gene amplification predicts a bad prognosis for these patients may help to guide their treatment. "The ovarian results are even more interesting," Lippman says. "I believe that if they are true, it will be extremely useful."

Ovarian and breast cancer have a number of common features. The female hormone estrogen influences the growth of both. Moreover, women who have one cancer are at increased risk of developing the other. The Slamon group's results now suggest that the same cellular derangements may contribute to the development of the two cancers.

Whether the *neu* gene amplification plays a causative role in the two human cancers remains to be established, but work with animals and cultured cells suggests that it may. For example, Philip Leder, William Muller, and their colleagues at Harvard Medical School and the Howard Hughes Medical Institute have introduced the active *neu* oncogene into mice. Expression of the transferred gene in the mouse mammary tissue is sufficient by itself to produce malignant mammary tumors in the animals.

Leder initiated these experiments as part of his continuing investigations of oncogene action and before the link between *neu* gene amplification and breast cancer prognosis was made. "But," he notes, "that correlation when it came made *neu* an extremely interesting oncogene in breast cancer.

Other investigators, including the NCI's Stuart Aaronson and Axel Ullrich of Genentech, Inc., in South San Francisco, have shown that overproduction of the protein encoded by the normal *neu* gene can give cultured cells malignant properties. As mentioned, the human breast cancers cells in which the gene is amplified make higher than normal amounts of the *neu* protein. So do the ovarian cancer cells.

Overproduction of the *neu* protein might cause tumor cells to behave aggressively by causing them to grow faster than they otherwise would. The *neu* protein has all the earmarks of a growth factor receptor. It is structurally similar to the product of another oncogene, called *erbB-1*, which encodes the receptor for epidermal growth factor. No one has yet identified the growth factor that activates *neu* protein, however.

Amplification of the *neu* gene is not the only genetic change that has been implicated in the etiology of breast cancer and may have prognostic value. "We're not saying that this is the only important gene in breast cancer," Slamon remarks. Nevertheless, *neu* gene amplification is turning out after all to be a reliable guide to the prognosis of breast cancer patients, and perhaps of ovarian cancer patients as well. ■ JEAN L. MARX

New Chip May Speed Genome Analysis

An unlikely marriage between a defense contractor and Leroy Hood's DNA lab at Caltech is providing a powerful new tool for analyzing complex biological patterns

JUST AS MOLECULAR BIOLOGISTS are becoming buried in data, computer scientists are offering a shovel. DNA sequence data are pouring in as labs around the world gear up to tackle the human genome and other genomes. So far, some 30 million nucleotide bases have been sequenced, and that number is growing by about 10 million bases a year. But getting the complete DNA sequence—the ultimate goal of the human gemome project—is the easy part; deciphering it is a far trickier task. Now help may be in sight from a new computer chip, originally designed for the Defense Department.

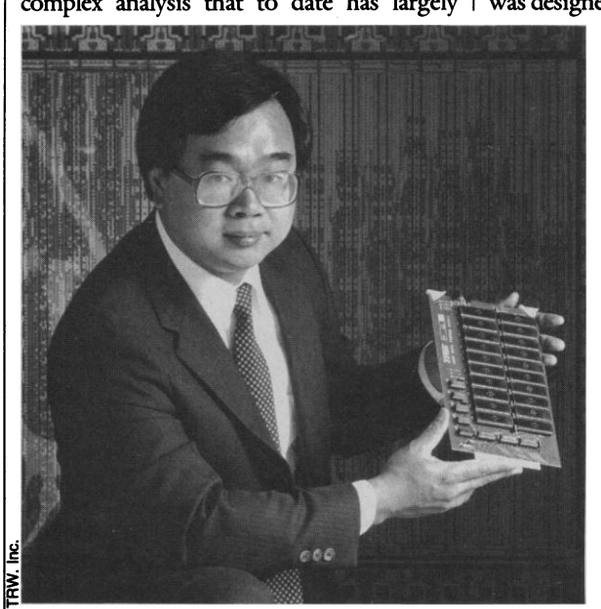
Last week Applied Biosystems, Inc., of Foster City, California, announced that it had obtained an exclusive license to this chip, heralded as the "world's fastest text scanning technology," from TRW, Inc., a collaboration that stemmed from work in Leroy Hood's Caltech laboratory, one of 11 National Science Foundation Science and Technology Centers.

