The use of microarrays to determine gene expression patterns is providing a wealth of new information that should aid in cancer diagnosis and ultimately in therapy

DNA Arrays Reveal Cancer in Its Many Forms

Over the years, cancer researchers and clinicians have learned that they are fighting not just one foe but a seemingly endless variety of adversaries. Indeed, the catchall name "cancer" hides the fact that there are actually hundreds of different cancers—each with its own unique characteristics. That's a major reason why cancer has been so hard to vanquish: Drugs that work against a cancer of one tissue—say, breast, or lung, or blood—often don't work against those of others. Even cancers of the same apparent type often vary widely in their responsiveness to a therapy; some retreat quickly while others relentlessly progress. Now, researchers have a powerful new tool that not only should help them sort out the differences that define the many types of cancer, but also should help identify new targets for therapeutic drugs.

The tool is the microarray—a slide or chip systematically dotted with DNA from thousands of genes that can serve as probes for detecting which genes are active in different types of cells (*Science*, 15 October 1999, p. 444). Researchers doing the work say such arrays are providing an unprecedented amount of information about the genetic changes underlying cancer. "When we started out, we like all other biologists were used to studying one gene at a time," says Paul Meltzer of the National Human Genome Research Institute (NHGRI) in Bethesda, Maryland. "When suddenly you don't have to do that, it changes things radically." Richard Klausner, director of the National Cancer Institute (NCI) and a proponent of using microarrays to study cancer, agrees. This "really represents a new type of data," he says.

Microarray technology is already providing insights into cancer that would be difficult, if not impossible, to obtain using the gene-by-gene approach. In the past several months, researchers in several labs have used it to identify specific subtypes of a variety of cancers, including leukemias and lymphomas, the dangerous skin cancer melanoma, and breast cancer. In some cases, they can determine which cancers are likely to respond to current therapies and which aren't. Such information, predicts NCI's Louis Staudt, "will rewrite the cancer textbooks over the next 3 to 4 years." In addition, the studies are giving researchers a fix on which genes are important for the development, maintenance, and spread of the various cancers, and are thus possible drug targets.

Subdividing the enemy

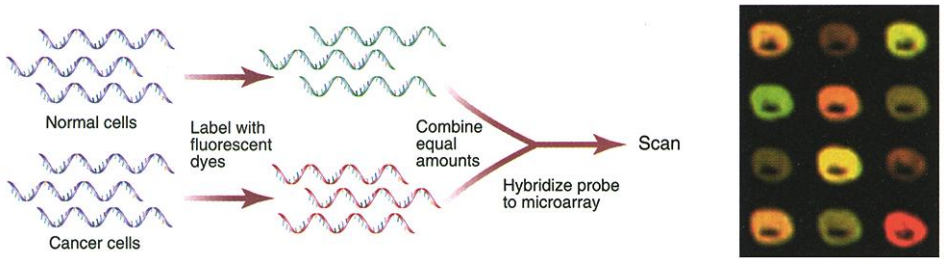
An early demonstration that it's possible to classify cancers based on their gene expression profiles came from a team led by Eric Lander and Todd Golub of the Whitehead Institute and the Massachusetts Institute of Technology (MIT) Center for Genome Research (*Science*, 15 October 1999, p. 531). They began by comparing the profiles of acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL), two blood cancers that are often hard to tell apart by standard pathological examination of the diseased cells. The researchers isolated messenger RNAs (mRNAs), which are produced only by active genes, from the bone marrow cells of 38 patients with either AML or ALL; labeled the mRNAs with biotin; and then applied each sample separately to microarray chips carrying the probes for more than 6800 human genes that were prepared by the biotech firm Affymetrix of Santa Clara, California.

After scanning the chips to determine how much mRNA had bound to each gene, the team used an algorithm to select the 50 genes whose level of expression differed most between AML and ALL cells. They selected the genes mathematically, "without human intervention," Golub says, "because we started with the notion that we weren't smart enough to know which were informative."

The researchers then went back and showed that the expression patterns of those genes could in fact identify which patients had AML and which had ALL in the original group of 38 and also in another group of 36 patients not previously studied. Although cancer specialists already knew that AML and ALL are separate diseases, "it was important to show that you could [tell them apart] by gene expression," says Jeff Trent of the NHGRI.

