Fast Tracks to Disease Genes

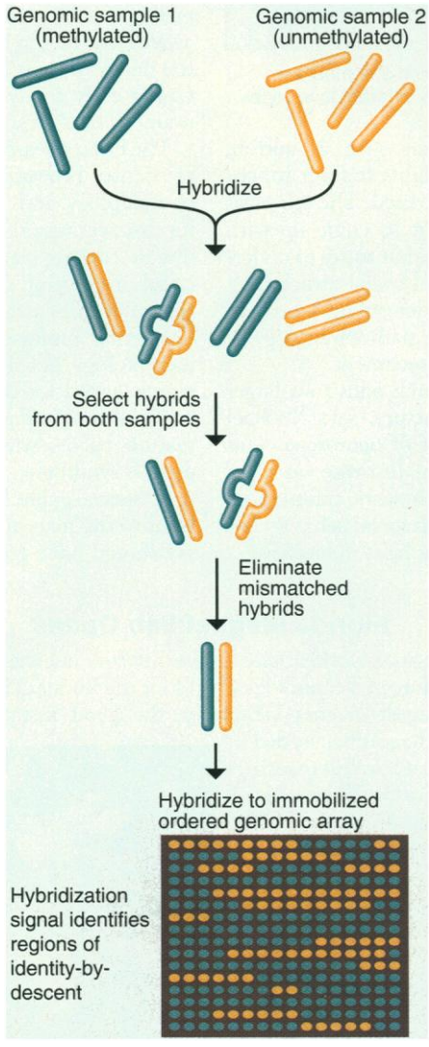
Two new techniques for scanning the genome promise great advances in tracking the roots of disorders caused by multiple genes

The community of researchers hunting for disease genes has good reason to feel pleased with itself: It has found the culprits responsible for a host of diseases caused by defects in a single gene, ranging from cystic fibrosis to Huntington's chorea. And recently, with the long-awaited identification of the breast cancer susceptibility gene *BRCA1* (*Science*, 23 September, p. 1796), geneticists took a major step toward tackling an even more difficult problem: diseases influenced by not one but several genes.

Exciting as the *BRCA1* discovery is, however, the task facing researchers interested in other multigene diseases remains daunting. While more than 80% of women unlucky enough to inherit *BRCA1* will develop breast cancer, for most complex diseases with a genetic component—such as diabetes, severe obesity, and schizophrenia—no single gene is thought to exert such a strong influence. Weaker influences pose problems for gene seekers, because the smaller the effect of an individual gene, the harder it is to map its position. "These are mammoth, labor-intensive, and expensive undertakings," says molecular biologist Patrick Brown, a Howard Hughes Medical Institute investigator at Stanford University.

But help may be on the way. Much of the spadework could be removed if two techniques which made their published debuts last year, called genomic mismatch scanning (GMS) and representational difference analysis (RDA), fulfill their early promise. Unlike existing methods of genome analysis, GMS and RDA allow geneticists to scan the entire genome rapidly in one fell swoop. "It's like sequencing the genome without having to sequence the whole genome," says geneticist Peter Goodfellow of the University of Cambridge.

This overcomes a drawback of the traditional gene-hunting technique, called linkage analysis. Researchers generally look for disease genes by scanning individuals from affected families for genetic marker sequences that vary from individual to individual. When, in a specific family, a particular variant of a marker is always inherited along with the disease, there's a strong likelihood that this marker lies near the culprit gene. The problem with this technique is that researchers must determine the sequence of each marker for each individual, and this genotyping must be done one marker at a time. Furthermore, the weaker the influence of an individual gene, the more families that must be included in the analysis—vastly increasing the amount of time, money, and labor involved.



Molecular matchmaking. Genomic mismatch scanning mixes DNA fragments from two genomes, then treats the mixture with a series of enzymes to leave only those fragments that are identical in both genomes.

Both RDA and GMS offer the prospect of speeding up the process by looking at the genome as a whole, rather than marker by marker. Although they share that goal, they approach it from opposite directions. RDA, invented by Michael Wigler and Nikolai Lisitsyn of the Cold Spring Harbor Laboratory in Cold Spring Harbor, New York, compares two genomes and identifies regions that differ between the two, while GMS, developed by Stanford's Brown, picks out identical regions.

The simpler of the two techniques is RDA, which picks out contrasting regions of DNA by exploiting DNA hybridization: the tendency of a strand of DNA to bind to another strand with a complementary sequence, so that each "C" in its sequence of bases lines up against a "G," each "A" pairs up with a "T," and vice versa.

For years geneticists have realized that this principle could be used to identify sequences present in one sample of DNA, but not in another. The general strategy, it was clear, was to treat the two samples with a restriction enzyme, which recognizes a specific sequence and cuts the DNA each time it occurs, resulting in DNA fragments of varying size. Next, the DNA would be "melted" so that the double-stranded DNA dissociates into single strands. After that the two samples would be mixed—but with much larger quantities of one sample than the other.

