Int. J. Cancer: **123,** 2007–2012 (2008)

Tissue culture-based breast cancer biomarker discovery platform

Vathany Kulasingam[1,2] and Eleftherios P. Diamandis[1,2,3]*

[1]*Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada*
[2]*Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, Canada*
[3]*Department of Clinical Biochemistry, University Health Network and Toronto Medical Laboratories, Toronto, Ontario, Canada*

Current cancer biomarkers suffer from low diagnostic sensitivity and specificity and have not yet made a major impact on reducing cancer burden. Proteomic methods based on mass spectrometry have matured significantly over the past few years and hold promise to deliver candidate markers for diagnosis, prognosis or monitoring therapeutic response. Because of the complex nature of biological fluids such as plasma, biomarker discovery efforts using proteomics have not as yet delivered any novel tumor markers. Recently, there has been a rise in the number of publications utilizing a cell culture-based model of cancer to identify novel candidate tumor markers. The secretome of cancer cell lines constitutes an important class of proteins that can act locally and systemically in the body. Secreted proteins, in addition to serving as serological markers, play a central role in physiology and pathophysiology. In this review, we focus on the proteomics of breast cancer and the different strategies to mine for biomarkers, with particular emphasis on a cell culture-based model developed in our laboratory.

Key words: breast cancer; mass spectrometry; tumor marker; cell lines; secretome

Renewed interest in biomarker discovery

The concept of early detection of various forms of cancer before they spread, and become incurable, has attracted the attention of physicians and scientists for decades.[1] Forty percent of breast cancer patients have regional or distant spread of their disease at the time of diagnosis.[2] Moreover, survival rates of patients diagnosed with advanced breast cancer have changed little over the past 20 years. It is known that survival is excellent for breast cancer when early-stage disease is treated with existing therapies.[1] Without doubt, shifting all cases to early detection will have a profound impact on overall mortality and economic burden. Unfortunately, other than definitive diagnosis by biopsy and histopathology, no diagnostic or screening test is presently suitable for the early detection of clinically relevant breast cancer. This is because sufficiently high sensitivity (the probability of the test being positive in individuals with the disease) and specificity (the probability of the test being negative in individuals without the disease) are usually both not attributes of the same test; an increase in sensitivity tends to result in a reduction in specificity, and vice versa. Newer diagnostic methods with improved sensitivity and specificity are clearly needed to identify women with early stage breast cancer.

The ability to detect human malignancy *via* a simple blood test has long been a major objective in medical screening. The advantages of such an easy to use, relatively noninvasive and operator-independent test are self-evident. In this respect, cancer biomarkers can be DNA, mRNA, proteins, metabolites, or processes such as apoptosis, angiogenesis or proliferation.[3] An ideal tumor marker should be measured easily, reliably and cost-effectively using an assay with high analytical sensitivity and specificity.[1] In particular, an ideal tumor marker should be produced by the tumor cells and enter the circulation and it should be present at low levels in serum of healthy or benign disease patients and increase significantly in cancer (preferably in one cancer type). Moreover, an ideal tumor marker should be present in detectable (or higher than normal) quantities at early or preclinical stages and the quantitative levels of the tumor marker should reflect tumor burden. Finally, it should demonstrate high diagnostic sensitivity and specificity. A caveat to currently used tumor markers is that generally, they suffer from low diagnostic specificity and sensitivity. Only a few markers have entered routine use, and only for a limited number of cancer types and clinical settings. In the majority of cases, the current markers are used in conjunction with imaging, biopsy and associated clinicopathological information before a clinical decision is made.

The past decade has witnessed an impressive growth in the field of large-scale and high-throughput biology, which is attributed to an era of new technology development. Most of the proteomic technology platforms for biomarker discovery focus on the implementation of mass spectrometric techniques, in conjunction with several other methods such as gel electrophoresis, isoelectric focusing and chromatography. Mass spectrometry (MS) is an analytical technique that measures the masses of individual ionized molecules and atoms. High detection sensitivity and molecular specificity are hallmarks of MS. The advantages of MS are its ability to provide molecular mass with high specificity, provide high detection sensitivity (detects a single molecule), determine structures of unknown compounds, its application to diverse samples (volatile, nonvolatile, polar, nonpolar, solid, liquid, gaseous materials) and its ability to analyze complex mixtures.[4] All information gathered from a mass spectrometer comes from the analysis of gas-phase ions. Therefore, the first step is to convert the analyte molecules into gas-phase ionic species (performed by an ionization source) because once gaseous, their motion can be manipulated (this cannot be done with neutral species).[5] Then, a mass analyzer separates the molecular ions and their charged fragments according to mass-to-charge (m/z) ratio. The ion currents due to these mass-separated ions are detected by a detector and displayed as a mass spectrum. All of these steps are carried out under high vacuum to enable ions not to collide with other species. The generated spectra are then analyzed by various algorithms.[6]

Furthermore, these MS-based technologies have matured over the past few years and hence are capable of identifying thousands of proteins simultaneously. These strategies should facilitate delivery of potential candidate molecules for cancer diagnosis, prognosis and prediction of therapy. These projected discoveries may be instrumental in substantially reducing the burden of cancer by providing prevention, individualized therapies and improved monitoring post-treatment.

