Protein microarrays: Meeting analytical challenges for clinical applications

Lance A. Liotta,[1] Virginia Espina,[1] Arpita I. Mehta,[1,4] Valerie Calvert,[2] Kevin Rosenblatt,[1] David Geho,[1] Peter J. Munson,[3] Lynn Young,[3] Julia Wulfkuhle,[2] and Emanuel F. Petricoin III[2]

[1]FDA-NCI Clinical Proteomics Program, Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, Bethesda, Maryland 20892
[2]FDA-NCI Clinical Proteomics Program, Center for Biologics Evaluation and Research, Food and Drug Administration, Bethesda, Maryland 20892
[3]Mathematical & Statistical Computing Laboratory, Center for Information Technology, National Institutes of Health, Bethesda, Maryland 20892
[4]NIH-Howard Hughes Research Scholar, Howard Hughes Medical Institute, Bethesda, Maryland 20892
*Correspondence: liottal@mail.nih.gov

Protein microarrays, one emerging class of proteomic technologies, have broad applications for discovery and quantitative analysis. A rapidly expanding use of this technology is the acquisition of information about the posttranslational modifications of proteins reflecting the activity state of signal pathways and networks, and is now employed for the analysis of biopsy samples in clinical trial research.

Signal pathway and network profiling

Cancer is functionally a proteomic disease

At a functional level, cancer is both a proteomic and a genomic disease. A cancer genetic defect is selected out during cancer progression because the defect ultimately alters the protein network generating a survival advantage for the cancer cell (Hanahan and Weinberg, 2000; Hunter, 2000). Defective, hyperactive, or dominating signal pathways may drive cancer growth, survival, invasion, and metastasis (Liotta and Kohn, 2001). An important goal of functional proteomics is to develop a "circuit map" of the normal and diseased state of the cellular protein network (Blume-Jensen and Hunter, 2001; Bowden et al., 1999; Celis and Gromov, 2003; Hunter, 2000; Jeong et al., 2000). Protein microarrays (Charboneau, 2002; Cutler, 2003; Ge, 2000; Lal et al., 2002; Liotta and Petricoin, 2000; MacBeath, 2002; MacBeath and Schreiber, 2000; Miller et al., 2003; Paweletz et al., 2001; Wilson and Nock, 2003; Zhu and Snyder, 2001, 2003), an emerging technology that examines protein-protein recognition events in a massively parallel mode, can be used to profile the working state of cellular signal pathways in a manner not possible by gene arrays. Herein we focus on the analytical challenges faced by protein arrays and propose a practical guide for optimizing construction and study design.

Individualized cancer therapy

We can envision a future technology that maps the state of a protein signal pathway within a patient's biopsied neoplastic lesion (Liotta et al., 2001; Petricoin et al., 2002). Assume that information flow through a specific node in the proteomic network requires the phosphorylation of a known protein at a specific amino acid sequence. By measuring the proportion of those protein molecules that are phosphorylated, we can infer the level of activity of that signal node. If we compare this measurement over time, or at stages of disease progression, or before and after treatment, a correlation can be made between the activity of the node and the biologic or disease state. A functional map of the state of key pathways within that patient's tumor cells will become the starting point for personalized therapy. Under this scenario, therapy can be tailored to the individual tumor's molecular defect. Moreover, it should be feasible to administer combination therapy targeting multiple interdependent points along a pathogenic pathway, or targeting separate pathways. Following rebiopsy or molecular imaging, the effect of the treatment can be monitored in real time (Liotta et al., 2001; Petricoin et al., 2002).

Protein microarray technology

Basic definitions

At an elementary level, a protein array contains an array of immobilized protein spots. Each spot can contain a homogeneous or heterogeneous set of "bait" molecules (Liotta and Petricoin, 2000; MacBeath, 2002; Zhu and Snyder, 2003). A spot on the array may display an antibody (Lal et al., 2002; Templin et al., 2002; Wilson and Nock, 2003), a cell or phage lysate (Paweletz et al., 2001; Zhu and Snyder, 2003), a recombinant protein or peptide (MacBeath and Schreiber, 2000), a drug (Humphery-Smith et al., 2002), or a nucleic acid (Petach and Gold, 2002; Schaeferling et al., 2002; Weng et al., 2002). The array is queried with (1) a probe (labeled antibody or ligand), or (2) an unknown biologic sample (e.g., cell lysate or serum sample) containing analytes of interest. By tagging the query molecules with a signal-generating moiety, a pattern of positive and negative spots is generated. For each spot, the intensity of the signal is proportional to the quantity of applied query molecules bound to the bait molecules. An image of the spot pattern is captured, analyzed, and interpreted.

Functional assay methodology

Discovery efforts fueled by genomic and proteomic technologies have led to long lists of potential drug targets. Most of these targets are proteins. There is an urgent need for function-based assays to prioritize and validate candidate targets. Protein arrays offer one means of conducting massive screening of drugs (Greenbaum et al., 2002; Wilson and Nock, 2003). The inhibitory properties of lead compounds on enzymatic drug targets can be analyzed in chip-based systems (Huels et al., 2002). Protein arrays have been employed to measure the enzyme activity profiles of the secreted and membrane pro-

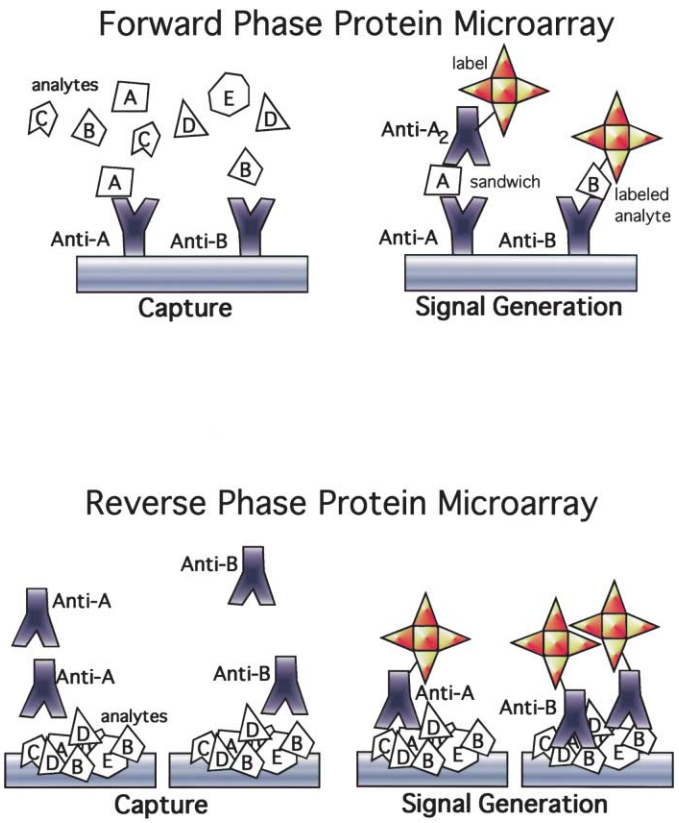

Figure 1. Classes of protein microarray platforms

Forward phase arrays (top) immobilize a bait molecule such as an antibody designed to capture specific analytes with a mixture of test sample proteins. The bound analytes are detected by a second sandwich antibody, or by labeling the analyte directly (upper right). Reverse phase arrays immobilize the test sample analytes on the solid phase. An analyte-specific ligand (e.g., antibody; lower left) is applied in solution phase. Bound antibodies are detected by secondary tagging and signal amplification (lower right).

