Minireview

Cancer Genomics: Promises and Complexities[1]

Wei Zhang,[2] Pilar M. Labordé, Kevin R. Coombes, Donald A. Berry, and Stanley R. Hamilton[2]

Cancer Genomics and Proteomics Program, Department of Pathology, Division of Pathology and Laboratory Medicine [W. Z., P. M. L., S. R. H.] and Department of Biostatistics [K. R. C., D. A. B.], The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030

Abstract

The impending final deciphering of the complete human genome, coupled with the advancement of high-throughput technologies, is positioned to bring about a fundamental transformation in cancer research. The era of molecular biology is transforming into the era of genomic biology, with an unprecedented promise of understanding multifactorial diseases and of identifying specific targets that can be used to develop patient-tailored therapies. Although the genomic approach is in an early phase of its development and its tools need to be honed, the application of genomic technologies to cancer research has already generated exciting results both in target identification and in disease classification. In this article, we review some of the developments pertinent to cancer research, discuss potentially problematic areas associated with them, and comment on future trends and issues.

Background

With the distinct possibility of surpassing cardiovascular disease as the number one killer of humans (1), cancer has become a more formidable public enemy than ever before. Despite the fact that the disease is now better understood, we are faced with the frustrating reality that this trend might continue as the life span of the general population continues to increase. The knowledge acquired over the years that cancer cells harbor genetic defects that alter the balance of cell proliferation and cell death (2–7) has led to the compilation of "cancer gene" lists, which have been steadily increasing over the past 10 years. Quite ironically, however, instead of obtaining a progressively better-defined picture of the disease with each new addition to the list, the image has often become obscured in the face of new scientific findings.

The only safe conclusion we can reach is that cancer is a highly variable disease with multiple heterogeneous genetic and epigenetic changes (8–10). Thus, the hit-and-run approach we have taken thus far has revealed itself to be, not surprisingly, less than effective for this complex and multifactorial disease. Instead, a zoom-in holistic approach appears to be needed to identify the genetic and epigenetic changes associated with malignancy onset and progression, to determine the structure of the gene network of effectors and modifiers involved, and to classify cancers based on those genetic networks.

The Human Genome Project and Cancer Genomics

The realization that a comprehensive roadmap was indispensable to gain a holistic insight into biology prompted scientists to establish an ambitious goal for the new century: to decipher the complete sequence of the human genome, the roadmap of life. Undaunted by the seeming enormity of the task, a major international effort was launched 10 years ago to sequence the complete human genome, which amounts to 3.2 billion bp (11, 12).

By mid-2000, a draft of the human genome was announced (13) after a fierce competition between the publicly funded Human Genome Project and a private genomics company, *Celera* (14, 15), and the results were recently reported in *Nature* and *Science* [see vol. 409 of *Nature* (2001) and vol. 291 of *Science* (2001)]. Gaining knowledge about the order in which the billions of bp that comprise the human genome are disposed, however, reaches far beyond providing a compilation of nucleotide lists. Subsequent comparison of sequences from different individuals has revealed the existence of single nucleotide polymorphisms that may provide new insight into the genetic modifications that characterize cancer onset (16–18).

With the anticipation of having the human genome fully sequenced by 2003, there is no doubt that we will experience an unprecedented exploration of genetic information in the first part of the 21st century. Simultaneously, the sequencing of the complete genomes of several other animal model systems has also been undertaken. Whereas efforts to unleash the secrets of the human genome continue, the genomes of several disease-causing microbes have been completed (19, 20), as well as those of yeast (21), *Caenorhabditis elegans* (22), and *Drosophila* (23). The mouse genome sequencing is likewise well under way.

The discovery of the Periodic Table had a major impact on chemistry and the pharmaceutical industry. So too, the complete decoding of the human genome will have far-reaching effects on human health and quality of life. Implications for clinical practice are becoming evident (24–29), and issues spanning all areas of medical practice have arisen. An example is prophylactic surgery for carriers of mutations of cancer susceptibility genes (30). The multidisciplinary field of cancer genomics has been born and will experience a rapid growth in the new decade (31).

Cancer genomics researchers approach cancer from a systemic perspective, and they view cancer holistically. The growth

Received 3/21/01; accepted 5/21/01.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] Supported in part by the Tobacco Settlement Funds as appropriated by the Texas State Legislature, by a generous donation from the Michael and Betty Kadoorie Foundation, and by a grant from the Texas Higher Education Coordination Board under Grant 003657-0039-1999.

[2] To whom requests for reprints should be addressed, at Department of Pathology, Box 85, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030.

of this field will involve shaping and reshaping of our scientific pursuits and our academic institutions. The result will be a better understanding of the disease process and its heterogeneity. We envision drugs that are exquisitely tailored to treat individual cancers. A new and potentially very complicated system of cancer classification and naming will augment the current descriptive system. How soon this will occur depends both on improving genomic technologies and bioinformatics and on the teamwork of cancer researchers, statisticians, mathematicians, engineers, and computer scientists.

Technology Development and the Role of Academic Institutes

Among the 3 billion bp of the human genome, recent estimates place the number of genes in the range of 26,000–31,000 (32). Currently, we only understand the function of a fraction of those genes. To survey all of the genes is a technological challenge. To face this challenge, new high-throughput genome screening tools have been developed. Both cDNA microarrays (33–36) and SAGE[3] (37, 38) can simultaneously assess the expression of a large pool of genes in a given cell population (typically one tumor sample). By applying similar principles to CGH (39), CGH arrays (40) have been developed to screen for gene deletion and amplification at the chromosomal level. These technologies are high throughput and offer a comprehensive and accurate picture. Other screening techniques can be used in a complementary way to confirm the results.

The use of arrays is not new in biology (41), but relatively recent innovations have increased the efficiency and reliability of the technique. High-density printing and miniaturization permit the simultaneous analysis of thousands of genes. Gene expression arrays come in different formats, but the basic principle is the same. Thousands of cDNAs or oligonucleotides can be robotically deposited (or printed) to a solid matrix such as a precoated glass microscope slide or a nylon membrane in an organized fashion. Oligonucleotides can also be synthesized directly on a silicon chip (42). Those attached DNAs, representing thousands of genes, are usually called probes. To use microarrays to simultaneously profile the expression of genes in a given cell population, one must first isolate total RNA. Next, one generates cDNA by reverse transcription. During reverse transcription, nucleotides labeled with either fluorescent dyes, such as cy5 and cy3, or radioactive isotopes are incorporated to make "hot" targets. The targets are then hybridized to the probes on the microarray. After detection of the hybridized signals with a laser scanner or a phosphorimager, an expression profile is produced by quantification and mathematical analysis of the image thus generated.

One limitation of microarrays is that they form a closed system: each microarray can only provide information about the particular genes that are included on the array. In contrast, SAGE (37, 38) is an open system that may uncover any genes expressed in the cells being tested. SAGE takes advantage of the fact that a 9–10-nucleotide sequence (a "tag") often contains enough information to uniquely identify a gene. Generation of gene tag concatamers and their subsequent sequencing reveals the relative frequency with which each tag appears in the whole library. This frequency represents the relative expression level of that gene in cells.

SAGE is more labor intensive than microarrays. Tags from more abundant genes are, naturally, more likely to be sequenced. In practice, only about 10–20% of all tags represent unique genes (43). To detect low abundance genes, one must sequence many more tags. For example, gaining information about 5,000 genes requires sequencing of 50,000 tags.