Different sources to mine for breast cancer biomarkers

In 1974, sera from normal volunteers, patients with a variety of non-neoplastic diseases and patients with malignant or benign tumors were examined by two-dimensional poly-acrylamide gel electrophoresis.[7] The authors did not identify specific proteins, but rather, observed a differential expression pattern discriminating the groups studied. However, despite the early searches for cancer

***Correspondence to:** Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, 600 University Ave., Toronto, Ontario M5G 1X5, Canada. Fax: 416-586-8628. E-mail: ediamandis@mtsinai.on.ca
Received 24 April 2008; Accepted after revision 26 June 2008
DOI 10.1002/ijc.23844
Published online 19 August 2008 in Wiley InterScience (www.interscience.wiley.com).

biomarkers and despite the rapidly advancing proteomic techniques with superior sensitivity, none of the potential biomarkers identified in proteomic experiments has found a niche for the management of breast cancer at the clinical level.[8] Very few, if any, serum tumor markers have been introduced to the clinic over the past 15 years.[9]

Nevertheless, proteomics, and in particular mass spectrometry, has been used to identify novel breast cancer biomarkers. Such studies have predominantly examined breast tumor tissues and biological fluids including serum, plasma, nipple aspirate or ductal lavage as well as cancer cell lines.[10] The intent of examining these different sources include obtaining a better understanding of mammary oncogenesis[11] and potentially leading to improvements in screening, diagnosis as well as prognosis and/or prediction of therapeutic response. Because breast cancer is a complex and heterogenous disease, no single model or biological source is expected to mimic all aspects of the disease.[12] For this reason, an approach to biomarker development should be well conceived and play to the strengths of current technologies while acknowledging and addressing the limitations.[8]

One of the sources to mine for potential biomarkers is serum or plasma of breast cancer patients, when compared with serum of healthy controls. Exploring biological fluids is an attractive way to look at secreted proteins. The analysis of plasma for breast cancer biomarkers (and other cancer markers) is currently ongoing.[9,13] It has been estimated that blood contains more than 100,000 different protein forms with abundances that span 10–12 orders of magnitude.[9] Unfortunately, the discovery of tumor-derived biomarkers by analyzing plasma is challenging because the 20 most abundant plasma proteins (concentration ranges in the mg/mL range) account for 99% of the total protein mass and impede detection of lower abundance tumor antigens.[9] Potential tumor markers are expected to exist in the low ng–pg/mL concentration range. Currently, without up-front fractionation techniques, the presence of major proteins in blood represents a technological challenge for the detection of the less abundant ones. The main concern is suppression of ionization of low abundance proteins by high abundance proteins such as albumin and immunoglobulins.

Fortunately for breast cancer, the mammary gland offers the possibility to access local fluids, which could be potential sources for breast cancer biomarker discovery. Fluid found within the ductal and lobular system of the breast can be extracted through the nipple using an aspiration device to obtain nipple aspirate fluid (NAF).[14] Nonpregnant and nonlactating women continuously secrete and reabsorb this fluid.[15] Consequently, NAF is a viable source to me because it surrounds the ducts and breast epithelial cells.[16–18] Despite this, only a limited number of proteins have been identified in NAF, predominantly owing to the presence of high abundance plasma proteins.[17,18]

Alternatively, another source to me for potential biomarkers is at the tissue level—examining normal mammary gland tissues and breast tumors.[19] Nevertheless, these structures are complex, incorporating different cell types with different proportions such as epithelial cells, adipocytes, myoepithelial cells and fibroblasts. Because of this multifaceted population, breast tumor cells comprise a minor fraction of this whole population of cells. Furthermore, tumor biopsies also contain blood components; therefore proteomic analysis of breast tumor tissues also identifies proteins from circulating cells and from plasma.[20] For tissue proteomics, the hypothesis is that certain proteins originating in the tissue could subsequently appear and be monitored in the bloodstream. Leaky capillary beds, local production of proteases and the high rates of cell death within the tumor mass are expected to facilitate shedding or secretion of tumor proteins into the bloodstream. But given the complexity of analyzing tissues, microdissection can be regarded as a reasonable alternative for selectively isolating individual cell types. The limitations with this approach, though, are the low amounts of material obtained, the large amount of sample that is needed to perform an experiment and the quality of the dissected material which interfere with proteomic experiments.[21]

Interestingly, there has also recently been an effort to take advantage of animal models in breast cancer research and their examination by proteomics.[22] For example, a conditional HER-2/neu-driven mouse model of breast cancer was used to examine the proteome of tumor and normal mammary tissue.[23] The authors identified over 700 proteins. A caveat to using an animal model to study human disease is whether the same genetic alterations transform both mouse and human epithelial cells.[12] Furthermore, some important aspects of breast cancer, particularly steroid hormone dependence, are not well modeled in mice.[24] Regardless of the species differences, examining the tissues and/or biological fluids in the rodent model has the same limitations as examining them in humans.