teome of cancer cell lines (Jessani et al., 2002). In addition, new chemical reagents have been developed that can profile and identify the catalytic activity of enzymes in complex mixtures. This combination of emerging technologies has the potential to functionally map the state of enzyme networks before and after perturbation (Ideker et al., 2001; Phizicky et al., 2003; Tyers and Mann, 2003). Nevertheless, the measurement of enzymatic activity levels, by themselves, is not sufficient to recapitulate the activity level of a protein network in situ. This is because signaling networks exist within the context of the inter- and intracellular microenvironment. Consequently, array-based technology must incorporate measurements related to the status of enzyme substrates and binding partners as they exist in vivo.

Spectrum of applications

Protein microarrays are currently being studied within several fields relevant to cancer research: (1) Discovery of novel ligands or drugs that bind to specific bait molecules on the array, (2) multiplexing immunoassays to develop a miniature panel of serum biomarkers or cytokines, and (3) profiling the state of specific members of known signal pathways and protein networks. For categories 1 and 2, a variety of competing technologies already exist for protein discovery and multiplexed clinical immunoassays. These include mass spectroscopy, ICAT, 2D gel electrophoresis, bead capture, micro-ELISA (Celis and Gromov, 2003; Chen et al., 2002; Delehanty and Ligler, 2002; Pandey and Mann, 2000), and 3rd generation clinical analyzers. In contrast, for category 3, the use of protein microarrays for network profiling of cellular samples and clinical material offers the greatest potential to gather knowledge not attainable by other methods. Consequently, we will narrow our analysis to this specific application for clinical specimen profiling.

Classes of protein array technology

Currently, protein microarray formats fall into two major classes, forward phase arrays (FPA) and reverse phase arrays (RPA), depending on whether the analyte is captured from solution phase or bound to the solid phase (Figure 1). In FPAs, capture molecules, usually an antibody, are immobilized onto the substratum and act as a bait molecule. Each spot contains one type of immobilized antibody or bait protein. In the FPA format, each array is incubated with one test sample (e.g., a cellular lysate from one treatment condition), and multiple analytes are measured at once. In contrast, the RPA format immobilizes an individual test sample in each array spot, such that an array is comprised of hundreds of different patient samples or cellular lysates. In the RPA format, each array is incubated with one detection protein (e.g., antibody), and a single analyte endpoint is measured and directly compared across multiple samples.

Protein microarrays: Analytical challenges

Dynamic range of the proteome

Protein microarrays pose a significant set of analytical challenges not faced by gene arrays (Celis and Gromov, 2003; Lal et al., 2002; Zhu and Snyder, 2003). The first serious obstacle is the vast range of analyte concentrations to be detected. Protein concentrations exist over a broad dynamic range (by up to a factor of 10^{10}). To make the analysis even harder, a low abundance analyte always exists in a complex biological mixture containing a vast excess of contaminating proteins. Imagine that the specificity of a detection antibody is 99%, but a crossreacting protein exists in a thousand fold (or greater) excess. For every one analyte molecule detected, there will be ten crossreacting contaminating molecules detected, and the signal over background will be unacceptable.

Sensitivity requirements

PCR-like direct amplification methods do not exist for proteins. Consequently, protein microarrays require indirect, and very stringent, amplification chemistries (King et al., 1997; Kukar et al., 2002; Morozov et al., 2002; Schweitzer et al., 2002). Adequate sensitivity must be achieved (at least femtomolar range), with acceptable background. Moreover, the labeling and amplification method must be linear and reproducible to insure reliable quantitative analysis. Finally, the amplification chemistry must be tolerant to the large dynamic range of the analytes and the complexity of the biologic samples. The biologic sample may naturally contain biotin, peroxidases, alkaline phosphatases, fluorescent proteins, and immunoglobulins, all of which can substantially reduce the yield or background of the amplification reaction.

Clinical samples

The clinical power of protein microarrays can only be realized if the technology can be directly applied to biopsies, tissue cell aspirates, or body fluid samples. In such cases, the input sample for protein microarrays is small in volume and low in analyte concentration. A cubic centimeter of tissue may contain approx-

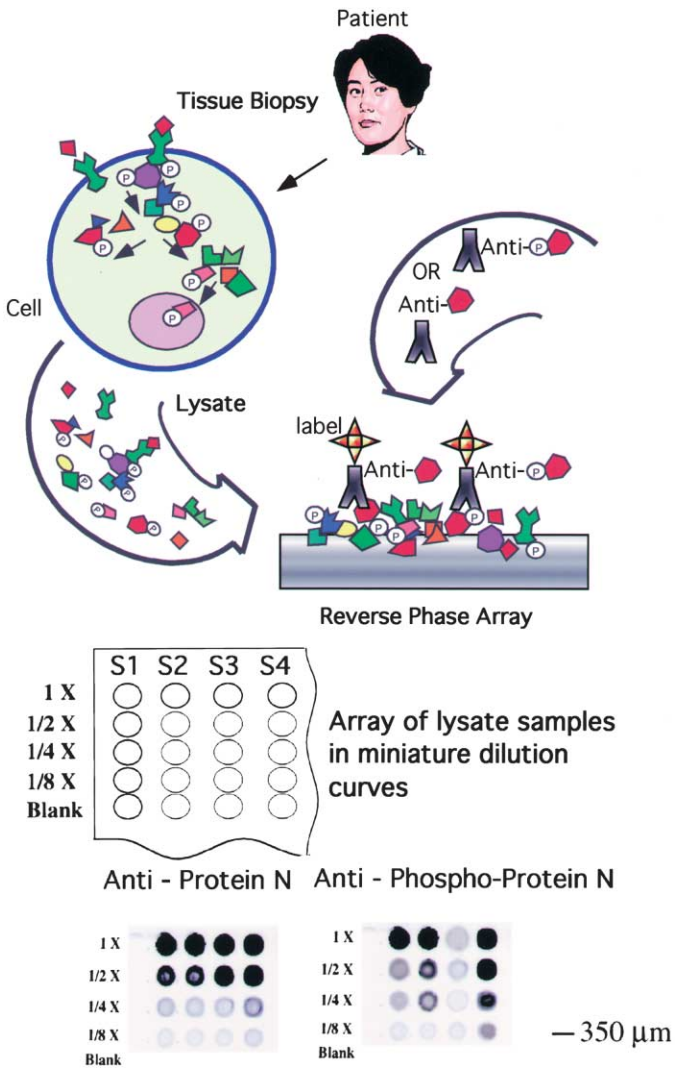

Figure 2. Reverse phase array design applied to analyze phosphorylation states of signal pathway proteins

