescent signal. The value of each spot on the y-axis presents the signal yield of each assay. A1BG: Alpha-1- β glycoprotein; SAP: Serum amyloid p component.

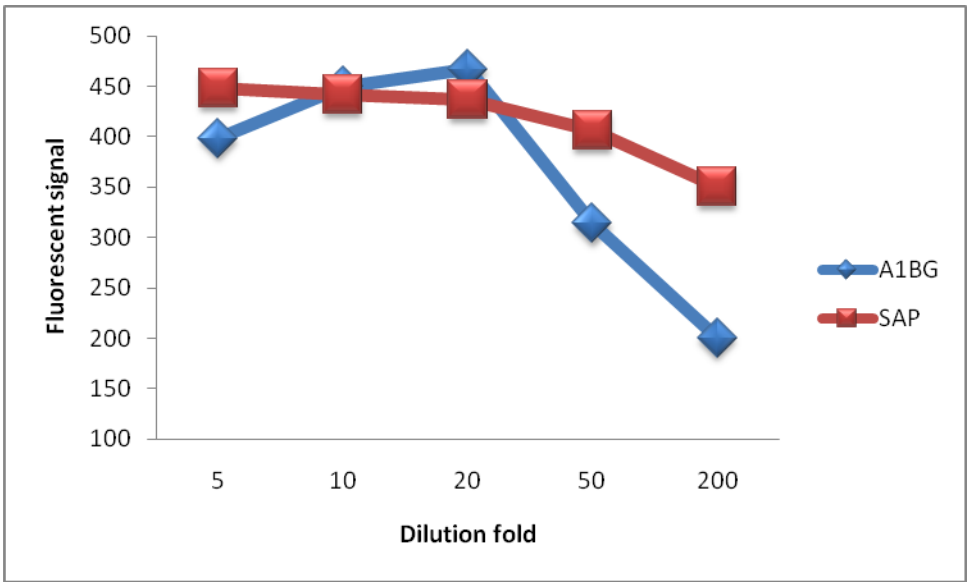
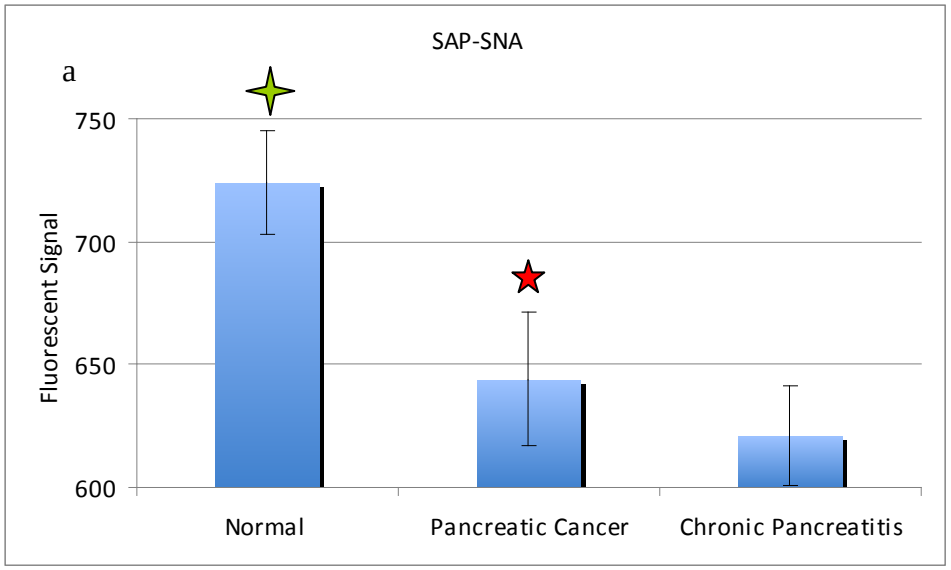


Figure 3.3. The experiment using lectin SNA to detect alpha-2,6 sialylation is performed with 20 samples from each of the three groups, pancreatic adenocarcinoma, chronic pancreatitis, and normal control. The bar graph and SEM (error bars) shows the average signal and variation for each group of samples. The group marked with a green star on top can be significantly distinguished from the group marked with a red star. a) Serum amyloid p component; b) Alpha-1-β glycoprotein.



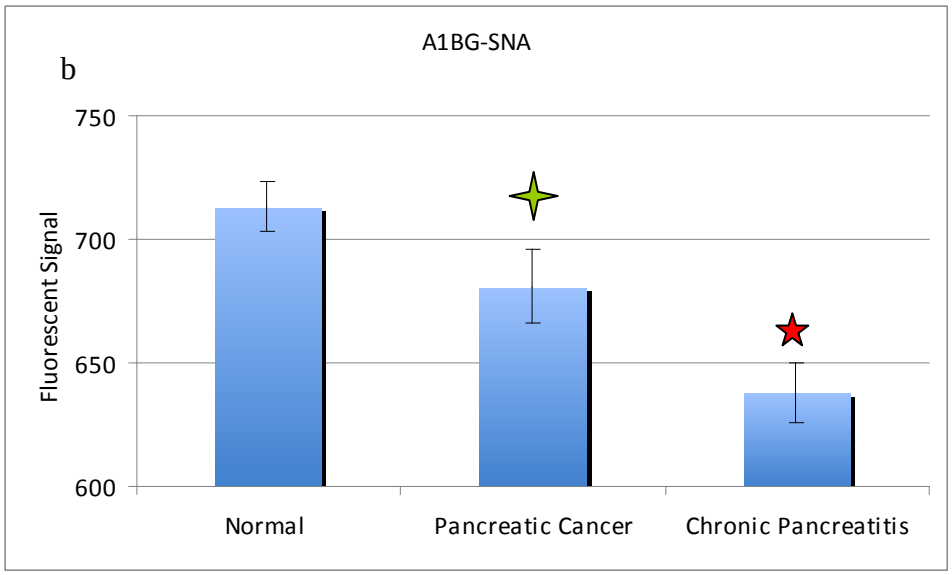
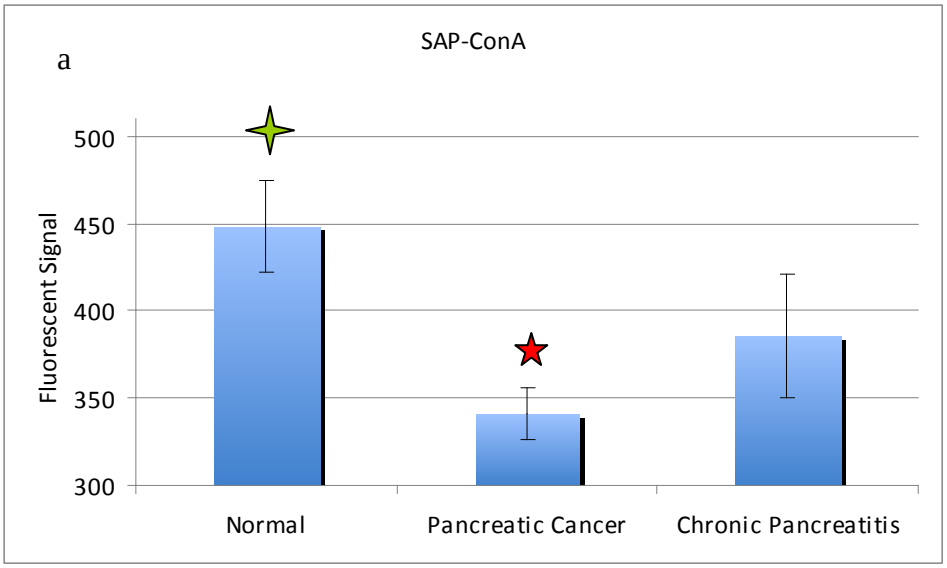


Figure 3.4. Result of an experiment using ten sera samples from patients with stage III and IV Pancreatic Adenocarcinoma , Chronic Pancreatitis, and Normal Healthy Controls. Mannosylation levels were probed with biotinylated ConA and detected by streptavidinylated Alexa555. The bar graph and SEM (error bars) shows the average signal and variation for each group of samples. The group marked with a green star on top can be significantly distinguished from the group marked with a red star. a) Serum amyloid p component; b) Alpha-1-β glycoprotein.



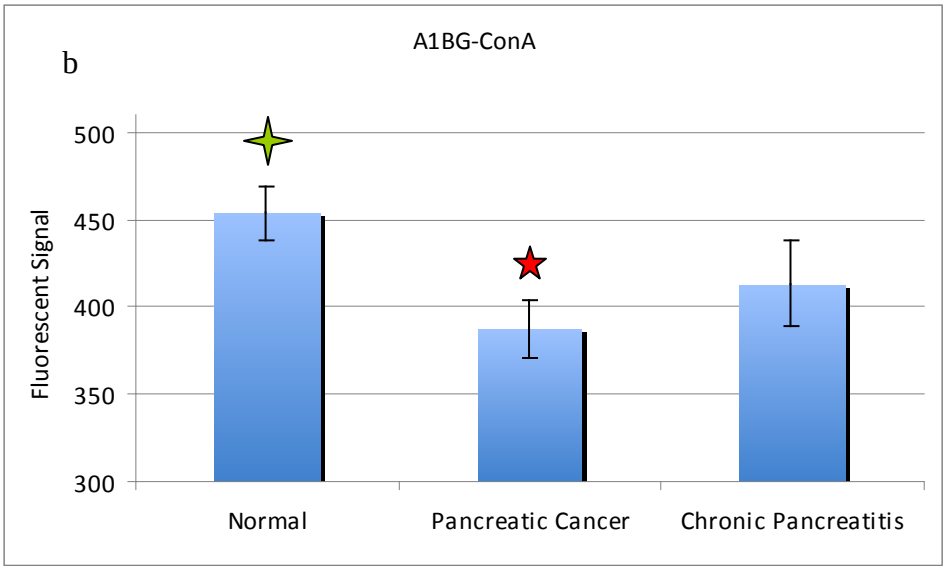


Table 3.1 Fluorescence signal of antibody-lectin experiment using either antibody-conjugated beads or microarray platform. The beads and microarray were hybridized with serum or sample buffer.

a) antibody-conjugated beads without glycan

Bead Unblocked	Without Serum		With Serum	
	A1BG	SAP	A1BG	Sap
SNA	15	23	636	660
MAL	14	57	43	184
AAL	276	353	317	403
ConA	12	5	418	385
PNA	2.5	16	13	11

b. antibody microarray with glycan

Microarray Unblocked	Without Serum		With Serum	
	A1BG	SAP	A1BG	SAP
SNA	3080	5674	11004	6312
MAL	3924	3072	6335	5329
AAL	19052	21449	19588	21569
ConA	22570	18172	19850	16191
PNA	985	405	1439	685

c. antibody microarray without glycan

Microarray Blocked	Without Serum		With Serum	
	A1BG	SAP	A1BG	SAP
SNA	284	195	6834	5973
MAL	144	172	3231	3735
AAL	1069	725	12521	9875
ConA	1543	1215	13067	11600
PNA	82	55	739	387

Table 2. Correlations of the results between the replicates in the microarray or bead assay and the correlation between the two assays. CVs indicate the variation in each group.

	Reproducibility				Validity	
	Microarray vs. Microarray		Bead vs. Bead		Microarray vs. Bead	
Antibody	A1BG	SAP	A1BG	SAP	A1BG	SAP
Pearson Correlation (r)	0.80	0.70	0.85	0.72	0.64	0.51
Variability (CV)	0.16	0.15	0.16	0.09	N/A	N/A

References

1. Plavina T, Wakshull E, Hancock WS, Hincapie M, J Proteome Res 2007, 2, 662
2. Rudd PM, Elliott T, Cresswell P, Wilson IA, Dwek RA, Science, 2001, 291, 2370
3. Gorelik E, Galili U, Raz A, Cancer Metastasis Rev, 2001, 20, 245
4. Zhao J, Simeone DM, Heidt D, Anderson MA, Lubman DM, J Proteome Res, 2006, 5, 1792
5. Block TM, Comunale MA, Lowman M, Steel LF, Romano PR, Fimmel C, Tennant BC, London WT, Evans AA, Blumberg BS, Dwek RA, Mattu TS, Mehta AS, Proc Natl Acad Sci USA, 2005, 102, 779
6. Drake RR, Schwegler EE, Malik G, Diaz J, Block T, Mehta A, Semmes OJ, Mol Cell Proteomics, 2006, 10, 1957
7. Peracaula R, Tabares G, Royle L, Harvey DJ, Dwek RA, Rudd PM, de Llorens RA, Glycobiol, 2003, 13, 457
8. Kellokumpu S, Sormunen R, Kellokumpu M, FEBS Lett 2002, 516, 217
9. Balzarini J, Nat Rev Microbiol, 2007, 8,583
10. Kobata A, Amano J, Immunol Cell Biol, 2005, 83, 429
11. Jemal A, Tiwari RC, Murray T, Ghafoor A, Samuels A, Ward E, Feuer EJ, and Thun MJ, CA Cancer J Clin 2004, 54, 8
12. Sohn TA, Lillemoe KD, Cameron JL, Huang JJ, Pitt HA, Yeo CJ, J Am Coll Surg, 1999, 188, 658
13. Larghi A, Verna EC, Lecca PG, Costamagna G, Clin Cancer Res, 2009, 15, 1907
14. Dalgleish AG, Br Med J, 2000, 321, 380
15. Nazli O, Bozdag AD, Tansug T, Kir R, Kaymak E, Hepatogastroenterology, 2000, 47, 1750
16. Zhao J, Patwa TH, Qiu WL, Shedden K, Hinderer R, Misek DE, Anderson MA, Simeone DM, Lubman DM, J Proteome Res 2007, 5, 1864
17. Li C, Simeone DM, Brenner DE, Anderson MA, Shedden KA, Ruffin MT, Lubman DM, J Proteome Res, 2009, 8, 483
18. Chen SM, LaRoche T, Hamelinck D, Bergsma D, Brenner D, Simeone D, Brand RE, Haab BB, Nat Methods, 2007, 5, 437
19. Wu YM, Nowack DD, Omenn GS, Haab BB, Pancreas, 2009, 37, 502
20. Morgan E, Varro R, Sepulveda H, Ember JA, Apgar J, Wilson J, Lowe L, Chen R, Shivraj L, Agadir A, Campos R, Ernst D, Gaur A, Clin Immunol, 2004, 110, 252
21. Joos TO, Stoll D, Templin MF, Curr Opin Chem Biol, 2002, 6, 76

Chapter 4

Quantitative proteomic profiling studies of pancreatic cancer stem cells

Abstract:

Analyzing subpopulations of tumor cells in tissue is a challenging subject in proteomic studies. Pancreatic cancer stem cells (CSC) are such a group of cells that only constitute 0.2-0.8% of the total tumor cells but have been found to be the origin of pancreatic cancer carcinogenesis and metastasis. Global proteome profiling of pancreatic cancer stem cells from xenograft tumors in mice is a promising way to unveil the molecular machinery underlying the signaling pathways. However, the extremely low availability of pancreatic tissue cancer stem cells (around 10,000 cells per xenograft tumor or patient sample) has limited the utilization of currently standard proteomic approaches which do not work effectively with such a small amount of material. Herein, we describe the profiling of the proteome of pancreatic cancer stem cells using a capillary scale shotgun technique by coupling offline capillary isoelectric focusing(CIEF) with nano reversed phase liquid chromatography(RPLC) followed by spectral counting peptide quantification. A whole cell lysate from 10,000 cells which corresponds to ~1ug protein material is equally divided for three repeated CIEF separations where around 300 ng peptide material is used in each run. In comparison with a non-tumorigenic tumor cell sample, among 1159 distinct proteins identified with FDR less than 0.2%, 169 differentially expressed proteins are identified after multiple testing correction where 24% of the proteins are upregulated in the CSC group. Ingenuity Pathway analysis of these differential expression signatures

further suggests significant involvement of signaling pathways related to apoptosis, cell proliferation, inflammation and metastasis.

4.1 Introduction:

Pancreatic cancer has the worst prognosis of any major malignancy and is currently ranked the fourth leading cause of cancer-related mortality with a five-year survival rate less than 5%. Delayed diagnosis, relative chemotherapy and radiation resistance and an intrinsic biological aggressiveness all contribute to the abysmal prognosis[1]. Attempts to better understand the molecular characteristics of pancreatic adenocarcinoma have focused on studying the gene and protein expression profiling of pancreatic cancer compared to either normal pancreas or pancreatitis. However, these studies have not accounted for the heterogeneity of the tumor cells, in particular, the existence of a small set of distinct cells termed cancer stem cells which are responsible for tumor initiation and propagation.

Cancer stem cells have been identified in pancreatic cancer and several other tumor types including colon, prostate, and brain. In Li *et al* [2], a subpopulation of pancreatic tumor cells with cell surface markers $CD44^{+}CD24^{+}ESA^{+}$ was isolated and functional studies were conducted to verify that this subpopulation possessed the ability of self-renewal and producing differentiated progeny. New strategies addressing this disease with a paradigm shift in the mechanism of the therapeutic resistance and recurrence of pancreatic tumor can be developed with an improved understanding of the cellular signaling pathways in CSCs at the protein level. The xenograft model of primary human pancreatic cancer represents a significant advance for the study of pancreatic cancer. Animal models using cancer cell lines often do not recapitulate human

diseases accurately, where the biological characteristics and histology of tumor-derived human pancreatic cancer tissue are preserved in the xenograft.