As with DNA microarrays, several variants of SAGE are being developed to address common problems. Mini-SAGE, for example, attempts to gain genetic information by gene expression profiling from 1 μg of total RNA (44). Generation of longer cDNA fragments from SAGE tags for gene identification (GLGI) attempts to address some limitations of SAGE. GLGI generates a longer cDNA fragment for further study and also enables the identification of a sequence from among many with the same tag (45). A public database, SAGEmap,[4] incorporates the information compiled to date and makes it available to the researcher.

CGH (39) and CGH arrays (40) can also be considered complementary genome screening tools. Developed in the early 1990s, CGH presents a physical map of chromosomal region gain or loss throughout the entire genome. To achieve this map, chromosomal DNA obtained from both the sample being tested and a normal control is labeled with two different fluorochromes and then cohybridized to normal chromosome spreads. Differences in signal intensity along the chromosome indicate gene deletion and/or amplification.

A limitation of CGH is that it is able only to detect general regions of chromosomal alterations. CGH arrays, on the other hand, possess a higher resolution capability and can further localize the altered region in the chromosome. Instead of arraying cDNAs on the solid matrix, microarray-based CGH involves printing large mapped DNA fragments, such as bacterial artificial chromosome (BAC) clones, on a glass slide, aligned according to their chromosomal locations. To search for regions that harbor deletion or amplification in tumor cells, genomic DNAs from tumor and control normal cells are labeled with two different fluorophores and then cohybridized to the CGH arrays. The skewed signal of certain clones hints at the possible deletion or amplification. The enhanced resolution of CGH arrays enables a rapid identification of the specific genes involved in the chromosomal aberration, particularly when coupled with other approaches such as customized cDNA arrays.

The CGH and CGH array approaches can complement other techniques such as fluorescence *in situ* hybridization or classical chromosome banding, a technique that cannot reliably analyze grossly abnormal karyotypes. For example, translocations in breast carcinoma cell lines that were previously unknown due to their intractability to banding analysis have been described using CGH (46).

[3] The abbreviations used are: SAGE, serial analysis of gene expression; CGH, comparative genomic hybridization; IGFBP2, insulin-like growth factor-binding protein 2; LCM, laser capture microdissection.

[4] Internet address: http://www.ncbi.nlm.nih.gov/SAGE.

Although CGH cannot be exclusively relied upon for the identification of chromosomal abnormalities (for example, it cannot detect polyploid karyotypes), it is able to quickly yield reliable information about local chromosomal alterations and can help generate tumor-specific fluorescence *in situ* hybridization probes to obtain prognostic information faster than other techniques (47).

Since the advent of these high-throughput techniques, the cytogenetic characterization of malignant cells that were previously resistant to identification has been enhanced. A recent study combined CGH, conventional G-banding, and whole chromosome painting to provide a genetic map of two colon cell lines whose chromosomal aberrations had not been identified previously (48).

The advent of genomics has signaled the beginning of a new era of medical promise, and the race to hone genomic tools is proceeding at a dazzling pace (49–53). In the initial stages, the technology has been dominated by biotechnology industries. The resulting high cost of the newly developed tools, coupled with issues of intellectual property and copyrights in the digital area (54), has hindered widespread application of the technology. Amid a general transformation of cancer and genomic centers (55, 56), academic institutions around the country have begun to build core genomic facilities.

There are several reasons why academic institutions have become involved in this technological arena. First, having in-house core facilities speeds up the application of the technology by providing low-cost cDNA microarray products to a broad range of cancer researchers. Second, quality can be controlled more vigorously. Third, customized microarrays can be easily built to meet the specific needs of research projects, which is particularly important because the genes of most interest vary from one project to another. Fourth, academic institutions with their own microarray facilities can better control their intellectual property resulting from genomic studies. During the last year, massive quantities of data have accumulated, problems have been identified, and new analysis tools are being built. With technology from both commercial and academic sources, success stories are emerging.

Application of Genomic Technology in Cancer Research

Genomic technology has been applied to several areas of cancer research. By profiling and comparing gene expression of tumors of different grades or primary and metastatic tumors, several genes involved in cancer progression or metastases have been found, new classification paradigms have been established, genes have been placed into pathways, and gene deletions and amplifications have been identified. During this process, bioinformatics has evolved from classical sequence analysis to a discipline of data mining and handling of multidimensional data.

Identification of Genes Differentially Expressed in Tumors. One of the major accomplishments of the newly developed genome screening tools has been the identification of cancer genes. More efficient gene identification will lead to better target selection for cancer therapeutics. This has been shown by the recent use of the drug STI571, which specifically targets bcr-abl tyrosine kinase, a hallmark of chronic myelogenous leukemia. (57) STI571 has been combined with leptomycin B to irreversibly induce cell apoptosis (58).

Several groups have demonstrated the power and utility of high-throughput gene expression profiling in cancer gene discovery, and many examples testify to the potential of these technologies.

Our recent glioma gene expression profiling studies of IGFBP2 revealed that IGFBP2, which is normally expressed in fetal cells and turned off in adult cells, becomes highly expressed in high-grade glioma-glioblastoma (59). This observation suggests that IGFBP2 expression might be associated with progression to glioblastoma, one of the deadliest brain tumors. Another independent study has confirmed that IGFBP2 is up-regulated in glioblastomas, using a combination of DNA and tissue microarrays (60).

In a study of melanoma variants with differing metastatic potential, the small GTPase *RhoC* has been identified as a key gene for the enhancement of tumor metastasis. This potential link was investigated in an *in vitro* system in which expression of a dominant-negative Rho inhibits metastasis (61).

Other transcripts associated with enhanced metastatic potential, such as keratin K5, cystatin S, serum amyloid A, the human homologue of yeast ribosomal S28, and the p32 subunit of human pre-mRNA splicing factor SF2, have been identified in colon tumor cell lines, using the SAGE technique (62).

SAGE was also used in a study of 85 patients with pancreatic cancer to identify markers for this malignancy (63). In this study, 183 genes were found to be differentially expressed in tumor cells in comparison with normal samples. One such gene is the tissue inhibitor of metalloproteinase type I, which encodes a secreted protein. This study opened the door for the use of SAGE to reveal novel serum markers. SAGE has also enabled the identification of expressed genes in human tumor endothelium (64), which can have an impact on the development of new therapeutic approaches to inhibit angiogenesis in a tumor-specific manner.

Pharmacogenomics. In association with these findings, another emerging field of drug development has been pharmacogenomics (65–67). As mentioned above, elucidation of the functions of cancer genes is an essential component of target validation and is the first step to drug discovery. This process also involves learning about epistasis and gene interactions. In pharmacology, high-throughput technologies offer the prospect of identifying mechanisms that affect therapeutic response to a specific chemotherapeutic agent in a given patient. Examination of gene expression is potentially able to show the molecular basis for drug resistance or sensitivity and can aid in the elucidation of the alterations in the intracellular metabolic pathway that are necessary for drug response.

For example, two glioblastoma U251 cell lines (one of which was chemoresistant, whereas the other was chemosensitive), were treated with 1,3-bis(2-chloroethyl)-1-nitrosourea, a chemotherapeutic agent. Gene expression profiles were then examined. Interestingly, a group of DNA repair genes was down-regulated in 1,3-bis(2-chloroethyl)-1-nitrosourea-treated chemosensitive U251 cells, whereas some of the DNA repair genes were up-regulated in chemoresistant U251 cells (68). Similarly, in a recent study of the apoptotic pathway induced by all-*trans* retinoic acid in T-cell lymphoma, four groups of genes

with differential biological functions were identified, and their expression varied in response to retinoic acid treatment (69).

