Understanding Cancer Metastasis

An Urgent Need for Using Differential Gene Expression Analysis

Murali D. Bashyam, Ph.D.

Centre for DNA Fingerprinting and Diagnostics, Nacharam, India.

The author thanks the laboratories of Raju Chaganti, Ph.D., and Raju Kucherlapati, Ph.D., at the Memorial Sloan-Kettering Cancer Center, New York, New York and the Albert Einstein College of Medicine, New York, New York, respectively, for training in the field of microarray technology. The author thanks Gayatri Ramakrishna, Ph.D., and Aruna Ramachandran, Ph.D., for critical reading of the article.

Address for reprints: Murali D Bashyam, Ph.D., Staff Scientist and Chief, Molecular Oncology, Centre for DNA Fingerprinting and Diagnostics, Nacharam, Hyderabad-500076, India; Fax: (011) 91-40-7155610; E-mail: bashyam@www.cdfd.org.in, bashyam69@lycos.com

Received August 23, 2001; revision received October 22, 2001; accepted October 24, 2001.

Cancer is a multistep process and occurs as a result of the loss of control of cell division, leading to the initial tumor formation, which is followed by metastatic spread. Recent years have witnessed a vast improvement in the understanding of the molecular mechanisms regulating cell division and their links to tumorigenesis. The process of metastasis involves an intricate interplay between cell adhesion, proteolysis, migration, and angiogenesis. However, there is little knowledge of how these events are coordinately regulated in the tumor cell. Given that the uncontrolled spread of the tumor to distant organs is usually lethal, a study of the molecular mechanisms regulating metastasis assumes great significance. Recently, several technologies have been developed for analyzing differential gene expression. The current review discusses the importance of these technologies in the molecular analyses of metastasis. Cancer 2002;94:1821–9.

© 2002 American Cancer Society.
DOI 10.1002/cncr.10362

KEYWORDS: metastasis, metalloproteinases, adhesion, microarray.

The method by which cells increase in number is similar in all somatic cells. Structural changes during different stages of the cell cycle have been known for a long time, and the molecular basis of these changes has recently been worked out to a large extent (Fig. 1). Sometimes the delicate control over cell growth is lost, leading to the complex process of tumorigenesis. It is well established that tumors arise due to defect(s) in the genetic constitution of a cell. Cancer genes are broadly classified into two categories: 1) oncogenes which have a positive effect on growth and proliferation, and 2) tumor suppressor genes or anti-oncogenes that have a negative effect. More than a hundred oncogenes have been reported, including growth factors and their receptors, signal transducers, membrane bound or cytoplasmic kinases, and transcription factors. The tumor suppressor gene retinoblastoma (Rb) codes for a nuclear phosphoprotein and is a target for phosphorylation by the cyclin dependent kinases. The protein gets progressively hyperphosphorylated during G1 and provides the signal for transition from the G1 phase to the S phase (Fig. 1). Another important tumor suppressor gene, p53, is also a nuclear phosphoprotein. Mutation in p53 is the single most frequent event in human cancers. The multi-functional p53 protein is a transcription factor that is induced by DNA damage. It induces the transcription of several genes including p21, GADD45, and bax, as well as others. The p21 product decreases the activity of cyclin dependent kinases, GADD45 is involved in DNA repair, and bax is a member of the bcl-2 family that regulates apoptosis. It can be said that in response to DNA
damage, the p53 protein is able either to block the cell cycle or to induce apoptosis. Besides these, there are several other tumor suppressor genes that have been identified (please see http://web.indstate.edu/thcme/mwking/tumor-suppressors.html for a comprehensive list).

Metastasis

Invasion is the process whereby tumor cells migrate into the surrounding stroma, thereby becoming free to disseminate to distant organs via lymphatic or vascular channels. Metastasis is the spread of tumor cells from the primary site to form secondary tumors at other sites in the body. Tumors pose less risk if they are localized (benign). However, a benign tumor may be harmful if it secretes hormones or biologically active substances or if its sheer bulk interferes with a vital function (e.g., a brain tumor). Benign tumors are usually treated with a fair degree of success by surgery or radiotherapy, with limited side effects. However, it is generally extremely difficult to treat a malignant tumor that has spread to distant organs. The most common treatment is chemotherapy, which has a low success rate and severe side effects.

