Application of Genomic and Proteomic Technologies to Early Detection of Cancer

Ali M. Ardekani PhD*, Mohammad-Mehdi Akhondi PhD*, Mohammad-Reza Sadeghi PhD*

Background

The completion of the Human Genome Project has led to a surge in the use of genomic and proteomic technologies for identification of markers for early cancer detection and molecular-targeted treatments. The number of defined human genes and expressed sequence tags continues to grow, and new tools are being developed for interrogation of these databases. High-throughput screening techniques are now widely available allowing investigators to rapidly screen and confirm new genes, mRNA transcripts, and proteins. Differential expression of these molecules between normal and malignant tissues allows identification of genes and pathways that are dysregulated in a variety of human cancers.

The National Cancer Institute-Food and Drug Administration (NCI-FDA) Clinical Proteomics Program was formed in late 1990’s to develop and apply novel technology to improve the ability to understand the biology of cancer. Putting this knowledge into practice is hoped to advance early cancer detection and identify molecular events that may be targets for prevention and treatment. Since its inception, new technologies have been developed at National Institutes of Health (NIH), and technologies have been applied for tumor marker discovery, profiling, and clinical application.

Laser capture microdissection (LCM) was the first of these new technologies. It has allowed accurate separation of tumor, stromal, and normal cells within a single biopsy specimen. High-throughput analysis of microdissected...
specimens allows for clean discrimination of events occurring in and between each of these tissue microcompartments. Application of these technologies to patient samples has allowed dissection of genomic changes, expression events, and differential expression, activation, and signaling of a variety of proteins in tumor samples.\textsuperscript{6,7} Additionally, profiling of low-molecular weight proteins in patients’ serum samples is now possible using high-tech instruments. Also such data can be queried with new, powerful bioinformatics tools to cluster unaffected and cancer patients into their respective groups. Validation of the specificity, sensitivity, and positive predictive value of these novel technologies for early cancer detection in clinical specimens of ovarian cancer has been reported.\textsuperscript{8} Screening and validation of several other malignancies using these technologies is already underway.

Continued development and application of these technologies to clinical materials holds the promise of improved early cancer detection, prevention, and tumor-specific treatments. Because the understanding of specific changes in gene expression and protein signaling pathways is increased, the ability to move away from histopathologic-based diagnosis of disease as groups and toward the treatment of the specific dysregulated components in the host-tumor interaction would be increased. Of course, this is not to say that such methods are expected to completely replace the current conventional methods of treatment of diseases. However, it is expected that the overall advances in genomic and proteomic technologies will help to guide our judgment with regard to the best treatment for each individual patient.\textsuperscript{9}

\textbf{Advances in molecular medicine}

In the past several decades great improvements have been shown in various microscopic techniques, such as light microscopy, immunohistochemistry, and now antibody-based assays for the diagnosis and evaluation of cancers. Such technologic advances have helped to improve both early diagnosis and treatment of an overwhelming number of cancers. For example, the introduction of the Papanicolaou smear in the late 1940’s has been a great tool in decreasing the incidence of invasive cancer of the cervix by over 70\%.\textsuperscript{10} However, in the past two decades the rate of intraepithelial neoplasia has increased while the incidence of invasive cervical cancer has plateaued and possibly begun to rise.\textsuperscript{11} This has led to the development of additional diagnostic testing for identification of high-risk human papillomavirus (HPV) subtypes using molecular-based techniques, such as polymerase chain reaction (PCR) and solution hybridization.\textsuperscript{12,13} Similarly, the early diagnosis of breast cancer has been improved by mammographic screening. While molecular imaging modalities are being developed for improved detection of breast abnormalities, we are also seeing changes in the way women are being treated once a diagnosis of breast cancer is made. New ‘signal pathway-targeted’ pharmacotherapeutics, such as trastuzumab for breast cancer, as well as imatinib for gastrointestinal stromal tumors and chronic myelogenous leukemia,\textsuperscript{14,15} have come from an enhanced understanding of the genetic and protein alterations within a tumor. Replacing or complementing cytotoxic chemotherapy with treatments targeting specific dysregulated pathways in a given tumor are possible because highly efficient ‘microtechniques’, such as microarrays and microcapillary liquid chromatography are now available.\textsuperscript{16} Continuing to develop new, high-throughput microtechniques should allow us to uncover many additional targets within a tumor’s genome and proteome.