Once mixed, the fragments of DNA will start hybridizing with complementary fragments, but because of the large excess of DNA from the second sample, most DNA fragments from the first sample will hybridize to sequences from the second. Indeed, the only fragments that will hybridize in large quantities with their compatriots from the first sample are those that cannot find a second-sample partner. By selecting out DNA in which both strands originate in the first sample, and repeating the whole procedure several times, it should theoretically be possible to purify sequences present in the first sample alone.

In practice, however, the technique struggles with large, complicated genomes. The problem is that the procedure only works well if, when mixed, almost every fragment in the two samples hybridizes onto another fragment. But in genomes like our own, that can take weeks. "The mammalian genome is very complex," explains Wigler, "so you don't get complete hybridization in [rapid] time." Wigler and Lisitsyn, however, realized that the process could be accelerated if they analyzed only a representative sample of the two genomes to be compared.

To select a sample, the researchers exploit

SOURCE: *CURRENT OPINION IN GENETICS AND DEVELOPMENT*

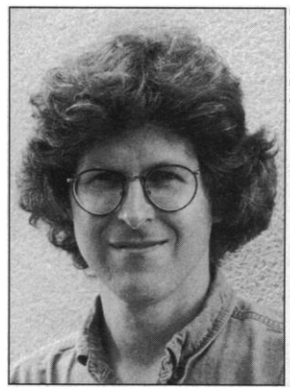
an idiosyncrasy of the polymerase chain reaction (PCR), the standard lab technique that is used to make a multitude of copies of DNA. PCR has a tendency to multiply small fragments of DNA more efficiently than larger ones. After first breaking up their two DNA samples with a restriction enzyme, Wigler and Lisitsyn subject the resulting fragments to PCR. The result is what the researchers call "amplicons," a collection of the fragments that are small enough to be easily amplified. These hybridize readily, allowing sequences unique to one amplicon to be purified rapidly.

One problem with this approach is that each amplicon contains only around 10% of the genome—so it is possible to miss genetic differences that fall in one of the larger fragments not included in the amplicon. This is easily solved, however, says Wigler, by employing a different restriction enzyme at the start of the procedure, which produces a different set of initial DNA fragments, and hence a different amplicon. "If you're worried," says Wigler, "you can take multiple [amplicons]."

RDA is already being used to search for mutations involved in cancer, by comparing DNA from tumor cells with that extracted from the same patient's healthy cells. That's a desirable capacity, but what gene mappers would like is the ability to identify genetic markers that are closely linked to an inherited disease gene. Working with Eric Lander of the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, and researchers at the Jackson Laboratory in Barr Harbor, Maine, Lisitsyn and Wigler have already demonstrated that RDA can find such linkages in mice. They crossed two different mouse strains, one of which carried two copies of a recessive gene and the other carrying two copies of a dominant gene for the same trait, and then bred the resulting progeny for another generation. By comparing DNA from mice in this last generation that showed the recessive trait, with DNA from the original strain with the dominant genes, the team was able to pinpoint marker sequences closely linked to the gene in question.

Obviously, human geneticists cannot conduct such controlled breeding experiments. And analyzing families in which a disease is inherited (as in conventional linkage analysis) is of little use, because in an outbred human population there are a multitude of differences between the genomes even of close relatives, making it difficult to pinpoint differences linked to a gene of interest. Some gene hunters, however, are studying inbred populations, which show less genetic variation. And in such populations, notes Geoffrey Duyk, director of genomics with Millennium Pharmaceuticals Inc., a Cambridge, Massachusetts–based company that is focusing on multigene diseases, it should be possible to map genes by RDA.

The idea, explains Duyk, would be to pool DNA from individuals affected by a genetic disease, and compare this with a similar pool of DNA from unaffected members of the same population. If the pools cover large enough numbers of people, most of the genetic variation in the population will be present in the samples. As a result, when the pooled samples are mixed, fragments from the first sample will be able to find and hybridize with complementary fragments in the second, except those which come from the region containing the gene of interest. In an outbred population, the number of people that must be sampled to cover most of the population's genetic variation is impracticably large. But with an inbred population the number is considerably smaller. And a population need only be modestly inbred for this strategy to be feasible. "These unusual populations don't have to be so weird," says gene mapper Richard Myers of Stanford University.



SUE KLAPHOLZ

Fine detail. According to Patrick Brown, genomic mismatch scanning could map genes with higher resolution than current methods.