Pre-malignant lesions arise due to the accumulation of mutations in genes that regulate growth. Subsequently, further genetic alterations arise in one or a few cells in the lesion, leading to malignancy. The population of cells in a tumor may therefore be heterogeneous and could consist of phenotypically distinct cells that are at different stages of malignancy. Only a few cells may expand to yield the malignant tumor. Invasion and metastasis are a dynamic multi-step process.

Steps in the invasive process

1. Loss of contact inhibition of growth
2. Tumor vascularization
3. Detachment of cells from the primary tumor
4. Penetration of the basement membrane
5. Migration through the stroma

Steps in the metastatic process

1. Penetration of lymphatics and vasculature
2. Escape from immune surveillance; survival in circulation
3. Arrest in capillary beds of distant organs
4. Penetration of lymphatic or blood vessel wall; extravasation into tissue
5. Growth of metastatic deposit in the new location
6. Tumor vascularization

As a first step, the potentially metastatic tumor cell loses its dependence on contact with identical cells for growth. E-cadherin is the main adhesion molecule of epithelia and has been termed “molecular glue.” Cell-cell adhesion brought about by E-cadherin transmits anti-growth signals via the cytoplasmic protein β-catenin and transcription factors belonging to the Lef/Tcf family. The E-cadherin function is frequently lost in many cancers by either inactivation of E-cadherin or the inactivation of β-catenin. Re-establishment of the functional cadherin molecule in tumor cell lines and in transgenic carcinogenesis mouse models results in a reversion from an invasive to a benign phenotype. An interesting link between loss of control over cell division and metastasis was revealed by Batsche and colleagues, who showed that Rb and c-myc can activate the expression of the E-cadherin gene in epithelial cells through interaction with the transcription factor AP-2.

Interactions between tumor and stromal cells control the two protease systems that are responsible for most of the proteolysis outside the cell: the urokinase plasminogen activator (uPA)/uPA receptor/plasminogen network, and the matrix metalloproteinases (MMPs). The complex interaction between the stromal cell and the tumor cell is illustrated in Figure 2. The MMPs are a family of about 17 zinc-dependent
endopeptidases, including collagenases, gelatinases, and stromelysins, and are collectively capable of degrading essentially all extracellular matrix (ECM) components. The tissue inhibitors of metalloproteinases inhibit the activity of the MMPs. The expression of MMPs is regulated in a paracrine manner by growth factors and cytokines secreted by tumor-infiltrating lymphocytes, stromal cells, and, in rare instances, by the tumor cells themselves. Recent studies have highlighted the importance of MMPs in tumor spread. Stromal MMP-2 and uPA are produced as inactive precursors and are subsequently activated on the surface of tumor cells (Fig. 2), thus allowing malignant cells to breach basement membranes. Transgenic mice expressing the stromelysin 1 gene in mammary gland epithelial cells developed malignant tumors in the mammary gland.

A link between cell adhesion and matrix degrading enzymes has also been established. Overproduction of MMP-3 leads to the cleavage of E-cadherin and thus a breakdown of the interactions between epithelial cells. It has been shown that the loss of E-cadherin function allows transcriptional activation of MMP-7.

An essential requirement for successful metastasis is the growth of new blood vessels (angiogenesis) at the tumor site. Tumor cells are able to: 1) induce angiogenesis through expression of vascular endothelial growth factor and members of the fibroblast growth factor family, and 2) suppress expression of anti-angiogenic factors such as thrombospondin. The establishment of secondary tumors in distant organs is achieved by changes in the expression of integrin genes to facilitate binding to the changed stromal and matrix microenvironment. Recently, it was shown that the disruption of matrix metalloproteinase 2 binding to integrin αvβ3 could inhibit tumor growth. The αvβ3 integrin on the melanoma cell surface interacts with the lymph node matrix protein vitronectin to induce the secretion of MMP-2, resulting in an enhancement in cellular invasion by the melanoma cells.