The major obstacle to the study of global proteome expression profiling of the CSCs is the extremely small number of cancer stem cells available per study per tumor sample. Due to the unique features of pancreatic tissue and low percentage of pancreatic cancer stem cells (0.2% to 0.8%), from a single human tumor xenografted in a mouse, we can typically obtain 10k antibody labeled cancer stem cells using flow cytometry, which corresponds to around 1ug of total protein. Current publications of proteomic studies using capillary scale shotgun approaches are based on the analysis of entire tissue sections or cell lines instead of a subpopulation of primary human cells[3]. The amount of material consumed in a proteomic study using shotgun approaches such as MudPIT or offline 2D-LC/MS/MS is higher than 20 ug[4-6]. Some studies targeting a certain subpopulation of cells have restricted the analysis to one dimensional separation before mass spectrometry[7, 8] to avoid the sample loss in a 2^{nd} dimension of separation, though limiting the ability of identifying proteins present in lower abundance.

In the present study, we employed a PPS facilitated lysis procedure combined with a high resolution two-dimensional separation to accommodate the small sample size of this study. After cell lysis, protein extracts were digested and equally divided into three aliquots. Each aliquot of around 300 ng total material was then introduced into a two-dimensional separation by CIEF and nano-RPLC followed by tandem mass spectrometry analysis to identify the proteins present. A label-free protein quantification using spectral counts was employed to measure the protein fold changes between the CSC group and the bulk tumor group. Ultimately, the signature proteins detected by the method were

then uploaded to Ingenuity Pathway Analysis (IPA) for functional analysis to identify signaling pathways and protein-protein interaction networks that were significant in CSCs compared to bulk tumor cells.

4.2 Experimental:

4.2.1. Starting Material Preparation

a. Primary tumor specimen implantation.

Samples of human pancreatic adenocarcinomas were obtained within 30 min following surgical resection according to Institutional Review Board–approved guidelines. Tumors were suspended in sterile RPMI 1640 and mechanically dissociated using scissors and then minced with a sterile scalpel blade over ice to yield two 2–mm pieces. The tumor pieces were washed with serum-free PBS before implantation. Eight-week-old male NOD/SCID mice were anesthetized using an i.p. injection of 100 mg/kg ketamine and 5 mg/kg xylazine. A 5-mm incision was then made in the skin overlying the mid-abdomen, and three pieces of tumor were implanted subcutaneously. The skin incision was closed with absorbable suture. The mice were monitored weekly for tumor growth for 16 weeks.

b. Preparation of single-cell suspensions of tumor cells.

Before digestion with collagenase, low passage primary human pancreatic xenograft tumors were cut up into small pieces with scissors and then minced completely using sterile scalpel blades. To obtain single-cell suspensions, the resultant minced tumor pieces were mixed with ultrapure collagenase IV (Worthington Biochemicals, Freehold, NJ) in medium 199 (200 units of collagenase per ml) and allowed to incubate at 37 C for 1.5 to 2 hrs for enzymatic dissociation. The specimens were further mechanically dissociated every 15 to 20 min by pipetting with a 10-ml pipette. At the end of the

incubation, cells were filtered through a 40-Am nylon mesh and washed with HBSS/20% fetal bovine serum (FBS) and then washed twice with HBSS.

c. Flow cytometry.

Dissociated cells were counted and transferred to a 5-mL tube, washed twice with HBSS containing 2% heat-inactivated FBS, and resuspended in HBSS with 2% FBS at a concentration of 10^6 per 100 ul. Sandoglobin solution (1 mg/ml) was then added to the sample at a dilution of 1:20 and the sample was incubated on ice for 20 min. The sample was then washed twice with HBSS/2% FBS and resuspended in HBSS/2% FBS. Antibodies were added and incubated for 20 min on ice, and the sample was washed twice with HBSS/2% FBS. When needed, a secondary antibody was added by resuspending the cells in HBSS/2%FBS followed by a 20-min incubation. After another washing, cells were resuspended in HBSS/2% FBS containing 4,6-diamidino-2-phenylindole (DAPI; 1 Ag/mL final concentration). The antibodies used were anti-CD44 allophycocyanin, anti-CD24 (phycoerythrin), and anti-H2K (PharMingen, Franklin Lakes, NJ) as well as anti–ESA-FITC (Biomeda, Foster City, CA), each at a dilution of 1:40. In all experiments using human xenograft tissue, infiltrating mouse cells were eliminated by discarding H2K (mouse histocompatibility class I) cells during flow cytometry. Dead cells were eliminated by using the viability dye DAPI. Flow cytometry was done using a FACSAria (BD Immunocytometry Systems, Franklin Lakes, NJ). Side scatter and forward scatter profiles were used to eliminate cell doublets. Cells were reanalyzed for purity, which typically was >97%.

d. Material

10k Pancreatic cancer stem cells and 100k bulk tumor cells were obtained from mice xenografts after sorting by flow cytometry and gently washed three times with cold PBS(pH 7.4) by repetitive pipetting, followed each time by centrifugation at 1000g for 5min at 4°C. In the third time of washing, excessive PBS was gently sucked off with extra caution when cell pellets were observed at the bottom of the tube.

4.2.2 Cell lysis and trypsin digestion

PPS (Protein Discovery, Knoxville, TN) powder was dissolved in 50 mM Ammonia Bicarbonate and was added to each tube at a final concentration of 0.2%(m/v). Around a 100 ul cell suspension was then vortexed and incubated at 60°C for 10min, followed by sonication in an ice-water bath for 2 hrs. An aliquot of 5 mM DTT was added and the mixture was incubated at 60°C for 30min. After cooling, 15 mM iodoacetamide was added and the mixture was placed in the dark at room temperature for 30 min in order to allow the carboxymethylation reaction of cysteine residues. 50 mM ammonia bicarbonate was then added at a dilution ratio of 1:5 and 1:50(w/v) L-1-tosylamido-2-phenyletyl chloromethylketone modified sequencing-grade porcine trypsin (Promega, Madison, WI) was added. The mixture was incubated at 37°C in a water bath with agitation. Formic acid (FA) was then added to make a final concentration of 2% to stop the proteolysis. Following termination, the acidified mixture was placed in a 37°C water bath again for 4 hrs to facilitate the hydrolysis and allow the cleavage of PPS. The acidified tryptic peptide mixture was then desalted by a peptide micro-trap (Michrom, Auburn, CA) and eluted with 98% acetonitrile (ACN) and 0.3% FA, followed by spinning to dryness using a SpeedVac concentrator (Labconco, Kansas City) and stored in the -80°C freezer for future use. All chemicals were purchased from Sigma unless mentioned otherwise.

4.2.3. First dimensional separation: CIEF

A Beckman CE instrument was modified for CIEF with fraction collection. A 70 cm CIEF (100 mm id, 365 mm od) capillary was coated as previously described[9]. Lyophilized peptides were first reconstituted in gel buffer containing 2% ampholyte (pH 3-10) and were injected hydrodynamically to fill the capillary. Peptide focusing was performed by applying 21kv voltage to the two ends of the capillary using 0.1M phosphate acid and 1mM sodium hydroxide as the anolyte and catholyte, respectively. The cathodic end of the capillary was kept in a stainless steel coaxial device. As the current reached its plateau, the focusing was complete and the focused peptides were mobilized under pressure and eluted into a 96-well auto-sampler plate by a 2uL/min flow of catholyte solution delivered by a syringe pump. The auto-sampler plate was moved from well to well automatically every 2 minutes by a Beckman fraction collector.

4.2.4 LC/MS/MS

When CIEF separation was completed, all CIEF runs with each containing ~30 pI fractions were injected in randomized order via Paradigm auto-sampler (Michrom Biosciences, Auburn, CA) and loaded onto a desalting nano-trap (300 µm x 50mm) (Michrom) connected to a nano-RP column (C18AQ, 5 mm 200A, 100 µm x 150 mm)(Michrom) by a Paradigm AS1 micro-pump (Michrom). The mobile phases A and B were composed of 0.3% FA in water and 0.3% FA in ACN, respectively. Peptides were first desalted and enriched starting at 100%A with a flow rate of 50 mL/min for 5 min. The sample was subsequently separated by a nano-RP column with a flow rate of 0.3 mL/min after splitting. The linear gradient for separation was as follows: from 3% ACN to 12% ACN in 5 min, from 12% ACN to 40% ACN in 30min, from 40% ACN to 80%

ACN in 15 min and finally decreased from 80% ACN to 3% ACN in 10min. The resolved peptides were then introduced into a ThermoFinnigan linear ion trap mass spectrometer (LTQ) (Thermo Electron, San Jose, CA) equipped with a nano-spray ion source (Thermo Electron). The LTQ was operated in data dependent mode in which one cycle of experiments consisting of one full MS scan was followed by five pairs of zoom scans and MS/MS scans with dynamic exclusion set to 30 s. The capillary temperature was set at 175°C, spray voltage was 2.8 kV, capillary voltage was 30 V and the normalized collision energy was 35% for the fragmentation.

4.2.5 Database search and protein identification

MS/MS spectra were then searched against the human UniProt FASTA database by TurboSEQUEST provided by Bioworks ver3.1 SR1 (Thermo-Finnigan). The following modification was allowed in the search: 15.99 Da shift for oxidized Met residues; 58.1 Da shift for carboxymethylated Cys residues. The identified peptides were subsequently processed through PeptideProphet and ProteinProphet incorporated in the trans-proteomic pipeline (TPP: http://proteinprophet.sourceforge.net/prot-software.html) where each protein was assigned with a probability indicating the significance level of the protein appearing in the original sample. In this study, we used a protein probability score of 0.99 as the threshold for protein identification and the FDR is below 0.2%.

4.2.6 Label-free protein quantitation and data transformation

Spectral counts were parsed out of TPP xml files after processing the SEQUEST data and used as a surrogate measure of protein abundance in our analysis. Global normalization was used to reduce technical bias when acquiring spectral count data from different runs between and across samples. The bias may come from instrument error or the inherent

random sampling nature of the LTQ. Three replicated datasets for the CSC group and four replicated datasets for the tumor group (denoted as CSC1, CSC2, CSC3 and tumor1, tumor2, tumor3, tumor4), containing proteins with 99% confidence or above, together with their spectral counts, were generated. The data were consolidated to form a matrix with seven columns; missing values were replaced with zero. To eliminate the discontinuity observed in simple count ratios when a protein shows spectral count 0, raw data were transformed according to Old *et al*[10]. The transformation uses the $\log_2$ scale quantity

$$N= \log_2[(n+f)/(t-n+f)]$$

(Eq.1)

for each protein, where n is the raw (globally normalized) spectral count value; t is the total number of spectra over all proteins in each dataset; and f is a correction factor. Larger values of f shrink the results for low spectral count proteins toward zero, thereby eliminating the discontinuity at zero and down weighting the results with greatest measurement error (i.e. the proteins with low spectral counts). Several procedures for setting the constant term f have been proposed; we devised a new approach that is suitable for experiments with technical replicates. We considered the following criterion:

$$R(f) = \Sigma \text{ cor(between replicates)}$$

(Eq.2)

Correction factor f is defined to be the value that maximizes the correlation given in (Eq. 2). This maximizing effect yielded the value f=3. A schematic view of this computational process is depicted in Figure1(b). After transformation, statistical significance levels between the CSC group and the tumor group were then determined by

student's t-test followed by multiple testing adjustment using FDR test. Differentially expressed proteins used for subsequent pathway analysis were declared at the level of q-value < 0.1. This FDR test was performed using R-package (http://cran.r-project.org/web/packages/fdrtool/index.html). Fold change (FC) is computed from transformed data using the mean of spectral counts from all replicates within a group: FC = (mean CSC group) – (mean tumor group), where a positive sign indicates over expression in the CSC group and a negative sign indicates over expression in the tumor group.

4.2.7 Ingenuity pathway analysis (IPA)

To obtain detailed molecular information and infer significant signaling pathways from our global profiling results, differentially expressed proteins from the CSC group and tumor group were uploaded to IPA. The uploaded Excel spreadsheet file contains the relevant proteins with their fold change, q-value and corresponding primary accession number. The significance values for canonical pathways were calculated using the right-tailed Fisher's Exact Test by comparing the number of proteins that participate in a given function or pathway relative to the total number of occurrences of these proteins in all functional/pathway annotations stored in the ingenuity pathway knowledge base (IPKB).

4.3 Results and Discussion:

4.3.1 Evaluation of CIEF+RPLC platform

Capillary isoelectric focusing is a powerful 1st dimensional separation for protein/peptides because of its high resolution and orthogonal separation mechanism versus RP-HPLC. This pI based separation provides an optimal resolution of 0.01 pH unit, which indicates a peak capacity of 700 in a pH range from 3 to 10, while strong cation

exchange only has a peak capacity of around 50. SCX also presents undesired retention of peptides with strong interaction with the chromatographic resin and results in poor sample recovery rate. In contrast, CIEF is performed in an open capillary which is usually neutrally coated to prevent electroosmotic flow (EOF) and absorption of samples. Thus CIEF usually provides a sample recovery rate of higher than 90% which is critical in analyzing the extremely small amount of sample in our pancreatic CSC study.

The quality of the CIEF separation in terms of the resolution and reproducibility is essential to the accuracy of the comparative proteomic study. The theoretical pI value for each identified peptide within each fraction was calculated after database searching. The pI distribution plot from the first replicate of the CSC group is shown in Figure 2(a). As expected, the pIs of the fractions decrease from 10 to 4 following a linear trend except the first 7 fractions where only a few proteins were identified. Peptides beyond the pH range of the ampholytes (pH 3–10) used in these experiments were not expected to be resolved. Overall, CIEF exhibited high separation resolution where more than 70% of the peptides were identified in no more than a single fraction. In addition, the off line collection method has the advantage of maintaining the separated peptide bands, whereas on line collection directly coupled to RPLC reduces the workload at the cost of sacrificing separation resolution and increased sample loss. The number of peptides identified from the four replicate runs of the same tryptic digest from the tumor sample is plotted against their pI values shown in Figure 2(b). The distribution of the peptides demonstrates excellent reproducibility of the experiment.

Reproducibility is also accessed by pairwise comparison of two selected replicates (CSC replicate run 1 and 2, Tumor replicate run 1 and 2) using the Pearson correlation

coefficient. No transformation was performed at this point. Common proteins identified in both replicate runs were used to calculate the correlation values (R) which are 0.87 with 95% confidence level between [0.84, 0.9] for CSC1 vs CSC2 shown in Figure2(c) and 0.91 for tumor1 vs tumor2 shown in Figure2(d) with 95% confidence level between [0.89, 0.93] shown in Figure2(d). A complete correlation matrix can be found in the supporting material1.

4.3.2 Spectral counting results and transformation

Detecting relative protein quantity change between various disease classes or different samples is central to understanding the molecular processes of the cell. Currently there are two widely used but fundamentally different label-free protein quantification strategies: spectral counting and peak-area measurement. The latter method requires aligning the retention time of the chromatogram peaks to accurately locate the same peptide and is preferred in experiments where peptides are subject to only one-dimensional separation before being analyzed by mass spectrometry for its accuracy, while the peak-area method is difficult to use with approaches containing two-dimensional separations where the same peptide might appear in more than one LC/MS/MS run. Spectral counting is relatively easy to apply to two-dimensional separation data. We have previously reported a study using spectral counting to analyze the protein expression levels in two ovarian cancer cell line samples and spectral counting results of selected proteins were all consistent with western blot experiments[9].

Different types of computational algorithms regarding the processing strategies of spectral counting data have been proposed by a number of groups[10-13]. Zybailov et

al.[11] proposed a NSAF method to normalize and transform spectral count data based on protein length whereas Old et al.[10] adopted a transformation method to avoid the discontinuity problem which was originally used for serial analysis of gene expression. In the present study, spectral counts were assigned to each identified protein followed by global normalization by adjusting the mean of each dataset to be equal. After consolidating all 7 datasets into a matrix and replacing missing values with zero, we proceeded to identify the correction parameter f that produces the most meaningful abundance measurements for our experimental data. The transformation (Eq. 1) that we applied has the effect of shrinking the expression scores for low spectral count proteins toward zero, compensating for their relatively greater uncertainty. By maximizing R(f) (Eq. 2), the data was transformed in such a way that low-abundance proteins were appropriately, but not excessively down weighted in the analysis. The approach to defining f by maximizing R(f) is based on the fact that technical replicates from the same biological sample should be intrinsically the same. The correction factor was calculated to be 3 and then the whole dataset was transformed accordingly. This is similar to Old's method[10] that f is adjusted according to higher correlation between expected and observed. The fitted transformation, shown in Figure3, indicated that above 300 expression units on the original scale, the measured data were sufficiently reliable to be used without adjustment; between 100 and 300 expression units on the original scale the data were substantially shrunk toward zero, but still contributed to the analysis; and below 100 units, the data were shrunk to the degree that they have little influence in the analysis.

An important issue with any data transformation scheme that uses the class labels or clinical outcome information is the possibility that it may induce artifactual correlations between protein abundance and outcomes. The transformation function (Eq. 1) was monotonic, meaning that if the spectral counts of protein A were larger than those of protein B, this relationship would continue to hold after the transformation is applied. A monotonic transformation is limited in its ability to induce spurious correlations. To further explore whether the transformation induced spurious correlations, we conducted a simulation study. We simulated spectral counts for 1000 proteins measured in two groups of samples, each with three replicates. All 6,000 data values were simulated independently from a standard exponential distribution. We applied the procedure described above for a range of f values from 0 to 100, and considered the number of Z-statistics (comparing the two groups of three samples for a given protein) that exceeded various thresholds (2, 2.5, and 3). We did not observe any inflation in the number of significant associations for f>0 compared to f=0. Furthermore, we observed in our experimental data that the results were not very sensitive to the specific value of f, as long as it was not too close to zero. Figure 4(a) shows the clustering results after transforming by f =3. The CSC group and the tumor group are well separated without introducing artificial interference.

