

THE MAGIC OF MICROARRAYS

Research tools known as DNA microarrays are already clarifying the molecular roots of health and disease and speeding drug discovery. They could also hasten the day when custom-tailored treatment plans replace a one-size-fits-all approach to health care

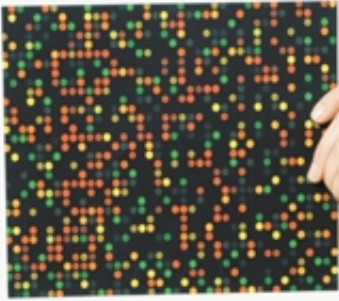
BY STEPHEN H. FRIEND
AND ROLAND B. STOUGHTON

DOT PATTERNS EMERGE when DNA microarrays analyze tissue samples. Individual differences in those patterns could one day help doctors match treatments to the unique needs of each patient.



MOST PEOPLE STRICKEN with a cancer called diffuse large B cell lymphoma initially respond well to standard therapy. Yet in more than half of cases, the cancer soon roars back lethally. Physicians have long assumed that the reason some individuals succumb quickly while others do well is that the disease actually comes in different forms caused by distinct molecular abnormalities. But until two years ago, investigators had no way to spot the patients who had the most virulent version and thus needed to consider the riskiest, most intensive treatment.

Then a remarkable tool known as a DNA microarray, or DNA chip, broke the impasse. It enabled a team of researchers from the National Institutes of Health, Stanford University and elsewhere to distinguish between known long- and short-term survivors based on differences in the overall pattern of activity exhibited by



hundreds of genes in their malignant cells at the time of diagnosis. That achievement should lead to a diagnostic test able to identify the patients in greatest danger.

DNA microarrays, first introduced commercially in 1996, are now mainstays of drug discovery research, and more than 20 companies sell them or the instruments or software needed to interpret the information they provide. The devices are also beginning to revolutionize how scientists explore the operation of normal cells in the body and the molecular aberrations that underlie medical disorders. The tools promise as well to pave the way for faster, more accurate diagnoses of many conditions and to help doctors personalize medical care—that is, tailor therapies to the exact form of disease in each person and select the drugs likely to work best, with the mildest side effects, in those individuals.

Tiny Troupers

THE ARRAYS COME IN several varieties, but all assess the composition of genetic material in a tissue sample, and all consist of a lawn of single-stranded DNA molecules (probes) that are tethered to a wafer often no bigger than a thumbprint. These chips also capitalize on a very handy property of DNA: complementary base pairing.

DNA is the material that forms the more than 30,000 genes in human cells—the sequences of code that constitute the blueprints for proteins. It is built from four building blocks, usually referred to by the first letter of their distinguishing

chemical bases: A, C, G and T. The base A in one strand of DNA will pair only with T (A's complement) on another strand, and C will pair only with G.

Hence, if a DNA molecule from a tissue sample binds to a probe having the sequence ATCGGC, an observer will be able to infer that the molecule from the sample has the complementary sequence: TAGCCG. RNA, which is DNA's chemical cousin, also follows a strict base-pairing rule when binding to DNA, so the sequence of any RNA strand that pairs up with DNA on a microarray can be inferred as well.

Complementary base-pairing reactions have been integral to many biological tests for years. But amazingly, DNA microarrays can track tens of thousands of those reactions in parallel on a single chip. Such tracking is possible because each kind of probe—be it a gene or a shorter sequence of code—sits at an assigned spot within a checkerboardlike grid on the chip and because the DNA or RNA molecules that get poured over the array carry a fluorescent tag or other label that can be detected by a scanner. Once a chip has been scanned, a computer converts the raw data into a color-coded readout.

Scientists rely on DNA microarrays for two very different purposes. So-called genotype applications compare the DNA on a chip with DNA in a tissue sample to determine which genes are in the sample or to decipher the order of code letters in as yet unsequenced strings of DNA. Frequently, however, investigators these days

use the devices to assess not merely the presence or sequence of genes in a sample but the expression, or activity level, of those genes. A gene is said to be expressed when it is transcribed into messenger RNA (mRNA) and translated into protein. Messenger RNA molecules are the mobile transcripts of genes and serve as the templates for protein synthesis.