Following tissue procurement and microdissection, the cancer cells are lysed and the entire cellular proteomic repertoire is immobilized onto a solid phase. The immobilized analyte proteins containing those phosphorylated during signal transduction are probed with two classes of antibodies that specifically recognize (1) the phosphorylated (modified) form of the protein, or (2) the total protein regardless of its modified state. Each test sample S1–S4 is arrayed and immobilized in a miniature dilution curve. Upon signal development and imaging, the relative proportion of the analyte protein molecules, which are phosphorylated, can be compared between test samples on the same array. For example, S3 has a low ratio of phosphorylated to total protein, while sample S4 has a high ratio.

imately 10^9 cells. For a core needle biopsy or a cell aspirate, the total number of cells available for analysis may be less than 100,000. Moreover, since tissues are highly heterogeneous, the population of target cells may comprise a small percentage of the total. Thus, for the analysis of cancer cells within a core biopsy, only a few thousand cancer cells may be procurable. Assuming that many proteins of interest, or their phosphorylated counterparts, exist in low abundance, the total concentration of the analyte protein in the sample is obviously very low.

If the sensitivity of an analytical system is s (moles per volume), and the number of analyte molecules per cell is x, then the threshold T for cell procurement per volume will be:

$$T = \frac{(A * s)}{x}$$

(x ≡ number of analyte molecules per cell [molecules/cell]; T ≡ threshold for cell procurement per volume [cells/volume]; s ≡ sensitivity of detection system [moles/volume]; A ≡ Avogadro's number [6.02 * 10^{23} molecules/mole]).

For example, if the abundance of a signal transduction or transcription factor protein is in the range of 10,000 copies per cell, and the analytical sensitivity of the detection is one femtomole/ml, then the number of cells T required for the assay will be approximately 60,000 cells per ml, or 60 cells per microliter. Consequently, if the analytic method does not have adequate sensitivity, the number of cells required for the assay may not exist within the range achievable for clinical utility.

Requirement for specific high affinity antibodies and ligands

Gene transcript profiling was catalyzed by the ease and throughput of manufacturing probes with known, specific, and predictable affinity constants. In contrast, the probes (e.g., antibodies, aptamers, ligands, drugs) for protein microarrays cannot be directly manufactured with predictable affinity and specificity. The availability of high quality, specific antibodies or suitable protein binding ligands is the limiting factor, and starting point, for successful utilization of protein microarray technology (Templin et al., 2002). Prior to use on any array format, antibody specificity must be thoroughly validated (e.g., single appropriate sized band on Western blot) using a complex biologic sample similar to that applied and analyzed on the array. The degree of posttranslational modifications or protein-protein interactions, for an individual analyte protein, will contain critical biologic meaning that cannot be ascertained by measuring the total concentration of the analyte. Consequently, a significant challenge for protein microarrays is the requirement for antibodies, or similar detection probes, that are specific for the modification or activation state of the target protein. Sets of high-quality modification state-specific antibodies are commercially available. Unfortunately, high-quality antibodies are currently available for only a small percentage of the known proteins involved in signal networks and gene regulation. A significant challenge for cooperative groups, funding agencies, and international consortia is the generation of large comprehensive libraries of fully characterized specific antibodies, ligands, and probes. A major initiative of HUPO (Human Proteome Organization) is the production and qualification of antibody libraries that will be made available to the scientific community (Hanash, 2003; Tyers and Mann, 2003).

Antibody affinity constrains array design

Each antibody-ligand interaction has its own unique affinity constants (association and dissociation rates) that must be determined empirically, and in some cases laboriously. A number of strategies are emerging that may speed up the process of discovering and mass producing antibody/ligands with high specificity and affinity. The affinity constants constrain the linear range of the assay. The linear detection range can only be attained if the concentration of the analyte and antibody/ligand are properly matched to the affinity. The analyte concentration in many situations is, by definition, unknown, and may be the experimental goal. A multiplexed format containing antibodies with a wide range of affinities may not be able to handle a sample containing a wide span of analyte concentrations. Such a

Table 1. Sources of technical errors and proposed solutions in microarray printing and detection

Issue	Resolution
Printing array	
Inconsistent spotting pattern	Ensure print head assembly is clean and dry.
	Check for debris under slide.
	Confirm level of platen.
Merging spots	Allow spots to dry in between printing multiple hits/spot.
Missing spots	Confirm calibration of print head assembly.
	Ensure adequate sample volume.
Compression of nitrocellulose	Reset pin calibration to a level consistent with membrane contact, but not membrane damage.
Variability in slide surface	Visually inspect commercial and in-house slide preparations for consistency in the slide membrane surface.
Processing/storage	
Water droplets on slide	Store printed slides in the dark at –20°C in a container with desiccant.
	Allow slides to reach room temperature prior to staining.
Detection technologies	
Little or absent staining	Inadequate protein content of sample.
	Confirm addition and concentration of primary and secondary antibodies.
	Confirm species cross-reactivity of antibodies for the tissue source.
	For horseradish peroxidase (HRP) detection systems, avoid reagents containing sodium azide in concentrations greater than 15 mM/l.
Expired/defective reagents	Confirm storage conditions of reagents.
	Discard expired reagents.
Dissociation of primary antibody	Optimize antibody dilution for type of antibody.
Background staining	Ensure adequate blocking of nitrocellulose membrane prior to detection protocols.
	Avoid slide drying during chromogenic detection.
	Use humidifiers or larger reagent volumes to prevent drying during incubations.
	Prepare a more dilute secondary antibody solution. Nonspecific binding to the Fc portion of IgG may be limited by using a F(ab) fragment as a secondary antibody.
No discrimination between background and target	Primary antibody may not be specific for protein of interest.
	Confirm specificity of primary antibody by Western blot.
	The antibody should produce a single band at the designated molecular weight on a Western blot.
Sample	
Viscous sample	Dilute samples for adequate sample delivery.
Limited protein quantity	Avoid degradation of proteins. Store samples at –80°C.
	Add protease inhibitors to extraction buffers.

format will achieve analytical linearity, and thus accuracy, only in a small subset of array points for which the analyte concentration and cognate antibody affinity happen to match. This problem is greatly compounded for multiplexed sandwich assays. Sandwich assays (Templin et al., 2002; Zhu and Snyder, 2003), traditionally used for clinical immunoassays, do not require direct tagging of the analyte. Instead, the analyte is captured between an immobilized antibody recognizing one epitope, and a second labeled antibody recognizing a different epitope on the same analyte. This obviously limits the assay to proteins containing two nonoverlapping accessible epitopes, and further presupposes the existence of two validated antibodies. Because there are two sets of affinity constants for each analyte, use of a sandwich format effectively doubles the stringency placed on linearity across the array.