\textbf{Analysis of the genome and proteome}

Understanding specific alterations in tumor DNA, RNA, and protein requires knowledge of what is occurring both in and around the tumor. The ability to localize such alterations has been confounded in the past by our inability to adequately isolate specific cell types from pathologic specimens. Cell scraping and affinity column purification,\textsuperscript{17} in vitro culture of immortalized cell lines,\textsuperscript{18} cell cocultures,\textsuperscript{19,20} and manual microdissection of tissues\textsuperscript{21,22} have all been used, each with their own benefits and disadvantages. While these techniques have allowed great strides in obtaining pure cell lines for evaluation of intracellular processes, an integral part of what confers a unique phenotype upon an individual tumor has been lost. For example, the tumor microenvironment of a carcinoma consists not only of the malignant epithelial component, but also the surrounding stroma and normal tissue. These distinct microcompartments use receptors, cell junctions, and inter- and intracellular signaling molecules to allow tumor cells to communicate with their surroundings and play
an active role in their own control or progression.\textsuperscript{23} Removing a subpopulation of these cells for growth in an \textit{in vitro} system interrupts potentially important cell-cell and cell-matrix interactions that may affect tumor behavior, thus giving scientists a false impression of \textit{in vivo} tumor composition and physiology.

The advent of LCM has enhanced the ability to remove specific subpopulations of cells from frozen or ethanol-fixed tissues under direct microscopic visualization.\textsuperscript{3,4} These tissues can be microdissected either stained or unstained. In fact, coupling rapid immunohistochemical staining techniques with LCM may allow for more accurate microdissection of cell subsets.\textsuperscript{24} LCM employs a pulsed infrared laser, 7.5 – 60 µm in diameter, to activate a thermoplastic film placed over the cells of interest causing the film to become fused to the cells. Laser “shots” are repeated until all cells of interest are collected onto a plastic cap to which the film was previously applied. The cap is then lifted away from the tissue and the cells that were fused to the cap are placed into a cell lysis buffer to solubilize DNA, RNA, or proteins for further analysis. It has been estimated that this technique allows 90 to almost 100% of the cells collected to be from the subpopulation of interest.\textsuperscript{25,26} Selection of cells by this method allows clean separation of malignant, \textit{in situ}, and various normal cell subpopulations within a single biopsy specimen. Most importantly collection of cells with LCM preserves the molecular composition and architecture of the cells so that direct comparisons of transcriptional and translational messages can be made between tissue microcompartments of the same sample. Therefore, LCM enables scientists to take a more realistic “snapshot” of a tumor’s \textit{in vivo} biologic and physiologic properties.

Ornstein et al. sought to objectively evaluate the impact of a tumors microenvironment on its proteome. They used LCM to isolate pure prostate tumor epithelium and stroma from prostatectomy specimens.\textsuperscript{26} They then compared protein profiles obtained from two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) of microdissected cell lysates, undissected whole cryostat section lysates, and immortalized cell lines from the same patients. Additional comparisons were made with protein profiles from two commercially available prostate cancer cell lines. Their findings clarify several important issues. Firstly, stromal and epithelial components of tumors may share less than half of the same proteins. Secondly, proteins differentially expressed between normal and malignant prostate tissue are isolated solely to the epithelial compartment. Thirdly, microdissection does not alter a protein’s integrity. Furthermore, they provide evidence that protein profiles of \textit{in vivo} and \textit{in vitro} cells vary significantly with as little as 25% overlap in their protein milieu. The role of LCM is then key in genomic and proteomic profiling facilitating direct comparison of tissue DNA, RNA, and protein content and function in both normal and neoplastic tissues.

**Genomic techniques**

Evaluation of the human genome has become quite efficient using techniques, such as loss of heterozygosity (LOH) screening\textsuperscript{27,28} and comparative genomic hybridization (CGH).\textsuperscript{29,30} These techniques can be used to identify and confirm both known and unknown alterations (deletions, mutations) in the genomes of a variety of tumors. For example, Sobol et al. demonstrated number and location of many genome alterations, by using 191 polymorphic markers in a genome-wide search for LOH in 31 Burkitt lymphoma cell lines and their normal counterparts.\textsuperscript{28} They were successful in distinguishing two types of altered allelic patterns: a bona fide LOH profile, indicative of deletion and a profile indicative of increased dosage (ID).\textsuperscript{27} Validation of these genomic changes at the transcriptional and translational levels is an important step in identifying biologically relevant candidates for further investigation as molecular targets in cancer diagnosis and treatment. High-throughput techniques to identify concomitant alterations in RNA and proteins have already being used to confirm the information that is extracted from interrogation of the genome.

