Amgen Award Lecture

A Perspective on DNA Microarrays in Pathology Research and Practice

Jonathan R. Pollack

From the Department of Pathology, Stanford University,
Stanford, California

DNA microarray technology matured in the mid-1990s, and the past decade has witnessed a tremendous growth in its application. DNA microarrays have provided powerful tools for pathology researchers seeking to describe, classify, and understand human disease. There has also been great expectation that the technology would advance the practice of pathology. This review highlights some of the key contributions of DNA microarrays to experimental pathology, focusing in the area of cancer research. Also discussed are some of the current challenges in translating utility to clinical practice. (*Am J Pathol* 2007, 171:375–385; DOI: 10.2353/ajpath.2007.070342)

Over the past decade, DNA microarray technology has fundamentally transformed the way much of modern pathology research is performed. Although the resultant advances have so far made only modest impact in clinical pathology practice, there remain expectations of much more to follow. This review highlights some of the key contributions of DNA microarrays to pathology research, focusing in the area of oncology, and also discusses some of the remaining challenges in translating its utility to pathology practice. The review is not meant to be comprehensive but rather aims to provide insight and perspective from a clinical pathology-trained investigative pathologist working in the field of cancer genomics and with interests in translating microarray technology and its derived discoveries to improve the practice of pathology. Comprehensive reviews of DNA microarray methods and data analysis can be found elsewhere.^{1–5}

DNA Microarrays

DNA microarray technology emerged in the early 1990s, made possible by the convergence of two advances. First, large-scale DNA sequencing efforts, preceding the full-scale Human Genome Project and focused on the

expressed component of the genome, provided DNA sequence information and physical clones for thousands of human genes. Second, technical advances provided methods to manufacture slides or chips containing thousands of DNA probes arrayed within a small surface area, for example, less than 1 cm². Dr. Patrick Brown and colleagues⁶ at Stanford University developed cDNA microarrays based on spotting polymerase chain reaction (PCR)-amplified gene fragments onto glass microscope slides, whereas at Affymetrix (Santa Clara, CA), scientists used light masks to direct the *in situ* synthesis of DNA oligonucleotide probes on silicon wafers to produce GeneChips.^{7,8} The number of studies using DNA microarrays has risen markedly over the past few years (Figure 1A), and at present, many different types of “homemade” and commercial DNA microarrays are in use.

Although there were many envisioned uses for DNA microarrays, profiling gene expression became the predominant application in the early years. The power of the technology derives from the ability to measure, in the case of gene expression, mRNA levels across thousands of genes simultaneously (Figure 1B). The resultant expression profiles, likened to “molecular portraits,”⁹ provide the researcher and pathologist a new tool to observe, describe, and understand the molecular variation within tissue specimens. Microarray data analysis methods can be generally divided into supervised and unsupervised approaches. Supervised analyses make up-front use of specimen annotations, for example, to identify genes differentially expressed between tumor and normal, whereas unsupervised analyses (eg, hierarchical clustering¹⁰) seek to organize data agnostic to sample annotation, which is useful in discovering previously unrecognized sample classes.

Accepted for publication May 23, 2007.

The ASIP-Amgen Outstanding Investigator Award is given by the American Society for Investigative Pathology to recognize excellence in experimental pathology research. Jonathan R. Pollack, a recipient of the 2006 Amgen Outstanding Investigator Award, delivered a lecture entitled “Genomic Views of Human Cancer,” on April 2, 2006 at the annual meeting of the American Society for Investigative Pathology in San Francisco, CA.

Address reprint requests to Jonathan R. Pollack, Department of Pathology, Stanford University School of Medicine, CCSR-3245A, 269 Campus Dr., Stanford, CA 94305-5176. E-mail: pollack1@stanford.edu.

Figure 1. Microarray analysis. **A:** Growth of microarray studies over the past decade, as evidenced by the number of publications in PubMed and in seven top-ranked (by impact factor) general pathology journals, using the search term “microarray.” The graph is meant to depict a trend; numbers likely represent substantial underestimates because of limitations of the search. **B:** Schematic depiction of a two-color microarray-based expression profiling method. mRNA isolated from test and reference samples are differentially labeled using two different fluoros (shown as red and green, respectively) and then co-hybridized to a DNA microarray comprising an ordered array of gene-specific DNA probes (left). Labeled mRNAs bind their cognate probes on the microarray by Watson-Crick base pairing. Following hybridization and imaging (center), the ratio of red to green fluorescence for each gene spot on the array reflects that gene’s relative expression level in the test compared with reference sample. The *ERBB2* gene, shown as a red spot in the scanned image, is more highly expressed in the test sample. Analysis of many samples produces a colorimetric table (“heat map”) of gene expression ratios (right), where each column represents a different sample, and each row represents a different gene on the array. The columns and rows here have been ordered by unsupervised hierarchical cluster analysis¹⁰ to reveal patterns in the data, where the dendrogram (tree) branches indicate relationships among samples and among genes.

In profiling across many genes and specimens, the resultant high-dimensional data sets have also brought new statistical challenges.^{11,12} In large data sets, the particular expression patterns sought are often found but might not be statistically meaningful. Another concern has been “over-fitting” of data, where data models are developed and tested on the same patient cohort. An additional common shortcoming has been the use of insufficiently large cohorts in both discovery and validation phases. Indeed, in the early years some inappropriate statistical analyses likely contributed to inflated expectations of DNA microarray technology.

Cancer Classification

Due in part to the ready availability of human tumor specimens, often excised as part of standard patient treatment, as well as the impact of the disease, many early microarray studies focused on human cancer. Leukemia specimens, devoid of many of the stromal components present in solid tumors, further simplified analysis. A major goal of many DNA microarray studies has been cancer classification. The pathologist classifies cancer, for example, based on its anatomical site of origin and histopathology, sometimes using ancillary tests like immunohistochemistry or cytogenetics. Classification systems provide important information for prognostication and selection of therapies. By defining hitherto unrecognized molecular variation in gene expression, DNA microarrays might provide a means for improved cancer classification.

An early landmark study by Golub et al¹³ laid the computational foundations for applying DNA microarrays to the problem of cancer classification, both to predict known tumor classes and to validate new classes. Using supervised methods, the investigators identified genes

differentially expressed between two classes of leukemia, acute myelogenous leukemia (AML) and acute lymphoblastic leukemia. Statistically significant differences could be defined as those occurring above what was expected by chance, estimated by comparison to differences observed in the same data set but after first randomly permuting class labels (AML versus acute lymphoblastic leukemia). Further, the expression of genes distinguishing AML and acute lymphoblastic leukemia could be used to classify new cases with high accuracy and quantifiable predictive strengths. In a proof-of-principle unsupervised analysis, the investigators could also “rediscover” the known leukemic classes from the expression data. Although AML and acute lymphoblastic leukemia are of course readily distinguishable by existing cytochemical staining and flow cytometry techniques, the concepts developed, and many variations on the original computational methods, are applicable to more difficult classification problems.

