

# A Map to Nowhere

THE GENOME ISN'T A CODE AND WE CAN'T READ IT

BY TOM BETHELL

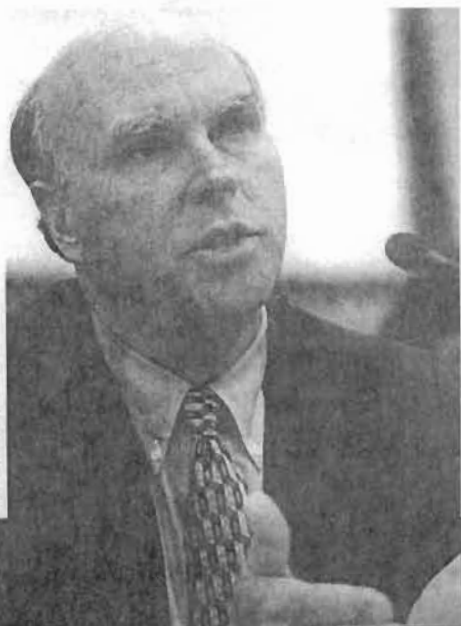
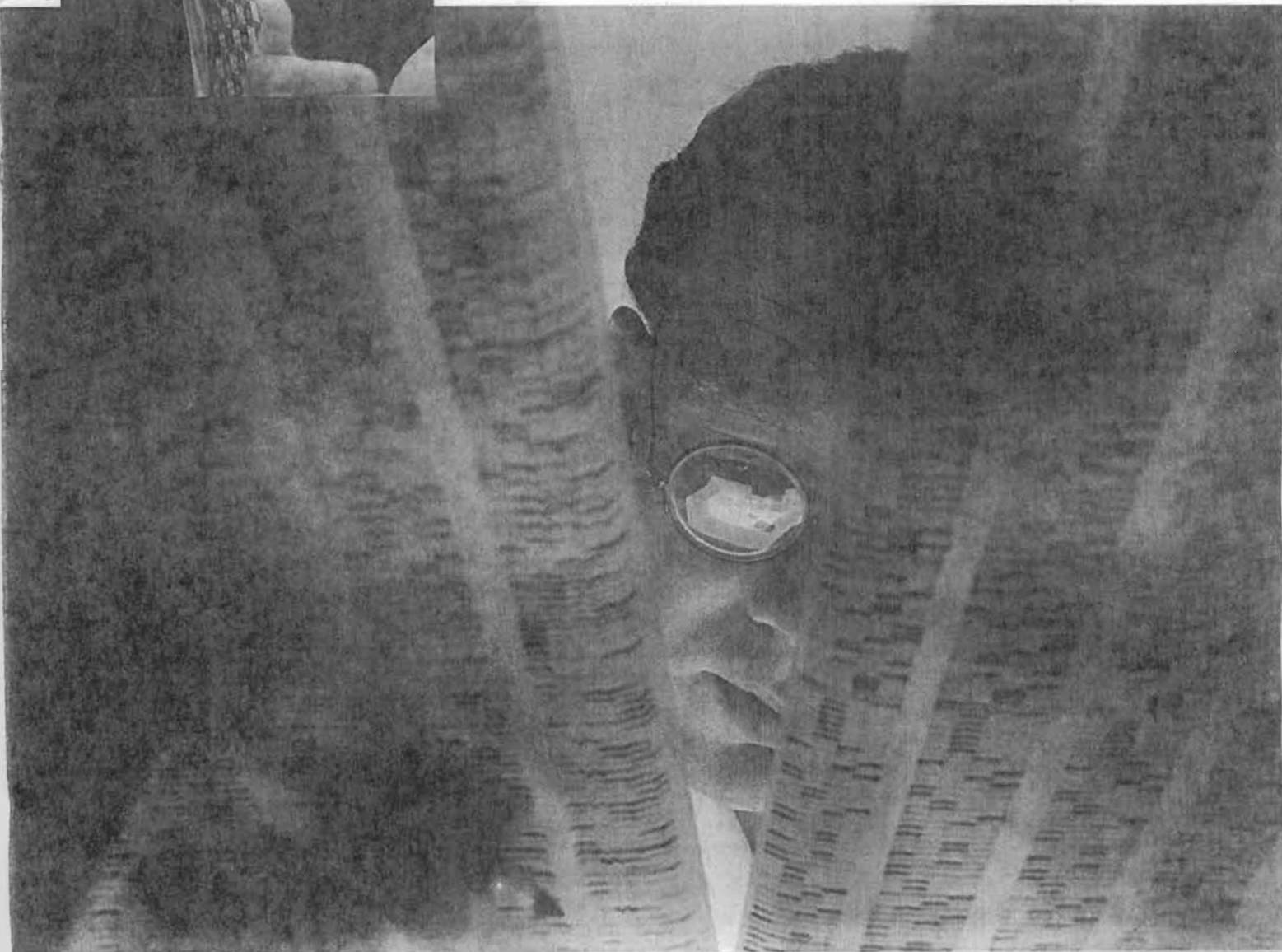