Drug development is a major problem area. There have been few advances in designing effective chemotherapeutic agents that selectively affect malignant cells. The hope is that identifying genes that are differentially expressed in tumor-derived cells will lead to the establishment of targets in the development of tumor-specific drugs. Further identification of tumor-specific gene expression may also uncover serum-based diagnostic markers that can detect malignancies earlier, at a stage when most are more sensitive to treatment. A recent study using cDNA arrays was undertaken for ovarian epithelial cells with this goal in mind (70). It was found that a large number of genes differentially expressed in tumor *versus* normal epithelial cells encode membrane-associated or secreted proteins, a finding that opens the field for research into finding diagnostic markers for this malignancy.

Identification of Chromosome and Gene Amplifications and Deletions. CGH is more sensitive than other techniques such as G-banding in identifying critical regions present in the malignant clone. This has been reported in several studies, such as those for myeloid neoplasias (71). CGH has likewise shown that certain human tumors undergo ploidy changes that correlate with different and complex chromosomal rearrangements. To mention just a few studies, CGH has revealed that overrepresentation of the short arm of chromosome 12 is related to invasive growth of human testicular seminomas and nonseminomas (72). It has also shown that amplification of the REL gene occurs in follicular lymphoma (73) and that previously undetected amplifications are present in fibroadenomas of the breast (74).

Detection of amplifications and deletions is closely related to new attempts at classifying and describing malignancies based on their molecular characteristics, as we discuss below. A common approach for the molecular description of malignancies consists of the identification of general areas of nonrandom chromosomal alterations. For example, one can use CGH or cross-species color banding [an approach that was chosen in a recent study of cholangiocarcinoma, a rare tumor (75)] and then zoom in with other genomic approaches to identify candidate genes. This will also enable the researchers to localize tumor suppressor genes or oncogenes. This has been accomplished recently in the identification of a tumor suppressor gene on chromosome 11q23 related to non-small cell lung carcinoma (76) and in the identification of PIK3CA as an oncogene in ovarian cancer (77).

Several studies have confirmed the ability of CGH arrays to measure chromosomal amplifications and deletions with high precision. Chromosomal targets may also be hybridized to cDNA arrays to identify deletion or amplification of genes, as demonstrated in a study in which a cDNA microarray-based CGH method was applied to analyze DNA copy number variation in breast cancer cell lines and tumors (78).

Molecular Classification of Cancers. The new technologies are also revealing their utility for a classification of malignancies that is no longer phenotypically based (*i.e.*, based on histopathological features) but genotypically based (79, 80). An early study by Golub *et al.* (81) using DNA microarrays showed that gene expression monitoring alone can suffice for cancer classification, an assertion that was based on their observation that different genetic signatures could be identified for acute myeloid leukemia and acute lymphoblastic leukemia.

A recent study by Alizadeh *et al.* (82) showed that two previously unknown subtypes of diffuse large B-cell lymphoma can be identified based on gene expression profiles and that the two subtypes have clear differences in survival. A molecular classification of cutaneous malignant melanoma revealed that mathematical analysis of gene expression is able to establish subtypes of this tumor based on gene expression profiles (83).

Data continue to accumulate on the genetic signature of dozens of malignancies. For example, the proliferation rate and clinical outcome of human synovial sarcoma has recently been correlated with the expression of SYT-SSXI, whereas until recently the genes associated with its characteristic translocation were unknown (84). More recently, hereditary breast cancers BRCA1 and BRCA2 have been studied in terms of their differing expression levels of certain genes, such as the keratin 8 gene, which was differentially expressed between the mutation-positive and mutation-negative breast cancer groups. The authors of this study conclude that "a heritable mutation will leave an identifiable trace on a tumor's expression profile" (85). Thyroid carcinomas have likewise been profiled, and the degree of differentiation of cells in normal, differentiated, and undifferentiated tissue has been correlated with the expression levels of genes such as osteonectin, α-tubulin, and glutathione peroxidase (86).

cDNA microarrays have also been used in a large and comprehensive study whose goal was to reveal systematic variation in gene expression. For this study, 8000 genes and 60 cell lines from central nervous system, renal, ovarian, leukemia, colon, and melanoma neoplasms were studied. The results establish the validity of the genomic approach to differentiate among tumor subtypes (87).

Tumor profiling goes beyond mere attempts at classification and offers insight into the dynamics of the cell differentiation process. In a recent study by Lossos *et al.* (88), gene expression profiling was not only able to distinguish between two tumor subtypes of B-cell diffuse large cell lymphoma but also to validate assumptions on the development of these malignancies at discrete stages of cell maturation. The new term "molecular portrait" (89), coined for the case of human breast tumors, appropriately reflects the recent findings that gene expression profiles relate to physiological differences for a diversity of cancers.

Because the ultimate goal of the new emerging paradigm is to benefit cancer patients, we stress that molecular classification of cancers can be translated into better therapeutic approaches. Indeed, explaining histological tumor heterogeneity by reference to molecular tumor subtypes has generated several important cancer therapies. For example, patients classified by the genomic approach as having acute promyelocytic leukemia who present a specific chromosomal translocation [t(15,17)] selectively respond to treatment with all-*trans* retinoic acid and arsenic trioxide (90, 91). It is now also known that oligodendrogliomas with deletion of chromosome 1p are much more responsive to chemotherapy than those that do not harbor the deletion (92). In addition, treating primary breast cancer patients with moderately high doses of Adriamycin has been shown to benefit only those patients whose tumors overexpress HER-2/neu or p53 (93).

Clustering Algorithms. The molecular classification of tumor type and subtypes using gene expression levels, as in the examples above, is performed by clustering algorithms. The goal of cluster analysis is to separate a collection of complex samples into groups in such a way that the most "similar" samples are grouped together. Unsupervised clustering methods are used to discover the group structure directly from the data. These methods do not use any information from a known classification of the samples (histology, for example), thus they are appropriate for discovering new classes. Supervised clustering methods, by contrast, start from a known classification and try to find the best methods for distinguishing the classes; these methods are more appropriate for trying to predict (or diagnose) the class of a new sample. A wide variety of clustering algorithms have been developed in the analysis of experiments from other scientific disciplines. Many of these methods have been applied to the analysis of large-scale gene expression experiments (94–96).

Potential Problems

The emergence of any new technological advance is inevitably accompanied by expanded enthusiasm and anticipatory eagerness, with the unwanted consequence that problems that might arise are downplayed and hidden. For the genomic technology to fulfill its promise, problems must be identified, and solutions must be found. At a minimum, cancer researchers have to recognize the limitations of the technology.

Quality Control Issues of Microarray Technology. Unfortunately, when high density and high throughput are the chief goals, quality controls are often ignored. To generate a cDNA microarray, thousands of cDNA clones have to be processed through many error-prone steps. Examples include bacterial stock library culture replication in 96-well plates, PCR amplification, purification of DNA, and transfer of DNA into 384-well plates for printing. Most high-density cDNA arrays contain an error rate in excess of 10%. In other words, at more than 10% of spots, the DNA on the microarray is not what the label says it is. It should therefore not be surprising that in a number of cases the microarray results cannot be confirmed by other approaches.

