

Commentary

Beyond the genome to tissue proteomics

Lance A Liotta and Emanuel F Petricoin

National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA

Received: 30 November 1999
Accepted: 3 December 1999
Published: 17 December 1999

Breast Cancer Res 2000, **2**:13–14

© Current Science Ltd

The ongoing revolution in molecular medicine can be divided into three phases. The first phase is gene discovery, in which the tools of molecular biology are applied to identify and sequence previously unknown genes. Identification of most of the expressed human genes will be accomplished before 2005. The second phase is molecular fingerprinting, which correlates the genomic state, the complementary DNA expression pattern, and the protein repertoire with the functional status of the cells or tissue. The promise of this phase is that expression profiles can uncover clues to functionally important molecules, and will generate information to tailor a treatment to the individual patient. The third phase is the synthesis of proteomic information into functional pathways and circuits in cells and tissues. This must take into account the dynamic state of protein post-translational modifications and protein–protein or protein–DNA interactions. Through an integrated genomic/proteomic analysis, the ultimate outcome will be an actual functional understanding of the molecular events that underlie normal development and disease pathophysiology. This higher level of functional understanding will be the basis for true rational therapeutic design.

Progress in these three phases of molecular medicine is largely driven by new technologies. The development of polymerase chain reaction, high throughput sequencing, and bioinformatics has been a driving force in the first phase. In the second phase, microhybridization arrays applied to genetic analysis and gene expression [1] is a powerful new tool that has entered the commercial sector, and is becoming widely available to researchers. As more genes are identified, it is likely that specialized arrays will be offered that are specific for a tissue type (eg mammary gland chip), physiologic process (eg apoptosis chip, angiogenesis chip, invasion chip) or class of genes (eg suppressor gene chip, oncogene chip).

Whereas DNA is an information archive, proteins do all the work of the cell. The existence of a given DNA sequence does not guarantee the synthesis of a corresponding protein [2,3]. The DNA sequence is also not sufficient to

describe protein structure, function, and cellular location. This is because protein complexity and versatility stems from context-dependent post-translational processes such as phosphorylation, sulfation, and glycosylation. Moreover, the DNA code does not provide information about how proteins link together into networks and functional machines in the cell. In fact, the activation of a protein signal pathway causing a cell to migrate, die, or initiate division can immediately take place before any changes occur in DNA/RNA gene expression. Consequently, the technology to drive the molecular medicine revolution into the third phase is emerging from protein analytic methods.

The term 'proteome', which denotes all the proteins expressed by a genome, was first coined in late 1994 at the Siena two-dimensional gel electrophoresis meeting [4]. Proteomics is proclaimed as the next step after genomics. A goal of investigators in this exciting field is to assemble a complete library of all of the proteins. Only a small percentage of the proteome has been cataloged to date [2,3]. Because 'polymerase chain reaction for proteins' does not exist, sequencing the order of 20 possible amino acids in a given protein remains relatively slow and labor intensive, compared with nucleotide sequencing. Although a number of new technologies are being introduced for high throughput protein characterization and discovery [3,5], the mainstay of protein identification continues to be two-dimensional gel electrophoresis. Two-dimensional electrophoresis can separate proteins by molecular weight in one dimension and charge in the second dimension. When a mixture of proteins is applied to the two-dimensional gel, individual proteins in the mixture are separated out into signature locations on the display, depending on their individual size and charge. Each signature is a 'spot' on the gel, which can constitute a unique single protein species. The protein spot can be procured from the gel and a partial amino acid sequence can be read. In this manner known proteins can be monitored for changes in abundance under treatment or new proteins can be identified. An experimental two-dimensional gel image can be captured and overlaid digitally with known archived two-dimensional gels. In this

way it is possible to immediately highlight proteins that are differentially abundant in one state versus another (eg tumor versus normal, or before and after hormone treatment).

Two-dimensional gels have traditionally required large amounts of protein starting material, equivalent to millions of cells. Thus, their application has been limited to cultured cells or ground-up heterogeneous tissue. Not unexpectedly, this approach does not provide an accurate picture of the proteins that are in use by cells in real tissue. Tissues are complicated structures composed of hundreds of interacting cell populations in specialized spatial configurations. The fluctuating proteins expressed by cells in tissues may bear little resemblance to the proteins made by cultured cells that are torn from their tissue context and reacting to a new culture environment. Proteins extracted from ground-up tissue will represent an averaging-out of proteins from all of the heterogeneous tissue subpopulations. For example, in the case of breast tissue the glandular epithelium constitutes a small proportion of the tissue; the vast majority is stroma and adipose. Thus, it has previously been impossible to obtain a clear snapshot of gene or protein expression within normal or diseased tissue cell subpopulations.