Tissue culture based biomarker discovery platform using proteomics

Despite optimistic views that many more protein cancer biomarkers will be discovered through various high-throughput techniques, very few, if any, serum cancer biomarkers have been introduced at the clinic. These molecules have not yet been identified presumably because their concentration in serum and/or biological fluids are too low and therefore cannot be measured or purified, unless specific immunological reagents and highly sensitive ELISA methods are available. Therefore, in the initial discovery phase for novel cancer biomarkers, a less complex sample (elimination of high abundance proteins) is essential. Although clinical validation of biomarkers must address variability arising from genetic, environmental, and behavioral differences among humans, optimization of the discovery and candidate verification processes involves controlling as many biological variables as possible so that the current technologies being used can be directly evaluated. Moreover, given that a secretome in a tumor microenvironment contains the extracellular matrix, constituted by proteins, receptors and adhesion molecules as well as a whole host of secreted proteins such as cytokines, chemokines, growth factors and proteases—all of which can be potential biomarkers,[25] a cell culture based model for sampling the secretome associated with breast cancer appears promising. Secreted proteins play important roles in physiology and pathophysiology and they can act locally and systemically in the body. The secretome reflects the functionality of a cell in a given environment.[26] For cell culture-based proteomic studies, the hypothesis is that proteins or their fragments originating from cancer cells (hence present in the conditioned media (CM)) may eventually enter the circulation. Therefore, novel candidate tumor markers for breast cancer are secreted or shed proteins that can be harnessed from tissue culture supernatants of human breast cancer cell lines using mass spectrometry. CM as a source to mine for biomarkers is increasingly gaining popularity, as illustrated by the rise in publications over the past few years.

Breast cancer cell lines have been the most widely used models to investigate how proliferation, apoptosis and migration become deregulated during the progression of breast cancer.[27] A number of studies have used a cell culture model system where the cells were grown in serum-free media (SFM) to perform proteomic analysis.[28–33] The clinical relevance of using a cell culture model to understand biological processes and functions has been examined. Using DNA microarrays, the molecular subtypes of 31 breast cell lines yielded 2 discriminating clusters corresponding to luminal cell lines and basal/mesenchymal cell lines.[34] The basal subtype was further subdivided into Basal A and Basal B; this subdivision was not observed in primary tumors. In primary tumors, gene expression patterns have been used to classify breast tumors into 5 clinically relevant subgroups (luminal A, luminal B, basal, ERBB2-overexpressing and normal-like).[35,36] In general, the luminal subtypes are estrogen receptor (ER) positive and grow slowly, whereas basal-type lack ER and are usually high-grade cancers that grow rapidly. Recently, the molecular taxonomy has been confirmed by protein expression profiling.[37,38] Also recently,

TABLE I – ADVANTAGES AND DISADVANTAGES OF A CELL CULTURE-BASED MODEL FOR BIOMARKER DISCOVERY

Advantages
Cell lines are readily available.
Cost-effective.
High-throughput.
Easily modified, versatile.
Easily propagated.
Enables secretome analysis.
Permits detection of low abundance proteins (do not represent the dynamic range problem associated with plasma; less complex mixture).
Allows for reproducibility (under well-defined experimental conditions, it yields reproducible and quantifiable results), growth standardized.
The proteome of cancer cells should reflect the genetic alterations they harbor.
Cancer cells can be grown as xenografts.
Disadvantages
No single cell line will reflect the heterogeneity of cancer.
Multiple variants of the same cell line exist.
Host stromal environment influencing tumor development and progression is absent.
A reductionist approach; cannot mimic complexity of mammary gland; does not take into account the complex interplay between cell types and the tissue microenvironment.
Does not provide insight into the evolution of breast cancer from benign lesions and normal breast epithelial cells.
Subclassification of breast cancer cell lines based on microarray data did not yield similar clusters as primary tumors.

it was found that cell lines display the same heterogeneity in copy number and expression abnormalities as the primary tumors.[39] Indeed, cancer cell lines that are invasive in culture do form tumors in immune deficient mice. This is primarily because the cancer cells in culture represent the tumor-forming cells *in vivo*. Although no single cell line is truly representative, a panel of cell lines show the heterogeneity that is observed in primary breast cancers.[39] In Table I, we outline the advantages and disadvantages to using a cell culture-based approach to discover biomarkers using proteomics.

Important factors to consider when performing a conditioned media proteomic analysis

Typically, proteomic analysis of CM involves culturing the cells in SFM to ensure that the collected CM contain no other extraneous proteins, except for the secreted or shed proteins from the cancer cells, thereby facilitating their identification through MS. Furthermore, given that the cell lines to be used are specific to epithelial breast cancer cells, the proteins present in the CM must originate from the cancer cell and not from the surrounding stroma, thereby avoiding unnecessary complications in the analyses. Seeding density, incubation time in SFM, volume of media used, type of SFM, type of tissue culture flasks are all variables that need to be explored thoroughly to select the most optimal conditions for growth. Culture conditions also need to determine the amount of cell death and autolysis occurring in SFM. Measurements of lactate dehydrogenase, an intracellular protein which if measured in the CM of cell lines is an indicator of cell death, can be utilized. Alternatively, the amount of major cytosolic proteins such as beta-actin or beta-tubulin can be used to optimize incubation times. Also, when performing a proteomic analysis of CM from cell lines, it is important that the cells are extensively washed to remove protein components arising from the fetal bovine serum (FBS) used in culture. Therefore, an essential experiment to perform is to incubate tissue culture flasks with no cells added, but treated with the same wash conditions and incubation times to serve as negative controls. The proteins identified in the negative controls, predominantly FBS-derived proteins, can be deleted from the list of proteins identified in the CM as arising from incomplete washing of the flasks. Moreover, it is known that cell growth is slower in SFM and that cells are prone to autolysis resulting in nonspecific release of intracellular proteins into the culture supernatant. Hence, it is important to obtain a proteome of the whole cell lysate derived from the same cell samples that were the sources of CM to consider only proteins found uniquely in the CM *versus* those that were found in the cell lysate. Finally, given the selective ionization process of mass spectrometry, it is important to have at least biological triplicates (same cell line prepared independently) in the analysis.

