



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Predictive Assay For Cancer Targets

A. Suess, C. Nguyen, K. Sorensen, J. Montgomery, B.
Souza, K. Kulp, L. Dugan, A. Christian

September 23, 2005

OPTICS-East
Boston, MA, United States
October 23, 2005 through October 26, 2005

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

UCRL-PROC-215608

Predictive assay for cancer targets

Amanda Suess¹, Christine Nguyen¹, Karen Sorensen¹, Jennifer Montgomery¹, Brian Souza¹, Kris Kulp², Larry Dugan¹, Allen Christian^{1*}

¹Defense Biology Division, ²Biomedical Division, Biosciences Directorate, Lawrence Livermore National Laboratory, P.O. Box 808 Livermore, CA 94551

*Please address correspondence to: Allen T. Christian, Ph.D.

Defense Biology Division
Biosciences Directorate
Lawrence Livermore National Laboratory
7000 East Avenue PO Box 808, L-452
Livermore, CA. 94551
Phone: (925) 424-5909
Fax: (925) 424-3130
E-mail: christian4@llnl.gov

ABSTRACT

Early detection of cancer is a key element in successful treatment of the disease. Understanding the particular type of cancer involved, its origins and probable course, is also important. PhIP (2-amino-1-methyl-6 phenylimidazo [4,5-*b*]pyridine), a heterocyclic amine produced during the cooking of meat at elevated temperatures, has been shown to induce mammary cancer in female, Sprague-Dawley rats. Tumors induced by PhIP have been shown to contain discreet cytogenetic signature patterns of gains and losses using comparative genomic hybridization (CGH). To determine if a protein signature exists for these tumors, we are analyzing expression levels of the protein products of the above-mentioned tumors in combination with a new bulk protein subtractive assay. This assay produces a panel of antibodies against proteins that are either on or off in the tumor. Hybridization of the antibody panel onto a 2-D gel of tumor or control protein will allow for identification of a distinct protein signature in the tumor. Analysis of several gene databases has identified a number of rat homologs of human cancer genes located in these regions of gain and loss. These genes include the oncogenes *c-MYK*, *ERBB2/NEU*, *THRA* and tumor suppressor genes *EGR1* and *HDAC3*. The listed genes have been shown to be estrogen-responsive, suggesting a possible link between delivery of bio-activated PhIP to the cell nucleus via estrogen receptors and gene-specific PhIP-induced DNA damage, leading to cell transformation. All three tumors showed similar silver staining patterns compared to each other, while they all were different than the control tissue. Subsequent screening of these genes against those from tumors known to be caused by other agents may produce a protein signature unique to PhIP, which can be used as a diagnostic to augment optical and radiation-based detection schemes.

Keywords: Biomarkers, Proteomics, Serum, Peptide, Mass Spectrometry, MALDI-TOF, SELDI-TOF, Microarray, Laser Capture Microdissection, Transcript

1. INTRODUCTION

Proteins are the functional result of genetic expression. Therefore they can be used to study the condition and health of the body and its components. This is the emerging field of proteomics, which studies the proteins within a biosystem, their interactions, and patterns of expression in health and disease [1]. Proteins can be sampled from many biological sources including saliva, urine, hair, and blood. Serum (blood minus cells and clotting factors) is the most valuable source for protein studies. Blood supplies all of the body with necessary nutrients and also is a dumping ground

for cellular waste and secretions. Therefore, it contains a diverse collection of proteins that are representative of the condition of the body at a given time. Secretions resulting from cell death or damage, tissue leakage proteins and aberrant proteins released from tumor cells can act as biomarkers for the disease or problem present [2]. Identification and classification of specific biomarkers resulting from a disease such as cancer can improve the chances of early diagnosis and therefore increase the survival rate. Ordinarily, by the time symptoms arise or a malignant tumor is palpable, the cancer has already begun to metastasize. As is the case with ovarian cancer 80% of affected individuals are not diagnosed until the later stages of the disease, at which time the mortality rate is 35% after five years. However, if this cancer is detected early on in stage I, 90% of patients survive, requiring only surgery for treatment [3]. Finding cancer before it finds the blood stream is an emerging field that hasn't been substantially validated and therefore cannot yet be beneficial in a clinical setting. The promise of a faster, non-invasive technique that can be used to diagnose pathology and determine health is the future of proteomic studies.