4.3.3 Protein profiling results

Accepted protein identifications are obtained after applying a cutoff protein probability of 0.99 by TPP which ensures the FDR is below 0.2%. A total of 763 and 1031 distinct proteins were identified from three CSC replicates and four tumor replicates, respectively. The detailed information about these identifications can be found from supporting

material2. Each identified protein was assigned a cellular location based on information from IPKB. Figure 5(a) and Figure5(b) show the cellular distribution of 763 and 1031 identified proteins from the CSC group and the tumor group, respectively. The cellular distribution is consistent for both of these two groups: the majorities are cytoplasmic and nuclear proteins; plasma membrane proteins occupy 10% of each total proteome, suggesting PPS has the ability to extract hydrophobic proteins. Multiple correcting testing (FDR) was then performed to capture the differentially expressed molecules and 161 out of 1159 proteins were identified by using a threshold of q-value < 0.1. 24% of these differentially expressed proteins show up-regulation in the CSCs group and a few of them are related to the key signaling pathways of CSCs. For example, inter-alpha trypsin inhibitor H3 (ITIH3) was identified in all 3 runs of the CSC group with two unique peptides and an average spectral count larger than 10, while it was only identified in one of the tumor runs with low spectral counts. This protein associated with inflammatory response in local tissue[13]was previously reported to be one of the downstream target genes of Sonic Hedgehog (Shh), which plays a key role in signaling pathways that directly activate the genes involved in the self-renewal and apoptosis-inhibition functions of CSC[14] . The over-expression of ITIH3 is consistent with previous finding of the up-regulated Shh at the mRNA level in pancreatic CSCs[2]. In contrast, a mitochondrial apoptosis-inducing factor (AIFM1) was down-regulated in the CSC group. This protein was identified in all 4 runs of the tumor group with 5 unique peptides, whereas it was only identified in one of the CSC runs. The decreased protein level detected in the stem cell group from our study agrees with previous reports that inactivation of AIFM1 renders embryonic stem cells resistant to cell death [15]. Beside these significantly

differentially expressed proteins, some important low-abundant proteins were also detected although they were not found to be significantly different based on the FDR test. For example, NF-κβ was identified in one of the CSC runs with two unique peptides but not in any of the tumor runs, as well as c-MET and CXCL5. Their absence in the tumor group is mainly due to their low abundance level which was below the detection limit, however their relative over expression in the CSC group agrees with previous findings [16-20] that the elevated expressions of these proteins are related to the properties of CSCs.

4.3.4 Signaling Pathway and Connectivity Network Analysis

A list of 169 differentially expressed proteins was uploaded into IPA for functional annotation and pathway analysis. The detailed information about these molecules can be found in supporting material2. The most variant and relevant canonical signaling pathways enriched with differentially expressed molecules between the CSC group and the tumor group were generated by IPA and are ranked by significance shown in figure6 with a threshold of p-value < 0.1. The length of the bars indicate the significance of the signaling pathways to which the differentially expressed proteins are related.

These enriched pathways can be grouped into four main categories related to the characteristics of cancer stem cells: resistance to apoptosis, dysregulation of cell proliferation, association with inflammation and metastasis. The top pathway, Mitochondrial Dysfunction, is related to both apoptosis and tumorigenesis[21, 22] and a recent report has linked mitochondrial dysfunction to ovarian cancer stem cells[23]. Pathways mapped to cellular growth, proliferation and development by IKGB include ILK Signaling, RhoA Signaling and Integrin Signaling. CXCR4 signaling and Acute

Phase Response signaling pathways categorized under cellular immune/inflammatory response are also shown to be significantly involved. The connection between inflammation and tumorigenesis has been recognized in many pathologic conditions including pancreatic cancer where some clues are suggesting that inflammation might induce an accelerated process of mutagenesis and mutation accumulation[1, 24]. VEGF signaling associated with angiogenesis is also shown to be significant.

To further infer the functional relevance between these differentially expressed molecules, we have constructed connectivity networks by IPA. The top network which has the highest score of 54 (p-value < $1x10^{-54}$ from Fisher Exact Test) consists of 27 signature proteins shown in figure7. Interestingly, NF-κβ is imported as the central node to generate this interaction network by IPA, suggesting a potential involvement of this transcription factor although this molecule is not in the experimental signature protein list.

4.4 Conclusion:

This work represents the first proteome profiling study on pancreatic cancer stem cells from xenograted tumor in mice. We overcome the difficulty of analyzing the extremely small number of cancer stem cells from the xenografted tumor by using an ultrasensitive sample preparation procedure and CIEF as the 1^{st} dimensional separation before LC/MS/MS to minimize sample loss and obtain high resolution of peptides. A modified transformation algorithm was also devised to handle spectral counting data with technical replicates in order to weight proteins with a large range of spectral counts appropriately in the subsequent statistical analysis. 169 proteins have been captured as differentially expressed signatures between the CSC group and the bulk tumor group. Pathway analysis

and network modeling by IPA has further revealed significantly involved signaling cascades relevant to the characteristics of CSCs.

It will be important to validate proteome profiling results using alternative technologies such as the Western blotting or RT-PCR. However, currently, no other techniques work effectively with less than 1ug of protein material. To compensate for the lack of validation, we increased the confidence level of protein identifications by adopting a more stringent threshold. We set the cutoff to a protein probability score of 0.99 which has a FDR less than 0.2% compared to the commonly used score of 0.9 which gives FDR less than 1%. Moreover, multiple testing corrections are employed to control false positives induced by performing significance tests on a large number of proteins. Thus, differentially expressed proteins identified in this way are more likely to be true positives.

Figure 4.1a Experimental flow chart

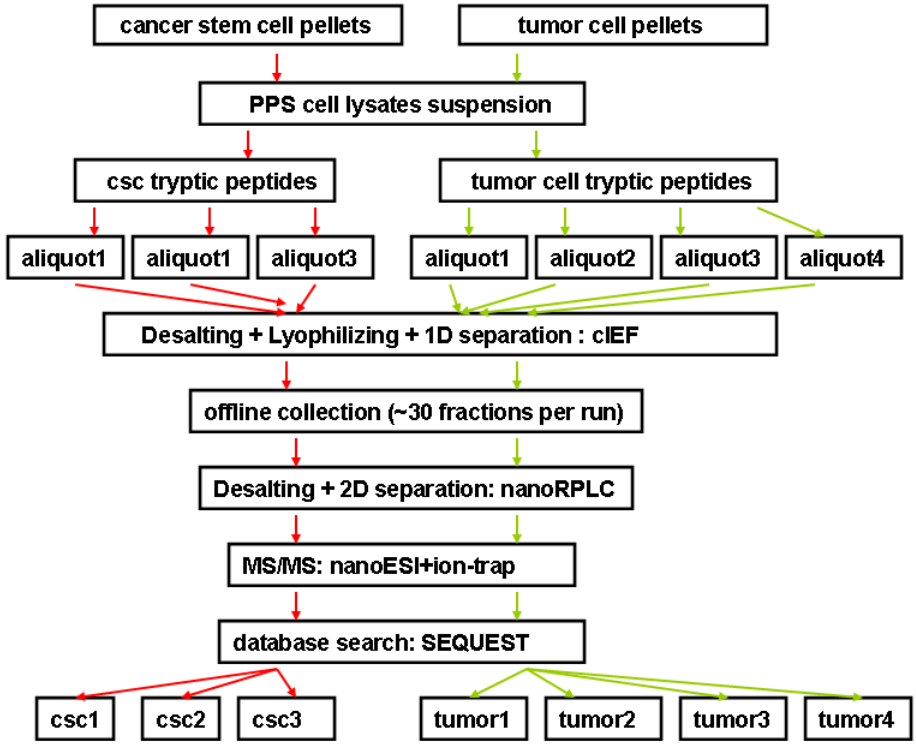


Figure 4.1b Data Processing Strategy. Upper left matrix is the consolidated dataset. Lower left flowchart is the correction factor searching scheme and transformation algorithm.

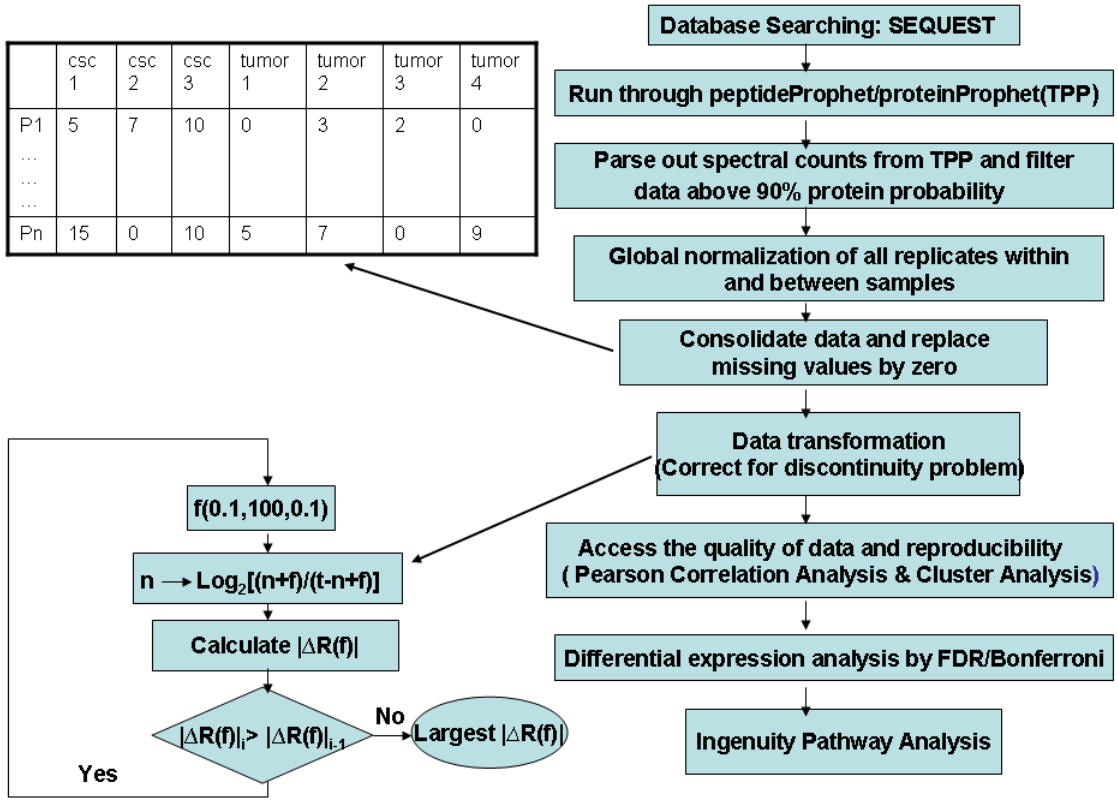


Figure 4.2a Theoretical pI distribution plot of the first run of CSC group. Fraction number shown in the X-axis is plotted against the average of peptides' pI value within each fraction shown in the Y-axis.

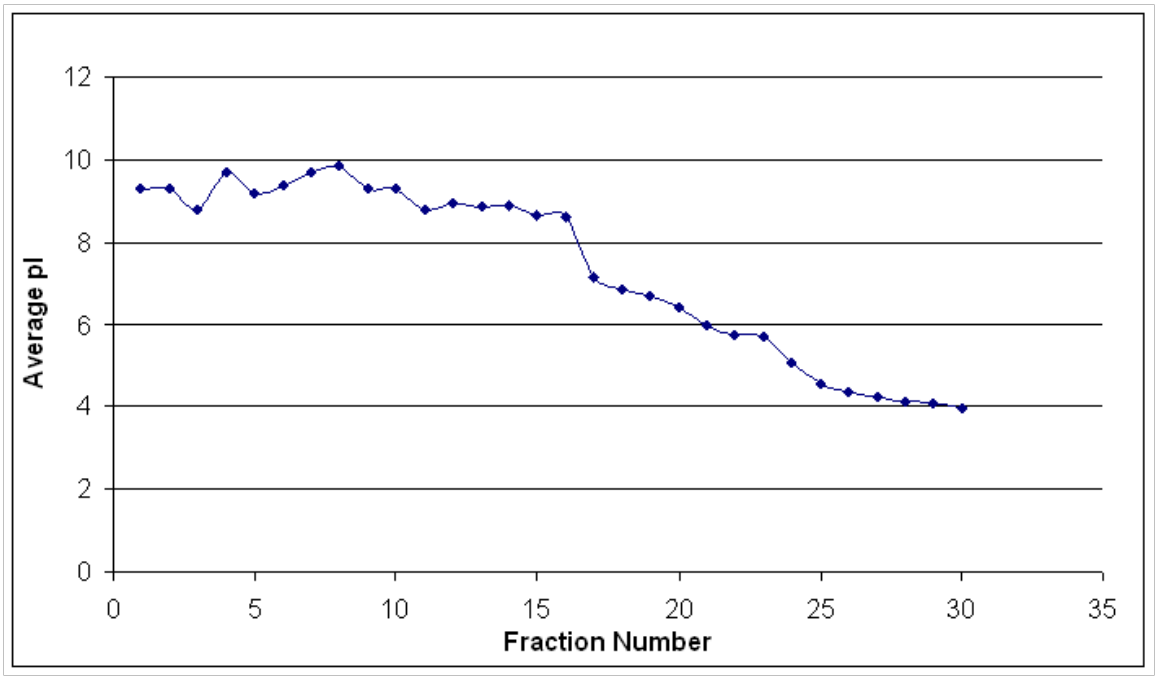


Figure 4.2b Distribution of number of identified peptides from each run of tumor group in pI range 3.5-10. X-axis shows their pI values and Y-axis shows co the number of identified peptides. Different tumor replicate runs are represented by different colors.

.

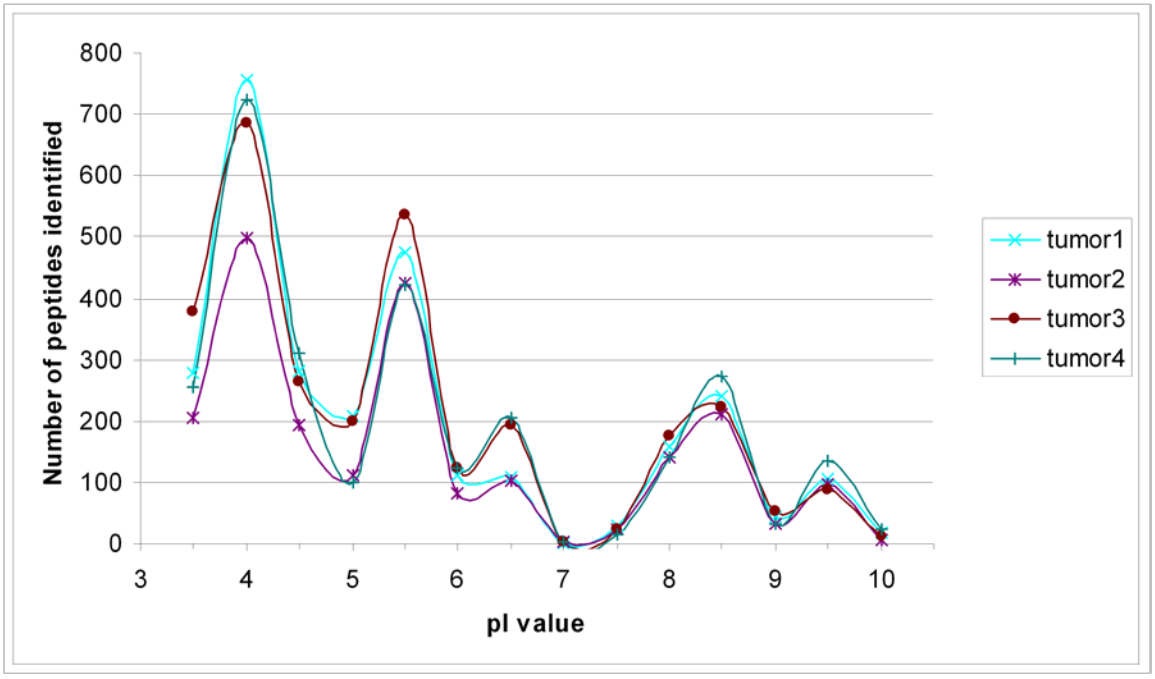


Figure 4.2c Pearson correlation plot of all proteins detected with single or more spectral counts in the first and the second run of CSC group.