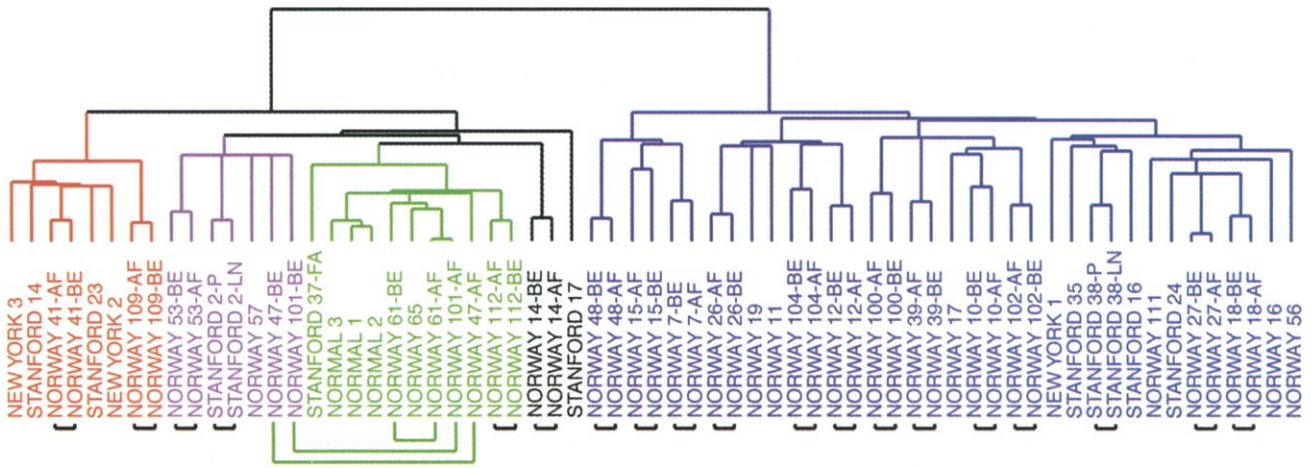
Since then, researchers have used gene expression patterns to reveal previously unknown cancer categories. Some of this work comes from Staudt's team, working in collaboration with Patrick Brown and David Botstein of Stanford University School of Medicine and their colleagues. These researchers focused on patients with diffuse large B cell lymphoma, a common type of non-Hodgkin's lymphoma that affects more than 15,000 new patients annually in the United States and follows a highly variable clinical course. "From the first," Staudt says, "it was clear that 40% did wonderfully and were cured, while 60% succumbed to the disease." The microarray analysis points to a possible reason why.

In the first phase of the work, which was reported in the 3 February issue of *Nature*, the researchers created what they call a "Lymphochip" containing nearly 18,000 genes, most of which are expressed in normal



A shade of difference. In some experiments, cDNAs copied from the mRNAs of cancer cells are given a red fluorescent label, while those from normal cells get a green label. When equal amounts of the two cDNAs are applied to an array, red indicates higher expression of a gene in cancer cells, green denotes higher expression in normal cells, and yellow shows that the expression levels are the same.

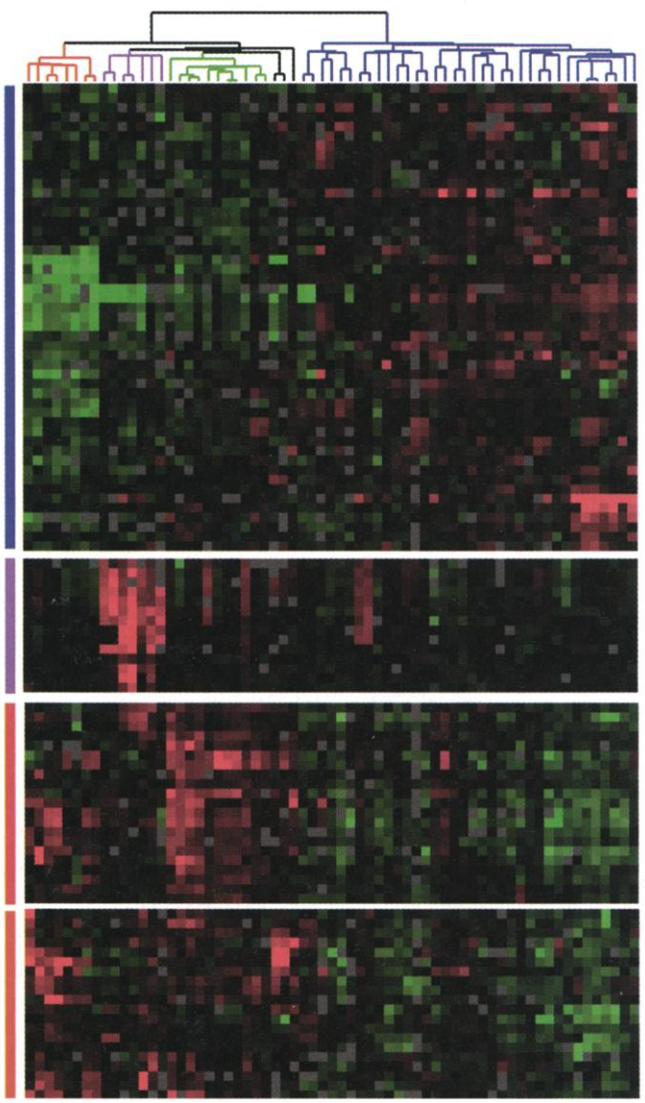
CREDIT: NHGRI



and malignant lymphoid cells. They then prepared mRNAs from biopsy samples from 40 lymphoma patients, copied them into DNAs that were fluorescently labeled, and then applied the cDNAs from each patient to a Lymphochip. "We found a great diversity in gene expression among the patients, despite [their having] the same diagnosis," Staudt says.