Most biomarkers indicative of disease or tissue damage are smaller than the kidney cut-off point for filtration. These proteins are low in abundance compared to many common plasma proteins (i.e. albumin and immunoglobulins) [3, 4]. Therefore, the study of the plasma proteome usually will include an efficient method to remove high abundance proteins, resulting in an enrichment of the remaining compounds [4]. After the serum has been depleted the remaining contents can be studied with presently available technologies, most commonly mass spectrometry or microarray analysis.

2. BIOMARKER DETECTION AND CLASSIFICATION

2.1 Mass spectrometry

Classically proteins have been studied using electrophoresis techniques of which 2-dimensional electrophoresis (2-DE) has been the most common approach. The first dimension separates proteins by charge, also known as isoelectric focusing (IEF). Proteins migrate through an immobilized pH gradient (IPG) until they encounter their point of isoelectricity. The second dimension then separates the proteins by their molecular weight, following denaturation, using SDS PAGE [5]. This technique has been highly successful in the identification of genetic variants among the high abundance proteins within serum [4]. However, the study of biomarkers resulting from pathogenesis such as cancer requires the analysis of the serum proteins that are in low abundance. This therefore must be performed after the common serum proteins are removed by fractionation or chromatographic methods resulting in an enrichment of the remaining species [4]. 2-DE, although useful in separation ability, is unable to identify proteins beyond their size and charge. To further classify proteins already characterized by 2-DE, mass spectrometry (MS) has offered promising results.

Proteins resolved by two-dimensional gel electrophoresis can be digested from the gel with a proteolytic enzyme after which the sample can be analyzed by MS [6]. SELDI-TOF (surface-enhanced laser desorption and ionization with time-of-flight) and MALDI-TOF (matrix-assisted laser desorption and ionization with time-of-flight) have been used to study protein profiles from whole and fractionated protein samples, respectively [1]. Like 2-DE, MALDI and SELDI produce spectra dependent on the protein or fragment's size and electrical charge [3]. However, the determination is made as a result of the individual protein or fragment's time of flight, which results in the position held along the sample's spectra [3].

With MALDI-TOF MS the protein sample is immobilized on a matrix that all the proteins within a sample interact with, but only a select few are bound to as a result of the matrix's composition. SELDI-TOF MS utilizes surfaces that bind specific proteins based on their particular absorption, partition, and electrostatic properties. In both MS technologies the surfaces with bound proteins are pulsed with a laser causing ionization of the peptides, which then induces flight. The TOF measurement is dependent on the charge and mass of the species being studied [1]. In order to account for such variables as the number of different compounds and the number of mass peaks observed, institution of appropriate bioinformatic tools is necessary [7]. Available online, Mascot software (www.matrixscience.com) can be used along with relevant sequence information obtained from NCBI (www.ncbi.nlm.nih.gov) to identify proteins from MALDI MS experiments [6]. MALDI is ideally used for proteins of low molecular weight (<20 kD) [1]. SELDI protein samples are purified of contaminants by use of an affinity-based surface from which wash steps can be applied to further purify the sample. To analyze data received from SELDI MS ProPeak (3Z Informatics) offers a tool that applies Unified Maximum Separability Analysis (UMSA) based modules to produce linear results [7]. This tool is useful for identifying possible biomarkers for disease by locating variations between protein samples from affected and unaffected individuals. MALDI suffers from higher background signals as a result of the non-specific binding of materials to the matrix [7].

Both of these technologies lack the ability to provide direct sequence data from protein mixtures [2, 6]. MALDI and SELDI can however identify proteins accurately when the sample is digested and not extremely complex [6].

2.2 DNA microarrays

To facilitate the study of expression patterns within specific cells or tissues of interest microarray technology has offered a high-throughput method for delivering rapid results [8]. The success of microarray technologies can be attributed primarily to databases containing sequence information that can be used to develop targets for investigation [8, 9]. Microarrays can be used to study the genes expressed within a single cell, a collection of cells from a certain tissue, or even areas within a single cell where expression varies by location. Individual cells can be collected using laser capture microdissection (LCM). LCM utilizes a pulsed infrared laser that causes a thermoplastic film to adhere to selected cells of interest contained within a tissue sample. This technique is especially useful in the segregation and collection of malignant cells that are surrounded by normal cells and stroma. LCM is ideally suited for subsequent microarray analyses as it preserves subcellular compartments and compositions, enabling direct studies of the transcriptional and translational properties within the selected cell(s) [1].