At the M. D. Anderson Cancer Genomics Core Laboratory, quality control of the microarray is crucial. To ensure the correct gene identity on the arrays, all of the cDNAs on the microarrays are sequenced before being printed (97). To avoid propagation to later analysis of errors generated from artifacts, cDNA clones are duplicated on the array. If the replicate spots do not agree, the corresponding data points can be flagged as likely artifacts and discarded. Although high-density microarrays can be tempting, without vigorous quality control, more can be less.

Sample Size and Replication of Experiments. Due to the high cost of microarrays in the past, the sample size of the studies has been small, and replication of experiments has been cost prohibitive. These stumbling blocks have posed serious difficulties for analysis, especially in light of the multidimensional nature of the data. With inexpensive arrays produced by academic institutions, researchers can design experiments with a larger number of samples to achieve better results, and experiments can be repeated to better understand the sources of errors.

Biological Variability. Many potential sources of variability are involved in microarray data acquisition and analysis. First, the biological materials considered can be—and usually are—highly heterogeneous. Tumors of the same type from different patients can have very different genetic and epigenetic profiles (9, 10). Even normal tissues from different individuals exhibit differences, although the degree of difference may be smaller than that in the presence of neoplasms. We also know that tumor tissues surgically removed from patients can differ in size and grade, contain different compositions of cell types, present different levels of infiltrating lymphocytes, and also differ in the extent to which necrosis is present. As an added source of heterogeneity, samples being studied may come from different locations within the same tumor.

Based on pathological evaluation, heterogeneous patterns can be observed in different sections of tumor tissues. Therefore, effort must be expended by pathologists to characterize the tissues under study. Protocols allowing surgical samples to be rapidly processed and appropriately stored from operating room to laboratories need to be put in place in research and clinical facilities. This process requires a coordinated operation between researchers, pathologists, and clinicians.

Experiments need to evaluate the degree of nonpathological variability caused by biological factors. But not all of the validation experiments need to be carried out using a microarray platform. Because of the variability associated with microarray technologies, independent assays such as Northern blotting, real-time PCR, and tissue arrays are commonly performed to confirm the findings of potential interest from microarray assay. Toward this end, high-throughput real-time PCR assays as well as tissue arrays are being established to screen multiple samples for a single gene for diagnostic and prognostic purposes (98–101).

LCM and Amplification. One approach to addressing the tissue heterogeneity problem is to isolate specific cells for analysis. LCM technology (102, 103) has been developed for this purpose. Before the advent of this technology, manual dissection, selection by cell sorting, surface marker selection, and culture in selective conditions were carried out under less than optimal circumstances. With LCM, this isolation can be accomplished with a much greater precision and under conditions that only minimally perturb the cells for study. One school of thought is that to ascertain an accurate cancer gene expression profile by microarray analysis, pure cancer cells have to be isolated and used. LCM requires a higher sensitivity of the microarray protocol to compensate for the limited amount of tissue that can be isolated by this approach.

Thus, using LCM requires adopting an amplification procedure. Several protocols that involve a limited PCR amplification and T7 *in vitro* transcription step have been described. For example, antisense RNA amplification based on T7 transcription reportedly presents an amplification method for analyzing the mRNA population from a few hundred living cells (104). Other techniques combine antisense RNA amplification with a template switching effect (105). Both the linearity of these amplification procedures and the advantages and disadvantages associated with them have yet to be adequately assessed. Because amplification methods can yield false positives, confirmatory studies are a necessity. On the other hand, targeted analysis of purified cancer cells may be too simple-minded to reflect

the whole picture of cancer. After all, cancer cells do not exist in isolation, and cancer pathology deals with a mixture of cancer cells with surrounding stroma cells, inflammatory cells, and endothelial cells. These so-called "normal" cells among tumors are clearly distinct from normal cells residing in the non-tumor region, as confirmed by the recent finding that tumor endothelium cells have a unique gene expression signature (64).

Interface between Biologists and Bioinformaticians

Although reviewing the bioinformatics involved in analyzing genomics data is beyond the scope of this article, we assert that it is a critical component of the process (106–109). Cancer genomics demands a new level of interaction between clinicians, biologists, statisticians, mathematicians, programmers, and computer scientists. Biologists in search of responses to biological questions are faced daily with new challenges. Individual experiments—whether they use cDNA microarrays, SAGE, CGH arrays, or tissue arrays—generate massive amounts of data that must be stored, catalogued, and annotated. The problem of storing this information is manageable. The challenge lies in understanding it and interpreting it in terms of biological meanings. The promise of high-throughput technologies is that they can produce a comprehensive, holistic picture of the genotypic state of a particular cancer cell. However, new tools are necessary to extract this picture from the mountains of data being collected; it is not feasible to sift through the data by hand and find interesting genes. Biological protocols for collecting the data must be integrated with mathematical and statistical protocols for analyzing it accurately and with the computer algorithms implementing the protocols (110–112).

To solve these problems, scientists face another challenge: reconciling the different perspectives from which biologists and nonbiologists approach biological questions. Blending various views can be difficult because the various groups use different languages. Training a new breed of scientists with expertise in all relevant areas will bridge the gap. At present, close interactions and ongoing communication are needed to make this enterprise work. Biologists must become more quantitative, and informaticians must learn to appreciate the sometimes idiosyncratic nature of biology. Together, they must design methodologies suitable for this "noisy" and "inexact" science.

The Human Genome Project brought two kinds of computer scientists into closer contact with biologists: database experts who could design the structures needed to store the data in a useful form, and pattern matching experts who could develop tools to recognize and measure the similarities between different genetic profiles. Making sense of the data from microarrays or other high-throughput technologies will require a wider range of expertise, drawing on mathematicians, statisticians, and engineers in addition to computer scientists. Experts in pattern recognition will help assess transcript changes induced by carcinogenic events, for example, to link oncogene activity to specific gene expression patterns or to establish other interrelations. Biostatisticians will develop new tools and new experimental designs to help distinguish the true differential expression of genes on a microarray from the normal random variation between biologically similar samples. New models will be built, and extensive computer simulations will be carried out to ascertain the sensitivity of classifying tumor samples into disease subtypes or assigning genes to functional categories to varying levels of noise. By collaborating closely with scientists from other disciplines, cancer researchers will be able to turn the promise of genomics technology into a reality.

Concluding Remarks and Perspectives

Amid constructive debate on the relative usefulness of genomics and the complementary applications of genomics and proteomics (111–122), there is a constant improvement of genomic tools and an outpouring of new proteomic techniques such as display technology and protein chips (124, 125). The future will bring a new combination of arrays and powerful amplification strategies, vaccines, gene chips, and translational research. However, behind the hyped enthusiasm, there is pain, and there are problems. Out of those problems, successes will occur, technologies will improve, and a new breed of scientists will emerge. Looking ahead, one cannot help but feel renewed enthusiasm about our knowledge of cancer and the promise of using this knowledge to cure cancer one patient at a time.

Acknowledgments

We acknowledge Drs. Keith Baggerly and Latha Ramdas for valuable critical review of the manuscript.

References

1. Murphy, S. L. Deaths. Final data for 1998. National Vital Statistics Reports, *48:* 1–106, 2000.

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

3. Rowley, J. D. The critical role of chromosome translocation in human leukemia. Annu. Rev. Genet., *32:* 493–519, 1998.