Computer analysis of the expression patterns showed, however, that the patients could be divided into two groups. One group expressed a set of genes characteristically turned on in B cells in the spleen and lymph nodes during an immune response. The other set didn't express those genes but did show activity of a set of genes that are turned on when blood B cells are stimulated to divide by an antigen. "On this basis," Staudt says, "the patients could be thought of as having two different diseases." Indeed, their clinical pictures also varied: Those with the expression pattern of the spleen–lymph node B cells fared much better, with 75% alive 5 years after diagnosis, while 75% of the other group did not make it to that milestone.

Staudt's team is now participating in a collaborative study called the Lymphoma/Leukemia Molecular Profiling Project, which will look at hundreds of patients with B cell lymphoma to see whether the results hold up. But they already have possible therapeutic implications. Lymphoma patients are usually treated first with chemotherapy, and if they relapse, they become candidates for a bone marrow transplant. But in the future, those patients who have a gene expression profile indicating a poor prognosis might move directly to bone marrow transplant, avoiding the first-line chemotherapy regimen, which can be debilitating.



CREDIT: C. M. PEROU *ET AL.*, *NATURE* **406**, 747 (2000)

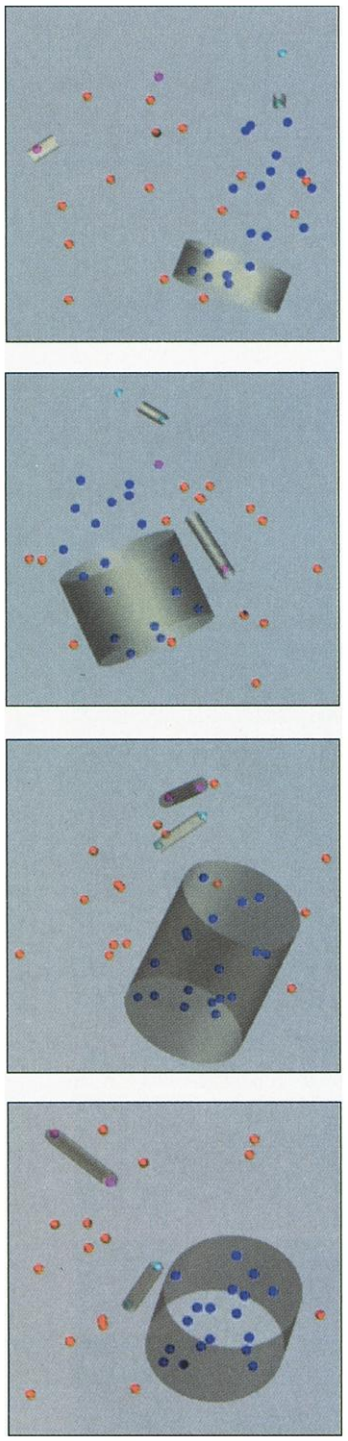
Tumor subdivision. The gene-expression patterns of human breast cancers showed that they could be grouped into an estrogen-receptor-positive branch (blue) and an estrogen-receptor-negative branch, which could be further subdivided. The tumors in the orange cluster have a pattern similar to that of the basal cells of the breast ducts, while the pink cluster is characterized by high expression of the *Erb-B2* oncogene, and the green denotes tumors with a pattern more like that of normal breast tissue.

Blood cancers aren't the only ones in which microarray analysis is picking up previously undetected subgroups. In the 3 August issue of *Nature*, the NHGRI team reported that 31 melanoma patients could be subdivided into two distinct groups based on their gene expression patterns, even though there were no obvious pathological distinctions between the patients' tumors. It's too soon to tell whether these categories are correlated with clinical outcome, NHGRI's Meltzer says, although there are hints that they might be.

For example, melanomas from the larger of the two clusters show reduced expression of a variety of genes involved in cell movement and, consistent with that observation, the tumor cells show reduced motility. That might mean they were less invasive. An indication that that might be the case comes from the fact that only three of the 10 patients in that cluster for whom survival data was known had died, compared with four out of five in the smaller cluster. Because the fates of some patients are unknown, however, the result might not be significant.