Relying on Watson and Crick base pairing for hybridization of nucleic acids, DNA microarrays can be used for analyses of transcript expression levels within single cells or whole tissue samples. Additionally, DNA microarrays can be used to determine if a gene of interest is being expressed at all in a sample. The key to this technology is using available sequence information to fabricate the probes that representative sample transcripts can be hybridized to. The sequence of a gene of interest is used as a template to derive probes approximately 0.1-2.5 kb in size. These probes are then spotted in a set orientation along the matrix of the microarray platform. The probes are usually amplified by PCR and contain only a fraction of the genetic sequence for a complete gene and are then arrayed in a progressive fashion. In this method not only is the presence/absence of the gene of interest being studied, but also the concentration of transcripts that are homologous to particular regions within the gene can be observed [8, 10]. Two techniques for spotting the probe onto the matrix are currently employed; they are mechanical microspotting and ink jetting [8]. With mechanical microspotting a pin loaded with the probe DNA makes contact with the matrix upon dispensing while ink jetting utilizes a propulsion mechanism to release the probe without making contact [8, 11-13]. The matrix the probe adheres to is usually made of either glass or nylon, each of which has certain advantages [8]. After the matrix is prepared with probe DNA, the targets must be manufactured. First, mRNA from the sample of interest is collected and then converted to cDNA by reverse transcriptase (RT). RT is an enzyme used by viruses to convert mRNA into DNA, the opposite of what occurs in the cellular process of making protein ($\text{DNA} \rightarrow \text{mRNA} \rightarrow \text{protein}$). The target cDNA must be labeled to quantify the results after hybridization to the probe. ^{33}P dCTP is a low emission radioactive label that can be incorporated into the target for data analysis. The use of a low energy emitter is preferred so that signals of adjacent weak and strong hybridization patterns can be differentiated [8, 10]. Fluorescence can also be used to label the target and is preferential to radioactive labeling when studying competitive hybridization reactions. Targets from two separate samples can be labeled with Cy3-dUTP and Cy5-dUTP respectively, from which the excitation wavelength for emission is widely separated, allowing optical filtration that is highly discriminatory [8, 10]. Although fluorescent labeling does have its advantages, it requires a large amount of target to be detectable ($>10 \mu\text{g}$), whereas radioactive ^{33}P labeling only requires a minimal amount for detection ($0.1 \mu\text{g}$) [8].

Extracting the data from a DNA microarray experiment for analysis has been approached with many different techniques and remains to be standardized. The first level of analysis involves the removal of background, and normalization from which the data can be statistically interpreted. Although the use of statistical methods is most common in microarray analysis there are many variants on this approach that are currently utilized. Commercially available bioinformatic programs are presently available to sort through the results of a DNA microarray experiment, however they too rely on complicated statistical calculations. A substantial amount of variables are included in a typical microarray experiment leading to the daunting task of locating and eliminating false positives and negatives with reference to transcript expression levels. In order to increase the reliability and validity of these experiments, the incorporation of replicates within a single experiment has been shown to reduce the amount of misclassifications [14].

2.3 Protein microarrays

DNA microarrays are only capable of revealing information about the mRNA content within a cell, which is a long step off from the functional side of protein expression. As a result protein and antibody arrays have been useful in elucidating protein specific interactions that are a result of post-transcriptional modifications. These alterations can't be

determined or predicted by the sequence information contained in the mRNA transcript [15]. Protein arrays can be used to study protein-DNA, protein-protein, and protein-small molecule interactions. The protein of interest is immobilized on a matrix and then incubated with the labeled potential substrate to ascertain interactions. Protein microarrays can also be used to study enzymatic activity with the enzyme, either functioning as the immobilized species or being added to immobilized substrate [16]. Although these applications have offered advanced methods for studying the behaviors and interactions of proteins, the potential for disease diagnostics and target identifications is where most of the research utilizing protein and antibody microarrays is currently focused.