Tandem mass spectrometry analysis of breast cancer conditioned media

Recently, we performed an extensive proteomic analysis of supernatants from 3 breast cancer cell lines: MCF-10A, BT474 and MDA-MB-468.[40] MCF-10A, a basal B subtype with intact p53, was derived by spontaneous immortalization of breast epithelial cells from a patient with fibrocystic disease and it has been used extensively as a near-normal control in breast cancer studies.[41] These cells do not survive when implanted subcutaneously into immunodeficient mice.[41] BT474, a luminal subtype, obtained from a stage II localized solid tumor, is positive for ER and progesterone receptor (PgR), which represent 50–60% of all breast cancer cases.[42] This cell line also displays amplification of HER-2/neu or erbB-2—which represents 30% of all breast cancer cases.[43] Finally, MDA-MB-468, a basal A-like subtype, obtained from a pleural effusion of a stage IV patient,[44] is ER and PgR negative (15–25% of breast cancer) and PTEN negative (30% of breast cancer).[45,46] Our comparative proteomic analysis of the CM of MCF-10A, BT474 and MDA-MB-468 identified over 600, 500 and 700 proteins, respectively. A large portion of the proteins was present in all 3 cell lines; however, a significant portion contained proteins that were unique to each of the lines. Various proteases, receptors, protease inhibitors, cytokines and growth factors were identified.

Another interesting study used tandem mass spectrometry to sample the CM of 4 isogenic breast cancer cell lines differing in aggressiveness.[47] Three independent secreted proteome preparations (biological replicates) for each of the cancer cell lines were performed. Using a protein fractionation strategy involving C2 columns previously reported to enrich for secreted proteins in the CM,[29] the authors identified over 250 proteins per cell line. From the 37 most significant secreted proteins across the isogenic cell lines, 31 (87%) were also observed in our breast cancer CM analysis of 3 cell lines. More recently, another group analyzed the CM of human mammary epithelial cells (HMEC) by LC-MS/MS and identified ~900 proteins.[48] To specifically focus on proteins that were secreted or shed, the authors compared their CM data with their previous work on analyzing the proteome of whole HMEC lysates. Approximately 150 proteins were identified to be enriched in the extracellular compartment of HMEC. Eighty-three of these proteins were also identified in our CM analysis (55% overlap). Table II depicts the 13 extracellular and membrane proteins that were common to all 3 CM analysis of breast cancer cell lines.

Quantitative proteomic approaches to discovery breast cancer biomarkers using cell culture

The use of quantitative proteomic technologies to identify promising candidate biomarkers is gaining popularity. Different compounds with equal concentrations do not necessarily ionize with the same efficiency. Therefore, techniques have been implemented to enable differential quantification of proteins with MS. In particular, the use of stable isotope labeling by amino acids in cell culture (SILAC) is a promising strategy for comparative proteomics when using cell culture.[49] SILAC relies on metabolic incorporation of an isotopically labeled amino acid. Two groups

TABLE II – OVERLAPPING EXTRACELLULAR AND MEMBRANE PROTEINS ACROSS THREE INDEPENDENT STUDIES WITH BREAST CANCER CELL LINE CONDITIONED MEDIA[1]

Protein name	Accession numbers
Clusterin precursor	IPI00291262, IPI00400826
Cystatin C precursor	IPI00032293
Dystroglycan precursor	IPI00028911
Galectin-3-binding protein precursor	IPI00023673
Lysyl oxidase homolog 2 precursor	IPI00294839
Mesothelin isoform 1 preproprotein	IPI00025110, IPI00645972
Metalloproteinase inhibitor 1 precursor	IPI00032292, IPI00642739
Peptidyl-prolyl cis-trans isomerase A	IPI00419585
Plasminogen activator inhibitor 1 precursor	IPI00007118
quiescin Q6 isoform a	IPI00003590
SPARC precursor	IPI00014572, IPI00654685
Splice Isoform 1 of SPARC-related modular calcium-binding protein 1 precursor	IPI00301812, IPI00412898
Thrombospondin-1 precursor	IPI00296099

[1]Data from previously published studies.[40,47,48]

of cells are grown in culture media that are identical except in 1 respect: the first media contains the "light" and the other a "heavy" form of a particular amino acid (for *e.g.* L-leucine or deuterated L-leucine). Through the use of essential amino acids (those not synthesizable by the cell-type) the cells are forced to use the particular labeled or unlabeled form of the amino acid. In each cell doubling, the cell population replaces at least half of the original form of the amino acid, eventually incorporating 100% of a given light or heavy form of the amino acid. A variety of amino acids are suitable in SILAC, including arginine, leucine, lysine, serine, methionine and tyrosine. The different cell line CM can then be combined and run together in a single MS run. The advantages of SILAC include the fact that the labeling process is highly efficient, it does not require additional purifications to remove excess labeling reagent, nor does it involve multi-step labeling protocols and the sample preparation bias introduced by the comparison of 2 separate preparation steps is avoided. As well, SILAC allows the experimenter to use any method of protein or peptide purification (after enzymatic digestion) without introducing error into the final quantitative analysis. In 1 study, SILAC was utilized to examine differential membrane expression between normal and malignant breast cancer cells.[50] Approximately 1,000 proteins were identified with more than 800 of these proteins being classified as membrane or membrane-associated. Although the majority of the proteins remained unchanged when compared with the corresponding normal cells, a number of proteins were found up-regulated or down-regulated by greater than 3-fold.