4. Fujiwara, T., Stolker, J. M., Watanabe, T., Rashid, A., Longo, P., Eshleman, J. R., Booker, S., Lynch, H. T., Jass, J. R., Kim, H., Jen, J., Vogelstein, B., and Hamilton, S. R. Accumulated clonal genetic alterations in familial and sporadic colorectal carcinomas with widespread instability in microsatellite sequences. Am. J. Pathol., *153:* 1063–1078, 1998.

5. Toyota, M., Shen, L., Ohe-Toyota, M., Hamilton, S. R., Sinicrope, F. A., and Issa, J. P. Aberrant methylation of the cyclooxygenase 2 CpG island in colorectal tumors. Cancer Res., *60:* 4044–4048, 2000.

6. Rashid, A., Zahurak, M., Goodman, S. N., and Hamilton, S. R. Genetic epidemiology of mutated *K-ras* proto-oncogene, altered suppressor genes, and microsatellite instability in colorectal adenomas. Gut, *44:* 826–833, 1999.

7. Boland, C. R., Thibodeau, S. N., Hamilton, S. R., Sidransky, D., Eshleman, J. R., Burt, R. W., Meltzer, S. J., Rodriguez, B., Fodde, R., Ranzani, G. N., and Srivastava, S. A National Cancer Institute Workshop on microsatellite instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res., *58:* 5248–5257, 1998.

8. Laird, P. W., and Jaenisch, R. The role of DNA methylation in cancer genetics and epigenetics. Annu. Rev. Genet., *30:* 441–464, 1996.

9. Lleonart, M. E., Martin-Duque, P., Sanchez Prieto, R., Moreno, A., and Ramon y Cajal, S. Tumor heterogeneity: morphological, molecular and clinical implications. Histol. Histopathol., *15:* 881–898, 2000.

10. Rew, D. A. The importance of heterogeneity in tumor pathology. Adv. Anat. Pathol., *5:* 156–163, 1998.

11. Dunham, I., Shimizu, N., Roe, B. A., Chissoe, S., Hunt, A. R., Coliins, J. E., Bruskiewich, R., Beare, D. M., Clamp, M., Smink, L. J., Ainscough, R., Almeida, J. P., Babbage, A., Bagguley, C., Bailey, J., Barlow, K., Bates, K. N., Beasley, O., Bird, C. P., Blakey, S., Bridgeman, A. M., Buck, D., Burgess, J., Burrill, W. D., and O'Brien, K. P.,

et al. The DNA sequence of human chromosome 22. Nature (Lond.), *402:* 489–495, 1999.

12. Hattori, M., Fujiyama, T. D., Taylor, H., Watanabe, T., Yada, H., Park, S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D. K., Soeda, M., Ohki, T., Takagi, Y., Sakaki, S., Taudien, S., Blechschmidt, K., Polley, A., Menzel, U., Delabar, J., Kumpf, K., Lehmann, R., Patterson, D., Reichwald, K., Rump, A., Schillhabel, M., Schudy, A., Zimmerman, W., Rosenthal, A., Kudoh, J., Shibuya, K., Kawasaki, K., Asakawa, S., Shintani, A., Sasaki, T., Nagamine, K., Mitsuyama, S., Antonarakis, S. E., Minoshima, S., Shimizu, N., Nordsiek, G., Hornischer, K., Brandt, P., Scharfe, M., Schon, O., Desario, A., Reichelt, J., Kauer, G., Blocker, H., Ramser, J., Beck, A., Klages, S., Hennig, S., Riesselmann, L., Dagand, E., Haaf, T., Wehrmeyer, S., Borzym, K., Gardiner, K., Nizetic, D., Francis, F., Lehrach, H., Reinhardt, R., and Yaspo, L. The DNA sequence of human chromosome 21. Nature (Lond.), *405:* 311–319, 2000.

13. Ferry, J. "Working draft" of human genome available by June. Lancet, *355:* 1337, 2000.

14. Collins, F. S., and Mansoura, M. K. The human genome project. Cancer (Phila.), *91* (Suppl.)*:* s221–s225, 2001.

15. Venter, J. C., Adams, M. D., Sutton, G. G., Kerlavage, A. R., Smith, H. O., and Hunkapiller, M. Shotgun sequencing of the human genome. Science (Wash. DC), *280:* 1540–1542, 1998.

16. Lindblad-Toh, K., Tanenbaum, D. M., Daly, M. J., Winchester, E., Lui, W. O., Villapakkam, A., Stanton, S. E., Larsson, C., Hudson, T. J., Johnson, B. E., Lander, E. S., and Meyerson, M. Loss-of-heterozygosity analysis of small-cell lung carcinomas using single-nucleotide polymorphism arrays. Nat. Biotechnol., *18:* 1001–1005, 2000.

17. Clifford, R., Edmonson, M., Hu, Y., Nguyen, C., Scherpbier, T., and Buetow, K. H. Expression-based genetic/physical maps of single-nucleotide polymorphisms identified by the cancer genome anatomy project. Genome Res., *10:* 1259–1265, 2000.

18. Smigielski, E. M., Sirotkin, K., Ward, M., and Sherry, S. T. dbSNP. A database of single nucleotide polymorphisms. Nucleic Acids Res., *28:* 352–355, 2000.

19. Baillie, L., and Read, T. D. *Bacillus anthracis*, a bug with attitude! Curr. Opin. Microbiol., *4:* 78–91, 2001.

20. Andersson, S. G., Zomorodipour, A., Andersson, J. O., Sicheritz-Ponten, T., Alsmark, U. C., Podowski, R. M., Naslund, A. K., Eriksson, A. S., Winkler, H. H., and Kurland, C. G. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature (Lond.), *396:* 133–140, 1998.

21. Mewes, H. W., Albermann, K., Bahr, M., Frishman, D., Gleissner, A., Hani, J., Heumann, K., Kleine, K., Maierl, A., Oliver, S. G., Pfeiffer, F., and Zollner, A. Overview of the yeast genome. Nature (Lond.), *387:* 7–65, 1997.

22. The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science (Wash. DC), *282:* 2012–2047, 1998.

23. Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., Chen, L. X., Brandon, R. C., Rogers, Y. H., Blazej, R. G., Champe, M., Pfeiffer, B. D., Wan, K. H., Doyle, C., Baxter, E. G., Helt, G., Nelson, C. R., Gabor, G. L., Abril, J. F., Agbayani, A., An, H. J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R. M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E. M., Beeson, K. Y., Benos, P. V., Berman, B. P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M. R., and Bouck, J., *et al*. The genome sequence of *Drosophila melanogaster*. Science (Wash. DC), *287:* 2185–2195, 2000.

24. Boenning, D. A., and Kalfoglou, A. L. The future of clinical research: the role of genomics, informatics, and evidence based medicine. J. Investig. Med., *49:* 10–11, 2001.

25. Gerzer, R. Functional genomics: from disease-oriented to personalized medicine. JMM, *78:* B54, 2000.

26. Olden, K., and Wilson, S. Environmental health and genomics: visions and implications. Nat. Rev. Genet., *1:* 149–153, 2000.

27. Pang, C. P., Baum, L., and Lam, D. S. C. Hunting for disease genes in multi-functional diseases. Clin. Chem. Lab. Med., *38:* 819–825, 2000.