To address the tissue-context problem, new technology is again coming to the rescue; creating 'tissue proteomics' as an exciting expanding discipline. Two major technologic approaches have been successfully used to sample macromolecules directly from subpopulations of human tissue cells. The first technology is laser capture microdissection. This is a technology for procuring specific tissue cell subpopulations under direct microscopic visualization of a standard stained frozen or fixed tissue section on a glass microscope slide. This technology was invented at the US National Institutes of Health and is commercially available through Arcturus Engineering (Mountain View, CA, USA; www.arctur.com). Tissue cells procured by laser capture microdissection have been used for highly sensitive and reproducible proteomic analysis using two-dimensional gels and other analytic methods [6–8].

A second major approach to isolate tissue cell subpopulations is affinity cell sorting of disaggregated cells from pieces of fresh tissue. A highly notable application of this technology in the field of breast physiology was recently reported [9] in a study resulting from a collaboration between Oxford Glycosciences (Oxford, UK) and the Ludwig Institute (London, UK). In that study the investigators separated and purified normal human breast luminal and myoepithelial tissue from reduction mammoplasty specimens using double antibody magnetic affinity cell sorting and Dynabead magnetic sedimentation (DynaL Inc, UK). After using enzymatic treatments and various incubation, separation, and washing steps, the investigators obtained purified luminal and myoepithelial cells in yields of 5×10^6 – 2×10^7 . Proteins from these cell populations

were then analyzed using two-dimensional gels. A master image for each cell type comprising a total of 1738 distinct proteins was derived. The investigators found 170 protein spots that were elevated twofold or more between the two populations. Of these, 51 were further characterized by tandem mass spectroscopy. The proteins preferential to the myoepithelial cells contained muscle-specific enzymes and structural proteins consistent with the contractile muscle-related derivation of these cell types.

Myoepithelial cells are a fascinating component of breast tissue. They are thought to play important roles in duct and lobule growth, matrix architecture, and remodeling after lactation and involution. A pathologic hallmark of early cancer progression from carcinoma *in situ* to invasive cancer is the loss or redistribution of myoepithelial cells. The conspicuous absence of myoepithelial cells in breast cancer progression could mean that these cells produce suppressor proteins that normally keep the malignant cells in check. Thus, one or more of the proteins identified in the study by Page *et al* [10] could be candidate cancer prevention molecules. The authors of that study concluded that 'These observations demonstrate that proteomics has the refinement and sensitivity to find proteins that are either uniquely or differentially expressed between different cell types, the consequences of which could enable new strategies for drug discovery.'

References

1. DeRisi J, Penland L, Brown P, *et al*: **Use of a cDNA microarray to analyse gene expression patterns in human cancer.** *Nature Genet* 1996, **14**:457–460.
2. Humphery-Smith I, Cordwell SJ, Blackstock WP: **Proteome research: complementarity and limitations with respect to the RNA and DNA worlds.** *Electrophoresis* 1997, **18**:1217–1242.
3. Hancock W, Appfel A, Chakel J, *et al*: **Integrated genomic/proteomic analysis.** *Anal Chem* 1999, **71**:743–748.
4. Williams K, Hochstrasser D: In: *Proteome Research: New Frontiers in Functional Genomics*. Edited by Wilkins M, Williams K, Appel R, Hochstrasser D. Berlin: Springer Press, 1997:1–12.
5. Gygi SP, Rist B, Gerber S, *et al*: **Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.** *Nature Biotechnol* 1999, **17**:994–999.
6. Emmert-Buck M, Bonner R, Smith P, *et al*: **Laser capture microdissection.** *Science* 1996, **274**:998–1001.
7. Banks RE, Dunn M, Forbes M, *et al*: **The potential use of laser capture microdissection to selectively obtain distinct populations of cells for proteomic analysis.** *Electrophoresis* 1999, **20**:689–700.
8. Emmert-Buck M, Paweletz C, Gillespie J, *et al*: **An approach to proteomic analysis of human tumors.** *Mol Carcinogen* (in press).
9. Paweletz C, Gillespie J, Ornstein D: **Rapid protein display profiling of cancer progression directly from human tissue using a protein biochip.** *Drug Dev Res* 2000 (in press).
10. Page MJ, Arness B, Townsend R, *et al*: **Proteomic definition of normal human luminal and myoepithelial breast cells purified from reduction mammoplasties.** *Proc Natl Acad Sci USA* 1999, **96**:12589–12594.

Author's affiliation: Lance A Liotta (Laboratory of Pathology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA) and Emanuel F Petricoin (Division of Therapeutic Proteins, CBER/Food and Drug Administration, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA)

Correspondence: Lance A Liotta, MD, PhD, Laboratory of Pathology, National Cancer Institute, National Institutes of Health, Bethesda, MA, USA. E-mail: lance@helix.nih.gov