Gene Hunters

RESEARCHERS have employed the genotype approach to compare the genes in different organisms (to find clues to the evolutionary history of the organisms, for example) and to compare the genes in tumors with those in normal tissues (to uncover subtle differences in gene composition or number). One day gene comparisons performed on DNA chips could prove valuable in medical practice as well.

Carefully designed arrays could, for instance, announce the precise cause of infection in a patient whose flulike symptoms (such as aches, high fever and breathing difficulty) do not point to one clear culprit. A surface could be arrayed with DNA representing genes that occur only in selected disease-causing agents, and a medical laboratory could extract and label DNA from a sample of infected tissue (perhaps drawn from the person's nasal passages). Binding of the patient's DNA to some gene sequence on the chip would indicate which of the agents was at fault. Similarly, chips now being developed could signal that bioterrorists have released specific types of anthrax or other exotic germs into a community.

For better or worse, gene-detecting microarrays could also identify an individual's genetic propensity to a host of disorders. Most genetic differences in people probably take the form of single nucleotide polymorphisms, or SNPs (pronounced "snips"), in which a single DNA letter substitutes for another. A chip bearing illness-linked gene variants could be constructed to reveal an individual's SNPs and thus predict the person's likelihood of acquiring Alzheimer's disease, diabetes, specific cancers and so on. Those people at greatest risk could then receive close

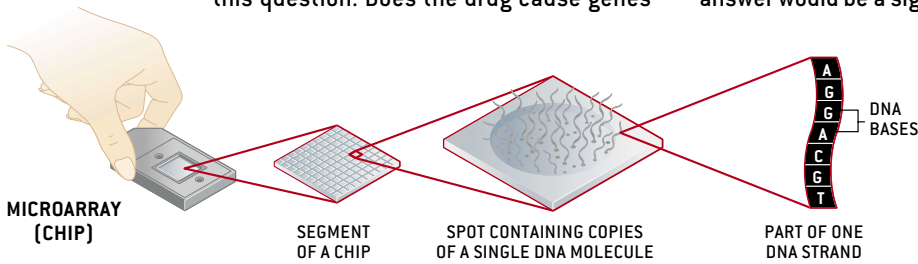
Overview/Microarrays

- DNA microarrays, also known as DNA or gene chips, can track tens of thousands of molecular reactions in parallel on a wafer smaller than a microscope slide. The chips can be designed to detect specific genes or to measure gene activity in tissue samples.
- These properties are proving immensely valuable to cell biologists, to scientists who study the roots of cancer and other complex diseases, and to drug researchers. Microarrays are also under study as quick diagnostic and prognostic tools.
- Protein arrays, which have great promise as diagnostic devices and as aids to biological research, are being developed as well.
- The research and diagnostic information provided by DNA chips and protein arrays should eventually help physicians provide highly individualized therapies.

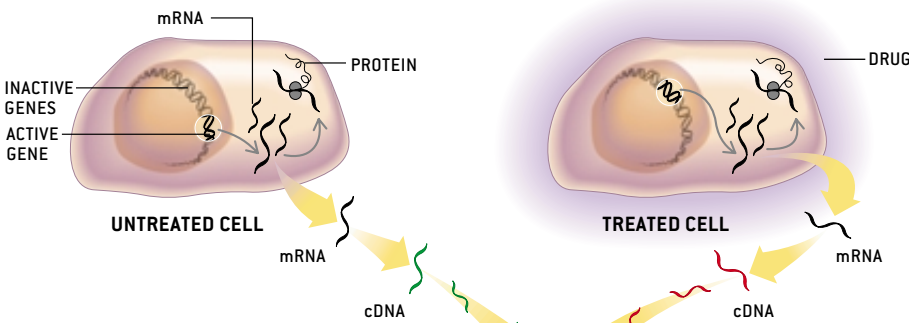
HOW ARRAYS WORK

TO DETERMINE QUICKLY whether a potential new drug is likely to harm the liver, a researcher could follow the steps below, asking this question: Does the drug cause genes

(the blueprints for proteins) in liver cells to alter their activity in ways that are known to cause or reflect liver damage? A “yes” answer would be a sign of trouble.



