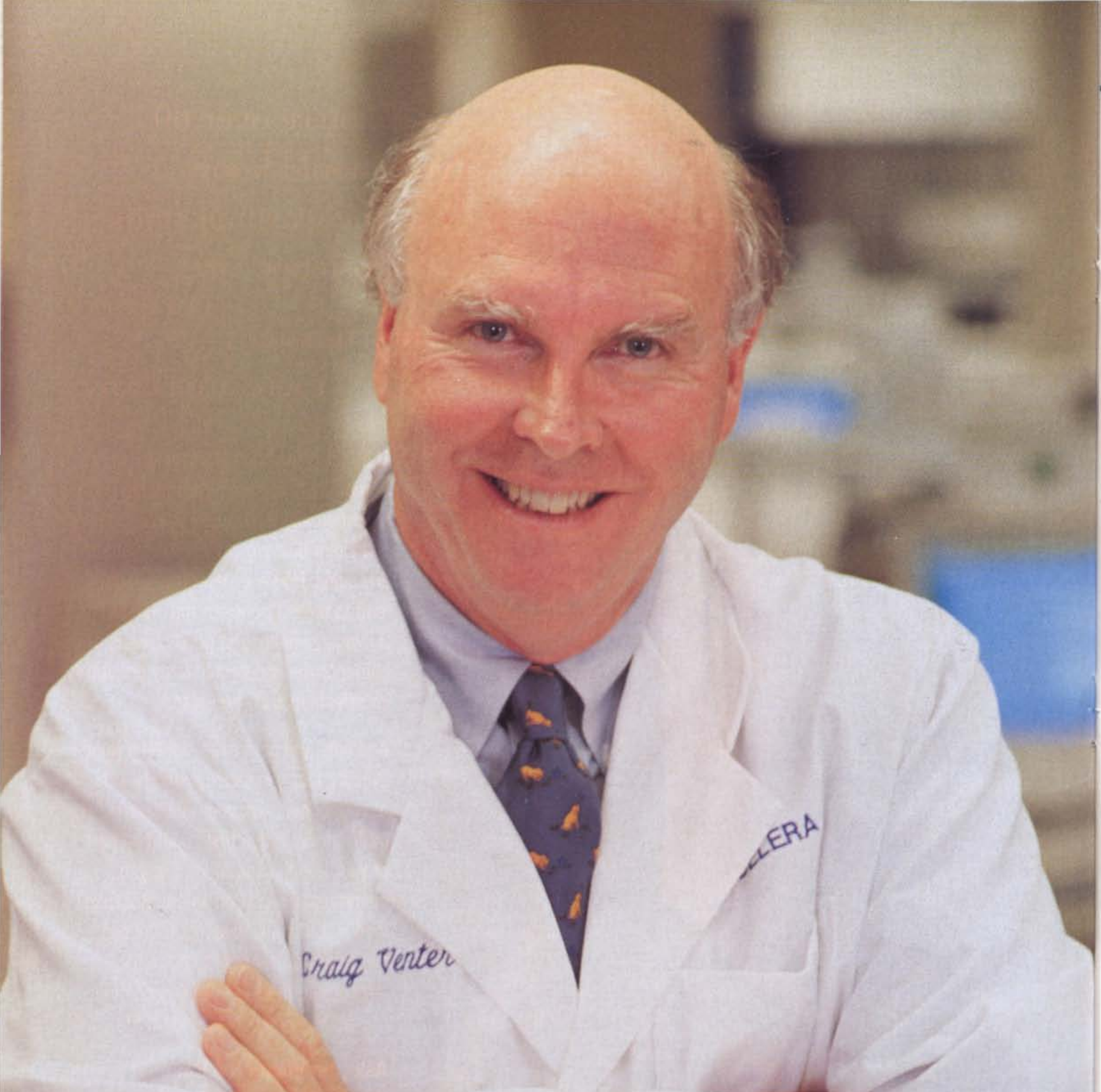

Craig Venter

# FOR DR. CRAIG VENTER, DISCOVERY CAN'T WAIT!

*Sequencing the human genome will signal one of the greatest biological accomplishments of our time.*

## by Cory SerVaas, M.D. and Patrick Perry

*Photography by Bill Fitz-Patrick*

In the world of medical discovery, speed matters. No one knows this better than the leader in the race to discover the blueprint of humanity, the human genome—Dr. Craig Venter. His company, Celera, is paving the way to the next revolution in medicine. The *Post* visited Dr. Venter and his distinguished colleagues at Celera's headquarters in Rockville, Maryland. Our mission: to learn what impact this new world of genetic breakthroughs will have on preventing disease in the years ahead.