While cellular invasion depends on adhesion, migration, and proteolytic activity, we are only just beginning to understand how these events are coordinated. Although there may be a preferred site of metastasis for a particular tumor, it is generally difficult to predict the location of metastatic spread for a specific tumor. Carcinomas (cancers of the epithelial or endothelial cells) mainly metastasize to lymph nodes, unlike sarcomas (cancers of the mesenchyme). Metastasis in lung, liver, and bone is common, as they have small blood vessels in which tumor cells may get trapped. However, muscles and the spleen also have small blood vessels but are rarely sites of tumor deposits. Loss of heterozygosity studies on various advanced cancers have revealed that the profiles of allelotypes are different among tumors of different origins, indicating the presence of multiple pathways for tumor spread. It has been suggested that while the primary tumor may result from mutations in the growth control genes, metastasis probably results mainly from changes in gene expression patterns in the cell. Thus it is imperative to use new technologies for the identification and characterization of such genes in order to understand metastasis and design better therapeutic regimes to combat it.

A Possible Solution: Analysis of Differential Gene Expression

Since tumor spread involves the coordinated regulation of different cellular processes, it is evident that we need to analyze global gene expression changes in order to understand this complex phenomenon. Several techniques have been used for the detection and quantification of differentially expressed genes. These include Northern blotting, S1 nuclease protection, subtractive hybridization, differential display reverse transcriptase polymerase chain reaction (DDRT-PCR), RNA arbitrarily primed-polymerase chain reaction
(RAP-PCR), cDNA-amplified fragment length polymorphism (AFLP), sequencing of cDNA libraries, serial analysis of gene expression (SAGE), suppression subtractive hybridization (SSH), and others. Representational difference analysis (RDA) has been used to detect differences between normal and tumor genomes. Other techniques, like rapid isolation of cDNA by hybridization\(^{11}\) and the sensitive and specific exon trapping system,\(^{12}\) have been used to determine transcribed regions in large chunks of DNA in genomic libraries.

Northern blotting and S1 nuclease protection are useful for the study of specific genes, but are not suitable for analyzing expression levels of several genes at a time. One of the earliest techniques for analyzing differential gene expression, subtractive hybridization, is not efficient for analyzing low abundance transcripts. Representational difference analysis\(^{13}\) and SSH\(^{14}\) have alleviated some of the lacunae of subtractive hybridization. In melanoma cells, RDA was used to identify the differentially expressed gene LAGE-1.\(^{15}\) Suppression subtractive hybridization has been used to identify a melanoma differentiation associated gene (mda-7).\(^{16}\) In addition, SSH was used to contrast differential gene expression profiles in immortalized, non-tumorigenic rat embryo fibroblasts and in H-Ras-transformed cells.\(^{17}\)

Serial analysis of gene expression was developed by Velculescu and colleagues and allows the simultaneous and quantitative analysis of a large number of transcripts.\(^{18}\) Recently SAGE was used to build the transcriptome map of the human genome.\(^{19}\) In addition, SAGE has been used to systematically analyze transcripts present in nonsmall cell lung cancer. The transcripts for PGP 9.5, B-myb, and human mutT were abundantly expressed in primary lung carcinomas.\(^{20}\) Lal et al. have created a public database, SAGEmap, to differentially quantify transcripts in glioblastomas and normal brain tissues.\(^{21}\) The technique of differential display has perhaps been the most frequently employed technique for analysis of differential gene expression in cancers.\(^{22}\) Differential display involves reverse transcription of mRNA from the samples to be compared using oligo dT primers, followed by PCR amplification using arbitrary primers. Su et al. identified the prostate tumor-inducing gene 1 that is also differentially expressed in breast, colon, and small cell lung carcinoma cell lines.\(^{23}\) Differential cDNA display was also used to identify the metastasis suppressor gene melastatin that was not detected in melanoma metastases.\(^{24}\) Zhang et al. identified the oncogenic ras-specific gene mob-5 through differential display.\(^{25}\)