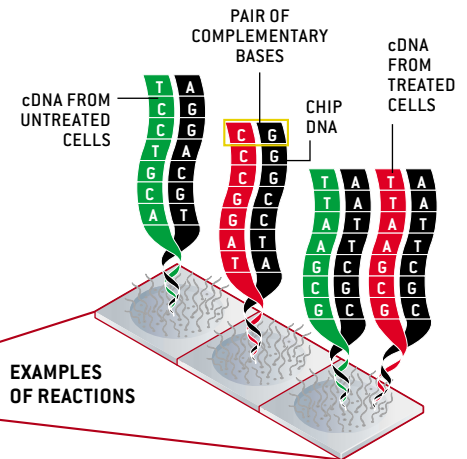
1 Construct or buy a microarray, or chip, containing single-stranded DNA representing thousands of different genes, each assigned to a specified spot on the one-by-three-inch or smaller device. Have every spot include thousands to millions of copies of a DNA strand.



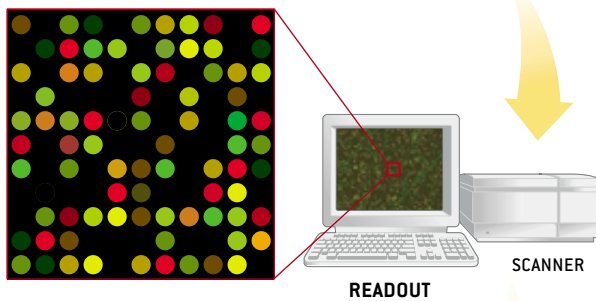
2 Obtain two samples of liver cells; apply the drug to one sample. Then, from each sample, collect molecules of messenger RNA (mRNA)—the mobile copies of genes and the templates for protein synthesis in cells.

3 Transcribe the mRNA into more stable complementary DNA (cDNA) and add fluorescent labels—green to cDNAs derived from untreated cells, red to those from treated cells.

4 Apply the labeled cDNAs to the chip. Binding occurs when cDNA from a sample finds its complementary sequence of bases on the chip (detail at right). Such binding means that the gene represented by the chip DNA was active, or expressed, in the sample.

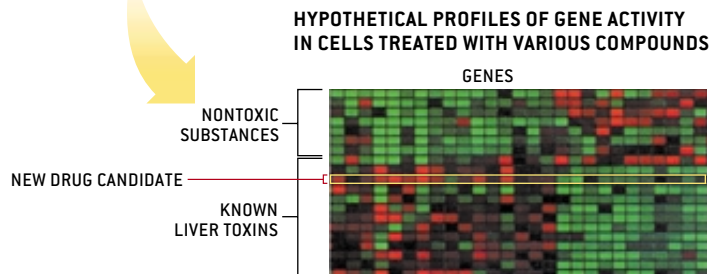


- GENE THAT STRONGLY INCREASED ACTIVITY IN TREATED CELLS
- GENE THAT STRONGLY DECREASED ACTIVITY IN TREATED CELLS
- GENE THAT WAS EQUALLY ACTIVE IN TREATED AND UNTREATED CELLS
- GENE THAT WAS INACTIVE IN BOTH GROUPS



5 Put the chip in a scanner. Have a computer calculate the ratio of red to green at each spot (to quantify any changes in gene activity induced by the drug) and generate a color-coded readout.

6 Determine whether any genes responded strongly to the drug in ways known to promote or reflect liver damage. Or compare the overall expression pattern produced by strong responders with the patterns produced when those genes react to known liver toxins (right). Close similarity would indicate that the new candidate was probably toxic as well. In the diagram, each box represents a single gene’s response to a compound.



monitoring, intensive preventive care and early intervention. Whether these kinds of tests would appeal to the public is an open question, though; the downside of such knowledge can be increased anxiety and the potential for discrimination by employers and insurers.

