From Genes to Genome Biology

No longer content to study one gene at a time, forward-looking researchers are developing a host of strategies—from lab techniques to computer programs—to analyze whole batches of genes at once

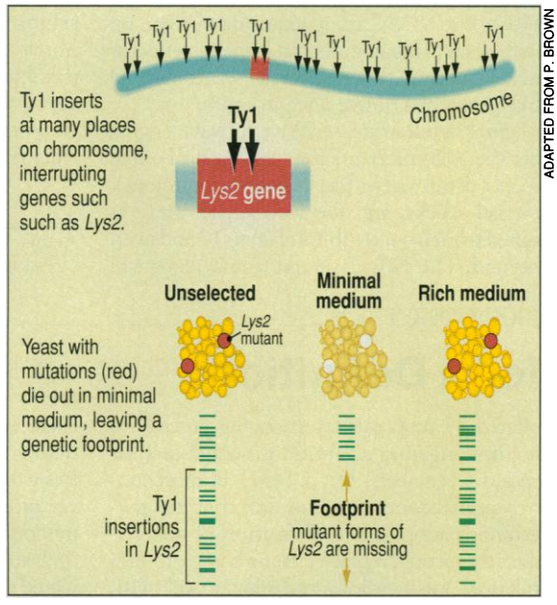
As a graduate student at Cambridge University in England 5 years ago, Victoria Smith and her colleagues were studying telltale bits of sequence that indicate the beginnings and endings of genes in the yeast genome, when they made a surprising discovery: Yeast had 6000 genes—about 2000 more than expected.

It was exciting news—but it also raised the prospect of years of painstaking labor, isolating and cloning each of the new genes, one by one, to learn what they do. After chasing down just one intriguing yeast gene, which resembled a human gene involved in autoimmune disease, Smith realized that "it was going to be a very slow process to look at the thousands of genes. We needed a faster way to get at their function."

If the yeast genome seemed daunting, then individually studying the 75,000-some genes buried in the human genome was overwhelming. And such single-gene studies wouldn't begin to address the biologically important question of how genes interact. So at last month's Genome Mapping and Sequencing Meeting at Cold Spring Harbor Laboratory in New York, scientists marveled over how Smith and other forward-looking researchers are finding clever ways to analyze genes on an industrial scale. While some groups find new ways to rapidly compare the genomes of different organisms, or to scale up the bread-and-butter mapping work, many others are learning to explore the functions of many genes simultaneously. Their approaches range from software tools that scan raw sequences for signs of genes, then tease out clues to those genes' functions, to new laboratory screening systems that monitor the function of an entire chromosome's worth of genes all at once.

All of these strategies allow researchers to achieve in one sweep what might have taken months or years to accomplish before—and also to ask completely new questions in genomic biology. Instead of seeking a single gene in a particular biochemical pathway, for example, these enterprising scientists want to reveal all the genes involved—as well as how those genes interact to control the pathway.

Such approaches will aid in probing diseases influenced by multiple genes, including cancer, says Richard Klausner, director of the National Cancer Institute (NCI). He expects that monitoring what genes are active in various tumors will help researchers understand why seemingly similar tumors respond differently to chemotherapy, for example, or why the same environment affects individuals differently. "The answer is going to be in the pattern of gene expression," he says. He wants NCI to invest $79 million in molecular diagnostics based on these sorts of genetic analyses.



ADAPTED FROM P. BROWN

Making tracks. Genetic footprinting reveals the function of the *Lys2* gene *(shown in red)* and of many other genes at the same time.

As a result of such efforts, says Eric Green, a geneticist at the National Center for Human Genome Research (NCHGR) in Bethesda, Maryland, "there's a culture that's being created, genomic approaches to studying biology." Shirley Tilghman, a molecular biologist at the Howard Hughes Medical Institute at Princeton University, agrees: "We're moving from gene-centric biology to genome-centric biology."

Footprints reveal function

At the genome meeting, the ferment of a field shifting focus was clearly evident. Smith, for example, talked to an interested audience about a new technique, genetic footprinting, which offers "a fast way to get [through] the first round of functional characterization," she says. "It's tremendous, a very clever approach," says Philip Hieter, a yeast geneticist at Johns Hopkins University.

Building on an idea proposed by Stanford University's Patrick Brown, Smith first grows yeast containing a mobile bit of DNA called the Ty1 transposable element. Ty1 inserts itself into many genes in the genome, leaving a genetic "footprint" that stamps out gene expression. That yields "a whole pool of mutations in one step," says Smith. She allows the yeast to multiply and exposes them to different environmental conditions. Over time, yeast with even subtly harmful mutations can't compete against the other strains growing with them and become extinct. By analyzing the DNA of each population at regular intervals, Smith is able to monitor the disappearance of these strains, identifying the debilitating mutations because yeast carrying them eventually die out.