28. Powledge, T. M. The limits of clinical genomics. EMBO Rep., *1:* 213, 2000.

29. Sander, C. Genomic medicine and the future of health care. Science (Wash. DC), *28:* 1977–1978, 2000.

30. Meijers-Heijboer, E. J., Verhoog, L. C., Brekelmans, C. T., Seynaeve, C., Tilanus-Linthorst, M. M., Wagner, A., Dukel, L., Devilee, P., van den Ouweland, A. M., van Gell, A. N., and Klijn, J. G. Presymptomatic DNA testing and prophylactic surgery in families with a BRCA1 or BRCA2 mutation. Lancet, *355:* 2015–2020, 2000.

31. Futreal, P. A., Kaspryzk, A., Birney, E., Mullikin, J. C., Wooster, R., and Stratton, M. R. Cancer and genomics. Nature (Lond.), *409:* 850–851, 2001.

32. Baltimore, D. Our genome unveiled. Nature (Lond.), *409:* 814–816, 2001.

33. De Risi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., and Trent, J. M. Use of cDNA microarrays to analyze gene expression patterns in human cancer. Nat. Genet., *14:* 457–460, 1996.

34. Lockhart, D. J., and Winzeler, E. A. Genomics, gene expression and DNA arrays. Nature (Lond.), *405:* 827–835, 2000.

35. Bowtell, D. D. Options available–from start to finish–for obtaining expression data by microarray. Nat. Genet., *21:* 20–24, 1999.

36. Penn, S. G., Rank, D. R., Hanzel, D. K., and Barker, D. L. Mining the human genome using microarrays of open reading frames. Nat. Genet., *26:* 315–318, 2000.

37. Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. Serial analysis of gene expression. Science (Wash. DC), *270:* 484–487, 1995.

38. Velculescu, V. E., Vogelstein, B., and Kinzler, K. W. Analyzing uncharted transcriptomes with SAGE. Trends Genet., *16:* 423–425, 2000.

39. Kallionemi, A., Kallionemi, O. P., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F., and Pinkel, D. Comparative genomic hybridization for cytogenetic analysis of solid tumors. Science (Wash. DC), *258*: 818–821, 1992.

40. Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y., Dairkee, S. H., Ljung, B. M., Gray, J. W., and Albertson, D. G. High resolution analysis of cDNA copy number variation using comparative genomic hybridization to microarrays. Nat. Genet., *20:* 207–211, 1998.

41. Lennon, G. G., and Lehrach, H. H. Hybridization analyses of arrayed cDNA libraries. Trends Genet., *7:* 314–317, 1991.

42. Lipshutz, R. J., Fodor, S. P., Gingeras, T. R., and Lockhart, D. J. High density synthetic oligonucleotide arrays. Nat. Genet., *21* (Suppl. 1)*:* 20–24, 1999.

43. Velculescu, V. E., Madden, S. L., Zhang, L., Lash, A. E., Yu, J., Tago, C., Lal, A., Wang, C. J., Beaudry, G. A., Ciriello, K. M., Cook, B. P., Dufault, M. R., Ferguson, A. T., Gao, Y., He, T. C., Hermeking, H., Hiraldo, S. K., Hwang, P. M., Lopez, M. A., Luderer, H. F., Mathews, B., Petroziello, J. M., Polyak, K., Zawel, L., Kinzler, K. W., *et al*. Analysis of human transcriptomes. Nat. Genet., *23:* 387–388, 1999.

44. Ye, S. Q., Zhang, L. Q., Zheng, F., Virgil, D., and Kwiterovich, P. O. Minisage: gene expression profiling using serial analysis of gene expression from 1 μg total RNA. Anal. Biochem., *287:* 144–152, 2000.

45. Chen, J. J., Rowley, J. D., and Wang, S. M. Generation of longer cDNA fragments from serial analysis of gene expression tags for gene identification. Proc. Natl. Acad. Sci. USA, *97:* 349–353, 2000.

46. Davidson, J. M., Gorringe, K. L., Chin, S. F., Orsetti, B., Besret, C., Courtay-Cahen, C., Roberts, I., Theillet, C., Caldas, C., and Edwards, P. A. W. Molecular cytogenetic analysis of breast cancer cell lines. Br. J. Cancer, *83:* 1309–1317, 2000.

47. Jacobsen, A., Arnold, N., Weimer, J., and Kiechle, M. Comparison of comparative genomic hybridization and interphase fluorescence *in situ* hybridization in ovarian carcinomas: possibilities and limitations of both techniques. Cancer Genet. Cytogenet., *122:* 7–12, 2000.

48. Masramon, L., Ribas, M., Cifuentes, P., Arribas, R., Garcia, F., Egozcue, J., Peinado, M. A., and Miro, R. Cytogenetic characterization of two colon cell lines by using conventional G-banding, comparative genomic hybridization, and whole chromosome painting. Cancer Genet. Cytogenet., *121:* 17–21, 2000.

49. De Pinho, R. A. The age of cancer. Nature (Lond.), *408:* 248–254, 2000.

50. Eisenberg, R. S. Genomics in the public domain: strategy and policy. Nat. Rev. Genet., *1:* 70–74, 2000.

51. Finkel, E. The post-genome era: medical promise with problems. Lancet, *349:* 1228, 2000.

52. Devine, K. Honing genomics tools. Scientist, *14:* 36, 2000.

53. Gershon, D. Structural genomics: from cottage industry to industrial revolution. Nature (Lond.), *408:* 273–274, 2000.

54. Rowen, L., Wong, G., Lane, R., and Hood, L. Intellectual property. Publication rights in the era of open data release. Science, *289:* 1881, 2000.

55. Long, H., Aldredge, T., Hebert, A., Perrin, S., Smith, M., August, P., Newcomb, R., and Call, K. Automation technology applications in a genomics center. Am. Lab. (Shelton), *32:* 30–35, 2000.

56. Lowe, H. J. Transforming the cancer center in the 21st century. MD Comput., *16:* 40–42, 1999.

57. Thiesing, J. T., Ohno-Jones, S., Kolibaba, K. S., and Druker, B. J. Efficacy of STI571, an abl tyrosine kinase inhibitor, in conjunction with other antileukemic agents against bcr-abl-positive cells. Blood, *96:* 3195–3199, 2000.

58. Vigneri, P., and Wang, J. Y. Induction of apoptosis in chronic myelogenous leukemia cells through nuclear entrapment of BCR-ABL tyrosine kinase. Nat. Med., *7:* 228–234, 2001.

59. Fuller, G. N., Rhee, C. H., Hess, K. R., Caskey, L. S., Wang, R., Bruner, J. M., Yung, W. A., and Zhang, W. Growth factor binding protein 2 expression in glioblastoma multiforme: a revelation by parallel gene expression profiling. Cancer Res., *59:* 4228–4232, 1999.

60. Sallinen, S. L., Sallinen, P. K., Haapasalo, H. K., Helin, H. J., Helen, P. T., Schrami, P., Kallionemi, O. P., and Kononen, J. Identification of differentially expressed genes in human gliomas by DNA microarray and tissue chip techniques. Cancer Res., *60:* 6617–6622, 2000.

61. Clark, E. A., Golub, T. R., Lander, E. S., and Hynes, R. O. Genomic analysis of metastasis reveals an essential role for RhoC. Nature (Lond.), *406:* 532–535, 2000.

62. Parle-McDermott, A., McWilliam, P., Tighe, O., Dunican, D., and Croke, D. T. Serial analysis of gene expression identifies putative metastasis-associated transcripts in colon tumor cell lines. Br. J. Cancer, *83:* 725–728, 2000.