Other valuable information provided by SNP chips would pose no threat to people's mental state, employability or insurability. The gene variants we possess influence how our bodies process the medicines we take, which in turn influences the effectiveness of the drugs and the intensity of their side effects. Chips that highlighted our unique genetic sensitivities would help physicians choose the drugs that work best and pose the fewest dangers in each of us. SNP chips displaying genetic mutations that increase the aggressiveness of tumors might also help pathologists determine whether benign-looking tumors are actually fiercer than they seem based on microscopic analyses. Both types of arrays are already being investigated for use in medical care.

Choice Expressions

AS EXCITING AS such applications are, it is the other major use of arrays—expression profiling—that has increasingly captivated researchers over the past few years. Laboratory workers produce these profiles by measuring the amounts of different mRNAs in a tissue sample. Generally, the more copies of mRNA a cell makes, the more copies of protein it will make, so the quantities of the various mRNAs in a sample can indirectly indicate the types and amounts of proteins present. Proteins are often of interest because they control and carry out most activities in our bodies' cells and tissues. Chips that directly measure protein levels are being developed [see box on page 52], but constructing them remains challenging.

By using the genome as a sensor pad to detect activity changes in a cell's various genes, scientists can gain exquisitely detailed "snapshots" of how a cell's functions have been altered by drugs or disease states. At times, knowing the overall on-off pattern of gene activity in a sam-

ple can actually be more useful than knowing which particular genes turn on and off in response to some influence. In those cases, as will be seen, the pattern serves as a shorthand "signature" reflecting the molecular state of a sample under some specific condition.

Expression profiling has proved invaluable on many fronts. Cell biologists like it because knowledge of the proteins that predominate after a tissue is exposed to different conditions can provide insight into how the tissue normally compensates for disruptions and what goes wrong when diseases develop.

These scientists are also using expression arrays to learn the functions of genes that have been discovered as a result of the recent sequencing of nearly all the DNA in the nucleus of the human cell. Several techniques that do not involve microarrays can reveal the jobs performed by newly discovered genes (or, more

properly, by the proteins those genes encode), but those approaches do not always work well or quickly. In what has come to be called the guilt-by-association application, expression arrays can help fill in the blanks, even in the absence of any prior clues to a gene's role in the body.

This method derives from the awareness that no gene is an island. If genes in a tissue switch on and off together in response to some influence—say, a drug, an infection or an induced gene mutation—workers can surmise that those like-acting genes operate in the same regulatory pathway; that is, the genes work together or in series to induce a cellular response. Investigators can reasonably guess, then, that the jobs of any originally mysterious genes in the group resemble those of genes whose responsibilities are already understood.

Drug Discovery Tools

DRUG RESEARCHERS, too, take advantage of the guilt-by-association method—to discover proteins not previously known to operate in biological pathways involved in diseases. Once those proteins are found, they can be enlisted as targets for the development of new and better medicines.

In one example, Peter S. Linsley, our colleague at Rosetta Inpharmatics, wanted to identify fresh targets for drugs that might combat inflammatory illnesses, in which the immune system perversely damages parts of the body. He therefore asked which genes in white blood cells of the immune system increase and decrease their protein production in parallel with the gene for a protein called interleukin-2 (IL-2), which is strongly implicated in inflammatory disorders.

He got the answer by producing expression profiles for white blood cells exposed to various chemicals and then having a computer run a sophisticated pattern-matching program to pinpoint a set of genes that consistently switched on or off when the IL-2 gene was activated. This set included a gene whose function in the body had not been determined by other means. At about the same time, investigators at the Pasteur Institute in

AN ARRAY OF COMPANIES

The following are just some of the companies that sell or are developing array-related products and services:

DNA MICROARRAYS

Affymetrix, Santa Clara, Calif.
 Agilent Technologies, Palo Alto, Calif.
 Amersham Biosciences, Piscataway, N.J.
 Axon Instruments, Union City, Calif.
 BioDiscovery, Marina del Rey, Calif.
 Clontech, Palo Alto, Calif.
 Genomic Solutions, Ann Arbor, Mich.
 Mergen, San Leandro, Calif.
 Motorola Life Sciences, Northbrook, Ill.
 Nanogen, San Diego, Calif.
 Partek, St. Peters, Mo.
 PerkinElmer, Boston, Mass.
 Rosetta Inpharmatics, Kirkland, Wash.
 Spotfire, Cambridge, Mass.
 Virtek Vision International, Ontario, Canada

PROTEIN ARRAYS

Biacore International, Uppsala, Sweden
 Biosite Diagnostics, San Diego, Calif.
 CIPHERGEN, Fremont, Calif.
 Large Scale Biology, Germantown, Md.

