A New Guide to the Human Genome

The construction of a detailed physical map of the human genome, with 15,000 sequence-based landmarks, opens the door to genome-wide sequencing

If you are about to embark on a long and difficult journey, one of the first things you'll want is a detailed map. Indeed, some journeys could never even get under way without the right kind of map. That's been the case, for example, for the Human Genome Project's (HGP's) ultimate trip: determining the exact sequence of the 3 billion bases that make up the genome. Now, an article in this issue brings the day that the entire human genome will be accessible to sequencing a big step closer.

On page 1945, a large research team led by molecular biologists Thomas Hudson, Lincoln Stein, and Eric Lander of the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, describes a new physical map of the human genome. The map, the product of nearly 3 years of work by some 50 mappers, sequencers, informaticists, and bioengineers, contains more than 15,000 specific sequence markers distributed over all the human chromosomes. That's about half the goal of 30,000 markers set by the HGP for a physical map, but it's already good enough to begin large-scale genome sequencing. Francis Collins, head of the National Center for Human Genome Research (NCHGR) at the National Institutes of Health, describes the map as "a big leap forward" for the sequencing effort as well as "an enormous boon to people who are hunting genes."

One reason for plaudits is the map's improved coverage of the genome—roughly 95% compared to about 75% for the previous best whole genome map, which was published in September in *Nature* by Daniel Cohen's team at the Centre d'Etude du Polymorphisme Humain (CEPH) in Paris, Jean Weissenbach's at Généthon, also in Paris, and Lander's Whitehead group. What's more, the kind of landmarks on the new map, DNA sequences known as sequence-tagged sites (STSs) that can be written into the databases, should make it far more versatile and easier to use than earlier maps. Indeed, genome researchers around the world are already using the information, which is available through the Internet.

In all physical mapping efforts, the goal is to find specific identifiable landmarks located throughout the genome. Traditionally, physical maps have consisted of collections of actual DNA containing the landmarks. Such maps, however, require laboratories to maintain libraries of cloned DNA. The Whitehead group opted instead for STSs, a strategy researchers have adopted for mapping some individual chromosomes, but not the entire genome. These short sequences of DNA, which usually contain from 200 to 500 base pairs, don't need to be stored as actual DNA, and they are easy to detect in a sample of unknown DNA. Once the marker sequence is known, researchers can simply screen for it with the polymerase chain reaction (PCR), and that will give them a stretch of DNA that they know comes from a specific location on the genome.

SAM OGDEN

Mapping crew. In this photograph of the Whitehead team, Eric Lander is in the center (with arms folded); Thomas Hudson is at his left and Lincoln Stein is second from the right.

Construction of the Whitehead team's map certainly required a lot of screening. Not only did they begin with a large collection of STSs—more than 20,000 in all, most of them identified by the Whitehead group itself but some developed elsewhere—but they actually assembled the markers into three different maps, each with a different scale. They then integrated all three maps into one.

To make the most fine-scale of the three, the researchers turned to the same DNA library that was the basis of the CEPH-Généthon map. This library contains the entire human genome, broken up into fragments of a million base pairs or so and cloned in yeast artificial chromosomes (YACs)—a total of about 30,000 clones. All 30,000 YACs had to be screened against the 20,000 STS markers. The idea was to arrange the STS landmarks in their proper order in the genome by looking for those that were present in many of the same clones. As Hudson describes the task, "There were 30,000 pieces of the puzzle on the table, and our job was to put the pieces together."

But Hudson says that the YAC analysis gave only local order—small sections of the puzzle. To make assignments to chromosomes and build the whole genomewide picture, the team needed the other two maps they constructed.

One of these is a so-called "radiation hybrid" map, which allowed the STSs to be ordered over longer distances, up to 10 million base pairs compared to 1 million for YACs. This type of mapping is based on the fact that chromosomes can be chopped into small fragments by exposing cells to radiation. Although the radiation kills the cells, they can still be fused with hamster cells, which acquire some, but not all, of the chromosome fragments. By following which STS markers ended up in which hybrids, the researchers were able to work out which STSs were likely to be close to each other in the genome. And because they used many of the same STSs as in the YAC map, they could assemble the STSs into still bigger puzzle pieces.

And finally, the team used a genetic map, which Weissenbach's team at Généthon has worked out by following inheritance patterns of a variety of markers. By putting STS markers on this map, they were able to build up still bigger pieces, containing as many as 30 million base pairs of DNA. With the radiation hybrid and genetic maps, Hudson says, "it's like you're in a helicopter looking down. It gives the big picture." After two-and-half years of work, the pieces of the puzzle finally fell into place, and the researchers had their map. "It was," Lander recalls, "a very satisfying experience when the puzzle came together."

The team members had good reason to be satisfied, because the technical challenge of carrying out their mapping strategy had been formidable. Because of the extremely high number of PCR analyses needed to detect and put the STSs on three maps—15 million so far—the team had to spend the first year and a half of the project automating the technology, devising a machine they call the "Genomatron," which can do 150,000 PCR assays per run. As geneticist Maynard Olson of the University of Washington, Seattle, notes, "15 million PCR assays is a lot of PCR assays, essentially impossible with standard

laboratory procedures." They also had to bring in the informaticists to produce a system for tracking and integrating the vast amount of data produced.

The genome community is also pleased, partly because the new map fills in some serious gaps. Even though researchers are producing STS-based maps of individual chromosomes, together these account for only about one third of the genome, says Eric Green of the NCHGR, whose team is mapping chromosome 7. The new map, he notes, "is a good start on the rest of the genome, which everyone was afraid would be incompletely covered."