Microarray technology has become a well-established method for comparing gene and expressed sequence tag (EST) expression in laboratory and clinical tumor specimens.\textsuperscript{5} Thousands of genes can be evaluated simultaneously by amplifying RNA with fluorescent labels and applying the labeled transcripts to array slides containing large numbers of oligonucleotides or cDNAs.\textsuperscript{31-33} Expression of the fluorescent label is indicative of the presence and magnitude of that particular cDNA transcript within the test population. Differences in expression are logged and patterns of gene expression can be identified by coupling microarrays with comparative
bioinformatic software.\textsuperscript{34,35} Information obtained from such inquiries into a tumor’s genome and transcriptome may lead to improved gene discovery and cancer detection methods.

**Proteomic techniques**

Evaluation of protein structure, function, and regulation has evolved rapidly over the past several decades. Large amounts of information about a particular protein’s activation status and interactions can now be obtained in a matter of minutes with new high-throughput approaches. Microarrays are now being used to profile the proteome of a given cell population using antigen-antibody interactions.\textsuperscript{5,36} Prior to the introduction of these high-throughput techniques, specific protein content in tissues was being evaluated using low-throughput western blotting, \textit{in situ} hybridization, and immunohistochemical staining.

Immunohistochemistry has long been used as an adjunct diagnostic method for a variety of cancers. Information gleaned from these studies has been used to aid clinicians in developing more accurate prognostic information. Overexpression of HER-2/neu in breast and ovarian cancers,\textsuperscript{37} p53 in endometrial adenocarcinomas,\textsuperscript{38} epithelial growth factor receptor (EGFR) in squamous cell carcinoma of the cervix,\textsuperscript{39} and underexpression of H-ras p21 in nonsmall-cell lung cancers,\textsuperscript{40} have all been associated with poorer outcomes. Identification of some of these protein alterations has helped lead to the development of our first group of molecularly targeted antitumor therapies. Testing for estrogen and progesterone receptor status and for HER-2/neu expression in biopsy specimens has enabled physicians to use tamoxifen and trastuzumab as adjunctive treatments in patients with breast cancer.\textsuperscript{14,15,41,42}

Unfortunately, immunohistochemistry is labor intensive and specimens must be evaluated by a pathologist one at a time. In addition, quantification of the staining is subjective and results cannot be easily generalized.

Protein microarrays also depend on specific antibody—ligand interactions, but avoid some of the complications of immunohistochemistry by allowing rapid, quantitative comparison of the proteins of interest using fluorescence-based imaging programs.\textsuperscript{43,44} Recent breakthroughs in protein microarrays have produced commercially available protein pathway arrays that are able to test cell lysates for both their phosphorylated and nonphosphorylated forms using specialized protein chips. Reverse phase protein arrays have been used to study disease progression and pathway activation in prostate and esophageal cancer specimens.\textsuperscript{5,45,46} Cell lysates from microdissected cell populations were applied to nitrocellulose-coated slides in serial dilutions and then incubated with antibodies of interest. This approach yields quantitative, informative comparisons of normal, dysplastic, and invasive carcinoma cells. It was shown that annexin-1 levels decrease significantly with increasing disease severity by comparing normal, \textit{in situ}, and malignant cell subpopulations microdissected from the same prostate and esophageal specimens. Furthermore, it was shown that worsening disease progression is associated with activation of the pro-survival pathways with a corresponding reduction in apoptosis.\textsuperscript{5} These studies show that profiling the complete protein repertoire of a cell population can give information about the specific pathways that are altered in carcinogenesis. Evaluation of signal pathway activation or quiescence, in the transition from the normal to malignant cell phenotype, can be done by comparing these two subpopulations of cells from the same patient specimen. Although these arrays allow comparison of known proteins in various cell populations, their dependence on sensitive and specific antibodies does not allow us to detect the presence or function of, as yet, uncharacterized and unknown proteins.

New methods for rapid identification of both known and unknown proteins are under development. Until now, patterns of protein expression using 2D-PAGE separation has provided the best “snapshot” of the protein repertoire of the cell or body fluid. However, this technology is limited in being low-throughput, labor intensive, time consuming, and also is problematic in detecting proteins that are basic in charge or smaller than 10,000 Da.\textsuperscript{47} This lower molecular weight range may contain cleaved proteins or peptides that are aberrantly shed or secreted from cells in response to a disease. In serum, the pattern of the smaller molecular weight proteins and peptides is an information archive previously unexplored. New technology has now been developed to characterize proteins in this lower molecular weight range.\textsuperscript{8,38,49} Matrix-assisted desorption and ionization time-of-flight (MALDI–TOF) and surface-enhanced laser desorption and ionization time-of-flight (SELDI–TOF) are two of the methods currently being employed for this purpose. Since MALDI–TOF
and SELDI–TOF analyses generate streams of data comprised of tens of thousands of data points, complex computational systems have been devised to discover subtle changes in protein expression patterns which could be diagnostic or predictive of cancer. Each of these modalities is finding its niche in gene discovery, signal pathway transduction mapping, and target selection for molecular-based treatments in a variety of cancers.