It holds out the tantalizing prospect that molecular biologists will soon be able to do at their workstation computers the type of complex analysis that to date has largely been limited to supercomputers—and to do so hundreds of times faster and at a fraction of the cost. All this remains to be seen, however, as work to date has been performed only on prototypes, and a commercial product is thought to be 2 years away.

This unlikely marriage between TRW and Hood's group had its genesis some 3 years ago, when TRW's B. K. Richards heard a lecture at Stanford on the mathematics of genetics. The problem in DNA analysis, as Richards learned, is that the sequence consists of just four letters, the four nucleotide bases, repeated over and over again. How, then, do you extract the biologically meaningful information from the 3 billion letters that make up the human genome?

"Where are the 100,000 or so genes?" asks Hood. "What is the nature of the regulatory machinery? What are the sequences responsible for compactly folding in each and every cell 2 meters of DNA and 24 different chromosomes?" The answers are encoded in the string of letters.

To Richards, this decoding task seemed ideally suited to TRW's new chip, which was designed not to sift through DNA bases but to filter out important information in real time from the scads of cables and reports coming in to the Defense Department each day. A few weeks after the lecture, Richards met Nobel laureate Joshua Lederberg, president of Rockefeller University, who confirmed his suspicions. Richards, a Caltech alum, called the university and was put in touch with Tim Hunkapillar, a computer scientist in Hood's lab. Says Richards: "Tim came down to see the technology and in about 10 seconds said, 'This is a great idea.' "



TRW, Inc.

Kwang-I Yu. *"This is not to say it is better than a supercomputer, but for this particular application, it has much more computing power."*

Using the TRW chip and a Sun 3 computer, Hunkapillar designed a prototype DNA analysis system and wrote the necessary software, which will be available free from Hood's lab. To commercialize the prod-

uct, the Caltech group put TRW in touch with Applied Biosystems, a longtime collaborator that earlier developed Hood's DNA sequencing machine.

Clearly, a new approach is in order. Already, it is impossible to search through the sequence data by hand, and traditional computers and pattern-searching software are barely adequate for the task. Even with a Cray-2 supercomputer, a simple comparative analysis of the sequence information that comes in each year with the existing DNA data base would take more than 5000 hours, asserts Hood. A few mathematicians and molecular biologists are working on faster algorithms for DNA pattern recognition, work that is considered promising but still quite preliminary.

Hood's group has taken a different tack. What they have come up with, in collaboration with Applied Biosystems and TRW, is "a hardware solution to what is normally handled by investigators as a software problem," says Mike Hunkapillar, vice president for research and development at Applied Biosystems—a relatively inexpensive parallel processing system that can scan up to 10 million characters a second.

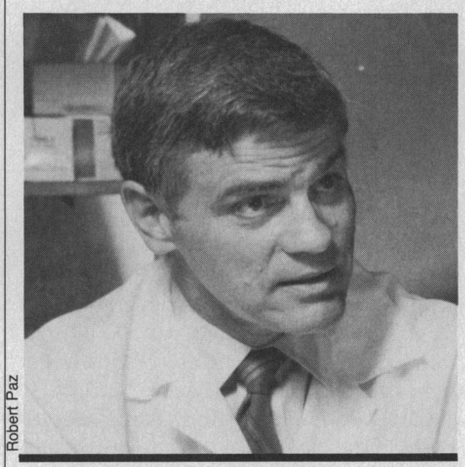
"The machine is incredibly fast. It has some limitations, but not a lot. It is very, very impressive," comments Temple Smith of Harvard School of Public Health, who recently tried out the prototype in Hood's lab. Daniel Davison at Los Alamos National Laboratory tested this new technology, installed in a Sun 3 workstation, to compare a 10,000–base gene with the 30 million bases now in Genbank, the DNA database at Los Alamos. It took 1 day on a Cray-2, says Hood, and 10 days on a VAX. "With the new technology, it took 10 minutes."