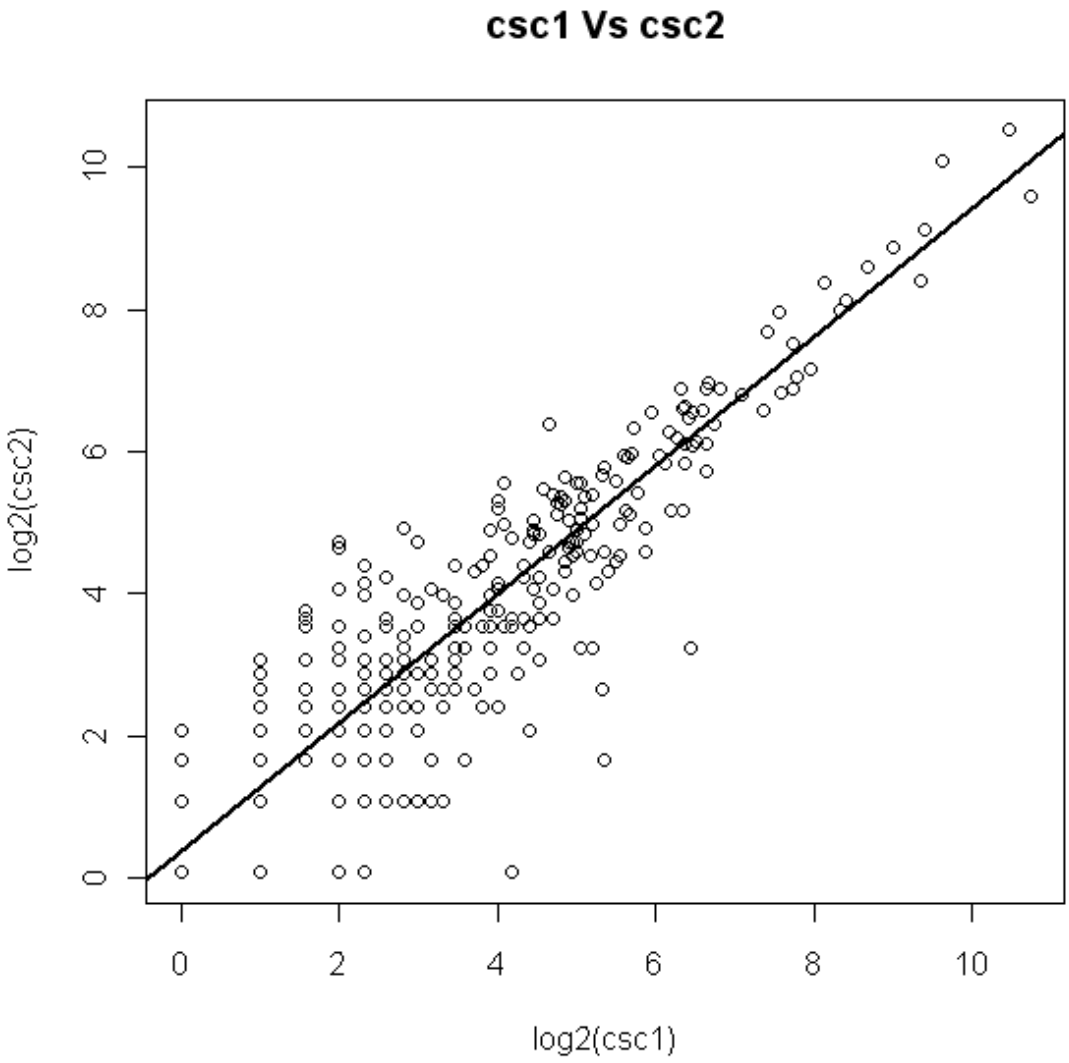


Figure 4.2d Pearson correlation plot of all proteins detected with single or more spectral counts in the first and second replicate of tumor group.

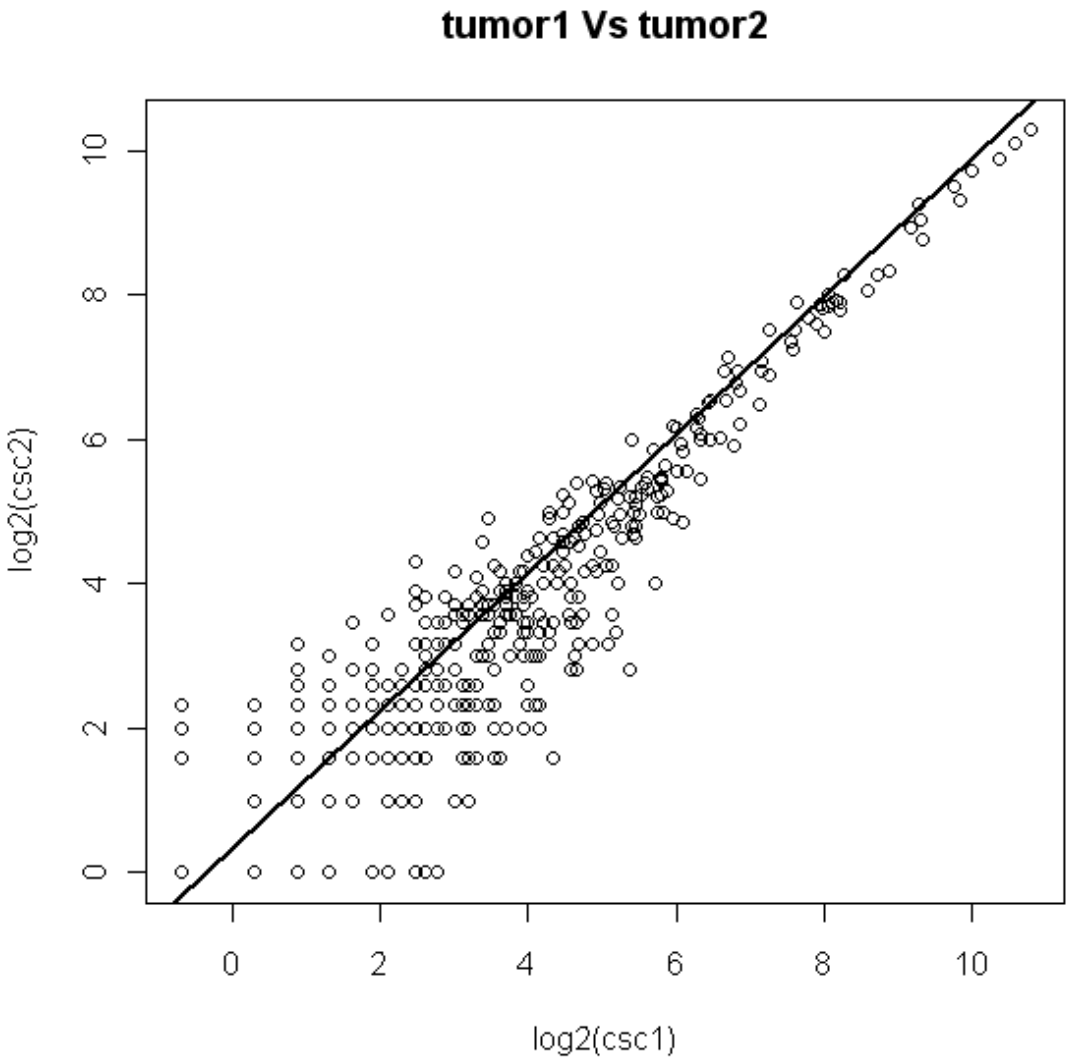


Figure 4.3 Monotonic plot of original data Vs transformed data. Different color and different shapes represent csc1, csc2, csc3, tumor1, tumor2, tumor3, tumor4, respectively. Y and X-axis represent transformed data on log2 scale.

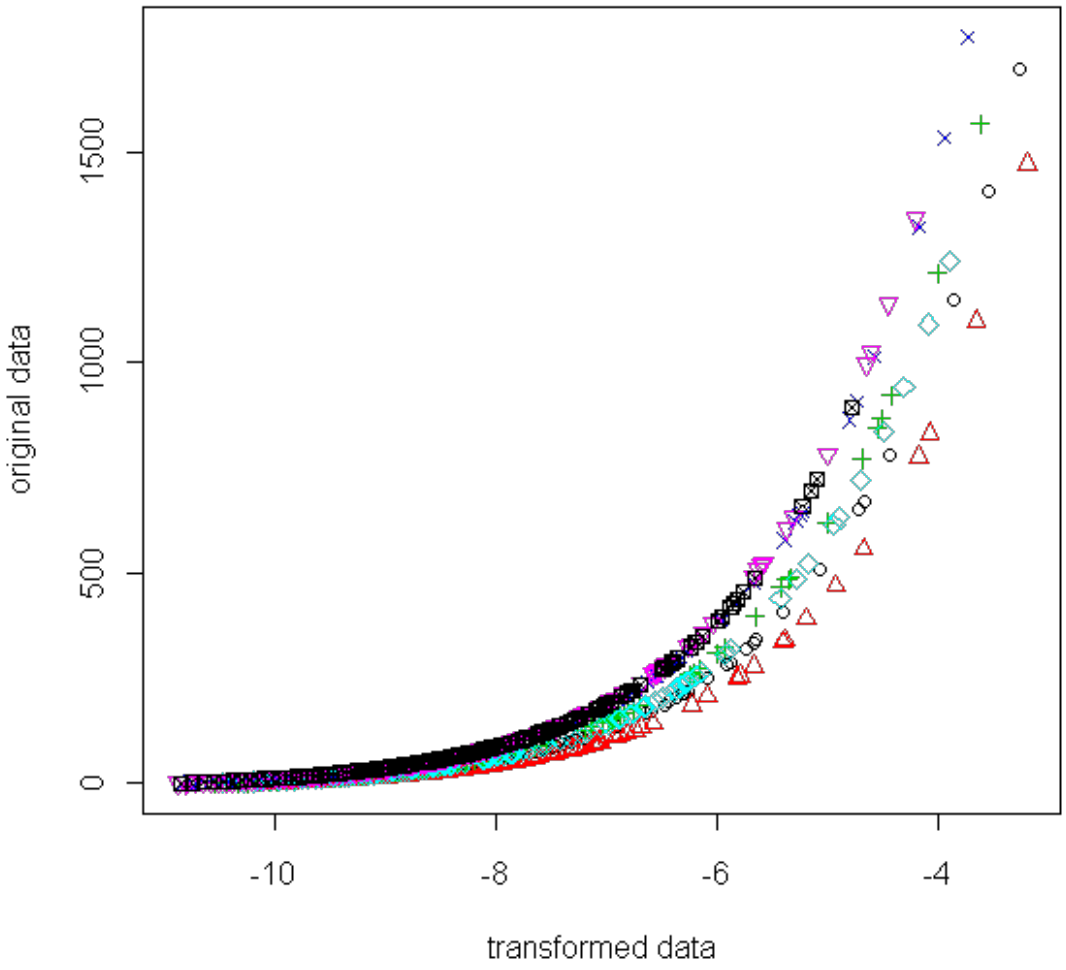


Figure 4.4 Clustering results of the CSC group and tumor group after transforming by f=3.

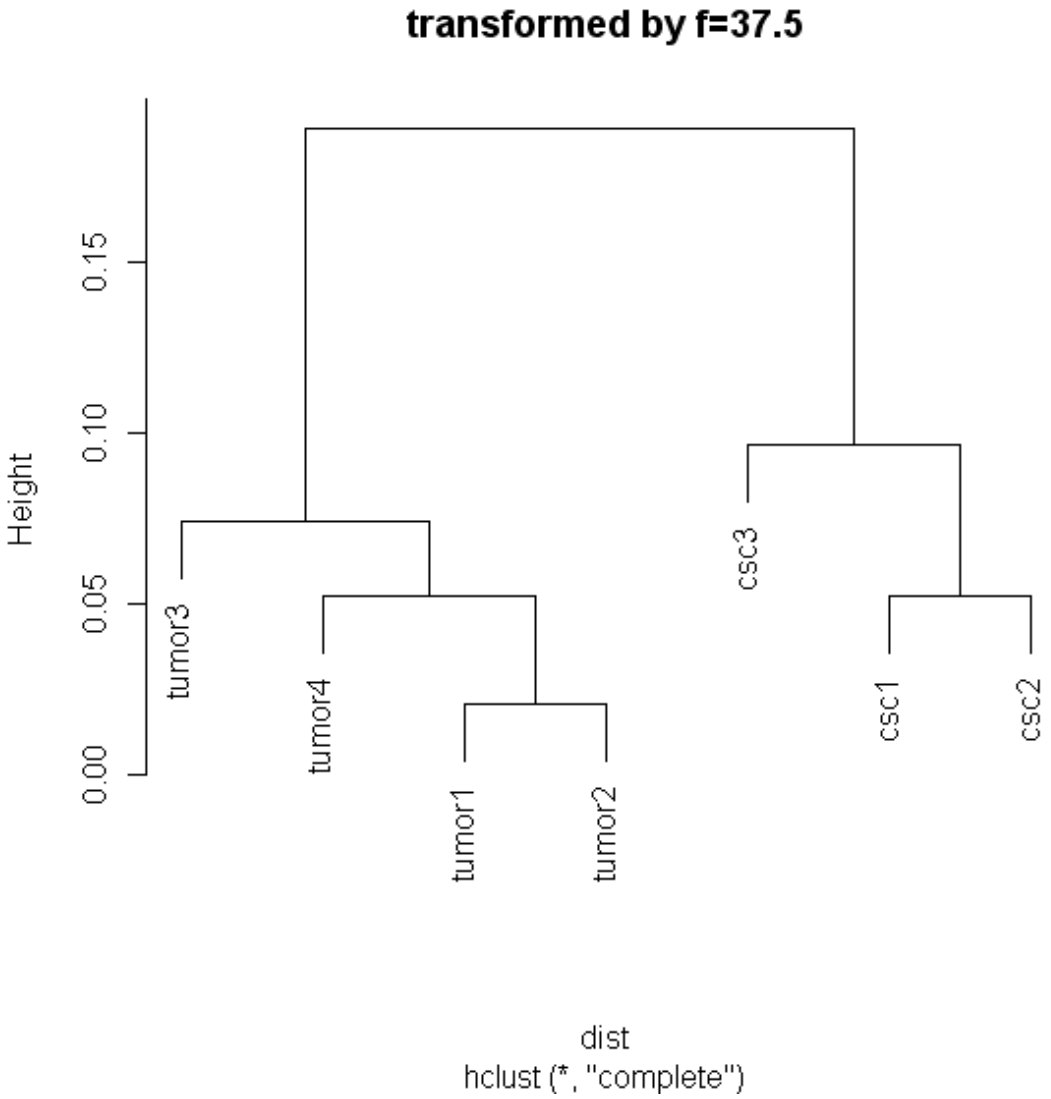


Figure 4.5a Cellular Distribution of identified proteins from pooled CSC group

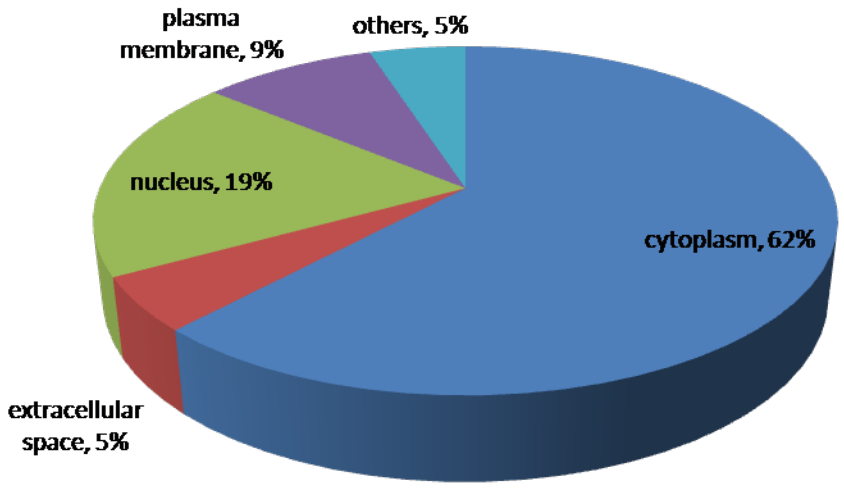


Figure 4.5(b): Cellular Distribution of identified proteins from pooled tumor group.

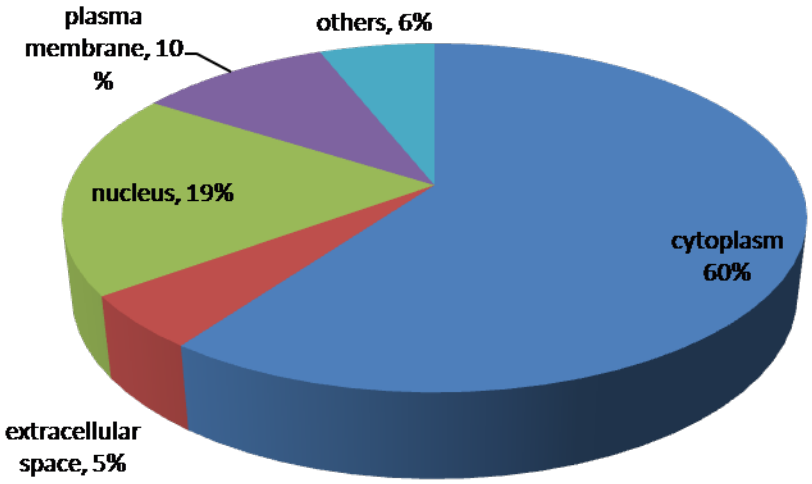


Figure 4.6 Canonical signaling pathways enriched with differentially expressed proteins ranked significance. A threshold p-value < 0.1 is applied.