Native versus denatured proteins

The native versus denatured state of a protein analyte is another critical factor that will affect the design and success of protein array experiments. Many antibodies are made against peptides, and require antigen denaturation in order to linearize the epitope. On the other hand, denaturation will destroy informative protein-protein interactions, and will abolish the binding of antibodies that require native 3D conformation. A common experimental strategy, which employs antibody arrays, requires direct conjugation of the analyte proteins with a fluorescent, nucleic acid, or biotin tag (Knezevic et al., 2001; Kukar et al., 2002; Lal et al., 2002). The tag becomes the basis for subsequent detection or signal amplification. Unfortunately the conjugation method itself may denature, damage, or mask the epitope. Since every tagging reaction becomes an uncontrolled experimental variable, the reproducibility of the tagging and the ultimate effect on the analyte(s) of interest are inherently imprecise. In addition, preservation of the native state may be required for one set of experiments, while a second protocol demands denaturation. The two different experiments will require completely different lysing conditions, buffers, antibody reagents, and substratum immobilization chemistries.

Protein array substrata

Most gene array platforms use planar silica substrata to immobilize the probes (Miller et al., 2002). By analogy, initial attempts

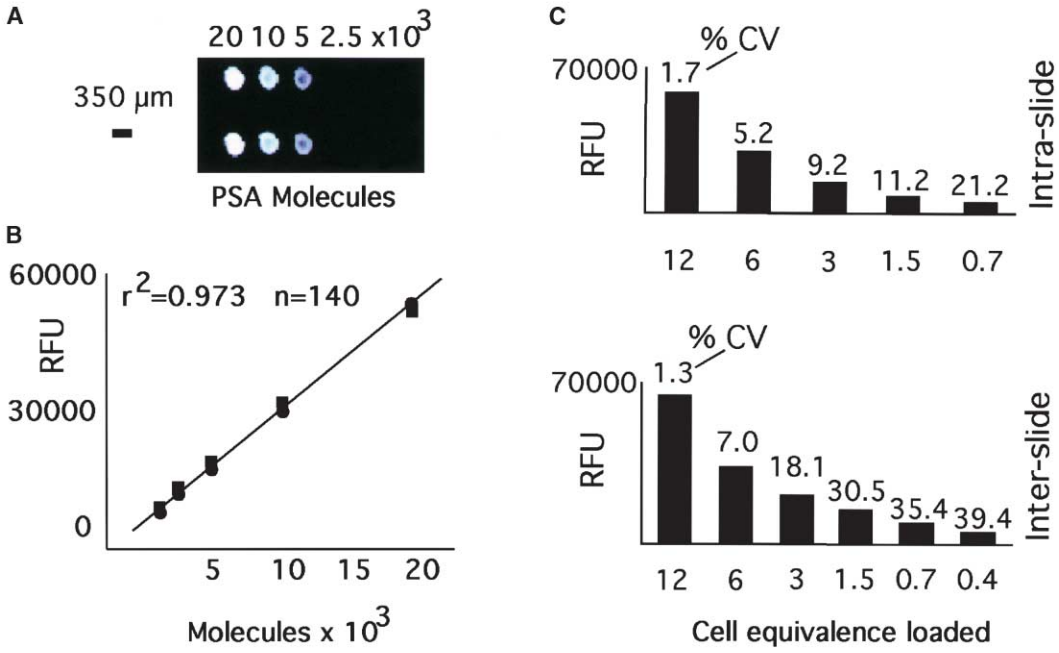

Figure 3. Validation of reverse phase protein microarrays for microdissected lysates

A: Annexin-1 stained slide of triplicate arrayed dilution curves ranging over one log of cellular equivalents loaded per spot.
B: Linearity between immobilized protein lysate and signal intensity between seven replicate slides.
C: Intra- and interspot coefficient of variance % CV. RFU, relative fluorescence units.
See Paweletz et al., 2001.

at protein microarrays have used the same substrata and spotting equipment. While this approach may be successful for high abundant or recombinant analytes, a planar substratum cannot attain sufficient surface area per spot for femtomolar detection of analyte proteins in biologically relevant samples. Optimal substrata for protein microarrays must have high binding capacity, high surface area, and intrinsically low background signal (Seong, 2002; Wilson and Nock, 2003). The choice of substratum will dictate the immobilization chemistries employed. This, in turn, will impact on the native state, or appropriate orientation of the immobilized protein bait or capture molecule. For example, if the immobilized capture molecule is an antibody, it is highly desirable to position each antibody with its ligand binding domain upright.

Need for quantitation

Accuracy and precision are paramount for protein microarrays. A small shift in a signaling pathway in the protein circuitry can dramatically affect the ultimate biologic outcome. This issue is pronounced for specific pathologic states such as premalignant progression. A subtle shift in the birth or death rate of hyperplastic cells, integrated over many years, can determine the fate of a premalignant lesion. The strength of immunohistochemistry (and by extension multiplexed tissue arrays) is to provide information about histomorphological and subcellular localization. However, immunohistochemistry is inherently subjective and nonquantitative, and may not be adequately sensitive for low abundance analytes. Variability in immunohistochemistry can be imparted by tissue handling and fixation, accessibility of the antigen epitope (antigen retrieval treatments), and cellular heterogeneity. In the face of the quantitative limitations of immunohistochemistry, investigators are intrigued with the promise of protein microarrays and the hope of obtaining high-throughput multiplexed quantitative analysis of biologic samples.

Reverse phase arrays (RPA) offer advantages for translational research

RPAs address analytical challenges

We have developed RPAs as a format that addresses the analytical challenges stated above (Charboneau, 2002; Paweletz et al., 2001; Petricoin et al., 2002) (Figure 2). RPAs can achieve adequate sensitivity for small samples. This format has been successfully applied to analyze the state of the prosurvival, apoptosis, and mitogenesis pathways within microdissected premalignant lesions, compared to adjacent normal epithelium, invasive carcinoma, and host stroma (Paweletz et al., 2001). Each spot within an RPA contains an immobilized bait zone measuring only a few hundred microns in diameter and containing a solubilized sample of the cellular material itself. The high sensitivity of RPAs is in part because the detection probe (e.g., antibody) can be tagged and the signal amplified, independent from the immobilized analyte protein. Amplification chemistries available take advantage of methods developed for highly sensitive 3rd generation commercial clinical immunoassays (Bobrow et al., 1989, 1991; Hunyady et al., 1996; King et al., 1997). For example, coupling the detection antibody with highly sensitive tyramide-based avidin/biotin signal amplification systems can yield detection sensitivities down to fewer than 1000–5000 molecules/spot. A biopsy of 10,000 cells can yield 100 RPA arrays. Each array can be probed with a different antibody.