The first microarray-based discovery of novel tumor classes was reported by Alizadeh et al.¹⁴ By unsupervised hierarchical cluster analysis of variably expressed genes in diffuse large B-cell lymphoma, the investigators identified two subclasses with distinct expression patterns. One pattern was similar to that of normal germinal center B cells, whereas the other was to that of activated B cells. The latter diffuse large B-cell lymphoma subtype was also associated with constitutive nuclear factor- κ B activity and a less favorable prognosis.^{14–15} Therefore, although indistinguishable by histology, expression profiling nonetheless suggested a refined classification of diffuse large B-cell lymphoma, which might improve outcome prediction and possibly selection of therapies. Indeed, *BCL6* gene expression, a surrogate indicator of the germinal center B cell-like subtype, has since been

Figure 2. Microarray analysis identifies clinically relevant prostate cancer subtypes. **A:** Hierarchical cluster analysis of gene expression patterns of normal and cancerous prostate specimens. The dendrogram indicates relationships among samples based on gene expression profiles; only a select subset of gene clusters is shown. Cluster analysis is seen to distinguish malignant from normal prostate (pink branches). Note that *AMACR* (open arrow) is among the genes more highly expressed in prostate cancer. Cluster analysis also defines three subtypes of prostate cancer (numbered above) not distinguishable histologically. *AZGP1* and *MUC1* (closed arrows) are located within gene expression patterns that characterize subtype-1 and subtypes-2/3, respectively. **B:** Prostate cancer gene expression subtypes are prognostically relevant. IHC staining (tissue microarray, left) of *AZGP1* and *MUC1*, surrogate markers for subtype-1 and subtype-2/3 expression patterns, predict lower and higher tumor recurrence rates, respectively, independent of tumor stage, Gleason grade, and preoperative serum prostate-specific antigen.¹⁸ Findings identify subtype-1 as a clinically favorable prostate cancer subclass.

shown to predict survival independent of the currently used International Prognostic Index score.¹⁶

Soon thereafter, microarray analysis of breast cancer also identified multiple tumor subclasses, refining

the existing classification.^{9,17} Estrogen receptor (ER)-negative tumors included those with *ERBB2* amplification as well as a previously underappreciated subclass with basal epithelial markers and poor prognosis. ER-

positive breast tumors could be subdivided into luminal A and B subtypes, with the former associated with more favorable outcome. Likewise, in prostate cancer, our own microarray studies have defined three clinically relevant tumor subtypes indistinguishable by histology (Figure 2A).¹⁸ Surrogate immunohistochemical

markers (AZGP1 and MUC1) for these subtypes are predictive of tumor recurrence, independent of tumor stage, Gleason grade, and serum prostate-specific antigen levels (Figure 2B). Similar microarray studies have identified tumor subclasses within other tumor types as well.¹⁹

Outcome Prediction

Microarray analysis has also been applied directly to define gene signatures for prognostication and for prediction of response to therapies. In a landmark study, van't Veer et al²⁰ compared tumor gene expression profiles between two groups of patients with surgically excised lymph node-negative breast cancer, those who did or did not develop distant metastases within 5 years of follow-up. Supervised analysis based on the group distinction defined a 70-gene signature that could predict disease-free and overall survival in two independent cohorts of breast cancer patients,^{20,21} outperforming current prognostic indices based on clinical and histological parameters such as the St. Galen and National Institutes of Health consensus criteria (but see Ref. 22). The poor-prognosis signature might therefore improve the selection of patients who would benefit from adjuvant therapy.

Our own microarray studies of AML have defined a 133-gene signature that predicts overall survival independent of cytogenetics, itself a strong prognosticator.²³ This signature, recently validated in another AML patient cohort,²⁴ may prove particularly applicable for selecting appropriate risk-adapted therapy within the large subset of AML cases with no karyotypic abnormality. Gene signatures have also been reported that define responses to specific therapies, for example, to rituximab (an anti-CD20 antibody) treatment for patients with follicular lymphoma.²⁵

Insights into Cancer Pathogenesis

In addition to the discovery of previously unrecognized tumor subclasses, expression profiling has made many other significant contributions to our understanding of cancer biology. For example, Ramaswamy et al²⁶ explored genes differentially expressed between primary and metastatic tumors across a spectrum of solid tumor types. The investigators defined a 17-gene signature of metastasis that, surprisingly, was also expressed in a subset of primary tumors. The signature, which included genes with expected expression mainly in the stromal compartment, was predictive of patient outcome across several tumor types. It is noteworthy that this study challenged the existing paradigm that metastases arise from rare cells in the primary tumor that have acquired additional genetic alterations, suggesting rather that the propensity to metastasize is determined early in tumor pro-

gression and characterizes the bulk population of tumor cells. The findings also underscored the contribution of tumor stroma to cancer progression.

Another compelling contribution of microarray analysis was the recent discovery of recurrent gene fusions in prostate cancer. By analyzing "outlier" values of gene expression in prostate tumor microarray data sets, Tomlins et al²⁷ identified the ETS family of oncogenic transcriptional factors *ERG* and *ETV1* to be highly expressed in a subset of prostate tumors. Further characterization revealed chromosome rearrangement and gene fusion, resulting in the promoter of the prostate-expressed gene *TMPRSS2* driving androgen-regulated overexpression of *ERG* or *ETV1*. This finding not only provides novel insight into prostate tumorigenesis but also challenges the long-standing assumption that recurrent chromosomal alterations, frequent in hematogenous and mesenchymal malignancies, are rare in common epithelial tumor types.

In both of the aforementioned examples, it is worth noting that these discoveries resulted from exploratory rather than "hypothesis-driven" investigations. Such non-hypothesis-driven research, sometimes derogatorily labeled as "fishing expeditions," had initially received less favorable enthusiasm among grant-funding agencies, although many agencies now recognize its value. In addition, both of the above studies benefited from the public availability of clinically annotated microarray data, increasingly but not universally a requirement for peer-reviewed publication.