That allows her to quickly determine the crucial genes and identify the roles other, still-unknown genes play in yeast fitness. For example, in a wholesale analysis of yeast chromosome 5, she found that 13% of the chromosome's genes are essential, because mutations in them prevent or severely retard growth. Another 28% led to changes in growth rate. And even for the known genes, novel functions popped up. For example, mutations in a gene called *cho1* alter membrane synthesis and generally retard growth, but Smith found to her surprise that in high salt conditions, yeast with this mutation thrive.

She plans to characterize the entire yeast genome in this way and put the information in a public database, to "reveal for yeast biologists genes they didn't know about before," she says. "And now that they know about them, they can study them."

Wholesale expression

Smith wasn't the only scientist at the meeting to demonstrate the power of genome-centric thinking—at least three groups are developing different approaches. Two of the groups farthest along are Brown's team, which focuses on gene expression, and Eric Lander's at the Whitehead Institute in Cambridge, Massachusetts, whose latest genetic-mapping technique may have much broader applications (see box).

Brown's team at Stanford and their collaborators had already made their mark with technology called genomic mismatch scan-

ning and microarray technology, first described last fall (*Science*, 20 October 1995, pp. 368 and 467). The technique sorts various cells' messenger RNAs (mRNAs), which are the hallmarks of active genes. The mRNAs are allowed to bind to an array of different gene fragments fixed to a surface, and those in each cell extract are labeled with fluorescent dye, so different colors on the array indicate active genes from different cells.

The technique debuted in an analysis of differences in gene expression between roots and leaves of the model plant *Arabidopsis*. But human and even yeast DNA has many more genes, so the broad utility of Brown's approach was unclear. Now the Stanford group has refined its microarray machine to create a finer grid that can evaluate more genes—and do large-scale analyses of yeast and human genes. "The initial paper was 45 genes from a plant; [this] report is about 1200 genes from a human," says one of Brown's collaborators, NCHGR's Jeffrey Trent.

Brown plans to scale up to 10,000 bits of DNA soon and eventually produce an array that packs in cDNA for all the known human genes. These scale-ups will allow more comprehensive comparisons of genetic expression not only between different types of tissues but also between the same type of tissue in different conditions. "This microarray technology is quite a general tool for looking quantitatively at a large number of sequences in a very complex DNA environment," Brown says.

Indeed, working with only the current 1200-speck array, Brown and Trent did exactly what Klausner would like to see oncologists eventually do regularly: compare gene expression in normal and cancerous tissue. With Paul Meltzer and Michael Bittner at NCHGR and Stanford's Joseph DeRisi, they studied how gene expression changes when a skin pigment cell turns into malignant melanoma. Such changes should reveal the genetic defects behind the malignant transformation and eventually point to the best kinds of drugs for that tumor, Trent says.

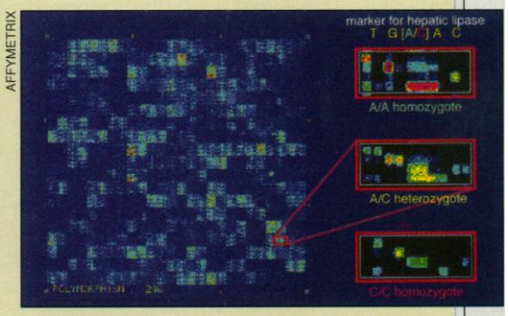
Brown's team is fielding calls from scientists eager to adopt these methods, and Klausner is so excited by these approaches that he has met with the developers of all these technologies. Eventually, he wants NCI to provide researchers with an easy-to-use version of large-scale genome scanning or genotyping technology. But not everyone is sure Brown's technique is ready for widespread use. "People really don't know how [the Stanford approach] is going to work," cautions David Bentley from the Sanger Center in Cambridge, England, who worries how quantifiable the expression data will be.

Even if takes a while for microarrays or genotyping chips to hit the market, armchair genomic biology is already taking hold in laboratories worldwide. Already, many researchers

Chipping Away at the Human Genome

Eric Lander, a mathematician turned geneticist and director of the Center for Genome Research at the Whitehead Institute in Cambridge, Massachusetts, is legendary at the annual genome-mapping meeting at Cold Spring Harbor Laboratory for dazzling his audience with innovative uses of technology. One year he wowed them with the idea of a gene-mapping machine—and 2 years later, the "Genomatron" had provided human and mouse maps ahead of schedule.