Precision and linearity of RPAs

Using commercially available automated equipment, RPAs exhibit excellent within-run and between-run analytical preci-

sion with greater than 6 cell equivalents (3%–10% c.v.) (Figure 3). RPAs do not require direct labeling of the sample analyte, and do not utilize a two-site antibody sandwich. Therefore, there is no experimental variability introduced due to labeling yield, efficiency, or epitope masking. As each array is comprised of dozens or hundreds of experimental samples, subtle differences in an analyte can be measured, because each sample is exposed for the same amount of time to the same concentration of primary and secondary antibody and amplification reagents. RPA platforms can utilize reliable commercially available automated strainers designed for immunohistochemistry.

Miniature dilution curves

A critical factor in determining the linearity of a protein array, common to all immunoassays, is the match between the antibody probe concentration (affinity constant) and the unknown concentration of the analyte (Humphery-Smith et al., 2002). Another special and important attribute of the RPA is that each sample is applied in a miniature dilution curve on the array (Figure 2). In principle, a calibration curve is developed for each antibody, for each sample, and for each analyte concentration. This provides an improved means of matching the antibody concentration with the analyte concentration so that the linear range of each analyte measurement is insured.

Protein microarrays: A practical guide

Selection and validation of antibodies

The most important starting point for the development of any protein array method is the selection of antibodies with high specificity and adequate affinity. If the experimental focus is cellular signaling analysis, then the platform will require at least two different kinds of antibodies for each protein. One antibody is chosen to recognize the phosphorylated form of the protein, and the second recognizes the protein regardless of its phosphorylation status. Prior to use on any array, each antibody must be validated for specificity and sensitivity. Specificity should be performed by Western blotting, preferably using the same class of material that will be applied to the array. The optimal criterion for specificity is a single band at the appropriate molecular weight, which can be competed with the immunizing antigen. The optimal dilution of the primary antibody will be dictated by the relative binding affinity. This can only be determined empirically using a known concentration of reference antigen or control lysate. We have validated over 100 (out of approximately 400) commercially available phosphospecific or modification specific antibodies covering mitogenesis, survival, apoptosis, differentiation, and motility related pathways. The full list of antibodies, the commercial suppliers, and initial working dilutions is available for download at http://clinicalproteomics.steem.com.

Tissue microdissection and lysate conditions

Cellular signaling processes are dominated by the context of the cell type and the tissue microenvironment (Liotta and Kohn, 2001). Microdissection (Emmert-Buck et al., 1996) of the heterogeneous tissue sample is therefore required to obtain meaningful information. In order to reduce protein degradation by phosphatases and proteinases, the optimal protocol is to snap-freeze the tissue immediately after procurement. Protein extraction buffers with ionic and nonionic detergents will effectively solubilize the cells, while preserving phosphorylated proteins. Moreover, laser capture microdissection can employ proteinase and phosphatase inhibitors directly in the fixation and staining baths (Simone et al., 2000).

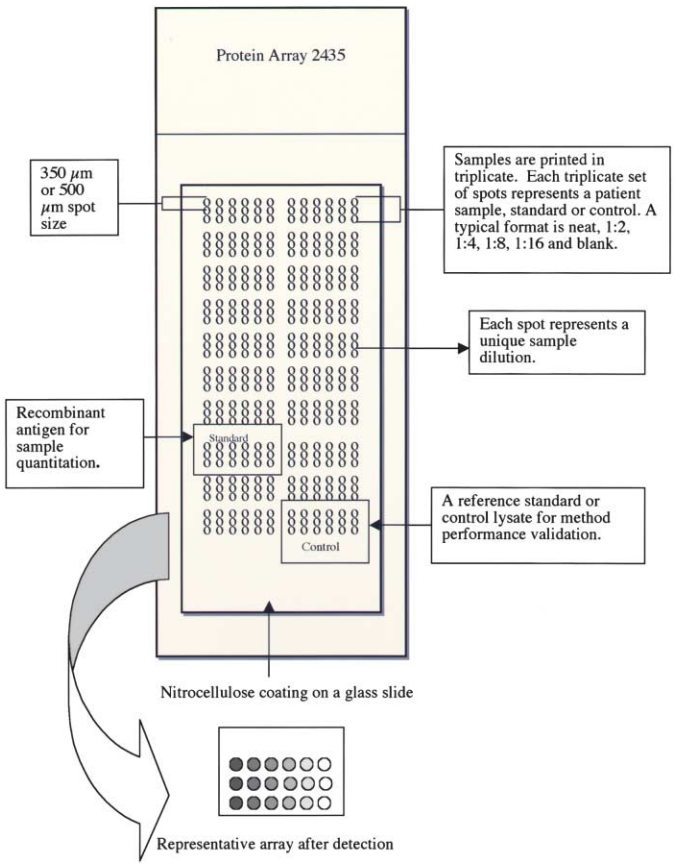

Figure 4. Idealized reverse phase array format

Triplicate samples are printed in dilution curves representing undiluted, 1:2, 1:4, 1:8, and 1:16 dilutions. The sixth spot represents a negative control, consisting of extraction buffer without sample. Each set of triplicate spots represents a patient sample before or after treatment, or microdissected normal, premalignant, or stromal tissue cells. A reference lysate or control lysate is printed on each array for monitoring assay performance. A recombinant antigen is printed on each array for comparative quantitation of patient samples. Altering spot size and/or spot spacing may vary array capacity.

Arraying and arrayers

Protein microarrays are printed using the same technology used for DNA microarrays, but the protein array layout is vastly different from a typical DNA array. Both printing technologies transfer sample fluid from a microtiter plate onto a substratum, usually a coated glass slide. The substratum requirements for protein arrays are (1) high binding capacity, (2) minimum effect on the protein structure, and (3) low background (Table 1). Nitrocellulose coated glass slides are a common substratum for protein arrays (FAST slides, Schleicher & Schuell BioSciences or Grace Biolabs). Proteins bind to nitrocellulose via electrostatic interactions in an irreversible manner (Stillman and Tonkinson, 2000). The nitrocellulose polymer coating of FAST slides permits protein binding capacities of 75–150 ug/cm^2 in a volume of 0.3–2 nl/spot. Protein arrays may also be printed in sector formats. A sector array consists of multiple small pads of substratum on a slide. A reservoir placed around each sector permits a different antibody to be used for probing the samples. The sector format miniaturizes the array, providing an increased signal/noise ratio.