Researchers are also more confident that the markers are arranged correctly. "Previous maps were very good at having clone coverage," says molecular geneticist David Cox of Stanford University. "We had YACs and some idea of how these YACs went together, but we weren't so comfortable about the order of the markers." Here, he adds, they are "really delineating the position of these unique markers and delineating them in a way that is easy to use."

The STS sequences, after all, allow the map to be made available to anyone who wants to use it, because they can be maintained as a database rather than as clones in a freezer. What's more, says Olson, "this whole map, because it is STS-based, can be put up on the World Wide Web." In fact, the Whitehead team has been putting the data on the Web from the start (http://www-genome.wi.mit.edu). "They've been very good about that, and they should be commended," says Peter Goodfellow of the University of Cambridge, U.K., who provided the cell lines used to make the radiation hybrid map. In fact, people trying to find disease or other genes were using the data even before the map coalesced. Says Collins, "You can't find anyone who is hunting genes who doesn't know the Whitehead World Wide Web address by heart."

And the map should remain useful even as genome technology changes. Because STS markers are sequences, they can be used to orient gene searchers or sequencers no matter what type of DNA clones they are working with. "The neat thing about STSs is the universality of the landmarks. You can move quickly from one mapping resource to another," Green says. That's important, because even though YACs have proved very useful for map-building, they are not likely to be used for the eventual sequencing effort, because the DNA they contain sometimes gets rearranged so that it doesn't accurately reflect what's in the genome. Researchers will instead use smaller, more stable forms of cloned DNA such as that made in cosmids or bacterial artificial chromosomes.

For all its advantages, Collins and others, including the Whitehead team members themselves, note that the map does not yet provide dense enough coverage of the genome to finish the project. The average distance between the markers is now about 200 kilobases, and Collins says "to finish [sequencing] we really need [marker] spacing of 100 kilobases, or better yet 50 or 30 kilobases." Lander predicts, however, that the combined effort of his lab and others ought to bring the project up to the 100-kilobase target sometime next year. The genome mappers, it seems, will soon have the detailed map they need to guide them to the end of their long and winding road.

–Jean Marx

MICROSCOPY

Making a New Ruler for the Nanoworld

GAITHERSBURG, MD—A good machinist needs precise measuring tools, and so do craftspeople whose wares are very, very small. The etched circuits on a computer chip, for example, won't function well unless their height, width, and smoothness fall within tight tolerances. And those tolerances are getting even tighter, as devices shrink to the nanometer (a billionth of a meter) scale. But the rulers used to measure these minute features and the distances between them aren't keeping pace. "We're getting to a situation where we have a capability to manufacture smaller dimensions than we can accurately measure and certify," says James Greed, president of VLSI Standards Inc., a company in San Jose, California, that makes grids used to calibrate a cutting-edge measuring device, the scanning tunneling microscope (STM).

Last week, however, researchers at the National Institute of Standards and Technology (NIST) in Gaithersburg unveiled a ruler that can measure the width of a circuit feature to within a few tens of nanometers and tell you how far that feature is from another one, which can be millions of nanometers away. That's comparable to combining a microscope that can size up a ladybird on one hilltop with a telescope that can measure the distance to the next hilltop with almost the same precision. The device, called the "molecular measuring machine," or M^3, combines a probe that senses atomic contours with a precise system for tracking that probe. In the nanoworld, says metrologist Joe Griffith of AT&T Bell Labs in Murray Hill, New Jersey, "there is no machine in existence now that has higher precision over those distances."

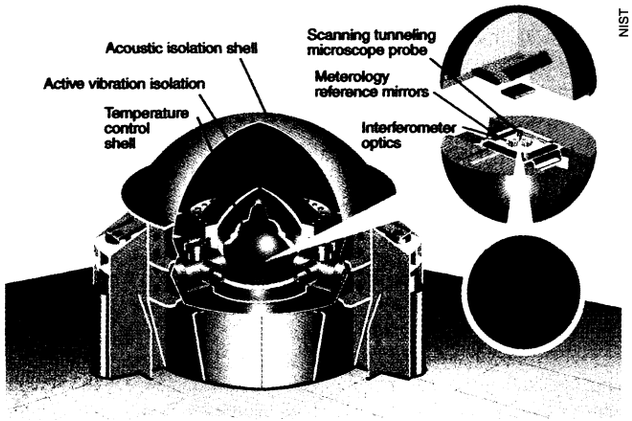
The present nanochampion, the STM, works by moving a tip over a sample, maintaining a slight but constant electrical current between the tip and the sample surface. This constancy requires the tip to move up and down as it encounters atomic hills and valleys, allowing it to pick out and measure atomic-scale features. But the microscope isn't designed to measure precise distances between such features when they are farther than 100 micrometers apart.

So physicist Clayton Teague of NIST designed a machine that combines an STM with a technique for tracking the position of the STM tip over relatively vast distances: laser interferometry. By training a laser on the STM tip as it moves across a vaster landscape, M^3 can measure small features and indicate just how far apart those features are.

To date, M^3 can measure the distance between points 1 millimeter apart to within 20 to 40 nanometers. Teague's team is now shooting for an accuracy of one nanometer over 50 millimeters. Alignment problems between the interferometer and the STM, and errors that arise because mirrors that channel the laser don't move exactly in sync with the STM tip, are currently preventing this. The researchers believe they can solve these problems within 2 to 3 years.

Still, the M^3 can already do something a standard STM can't: measure the accuracy of the 600-micrometer grids, divided into 1.8 micrometer squares, that are used by semiconductor manufacturers to calibrate their own STMs and other microscopes. And that's no small achievement.

–Jocelyn Kaiser



NIST

Multipurpose tool. The "molecular measuring machine" can measure atom-scale features and vast distances between them.