**Microarray Technology**

Microarray technology is gaining popularity among scientists worldwide as the most favored technique for global differential gene expression analysis. DNA microarrays are arrays of specific DNA sequences immobilized on a solid surface as spots that are used to identify or quantitate several specific DNA/RNA sequences in a heterogeneous pool of nucleic acid samples. These arrays are widely used for genetic mapping, identification of mutations, and monitoring global gene expression changes. In principle, microarrays are based on the standard hybridization technique pioneered by Ed Southern in 1975 and utilize the quintessential property of a DNA or RNA strand to hybridize only to its complementary strand.\(^{26}\)

Currently there are two different kinds of DNA array technologies being extensively used for gene expression and genotypic analysis. One has been pioneered and patented by Affymetrix Inc. and involves the direct synthesis of short oligonucleotides at a very high density on a glass surface using photolithography. The strength of the technique is primarily based on the fact that each gene is represented by several oligonucleotides complementary to different regions of the gene. Further, each oligonucleotide is paired with a sister oligonucleotide with one base mismatch in order to account for background and cross/mismatch hybridization. Each oligonucleotide is selected in silico based on empirically derived, composition dependent design rules. RNA samples to be studied are reverse transcribed and then again transcribed in vitro by using the phage T7 RNA polymerase in the presence of biotin labeled primers, thus generating large amounts of biotin-labeled RNA. The labeled RNA is then allowed to hybridize to the DNA arrays and is detected by a laser confocal fluorescent scanner. Affymetrix also provides arrays for the detection of single nucleotide polymorphisms in the human genome, thus taking an important step toward identification of genotypic variations responsible for specific phenotypes in normal and disease states. A distinct disadvantage in this system is the absolute necessity of knowing the sequence of the gene to be investigated. Moreover, the test and reference samples are processed separately, and although several controls are incorporated in the experimental design, one cannot completely rule out the effect of experimental variations during separate treatments.

An alternative approach to this technology was developed around the same time by Mark Schena and Patrick O. Brown,\(^{27,28}\) termed cDNA microarray technology. This approach uses PCR amplification of
cloned cDNAs (instead of oligonucleotides) that are robotically printed onto a microscopic glass slide, with each PCR product representing a gene. Although it is not possible to achieve as high a density as in the Affymetrix gene chips, cDNA arrays do not require prior knowledge of the sequence of the gene to be arrayed. The technology involves differential fluorescent labeling of RNA samples isolated from two or more different states of the tissue or cell that one wishes to compare (e.g., diseased vs. normal; activated vs. steady state; pathogenic vs. non-pathogenic, resistant vs. sensitive, etc.). The differentially labeled cDNAs are hybridized to the microarray slide. An image of the intensities of each spot is acquired by using a confocal laser fluorescence scanner, and the ratio of the two samples for each spot is deduced, providing an estimate of the changes in expression levels for each gene.

It is important to highlight how microarray analysis scores over all the other technologies mentioned above. First, separate cloning and DNA sequencing is not required to determine the identity of the differentially expressed transcript. Second, the test and reference samples are processed together, thereby ruling out experimental errors from separate handling. Third, microarray analysis can compare more than two samples in a single experiment. Fourth, information regarding variations in expression levels of thousands of genes can be obtained with relative ease in a single experiment.

There are four main steps involved in this technique:

Making the array
The microarray is generated by PCR amplification of all available genes followed by printing on a glass slide. Owing to the presence of interrupted genes in eukaryotes, cDNA libraries are used as templates for generating PCR amplified DNA samples for arraying. Each clone is PCR amplified and purified before spotting. Two kinds of coated slides are available: polylysine-coated or amino silane-coated. The former is generally more sensitive and consequently gives more background fluorescence. The arrayer is a robot fitted with contact or noncontact printing pins. The robotic spotter/arrayer is usually mounted on a flat vibration-free workstation table (Fig. 3). The spotter essentially draws about 250-500 nL of 100-500 ng/μL PCR-generated DNA samples into each pen from 96/384 well plates and deposits a few nLs on a glass slide. Spot diameters range from 50 to 250 μm with about 200-250 μm separating the center of adjacent spots. The precision is about 10 μm.