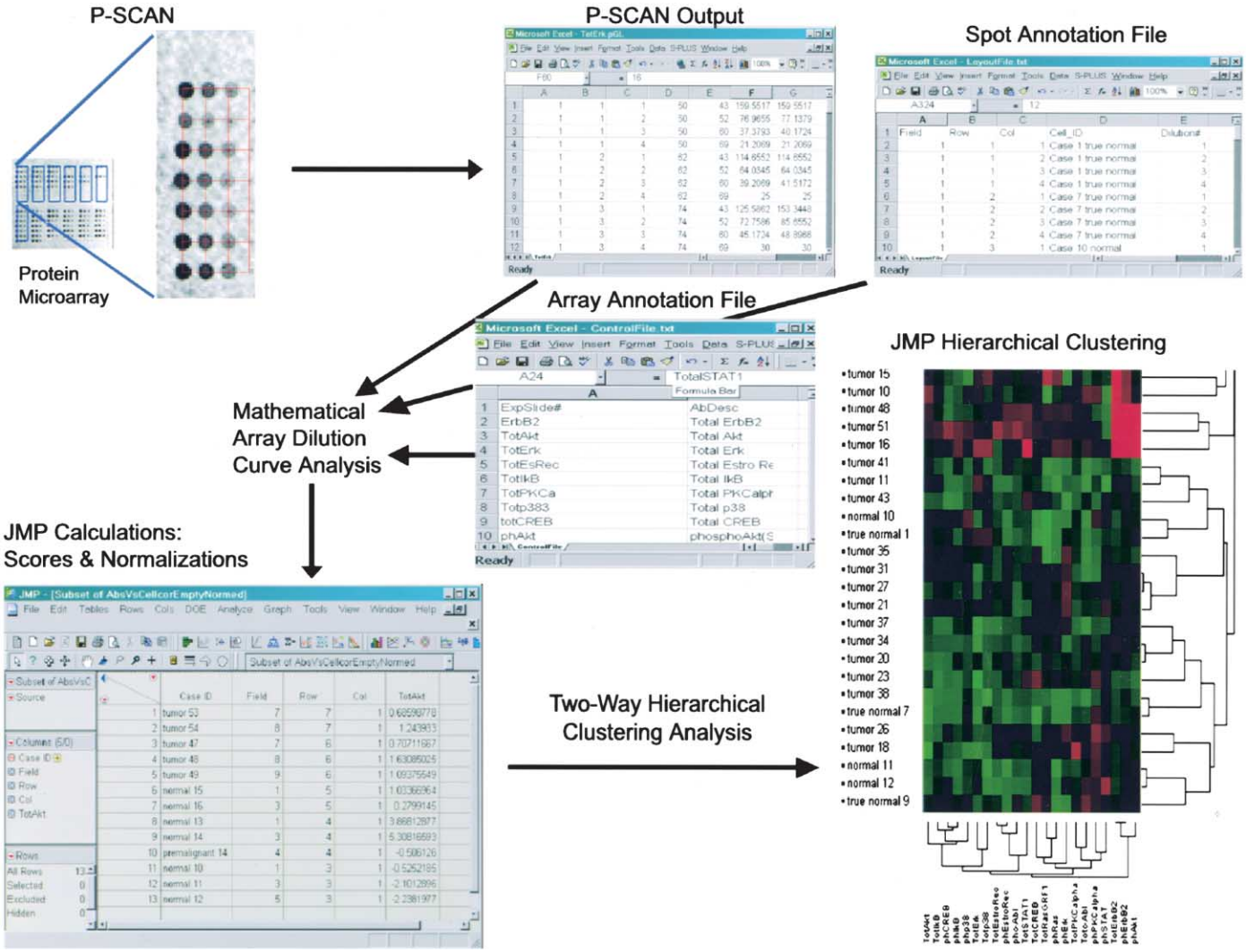

Figure 5. Scanning and bioinformatic analysis of reverse phase arrays

The example shown displays a cluster analysis of microdissected human breast cancer and normal breast epithelium (vertical axis) compared across the phosphorylation states of a series of proteins within the EGF-receptor family signal pathway (horizontal axis lower right). The data collection steps depicted are described in the text.

Signal amplification and detection

A variety of protein array labels and amplification chemistries are available (King et al., 1997; Kukar et al., 2002; Morozov et al., 2002; Schweitzer et al., 2002; Wiese, 2003). These include fluorescent, radioactive, luminescent, and colorimetric readouts. Chromogenic, fluorometric, and luminescent detection methods may be used with FAST slides with an adequate signal/noise ratio. Amplification can be achieved by enzymatic cleavage of colorimetric, luminescent, and fluorescent substrates. Several amplification systems employ coupling of nucleic acids to proteins, and thereby take advantage of the PCR method. Amplification is absolutely required to achieve the sensitivity adequate for routine analysis of relatively low abundance proteins.

Determination of baseline sensitivity

The baseline sensitivity of detection dictates the minimum number of cells T (see above) that can be profiled on the array. Using a dilution curve of a known concentration of reference antigen, the first step is to determine the intra- and inter-slide coefficients of variance. Baseline sensitivity is defined as a signal two standard deviations above background. Using the same amplification or labeling chemistries, the baseline sensitivity is different for each antibody choice. This is because each antibody has a different affinity and background binding, and its cognate antigen falls within a different dynamic range.

Bioinformatic analysis of protein array data

A variety of bioanalytical methods have been successfully used for protein microarrays (Brazma et al., 2001; Carlisle et al., 2000; Cutler, 2003; Miller et al., 2003; Sreekumar et al., 2001). Analyzing RPAs presents a new set of challenges, compared with conventional spotted arrays. Using the flexible, open-source program PSCAN (Carlisle et al.), we developed an analysis strategy tailored specifically to these arrays (Figure 5).

Array spot placement

The first challenge relates to detecting the placement of sample spots on the array. The spot placement is constrained by the pin configuration of the printer, and by the organization of serial dilutions in the original plates, and may be controlled by programming the printer robot. Currently, determining the spot locations within each image requires operator assistance, but this process could become fully automated with the inclusion of reference spots. Determination of background intensity is facilitated by reserving an adequate number of "empty" spots within the image. Spot intensity is usually estimated by integrating over a fixed spot area. Spot size should also be monitored. Systematic change in spot size with dilution may signal a change in viscosity of the printed sample, and may require a more sophisticated intensity estimation algorithm.

Normalizing spot intensity

Spot intensities for points on each dilution curve need to be reduced to a single value corresponding to the amount of protein present in the sample. Assuming that intensity above background is proportional to protein concentration, the intensity versus log-dilution plot should ideally appear as an exponential decay curve. With a sufficient number and quality of data points, well-known curve-fitting techniques are ideal for determining the parameters of relative protein concentration, background intensity, and decay constant, as well as for checking the overall fit to the curve. Deviations from ideality (nonmonotonicity, high-dose "hook" effects, nonexponential behavior, too few points, trends, or gradients in background intensity) are sometimes observed in RPA assays, and require special attention. Robust estimation techniques can be used to address the presence of outliers in the data, caused for example by printing dropouts or flares. Observing an exponential decay rate different from unity suggests nonlinearity in the intensity-dilution curve. Parallelism between multiple curves should be checked, and samples that deviate should be flagged. Apparent saturation of intensity for high concentration (low dilution) points may suggest use of only the low-dose portion of the curve.