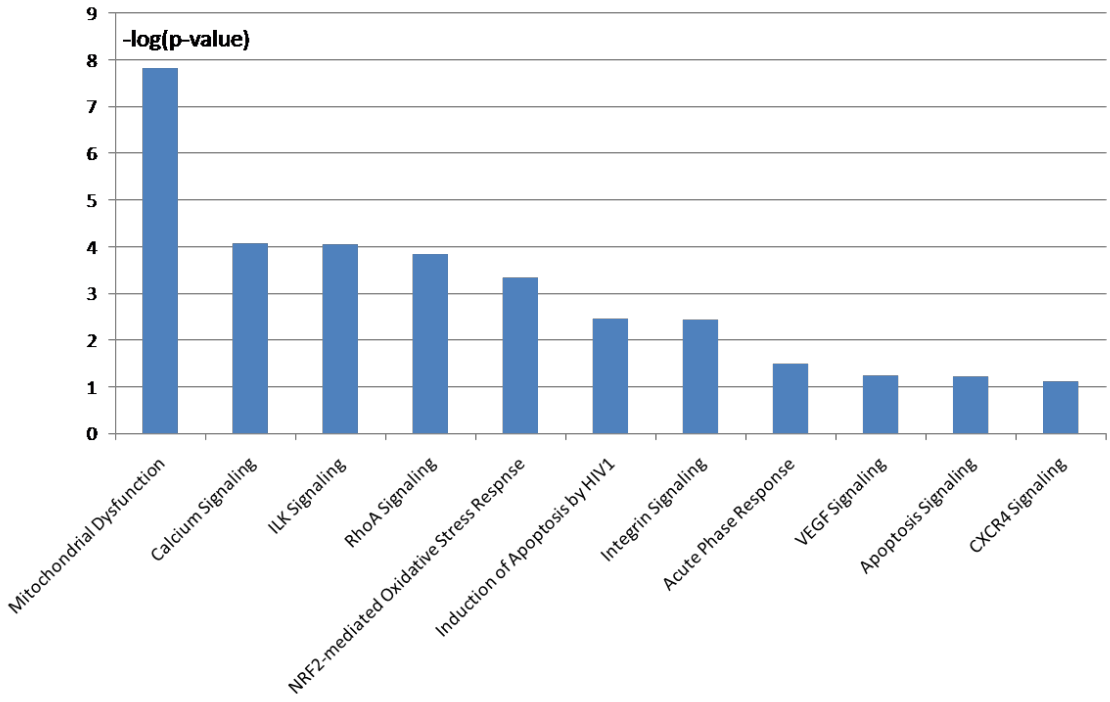
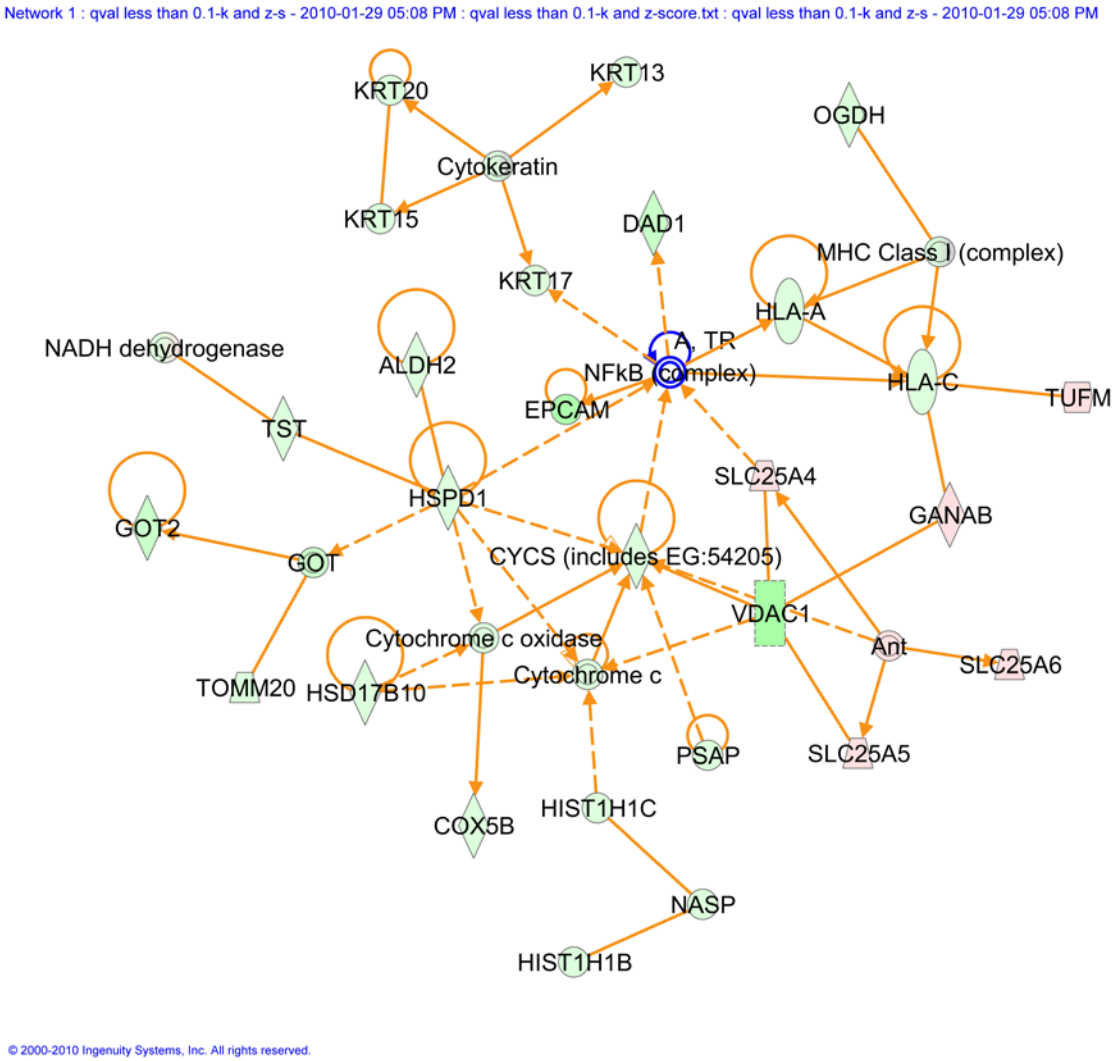


Figure 4.7 The connectivity network constructed ranked on top by IPA. This network only consists of differentially expressed proteins from experimental data. Red and green circles indicate overexpression and underexpression in the CSC group versus the bulk tumor group, respectively.



Reference

1. Lee CJ, Dosch J, Simeone DM, J Clin Oncol 2008, 26, 2806
2. Li C, Heidt DG, Dalerba P, Burant CF, Zhang L, Adsay V, Wicha M, Clarke MF, Simeone DM, Cancer Research 2007, 67, 1030
3. Fang X, Balgley BM, Wang W, Park DM, Lee CS Electrophoresis 2009, 30, 4063
4. Gortzak-Uzan L, Ignatchenko A, Evangelou AI, Agochiya M, Brown KA, St Onge P, Kireeva I, Schmitt-Ulms G, Brown TJ, Murphy J, Rosen B, Shaw P, Jurisica I, Kislinger T, J Proteome Res 2008, 7, 339
5. Sodek KL, Evangelou A, Ignatchenko A, Agochiya M, Brown TJ, Ringuette MJ, Jurisica I, Kislinger T Mol Biosyst 2008, 4, 762
6. Elschenbroich S, Ignatchenko V, Sharma P, Schmitt-Ulms G, Gramolini AO, Kislinger T J Proteome Res 2009, 8, 4860
7. Bagnato C, Thumar J, Mayya V, Hwang SI, Zebroski H, Claffey KP, Haudenschild C, Eng JK, Lundgren DH, Han DK, Mol Cell Proteomics 2007, 6, 1088
8. Hwang SI, Thumar J, Lundgren DH, Rezaul K, Mayya V, Wu L, Eng J, Wright ME, Han DK, Oncogene 2007, 26, 65
9. Dai L, Li C, Shedden KA, Misek DE, Lubman DM, Electrophoresis 2009, 30, 1119
10. Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, Resing KA, Ahn NG, Mol Cell Proteomics 2005, 4, 1487
11. Zybailov B, Mosley AL, Sardiu ME, Coleman MK, Florens L, Washburn MP, J Proteome Res 2006, 5, 2339
12. Choi H, Fermin D, Nesvizhskii AI, Mol Cell Proteomics 2008, 7, 2373
13. Zhuo L, Kimata K, Connect Tissue Res 2008, 49, 311
14. Kato M, Seki N, Sugano S, Hashimoto K, Masuho Y, Muramatsu M, Kaibuchi K, Nakafuku M, Biochem Biophys Res Commun 2001, 289, 472
15. Joza N, S S, Daugas E Nature 2002, 410, 549
16. Donahue TR, Hines OJ, World J Surg 2009, 33, 710
17. Wente MN, Keane M, Burdick MD, Friess H, Büchler MW, Ceyhan GO, Reber HA, Strieter RM, Hines OJ, Cancer Lett 2006, 241, 221
18. Hansel DE, Rahman A, House M, Ashfaq R, Berg K, Yeo CJ, Maitra A, Clin Cancer Res 2004, 10, 6152
19. Alvero AB, Chen R, Fu HH, Montagna M, Schwartz PE, Rutherford T, Silasi DA, Steffensen KD, Waldstrom M, Visintin I, Mor G, Cell Cycle 2009, 8, 158
20. Birnie R, Bryce S, Roome C, Dussupt V, Droop A, Lang SH, Berry PA, Hyde CF, Lewis JL, Stower MJ, Maitland NJ, Collins AT, Genome Biol 2008, 9, R83
21. Newmeyer DD, Ferguson-Miller S, Cell 2003, 112, 481
22. Lu J, Sharma L, Bai Y, Cell Res 2009, 19, 802
23. Bapat, S, Semin Cancer Bio 2007, 17, 204
24. Guerra C, Schuhmacher A, Cañamero M, Grippo PJ, Verdaguer L, Pérez-Gallego L, Dubus P, Sandgren EP, Barbacid M, Cancer Cell 2007, 11, 291

Chapter 5

The identification of auto-antibodies in pancreatic cancer patient sera using a naturally fractionated panc-1 cell line

Abstract:

The immunogenic nature of cancer can be explored to distinguish pancreatic cancer from related non-cancer conditions. We describe a liquid-based microarray approach followed by statistical analysis and pre-validation for discovery of auto-immune biomarkers for pancreatic cancer. Proteins from the Panc-1 pancreatic cancer cell line were fractionated using a 2-D liquid separation method into over 1052 fractions and spotted onto nitrocellulose coated glass slides. The slides were hybridized with 37 pancreatic cancer sera, 24 chronic pancreatitis sera and 23 normal sera to detect elevated levels of reactivity against the proteins in spotted fractions. The response data obtained from protein microarrays was first analyzed by Wilcoxon Rank-Sum Tests to generate two lists of fractions that positively responded to the cancer sera and showed p-values less than 0.02 in the pairwise comparison between cancer specimens and normal and chronic pancreatitis specimens. The top 20 fractions in the cancer vs. normal category were combined and demonstrated the ability to detect 85% of cancer samples with 95% specificity. Outlier-Sum statistics were then applied to the microarray data to determine the existence of outliers exclusive in cancer sera. The selected fractions were identified by LC-MS/MS. We further confirmed the occurrence of outliers with three proteins among cancer samples in a validation experiment using a separate dataset of 168 serum samples containing 48 cancer sera and 120 non-cancer controls. Phosphoglycerate kinase

1 (PGK1) elicited greater reactivity in 20.9% (10 in 48) of the samples in the cancer group, while no outlier was present in the non-cancer groups.

5.1 Introduction:

Pancreatic adenocarcinoma is the 4th leading cause of cancer-related death in the United States [1]. Pancreatic ductal adenocarcinoma (PDAC) has one of the poorest survival rates of any cancers, where according to the American Cancer Society, for all stages of pancreatic cancer combined, the one-year relative survival rate is 20%, and the five-year rate is 4% [1]. These low survival rates result from the fact that fewer than 10% of patients' tumors are confined to the pancreas when in most cases, diagnosis of 80% to 90% of PDAC cases are too late for surgical procedures to have a positive outcome. Unfortunately there are not any available diagnostic tools that allow for detection of early stage pancreatic cancer. Although there has been an effort to find protein markers in serum, none have shown sufficient sensitivity and specificity for early diagnosis including the commonly used CA 19-9 test [2,3] which may be significantly increased in pancreatitis in addition to pancreatic cancer, and is not reliably elevated in early stage cancer.

There remains a need for the discovery of innovative serological biomarkers that effectively improve diagnosis and prognosis of human cancer. Antibody responses associated with the occurrence and progression of solid tumors have been identified in multiple cancer types [4-6]. The underlying mechanism of the auto-immune response is still not fully understood [7]. However, the known molecular changes that can induce auto-immune response include proteins expressed at an aberrant level, and mutated gene products and isoforms of proteins with abnormal post-translational modifications (PTMs)

[8-10]. The immunogenic proteins are often found to be intracellular proteins whose functions are linked to the onset and growth of malignant tumors, such as oncoproteins HER-2/Neu and c-MYC [11-13] and, tumor suppression proteins such as p53 [14].

Although research has suggested strong correlation between the presence of some auto-antibodies and the process of tumorgenesis, the frequency of the appearance of auto-antibodies in cancer patients varies. A mutation in the p53 gene elicits an auto-immune response in 4-30% of patients in several types of cancer [14]. Around 30% of patients with lung adenocarcinoma exhibit a humoral response to glycosylated annexins I and/or II whereas none of the noncancerous standards exhibit such a response [15]. In PDAC, auto-antibodies to DEAD-box protein 48 were observed in 33.3% of pancreatic patient sera, while none of the patients with benign disease and healthy controls showed reactivity against the antigen [16]. MUC1 [17,18], p53 [19] and Rad51 [18] have also shown restricted immune reactivity in a subset of cancer samples. The typical frequency of the detection of a particular auto-antibody in a cancer type is 10-20% and may not be sufficient when used as a biomarker individually, but may be combined as a panel for improved performance [20-22].

Adding new cancer specific auto-antigens to the existing biomarker repertoire is the impetus of developing analytical and statistical techniques for auto-immune response study. There are several approaches currently available for the identification of auto-antigens. One is targeting specific proteins or gene products that are known for their roles in cancer including p53, c-myc, and erB-2. This method only provides limited candidates for biomarker. Recombinant protein microarray produced from cDNA expression library has been used as a comprehensive antigen substrate to profile the auto immune reactivity,

however, it is unable to profile PTM dependent antibody-antigen interactions. The development of proteomic separation and identification techniques has benefited the discovery of auto-antibody biomarker, where proteins from tissue or cell line are fractionated by gel or liquid based multidimensional separations, while maintaining the natural PTMs and further identified by mass spectrometry. Few studies have examined the change of the auto-antibody profile in pancreatic cancer using such techniques. In current work, we have used liquid fractionation methods to produce microarrays for the humoral response experiment against a Panc-1 pancreatic cancer cell line. The methods that are used involve separating intact proteins from cell lysates using two dimensions. A total cell fractionation can be performed using chromatofocusing separation in the first dimension where the proteins are fractionated according to pI. Each fraction is then separated in a second dimension by nonporous silica RP HPLC [23]. Using this method isolated proteins in the liquid phase can be collected for spotting on coated glass slides [24]. The protein spots are probed for their humoral response by exposing them to sera from cancer and chronic pancreatitis patients, and normal individuals. This method offers a means for comprehensive proteomic analysis of proteins from large numbers of purified proteins as expressed in cancer cells while maintaining their PTMs that are often critical to the humoral response [25]. The method can produce arrays with over a thousand spots and can produce large numbers of slides for testing the response against a large number of patients.

In order to account for the limited subset of patients with a specific tumor type that develop a response to a specific antigen, we attempt to apply two statistical methods, Wilcoxon Rank-Sum Test and Outlier Sum Statistics [26], to find potential markers that

show different types of immune reactivity patterns. Based on our results we perform a prevalidation study of 5 potential markers for pancreatic cancer against recombinant proteins on a microarray based format against samples from pancreatic cancer, pancreatitis, diabetes type 2 and normal controls. For three of the five proteins, a substantial number of samples from the cancer group show higher reactivity than the non-cancer sample groups.

5.2 Experimental

5.2.1 Chemicals:

Methanol, acetonitrile, urea, thiourea, iminodiacetic acid, dithiothreitol (DTT), n-octyl-D-glucopyranoside (OG), glycerol, bis-tris, Trifluoroacetic acid (TFA), and PMSF (Phenylmethanesulfonyl fluoride) were purchased from Sigma (St. Louis, MO). Water was purified using a Milli-Q water filtration system (Millipore Inc., Bedford, MA) and all solvents were HPLC grade unless otherwise specified. Reagents used were in the purest form commercially available. Polybuffer 74 and polybuffer 96 were purchased from GE Healthcare Bio-Sciences Corp. (Piscataway, NJ). 1x PBS and ultra-pure DNase/RNase free distilled water were obtained from Invitrogen (Carlsbad, CA).

Eighty six serum samples were obtained at the time of diagnosis following informed consent using IRB-approved guidelines. Sera were obtained from patients with a confirmed diagnosis of pancreatic adenocarcinoma in the Multidisciplinary Pancreatic Tumor clinic at the University of Michigan Hospital. Inclusion criteria for the study included patients with a confirmed diagnosis of pancreatic cancer, the ability to provide written, informed consent, and the ability to provide 40ml of blood. Exclusion criteria included inability to provide informed consent, patient’s actively undergoing

chemotherapy or radiation therapy for pancreatic cancer, and patients with other malignancies diagnosed or treated within the last 5 years. Sera were also obtained from patients with chronic pancreatitis who were seen in the Gastroenterology Clinic at University of Michigan Medical Center and from control healthy individuals collected at the University of Michigan under the auspices of the Early Detection Research Network (EDRN). The mean age of the tumor group was 65.4 years (range 54-74years) and the chronic pancreatitis group was 54 years (range 45-65). The sera from the normal subject group was age and sex-matched to the tumor group. The chronic pancreatitis group was sampled when there were no symptoms of acute flare of their disease. All sera were processed using identical procedures. The samples were permitted to sit at room temperature for a minimum of 30 minutes (and a maximum of 60 minutes) to allow the clot to form in the red top tubes, and then centrifuged at 1,300 x g at 4^{o}C for 20 minutes. The serum was removed, transferred to polypropylene, capped tubes in 1 ml aliquots, and frozen. The frozen samples were stored at -70^{o}C until assayed. All serum samples were labeled with a unique identifier to protect the confidentiality of the patient. None of the samples were thawed more than twice before analysis. This study was approved by the Institutional Review Board at the University of Michigan Medical School.