At last month's presentation, Lander came through again, spewing out data and ideas almost too fast for his audience to take in. His next novel technology: a "genotyping chip," made with the same techniques used in manufacturing computer chips, that can quickly analyze thousands of stretches of DNA. While other innovators explore new ways to examine the function of many genes at once (see main text), this new technology offers an industrial-strength scale-up of fundamental genetic-mapping work. And although one of the chip's first tasks is to help evaluate the thousands of markers needed in gene mapping, it may one day be used to quickly scan entire genomes at a single pass, determining which versions of various genes are present in an individual, for example, or rapidly comparing the genomes of two people.



To the letter. This DNA chip holds thousands of tailor-made DNA sequences.

To evaluate potential markers en masse, Lander's group relied on DNA chip technology developed by Affymetrix, a biotechnology company based in Santa Clara, California. Affymetrix scientists use masks and light-sensitive reagents—the workhorses of silicon chip–making—to lay down arrays of short DNA strands on glass wafers. A completed 2-cm by 2-cm chip contains more than 100,000 different sequences, each precisely placed.

Lander and Robert Lipshutz of Affymetrix realized that such an array of DNA fragments could provide an efficient way of screening a DNA sample for the presence of small stretches of DNA that differ in only one base. Those differences, called single nucleotide polymorphisms (SNPs), are "spelling" mistakes in the DNA code, Lander explains. Such mistakes turn up rather often along the genome and can help distinguish one individual's genetic material from that of another. So they can be used to trace how DNA gets recombined as it passes from parents to offspring and also provide the clues needed to build a better genetic map.

Whitehead's David Wang and Affymetrix's Ron Sapolsky first identified some SNPs, then created a chip that contains many stretches of DNA, each bearing a particular form of an SNP. To find out how much the SNPs actually vary in human populations—and hence how useful they may be for genetic mapping—the researchers take samples of human DNA and expose them to the chip. Each different DNA piece—with its particular SNP—binds to the matching DNA on the chip. By observing where on the chip the sample binds, researchers can determine what SNPs are present and, to some extent, in what proportions.

So far, Lander and his colleagues have identified about 150 SNPs that work "exceedingly well" as markers. Such new markers will speed the work of mapping the genome. But, even more importantly, says Lander, once they have in hand 2000 such signposts, they plan to create a "genomic scanner" by putting all the SNPs on a single chip. Then they can use the variation among the markers to look at genomic differences among individuals. Already they have assessed such variation at more than 100 SNPs in 16 family members—in a single afternoon, Lander told his spellbound audience. And by changing the DNA on the chip, researchers could apply this technology to everything from DNA forensics to alcoholism studies to cancer research.

At the meeting, Lander's audience was eager for even more results. "It's terribly exciting, but it was as much a provocative proposal as anything else," says Eric Green of the National Center for Human Genome Research in Bethesda, Maryland. Lander agrees: "We've made a prototype—it's a clear proof of principle, but it's going to take a hell of a lot of work to scale it up." **–E.P.**

have recognized that there is a mother lode of information about gene function to be mined from the data now stockpiled in computer databases. "The genome project has taught us the power of information and the power of collective action," Tilghman says. But as the burgeoning commercial interest in bioinformatics experts indicates (see p. 1730), striking gold in veins of genetic data requires search programs and well-organized databases, and programmers are scrambling to provide them.

The first step in computer genome analysis is to pick out likely genes from raw sequence data. One strategy comes from Gregory Schuler of the National Center for Biotechnology Information, who uses expressed sequence tags (ESTs), bits of DNA that signal the presence of a functionally active stretch of DNA and have been identified in huge numbers by mappers. So he and colleagues have spent the past year creating UniGene, a computer program that organizes these known coding stretches of DNA into groups that likely represent unknown genes based on how similar they are at one end. Thus far, they have converted ESTs into some 46,000 clusters that may include half the genes in the human genome.

Schuler calls these clusters "diamonds in the rough," which researchers are now "polishing" by using other databases to pin down the genes hidden in the clusters and find clues to their function. For example, Andrea Ballabio of the Telethon Institute of Genetics and Medicine in Milan, Italy, and her colleagues tapped a variety of such database programs, including UniGene, to find 66 human equivalents to known *Drosophila* genes. They then located these genes on the human genetic map and looked for diseases linked to mutations at those locations. Seven of the new genes seem to be tumor-related; about two dozen are involved in eye or neural development; and four relate to ion-channel proteins, the group reported in the June issue of *Nature Genetics*. Because researchers know how mutations in these genes affect fruit flies, they have a jump on understanding how the human counterparts may work, Ballabio says.