Correction for internal sources of variability

After estimating the parameters for each dilution curve, it may then be necessary to correct for intensity gradients between different printer pins or across the slide. Thoughtful planning of the slide layout will have a big impact in such situations (Figure 4). When each set of dilutions for a sample is printed contiguously, it may be possible to correct for such trends by subtracting locally averaged values for nearby samples, yielding relative, locally corrected values for protein concentration. Finally, as the overall intensity on each slide or chip varies significantly with the particular specific antibody, it may be necessary to further adjust the values for each array, before combining into a single study.

Public software for protein array analysis

Our analytical protocol for a specific RPA study based on public software is outlined below (Figure 5). Multiple RPAs, each analyzing a different phosphorylated protein, are scanned, spot intensities are calculated and normalized, and the dilution curve is collapsed to a single intensity value (Figure 5). This value is then assigned a relative normalized intensity value referenced to the other patient samples on the array. The data output is in the form suitable for traditional unsupervised and supervised learning systems. In this way, protein array data is displayed as traditional "heat maps" and can employ powerful Bayesian clustering analysis for signal pathway profiling.

Protein microarrays can be applied to clinical material

In the current art, protein microarrays can be applied to serum (Miller et al., 2003) and tissue samples (Knezevic et al., 2001; Paweletz et al., 2001) with acceptable reliability. FPAs containing up to 184 antibodies have been successfully applied to the analysis of microdissected biotinylated cellular lysates (Miller et al., 2003) and serum antibody or biomarker screening (Knezevic et al., 2001). The major weakness of current FPAs (antibody arrays) is the low sensitivity and the between-run variability (Lal et al., 2002; Wilson and Nock, 2003). RPAs, described above, have achieved the sensitivity and precision acceptable for the analysis of clinical tissue biopsy specimens. A perceived drawback of RPA is the fact that the arrays, since they are composed of experimental samples, cannot be printed and stored ahead of time. This potential drawback is overcome by the improved ability to match the antibody and analyte in the linear range, and the lack of the requirement for analyte tagging. A major weakness of both formats is the limited (but growing) availability of high-affinity specific antibodies for a multitude of important signal pathway molecules and cancer-related gene products.

Ongoing clinical research trials at the National Cancer Institute Warren G. Magnuson Clinical Center are currently employing RPAs. According to the IRB-approved study design, core needle biopsies are procured before and after therapy, and at the time of potential recurrence. For three trials underway, RPAs are being used to study the change in the phosphorylated or cleaved state of selected proteins within EGF-receptor-family triggered, prosurvival, mitogenesis, motility, and apoptosis-related pathways. The first goal of the trials is to fully evaluate the strengths and limitations of the technology, particularly for sample procurement and processing. The second goal is to transition from hypothesis generation to prospective individual selection of combinatorial therapy. Thus, step-by-step, the hope is to realize the concept of patient-tailored therapy.

References

Blume-Jensen, P., and Hunter, T. (2001). Oncogenic kinase signalling. Nature *411*, 355–365.

Bobrow, M.N., Harris, T.D., Shaughnessy, K.J., and Litt, G.J. (1989). Catalyzed reporter deposition, a novel method of signal amplification. Application to immunoassays. J. Immunol. Methods *125*, 279–285.

Bobrow, M.N., Shaughnessy, K.J., and Litt, G.J. (1991). Catalyzed reporter deposition, a novel method of signal amplification. II. Application to membrane immunoassays. J. Immunol. Methods *137*, 103–112.

Bowden, E.T., Barth, M., Thomas, D., Glazer, R.I., and Mueller, S.C. (1999). An invasion-related complex of cortactin, paxillin and PKCmu associates with invadopodia at sites of extracellular matrix degradation. Oncogene *18*, 4440–4449.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., et al. (2001). Minimum information bout a microarray experiment (MIAME)—toward standars for microarray data. Nat Genet *29*, 365–371.

Carlisle, A.J., Prabhu, V.V., Elkahloun, A., Hudson, J., Trent, J.M., Linehan, W.M., Williams, E.D., Emmert-Buck, M.R., Liotta, L.A., Munson, P.J., and Krizman, D.B. (2000). Development of a prostate cDNA microarray and statistical gene expression analysis package. Mol. Carcinog. *28*, 12–22

Celis, J.E., and Gromov, P. (2003). Proteomics in translational cancer research: Toward an integrated approach. Cancer Cell *3*, 9–15.

Charboneau, L. (2002). Utility of reverse phase protein arrays: Applications to signaling pathways and human body arrays. Brief. Func. Genomics Proteomics *1*, 305–315.

Chen, G., Gharib, T.G., Huang, C.C., Thomas, D.G., Shedden, K.A., Taylor, J.M., Kardia, S.L., Misek, D.E., Giordano, T.J., Iannettoni, M.D., et al. (2002). Proteomic analysis of lung adenocarcinoma: identification of a highly expressed set of proteins in tumors. Clin. Cancer Res. *8*, 2298–2305.

Cutler, P. (2003). Protein arrays: The current state-of-the-art. Proteomics *3*, 3–18.

Delehanty, J.B., and Ligler, F.S. (2002). A microarray immunoassay for simultaneous detection of proteins and bacteria. Anal. Chem. *74*, 5681–5687.

Emmert-Buck, M.R., Bonner, R.F., Smith, P.D., Chuaqui, R.F., Zhuang, Z., Goldstein, S.R., Weiss, R.A., and Liotta, L.A. (1996). Laser capture microdissection. Science *274*, 998–1001.

Ge, H. (2000). UPA, a universal protein array system for quantitative detection of protein-protein, protein-DNA, protein-RNA and protein-ligand interactions. Nucleic Acids Res. *28*, e3.

Greenbaum, D., Baruch, A., Hayrapetian, L., Darula, Z., Burlingame, A., Medzihradszky, K.F., and Bogyo, M. (2002). Chemical approaches for functionally probing the proteome. Mol. Cell. Proteomics *1*, 60–68.

Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. Cell *100*, 57–70.

Hanash, S. (2003). Disease proteomics. Nature *422*, 226–232.

Huels, C., Muellner, S., Meyer, H.E., and Cahill, D.J. (2002). The impact of protein biochips and microarrays on the drug development process. Drug Discov. Today *7*, S119–124.

Humphery-Smith, I., Wischerhoff, E., and Hashimoto, R. (2002). Protein arrays for assessment of target selectivity. Drug Discov. World *4*, 17–27.

Hunter, T. (2000). Signaling–2000 and beyond. Cell *100*, 113–127.