"This is not to say it is better than a supercomputer," adds TRW's Kwang-I Yu, the inventor of the chip, "but for this specific operation, it has much more computing power."

The heart of this new technology is TRW's Fast Data Finder chip, which, Yu explains, was "very specifically designed for complicated pattern matching." Yu likens the system to a garden hose, containing a long string of identical microprocessors. At this stage, each chip contains just eight microprocessors, but the chips can be arrayed on boards to create a system of essentially any size. So far, the largest prototype contains nearly 10,000 processors.

Different segments of the hose can be programmed to look for different patterns, Yu explains. "You then feed the data through the hose like water. In computer jargon, this is a pure pipeline. The data flow through at a constant rate—the system doesn't stop to do some crunching here or there." All the microprocessors work in unison, rather than sequentially, to search out patterns in assembly-line fashion.

What accounts for the speed of this system is that the instructions for pattern matching are hardwired into the processors. Most computers, by contrast, use general-purpose processors that can be programmed for different functions. Thus, with the TRW chip, the operator need only tell it which patterns to search for—say, to look for immunoglobin-type sequences, or regulatory regions, or to scan a DNA database to see if it contains anything similar to the piece of DNA you just sequenced.



Robert Paz

"It took 1 day on a supercomputer and 10 days on a VAX. With the new technology, it took 10 minutes."

—Leroy Hood

Hardwiring does bring a certain loss in flexibility, which Harvard's Smith compares to the relative advantages of using a cake mix over a cookbook. The mix is faster but can only make a chocolate cake and not a pie. Likewise, with a dedicated chip, says Smith, "you can only look for a pattern in a certain way." These limitations, however, are more than compensated for by its speed, he adds.

Initially, the machine will be used for DNA and protein analysis and to communicate among the various databases. But Hood anticipates its eventual application to other complex biological questions, such as protein folding and the patterns of communication among the 100 billion neurons in the brain.

Perhaps the biggest advantage of the new chip over current software approaches is that there is no penalty for asking complex questions, says Caltech's Hunkapillar. With most software, performing a complex search in real time—say, to look for "cat or dog but not parrot"—involves several iterations as the computer first searches for cat, then dog, and so on. Hunkapillar adds: "The more complex your pattern is, the longer it takes."

Similarly, some software approaches require indexing, in which the data set is broken down into words before it can be searched, with tremendous costs in both time and money. By contrast, the TRW chip requires no preprocessing of the data set and no iterations. The system is also very forgiving, explains Hunkapillar. It can handle both ambiguous questions and misspellings with no penalty in speed.

In fact, the more complex the pattern, the greater the speed advantage, says the Caltech computer scientist, who explains that "it ranges from a little bit faster to orders of magnitude faster, depending on the specific search. As long as you have enough processors, you can put in an enormously complex pattern and search in the same time it takes to search for just 'cat.' You can have a pattern that fills an entire typewritten page without affecting speed. And you can put in multiple patterns—90 patterns at once—on the same single reading of sequence."

The other advantage is cost. The Connection computer, a massive parallel processing machine made by Thinking Machines Inc., can tackle DNA analysis at similar speed. But compared with the $2 million or so price tag for that computer, the Applied Biosystems machine is likely to be cheap. Without a marketable product, Applied Biosystems is understandably vague about price, though they say it will be affordable. Others estimate the cost at $40,000 or so—cheap enough so every university if not every lab, could have one, says Smith.

Smith suspects that the ultimate solution to DNA pattern recognition may lie in more sophisticated pattern-searching algorithms combined with dedicated hardware, such as the TRW chip. His group and several others are now working on artificial intelligence approaches, which he suspects will be able to discern far more complex patterns than can the TRW chip. At this point, however, the software is painfully slow and cannot rival TRW's chip or a small parallel processing computer, made by Active Memory Technologies, recently adapted for DNA and protein pattern searches.

For now, the biggest gain from this new technology may be in the type of questions it lets you ask, predicts Hunkapillar. "You can ask questions that in the past you couldn't unless you had a Cray."

Smith agrees: "It gives molecular biologists the intellectual freedom to ask goofy questions." ■ **LESLIE ROBERTS**