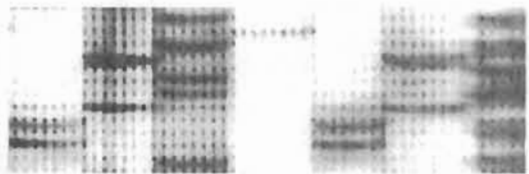
PHOTO: AP PHOTO/LESLIE KOSOFF © AP/ORBIS

**T**he principal actors had appeared in the White House last June—Francis Collins of the National Human Genome Research Institute, and J. Craig Venter of Celera Genomics. Now they were back with a supporting cast and a more detailed analysis, in the Capital Hilton Hotel, with the TV lights glinting off the ballroom chandeliers, 250 journalists packed into the hot room, and James Watson of DNA fame on hand to take a bow. There would be one more blaze of publicity about the project to decipher the human genome. The new findings were about to be published in long articles, with a comical abundance of co-authors, in the journals *Nature* and *Science*.

PHOTO: ORBIS PHOTO/ORBIS INTERNATIONAL, LLC



New Mexico's Sen. Pete Domenici, an early and eager supporter of the project on Capitol Hill, received a vigorous round of applause. He was sitting next to Watson, and in his remarks Domenici said that Watson had just whispered to him, "You must say that this project was congressionally driven." The senator added, "And that's true.... This project,



in terms of the U.S. government, was truly started in the Congress." One of the new buildings going up on the "campus" of the National Institutes of Health will surely be named after Domenici.

One news item was prominently reported. The number of human genes is now believed to be about 30,000, one-third or even one-fourth the number recently estimated. At first this was played as the familiar object lesson in humility for us self-satisfied anthropoids. We thought we were at the center of the universe. Silly old us! Now, our supposedly overweening pride receives another setback. For we have "only twice as many genes as a fruit fly, or a lowly nematode worm," said the ever-so-humble Eric Lander, head of genome research at the NIH-funded Whitehead Institute in Cambridge, Mass. "What a comedown!" The journalists roared on cue. That would be the sound-bite for National Public Radio, you knew, and the *Washington Post* would publish it the next day.

There was, however, a more disturbing implication. It took a few days to sink in. There followed a kind of appalled silence, and then the alarm bells began to ring, if only faintly. "The way these genes work must therefore be far more complicated than the mechanism long taught," whispered the *Washington Post*. The alarms will grow louder. For if what Craig Venter said is true—and it was accepted by James Watson when I

spoke to him immediately after the press conference—the genetics textbooks will have to be rewritten and the therapeutic breakthroughs promised by the map of the genome may not come for decades, if ever. No one at the press conference disputed Venter's claims. That included the editors of *Science* and *Nature*, who made brief remarks.

## The genetics textbooks will have to be rewritten and the therapeutic breakthroughs promised may not come for decades.

Craig Venter's opening statement contained the bombshell. Since last June, he said:

[O]ur understanding of the human genome has changed in the most fundamental ways. The small number of genes—some 30,000—supports the notion that we are not hard wired. We now know the notion that one gene leads to one protein, and perhaps one disease, is false.

One gene leads to many different protein products that can change dramatically once they are produced. We know that some of the regions that are not genes may be some of the keys to the complexity that we see in ourselves. We now know that the environment acting on our biological steps may be as important in making us what we are as our genetic code.