Array CGH and Integrative Genomics Analysis

Whereas DNA microarrays were first used widely to profile gene expression, other applications soon emerged. For example, in array-based comparative genomic hybridization (array CGH),^{28–30} tumor and normal genomic DNA are differentially labeled and compared by hybridization to microarrays comprising DNA probes of defined human genome map position, such as large genomic clones (eg, bacterial artificial chromosomes), gene fragments (cDNAs), or oligonucleotides (Figure 3A). The resultant tumor/normal ratios can be mapped onto the genome sequence to reveal in tumor genomes DNA amplifications and deletions, which drive the altered expression of cancer genes. Because genomic DNA comprises a more complex mixture of DNA sequences com-

Figure 3. Array CGH and integrative microarray analysis. **A:** Schematic depiction of array-based comparative genomic hybridization (array CGH) method. Genomic DNA (gDNA) isolated from tumor and normal samples is differentially labeled (shown as red and green fluor, respectively) and then co-hybridized to a microarray comprising DNA probes of known chromosome location. Following hybridization and imaging, the ratio of red to green fluorescence for each DNA spot on the array reflects that gene's relative copy number in the tumor genome. The *MYC* gene, shown as a red spot in the scanned image, is amplified in the tumor genome. Plotting fluorescence ratios by genome map position is useful in defining DNA amplifications and deletions. Illustrative data are shown (right) for chromosome 8 of the breast cancer cell line SKBR3. Peaks reflect DNA amplification (including an amplicon harboring *MYC*, arrow), and valleys DNA deletion. **B:** Integrative analysis of array CGH and gene expression data. Shown are heat map representations of DNA copy number (left) and gene expression (right) for the subset of genes identified by microarray to be both highly amplified and overexpressed in breast cancer cell lines or tumors. Samples are ordered identically in both panels; genes are ordered by chromosome location. This subset of amplified overexpressed genes is a rich source for breast cancer gene discovery; previously known bona fide and putative oncogenes are highlighted by red text and include *MYC*, *CCND1*, and *ERBB2*. Data abstracted from Ref. 35. **C:** Integrative analysis reveals distinct classes of DNA copy number alteration associated with the different gene expression subtypes of breast cancer. Box plots indicate 25th, 50th (median), and 75th percentiles of genome fraction exhibiting chromosome segment gain/loss (above) or high-level DNA amplification (below), for tumors stratified by expression subtype. Basal-like tumors exhibit significantly higher numbers of gain/loss, whereas luminal-B tumors display more high-level DNA amplification. Findings suggest that subtypes arise by different underlying mechanisms of genomic instability. Data abstracted from Ref. 36.

Table 1. Diverse Applications of Microarrays

Application	Arrayed probes	Selected citations
Gene expression	DNA	6, 8
DNA sequence variation	DNA	
SNP discovery and genotyping		67
Ultrahigh throughput DNA sequencing		68
DNA copy number variation (array CGH)	DNA	29, 30
Loss of heterozygosity	DNA	39
DNA methylation	DNA	41, 69
Transcription factors/chromatin (ChIP on Chip)	DNA	70, 71
Gene regulatory elements (DNase-chip)	DNA	72
Gene function	DNA	
Reverse transfection		42
Barcode shRNA screens		73
Protein levels	Antibodies	74
Antibody levels	Antigens	75
Protein interactions	Proteins	76
Cellular expression	Cell lysates	77
Tissue expression	Tissues	45

pared with the subset of expressed genes, array CGH presents additional technical challenges compared with expression profiling. Nevertheless, robust protocols have been developed, and array CGH analyses have pinpointed new cancer genes, for example, *PPM1D* in breast cancer³¹ and *MITF* in melanoma.³² In addition, patterns of DNA copy number alteration, analogous to signatures of gene expression, have been proposed for cancer classification and outcome prediction.^{33,34}

Although expression profiling and array CGH each provides important information, integrating data from both of these methods can reveal additional insight. For example, although many genes exhibit elevated expression in cancer, the subset that is also highly amplified is enriched for key genes driving tumorigenesis (Figure 3B); such an integrative analysis is therefore valuable for cancer gene discovery. Our own integrative analysis of breast cancers has also uncovered a significant impact of aneuploidy (chromosome copy number imbalances) on gene expression patterns.³⁵ This finding, since observed in many tumor types, suggests the possibility that aneuploidy contributes to tumor progression through the altered expression of many genes (perhaps even hundreds), in turn affecting cancer phenotypes like metastatic potential or drug resistance. Integrative analysis has also revealed distinct patterns of DNA copy number alteration underlying the above-mentioned gene expression subtypes of breast cancer (Figure 3C).^{36,37} This finding suggests that breast cancer subtypes arise via different genetic pathways and with distinct underlying mechanisms of genomic instability.

Other Microarray Applications

The DNA microarray platform has been adapted to other applications as well (Table 1), such as measuring germ-

line genetic variation (the originally envisioned application), occurring as single nucleotide polymorphisms (SNPs) or copy number variations. Genome-wide association studies using SNP arrays have identified genetic loci conferring cancer risk.³⁸ In tumor genomes, SNP arrays have been used to map somatic alterations in gene dosage (like array CGH), as well as loss of heterozygosity, where genetic information might be lost in a copy-number neutral state.^{39,40} DNA microarrays have also been used to identify altered patterns of DNA methylation and chromatin in cancers,⁴¹ and have even been applied to characterize gene function by "reverse transfection" of spotted full-length genes (cloned into expression cassettes) into overlaying cultured cells.⁴²

Beyond arraying DNA probes, arrayed antibodies and antigens have been used to respectively quantify levels of tumor proteins⁴³ and the antitumor humoral response.⁴⁴ Cell lysates and even tissue specimens have also been arrayed. Tissue microarrays comprise cylindrical tissue cores from several hundred different tumor specimens sectioned onto a single glass slide and permit the measurement of a single gene's expression across many samples (rather than a single sample's expression across many genes, as for DNA microarrays) simultaneously by immunohistochemistry (IHC) or RNA *in situ* hybridization.⁴⁵ Tissue microarrays, which also provide information on cellular localization of expression, have markedly sped the evaluation and validation of initial DNA microarray discoveries across larger patient cohorts (Figure 2B).

Applications in Pathology Practice

There are many challenges in using DNA microarrays as a platform for clinical diagnosis (discussed below). Not surprisingly then, most current diagnostic applications resulting from microarray discoveries rely on methods already in common use in histopathology and molecular pathology laboratories, such as IHC or reverse transcription (RT)-PCR. For example, *AMACR* (α -methylacyl-CoA racemase) was identified by microarray analysis to be more highly expressed in prostate cancer compared with normal prostate^{46–48} (Figure 2A). IHC analysis of *AMACR* expression is now in routine use in some pathology centers to evaluate difficult biopsy cases for the presence of prostate cancer. Many other promising single-gene biomarkers discovered using DNA microarrays are under evaluation but not yet in routine use (Table 2).