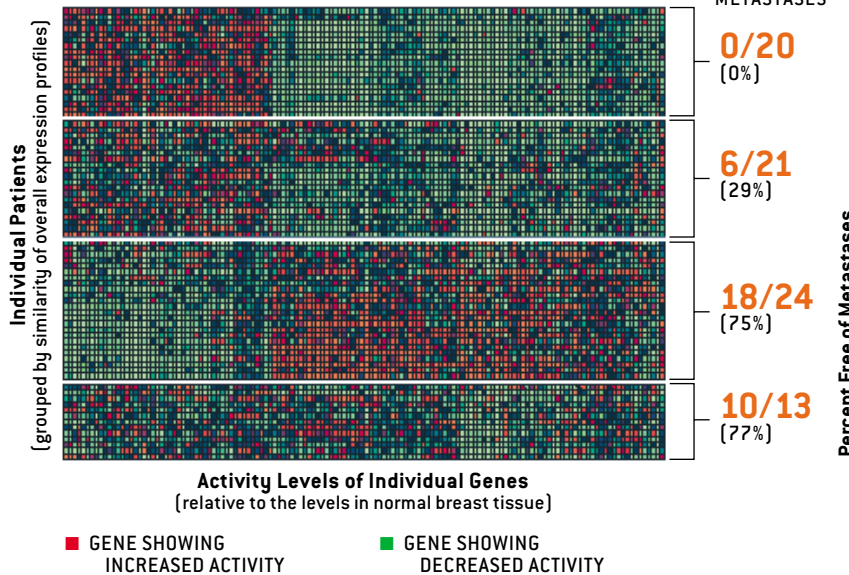
PREDICTING CANCER'S COURSE

WORK AT ROSETTA INPHARMATICS and the Netherlands Cancer Institute suggests that microarrays can help distinguish cancer patients with different prognoses. After determining the activity (expression) levels of genes in small, localized breast tumors from young women who were followed for at least five years after surgery, the researchers found that the expression profiles—the overall patterns of activity across a selection of genes in the

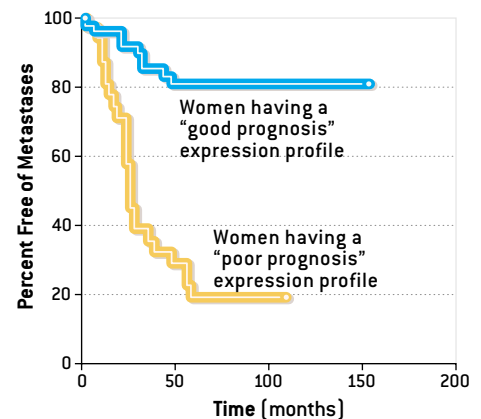
tumors—differed among the patients (left). A mathematical analysis (right) then revealed that patients whose expression profiles resembled a “poor prognosis” signature (the average pattern in tumors that metastasized) were much more likely to suffer a quick recurrence than were patients whose profiles resembled a “good prognosis” signature (the typical pattern in tumors that did not spread). If such results are confirmed by others,

doctors may one day be able to discern which patients need the most intensive therapy based in part on how closely their expression profiles match a standard good or poor prognosis profile.

EXPRESSION PROFILES



PATIENTS' FATES



Paris independently confirmed, with other methods, that this gene operates in the IL-2 pathway. Together the findings suggest that the protein encoded by the gene could be a good target for anti-inflammatory drugs.

Pharmaceutical scientists use expression profiling in a different way: to pick out—and eliminate—drug candidates that are likely to produce unacceptable side effects. Workers who want to determine whether a given compound could damage the heart, for example, can compile a compendium of expression profiles for heart cells exposed to existing drugs and other chemicals. If they also treat heart cells with the drug candidate under study, they can ask a computer to compare the resulting signature with those in the compendium. A signature matching those produced by substances already known to disrupt cardiac cells would raise a red flag.