The old dogma of genetics, prevailing in the textbooks to this day, was that one gene made one protein. George Beadle and Edward Tatum won the Nobel Prize in 1958 for their formulation of this doctrine. Usually it is put, "one gene, one enzyme," but an enzyme is a special kind of protein, and today it is most often expressed as "one gene, one protein." Now, in front of some of the country's most eminent molecular biologists, Venter was telling us that there may be ten times as many proteins

as genes. "Perhaps 300,000 proteins," he said at one point.

The genome consists of a string of four nucleotide bases, symbolized by the letters A, C, G, and T, and the string is over 3 billion letters long. Over 98 percent of the genome appears to be inactive, consisting of "non-coding regions," and some dismiss it as "junk DNA."

(Not Collins or Venter, however, who say we just don't know what it does.) The intermittent "coding" segments along the way are called genes, and they give instructions for the manufacture of the body's proteins. In people with heritable diseases, some of those proteins are defective. So what the science of genomics would do was find the defective genes along the genome string by comparing the nucleotide sequence of sick people with the genomes of well people. If the defect could then be corrected, the protein made by that gene would be restored to its healthy state. That is the underlying theory of gene therapy.

Seemingly intractable problems have arisen, however, and they have been well known to the gene hunters for several years. There are indeed diseases that are caused by a simple defect in the genome—just as in a rare case a single typographical error will radically alter the meaning of a text. (But most "typos" are immediately apparent as such and cause no defect in the reader's understanding.) The genetic character of these diseases—sickle cell anemia, cystic fibrosis, and Huntington's Disease are among the best known—was initially established not by searching the genome but by tracing them in family histories. This showed their predictably inheritable character. When both parents carry a copy of the "typo," there is a good

chance that their child will have the disease.

With many of these diseases the defective gene has indeed been discovered. The problem is that a cure is still no closer. "We've had our gene since 1989," Dr. Robert Beall, president of the Cystic Fibrosis Foundation told the *Wall Street Journal* last June. Yet no gene therapy for the 30,000 cystic fibrosis sufferers has yet emerged. In other words, while we have waited for the "breakthrough" in mapping the human genome, so that we may cure diseases, this "breakthrough" already occurred for cystic fibrosis a dozen years ago, to no effect. In the case of sickle cell anemia, the genetic defect has been known for over 30 years, yet the disease can only be treated by non-genetic therapy. It is the same with all the other heritable diseases. Because the genetic defect is in the germ-line, the "typo" occurs in every one of the one hundred trillion cells in the body. The problem for genetic engineering is how to get the "corrected" gene into enough cells to make a difference. It's an unsolved problem.

True genetic diseases are rare, appearing in only one or two percent of all births. Therefore, it is said, they "do not fit the business model." Nonetheless,

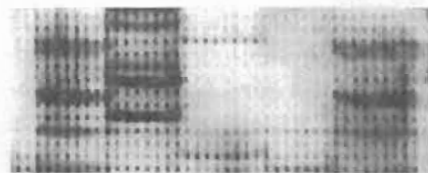
therapy quietly shifted toward far more common and potentially highly profitable diseases; in particular cancer, heart disease, AIDS, and Alzheimer's. Dr. Collins told us that his own lab at NIH is engaged in a "huge and very complicated" search for genes for adult-onset diabetes, the latest cause celebre for gene hunters.

All along, however, the idea that these very widespread conditions are caused by the sort of clear, isolated genetic "misspellings" that seem to explain sickle cell or cystic fibrosis was entirely speculative. In the case of cancer, nothing definite has been found after a 20-year search. Cancer researchers will tell you otherwise and mean it, but it is not at all reassuring to learn that they claim to have found over one hundred "oncogenes" that "predispose" us to, or are "associated with" cancer; and that about 30 defective "tumor suppressor" genes have also been located. That is far too many to be useful, and therapeutic benefits have been elusive. Some of these genes have already been patented, which means that companies can, once again, charge a monopoly price to "screen for" the presence of these oncogenes. But their causal role

one message, but are combinations of separated segments along the genome. Between them lie intervening segments which can be cut out by the cell, as it translates DNA into proteins, and the relevant or coding parts (called exons, as opposed to the intervening parts, which are called introns) can be put together in numerous different ways. Gene therapy send different messages and make a variety of proteins *as the occasion demands*.