The key to a protein microarray is to maintain native structure while immobilizing the arrayed species on a particular surface. These microarrays are specifically studying the proteins functions or characteristics, which can only be observed when the compounds withhold their natural characteristics. Therefore, a significant amount of time and research efforts have been focused on finding the optimal surface for proteins and antibodies to adhere to without interfering with detection of the results. Nitrocellulose offers a support that proteins can interact with in a non-covalent, and non-polar process, therefore not altering the structure of the immobilized species. However, these surfaces allow for high background signals as a result of non-specific binding by the test sample. This can be avoided with use of a blocking reagent such as BSA. Glass surfaces that are specifically treated with immobilized nickel ions or agarose to increase the binding capabilities of the protein have been used [17, 18]. Treated glass surfaces form covalent bonds with the arrayed protein, which can alter the structure of the protein resulting in possible reduction in binding affinity to the target. In order to prevent this, research has been performed on glass coated with polyacrylamide based hydrogels and poly-L-lysine treated glasses both of which have shown high binding efficiencies and can aid in the maintenance of native structure and reactivity [19]. PVDF, polystyrene and gold are other surfaces used in the binding of proteins in microarray research, however they are not easily attained or cost effective [15].

Protein microarrays can be performed either with antibodies immobilized in an array that protein mixtures can then be incubated onto, or with a protein sample immobilized on the surface to which antibodies are then targeted to. Forward phase arrays are those with the antibodies immobilized on a surface to which only a specific substrate can bind. By choosing the antibodies for immobilization, the specific substrate that will bind is already understood and the experiment is performed with the purpose of observing whether or not the analyte is present in a sample [1]. The detection and interpretation of the results of a forward phase antibody array can be performed by two methods. The sandwich method which requires a second antibody with a fluorescent label to be applied to the bound analyte, or by fluorescently labeling the analyte before incubation on the microarray [15]. Forward phase arrays have low sensitivity and suffer from the possibility of not containing a high enough concentration of antibodies for all the antigens in the test sample [1]. Reverse phase arrays have been more popular for studying a single biological component. In this set-up a complete tissue lysate can be contained in a single spot within an array so that multiple samples can be probed for the same compound. An optimized antibody with a fluorescent label is chosen to use as a probe for the arrayed tissue lysate. Therefore, only a single antigen is being targeted within a complete biological sample [20]. Protein microarrays are useful in identifying changes in protein expression and can even be used to classify new targets for disease diagnosis and treatment. Handling and maintaining the delicate structure of proteins requires further research to be performed in the field of protein microarrays.

3. FUTURE PROSPECTS

Many of the most common and deadliest diseases are known to be at some level under genetic control. So it would only be obvious to look at the genetic content of affected individuals for a target to design treatments against. However, it is understood that gene content, expression and functionality all play an important and individual role in disease development and therefore must be studied with a variety of different techniques. Before drugs can be manufactured, a target must be identified and thus far most changes related to disease are amongst compounds also present in the healthy state. Setting a baseline that can be directly correlated to the healthy state is the most challenging and controversial aspect of proteomic research. It has been shown that concentration variations have been observed within the plasma proteome based on the time of day, age, gender, season, diet, and sleep patterns of the individual being screened [21-25]. Even the posture of the subject under investigation can change the concentration of the protein within the blood by 10% due to arterial pressure increases when standing versus sitting [26]. Therefore measuring an increase or decrease in the concentration of a protein that is linked to disease is not enough to predict illness. With current technologies such as mass spectrometry and microarrays, identification of biomarkers for disease is possible. However, setting a baseline for healthy concentrations versus diseased levels will have to be based on the individual being tested. This still lacks 100% reliability, which begs the question of can we find biomarkers that are only expressed in the disease

state? This relies on finding genes and/or proteins that are only found in diseased tissue such as a cancer cell and then designing drugs against them. Of the available technologies not one can do the job by itself, they all must be combined in a time and economically efficient way to achieve any sort of progress in the discovery of potential biomarkers for disease diagnosis [27].

This work was performed under the auspices of the U. S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48 with support from NIH grant CA55861.