Nevertheless, what gene hunters would really appreciate is a rapid genome scanning technique that could be used to map genes in any study population, regardless of inbreeding. And that, says Stanford's Brown, is where GMS comes in. Because GMS detects similarities between pairs of genomes, it should be possible to use the technique to map genes by studying pairs of affected relatives. For any one pair of relatives, large chunks of the genome will be identical by descent. But comparing numerous pairs of affected relatives across many families using GMS should reveal regions that are more-or-less consistently identical by descent within each family—and therefore must be linked to a disease susceptibility gene.

As in the case of RDA, the main advantage of GMS is its ability to compare two genomes in a single procedure. Once again, the technique involves DNA hybridization (see figure, p. 2008), and starts out by treating the two genomes with a restriction enzyme, to produce two samples of DNA fragments of varying size. At this point, however, in contrast to RDA, one sample is treated so that methyl groups are tacked on at intervals all along the fragments. Once altered in this way, the two samples are melted to produce single-stranded DNA, mixed together, and allowed to hybridize. The result is a mixture of double-stranded DNA fragments, some containing two methylated strands from the first sample, some containing two unmethylated strands from the second sample, and some containing one strand from each.

Fragments derived from just one sample cannot be used to compare the two, so these must first be removed using a series of enzymes that digest DNA containing either two methylated or two unmethylated strands. The key step, however, is to remove those DNA fragments that—although their component strands were sufficiently similar in sequence to hybridize with the corresponding sequence from the other sample—contain mismatches between the two strands, corresponding to subtle sequence differences between the two genomes being compared. To do this, GMS employs what Brown calls "Nature's own tool for detecting differences between DNA sequences": three proteins which are used by the bacterium *Escherichia coli* to scan newly synthesized DNA for sequence errors.

In *E. coli*, these proteins look for mismatches between complementary strands of DNA, and whenever they find one, introduce a nick into the newly synthesized strand, which will be unmethylated. In GMS, DNA fragments nicked in this way can again be removed using a DNA-digesting enzyme, leaving behind only those fragments that are identical for the two genomes under comparison.

As Brown readily admits, however, some technical problems must be ironed out before GMS can really make its mark. For one, an efficient method for identifying the DNA fragments that have survived the GMS process is needed. This should be possible by pinning DNA sequences of the whole genome down onto an immobilized array. The GMS fragments could then be hybridized to the array of sequences to rapidly reveal identical regions between the two genomes being compared. But the major hurdle is that—while GMS works like a dream in yeast—it has yet to be proven in mammalian genomes.

One problem is that complex genomes contain numerous repetitive sequences, which interfere with the hybridization of the sequences in which they are embedded. Another difficulty is that the human genome is less variable than that of yeast—which means that genuine genetic differences detected by the *E. coli* mismatch repair proteins are harder to separate from the "noise" caused by the occasional tendency of these proteins to erroneously nick matching DNA sequences. Nevertheless, both Brown and Stanley Nelson, a former postdoc in his lab now at the University of California, Los Angeles, are confident that these problems

can be overcome by tinkering with the procedure. "There is no question that it should work in human DNA," says Nelson.

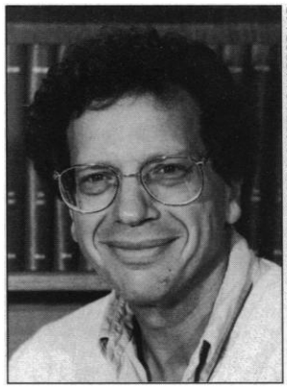
If so, GMS should lend itself to a powerful mapping method called linkage disequilibrium analysis, which compares the genomes of individuals from different families which nonetheless share a particular genetically determined trait. This method assumes that two apparently unrelated people sharing a trait will both have inherited the gene for it from a distant common ancestor. Over the intervening generations, their genomes will have become so thoroughly shuffled up that only those sequences very closely linked to the gene in question will be common between them. Linkage disequilibrium offers the possibility of ultrafine resolution mapping, and whereas many of the markers used in standard linkage analysis are too widely spaced to exploit this potential, says Brown, "the theoretical resolution of GMS is much higher."

For now, however, gene mappers are waiting for GMS's bugs to be ironed out. And while RDA has been shown to work with mammalian DNA, there is a similar reluctance to take up the technique, as some researchers have found it difficult to master. "You just have to fool around with it in the lab and not get discouraged," says Lorraine Flaherty, a mouse geneticist at the State University of New York in Albany, who is using RDA to map genes for a mouse model of polycystic kidney disease. The huge promise of GMS and RDA, however, says Stanford's Myers, makes it important to persevere with both. "You always have a signal-to-noise problem with new techniques," he says.