Another relatively newer quantitative technology is Isobaric Tagging Reagent for Absolute Quantitation (iTRAQ). It involves the isobaric and amine labeling of peptides. The iTRAQ tag consists of a reporter group with 4 different masses in the low m/z region (114, 115, 116 and 117), a balancer group and a peptide reactive group which covalently links the iTRAQ tag to each lysine and N-terminus. After digestion with a protease, the peptides are labelled with each of the respective iTRAQ tags and then combined together before MS. During MS/MS, the iTRAQ tag is cleaved and the reporter group is released (114–117 mass range), which is free from other common ions. The intensity of the reporter ions indicate the relative or absolute quantities of different states examined. One of the major advantages of using iTRAQ is that it enables comparison of up to 4 different samples in a single LC-MS/MS run. Although iTRAQ has been used to label cell lysates, in 1 study, the authors used iTRAQ to profile the tyrosine phosphorylation level of proteins in the MCF10AT model of breast cancer progression.[51]

Criteria for breast cancer candidate biomarker selection

Using a cell culture based model, a proteomic platform for biomarker discovery can be utilized, which consists of 3 major phases. The first phase involves the identification of markers using multi-dimensional protein identification technology (discovery phase). Following identification, the proteins must be prioritized to select a subset of marker candidates based on several criteria such as availability of reagent set for assay development and literature association to disease biology. The second phase consists of developing preliminary assays to measure the levels of the selected proteins in a relevant biological fluid comprising cancer and normal patients (verification phase). The final phase requires expanding the number of samples used to evaluate only the candidates that continued to show promise in discriminating cancer from normal from the verification phase (validation phase). This step involves the development of a robust analytical method such as immunoassay to measure the proteins accurately in clinical samples.

A major rate-limiting step for a biomarker discovery platform thus far has been in the selection of candidate molecules to investigate further as breast cancer biomarkers from the over 100–1,000s of proteins identified in the discovery phase. Given the different experimental questions being asked, a standard procedure for candidate selection from a discovery platform could not be established. However, a number of filtering criteria exists that can be applied to a dataset. In our proteomic analysis of breast cancer cell lines, we chose to focus on only the extracellular and membrane proteins identified because it is these proteins that have the highest chance of entering the circulation and hence serving as serological markers. As well, 1 of the properties of an ideal tumor marker is that it should be tissue specific. Currently, 1 of the few markers used in the clinic that are tissue-specific is prostate-specific antigen (PSA) for prostate cancer. Therefore, it may be important to examine tissue specificity of the narrowed list of proteins from the discovery phase using the Unigene database or in silico analysis[52] or comparing the proteome of breast cancer tissues to the proteome of CM and focusing only on the overlapping members. In fact, among the 54 proteins identified by Alldridge et al as being strongly over-expressed in breast tumors when compared with healthy control tissues, only a handful of proteins were classified by localization as extracellular or membrane proteins, which would have the highest chance of entering the circulation.[19] Nevertheless, our CM analysis did identify approximately half of the differentially expressed breast tissue-specific proteins (data not shown). Another criterion to select a more limited set of proteins for future exploration as potential breast cancer biomarkers is to compare the dataset with mRNA microarray databases[53] as well as to examine overlapping proteins with proteomes of relevant biological fluids such as NAF or tumor interstitial fluid (TIF).[54] Additionally, it is also equally important to perform literature searches to identify molecules that have not been previously studied as serological markers for breast cancer.

Some of the secondary filtering criteria that may assist in selection of candidate molecules include examining factors such as reagent availability (recombinant protein, antibodies and ELISA) and focusing on proteins known to participate in pathways or signaling cascades relevant to cancer progression. If different cell lines were used in the discovery phase differing in aggressiveness, then, proteins can be selected that are expressed only in the early or advanced stages of cancer and not in the normal cell line. Table III lists some of the criteria that can be applied to investigate further only the proteins that have the highest chance of being successful in the verification and validation phases.

Where the future lies

The classical tumor markers carcinoembryonic antigen and alpha-feto protein were discovered in the '60s mainly due to the introduction of novel and relatively sensitive immunological

TABLE III – PROPOSED FILTERING CRITERIA FOR CANDIDATE SELECTION

1. Focus on extracellular and membrane proteins.
2. Examine tissue specificity (Unigene database, SAGE[1]).
3. Compare to mRNA microarray databases.
4. Compare to proteins identified in other biological fluids.
5. Examine molecules not previously examined as serological markers for the cancer type (literature searches).
6. Perform tissue proteomics on relevant tissues to the cancer type and compare with cell line data to select only candidates that overlap between the two.
7. Reagent availability such as ELISA, antibodies etc.
8. Focus on proteins known to participate in pathways/signaling related to cancer.
9. Focus on differentially expressed proteins (based on label-free or quantitative proteomics).
10. Exclude high-abundance (μg/mL) serum proteins, especially liver products and acute-phase reactants.