63. Zhou, W., Sokoll, L. J., Bruzek, D. J., Zhang, L., Velculescu, V. E., Goldin, S. B., Hruban, R. H., Kern, S. E., Hamilton, S. R., Chan, D. W., Vogelstein, B., and Kinzler, K. W. Identifying markers for pancreatic cancer by gene expression analysis. Cancer Epidemiol. Biomark. Prev., *7:* 109–112, 1998.

64. St. Croix, B., Rago, C., Velculescu, V., Traverso, G., Romans, K. E., Montgomery, E., Lal, A., Riggins, G. J., Lengauer, C., Vogelstein, B., and Kinzler, K. W. Genes expressed in human tumor endothelium. Science (Wash. DC), *289:* 1197–1202, 2000.

65. Evans, W., and Relling, M. Pharmacogenomics: translating functional genomics into rational therapeutics. Science (Wash. DC), *286:* 487–491, 1999.

66. Garrett, M. D., and Workman, P. Discovering novel chemotherapeutic drugs for the third millennium. Eur. J. Cancer, *35:* 2010–2030, 1999.

67. Roses, A. D. Idiosyncratic reactions to drugs: can medicine response profiles provide a dynamic drug surveillance system? Clin. Chem. Lab. Med., *38:* 815–818, 2000.

68. Rhee, C. H., Ruan, S. B., Chen, S., Chenchik, A., Levin, V., Yung, A., Fuller, G. N., and Zhang, W. Characterization of cellular pathways involved in glioblastoma response to chemotherapeutic agent 1,3-bis(2-chloroethyl)-1-nitrosourea (BCNU) by gene expression profiling. Oncol. Rep., *6:* 393–401, 1999.

69. Wang, K., Cheng, A., Chuang, S., Hsu, H., and Su, I. Retinoic acid-induced apoptotic pathway in T-cell lymphoma. Identification of four groups of genes with differential biological functions. Exp. Hematol. (Charlottesv.), *28:* 1441–1450, 2000.

70. Ismail, R. S., Baldwin, R. L., Fang, J., Browning, D., Karlan, B. Y., Gasson, J. C., and Chang, D. D. Differential gene expression between normal and tumor-derived ovarian epithelial cells. Cancer Res., *60:* 6744–6749, 2000.

71. Castuma, M. V., Rao, P. H., Acevedo, S. H., and Larripa, I. B. Comparative genomic hybridization study of *de novo* myeloid neoplasia. Acta Haematol. (Basel), *104:* 25–30, 2000.

72. Rosenberg, C., van Gurp, R. J. H. L. M., Geelen, E., Osterhuis, J. W., and Looijenga, L. H. J. Overrepresentation of the short arm of chromosome 12 is related to invasive growth of human testicular seminomas and nonseminomas. Oncogene, *19:* 5858–5862, 2000.

73. Goff, L. K., Neat, M. J., Crawlwy, C. R., Jones, L., Jones, E., Lister, T. A., and Gupta, R. K. The use of real-time quantitative polymerase chain reaction and comparative genomic hybridization to identify amplification of the *REL* gene in follicular lymphoma. Br. J. Haematol., *111:* 618–625, 2000.

74. Ojopi, E. P. B., Rogatto, S. R., Caldeira, J. R. F., Barbieri-Neto, J., and Squire, J. A. Comparative genomics hybridization detects novel amplifications in fibroadenomas of the breast. Genes Chromosomes Cancer, *30:* 25–31, 2001.

75. Kim, D. G., Park, S. Y., You, K. R., Lee, G. B., Kim, H., Moon, W. S., Chun, Y. H., and Park, S. H. Establishment and characterization of chromosomal aberrations in human cholangiocarcinoma cell lines by cross-species color banding. Genes Chromosomes Cancer, *30:* 48–56, 2001.

76. Murakami, Y., Nobukuni, T., Tamura, K., Maruyama, T., Sekiya, T., Arai, Y., Gomyou, H., Tanigami, A., Ohki, M., Cabin, D., Frischmeyer, P., Hunt, P., and Reeves, R. H. Localization of tumor suppressor activity important in non-small cell lung carcinoma on chromosome 11q. Proc. Natl. Acad. Sci. USA, *95:* 8153–8158, 1998.

77. Shayesteh, L., Lu, Y., Kuo, W. L., Baldocchi, R., Godfrey, T., Collins, C., Pinkel, D., Powell, B., Mills, G. B., and Gray, J. W. *PIKCA* is implicated as an oncogene in ovarian cancer. Nat. Genet., *21:* 99–102, 1999.

78. Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. Genome-wide analysis of DNA copy number changes using cDNA microarrays. Nat. Genet., *23:* 41–46, 1999.

79. Lichter, P. New tools in molecular pathology-commentary. J. Mol. Diagn., *2:* 171–173, 2000.

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

81. Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science (Wash. DC), *286:* 531–537, 1999.

82. Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., Powell, J. I., Yang, L., Marti, G. E., Moore, T., Hudson, J., Lu, L., Lewis, D. B., Tibshirani, R., Sherlock, G., Chan, W. C., Greiner, T. C., Weisenburger, D. D., Armitage, J. O., Warnke, R., Staudt, L. M., *et al*. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature (Lond.), *403:* 503–511, 2000.

83. Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., Sampas, N., Dougherty, E., Wang, E., Marincola, F., Gooden, C., Lueders, J., Glatfelter, A., Pollock, P., Carpten, J., Gillanders, E., Leja, D., Dietrich, K., Beaudry, C., Berens, M., Alberts, D., and Sondak, V. Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature (Lond.), *406:* 536–540, 2000.

84. Dos Santos, N. R., de Bruijn, D. R., and van Kessel, A. G. Molecular mechanisms underlying human synovial sarcoma development. Genes Chromosomes Cancer, *30:* 1–14, 2001.

85. Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Raffeld, M., Ya-

khini, Z., Ben-Dor, A., Dougherty, E., Kononen, J., Bubendorf, L., Fehrle, W., Pittaluga, S., Gruvberger, S., Loman, N., Johannsson, O., Olsson, H., Wilfond, B., Sauter, G., Kallionemi, O. P., Borg, A., and Trent, J. Gene expression profiles in hereditary breast cancer. N. Engl. J. Med., *244*: 539–548, 2001.

86. Takano, T., Hasegawa, Y., Matsuzuka, F., Miyauchi, A., Yoshida, H., Higashiyama, T., Kuma, K., and Amino, N. Gene expression profiles in thyroid carcinomas. Br. J. Cancer, *83:* 1495–1502, 2000.

87. Ross, D. T., Scherf, U., Eisen, M. B., Perou, C. M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S. S., Van de Rijn, M., Waltham, M., Pergamenschikov, A., Lee, J. C., Lashkari, D., Shalon, D., Myers, T. G., Weinstein, J. N., Botstein, D., and Brown, P. O. Systematic variation in gene expression patterns in human cancer cell lines. Nat. Genet., *24:* 227–235, 2000.

88. Lossos, I. S., Alizadeh, A. A., Eisen, M. B., Chan, W. C., Brown, P. O., Botstein, D., Staudt, L. M., and Levy, R. Ongoing immunoglobulin somatic mutation in germinal center B cell-like but not in activated B cell-like diffuse large cell lymphomas. Proc. Natl. Acad. Sci. USA, *97:* 10209–10213, 2000.

89. Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lonning, P. E., Borresen-Dale, A. L., Brown, P. O., and Botstein, D. Molecular portraits of human breast tumours. Nature (Lond.), *406:* 747–752, 2000.