Multigene tests derived from microarray data are also being evaluated, although few are in routine use. A 70-gene breast cancer prognostic signature, described above and assayed by DNA microarray analysis of freshly frozen specimens (Agendia's MammaPrint; Amsterdam, The Netherlands), recently received Food and Drug Administration clearance for the prediction of breast cancer recurrence for node-negative tumors.⁴⁹ Likewise, a 21-gene signature (16 cancer-related and five reference genes), derived in part from microarray studies and assayed by quantitative-RT-PCR using formalin-fixed, paraffin-embedded specimens (Genomic Health's Oncotype DX; Redwood City, CA), predicts the risk of

Table 2. Selected Candidate Biomarkers Identified by DNA Microarray Analysis

Gene	Cancer	Use [§]	Assay [#]	Citation
<i>AMACR</i>	Prostate	D	IHC	47, 48
<i>AURKA (STK15)</i>	Medulloblastoma	P	IHC	78
<i>AZGP1</i>	Prostate	P	IHC	18
<i>BCL6</i>	DLBCL	P	Q-RT-PCR	16
<i>CK17; CK5/6</i>	Breast	P	IHC	79, 80
<i>DOG1</i>	GI stromal tumor	D	RISH, IHC	81
<i>EZH2</i>	Prostate	P	IHC	82
<i>HOXB13:IL17BR*</i>	Breast	P	Q-RT-PCR	83
<i>HPN</i>	Prostate	P	IHC	84
<i>MN1</i>	AML	P	Q-RT-PCR	85
<i>MUC1</i>	Prostate	P	IHC	18
<i>NBS1</i>	Uveal melanoma	P	IHC	86
<i>PLA2G2A</i>	Gastric	P	Q-RT-PCR	87
<i>S100P</i>	Bladder	D	IHC	88
<i>SPP1 (OPN)</i>	Colon	P	IHC	89
<i>HNSCC</i>		P	ELISA	90
<i>TLE1</i>	Synovial sarcoma	D	IHC	91
<i>TMPRSS2-ERG</i>	Prostate	D, P	FISH, Q-RT-PCR	92, 93
<i>ZAP70</i>	CLL (IgV _H mutation)	P	Flow cytometry	94

DLBCL, diffuse large B-cell lymphoma; ELISA, enzyme-linked immunosorbent assay; GI, gastrointestinal; CLL, chronic lymphocytic leukemia.

*Ratio of two genes.

[§]D, diagnosis; P, prognosis.

[#]FISH, fluorescence *in situ* hybridization; Q-RT-PCR, quantitative RT-PCR; RISH, RNA *in situ* hybridization.

tumor recurrence in ER-positive, node-negative breast cancers.⁵⁰ Such tests, in conjunction with other clinical and laboratory information, might be used to select patients who are likely to benefit from adjuvant chemotherapy. Both of these multigene tests are currently being evaluated in prospective clinical trials. Interestingly, although these two tests share few genes in common, a recent study indicates a high concordance in predictions, suggesting that they are tracking similar tumor biology.⁵¹ Other multigene signatures have been described, several reporting on key biological features of tumors and with prognostic value across multiple tumor types (Table 3).

Current Challenges

Many challenges remain in adopting DNA microarrays as a commonplace platform for diagnostic testing. Early concerns with expression profiling centered on discordances of microarray findings among different laborato-

ries. Gene expression signatures ostensibly reporting on the same biological or clinical parameter often shared few genes in common. Such discrepancies are likely attributable in large part to differences in specimen cohorts, array platforms (and probes), protocols, and analysis methods. More recently, investigators have shown high reproducibility of findings when standard operating procedures are followed, and improved interplatform concordance with careful matching of probe annotations between platforms.^{52,53} In addition, further scrutiny reveals that the disparate gene signatures reported by different laboratories might nonetheless reflect the same underlying biology⁵⁴ or provide comparable clinical utility.⁵¹ Many of the early claims attributable to overfitting of data or to insufficiently large sample sets remain unsubstantiated, but more rigorous statistical analysis has led to an increased likelihood of validation.

Evaluating and validating microarray testing in the clinical laboratory is far from straightforward. Foremost, DNA microarrays are multianalyte tests, where tens, hundreds, or even thousands of individual probes each report on the expression of a different gene with differing performance. Individual gene probe performance characteristics include analytic sensitivity (limit of detection), dynamic range and linearity, specificity (minimizing cross-hybridization to other genes), precision, and accuracy. Together, multiple gene probes comprise a diagnostic gene signature with its own set of performance characteristics, like sensitivity (minimizing false negatives), specificity (minimizing false positives), reproducibility, and robustness. Much effort in testing laboratories is currently being directed to standardize operating procedures for specimen collection and processing, RNA isolation and labeling, and microarray hybridization, imaging, and data analysis. The development of appropriate hybridization controls and standards is in progress,⁵⁵

Table 3. Selected Signatures, Many with Potential Clinical Utility across Multiple Tumor Types

Signature	Prognostic utility	Citation
Metastasis (17 genes)	BR, MD, PR	26
Wound response (512 genes)	BR, GA, LU	95
Stem cell-like (11 genes)	BL, BR, GL, LE, LU, LY, MD, MS, PR	96
Stromal (786 genes)	BR	97
Oncogenic pathways (60 to 230 genes)	BR, LU, OV	98
Hypoxia (168 genes)	BR, OV	99
Chromosomal instability (25 genes)	BR, GL, LU, LY, MD, MS	100
Invasiveness (186 genes)	BR, LU, MD, PR	101

BL, bladder; BR, breast; GA, gastric; GL, glioma; LE, leukemia; LU, lung; LY, lymphoma; MD, medulloblastoma; MS, mesothelioma; OV, ovarian; PR, prostate.

as well as appropriate quality-control metrics to assess specimen characteristics and hybridization quality.

Complicating matters, microarray technology has been rapidly evolving, with changing microarray platform/version releases, protocols, and even gene annotations themselves. There is also still no consensus on the optimal specimen type, freshly frozen (for high-quality RNA) versus paraffin (for the convenience of standard pathology processing), and whole specimen versus microdissected (to enrich for tumor cells). Nor is there agreement as to whether microarray tests should assay only a focused set of diagnostic genes or alternatively a wider set of genes, the latter of which might provide additional information but perhaps also additional risks, like unanticipated diagnoses. Although the above discussion has focused on expression profiling, many of the same considerations are relevant to other microarray applications.

Future Directions

Recent technical and informatic advances promise new opportunities for DNA microarrays in pathology research and practice. Currently available microarrays permit expression analysis of transcript variants with alternative exons and of microRNAs, a recently discovered and expanding class of small RNAs that regulate gene expression. Microarray platforms also now support the typing of several hundred thousand SNPs for whole-genome dosage, loss of heterozygosity, and linkage/association studies. On the informatics side, and although not required for diagnostic utility, the biological interpretation of gene expression patterns has long been a rate-limiting step, typically requiring painstaking literature searches. Recent advances in interpreting gene signatures include Gene Ontology vocabularies,⁵⁶ pathway analysis,⁵⁷ and gene set enrichment analysis.⁵⁴ Additional insights will derive from integrating data across diverse microarray applications, such as measurements of DNA copy number, gene expression and protein levels, and from integrating data across species, for example, leveraging data from genetically tractable mouse models of human cancer. Informatics analysis will also be increasingly used to discover new connections between genes, disease states, and candidate therapies.⁵⁸

Both technical and informatic advances in microarray analysis are expected to continue. However, whereas state-of-the-art microarray technology continues to be costly, informatics is by comparison the great equalizer. Anyone anywhere in the world with a computer, an Internet connection, and some basic knowledge can access and mine large publicly available microarray data sets to obtain new biological and clinical insight. We can expect an increasing number of such studies in the years to follow, as well as many more meta-analyses combining the results of multiple studies. The availability of infrastructure to support microarray data access and analysis, for example, the Stanford Microarray Database⁵⁹ and Oncomine,⁶⁰ will further facilitate such studies.