[1]SAGE: Serial analysis of gene expression.

techniques (such as radial immunodiffusion), which allowed for the detection of these antigens in cancer tissues with high specificity and reasonable sensitivity. The most contemporary cancer biomarkers used at the clinic today (such as carbohydrate antigen CA 125, CA 15.3, CA 19.9 and PSA) were mainly developed due to the emergence, in the late '70s, of the monoclonal antibody technology. Most of these tumor markers were discovered by using cell lines or tumor extracts as immunogens and then selecting specific hybridoma clones which recognized these tumor antigens.[55] Therefore, it is conceivable that novel tumor markers may be identified in the CM of cancer cell lines using newly emerging technologies such as mass spectrometry. The assumption that cancer biomarkers to be discovered will be secreted or shed proteins is reasonable, because it is expected that secreted or membrane-bound proteins, the latter having the potential to be cleaved, have a high chance of reaching the circulation and can be found in serum, where they can be measured with immunological techniques. All currently known cancer biomarkers are indeed secreted or shed proteins.

It is conceivable that either the tumor itself or its microenvironment could be sources for biomarkers that would ultimately be shed into the serum proteome, allowing for early disease detection and for monitoring therapeutic efficacy.[56] However, in a tissue culture-based model for discovery, the role of the microenvironment as being a source of biomarkers is currently not considered. A major limitation of *in vitro* cell culture studies is that the culture conditions used to grow the cells do not mimic the breast microenvironment. The use of 3D culture methods of breast epithelial and tumor cell lines may be considered as an alternative to the traditional 2D culturing on plastic. The modeling of the microenvironment in 3D cultures has previously been reviewed.[57,58] Likewise, during tumor growth, it is known that that the stroma plays an important role in establishing a paracrine interaction to facilitate tumor progression.[12] Given that 2D cell cultures consist of an interaction of cells of the same type, they do not represent the "true" interaction. Alternatively, coculturing of cells to build a heterotypic culture may allow for the interactions between tumor cells and stromal cells to occur and hence mimic an *in vivo* situation. Cocultures have been used in the elucidation of breast cancer signaling pathways[59] but as of yet, the use of cocultures to perform proteomic analysis on the resulting CM for biomarker identification has not been performed.

Furthermore, a commonly stated limitation for a tissue culture-based model of cancer is that no single cell line is truly representative of a complex and heterogenous disease such as breast cancer. However, proteomic analysis of a relatively large panel of cell lines may circumvent this limitation. Indeed, it is possible that performing a secretome analysis on at least 10 carefully selected breast cancer cell lines may yield a more complete spectrum of proteins to represent the complexity of the disease. Such an approach poses mass spectrometry workload and bioinformatic challenges. Pooling supernatants is a reasonable consideration.

Given that several breast cancer cell lines exhibit hormone dependence, the role of estrogen and progesterone-regulated signaling pathways in breast cancer can be elucidated by proteomics. Using a cell culture-based model, hormonal stimulation of the cells followed by proteomic analysis, to examine differentially expressed proteins may be promising. Another emerging area is drug stimulation of breast cancer cell lines to study responses related to specific pathways. The field of pharmaco-proteomics is particularly interested in identifying proteins that are expressed upon drug stimulation so that these proteins can serve as surrogate markers of drug response in the clinic.

Although a less complex mixture can be utilized in the discovery phase to identify hundreds of proteins using mass spectrometry, once the list of promising biomarkers is selected, verification and validation must take place with biological fluids such as serum or urine. As quantitative proteomic technologies mature, the prospects of discovering novel cancer markers are increasing. If these new technological advances prove to be successful in identifying cancer biomarkers for early cancer detection, the clinical benefits are likely to be enormous.

References

1. Etzioni R, Urban N, Ramsey S, McIntosh M, Schwartz S, Reid B, Radich J, Anderson G, Hartwell L. The case for early detection. Nat Rev Cancer 2003;3:243–52.
2. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T, Thun MJ. Cancer statistics, 2008. CA Cancer J Clin 2008;58:71–96.
3. Hayes DF, Bast RC, Desch CE, Fritsche H, Jr, Kemeny NE, Jessup JM, Locker GY, McDonald JS, Mennel RG, Norton L, Ravdin P, Taube S, et al. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. J Natl Cancer Inst 1996;88:1456–66.
4. Glish GL, Vachet RW. The basics of mass spectrometry in the twenty-first century. Nat Rev Drug Discov 2003;2:140–50.
5. Steen H, Mann M. The ABC's (and XYZ's) of peptide sequencing. Nat Rev Mol Cell Biol 2004;5:699–711.
6. Marcotte EM. How do shotgun proteomics algorithms identify proteins? Nat Biotechnol 2007;25:755–7.
7. Wright GL, Jr. Two-dimensional acrylamide gel electrophoresis of cancer-patient serum proteins. Ann Clin Lab Sci 1974;4:281–93.
8. Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. Nat Biotechnol 2006;24:971–83.
9. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics 2002;1:845–67.
10. Bertucci F, Birnbaum D, Goncalves A. Proteomics of breast cancer: principles and potential clinical applications. Mol Cell Proteomics 2006;5:1772–86.
11. Bose R, Molina H, Patterson AS, Bitok JK, Periaswamy B, Bader JS, Pandey A, Cole PA. Phosphoproteomic analysis of Her2/neu signaling and inhibition. Proc Natl Acad Sci USA 2006;103:9773–8.
12. Vargo-Gogola T, Rosen JM. Modelling breast cancer: one size does not fit all. Nat Rev Cancer 2007;7:659–72.
13. Pitteri SJ, Hanash SM. Proteomic approaches for cancer biomarker discovery in plasma. Expert Rev Proteomics 2007;4:589–90.
14. Shao ZM, Nguyen M. Nipple aspiration in diagnosis of breast cancer. Semin Surg Oncol 2001;20:175–80.
15. Wrensch MR, Petrakis NL, Gruenke LD, Ernster VL, Miike R, King EB, Hauck WW. Factors associated with obtaining nipple aspirate fluid: analysis of 1428 women and literature review. Breast Cancer Res Treat 1990;15:39–51.
16. Alexander H, Stegner AL, Wagner-Mann C, Du Bois GC, Alexander S, Sauter ER. Proteomic analysis to identify breast cancer biomarkers in nipple aspirate fluid. Clin Cancer Res 2004;10:7500–10.
17. Pawlik TM, Hawke DH, Liu Y, Krishnamurthy S, Fritsche H, Hunt KK, Kuerer HM. Proteomic analysis of nipple aspirate fluid from women with early-stage breast cancer using isotope-coded affinity