5.2.2 Cell culture

The Panc-1 PDAC cell line was cultured in Dulbecco's modified Eagle medium supplemented with 10% fetal bovine serum, 100 units/mL penicillin and 100 units /mL streptomycin (Invitrogen, Carlsbad, CA). Upon reaching 80% confidence, the cells were washed twice in 10mL 1X PBS containing 4mM Na_3VO_4, 10mM NaF and one half of a protease inhibitor cocktail tablet. The sample was then solubilized in 300ul lysis buffer

consisting of 7M urea, 2M thiourea, 100mM DTT, 0.5% biolyte ampholyte 3-10, 2% OG, 4mM Na_3VO_4, 10mM NaF and 1mM PMSF at room temperature for 30min, followed by centrifugation at 35000 rpm at 4^oC for 1hr. The supernatant was stored at -80°C until further use.

5.2.3 2-D intact protein separation

Prior to CF, a PD10 column(Amersham Biosciences) was used to exchange the cell lysate from the lysis buffer solution to the CF buffer solution according to the manufacturer's protocols. The start buffer consisted of 6M Urea, 0.2% OG, 25mM bis-tris. The elution buffer solution was composed of 6M urea, 0.2% OG, and a 10 fold dilution of polybuffer 96 and polybuffer 74 in a ratio of 3:7. The pH of both buffer solutions (7.9, 4.0) was adjusted with saturated imminodiacetic acid. A chromatofocusing column (weak anion exchange HPCF-1D prep column, 250mm x 4.6mm ID, Eprogen, Darien, IL) was pre-equilibrated with the start buffer solution and 13mg of the cell lysate was injected into the CF column with multiple injections. Fractionation was started after switching elution buffer and a stable base line achieved. The pH fractions were collected in 0.3 pH intervals and pH was monitored using a flow-through on-line pH probe. UV absorption was recorded at 280 nm. When a pH of 4.0 was reached, elution buffer solution was switched to a 1M NaCl solution to wash the column followed by Isopropanol to elute out strongly bound proteins from the column. The collected fractions were stored at -80C°.

An ODSI-1 (8 x 33 mm) column (Eprogen, Inc.) was used to separate the pH fractions of the Panc-1 cell line after CF. Solvent A was 0.1% TFA in water and solvent B was 0.1%TFA in acetonitrile. The gradient was run from 5% to 15% B in 1min, 15% to 25% in 2 min, 25% to 31% in 2min, 31% to 45% in 10min, 41% to 47% in 6min, 47% to

67% in 4min, 67% to 100% B in 3min, and reduced to 5% B in 1min after maintaining 100% B for 1min. The flow rate was 1ml/min and the column temperature 65°C. UV absorption was monitored at 214 nm. The fractions were collected in 96 well plates and stored at -80°C.

5.2.4 Protein Microarrays

Approximately 30% of the total sample of the fractionated Panc-1 proteins obtained using 2D separation were transferred into 96-well printing plates (Bio-Rad) and lyophilized to dryness. The fractions were reconstituted in printing buffer which was composed of 62.5mM Tris-HCl (pH 6.8), 1% w/v sodium dodecyl sulfate (SDS), 5% w/v dithiothreitol(DTT) and 1% glycerol in 1 X PBS. Reconstituted fractions in the printing plate were placed in a shaker overnight at 4°C. The fractions from the printing plate were spotted onto nitrocellulose slides using a non-contact piezoelectric printer (nanoplotter 2 GeSiM). Each spot contained 2.5nL of liquid of ~450μm diameter, and the distance between spots was 600μm. Printed slides were dried on the printer deck overnight and stored in a fridge desiccated at 4C if the slides were not used immediately.

5.2.5 Hybridization of slides

The printed slides were blocked in a solution of 1% BSA in PBS-T (0.1%) overnight. Each serum sample was diluted 1:400 in probe buffer which consisted of 1% BSA, 0.5mM DTT, 5mM magnesium chloride, 0.05% Triton X-100, 5% glycerol in 1 X PBS. The slides were hybridized in diluted serum for 2hrs using a mini-rotator at 4°C. After hybridization, slides were washed five times using probe buffer for 5min each time, and then re-hybridized with goat-anti-human IgG conjugated with Alexafluor 647 (1μg/mL, Invitrogen, Calsbad, CA) for 1hr at 4°C. The slides were washed five times again with

probe buffer for 5min each and dried. All slides were scanned using an Axon 4000B microarray scanner (Axon Instruments Inc., Foster City, CA).

5.2.6 Data acquisition and analysis.

The residual 2/3rd of the sample in 96 well plates which was not used in microarray experiments were dried down to approximately 10μL and mixed with 10%(v/v) ammonium bicarbonate, 10% (v/v) DTT, and 1:50 ratio (v/v) TPCK-treated trypsin (Promega, Madison, WI). The solution was incubated at 37°C overnight and the tryptic digestion was terminated by addition of 2.5% (v/v) of TFA. The digested peptide mixture was analyzed by nano-flow reverse-phase LC/MS/MS using the LTQ mass spectrometer with a nano-spray ESI ion source (Thermo, San Jose, CA). The samples were separated using a (0.1 x 150mm) capillary reverse phase column (MichromBioresources, Auburn, CA) with a flow rate of 5ul/min. An acetonitrile:water gradient method was used, starting with 5% acetonitrile which was ramped to 60% in 25min and to 90% in another 5min. Both solvent A (water) and B (ACN) contained 0.3% formic acid. The electrospray voltage was 2.6kV, with a capillary temperature of 200°C and a capillary voltage of 4kV. The normalized collision energy was set at 35% for MS/MS. The MS/MS spectra obtained were analyzed using the Sequest feature of Bioworks 3.1 SR1, allowing only one missed cleavage during SwissProt human protein database searching. To further validate data obtained from Sequest, Protein prophet/peptide prophet software modified in house was used to provide a confidence level in identification of 95%. Since there might be more than one protein in a protein spot on the microarray slide, we compared proteins identified in adjacent fractions. If the spot that responded to the humoral response was unique and did not have an adjacent spot that lit up then the highest scoring

protein based on LC-MS/MS analysis and protein prophet/peptide prophet was considered as the likely identification. If more than one protein was identified in the spot, then we also performed mass spec analysis on the adjacent spots. If the proteins were identified in the adjacent spots that did not respond then they were likely not to be the protein with the humoral response in our unique spot. However, if adjacent spots also showed a humoral response then the protein present in all spots was considered as the most likely candidate.

5.2.7 Statistical Analysis:

GenePix 6.0 software was used to grid all spots, to determine the fluorescent intensities at wavelength 635nm and median local background intensities for each spot. Background subtracted data of the spots was taken into analysis if the foreground measure was at least 2X the background intensity measure. The signal intensities from all the slides are normalized to minimize experimental slide-to-slide variation. The data for each individual sample in the columns is centered by the median and scaled by the interquartile range (IQR). Two types of statistical analysis were applied to the normalized data in search of biomarkers with up-regulated response in the cancer samples compared to the normal and pancreatitis samples. The non-parametric Wilcoxon rank-sum test was employed to identify fractions showing a universally increased reactivity in the cancer samples. The Outlier-sum test was performed to select the fractions that react with only a subset of the samples in the cancer group.

Heatmap

Cancer and normal samples are clustered and presented in a heatmap containing normalized data of all the fractions. The heatmap and dendrogram are drawn in R.

Wilcoxon Rank-Sum Test

Two pair-wise Wilcoxon Rank-Sum Tests were performed between cancer versus normal and cancer versus pancreatitis. The fractions with a p-value less than 0.02 were further studied with a Receiver Operation Characteristic (ROC) curve to determine their sensitivity and specificity in differentiating the samples groups. The top ranked fractions are combined in summary ROC curves to improve the sensitivity and specificity. The Wilcoxon Rank-Sum Tests and the ROC curves are programmed in R. The combination of ROC curves is performed using Canonical Discriminal Analysis in Tanagra software.

Outlier Sum Statistics (OS)

The dataset is first standardized for each fraction by subsequently subtracting the median and dividing the median absolute deviation (MAD). The 75% quartile (q(75)) plus the interquartile range (q(75)+IQR) is used as a threshold. The data points beyond this threshold are defined as the outliers. The outlier-sum statistic is the sum of the values of these data points in the disease groups. Fractions with outlier sum statistics ranked top 5% and no outliers in the normal groups were considered to be differential. The overlapping fractions found in the comparisons between cancer/normal and cancer/pancreatitis are presented in bar graph form (Figure 4) (made in R with COPA package).

5.2.8 Validation Using Recombinant Proteins

Recombinant proteins were purchased from Abnova Coporation (Taiwan), and Genway Biotech Inc., (SanDiego, CA). The concentration of each recombinant protein was 10ug/mL. A piezoelectric non-contact printer (Nano Plotter, GeSIM) was used to print all the recombinant protein arrays on ultra-thin nitrocellulose slides (PATH slides, GenTel

Bioscience). Each spotting event that resulted in 500 pL of solution being deposited was programmed to occur 5 times/spot to ensure that 2.5 nL was deposited on each spot. Each recombinant protein was printed in triplicate and 14 identical blocks were printed on each slide. The slides were washed three times with 0.1% Tween in PBS buffer (PBS-T 0.1) and then blocked with 1% bovine serum albumin (Roche) in PBS-T 0.1 for one hour. The blocked slides were dried by centrifugation and inserted into a SIMplex (GenTel Bioscience) multi-array device which divides each of the slides by 16 small wells. The wells separate the neighboring blocks and prevent cross contamination. Serum samples were diluted 10 times with PBS-T 0.1 containing 0.1% Brij. One hundred microliters of each diluted sample was applied to the recombinant protein array and the hybridization was performed in a humidified chamber for one hour. The 168 samples from different groups were perfectly balanced on each slide to eliminate bias from block-to-block variation and slide-to-slide variation. Two blocks on each of the slides were hybridized with two specific samples and used as control blocks for data normalization. The slides were then rinsed three times to remove unbound proteins. 1ug/mL goat anti-human IgG conjugated with Alexafluor647 (Invitrogen, Carlsbad, CA) solution was used for detection. After a second one-hour hybridization with anti-human IgG, the slides were washed and dried again, then scanned with a microarray scanner (Axon 4000A). The program Genepix Pro 6.0 was used to extract the numerical data. The signals from different slides were normalized with the averaged signal of the control blocks on each slide.

5.3 Results and Discussion

The proteins from Panc-1 human pancreatic ductal adenocarcinoma (PDAC) cell line were used as bait to study the humoral response in pancreatic cancer since the Panc-1 cell line has been used as a good representative sample of human pancreatic cancer[27]. The analytical work flow is illustrated in Figure 1. The solubilized protein solution extracted from Panc-1 cell line was fractionated using 2-D liquid separation methods as described consisting of chromatofocusing in the first dimension followed by nonporous reversed phase HPLC where intact proteins were collected as the final product. Fraction collection was performed where liquid eluent from each chromatographic peak was collected into 96 well plates. Each collected protein fraction was separated into two parts for further work. One portion was used for spotting the microarray plates and a second portion was used for protein identification based on LC-MS/MS. There were 1052 protein peaks obtained over a pH range of 8.0-4.0 spotted using the microarray device onto each nitrocellulose coated glass slide. Each slide was hybridized against a patient serum sample where the humoral response was run in this work against 38 cancer serum samples, 23 pancreatitis serum samples and 25 normal controls. Statistical analysis including non-parametric Wilcoxon Rank-Sum Tests and Outlier-Sum Statistics were then performed over this sample set to determine which proteins provided a significant response to patient sera. For the selection of identified proteins, a pre-validation study using a second, independent set of 168 serum samples was performed where five recombinant proteins were arrayed on nitrocellulose slides and probed with serum from a separate cohort of normal, pancreatitis, diabetes and pancreatic cancer patients.

5.3.1 Microarray Result of Humoral Response

The heterogeneity of humoral response has been displayed in a substantial percentage of patients with increased antibody expression to disease-related antigens, where only a subset of patients has an autoimmune response to a particular antigen. We herein assume that auto-immune markers show either an increased level of reactivity against most of the patient sera or an outlier pattern that exclusively appear in the cancer group. Two statistical methods, Wilcoxon Rank-Sum and Outlier Sum Test, were applied to the dataset to search fractions for auto-antibody response.

5.3.2 Statistical Analysis

Compared to traditional T-Test, Wilcoxon Rank-Sum Test was preferred in several previous humoral response studies because the dataset do not always fit a Gaussian distribution. The test generates a list of fractions with significantly greater intensities in the cancer group (p-value set at <0.02) in the pairwise comparisons in cancer versus normal and cancer versus pancreatitis. Twenty-nine fractions were selected in the cancer/normal pair and only seventeen passed the threshold in the cancer/pancreatitis pair. Figure 2 shows two heatmaps of these fractions after they are clustered using a hierarchical clustering algorithm. The clustering tree is added on top of the heatmaps. In the first heatmap/dendrogram, 65% (24 out of 37) of the cancer samples and only 17% (4 out of 23) normal samples are clustered on the left side with more blue bands which indicate increased reactivity with serum. Similarly the left side of the second heatmap/dendrogram includes 70% (26 out of 37) of the cancer samples and 29% (7 out of 31) of pancreatitis samples. Most of the samples are clustered with their own groups while a portion of the samples are not. Several reasons can be taken into account for this result. The incorrectly clustered cancer samples may not contain the antibodies to some

particular antigens in the Panc-1 cell line. Additionally, the non-cancer samples that incorrectly clustered with the majority of the cancer samples may be reactive to the co-eluted proteins in some fractions containing cancer-related antigens.

In a 2-dimensional separation, a protein often appears in two or more subsequent fractions rather than one because of the limited chromatographic resolution and also the post-column diffusion. In Figure 2, it is worth noting that the fractions in the heatmap are often accompanied by their adjacent fractions (red circled) ex. 7B1-7B3 and 4E4-4E11. The consecutive bands with a smooth reactivity profile are better candidates for further investigation and confer important information for protein identification.

The result of the Wilcoxon Rank-Sum Tests can be transformed to a ROC curve which estimates the ability of the selected biomarkers to distinguish case from non-case. For each ROC curve, an area under the curve (AUC) is reported, where 1.0 represents perfect separation of one group from the other, 0.5 represents a completely random result or no separation. In both cancer vs. normal and cancer vs. pancreatitis categories, the fractions top-ranked in Wilcoxon Rank-Sum Tests, exhibit AUC values of 0.70-0.72. The samples which positively responded to these two fractions in the cancer group do not completely overlap, which indicates the possibility of obtaining higher sensitivity and specificity by using them together. The ROC curves are combined using Canonical Discriminant Analysis, which maximizes the separation of the existing groups. Combined ROC curves of these top two fractions in both categories are drawn in Figure 3a and 3b, suggesting an increased AUC of 0.748 and 0.746 respectively. The top 10 and top 20 ranked fractions in the Wilcoxon Rank-Sum Tests are combined and the resulting ROC curves are shown in Figure 3c and Figure 3e. In the cancer vs. normal category, the ROC

curve of the top 10 fractions present the AUC value increased to 0.83 and combining the top 20 fractions further increases the AUC to 0.92. In the cancer vs. pancreatitis category, the combined ROC (figure 3d and 3f) curves of the top 10 and 20 fractions also give increased AUC values of 0.777 and 0.791 respectively.

Recently, statistical outlier sum methods, such as COPA [28] and OS [26] have been proposed as methods for searching cancer related genes with microarray techniques. The outlier sum analysis is able to detect a small number of significantly up-regulated signals from microarray data in the disease group while the signal from the majority may not necessarily change. Since the majority of cancers have heterogeneous activation for different individuals, it appears that the application of this method using the "subset" idea where some cancers respond to the humoral response and others do not respond may result in an improved performance for microarray data. Of the two outlier sum methods, COPA identifies pairs of biomarkers with mutually exclusive up-regulated samples because it was designed to search for gene activation with a mutually exclusive mechanism, while the protein biomarkers in this work may not have the same feature. OS identifies outliers in a similar manner as COPA, but it calculates an outlier score for each individual. Therefore, OS is the preferred method in this study.

After OS analysis was applied to the dataset, we found 9 fractions (listed in Table 1) that ranked in the top 5% in both of the comparisons of cancer/normal and cancer/pancreatitis. The reactivity profiles for the top 3 fractions are shown in bar graphs in Figure 4. It appears that only subsets of the cancer group show increased reactivity against these fractions, while the signals from the other samples remain the same. The signals of subsets with increased reactivity in the cancer group are much higher than the

range of the non-cancer groups, which remain close to the baseline. Such an outlier pattern for these fractions indicates that they contain proteins that are only immunogenic for a subgroup of the cancer samples and not immunogenic for all the non-cancer samples. In clinical application, these fractions can provide information for accurate diagnosis as the immunogenic cancer samples distinguish themselves with a high signal.

The results from Wilcoxon Rank-Sum test and Outlier Sum analysis are compared, where the lists of marker fractions do not overlap. We are more interested in the candidates given by OS, as those fractions exhibiting an outlier pattern exclusively in the cancer group would be more useful in diagnosis. The list of candidate fractions from the OS were thus identified with mass spectrometry and their IDs and performance were validated with the recombinant protein array.

5.3.3 Mass Spectrometry Identification

LC-MS/MS is used to identify the proteins in these fractions and their adjacent fractions. As expected, multiple proteins were identified in each of the fractions. The identified proteins were screened based on an assumption that their reactivity profile should be consistent with their appearance in the neighboring fractions. The resulting protein IDs are listed in Table 1 for each of the fractions.

5.3.4 Recombinant protein validation:

Due to the large number of fractions from the 2-dimension separation, a set of only 91 serum samples was used to search for the fractions that could be potential biomarkers, where a larger set is usually required for confident biomarker discovery. It is also necessary to confirm the protein IDs identified for the candidate fractions. To pre-validate these potential markers, recombinant proteins were used to validate the result

with a different sample set and also confirm the protein IDs for the candidate fractions. Five commercially available recombinant proteins were selected for the validation experiment with 48 samples from the cancer group and 40 samples from each of pancreatitis, diabetes and normal control groups. Type 2 diabetes samples are included since some pancreatic patients also develop this condition which might be responsible for the autoimmune reactivity.