In regard to clinical testing, as performance concerns are adequately addressed, we can expect microarrays to

be increasingly used as a platform for clinical diagnosis. Microarray testing will emerge for indications where i) microarrays provide additional information or outperform standard histopathological markers, ii) many genes provide more information than one or a few, iii) adequate performance characteristics are demonstrated, iv) testing impacts a patient management decision, v) there has been appropriate validation of clinical utility (ideally including prospective clinical trials), and vi) testing is cost-worthy. Although the path to clinical acceptance, Food and Drug Administration approval and reimbursement remains largely untrodden, regulatory authorities are aware of the need to meet the many challenges.⁶¹ Also uncertain are which microarray platforms will become preferred for testing (where high performance, ease of use, automation, and adaptability are desirable) and whether testing will be performed predominantly at many sites, eg, using kits, or alternatively at central labs.

Microarray-based applications likely to have clinical utility in the future include cancer classification and subtyping. For example, approximately 5% of tumors present as metastatic cancers of unknown primary.⁶² Gene expression signatures have potential utility in classifying a tumor's anatomical site of origin,^{63,64} which would be useful in selecting the optimal treatment regimen. Likewise, such an assay could be applicable to other diagnostically challenging cases, eg, primary lung cancer versus metastatic cancer to lung.

Microarray analysis should also prove useful in prognostication where improved outcome prediction impacts patient management, for example identifying the subset of prostate cancer patients who can be safely followed without therapeutic intervention (ie, "watchful waiting"). Microarray analysis will also define germline DNA sequence variants, as well as somatic changes and deregulated pathways in individual patient tumors, thereby informing the selection of new molecularly directed therapies to realize "personalized" medicine. Other possible applications of microarray analysis include monitoring treatment response and toxicity. Outside of oncology, microarrays should have significant impact in microbiology in the identification of pathogens⁶⁵ and in cytogenetics, where array CGH can reveal microdeletions associated with mental retardation and other developmental disabilities.⁶⁶ Analysis of SNPs, copy number variations, and soon whole-genome DNA sequences, should also improve assessments of an individual's disease risks.

However, although we can expect that microarrays will become useful ancillary tests for many specific applications, microarrays are unlikely in the foreseeable future to replace most existing tests or to render obsolete the trained pathologist. Histopathological analysis by the trained eye can render quick, accurate, and cost-effective diagnoses. Nevertheless, the future looks bright for microarray technology in both pathology research and practice.

References

1. Tefferi A, Bolander ME, Ansell SM, Wieben ED, Spelsberg TC: Primer on medical genomics. Part III: Microarray experiments and data analysis. *Mayo Clin Proc* 2002; 77:927-940

2. Butte A: The use and analysis of microarray data. *Nat Rev Drug Discov* 2002, 1:951-960
3. Elvidge G: Microarray expression technology: from start to finish. *Pharmacogenomics* 2006, 7:123-134
4. Quackenbush J: Microarray analysis and tumor classification. *N Engl J Med* 2006, 354:2463-2472
5. Lee NH, Saeed AI: Microarrays: an overview. *Methods Mol Biol* 2007, 353:265-300
6. Schena M, Shalon D, Davis RW, Brown PO: Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995, 270:467-470
7. Fodor SP, Rava RP, Huang XC, Pease AC, Holmes CP, Adams CL: Multiplexed biochemical assays with biological chips. *Nature* 1993, 364:555-556
8. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL: Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 1996, 14:1675-1680
9. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D: Molecular portraits of human breast tumours. *Nature* 2000, 406:747-752
10. Eisen MB, Spellman PT, Brown PO, Botstein D: Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998, 95:14863-14868
11. Ransohoff DF: Rules of evidence for cancer molecular-marker discovery and validation. *Nat Rev Cancer* 2004, 4:309-314
12. Allison DB, Cui X, Page GP, Sabripour M: Microarray data analysis: from disarray to consolidation and consensus. *Nat Rev Genet* 2006, 7:55-65
13. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999, 286:531-537
14. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000, 403:503-511