A compendium of expression profiles can also help explain why a drug pro-

duces particular side effects. A pressing question today, for instance, is why protease inhibitors, which are lifesavers to people infected with HIV (the virus that causes AIDS), can lead to high cholesterol and triglyceride levels in the blood, strange redistributions of body fat, and insulin resistance. Aware that the liver influences the production and breakdown of lipids (the group that includes cholesterol and triglycerides) and of lipid-containing proteins, we and others at Rosetta, in collaboration with Roger G. Ulrich and his team at Abbott Laboratories, de-

cid to see whether one protease inhibitor—ritonavir—produced some of its side effects by acting on the liver.

With an array representing about 25,000 rat genes, we produced expression profiles of rat liver tissue exposed to an assortment of compounds that can be toxic to the liver. After that, we grouped the compounds according to similarities of expression signatures across some 2,400 genes that responded strongly to those substances. Next we delivered ritonavir to rat livers and compared the resulting expression profiles with those generated earlier.

THE AUTHORS

STEPHEN H. FRIEND and ROLAND B. STOUGHTON are colleagues at Rosetta Inpharmatics in Kirkland, Wash., which was founded in 1996 to develop molecular profiling methods involving computers and DNA microarray technology. Merck & Co. acquired the company last year. Friend is vice president of basic research at Merck and president of Rosetta. He was a pediatric oncologist and molecular biologist at Harvard University before becoming director of molecular pharmacology at the Fred Hutchinson Cancer Research Center and co-founding Rosetta. Stoughton, who has a Ph.D. in physics, is senior vice president for informatics at Rosetta. Before turning his attention to biotechnology, he worked on developing signal-processing and pattern-recognition tools for geophysics and astrophysics.

Protein Arrays—A New Option

by N. Leigh Anderson and Gunars Valkirs

LIKE DNA MICROARRAYS, protein-based chips—which array proteins instead of DNA molecules on a small surface—can measure the levels of proteins in tissues. In fact, they do the job more directly and, some evidence says, more accurately. Protein arrays also stand alone in being able to reveal which of thousands of proteins in a tissue interact with one another.

All these properties make protein arrays quite appealing to biological researchers. But the average person would most likely be intrigued for a different reason. Hope is high that such chips will dramatically expand the number of conditions that doctors can diagnose quickly in their offices.

These devices should be very useful as diagnostic tools in part because, unlike DNA microarrays, they can glean information from blood plasma, which is easy to obtain. Most medical disorders—from infectious diseases to heart or kidney damage—leave identifiable traces in the blood, in the form of secreted or leaked proteins. Moreover, in a single test, the arrays might measure many or all of the proteins known to flag the presence of medical problems. In contrast, standard diagnostic tests detect only one or a few disease-specific proteins at a time.

The design of protein arrays resembles that of DNA chips. Hundreds to thousands of distinct proteins sit (in millions of copies) at specified spots in a grid on a wafer-thin plate. Binding of proteins from a blood sample to proteins on a chip reveals the nature and quantities of the sample proteins.

The kinds of proteins displayed on the chips can vary depending on the questions being asked. But the chips closest to commercialization (initially for use by researchers) rely on the

remarkable immune system molecules called antibodies—each of which recognizes and binds to one specific protein or, more precisely, to a specific segment of a protein. Some of these antibody chips work by what is called the sandwich method: proteins recognized by a chip get sandwiched between two different antibodies, one that grabs the protein and a second that attaches a fluorescent label to the snagged molecule (diagram below).