**Cancer-specific blood patterns using SELDI–TOF**

SELDI–TOF coupled with advanced bioinformatics tools is proving itself to be a useful molecular tool for improving cancer screening. Current methods for screening women at risk for ovarian cancer relies on cancer antigen 125 (CA 125) levels and transvaginal ultrasonography. CA 125 has been used both for screening women at risk for ovarian cancer and for follow-up of women treated for ovarian cancer. Unfortunately, CA 125 detects only 35 – 50% of early-stage ovarian cancers, when patients have the best chance for survival. More than two-thirds of women are diagnosed when advanced stage of disease becomes clinically apparent. The five-year survival for patients with extrapelic ovarian cancer is still only 5 – 20% even with improved surgical techniques and chemotherapy regimens.

A recent study of the second-generation CA 125 assay (CA 125 II) showed that 66% of stage I ovarian cancers can be identified by this test alone. However, false-positive tests and lack of specificity for ovarian cancer, particularly in premenopausal women, makes it difficult to use CA 125 as a single serum marker. Additional serum markers for ovarian cancer, such as lysophosphatidic acid (LPA), have been discovered and validated with clinical patient samples, but improvement in early detection has come with higher false-positive rates. Transvaginal ultrasonography has been used both as a single screening method and an adjunct to serum tumor marker screening. van DePriest et al. showed a sensitivity of 81% and positive predictive value (PPV) of 9.4% using annual transvaginal ultrasonography alone for screening women at risk for epithelial ovarian cancer. Combining serial serum screening with ultrasonography does increase sensitivity, specificity, and PPV, but still remains inadequate.

The NCI-FDA Clinical Proteomics Program has analyzed sera from patients with ovarian cancer and sera of unaffected individuals using SELDI–TOF technology. A pattern matching heuristic algorithm was trained with knowns, cancer, and unaffected individuals. It randomly iterated small (five to 20 data point) groups of mass/charge (m/z) values until it identified an optimal pattern that could completely discriminate cancer from non-cancer. Blinded patients samples were then used to test the defined pattern, sensitivity, specificity, and PPV were determined. Using an optimal discriminatory group of five m/z values, all 50 ovarian cancer samples were correctly clustered including 18 that were stage I in this blinded cohort. Overall, this study yielded a sensitivity of 100%, and specificity of 95%. However, for any mass-spectrometry method such as SELDI to become a standard laboratory diagnostic method for early detection of cancer, many hurdles must be passed including repeatability of such studies in many laboratories and follow-up studies in large populations.

**What are the implications of the application of proteomic approach to cancer screening?**

The underlying pathologic state of an organ such as the ovary may be reflected in serum proteomic patterns. These pilot results demonstrate the capacity to produce a protein spectrum fingerprint from a tiny aliquot of unmodified serum and, with the assistance of a high-order bioinformatics approach, to extract a discriminatory fingerprint. Further, the specific bioinformatics algorithm used herein has the capacity to learn from each input datum and therefore will become stronger with continued application.

Following proper validation, serum proteomic pattern analysis might be ultimately applied in medical screening clinics, as a supplement to the diagnostic workup and evaluation. A negative value, if the sensitivity remains at 100% on further trials, could be used for reassurance, whereas a positive value may be sufficient to warrant further evaluation. An important future goal is confirmation of sensitivity and specificity for the prospective detection of stage I ovarian cancer in trials of high- and low-risk women, respectively. It will be important to design the trial to evaluate the efficacy of the approach as a stand-alone approach or one to be combined with current screening options. It is important to note that before any new cancer screening method is implemented in any cancer screening trials, the
cancers must be sensitively and specifically detected. Also in cancer control trials, it must become apparent that early detection actually improves outcomes for the population as a whole before the method is implemented. Such trials are underway at the NCI.

Future directions

Large-scale genomics and proteomics studies based on DNA and protein arrays are improving our understanding of the pathophysiology of cancers. The use of these research findings will soon lead to improvements in diagnosis, prognosis, and treatment of cancer. Genomics and Proteomics technologies not only have the potential to have a major impact on diagnostics and prognostics but also on the preclinical and clinical drug developments. The co-evolution of genomics and proteomics as complementary approaches to cancer will allow us to move closer to the goals of earlier detection, improved prevention, and institution of a molecular-based, tumor-specific approach to the treatment of each individual patient.

References


Application of genomic and proteomic technologies
to early detection of cancer


patterns in sera serve as biomarkers of ovarian cancer. *Gynecol Oncol*. 2003; 88: 25–42.