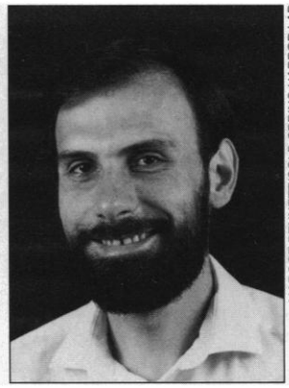
In the short term, say most gene mappers, the best hope of identifying genes underlying complex genetic disorders lies in developing high-throughput, automated versions of existing marker-by-marker genotyping methods. "I think we are going to see a flood of results with this [existing] technology," says John Todd of the Wellcome Trust Center for Human Genetics in Oxford, whose group published a paper in *Nature* earlier this month describing the use of automated genotyping to conduct a genomewide screen for genes underlying insulin-dependent diabetes.

Maybe so, but Millennium's Duyk, for one, believes that for some particularly intractable diseases, this technology probably won't be efficient enough, no matter how automated. And even if the next generation of gene-mapping technology is not based directly on GMS or RDA, argues Duyk, the obvious way forward is to look for methods that—like these techniques—can scan the entire genome in one pass. "The standard way we've survived for 10 years," he concludes, "just ain't going to cut it."

–Peter Aldhous



MARLENA EMMONS/COLD SPRING HARBOR LAB



MARGOT BENNETT/COLD SPRING HARBOR LAB

Mouse mappers. Michael Wigler *(left)* and Nikolai Lisitsyn have used representational difference analysis to map genes in mice.

Additional Reading

N. Lisitsyn *et al.*, "Cloning the differences between two complex genomes," *Science* **259**, 946 (1993).

R. M. Myers, "The pluses of subtraction," *Science* **259**, 942 (1993).

S. F. Nelson *et al.*, "Genomic mismatch scanning: A new approach to genetic linkage mapping," *Nature Genet.* **4**, 11 (1993).

E. S. Lander, "Finding similarities and differences among genomes," *Nature Genet.* **4**, 5 (1993).

P. O. Brown, "Genome scanning methods," *Current Opin. Genet. Devel.* **4**, 366 (1994).

N. A. Lisitsyn *et al.*, "Direct isolation of polymorphic markers linked to a trait by genetically directed representational difference analysis," *Nature Genet.* **6**, 57 (1994).

J. L. Davies *et al.*, "A genome-wide search for human type 1 diabetes susceptibility genes," *Nature* **371**, 130 (1994).

MICROSCOPY

A Sideways Look at Chemical Activity

Microscopists, who generally look down on the objects they're trying to image, are beginning to peer sideways to get a better view. In recent years, scientists have been using the up and down movements of force microscopes to scan the atomic-scale features of material landscapes. The instrument has a tiny tip sensitive to attractive or repelling atomic forces. As it skims the contours of molecules or atoms the tip moves vertically, and the plot of those movements forms a topographical map. But the atomic force microscope (AFM) can also yield another important piece of information—the frictional, or lateral, force on the tip.

On page 2071 of this issue, C. Daniel Frisbie and his colleagues in the lab of Charles Lieber at Harvard University, and others at the Massachusetts Institute of Technology, show how these sideways forces can reveal clues about a surface's chemistry—something topographic maps can't do. Such chemical information "together with the [spatial] image could be very important," notes Richard Colton, an AFM researcher at the Naval Research Lab in Washington, D.C. It could, for example, allow researchers prospecting for new drug compounds to quickly search for reactive molecules by swiping an AFM tip across a sample containing many such compounds.

The key to this work was to chemically modify the AFM tip and surface in a controlled way. The group used self-assembled monolayers to attach two kinds of hydrocarbon molecules to the tip. One is capped with a carboxylic acid that's hydrophilic, or attracted by water; the other ends with a methyl group, which is hydrophobic, or repelled by water. Using a lithographic technique, they also applied these hydrophilic and hydrophobic molecules to a gold surface in a pattern resembling concentric boxes.

When the scientists measured the topography of the surface, it appeared to be smooth. But when they plotted frictional forces, their map matched the box pattern. Friction was highest when hydrophilic molecules (which form hydrogen bonds) on the tip touched hydrophilic molecules on the surface. It was lower when hydrophobic areas on the tip encountered their counterparts on the surface or where the hydrophobic and hydrophilic molecules came in contact. Furthermore, the force that was required to pull the tip from the surface, measured for each combination, correlated quite nicely with the friction measurements.

Lieber has no trouble pointing to examples of how "chemical force microscopy," as he dubs this AFM application, could be useful. One is to study the principles of adhesion and lubrication, which are still rather elusive on the molecular level. But it's biological applications that really fire the imagination of AFM researchers. In medical diagnostics, for example, Lieber suggests that an array of potential drug molecules could be put on a surface, then scanned using a tip coated with a receptor molecule; the scanning could quickly reveal the strongest interaction, and thus the molecule best suited to bind to that receptor. AFM researchers have already begun to explore these uses, showing that sometimes a sideways move can be a real step forward.

–Jocelyn Kaiser