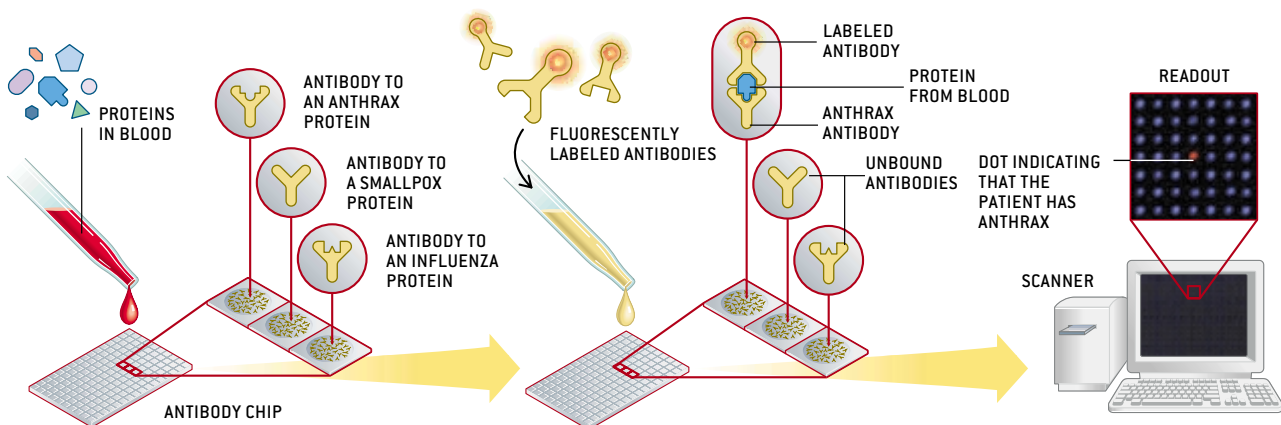
For antibody-based arrays to deliver fully on their potential for advancing research and diagnostics, scientists will have to topple at least two major impediments. One is the need for techniques that mass-produce many different antibodies at once, and not just any antibodies—those that bind tightly to one target, so as to reveal even small quantities in a sample. This problem is already being surmounted. The second obstacle is more fundamental. Medical science has so far uncovered only dozens of the perhaps thousands of proteins able to signal the presence or progress of a disease. Until chipmakers know which proteins to look for, they will be able to seek only a limited number of disease markers in a tissue sample. Fortunately, droves of investigators are now hunting for new disease-specific proteins. As advances in antibody manufacture and protein discovery converge, they will yield a second generation of protein arrays that could well transform both medical research and clinical practice.

N. Leigh Anderson and Gunars Valkirs collaborate on protein array research. Anderson is chief scientific officer at Large Scale Biology Corporation in Germantown, Md. Valkirs is chief technology officer at Biosite Diagnostics in San Diego, Calif.

A PROTEIN ARRAY IN ACTION

DOCTORS MIGHT ONE DAY use a “sandwich assay” to identify the infectious agent responsible for a patient’s illness. Is it a common flu bug or a new, deadly variety? Might the

tuberculosis bacterium be at fault—or even anthrax, smallpox or Q fever microorganisms unleashed by bioterrorists? Following the steps below would reveal the answer.



1 Apply blood from a patient to a chip, or array, consisting of antibodies assigned to specific squares on a grid. Each square includes multiple copies of an antibody able to bind to a specific protein from one organism and so represents a distinct disease-causing agent.

2 Apply fluorescently labeled antibodies able to attach to a second site on the proteins recognizable by the antibodies on the chip. If a protein from the blood has bound to the chip, one of these fluorescent antibodies will bind to that protein, enclosing it in an antibody “sandwich.”

3 Feed the chip into a scanner to determine which organism is present in the patient’s body. In this case, the culprit is shown to be a strain of anthrax.

Ritonavir, we learned, leads to activation of genes that are usually quieted in response to a well-known lipid-lowering agent; ritonavir also decreases the production of proteins that normally assemble into proteosomes, structures that break down no-longer-useful proteins, including lipid-containing types. These findings suggest that ritonavir raises lipid levels in the liver—and hence in the blood—in part by elevating the liver’s synthesis of lipids and inhibiting its breakdown of lipid-containing proteins. Further study of exactly how ritonavir interacts with the lipid- and proteosome-producing pathways will provide ideas for reducing its side effects.