15. Davis RE, Brown KD, Siebenlist U, Staudt LM: Constitutive nuclear factor κ B activity is required for survival of activated B cell-like diffuse large B cell lymphoma cells. *J Exp Med* 2001, 194:1861-1874
16. Lossos IS, Jones CD, Warnke R, Natkunam Y, Kaizer H, Zehnder JL, Tibshirani R, Levy R: Expression of a single gene, BCL-6, strongly predicts survival in patients with diffuse large B-cell lymphoma. *Blood* 2001, 98:945-951
17. Sørli T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lønning P, Borresen-Dale AL: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 2001, 98:10869-10874
18. Lapointe J, Li C, Higgins JP, van de Rijn M, Bair E, Montgomery K, Ferrari M, Egevad L, Rayford W, Bergerheim U, Ekman P, DeMarzo AM, Tibshirani R, Botstein D, Brown PO, Brooks JD, Pollack JR: Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc Natl Acad Sci USA* 2004, 101:811-816
19. Chung CH, Bernard PS, Perou CM: Molecular portraits and the family tree of cancer. *Nat Genet* 2002, 32(Suppl):533-540
20. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH: Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002, 415:530-536
21. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernards R: A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002, 347:1999-2009
22. Edén P, Ritz C, Rose C, Ferno M, Peterson C: "Good Old" clinical markers have similar power in breast cancer prognosis as microarray gene expression profilers. *Eur J Cancer* 2004, 40:1837-1841
23. Bullinger L, Döhner K, Bair E, Frohling S, Schlenk RF, Tibshirani R, Döhner H, Pollack JR: Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N Engl J Med* 2004, 350:1605-1616
24. Radmacher MD, Marcucci G, Ruppert AS, Mrozek K, Whitman SP, Vardiman JW, Paschka P, Vukosavljevic T, Baldus CD, Kolitz JE, Caligiuri MA, Larson RA, Bloomfield CD: Independent confirmation of a prognostic gene-expression signature in adult acute myeloid leukemia with a normal karyotype: a Cancer and Leukemia Group B study. *Blood* 2006, 108:1677-1683
25. Bohlen SP, Troyanskaya OG, Alter O, Warnke R, Botstein D, Brown PO, Levy R: Variation in gene expression patterns in follicular lymphoma and the response to rituximab. *Proc Natl Acad Sci USA* 2003, 100:1926-1930
26. Ramaswamy S, Ross KN, Lander ES, Golub TR: A molecular signature of metastasis in primary solid tumors. *Nat Genet* 2003, 33:49-54
27. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM: Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* 2005, 310:644-648
28. Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Döhner H, Cremer T, Lichter P: Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* 1997, 20:399-407
29. Pinkel D, Segreaves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG: High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* 1998, 20:207-211
30. Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO: Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* 1999, 23:41-46
31. Li J, Yang Y, Peng Y, Austin RJ, van Eynhoven WG, Nguyen KC, Gabriele T, McCurrach ME, Marks JR, Hoey T, Lowe SW, Powers S: Oncogenic properties of PPM1D located within a breast cancer amplification epicenter at 17q23. *Nat Genet* 2002, 31:133-134
32. Garraway LA, Widlund HR, Rubin MA, Getz G, Berger AJ, Ramaswamy S, Beroukhi M, Milner DA, Granter SR, Du J, Lee C, Wagner SN, Li C, Golub TR, Rimm DL, Meyerson ML, Fisher DE, Sellers WR: Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* 2005, 436:117-122
33. Carrasco DR, Tonon G, Huang Y, Zhang Y, Sinha R, Feng B, Stewart JP, Zhan F, Khatri D, Protopopova M, Protopopov A, Sukhdeo K, Hanamura I, Stephens O, Barlogie B, Anderson KC, Chin L, Shaughnessy JD Jr, Brennan C, Depinho RA: High-resolution genomic profiles define distinct clinico-pathogenetic subgroups of multiple myeloma patients. *Cancer Cell* 2006, 9:313-325
34. Fridlyand J, Snijders AM, Ylstra B, Li H, Olshen A, Segreaves R, Dairkee S, Tokuyasu T, Ljung BM, Jain AN, McLennan J, Ziegler J, Chin K, Devries S, Feiler H, Gray JW, Waldman F, Pinkel D, Albertson DG: Breast tumor copy number aberration phenotypes and genomic instability. *BMC Cancer* 2006, 6:96
35. Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE, Tibshirani R, Botstein D, Borresen-Dale AL, Brown PO: Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci USA* 2002, 99:12963-12968
36. Bergamaschi A, Kim YH, Wang P, Sorlie T, Hernandez-Boussard T, Lonning PE, Tibshirani R, Borresen-Dale AL, Pollack JR: Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. *Genes Chromosomes Cancer* 2006, 45:1033-1040
37. Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, Lapuk A, Neve RM, Qian Z, Ryder T, Chen F, Feiler H, Tokuyasu T, Kingsley C, Dairkee S, Meng Z, Chew K, Pinkel D, Jain A, Ljung BM, Esserman L, Albertson DG, Waldman FM, Gray JW: Genomic and transcriptional aberrations linked to breast cancer pathophysiology. *Cancer Cell* 2006, 10:529-541