tags and tandem mass spectrometry reveals differential expression of vitamin D binding protein. BMC Cancer 2006;6:68.
18. Varnum SM, Covington CC, Woodbury RL, Petritis K, Kangas LJ, Abdullah MS, Pounds JG, Smith RD, Zangar RC. Proteomic characterization of nipple aspirate fluid: identification of potential biomarkers of breast cancer. Breast Cancer Res Treat 2003;80:87–97.
19. Alldridge L, Metodieva G, Greenwood C, Al Janabi K, Thwaites L, Sauven P, Metodiev M. Proteome profiling of breast tumors by gel electrophoresis and nanoscale electrospray ionization mass spectrometry. J Proteome Res 2008;7:1458–69.
20. Hondermarck H, Tastet C, Yazidi-Belkoura I, Toillon RA, Le Bourhis X. Proteomics of breast cancer: the quest for markers and therapeutic targets. J Proteome Res 2008;7:1403–11.
21. Hondermarck H. Breast cancer: when proteomics challenges biological complexity. Mol Cell Proteomics 2003;2:281–91.
22. Pitteri SJ, Faca VM, Kelly-Spratt KS, Kasarda AE, Wang H, Zhang Q, Newcomb L, Krasnoselsky A, Paczesny S, Choi G, Fitzgibbon M, McIntosh MW, et al. Plasma proteome profiling of a mouse model of breast cancer identifies a set of up-regulated proteins in common with human breast cancer cells. J Proteome Res 2008;7:1481–9.
23. Whiteaker JR, Zhang H, Lei Z, Wang P, Kelly-Spratt KS, Ivey RG, Piening BD, Feng LC, Kasarda E, Gurley KE, Eng JK, Chodosh LA, et al. Integrated pipeline for mass spectrometry-based discovery and confirmation of biomarkers demonstrated in a mouse model of breast cancer. J Proteome Res 2007;6:3962–75.
24. Nandi S, Guzman RC, Yang J. Hormones and mammary carcinogenesis in mice, rats, and humans: a unifying hypothesis. Proc Natl Acad Sci USA 1995;92:3650–7.
25. Chen ST, Pan TL, Juan HF, Chen TY, Lin YS, Huang CM. Breast tumor microenvironment: proteomics highlights the treatments targeting secretome. J Proteome Res 2008;7:1379–87.
26. Hathout Y. Approaches to the study of the cell secretome. Expert Rev Proteomics 2007;4:239–48.
27. Lacroix M, Leclercq G. Relevance of breast cancer cell lines as models for breast tumours: an update. Breast Cancer Res Treat 2004;83:249–89.
28. Martin DB, Gifford DR, Wright ME, Keller A, Yi E, Goodlett DR, Goodlett DR, Aebersold R, Nelson PS. Quantitative proteomic analysis of proteins released by neoplastic prostate epithelium. Cancer Res 2004;64:347–55.
29. Mbeunkui F, Fodstad O, Pannell LK. Secretory protein enrichment and analysis: an optimized approach applied on cancer cell lines using 2D LC-MS/MS. J Proteome Res 2006;5:899–906.
30. Canelle L, Bousquet J, Pionneau C, Hardouin J, Choquet-Kastylevsky G, Joubert-Caron R, Caron M. A proteomic approach to investigate potential biomarkers directed against membrane-associated breast cancer proteins. Electrophoresis 2006;27:1609–16.
31. Xiang R, Shi Y, Dillon DA, Negin B, Horvath C, Wilkins JA. 2D LC/MS analysis of membrane proteins from breast cancer cell lines MCF7 and BT474. J Proteome Res 2004;3:1278–83.
32. Adam PJ, Boyd R, Tyson KL, Fletcher GC, Stamps A, Hudson L, Poyer HR, Redpath N, Griffiths M, Steers G, Harris AL, Patel S, et al. Comprehensive proteomic analysis of breast cancer cell membranes reveals unique proteins with potential roles in clinical cancer. J Biol Chem 2003;278:6482–9.
33. Patwardhan AJ, Strittmatter EF, Camp DG, Smith RD, Pallavicini MG. Comparison of normal and breast cancer cell lines using proteome, genome, and interactome data. J Proteome Res 2005;4:1952–60.
34. Charafe-Jauffret E, Ginestier C, Monville F, Finetti P, Adelaide J, Cervera N, Fekairi S, Xerri L, Jacquemier J, Birnbaum D, Bertucci F. Gene expression profiling of breast cell lines identifies potential new basal markers. Oncogene 2006;25:2273–84.
35. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci USA 2003;100:8418–23.
36. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thornsen T, Quist H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci USA 2001;98:10869–74.
37. Abd El-Rehim DM, Ball G, Pinder SE, Rakha E, Paish C, Robertson JF, Macmillan D, Blamey RW, Ellis IO. High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses. Int J Cancer 2005;116:340–50.
38. Jacquemier J, Ginestier C, Rougemont J, Bardou VJ, Charafe-Jauffret E, Geneix J, Adélaide J, Koki A, Houvenaeghel G, Hassoun J, Maraninchi D, Viens P, et al. Protein expression profiling identifies subclasses of breast cancer and predicts prognosis. Cancer Res 2005;65:767–79.
39. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe JP, Tong F, Speed T, Spellman PT, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell 2006;10:515–27.
40. Kulasingam V, Diamandis EP. Proteomics analysis of conditioned media from three breast cancer cell lines: a mine for biomarkers and therapeutic targets. Mol Cell Proteomics 2007;6:1997–2011.
41. Soule HD, Maloney TM, Wolman SR, Peterson WD, Jr, Brenz R, McGrath CM, Russo J, Pauley RJ, Jones RF, Brooks SC. Isolation and characterization of a spontaneously immortalized human breast epithelial cell line. MCF-10. Cancer Res 1990;50:6075–86.
42. Lasfargues EY, Coutinho WG, Redfield ES. Isolation of two human tumor epithelial cell lines from solid breast carcinomas. J Natl Cancer Inst 1978;61:967–78.
43. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science 1987;235:177–82.
44. Cailleau R, Olive M, Cruciger QV. Long-term human breast carcinoma cell lines of metastatic origin: preliminary characterization. In Vitro 1978;14:911–5.
45. She QB, Solit D, Basso A, Moasser MM. Resistance to gefitinib in PTEN-null HER-overexpressing tumor cells can be overcome through restoration of PTEN function or pharmacologic modulation of constitutive phosphatidylinositol 3′-kinase/Akt pathway signaling. Clin Cancer Res 2003;9:4340–6.
46. Panigrahi AR, Pinder SE, Chan SY, Paish EC, Robertson JF, Ellis IO. The role of PTEN and its signalling pathways, including AKT, in breast cancer; an assessment of relationships with other prognostic factors and with outcome. J Pathol 2004;204:93–100.
47. Mbeunkui F, Metge BJ, Shevde LA, Pannell LK. Identification of differentially secreted biomarkers using LC-MS/MS in isogenic cell lines representing a progression of breast cancer. J Proteome Res 2007;6:2993–3002.
48. Jacobs JM, Waters KM, Kathmann LE, Camp Ii DG, Wiley HS, Smith RD, Thrall BD. The mammary epithelial cell secretome and its regulation by signal transduction pathways. J Proteome Res 2008;7:558–69.
49. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Stable isotope labeling by amino acids in cell culture. SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 2002;1:376–86.
50. Liang X, Zhao J, Hajivandi M, Wu R, Tao J, Amshey JW, Pope RM. Quantification of membrane and membrane-bound proteins in normal and malignant breast cancer cells isolated from the same patient with primary breast carcinoma. J Proteome Res 2006;5:2632–41.
51. Chen Y, Choong LY, Lin Q, Philp R, Wong CH, Ang BK, Tan YL, Loh MC, Hew CL, Shah N, Drujer BJ, Chong PK, et al. Differential expression of novel tyrosine kinase substrates during breast cancer development. Mol Cell Proteomics 2007;6:2072–87.
52. Porter DA, Krop IE, Nasser S, Sgroi D, Kaelin CM, Marks JR, Riggins G, Polyak K. A SAGE (serial analysis of gene expression) view of breast tumor progression. Cancer Res 2001;61:5697–5702.
53. Dombkowski AA, Cukovic D, Novak RF. Secretome analysis of microarray data reveals extracellular events associated with proliferative potential in a cell line model of breast disease. Cancer Lett 2006;241:49–58.
54. Celis JE, Gromov P, Cabezon T, Moreira JM, Ambartsumian N, Sandelin K, Rank F, Gromova I. Proteomic characterization of the interstitial fluid perfusing the breast tumor microenvironment: a novel resource for biomarker and therapeutic target discovery. Mol Cell Proteomics 2004;3:327–44.
55. Bast RC, Jr, Klug TL, St John E, Jenison E, Niloff JM, Lazarus H, Berkowitz RS, Leavitt T, Griffiths CT, Parker L, Zurawski VR, Jr, Knapp RC. A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer. N Engl J Med 1983;309:883–7.
56. Petricoin EF, Liotta LA. Clinical applications of proteomics. J Nutr 2003;133:2476S–84S.
57. Shaw KR, Wrobel CN, Brugge JS. Use of three-dimensional basement membrane cultures to model oncogene-induced changes in mammary epithelial morphogenesis. J Mammary Gland Biol Neoplasia 2004;9:297–310.
58. Lee GY, Kenny PA, Lee EH, Bissell MJ. Three-dimensional culture models of normal and malignant breast epithelial cells. Nat Methods 2007;4:359–65.
59. Hagemann T, Robinson SC, Schulz M, Trumper L, Balkwill FR, Binder C. Enhanced invasiveness of breast cancer cell lines upon cocultivation with macrophages is due to TNF-alpha dependent up-regulation of matrix metalloproteases. Carcinogenesis 2004;25:1543–9.