Treatment Tailors

HAVING AN ENLARGED arsenal of drugs, and more drugs with fewer side effects, would be a great outcome of the molecular profiling made possible by DNA array studies. But many physicians are hoping for an even better result: rapid diagnostic tools that would divide patients with similar symptoms into separate groups that would benefit from different treatment plans. As the lymphoma study mentioned at the start of this article demonstrated, cancer specialists in particular desperately need ways to identify patients who require maximally aggressive treatment from the beginning.

Research into breast cancer by our group at Rosetta, working with collaborators from the Netherlands Cancer Institute in Amsterdam, demonstrates how expression arrays can help [see box on page 49]. In this case, we wanted to invent a test able to determine which young patients with early-stage breast cancer (with no evidence of cancer in the lymph nodes) need systemic drug therapy to prevent tumor spread (metastasis) after surgery and which do not. Although current guidelines recommend systemic treatment for about 90 percent of these women, a good many of them would probably avoid distant metastases even if they did not have such treatment. Unfortunately, standard tools cannot single out the women at greatest risk.

We began by generating expression profiles for tumors from close to 100 women under age 55 whose clinical course had been followed for more than five years after surgery. We initially worked with a microarray representing 25,000 human genes. In the end, we found that one particular signature produced by about 70 genes strongly indicated that metastases would soon appear. In addition, the opposite pattern was strongly indicative of a good prognosis. Clearly, some tumors are programmed to metastasize before they grow to a size smaller than half a dime, whereas other, larger masses are programmed not to spread.

Our results have to be confirmed by others before expression profiling can become a routine part of breast cancer workups. Within two years, many medical centers will probably begin to test expression profiling as a guide to therapy, not just for breast cancer but for other types as well. Other diseases need improved diagnostic tools, too. Expression profiling might help distinguish subgroups of patients with such disorders as asthma, diabetes or obesity who have special treatment needs. Those applications are now under study.

Before microarrays can live up to their full potential as research and diagnostic tools, several roadblocks have to be toppled. The chips, scanners and other accoutrements remain expensive (engendering “array envy” in many underfunded academics). Presumably, however, costs will drop with time.

Yet even if prices fall, the technologies may prove infeasible, at least initially, for doctors’ offices or standard medical laboratories. Few physicians or tech-

nicians have the equipment and the skill to prepare tissue samples properly for use with arrays. What is more, to diagnose, say, liver disease based on changes in gene expression in liver cells, a doctor would ideally need to obtain tissue from the liver. But that organ is not readily accessible.

These problems loom large right now but are probably surmountable with ingenuity. At times, for instance, accessible tissues might function as acceptable stand-ins for inaccessible ones. Moreover, in some instances, microarrays themselves may not have to be used; they might provide the research information needed for devising new diagnostic tests, which can then take other forms.

As the operations of cells and the entire body become better understood, physicians will be able to make more precise diagnoses, to offer patients more sophisticated therapies (possibly including gene therapies), and to tailor these interventions to an individual’s genetic background and current state of physiological functioning. By the year 2020, health maintenance organizations and their ilk could conceivably keep *in silico* models of the personal molecular states of their subscribers—virtual simulations that could be updated constantly with microarray and other data from doctor visits and with new scientific information about cell biology. Perhaps some subscribers won’t like that idea and will forgo a rate discount—and quite possibly the best care—in return for a feeling of privacy. Those who go along with the program, though, will probably delay the effects of aging more successfully and lead healthier lives. SA

MORE TO EXPLORE

- The Chipping Forecast.** Supplement to *Nature Genetics*, Vol. 21, pages 1–60; January 1999.
 - Genomics, Gene Expression and DNA Arrays.** David Lockhart and Elizabeth Winzeler in *Nature*, Vol. 405, pages 827–836; June 15, 2000.
 - Experimental Annotation of the Human Genome using Microarray Technology.** D. D. Shoemaker et al. in *Nature*, Vol. 409, pages 922–927; February 15, 2001.
- Web sites listing links and publications on microarrays can be found at:
- <http://bioinformatics.phrma.org/microarrays.html>
 - <http://industry.ebi.ac.uk/~alan/MicroArray/>
 - www.rii.com/publications/default.htm
 - http://ihome.cuhk.edu.hk/~b400559/2001j_mray.html
 - www.biologie.ens.fr/en/genetiqu/puces/links.html#news