Stanford's Brown and Botstein and their colleagues have also found that breast cancers show distinguishable patterns of gene expression. Again, the researchers could pick out two broad groups of the tumors, one of which was

marked by expression of the gene for the receptor for the hormone estrogen while the other one wasn't. That was no surprise, because cancer physicians have long known that breast cancer cells that lack estrogen receptors tend to be more aggressive. But as the researchers reported in the 17 August issue of *Nature*, those broad groups of tumors may contain a variety of subgroups. For example, the branch containing the estrogen-receptor-negative cells included one subgroup showing a gene expression pattern similar to that of the basal cells of the mammary ducts and another characterized by high expression of the *Erb-B2* oncogene. "Not described," Botstein says, "is whether there is a difference in outcomes [for the subgroups]. We're addressing that now."



Another view. The figure includes four still frames from a movie (also see www.nhgri.nih.gov/DIR/Microarray) showing that 19 melanoma tumors (blue, within cylinder) remain grouped together in a three-dimensional multiscaling plot that reflects the degree of correlation of their gene-expression patterns. The other tumors never enter the cylinder space.

Cause and effect

These early studies clearly show that it's now possible to detect wholesale changes in gene expression patterns, but researchers want to do more than just identify genes whose activity is turned up or down. They also want to find out which of those changes are important for cancer development and progression—the causes and not just the effects. Microarrays are helping out there, too. For example, also in the 3 August issue of *Nature*, Golub and Lander, with MIT colleagues Edwin Clark and Richard Hynes, used arrays to compare the gene expression patterns of highly metastatic melanoma cells with those of the much less metastatic cells from which they were derived. The comparison identified a suite of genes whose activity was apparently turned up as melanoma cells progressed to malignancy.

Many of the genes were of the type that would be expected to contribute to metastasis—genes involved either directly or indirectly in the cell's ability to move to and invade new tissues, for example. The MIT team looked in further detail at one such gene, called *RhoC*, because work by others had shown that expression of this gene correlates with progression of a pancreatic cancer to metastasis. When the researchers introduced the human *RhoC* gene into a line of human melanoma cells that normally shows little tendency to spread and then inoculated mice with these cells, they found that the cells had become highly metastatic.

In a similar fashion, researchers are using the arrays to determine how the activation of cancer-promoting oncogenes or the inactivation of tumor-suppressor genes perturbs the expression of other genes. As NCI's Klausner says, the arrays provide "a tool for understanding the relation between specific genetic defects and how they play out." One recent example comes from Staudt and his colleagues, who used their Lymphochip to study the consequences of abnormal activation of an oncogene called *BCL-6*, a situation that commonly occurs in lymphomas. Although researchers knew that *BCL-6*'s protein product represses the expression of certain genes, the NCI team's microarray analysis pinpointed the changes that could be leading to cancer.

They found that *BCL-6* activation leads to repression of a gene called *blimp-1*, which normally promotes the differentiation of B cells to become antibody-producing plasma cells, and also of a gene called *p27kip1*, which inhibits the cell division cycle. This two-part repression is effectively a double whammy when it comes to cancer development, because the net result is to lock cells in an undifferentiated, continuously dividing state. (The results appeared in the August issue of *Immunity*.)

The MIT group, in collaboration with Robert Eisenman's team at the Fred Hutchinson Cancer Research Center in Seattle, has also looked at the changes elicited in cells by activation of the *MYC* oncogene. In results reported in the 28 March issue of the *Proceedings of the National Academy of Sciences*, they found that *MYC* activity turned up the expression of 27 genes, including some that promote cell division, and turned down the activity of another nine. "We're discovering the malignant pathways of these tumors," Staudt says. "Now we can ask whether interfering with these pathways can help."

Other applications of microarray technology to cancer that are also getting under way include studying how cancer cells respond to various chemotherapeutic agents and determining why some cells respond and some don't. "The applications are almost endless," enthuses NHGRI's Meltzer.

A flood of data

As the current trickle of microarray studies swells to what is likely to be a flood, researchers will have to face what Klausner calls a "profound" challenge: how to deal with the masses of data that will come pouring out. "It's not just how to analyze the data," Klausner says, "but how to share and compare them." Right now, for example, people are using different "platforms," as the arrays are called, as well as different methods of analyzing the data those platforms produce. This lack of standardization makes it difficult to relate the findings of the different labs and assess their quality.

To deal with such issues, NCI has set up a Gene Expression Analysis Working Group, which will hold several workshops over the coming years similar to those held by the human genetics community to sort out the problems their field faced. The first workshop is in the planning stage, says NCI's Kenneth Buetow, and the questions it will deal with are still being worked out. "It's actually quite easy to find issues," he says. "The hardest part is to identify where to start."

But it's important for the community to get together and solve the challenges so that it can get a better understanding of the changes that produce cancer and how to counteract them. "It's an amazing time," Buetow says. "We're all having fun [with microarrays]. But at the end of the day, we have to translate this into clinical benefit."

—JEAN MARX

CREDIT: *NATURE* 406, 536 (2000)