90. Chen, Z., Tong, J. H., Dong, S., Zhu, J., Wang, Z. Y., and Chen, S. J. Retinoic acid regulatory pathways, chromosomal translocations, and acute promyelocytic leukemia. Genes Chromosomes Cancer, *15:* 147–156, 1996.

91. Chen, G. Q., Shi, X. G., Tang, W., Xiong, S. M., Zhu, J., Cai, X., Han, Z. G., Ni, J. H., Shi, G. Y., Jia, P. M., Liu, M. M., He, K. L., Niu, C., Ma, J., Zhang, P., Zhang, T. D., Paul, P., Naoe, T., Kitamura, K., Miller, W., Waxman, S., Wang, Z. Y., de The, H., Chen, S. J., and Chen, Z. Use of arsenic trioxide (As203) in the treatment of acute promyelocytic leukemia (APL). I. As203 exerts dose-dependent dual effects on APL cells. Blood, *89:* 3345–3353, 1997.

92. Bauman, G. S., Ino, Y., Ueki, K., Zlatescu, M. C., Fisher, B. J., Macdonald, D. R., Stitt, L., Louis, D. N., and Cairncross, J. G. Allelic loss of chromosome 1p and radiotherapy plus chemotherapy in patients with oligodendrogliomas. Int. J. Radiat. Oncol. Biol. Phys., *48:* 825–830, 2000.

93. Thor, A., Berry, D. A., Budman, D., Muss, H., Kute, T., Henderson, I. C., Barcos, M., Cirrincione, C., Edgerton, S., Allred, C., Norton, L., and Liu, E. *erb*B-2, p53 and efficacy of adjuvant therapy in lymph node-positive breast cancer. J. Natl. Cancer Inst. (Bethesda), *90:* 1346–1360, 1998.

94. Burke, J., Davidson, D., and Hide, W. d2 cluster: a validated method for clustering EST and full-length cDNA sequences. Genome Res., *9*: 1135–1142, 1999.

95. Herwig, R., Poustka, A. J., Muller, C., Bull, C., Lehrach, H., and O'Brien, J. Large-scale clustering of cDNA-fingerprinting data. Genome Res., *9:* 1093–1105, 1999.

96. Hartuv, E., Schmitt, A. O., Lange, J., Meier-Ewert, S., Lehrach, H., and Shamir, R. An algorithm for clustering cDNA fingerprints. Genomics, *66:* 249–256, 2000.

97. Taylor, E., Cogdell, D., Coombes, K., Hu, L., Ramdas, L., Hamilton, S. R., and Zhang, W. Sequence verification as quality control step for production of cDNA microarray, in press, 2001.

98. Moch, H., Kononen, T., Kallionemi, O. P., and Sauter, G. Tissue microarrays: what will they bring to molecular and anatomic pathology? Adv. Anat. Pathol., *8:* 14–20, 2001.

99. Kononen, J., Bubendorf, L., Kallionemi, A., Barlund, M., Schraml, P., Leighton, S., Tothorst, J., Mihatsch, M. J., Sauter, G., and Kallionemi, O. P. Tissue microarrays for high-throughput molecular profiling of tumor specimens. Nat. Med., *4:* 844–847, 1998.

100. Myakishev, M. V., Khripin, Y., Hu, S., and Hamer, D. H. High-throughput SNP genotyping by allele-specific PCR with universal energy transfer-labeled primers. Genome Res., *11:* 163–169, 2001.

101. Van Schie, R. C., Marras, S. A., Conroy, J. M., Nowak, N. J., Catanese, J. J., and de Jong, P. J. Semiautomated clone verification by real-time PCR using molecular beacons. Biotechniques, *29:* 1296–1300, 1302–1304, 1306 passim, 2000.

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

103. Banks, R. E., Craven, R., Harnden, P., and Selby, P. J. The use of laser capture microdissection to obtain distinct populations of cells for proteomic analysis. JMM, *78:* B28, 2000.

104. Phillips, J., and Eberwine, J. H. Antisense RNA amplification. A linear amplification method for analyzing the mRNA population from single living cells. Methods (Duluth), *10:* 283–288, 1996.

105. Wang, E., Miller, L. D., Ohnmacht, G. A., Liu, E. T., and Marincola, F. M. High-fidelity mRNA amplification for gene profiling. Nat. Biotechnol., *18:* 457–459, 2000.

106. Searls, D. B. Bioinformatics tools for whole genomes. Annu. Rev. Genomics Hum. Genet., *1:* 251–279, 2000.

107. Meyer, D., and Cook, D. Visualization of data. Curr. Opin. Biotechnol., *11*: 89–96, 2000.

108. Kuklin, A., Shah, S., Hoff, B., and Shams, S. Information processing issues and solutions associated with microarray technology. Lab. Robotics Automation, *12:* 317–327, 2000.

109. Hagen, J. B. The origins of bioinformatics. Nat. Rev. Genet., *1:* 231–236, 2000.

110. Chen, Y., Dougherty, E. R., and Bittner, M. L. Ratio-based decisions and the quantitative analysis of cDNA microarray images. J. Biomed. Optics, *24:* 364–374, 1997.

111. Ermolaeva, O., Rastogi, M., Pruitt, K. D., Schuler, G. D., Bittner, M. L., Chen, Y., Simon, R., Meltzer, P., Trent, J. M., and Boguski, M. S. Data management and analysis for gene expression arrays. Nat. Genet., *20:* 19–23, 1998.

112. Aldhous, P. Genomics: beyond the book of life. Nature (Lond.), *408:* 894–896, 2000.

113. Hess, K., Zhang, W., Baggerly, K., Stivers, D., and Coombes, K. Microarrays: handling the deluge of data and extracting reliable information, in press, 2001.

114. Black, D. L. Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology. Cell, *103:* 367–370, 2000.

115. Epstein, C., and Butow, R. Microarray technology-enhanced versatility, persistent challenge. Curr. Opin. Biotechnol., *11:* 36–41, 2000.

116. Shapiro, L., and Harris, T. Finding function through structural genomics. Curr. Opin. Biotechnol., *11:* 31–35, 2000.

117. Weinstein, J. N. Fishing expeditions. Science (Wash. DC), *282:* 628–629, 1998.

118. Asthuler, D., Daly, M., and Kruglyak, L. Guilt by association. Nat. Genet., *26:* 135–137, 2000.

119. Savage, J. R. Cancer. Enhanced: proximity matters. Science, *290:* 62–63, 2000.

120. Hadden, S. How much use is the human genome project? Nature, *404:* 541–542, 2000.

121. Lander, E. S., and Weinberg, R. A. Genomics: journey to the center of biology. Science (Wash. DC), *287:* 1777–1782, 2000.

122. Lander, E. S. Scientific commentary: the scientific foundations and medical and social prospects of the human genome project. J. Law Med. Ethics, *26:* 178, 184–188, 1998.

123. Childs, B., and Valle, D. Genetics, biology and disease. Annu. Rev. Genomics Hum. Genet., *1:* 1–19, 2000.

124. Li, M. Applications of display technology in protein analysis. Nat. Biotechnol., *18:* 1251–1256, 2000.

125. Fung, E. T., Thulasiraman, V., Weinberger, S. R., and Dalmasso, E. A. Protein biochips for differential profiling. Curr. Opin. Biotechnol., *12:* 65–69, 2001.