Essential features of a spotter include:
1. A computer controlled, high precision, high speed, three-axis robot
2. A unique pen tip assembly
3. Humidity and temperature control for better spot uniformity

Preparation of the target for analysis of gene expression
For gene expression analysis, one can start with total RNA or purified mRNA. Differential labeling is carried
out by direct incorporation of fluorescent dyes during reverse transcription of the RNA samples. The most popular dyes are cyanine 3 (cye-3)-dUTP and cyanine 5 (cye-5)-dUTP. They have good incorporation efficiency with reverse transcriptase, good photostability, and widely separated emission spectra. Several kits are now available as alternatives to direct labeling, which increase the efficiency of labeling (3DNA™ technology, Genisphere Inc., NJ) and lower the amount of starter RNA to 1-2 μg. Following differential labeling, test and reference samples are mixed and hybridized to the slide containing the printed array.

**Hybridization**

Hybridization solutions, both with and without formamide at hybridization temperatures of 42 °C and 65 °C respectively, have been successfully used. Standard blocking solutions are used, including Denhardt’s reagent, sodium dodecyl sulfate, salmon sperm DNA, etc. Hybridization is followed by washing procedures.

**Image analysis**

A sophisticated laser fluorescence microscope is used to analyze the slides. The reactive substrate is scanned at the appropriate wavelength(s), and the ratio of the fluorescence emission of the different wavelengths provides a measure of differential gene expression. Software packages are available to calculate the exact ratios between the intensities of the two colors while subtracting background noise and correcting for different efficiencies of detection of the two colors. Some essential features of a scanner are:

1. Simultaneous dual laser scanning (for cye3 and cye5 dyes)
2. User-selectable laser power and focus position
3. Precise control of the scan area
4. Automatic reading of barcodes
5. Pre-set hardware controls
6. Automatic calculation and subtraction of local background
7. High resolution acquisition
8. Integrated hardware and software.

**Data management**

This involves statistical and mathematical approaches to constructing algorithms to compare the variations in expression levels of particular genes under different conditions. High-density arrays generate tremendous amounts of data, and software tools are required to sort and cluster this vast information and to compare it with other sources of information. Some useful websites describing important features of the microarray technology are listed in Appendix I.

Despite its recent introduction in the mid-90s, microarray technology has already found a wide application in the analysis of gene expression in various types of cancers. Gene expression patterns have been compared between a human epithelial cell line and a breast carcinoma cell line. Oligonucleotide arrays were used to compare gene expression profiles between normal and tumor colon tissues. Gene expression patterns between different cancer cell lines have also been compared. These data were then linked to their response to specific drugs. An array-based comparative genomic hybridization assay has been developed to determine DNA copy number changes across the whole genome in breast carcinoma cell lines. Gene expression in normal and tumor endothelial cells was compared, and the results revealed elevated levels of several transcripts in the tumor endothelium. Many of these transcripts belonged to metalloproteinases, genes involved in wound healing and in corpus luteum formation. Molecular portraits of different surgical breast carcinoma specimens were determined by microarray analysis. A modification of the standard in situ hybridization technique (or tissue microarrays [TMAs]) has enabled the analysis of hundreds of tumor samples together. Separate minute cylindrical sections are taken from several paraffin blocks (each a different tumor sample; donors) and transferred onto a new paraffin block (the recipient). This new block now has representative tissue sections from several different tumor samples and can be processed for histologic staining/gene expression analysis/protein analysis. Moreover, several replicate blocks can be made. A single TMA experiment can yield information about the molecular characteristics of up to 1000 specimens at once, whereas in conventional microarray analysis a single slide yields information about only a single tissue sample.