In order to measure the auto-antibody response that is elicited against the recombinant proteins correctly, care must be taken to avoid saturating the signal. Hence, the serum must be diluted sufficiently so that the amount of available auto-antibody in the serum is lower than the binding capacity of the specific recombinant protein. Therefore, a saturation curve was made using different dilutions of serum to hybridize against identical blocks of the recombinant proteins. The result of the saturation test showed that with ten-fold dilution, the recombinant proteins were not saturated and yielded a signal/background ratio of >5. Higher or lower dilution resulted in partial saturation or decreased signal intensity. A tenfold dilution factor was therefore used in the current pre-validation experiment using recombinant proteins. The microarray data (background subtracted) was also adjusted by the average signal of control blocks on each slide and standardized for each recombinant protein.

In Figure 5, we show the plot of the distribution of the reactivity for each of the recombinant proteins against the sera. The recombinant protein that produces the best result is Phosphoglycerate kinase 1 (PGK1), where 10 outliers out of 48 total samples are observed in the cancer group, while no outlier is present in the other three non-cancer groups. PGK1 protein is a kinase in the glycolytic pathway and can be up-regulated by

HIF-1α in the cellular response to hypoxia to provide energy for tumor cell proliferation [29]. Genomics-based studies have found that it acts as a suppressor of proangiogenic factor such as, VEGF and triggers metastasis due to its effect on the increased expression level of β-catenin, chemokine CXCR4 and CXCL12 [30-32]. At the protein level, PGK-1 has been found overexpressed in pancreatic cancer tissue versus adjacent controls and also elevated significantly in the sera of pancreatic cancer patients (19% strongly up-regulated, 50% weakly-moderately up-regulated) [33]. The performance of PGK-1 in the prevalidation experiment using recombinant protein indicates that a lower percentage of patients elicit auto-response. In the future study, it would be interesting to see whether there is a correlation between serum level of PGK-1 and auto-antibody level and how the production of antibody affects the development of the cancer.

For both Malate dehydrogenase (MDH1) and ADP-ribosylation factor interacting protein 2 (ARFIP2), there are 4 such outliers in the cancer group. The absence of outliers in the non-cancer group indicates that these 3 recombinant proteins are exclusively antigenic in cancer sera and could be tumor-associated. In Table 2, the performance of these 3 biomarkers used together to distinguish cancer is estimated. A cutoff equal to the highest signal in a certain non-cancer group is applied to define the reactive samples in the cancer group. The 3 recombinant proteins together distinguish more than 40% of the cancer samples from the normal and diabetes group, while only 29.2% from the pancreatitis group.

For Annexin A2 (ANXA2), the cancer group only has one outlier that is above all the other groups. This is not consistent with the OS analysis which showed differential humoral response in the fraction where ANXA2 was identified in pancreatic cancer sera.

It could be due to the use of the recombinant proteins, which may lack the required PTMs to induce a humoral response or the protein may not be in a form to induce a humoral response [17]. Also, since multiple proteins are identified in the fraction, the protein that was responsible for the observed humoral response may not be ANXA2. Heterogeneous nuclear ribonucleoprotein A2 (HNRPA2) produced more unexpected result in the validation where it showed a universal increase in the reactivity against diabetes samples, while the signals of the other three groups remained at the same level.

5.4 Conclusion

We have presented a study of the cancer-related humoral response on pancreatic adenocarcinoma using 2-dimensional separation and protein microarray techniques. After analyzing the data with two statistical tools, the fractions showing outlier patterns in Outlier Sum Test was chosen for identification of the proteins and prevalidation using recombinant proteins. In the prevalidation experiment, 20.8% of the cancer samples demonstrated strongly elevated reactivity for PGK-1,while no proteins in the non-cancer groups were found to react. This result suggests that auto-antibody level of PGK-1 in the serum is useful as a diagnostic biomarker indicating the presence of cancer. Future study of the correlation between the protein level and auto-antibody level of PGK-1 in cancer patients may provide better understanding of the role of PGK-1 in cancer development.

Figure 5.1 Flowchart of the experiment.

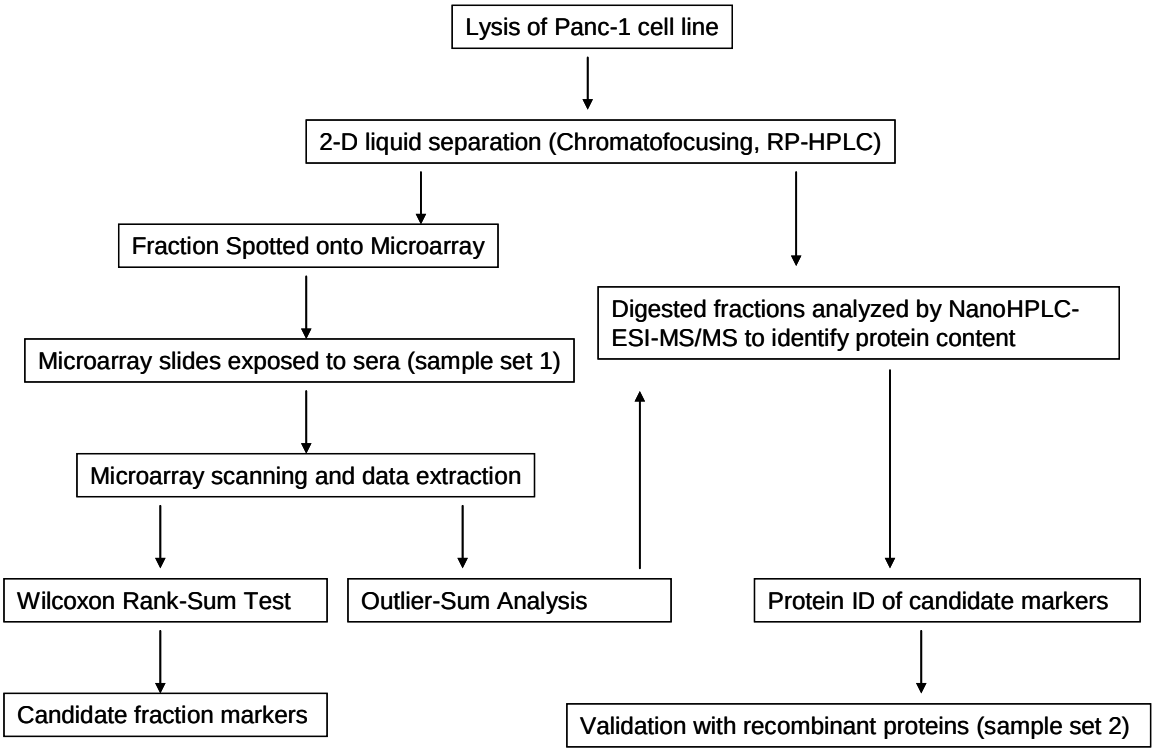
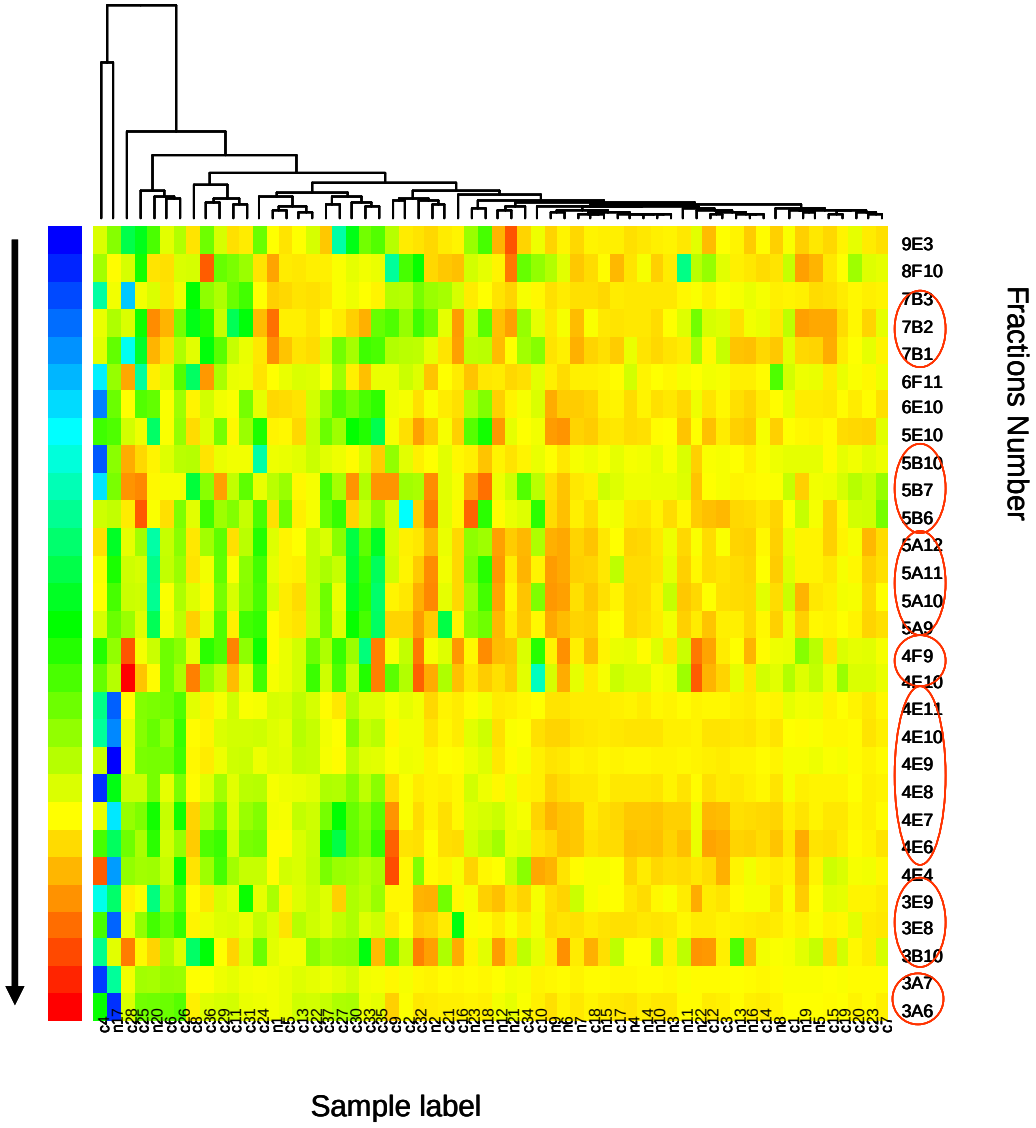


Figure 5.2 Heatmap with dendrogram of the microarray data. The colors of the bands indicate the normalized intensities of the microarray signal of the fractions. Only the fractions with a p-value less than 0.02 in the pairwise comparisons using Wilcoxon Rank-Sum Tests are shown in the heatmap. The fractions appear with their adjacent ones are red-circled. a. cancer vs. normal; b. cancer vs. pancreatitis.

b



b

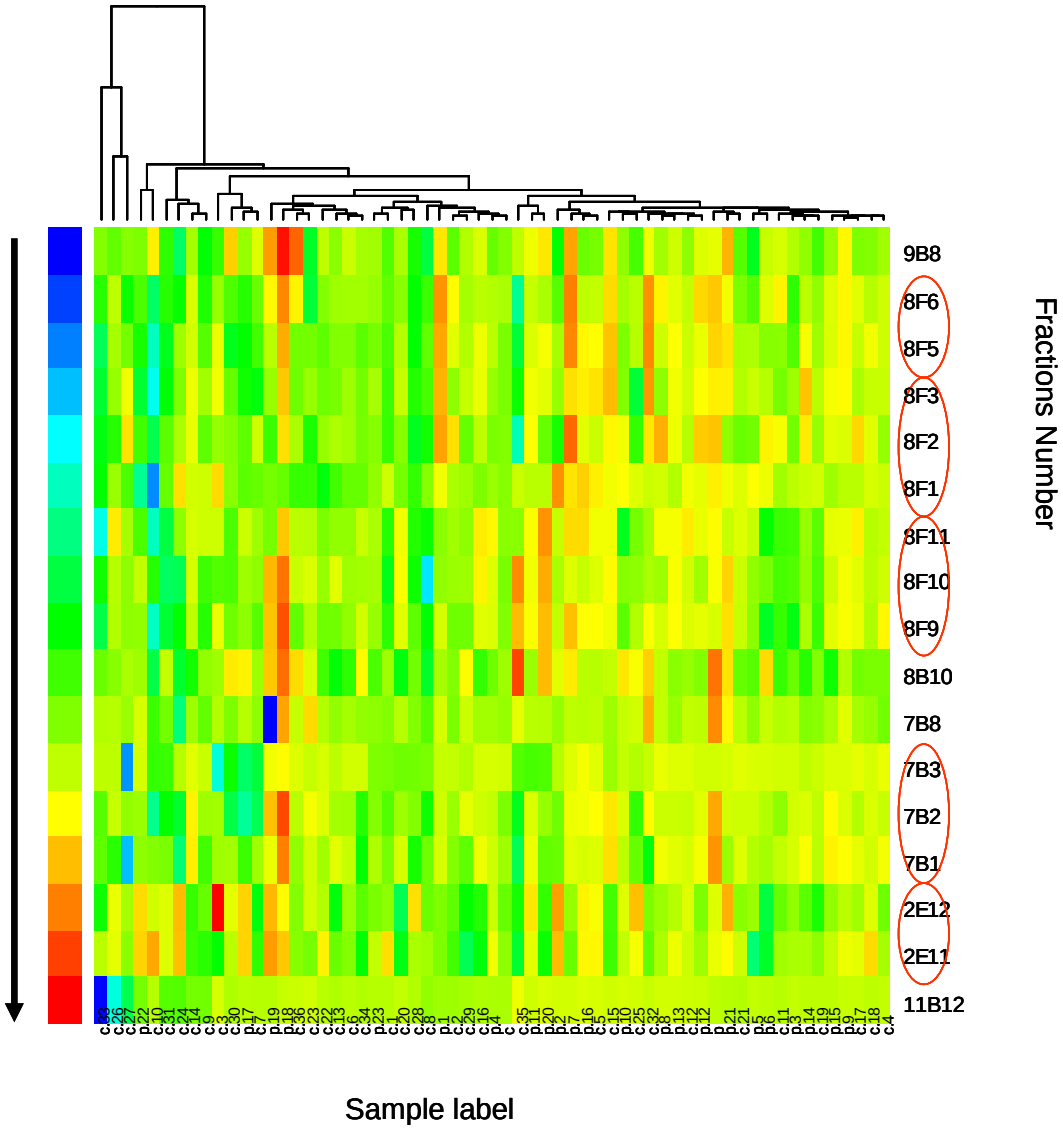

9B8
8F6
8F5
8F3
8F2
8F1
8F11
8F10
8F9
8B10
7B8
7B3
7B2
7B1
2E12
2E11
11B12
Fractions Number
Sample label

Figure 5.3 Combined ROC curves for the top-ranked fractions in the Wilcoxon Rank-Sum Tests. a) top 2 fractions combined, cancer vs. normal; b) top 2 fractions combined, cancer vs. pancreatitis; c) top 10 fractions combined, cancer vs. normal; d) top 10 fractions combined, cancer vs. pancreatitis; e) top 20 fractions combined, cancer vs. normal; f) top 20 fractions combined, cancer vs. pancreatitis.

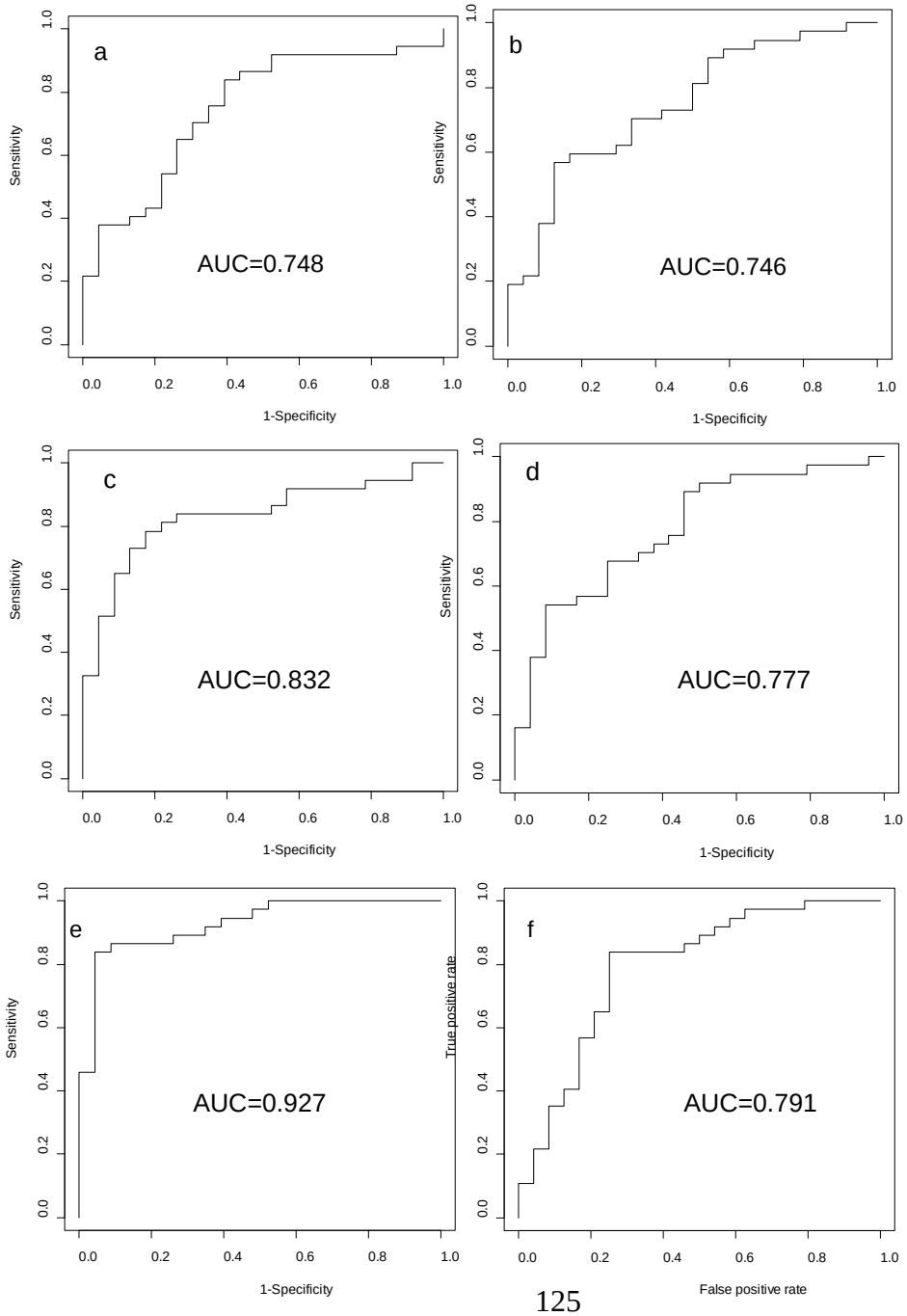


Figure 5.4 Colored bar graphs of three fractions found responded exclusively to some cancer sera in both pairwise comparisons between cancer vs. normal and cancer vs. pancreatitis. The y-axis is the normalized microarray signal for each sample.