As the number of databases continues to explode, scientists are relying more on clever software to help tease out the meaning behind all the data. "The time is coming when a graduate student thinking about a project will begin with a fairly long time at the computer coming up with a problem through sequence analyses," Hieter says. And thanks to the rapidly expanding tool kit of both computer and laboratory analysis, when students do choose a problem, the project will likely be quite different from the old model of finding and analyzing a single interesting gene. "There's going to be a paradigm shift in the way we think about biology," Tilghman predicts. "We're going to be able to ask questions that I could only dream about as a graduate student."

–Elizabeth Pennisi

SOLAR PHYSICS

Putting Some Sizzle in the Corona

MADISON, WISCONSIN—The most eloquent reaction of all came from Robert Bless, the University of Wisconsin astronomer who hosted last week's American Astronomical Society meeting here. When a reporter told him that an instrument aboard the Solar and Heliospheric Observatory (SOHO) satellite had detected 100-million-degree oxygen ions in the sun's atmosphere, he silently dropped his jaw. That temperature is tens of times higher than has ever been measured before in the corona, the sun's halo of ionized gases. It's fully 100 times hotter than the temperature of electrons in the same part of the corona. SOHO project scientist Art Poland of NASA's Goddard Space Flight Center had a more vocal reaction: The results "just blew me away."

ASP

Hot zone. The corona, seen in white light during a solar eclipse.

Announced here on 11 June by members of the team operating SOHO's Ultraviolet Coronagraph Spectrometer (UVCS), the findings are all the more surprising because they turned up in what was thought to be a cool spot: a coronal hole—a region of the corona where the sun's magnetic field lines wander into space rather than arching back to its surface. Coronal holes look dark in x-ray images of the sun because the free electrons there are cooler than elsewhere in the corona. But a handful of solar physicists say the detection of superhot oxygen provides crucial support for theories that would solve a long-standing mystery: what heats the corona to millions of kelvin when the surface of the sun is a mere 6000 degrees.

Details of the corona's temperature are hard to get from the ground, because many of the bright lines in its spectrum, given off by ions like hydrogen and oxygen, fall in the ultraviolet region, which is blocked by Earth's atmosphere. So when SOHO was launched last December as an all-purpose solar observatory (*Science*, 10 November, p. 921), it carried the UVCS to fill in the picture. The UVCS blocks the disk of the sun, creating an artificial solar eclipse that allows it to analyze the ultraviolet lines at distances out to several times the sun's radius, explains John Kohl of the Harvard-Smithsonian Center for Astrophysics (CfA), principal investigator for the UVCS.

To determine the temperature of the ions responsible for the lines, the instrument relies on the Doppler effect, which broadens the lines in proportion to the ions' velocities toward or away from an observer. In a coronal hole above the sun's north pole, early results showed that "the higher we went, the higher the oxygen temperature went up," says Kohl. It still hadn't peaked at the limit of the measurements 0.9 solar radii above the surface, where the temperature soared to 100 million degrees.

In the same structure, hydrogen nuclei reached a peak of 6 million degrees at about one solar radius and then gradually fell off, and separate measures had the electrons topping out at about a million degrees. Meanwhile, in the brighter regions of arching field lines called helmet streamers, UVCS generally found that both oxygen and hydrogen peaked at about 2 million degrees.

The results are "grist for my mill," says Jack Scudder, an astronomer at the University of Iowa who has proposed a controversial theory of coronal heating that relies on gravity to confine cooler particles near the surface, allowing hotter ones to escape into the distant corona—a filtering mechanism that would act more strongly on heavier species (*Science*, 11 February 1994, p. 757). But Scudder's isn't the only theory predicting hot heavy ions.

In a picture developed by Ian Axford and James McKenzie at the Max-Planck-Institut für Aeronomie in Lindau, Germany, for example, so-called magnetohydrodynamic waves, which could be driven by disturbances near the solar surface, would pump energy into particles by driving their gyrations around magnetic field lines. Such waves would tend to be more intense at low frequencies, best matched to the gyration frequencies of heavier ions. As Axford puts it, hot, heavy ions are "something we have expected for a long time." Both theories also predict that the preferential heating would be less noticeable in the denser, trapped gases of the streamers than in coronal holes.

To figure out what kinds of measurements could distinguish between these and several other competing theories, Kohl has organized a meeting to take place at CfA next week. What's already clear, says solar physicist Joseph Hollweg of the University of New Hampshire, is that the new observations "are going to get very close" to unraveling the mystery of the hot solar corona.

–James Glanz