Imagine that an intelligence service were to discover some unintelligible messages being sent by a spy. At first the intelligence agents naturally assume they are looking at a code. They assume the task of decoding will be straightforward. But on closer analysis it turns out that the message means one thing if the signal has been received and acted upon, another thing if it has been received and not acted upon, another thing if the receiving apparatus is not switched on, and so on. Rather than just a code the message is a bit like a set of rules for a rather complex interactive game. There are feedback loops, and circuits within circuits, and a lot of things happening inside the cell but outside the genome, in the unfashionable realm of cytogenetics. NIH-funded geneticists don't even

## About 30 defective "tumor suppressor" genes have been located. That is far too many to be useful.



millions of dollars were spent by medical research institutions to locate these genetic defects on the genome. Still, nothing has come of these findings, beyond the patenting of screening tests, which can be used to warn couples that any of their children might have a one-in-four chance of being born with a disease.

But the field of biotechnology can expect little or no payoff from diseases that affect anywhere from a handful to a few thousand people worldwide. As a result, some years back the focus of gene

has never been established and probably never will be. There has been a concerted effort to blur the difference between those rare diseases that are plainly and predictably heritable, and the common ones that are not.

Contrary to what the headlines say, the genome has not yet been decoded. It might never be, as the genome now does not appear to be a code at all in the conventional sense. It turns out that genes are not simple "strings," each one encoding for

want to think about that, because they thought that by sticking to the four nucleotide bases, they had the problem neatly "digitized." Computers would hum away unaided, 24 hours a day, and unravel the mysteries for them while they slept.

We should have known that it would not be so simple. Successful biological systems resist simple analysis for the very same reason that they are successful. Every time we gain greater knowledge of any such system we discover that it is far more complicated,



redundant, self-healing, adaptable, and resistant to "single points of failure" than it first appeared. If the functioning of the genome were as simple—and therefore easily manipulated—as the advocates of the genome project have been implying, it would be impossibly fragile. Significant genetic defects would be far more common, assuming any organism based on such an easily cracked and therefore easily corrupted code could survive in the first place.

In the case of genetics, the illusion of simplicity arises in large part from the genius of Mendel's insight in constructing the original genetic metaphor. Studying peas in a monastery garden, Gregor Mendel sorted them by their outward traits (size, shape, and color), and, observing that these traits appeared in the regular ratio of three to one, he ingeniously posited internal "factors" which occurred in "dominant" and "recessive" form. These were the genes. No one actually observed them, for no one then had the microscopes or the machinery to do so. But the theory was that when the parental contributions to reproduction are combined, one set of genes from each parent, the factors or genes do not "blend" but live on internally, one "dominating" the other in the expression of the trait that each gene controls. One gene, one trait. That was how a gene was defined. Hence, it was said, we have genes "for" this trait or for that; genes for skin color, for example. Where blending is obviously real, as in the case of skin color (a black mother and a white father have a child of intermediate color), geneticists could just say that there were several genes "for" that trait, and one or more from each parent was expressed in the child. No one has yet located the "skin color" genes on the DNA, incidentally.

Mendel made no claims about the structure of genes or how they might accomplish their task. The Mendelian gene was a hypothetical construct, possibly standing for an infinitely more complex set of processes.

Along came Thomas Hunt Morgan at Columbia University. The chromosomes had been discovered and in 1902 Walter Sutton noticed that they came in pairs—a convenient fit with the dominant-recessive understanding of the gene. Morgan set out to "map" the chromosomes, which is where the genes would have to be located. He didn't get very far with that but he won the Nobel for his valiant effort. The word *Drosophila* began to appear in the newspapers. Hermann J. Muller zapped a lot of fruit flies with x-rays and heritable body changes were observed—more Nobels for that. Enter Watson and Crick. Their discovery of the double helical structure of DNA worked nicely too, because it showed how a complex message could be replicated. "Genes," henceforth, would be thought of as nucleotide sequences along the string of DNA, which in turn was packed inside the chromosomes.