Hunyady, B., Krempels, K., Harta, G., and Mezey, E. (1996). Immunohistochemical signal amplification by catalyzed reporter deposition and its application in double immunostaining. J. Histochem. Cytochem. *44*, 1353–1362.

Ideker, T., Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K., Bumgarner, R., Goodlett, D.R., Aebersold, R., and Hood, L. (2001). Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science *292*, 929–934.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., and Barabasi, A.L. (2000). The large-scale organization of metabolic networks. Nature *407*, 651–654.

Jessani, N., Liu, Y., Humphrey, M., and Cravatt, B.F. (2002). Enzyme activity profiles of the secreted and membrane proteome that depict cancer cell invasiveness. Proc. Natl. Acad. Sci. USA *99*, 10335–10340.

King, G., Payne, S., Walker, F., and Murray, G.I. (1997). A highly sensitive detection method for immunohistochemistry using biotinylated tyramine. J. Pathol. *183*, 237–241.

Knezevic, V., Leethanakul, C., Bichsel, V.E., Worth, J.M., Prabhu, V.V., Gutkind, J.S., Liotta, L.A., Munson, P.J., Petricoin, E.F., 3rd, and Krizman, D.B. (2001). Proteomic profiling of the cancer microenvironment by antibody arrays. Proteomics *1*, 1271–1278.

Kukar, T., Eckenrode, S., Gu, Y., Lian, W., Megginson, M., She, J.X., and Wu, D. (2002). Protein microarrays to detect protein-protein interactions using red and green fluorescent proteins. Anal. Biochem. *306*, 50–54.

Lal, S.P., Christopherson, R.I., and dos Remedios, C.G. (2002). Antibody arrays: an embryonic but rapidly growing technology. Drug Discov. Today *7*, S143–149.

Liotta, L.A., and Kohn, E.C. (2001). The microenvironment of the tumour-host interface. Nature *411*, 375–379.

Liotta, L., and Petricoin, E. (2000). Molecular profiling of human cancer. Nat. Rev. Genet. *1*, 48–56.

Liotta, L.A., Kohn, E.C., and Petricoin, E.F. (2001). Clinical proteomics: personalized molecular medicine. JAMA *286*, 2211–2214.

MacBeath, G. (2002). Protein microarrays and proteomics. Nat. Genet. *32* (*Suppl*), 526–532.

MacBeath, G., and Schreiber, S.L. (2000). Printing proteins as microarrays for high-throughput function determination. Science *289*, 1760–1763.

Miller, L.D., Long, P.M., Wong, L., Mukherjee, S., McShane, L.M., and Liu, E.T. (2002). Optimal gene expression analysis by microarrays. Cancer Cell *2*, 353–361.

Miller, J.C., Zhou, H., Kwekel, J., Cavallo, R., Burke, J., Butler, E.B., Teh, B.S., and Haab, B.B. (2003). Antibody microarray profiling of human prostate cancer sera: Antibody screening and identification of potential biomarkers. Proteomics *3*, 56–63.

Morozov, V.N., Gavryushkin, A.V., and Deev, A.A. (2002). Direct detection of isotopically labeled metabolites bound to a protein microarray using a charge-coupled device. J. Biochem. Biophys. Methods *51*, 57–67.

Pandey, A., and Mann, M. (2000). Proteomics to study genes and genomes. Nature *405*, 837–846.

Paweletz, C.P., Charboneau, L., Bichsel, V.E., Simone, N.L., Chen, T., Gillespie, J.W., Emmert-Buck, M.R., Roth, M.J., Petricoin, I.E., and Liotta, L.A. (2001). Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. Oncogene *20*, 1981–1989.

Petach, H., and Gold, L. (2002). Dimensionality is the issue: use of photoaptamers in protein microarrays. Curr. Opin. Biotechnol. *13*, 309–314.

Petricoin, E.F., Zoon, K.C., Kohn, E.C., Barrett, J.C., and Liotta, L.A. (2002). Clinical proteomics: translating benchside promise into bedside reality. Nat. Rev. Drug Discov. *1*, 683–695.

Phizicky, E., Bastiaens, P.I., Zhu, H., Snyder, M., and Fields, S. (2003). Protein analysis on a proteomic scale. Nature *422*, 208–215.

Schaeferling, M., Schiller, S., Paul, H., Kruschina, M., Pavlickova, P., Meerkamp, M., Giammasi, C., and Kambhampati, D. (2002). Application of self-assembly techniques in the design of biocompatible protein microarray surfaces. Electrophoresis *23*, 3097–3105.

Schweitzer, B., Roberts, S., Grimwade, B., Shao, W., Wang, M., Fu, Q., Shu, Q., Laroche, I., Zhou, Z., Tchernev, V.T., et al. (2002). Multiplexed protein profiling on microarrays by rolling-circle amplification. Nat. Biotechnol. *20*, 359–365.

Seong, S.Y. (2002). Microimmunoassay using a protein chip: optimizing conditions for protein immobilization. Clin. Diagn. Lab. Immunol. *9*, 927–930.

Simone, N.L., Remaley, A.T., Charboneau, L., Petricoin, E.F., 3rd, Glickman, J.W., Emmert-Buck, M.R., Fleisher, T.A., and Liotta, L.A. (2000). Sensitive immunoassay of tissue cell proteins procured by laser capture microdissection. Am. J. Pathol. *156*, 445–452.

Sreekumar, A., Nyati, M.K., Varambally, S., Barrette, T.R. Ghosh, D., Lawrence, T.S., and Chinnaiyan, A.M. (2001). Profiling of cancer cells using protein microarrays: discovery of novel radiation-regulated proteins. Cancer Res. *61*, 7585–7593.

Stillman, B.A., and Tonkinson, J.L. (2000). FAST slides: a novel surface for microarrays. Biotechniques *29*, 630–635.

Templin, M.F., Stoll, D., Schrenk, M., Traub, P.C., Vohringer, C.F., and Joos, T.O. (2002). Protein microarray technology. Trends Biotechnol. *20*, 160–166.

Tyers, M., and Mann, M. (2003). From genomics to proteomics. Nature *422*, 193–197.

Weng, S., Gu, K., Hammond, P.W., Lohse, P., Rise, C., Wagner, R.W., Wright, M.C., and Kuimelis, R.G. (2002). Generating addressable protein microarrays with PROfusion covalent mRNA-protein fusion technology. Proteomics *2*, 48–57.

Wiese, R. (2003). Analysis of several fluorescent detector molecules for protein microarray use. Luminescence *18*, 25–30.

Wilson, D.S., and Nock, S. (2003). Recent developments in protein microarray technology. Angew Chem Int Ed Engl *42*, 494–500.

Zhu, H., and Snyder, M. (2001). Protein arrays and microarrays. Curr. Opin. Chem. Biol. *5*, 40–45.

Zhu, H., and Snyder, M. (2003). Protein chip technology. Curr. Opin. Chem. Biol. *7*, 55–63.