REFERENCES

1. Posadas, E.M., et al., *Proteomic analysis for the early detection and rational treatment of cancer--realistic hope?* Ann Oncol, 2005. **16**(1): p. 16-22.
2. Thadikaran, L., et al., *Recent advances in blood-related proteomics*. Proteomics, 2005.
3. Petricoin, E.F., et al., *Use of proteomic patterns in serum to identify ovarian cancer*. Lancet, 2002. **359**(9306): p. 572-7.
4. Anderson, N.L. and N.G. Anderson, *The human plasma proteome: history, character, and diagnostic prospects*. Mol Cell Proteomics, 2002. **1**(11): p. 845-67.
5. Gorg, A., W. Weiss, and M.J. Dunn, *Current two-dimensional electrophoresis technology for proteomics*. Proteomics, 2004. **4**(12): p. 3665-85.
6. Zeng, R., et al., *Identification of proteins separated by one-dimensional sodium dodecyl sulfate/polyacrylamide gel electrophoresis with matrix-assisted laser desorption/ionization ion trap mass spectrometry; comparison with matrix-assisted laser desorption/ionization time-of-flight mass fingerprinting*. Rapid Commun Mass Spectrom, 2003. **17**(17): p. 1995-2004.
7. Li, J., et al., *Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer*. Clin Chem, 2002. **48**(8): p. 1296-304.
8. Murphy, D., *Gene expression studies using microarrays: principles, problems, and prospects*. Adv Physiol Educ, 2002. **26**(1-4): p. 256-70.
9. Venter, J.C., et al., *The sequence of the human genome*. Science, 2001. **291**(5507): p. 1304-51.
10. Duggan, D.J., et al., *Expression profiling using cDNA microarrays*. Nat Genet, 1999. **21**(1 Suppl): p. 10-4.
11. Lashkari, D.A., et al., *Yeast microarrays for genome wide parallel genetic and gene expression analysis*. Proc Natl Acad Sci U S A, 1997. **94**(24): p. 13057-62.
12. Schena, M., et al., *Quantitative monitoring of gene expression patterns with a complementary DNA microarray*. Science, 1995. **270**(5235): p. 467-70.
13. Shalon, D., S.J. Smith, and P.O. Brown, *A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization*. Genome Res, 1996. **6**(7): p. 639-45.
14. Lee, M.L., et al., *Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations*. Proc Natl Acad Sci U S A, 2000. **97**(18): p. 9834-9.
15. Hudelist, G., et al., *Use of high-throughput arrays for profiling differentially expressed proteins in normal and malignant tissues*. Anticancer Drugs, 2005. **16**(7): p. 683-689.

16. Merkel, J.S., et al., *Functional protein microarrays: just how functional are they?* Curr Opin Biotechnol, 2005. **16**(4): p. 447-52.
17. Afanassiev, V., V. Hanemann, and S. Wolfl, *Preparation of DNA and protein micro arrays on glass slides coated with an agarose film.* Nucleic Acids Res, 2000. **28**(12): p. E66.
18. Lal, S.P., R.I. Christopherson, and C.G. dos Remedios, *Antibody arrays: an embryonic but rapidly growing technology.* Drug Discov Today, 2002. **7**(18 Suppl): p. S143-9.
19. Miller, J.C., et al., *Antibody microarray profiling of human prostate cancer sera: antibody screening and identification of potential biomarkers.* Proteomics, 2003. **3**(1): p. 56-63.
20. Petricoin, E.F., et al., *Clinical proteomics: translating benchside promise into bedside reality.* Nat Rev Drug Discov, 2002. **1**(9): p. 683-95.
21. Brohult, J., *Effects of high protein and low protein diets on ornithine carbamoyl transferase activity in human serum (S-OCT).* Acta Med Scand, 1969. **185**(4): p. 357-62.
22. Buckley, C.E., 3rd and F.C. Dorsey, *The effect of aging on human serum immunoglobulin concentrations.* J Immunol, 1970. **105**(4): p. 964-72.
23. Letellier, G. and F. Desjarlais, *Study of seasonal variations for eighteen biochemical parameters over a four-year period.* Clin Biochem, 1982. **15**(4): p. 206-11.
24. Linkowski, P., et al., *Genetic and environmental influences on prolactin secretion during wake and during sleep.* Am J Physiol, 1998. **274**(5 Pt 1): p. E909-19.
25. Macy, E.M., et al., *Sample preparation for plasma measurement of plasminogen activator inhibitor-1 antigen in large population studies.* Arch Pathol Lab Med, 1993. **117**(1): p. 67-70.
26. Statland, B.E., H. Bokelund, and P. Winkel, *Factors contributing to intra-individual variation of serum constituents: 4. Effects of posture and tourniquet application on variation of serum constituents in healthy subjects.* Clin Chem, 1974. **20**(12): p. 1513-9.
27. Anderson, N.L., *The roles of multiple proteomics platforms in a pipeline for new diagnostics.* Mol Cell Proteomics, 2005.