38. Freedman ML, Haiman CA, Patterson N, McDonald GJ, Tandon A, Waliszewska A, Penney K, Steen RG, Ardlie K, John EM, Oakley-Girvan I, Whittemore AS, Cooney KA, Ingles SA, Altshuler D, Henderson BE, Reich D: Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc Natl Acad Sci USA* 2006, 103:14068–14073
39. Mei R, Galipeau PC, Prass C, Berno A, Ghandour G, Patil N, Wolff RK, Chee MS, Reid BJ, Lockhart DJ: Genome-wide detection of allelic imbalance using human SNPs and high-density DNA arrays. *Genome Res* 2000, 10:1126–1137
40. Zhao X, Li C, Paez JG, Chin K, Jänne PA, Chen TH, Girard L, Minna J, Christiani D, Leo C, Gray JW, Sellers WR, Meyerson M: An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Res* 2004, 64:3060–3071
41. Yan PS, Chen CM, Shi H, Rahmatpanah F, Wei SH, Caldwell CW, Huang TH: Dissecting complex epigenetic alterations in breast cancer using CpG island microarrays. *Cancer Res* 2001, 61:8375–8380
42. Ziauddin J, Sabatini DM: Microarrays of cells expressing defined cDNAs. *Nature* 2001, 411:107–110
43. Sreekumar A, Nyati MK, Varambally S, Barrette TR, Ghosh D, Lawrence TS, Chinnaiyan AM: Profiling of cancer cells using protein microarrays: discovery of novel radiation-regulated proteins. *Cancer Res* 2001, 61:7585–7593
44. Scanlan MJ, Welt S, Gordon CM, Chen YT, Gure AO, Stockert E, Jungbluth AA, Ritter G, Jager D, Jager E, Knuth A, Old LJ: Cancer-related serological recognition of human colon cancer: identification of potential diagnostic and immunotherapeutic targets. *Cancer Res* 2002, 62:4041–4047
45. Kononen J, Bubendorf L, Kallioniemi A, Bärnlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G, Kallioniemi OP: Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med* 1998, 4:844–847
46. Xu J, Stolk JA, Zhang X, Silva SJ, Houghton RL, Matsumura M, Vedvick TS, Leslie KB, Badaro R, Reed SG: Identification of differentially expressed genes in human prostate cancer using subtraction and microarray. *Cancer Res* 2000, 60:1677–1682
47. Luo J, Zha S, Gage WR, Dunn TA, Hicks JL, Bennett CJ, Ewing CM, Platz EA, Ferdinandusse S, Wanders RJ, Trent JM, Isaacs WB, De Marzo AM: α -Methylacyl-CoA racemase: a new molecular marker for prostate cancer. *Cancer Res* 2002, 62:2220–2226
48. Rubin MA, Zhou M, Dhanasekaran SM, Varambally S, Barrette TR, Sanda MG, Pienta KJ, Ghosh D, Chinnaiyan AM: α -Methylacyl coenzyme A racemase as a tissue biomarker for prostate cancer. *JAMA* 2002, 287:1662–1670
49. Glas AM, Floore A, Delahaye LJ, Witteveen AT, Pover RC, Bakx N, Lahti-Domenici JS, Bruinsma TJ, Warmoes MO, Bernards R, Wesels LF, Van't Veer LJ: Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics* 2006, 7:278
50. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, Hiller W, Fisher ER, Wickerham DL, Bryant J, Wolmark N: A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004, 351:2817–2826
51. Fan C, Oh DS, Wessels L, Weigelt B, Nuyten DS, Nobel AB, van't Veer LJ, Perou CM: Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 2006, 355:560–569
52. Bammler T, Beyer RP, Bhattacharya S, Boorman GA, Boyles A, Bradford BU, Bumgarner RE, Bushel PR, Chaturvedi K, Choi D, Cunningham ML, Deng S, Dressman HK, Fannin RD, Farin FM, Freedman JH, Fry RC, Harper A, Humble MC, Hurban P, Kavanagh TJ, Kaufmann WK, Kerr KF, Jing L, Lapidus JA, Lasarev MR, Li J, Li YJ, Lobenhofer EK, Lu X, Malek RL, Milton S, Nagalla SR, O'Malley JP, Palmer VS, Pattee P, Paules RS, Perou CM, Phillips K, Qin LX, Qiu Y, Quigley SD, Rodland M, Rusyn I, Samson LD, Schwartz DA, Shi Y, Shin JL, Sieber SO, Slifer S, Speer MC, Spencer PS, Sproles DI, Swenberg JA, Suk WA, Sullivan RC, Tian R, Tennant RW, Todd SA, Tucker CJ, Van Houten B, Weis BK, Xuan S, Zarbl H; Members of the Toxicogenomics Research Consortium: Standardizing global gene expression analysis between laboratories and across platforms. *Nat Methods* 2005, 2:351–356
53. Irizarry RA, Warren D, Spencer F, Kim IF, Biswal S, Frank BC, Gabrielson E, Garcia JG, Geoghegan J, Germino G, Griffin C, Hilmer SC, Hoffman E, Jedlicka AE, Kawasaki E, Martinez-Murillo F, Morsberger L, Lee H, Petersen D, Quackenbush J, Scott A, Wilson M, Yang Y, Ye SQ, Yu W: Multiple-laboratory comparison of microarray platforms. *Nat Methods* 2005, 2:345–350
54. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005, 102:15545–15550
55. Baker SC, Bauer SR, Beyer RP, Brenton JD, Bromley B, Burrill J, Causton H, Conley MP, Elespuru R, Fero M, Foy C, Fuscoe J, Gao X, Gerhold DL, Gilles P, Goodsaid F, Guo X, Hackett J, Hockett RD, Ikonomi P, Irizarry RA, Kawasaki ES, Kayser-Kranich T, Kerr K, Kiser G, Koch WH, Lee KY, Liu C, Liu ZL, Lucas A, Manohar CF, Miyada G, Modrusan Z, Parkes H, Puri RK, Reid L, Ryder TB, Salit M, Samaha RR, Scherf U, Sendera TJ, Setterquist RA, Shi L, Shippy R, Soriano JV, Wagar EA, Warrington JA, Williams M, Wilmer F, Wilson M, Wolber PK, Wu X, Zadro R: The External RNA Controls Consortium: a progress report. *Nat Methods* 2005, 2:731–734
56. Gene Ontology Consortium: Creating the gene ontology resource: design and implementation. *Genome Res* 2001, 11:1425–1433
57. Dahlquist KD, Salomonis N, Vranizan K, Lawlor SC, Conklin BR: GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat Genet* 2002, 31:19–20
58. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich M, Hieronymus H, Wei G, Armstrong SA, Haggarty SJ, Clemons PA, Wei R, Carr SA, Lander ES, Golub TR: The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 2006, 313:1929–1935
59. Demeter J, Beauheim C, Gollub J, Hernandez-Boussard T, Jin H, Maier D, Matese JC, Nitzberg M, Wymore F, Zachariah ZK, Brown PO, Sherlock G, Ball CA: The Stanford Microarray Database: implementation of new analysis tools and open source release of software. *Nucleic Acids Res* 2007, 35:D766–D770
60. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincaid-Beal C, Kulkarni P, Varambally S, Ghosh D, Chinnaiyan AM: Oncome 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* 2007, 9:166–180
61. Petricoin 3rd EF, Hackett JL, Lesko LJ, Puri RK, Gutman SI, Chumakov K, Woodcock J, Feigal DW Jr, Zoon KC, Sistare FD: Medical applications of microarray technologies: a regulatory science perspective. *Nat Genet* 2002, 32(Suppl):474–479
62. Lembersky BC, Thomas LC: Metastases of unknown primary site. *Med Clin North Am* 1996, 80:153–171
63. Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, Angelo M, Ladd C, Reich M, Latulippe E, Mesirov JP, Poggio T, Gerald W, Loda M, Lander ES, Golub TR: Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci USA* 2001, 98:15149–15154
64. Tothill RW, Kowalczyk A, Rischin D, Bousioutas A, Haviv I, van Laar RK, Waring PM, Zalcberg J, Ward R, Biankin AV, Sutherland RL, Henshall SM, Fong K, Pollack JR, Bowtell DD, Holloway AJ: An expression-based site of origin diagnostic method designed for clinical application to cancer of unknown origin. *Cancer Res* 2005, 65:4031–4040
65. Wang D, Coscoy L, Zylberberg M, Avila PC, Boushey HA, Ganem D, DeRisi JL: Microarray-based detection and genotyping of viral pathogens. *Proc Natl Acad Sci USA* 2002, 99:15687–15692
66. Shaffer LG, Bejjani BA: Medical applications of array CGH and the transformation of clinical cytogenetics. *Cytogenet Genome Res* 2006, 115:303–309
67. Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lipshutz R, Chee M, Lander ES: Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 1998, 280:1077–1082
68. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM: Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 2005, 309:1728–1732

69. Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, Schubeler D: Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* 2005, 37:853–862
70. Weinmann AS, Yan PS, Oberley MJ, Huang TH, Farnham PJ: Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes Dev* 2002, 16:235–244
71. Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas 3rd EJ, Gingeras TR, Schreiber SL, Lander ES: Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* 2005, 120:169–181
72. Crawford GE, Davis S, Scacheri PC, Renaud G, Halawi MJ, Erdos MR, Green R, Meltzer PS, Wolfsberg TG, Collins FS: DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. *Nat Methods* 2006, 3:503–509
73. Berns K, Hijmans EM, Mullenders J, Brummelkamp TR, Velds A, Heimerikx M, Kerkhoven RM, Madiredjo M, Nijkamp W, Weigelt B, Agami R, Ge W, Cavet G, Linsley PS, Beijersbergen RL, Bernards R: A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* 2004, 428:431–437
74. Haab BB, Dunham MJ, Brown PO: Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol* 2001, 2:RESEARCH0004
75. Robinson WH, DiGennaro C, Hueber W, Haab BB, Kamachi M, Dean EJ, Fournel S, Fong D, Genovese MC, de Vegvar HE, Skinner K, Hirschberg DL, Morris RI, Muller S, Pruijn GJ, van Venrooij WJ, Smolen JS, Brown PO, Steinman L, Utz PJ: Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nat Med* 2002, 8:295–301
76. MacBeath G, Schreiber SL: Printing proteins as microarrays for high-throughput function determination. *Science* 2000, 289:1760–1763
77. Pawelcz CP, Charboneau L, Bichsel VE, Simone NL, Chen T, Gillespie JW, Emmert-Buck MR, Roth MJ, Petricoin IE, Liotta LA: Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. *Oncogene* 2001, 20:1981–1989
78. Neben K, Korshunov A, Benner A, Wrobel G, Hahn M, Kokocinski F, Golanov A, Joos S, Lichter P: Microarray-based screening for molecular markers in medulloblastoma revealed STK15 as independent predictor for survival. *Cancer Res* 2004, 64:3103–3111
79. van de Rijn M, Perou CM, Tibshirani R, Haas P, Kallioniemi O, Kononen J, Torhorst J, Sauter G, Zuber M, Kochli OR, Mross F, Dieterich H, Seitz R, Ross D, Botstein D, Brown P: Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. *Am J Pathol* 2002, 161:1991–1996
80. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, Hernandez-Boussard T, Livasy C, Cowan D, Dressler L, Akslen LA, Ragaz J, Gown AM, Gilks CB, van de Rijn M, Perou CM: Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin Cancer Res* 2004, 10:5367–5374
81. West RB, Corless CL, Chen X, Rubin BP, Subramanian S, Montgomery K, Zhu S, Ball CA, Nielsen TO, Patel R, Goldblum JR, Brown PO, Heinrich MC, van de Rijn M: The novel marker, DOG1, is expressed ubiquitously in gastrointestinal stromal tumors irrespective of KIT or PDGFRA mutation status. *Am J Pathol* 2004, 165:107–113
82. Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C, Sanda MG, Ghosh D, Pienta KJ, Sewalt RG, Otte AP, Rubin MA, Chinnaiyan AM: The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* 2002, 419:624–629
83. Ma XJ, Wang Z, Ryan PD, Isakoff SJ, Barnettler A, Fuller A, Muir B, Mohapatra G, Salunga R, Tuggle JT, Tran Y, Tran D, Tassin A, Amon P, Wang W, Wang W, Enright E, Stecker K, Estepa-Sabal E, Smith B, Younger J, Balis U, Michaelson J, Bhan A, Habin K, Baer TM, Brugge J, Haber DA, Erlander MG, Sgroi DC: A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell* 2004, 5:607–616
84. Dhanasekaran SM, Barrette TR, Ghosh D, Shah R, Varambally S, Kurachi K, Pienta KJ, Rubin MA, Chinnaiyan AM: Delineation of prognostic biomarkers in prostate cancer. *Nature* 2001, 412:822–826
85. Heuser M, Beutel G, Krauter J, Dohner K, von Neuhoff N, Schlegelberger B, Ganser A: High meningioma 1 (MN1) expression as a predictor for poor outcome in acute myeloid leukemia with normal cytogenetics. *Blood* 2006, 108:3898–3905
86. Ehlers JP, Harbour JW: NBS1 expression as a prognostic marker in uveal melanoma. *Clin Cancer Res* 2005, 11:1849–1853
87. Leung SY, Chen X, Chu KM, Yuen ST, Mathy J, Ji J, Chan AS, Li R, Law S, Troyanskaya OG, Tu IP, Wong J, So S, Botstein D, Brown PO: Phospholipase A2 group IIA expression in gastric adenocarcinoma is associated with prolonged survival and less frequent metastasis. *Proc Natl Acad Sci USA* 2002, 99:16203–16208
88. Higgins JP, Kaygusuz G, Wang L, Montgomery K, Mason V, Zhu SX, Marinelli RJ, Presti JC, van de Rijn M, Brooks JD: Placental S100 (S100P) and GATA3: markers for transitional epithelium and urothelial carcinoma discovered by complementary DNA microarray. *Am J Surg Pathol* 2007, 31:673–680
89. Agrawal D, Chen T, Irby R, Quackenbush J, Chambers AF, Szabo M, Cantor A, Coppola D, Yeatman TJ: Osteopontin identified as lead marker of colon cancer progression, using pooled sample expression profiling. *J Natl Cancer Inst* 2002, 94:513–521
90. Le QT, Sutphin PD, Raychaudhuri S, Yu SC, Terris DJ, Lin HS, Lum B, Pinto HA, Koong AC, Giaccia AJ: Identification of osteopontin as a prognostic plasma marker for head and neck squamous cell carcinomas. *Clin Cancer Res* 2003, 9:59–67
91. Terry J, Saito T, Subramanian S, Ruttan C, Antonescu CR, Goldblum JR, Downs-Kelly E, Corless CL, Rubin BP, van de Rijn M, Ladanyi M, Nielsen TO: TLE1 as a diagnostic immunohistochemical marker for synovial sarcoma emerging from gene expression profiling studies. *Am J Surg Pathol* 2007, 31:240–246
92. Demichellis F, Fall K, Perner S, Andren O, Schmidt F, Setlur SR, Hoshida Y, Mosquera JM, Pawitan Y, Lee C, Adami HO, Mucci LA, Kantoff PW, Andersson SO, Chinnaiyan AM, Johansson JE, Rubin MA: TMPRSS2:ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort. *Oncogene* 2007 [Epub ahead of print]
93. Laxman B, Tomlins SA, Mehra R, Morris DS, Wang L, Helgeson BE, Shah RB, Rubin MA, Wei JT, Chinnaiyan AM: Noninvasive detection of TMPRSS2:ERG fusion transcripts in the urine of men with prostate cancer. *Neoplasia* 2006, 8:885–888
94. Orchard JA, Ibbotson RE, Davis Z, Wiestner A, Rosenwald A, Thomas PW, Hamblin TJ, Staudt LM, Oscier DG: ZAP-70 expression and prognosis in chronic lymphocytic leukaemia. *Lancet* 2004, 363:105–111
95. Chang HY, Sneddon JB, Alizadeh AA, Sood R, West RB, Montgomery K, Chi JT, van de Rijn M, Botstein D, Brown PO: Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol* 2004, 2:e7
96. Glinsky GV, Berezovska O, Glinskii AB: Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer. *J Clin Invest* 2005, 115:1503–1521
97. West RB, Nuyten DS, Subramanian S, Nielsen TO, Corless CL, Rubin BP, Montgomery K, Zhu S, Patel R, Hernandez-Boussard T, Goldblum JR, Brown PO, van de Vijver M, van de Rijn M: Determination of stromal signatures in breast carcinoma. *PLoS Biol* 2005, 3:e187
98. Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, Joshi MB, Harpole D, Lancaster JM, Berchuck A, Olson JA Jr, Marks JR, Dressman HK, West M, Nevins JR: Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 2006, 439:353–357
99. Chi JT, Wang Z, Nuyten DS, Rodriguez EH, Schaner ME, Salim A, Wang Y, Kristensen GB, Helland A, Børresen-Dale AL, Giaccia A, Longaker MT, Hastie T, Yang GP, van de Vijver MJ, Brown PO: Gene expression programs in response to hypoxia: cell type specificity and prognostic significance in human cancers. *PLoS Med* 2006, 3:e47
100. Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z: A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet* 2006, 38:1043–1048
101. Liu R, Wang X, Chen GY, Dalerba P, Gurney A, Hoey T, Sherlock G, Lewicki J, Shedden K, Clarke MF: The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N Engl J Med* 2007, 356:217–226