The obvious approach to identifying genes responsible for metastasis in various cancers is to separate malignant cells from non/pre-malignant cells in a tumor biopsy sample. Once suitable markers are identified, it should be possible to separate such cells based on fluorescence activated cell sorting or laser capture microdissection (LCM). In LCM, a thin polymer is first placed in direct contact with a frozen or fixed tissue section under a microscope. A laser activates the polymer, resulting in the transfer of specific cells from the tissue section onto the polymer film. Nucleic acids can be extracted from these isolated cells and used for the desired analysis. The only drawback is the limiting starting material for RNA/
DNA isolation. The ability to separate high and low metastatic clones from a cancer cell line has revealed a few genes that have been shown to induce or suppress the process of metastasis. For example, the \textit{nm23}\textsuperscript{40} and \textit{Elm1}\textsuperscript{41} genes have been shown to suppress metastasis, whereas the \textit{mta1} gene has the opposite effect.\textsuperscript{42} The \textit{RhoC} gene, a small GTPase, was identified as an important player in the metastatic spread of melanoma cells by using an in vivo selection scheme.\textsuperscript{43} However, analyses would be improved if we could identify suitable markers in the tumor samples, rather than resorting to analyzing cell lines, which may not give a true picture of the tumor.

Although analysis of mRNA profiles is extremely important, it is obvious that steps that follow transcription (like capping, polyadenylation, splicing, mRNA turnover, efficiency of translation, post translational modifications, etc.) may also result in differences in protein levels. Therefore, during the past few years, cancer research has focussed on proteomics, the direct analysis of protein profiles. A proteomics study usually involves a two-dimensional electrophoresis comparison of protein profiles, and the identity of protein bands exhibiting differences is determined by mass spectrometric analysis.\textsuperscript{44} Paweletz et al. have recently employed a reverse phase protein microarray to analyze prostate carcinoma samples belonging to different stages of progressing tumors.\textsuperscript{45}

**SUMMARY**

The process of invasion and metastasis involves changes in the expression of genes encoding molecules involved in cell adhesion and migration and in the production of proteinases that degrade the barriers surrounding the tumor. The interaction of the spreading tumor cells with the vasculature and the immune system is also likely to be altered. Several questions relating to the process of metastasis remain to be answered. For example, is the activation of genes required for various steps in metastasis regulated in an ordered cascade, or are they random, independent events? How are the genes for homing of the cancer cells into specific organs activated? Is this event linked to earlier events like the ability to grow independent of anchorage and contact with similar cell types, or is it independent of these processes? How is the process of angiogenesis regulated? Are the metalloproteinase and integrin genes coordinately regulated? Recent advances in analyzing global gene expression have now made it possible to identify various genes that are activated (or repressed) during the transition of the primary tumor to a malignant phenotype. This should help in achieving the ultimate goal of identifying the cascade of molecular events that define the generation and maintenance of the cancer cell. It would also help in identifying markers on the tumor cell that could predict the prognosis of the disease. The work of Golub et al.\textsuperscript{46} and others\textsuperscript{47} has already shown that tumors can be classified on the basis of an RNA fingerprint by using microarray technology. Microarray technology can now be used to type tumors on the basis of the extent of malignant spread. Once a large number of tumors have been studied, it should be possible to predict, at a very early stage, whether a tumor is malignant, and if so, to what extent. Therefore, differential gene expression profiles may not only provide answers to the long unsolved question of the molecular basis for metastasis, but will also help in classifying tumors based on the extent (and maybe specific location) of spread. In the long run, microarray technology could prove an indispensable aide to the clinician in deciding the correct treatment regime.

**APPENDIX I**

Build your own arrayer and scanner:

http://cmgm.stanford.edu/pbrown/mguide
http://sequence.aecom.yu.edu/bioinf/microarray/printer.html

Arrayer and scanner descriptions:

**Institutional**

http://www.nhgri.nih.gov/DIR/LCG/15K/HTML
http://sequence.aecom.yu.edu/bioinf/funcgenomic/html
http://w95vcl.neuro.chop.edu/vcheung

**Commercial**

http://www.genomicsolutions.com
http://www.apbiotech.com
http://www.packardbiochip.com

**Commercial arrayers**

http://www.virtek.org
http://www.genemachines.com

**Commercial scanners**

http:// www.axon.com

**REFERENCES**