This reformulated gene was not an entirely satisfactory fit with the old Mendelian gene. But labels are powerful. In the decades since, genetics has largely consisted of an awkward attempt to combine under one name the Mendelian gene of the late 19th century and the DNA of the 20th century.

What DNA actually did was carry the instructions for making proteins, as we have seen. And proteins were not the same thing as the visible and outward bodily traits, such as chins and noses and skin color, that the Mendelian genes were said to make. Still, it was close enough for government work, and university work; it sustained the bandwagon of forward progress, and the truth was that no one really had much more than a hazy knowledge of these things anyway. And there was this—chins and noses and skin pigments *were* made of protein, so the different versions of the gene could be cobbled together, rather like a pantomime horse. Actually we had Mendel and Morgan and Watson and Crick all galumphing around onstage together under this gene umbrella, and it

held together pretty well throughout the 20th century.

So the idea of one gene, one protein was a carry-over from the earlier one gene, one trait formulation. And in their work with bread molds in the 1940s, Beadle and Tatum had seemed to confirm the idea that a gene is something that performs a single task. Now, it begins to look as though the gene is going to have to be rethought completely. It is a task that has long been postponed. The pantomime horse has begun to look ridiculous. Not only has our understanding of the gene become massively more complicated, the analysis of protein structure, which now moves to the fore, is truly a daunting prospect for biology. Proteins have 20 building blocks, not four, and are arrayed in three dimensions, not along one, like genes.

**B**oth Celera Genomics and the government-funded consortium have an interest in sustaining the old "breakthrough" refrain; Venter to attract investors, Collins to keep Congress happy. According to a spokesman at NIH's National Human Genome Research Institute in Bethesda, federal appropriations for the human genome project have totaled \$1.5 billion to date. As government projects always do, it started small, with \$10.7 million appropriated for the Department of Energy and \$17.2 million for NIH, in 1988. "Frankly, it was a gamble that we would be able to expand the [research-dollar] pie," said Maynard Olson, head of the publicly funded genome center at the University of Washington in Seattle. But expand it they did. The amount appropriated in fiscal year 2000 was \$271 million, and the estimate for this year is \$291 million. In the process, a whole new institute at NIH was created. "A lot of people don't mention those numbers," said the NIH spokesman. In contrast, at the time of the White House announcement last June, Celera said that it had spent a total of \$200 million on the project. The government is now on a



genome-spending path that will consume more dollars every year than Celera has spent overall.

And yet it was from Celera that the new understanding came, not from those with government funds. The contrived truce between the participants concealed this: It was Celera that was willing to advance the new insight even though it may end up undermining the company's original premise and business plan. (And it was the Whitehead Institute's Lander who had tried to stop *Science* from publishing Celera's article.) With the real world of investors to consider, Celera can less easily survive in the cushioned, sometimes make-believe world that government science has fashioned for itself. If there was bad news, Celera needed it sooner rather than later. Thus armed, perhaps Venter can adjust the business model.

Nonetheless, Celera's message is not likely to comfort investors. Gene therapy holds out less promise as a result of this new understanding. At the press conference, a journalist asked Francis Collins if the smaller number of genes would make medical advances easier or more difficult. "I would say easier," he said. Every gene search is like trying to find a needle in a haystack. "Guess what? The haystack just got three times smaller." But when another journalist asked a similar question about the genes interacting combinatorially, Collins retreated from the haystack metaphor. The straw interacts with itself, and the needle has other objects bumping into it, he allowed. Craig Venter said more simply that when you consider there is maybe a "tenfold expansion" in the number of proteins compared to the number of genes, it "does indicate increased complexity."

Celera will no doubt continue to sell its genome information to the big research institutions, to the pharmaceuticals, and to other biotechs—for a while at least. And the biotechs with patents will continue to charge high prices to screen for "predisposing"

genes; or for the rare but real disease-causing defects.