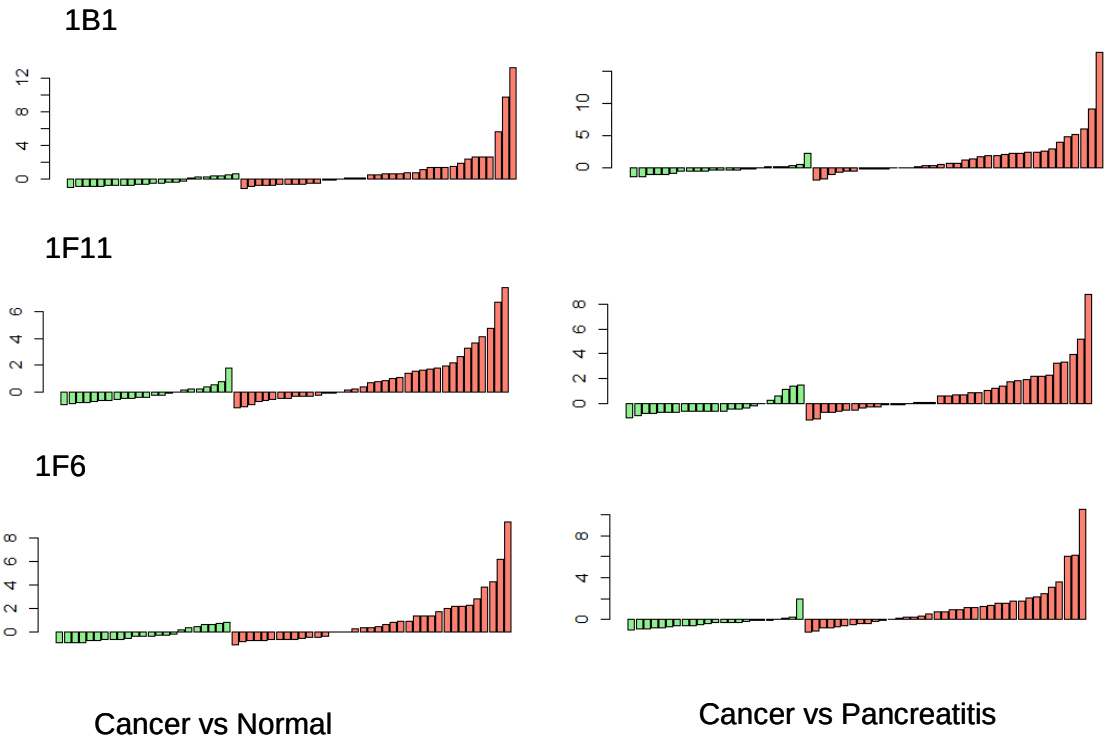


Figure 5.5 Distribution of the level of reactivity of five biomarker candidates examined in the pre-validation experiment. The plots present the normalized fluorescent intensities of the auto-antibody capture by the five recombinant proteins.

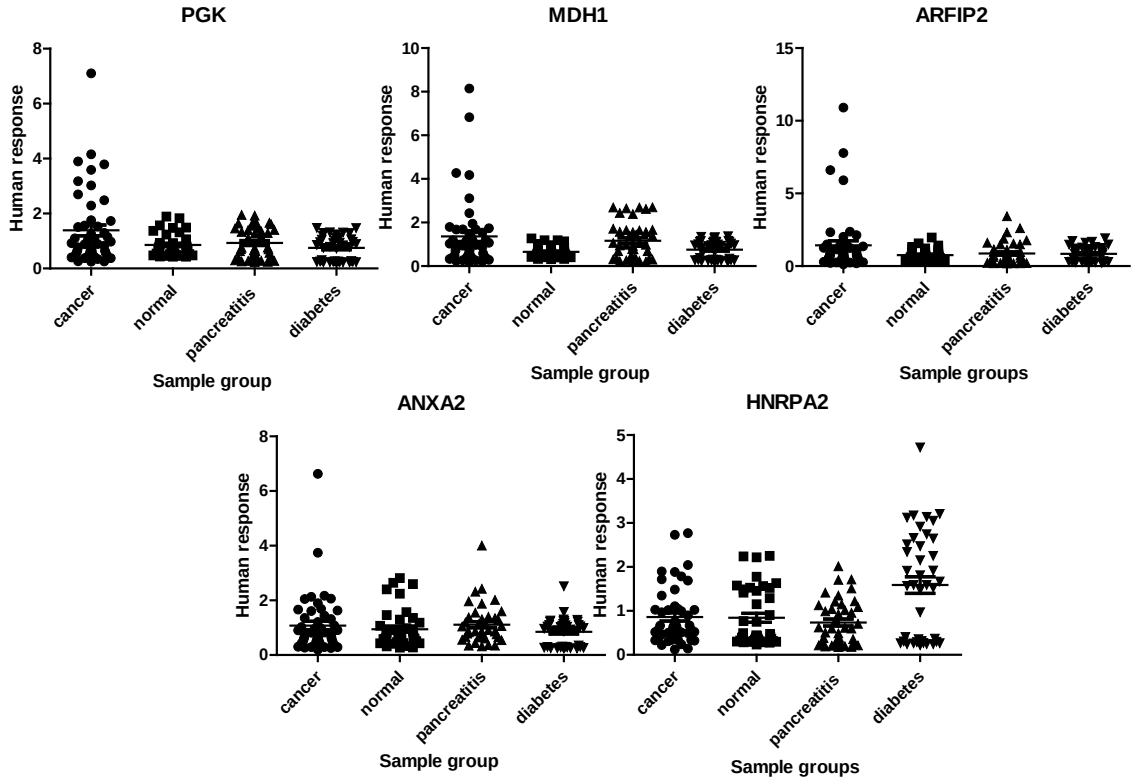


Table 5.1 Lists of differentiated fractions picked by OS analysis in both pairwise comparison between cancer vs. normal and cancer vs. pancreatitis. The information about the identified proteins in these fractions is also included.

Fraction	Acess Number	Protein Name	Fraction pH	MW	Seq Cov %	Theoretical pI	Unique peptides
1B1	P62847	40S ribosomal protein S24	7.9-7.6	15414	28.59	10.79	3
1F11	P00558	Phosphoglycerate kinase 1	7.9-7.6	44587	18.21	8.3	5
1F6	Q15369	Transcription elongation factor B polypeptide 1	7.9-7.6	12466	17.74	4.74	2
3E5	P04406	Glyceraldehyde-3-phosphate dehydrogenase	7.0-6.7	35900	18.97	8.58	4
4D6	Q9Y6N5	Sulfide:quinone oxidoreductase, mitochondrial precursor	6.7-6.4	49929	11.62	9.18	4
5G2	Q06830	Peroxiredoxin 1 (Thioredoxin peroxidase 2)	6.1-5.8	22097	29.35	8.27	6
8C3	O95881	Thioredoxin domain containing protein 12 precursor	4.9-4.6	19194	37.98	5.25	5
9A9	Q99729	Heterogeneous nuclear ribonucleoprotein A/B	4.6-4.3	36590	9.15	9.04	3
11D5	Q8NC51	Plasminogen activator inhibitor 1 RNA-binding protein	IPA wash	44539	18.43	8.66	5

Table 5.2 Numbers of samples with reactivity above the cutoff (cutoff = highest signal in the non-cancer group) against recombinant proteins in the cancer group compared to 3 non-cancer groups. The numbers in the parentheses are the percentage of the positive reactors in the cancer category i.e. sensitivity at 100% specificity.

Recombinant proteins	Pair of sample groups		
	Cancer vs Normal	Cancer vs. Pancreatitis	Cancer vs. Diabetes
PGK1	10 (20.8)	10 (20.8)	12 (25)
PGK1 or MPH1	18 (37.5)	12 (25)	19 (39.6)
PGK1 or MPH1 or ARFIP2	22 (45.8)	14 (29.2)	21 (43.8)

Reference

1. Jemal A, Siegel R, Ward E, Murray T, Xu JQ, Smigal C, Thun MJ, CA Cancer J Clin, 2006, *56*, 106
2. Rosty C, Goggins M Hematol Oncol Clin North Am 2002, 16, 37
3. Garcea G, Neal C P, Pattenden C J Steward W P, Berry D P Eur J Cancer, 2005, 41(15), 2213
4. Tan, EM J Clin Investig 2001, 108, 1411
5. Casiano CA, Mediavilla-Varela M, Tan EM, Mol & Cel Proteomics, 2006, 5, 1745
6. Desmetz C, Maudelonde T, Mange A, Solassol J, Expert Review of Proteomics, 2009, 6, 377
7. Hall JC, Casciola-Rosen L, Rosen A, Rheum Dis Clin North Am, 2004, 30, 455
8. Pollard KM, Lee DK, Casiano CA, Bluthner M, Johnston MM, Tan EM, J Immunol, 1997, 158, 3521
9. Utz PJ, Anderson P, Arthritis Rheum, 1998, 41, 1152
10. Rosen A, Casciola-Rosen C, Wigley F, Curr Opin Rheumatol, 1997, 9, 538
11. Ben-Mahrez K, Thierry D, Sorokine I, Danna-Muller A, Kohiyama M, Br J Cancer, 1988, 57, 529
12. Disis ML, Calenoff E, McLaughlin G, Murphy AE, Chen W, Groner B, Jeschke M, Lydon N, McGlynn E, Livingston RB, Moe R, Cheever MA, Cancer Res, 1994, 54, 16
13. Disis ML, Cheever MA, Curr Opin Immunol, 1996, 8, 637
14. Soussi T, Cancer Res, 2000, 60, 1777
15. Brichory FM, Misek DE, Yim AM, Krause MC, Giordano TJ, Beer DG, Hanash SM, Proceedings of The National Academy of Science of The United States of America, 2001, 98, 9834
16. Gnjatic S, Wheeler C, Ebner M, Ritter E, Murray A, Altorki NK, Ferrara CA, Hepburne-Scott H, Joyce S, Koopman J, McAndrew MB, Workman N, Ritter G, Fallon R, Old LJ, Clin Canc Res, 2009, 15, 4733
17. Desmetz C, Bascoul-Mollevi C, Rochaix P, Lamy PJ, Kramar A, Rouanet P, Maudelonde T, Mange A, Solassol J, Journal of Immunology Methods, 2009, 341
18. Looi KS, Nakayasu ES , de Diaz R, Tan EM, Almeida IC, Zhang JY, Journal of Proteome Res, 2008, 7, 4004
19. Xia Q, Kong XT, Zhang GA, Hou XJ, Qiang H, Zhong RQ, Biochem Biophys Res Commun, 2005, 330, 526
20. Hamanaka Y, Suehiro Y, Fukui M, Shikichi K, Imai K, Hinoda Y, Int J Cancer, 2003, 103, 97
21. Kotera Y, Fontenot JD, Pecher G, Metzgar RS, Finn OJ, Cancer Res, 1994, 54, 2856
22. Raedle J, Oremek G, Welker M, Roth WK, Caspary WF, ZeuzeFm S, Pancreas 1996, 13, 241
23. Yan F, Sreekumar A, Laxman B, Chinnaiyan AM, Lubman DM, Barder TJ, Proteomics, 2003, 3, 1228
24. Patwa TH, Li C, Poisson LM, Kim HY, Pal M, Ghosh D, Simeone DM, Lubman DM Eletrophoresis, 2009, 12, 2215

25. Canelle L, Bousquet J, Pionneau C, Deneux L, Imam-Sghiouar N, Caron M, Joubert-Caron R, Journal of Immunological Methods, 2005, 229, 77
26. Tibshirani R, Hastie T, Biostatistics, 2007, 8, 2
27. Lieber M, Mazzetta J, Nelson-Rees W, Kaplan M, Todaro G, Int J Cancer, 1975, 15, 741
28. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao XH, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM, Science, 2005, 310, 644
29. Dayan F, Roux D, Brahimi-Horn MC, Pouyssegur J, Mazure NM, Cancer Res, 2006, 66, 3688
30. Wang J, Wang J, Dai J, Jung Y, Wei CL, Wang Y, Havens AM, Hogg PJ, Keller ET, Pienta KJ, Nor JE, Wang CY, Taichman RS, Cancer Res, 2007, 67, 149
31. Kurayoshi M, Oue N, Yamamoto H, Kishida M, Inoue A, Asahara T, Yasui W, Kikuchi A, Cancer Res, 2006, 66, 10439
32. Zieker D, Konigsrainer I, Traub F, Nieselt K, Knapp B, Schillinger C, Stirnkorb C, Fend F, Northoff H, Kupka S, Brucher BL, Konigsrainer A, Cell Physiol Biochem, 2008, 21, 429
33. Shichijo S, Azuma K, Komatsu N, Ito M, Maeda Y, Ishihara Y, Itoh K, Clin Cancer Res 2004, 10, 5828

Chapter 6

Conclusion

Non-invasive protein biomarkers for early stage cancer detection and cancer therapy have attracted a significant attention from clinicians. They bring a great motivation and support for proteomic study, and as well as some critical challenges to the current techniques. The first problem for clinical proteomic study is the complexity of the clinical samples, which is often referred to as “dirty sample”. As typical clinical sample, resected malignant tissue often contains a mixture of different kinds of cells and intercell media which cause difficulty in protein extraction and compromise the specificity of the analysis. Multiple types of samples clean-up or isolation of specific cells can be done at the cost of introducing technical bias and extra time and labor. Sample clean-up results in reduced amount of sample which in some cases requires highly sensitive protein assay. Higher sensitivity in protein detection has been pursued since the very beginning and still under intense development. The CIEF-RPLC-MS 2-D shotgun technique described previously in this thesis is an effort of further lowering the limit of sample needed for a protein profiling experiment.

Besides digging deeper into the once forbidden area of low abundance proteins, people also look into the PTM that the proteins carry, such as glycosylation. The complexity of carbohydrate structures and their derivatives makes the study of the glycome a challenging subset of proteomic research. The microarray platform has become an essential tool to characterize glycan structure and to study glycosylation-related biological interactions, by using probes as a means to interrogate the spotted or captured glycosylated molecules on the arrays. Antibody-lectin sandwich microarray

measures glycosylation change on specific proteins captured from complex samples with lectin probes in a high-throughput array format. Compared to HPLC-MS, it is not only a simple and reproducible way of investigating large number of samples but also a type of assay that analyzes proteins in their native form which can be crucial for understanding the functional role of glycoproteins in the biological environment. Although the antibody-lectin sandwich microarray platform is able to distinguish glycoforms with different glycan structures, the use of fluorescence provides less detailed structural information than MS/MS. The introduction of mass spectrometry-based label-free detection would provide a means for further characterization of the glycan structure. It should be noted that, insufficient blocking efficiency with some of the lectins such as ConA limits the detection of several interesting glycan structures. To reduce the noise or background, we are searching for other chemical blocking strategies and even other platforms for the next generation of antibody-lectin assays.

Autoimmune response is another field with many remaining questions to answer. There is no doubt about the existence of tumor specific antigens. However, when the auto-antibodies are exploited for biomarkers to distinguish cancer, the serum level of auto-antibody against tumor specific antigens presents significant variation among individual patients. Auto-antibodies are typically up-regulated in a subset of the patient group, while the percentage of patient with positive response varies from 10-40%. Though the sensitivity of single auto-antibodies is not very satisfactory, combining several positive ones increases the performance of distinguishing cancer cases significantly. Unlike the protein glycosylation, the generation of auto-antibody is not regulated by certain enzymes, but rather randomly occurs in patients presenting abnormal

proteins from the tumor. Thus the overlapping positives between the different auto-antibodies are low. Down-regulation of some auto-antibodies has also been observed in some cancer patients with a specific stage of tumor. It is unclear that whether the production of auto-antibodies against tumor specific antigens is beneficial or harmful to the patients. Though auto-antibodies can already be used as excellent biomarker to help cancer diagnosis, the potential of auto-antibodies as biomarkers can only be fully opened up after more relevant biological experiment is done to unveil the mechanism underneath the initiation of auto-immune response which produces circulating auto-antibodies.