To some biotechs the new announcement came as something of an embarrassment. As Andrew Pollack pointed out in the *New York Times*: "Incyte Genomics advertises access to 120,000 human genes, including 60,000 not available from any other source. Human Genome Sciences says it has identified 100,000 human genes, and Double Twist 65,000 to 105,000. Affymetrix sells DNA-analysis chips containing 60,000 genes." Some of these genes have already been patented, but "if genes are not the whole story," Pollack added, "it also means those patents could be worth less." Or worthless. Venter told the London *Observer* that the head of a biotech company had phoned him in some distress because he had already done a deal with SmithKline Beecham to sell them the details of 100,000 genes. "Where am I going to get the rest?" the man asked. How long before someone starts comparing genes to tulips?

Last June, under the headline "50,000 Genes and we Know Them All (Almost)," David Baltimore, the president of Caltech and the winner of the 1975 Nobel Prize for medicine, wrote that "humans have no more genetic secrets; our genes are a book open to all to read." But in the issue of *Nature* announcing the new analysis of the human genome he wrote more soberly:

We wait with bated breath to see the chimpanzee genome. But knowing now how few genes humans have, I wonder if we will learn much about the origins of speech, the elaboration of the frontal lobes and the opposable thumb, the advent of upright posture, or the sources of abstract reasoning ability, from a simple genomic comparison of human and chimp. It seems likely that these features and abilities have mainly come from subtle changes...that are not now easily visible to our computers, and will require much more experimental study to tease out. Another half century of

work by armies of biologists may be needed before this key step of evolution is fully elucidated.

The old dream of reducing biology to physics, or of believing that something simple—four nucleotides on a string—could explain the vast complexity of the human (or any other) body, has



received a serious blow. A lot of the companies involved in this search for meaning in the genome were inordinately impressed with the idea that, with just four "building blocks," an abundance of computing power was all it would take. Strings of code with tens of millions of bases could be searched and analyzed in hours. Now, we may be looking at the beginning of the end for genomics. The word "proteomics" is already beginning to appear in science-journal headlines.

Writing in the *New York Times*, Stephen Jay Gould says that we may be now liberated from the "simplistic and harmful idea" that each aspect of our being, "either physical or behavioral, may be ascribed to the action of a particular gene 'for' the trait in question." The collapse of the doctrine of one gene, one protein, he added, "and one direction of causal flow from basic codes to elaborate totality, marks the failure of reductionism for the complex system that we call biology.... Organisms must be explained as organisms, not as a summation of genes." I think he is right.

Francis Crick was invited to attend the event. He wasn't feeling up to it—he is 84—so he sent along a videotaped message instead. It was played that afternoon at the packed, 500-seat Masur Auditorium at NIH. "We foresaw very little of what happened in molecular biology," Crick said. But the impact of

the latest development on medicine "will be enormous," and we can only hope that "it will bring us on balance more good than evil." Watson then stood up before the NIH employees and with his customary tactlessness noted that Crick was "in perfect health, as you could see." He just doesn't like to travel. "He is one of the 20 percent who has

Crick. I asked him if the two concepts of the gene could continue to coexist. He said he didn't think they differed all that much. He analogized the shift in understanding to Einstein's modification of Newtonian physics. "We have relativity but it doesn't affect the way artillery works," he said. Newton was still largely right. When I asked Venter the same

## Now, we may be looking at the beginning of the end for genomics.

had a heart bypass and whose brain isn't affected," Watson added.

After the press conference, I spoke to Watson briefly. A crowd was clustered around him, some of the journalists getting him to sign the issue of *Nature* in their press packets. A black woman asked him how long before we would see medical results from the genome. He said something about our children, then corrected himself: "Our grandchildren will have better lives because of what we are doing," he said. "You're worried about breast cancer," he said to the woman. "I'm worried about senility." He is 73. Forty-eight years have passed since his publication of the double helix with

question, he gave me to understand that this would take too long to discuss, but that the question was pointing us "in the right direction."

I asked Watson if one gene can give rise to ten proteins. "Some genes can give rise to 50 different proteins," he said. No problem! He was unruffled, content, still in control at mission central. The new knowledge about genes and proteins would be smoothly integrated into the received wisdom of molecular biology, apparently. The higher councils were taking it all in stride. There was no cause for alarm. But those who would like to put the news to medical use are frowning their brows.