Witty and personable, Dr. J. Craig Venter learned early in life that time is short and each day precious. In his office hangs a plaque that sums up his personal philosophy. "*Time and tide wait for no man.*" The tide, of course, refers to Venter's avocation—sailing the seas aboard his sloop with his wife of 21 years biologist Claire Frasier, Ph.D.

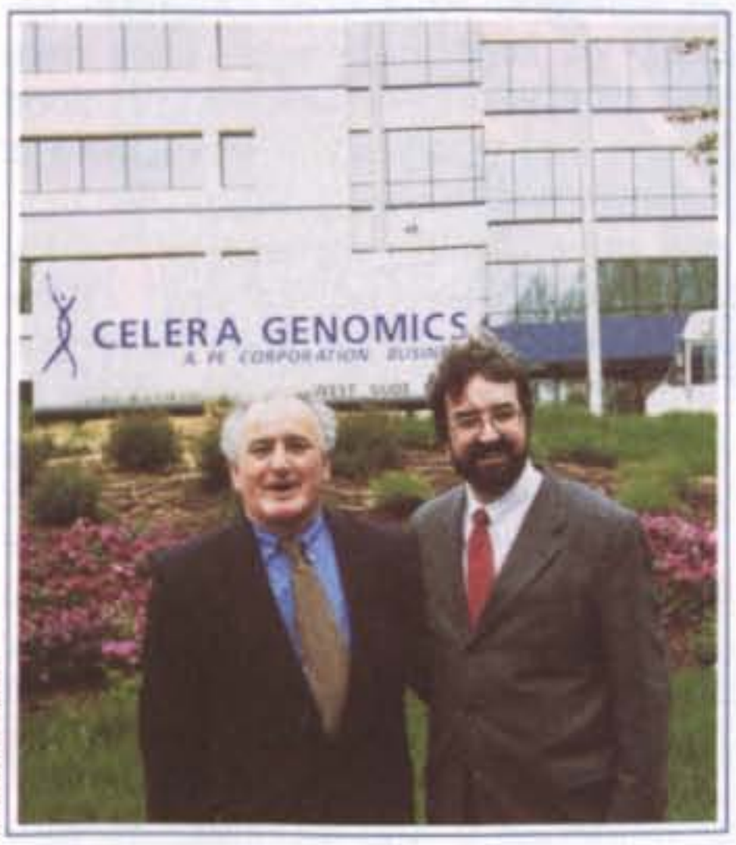



PHOTO: PAT FERRY

**Post *editors Cory SerVaas and Patrick Perry invited two prominent genetic researchers to assist them in the Celera interview: Dr. Michael Conneally (left), distinguished professor of genetics at the Indiana University School of Medicine, and Keith March, M.D., Ph.D., director of the Center for Vascular Biology and Medicine and associate professor of medicine, Indiana University School of Medicine.***

PERRY: We noticed the photos on your office wall of you as a young medic working in the hospital during the Vietnam War, a chapter in your life that inspired you to go into medical research.

VENTER: Yes, I'm the skinny guy there with hair. It's hard to imagine. [Laughter]

SERVAAS: Do you get your baldness from your mother?

VENTER: Yes.

CONNEALLY: Discovering a cure for baldness would be a lucrative area.

SERVAAS: Can you do anything about it early on?

VENTER: You can be proud of it. [Laughter]

SERVAAS: We like your slogan: "Speed Matters: Discovery Can't Wait." You have been quoted as saying that soon, people may be able to send in a swab of their cheek cells to Celera to discover their individual genetic profiles and to check for predisposition to various diseases. You predicted that this service might soon be available and that people could subscribe to this service for approximately ten dollars a month. That sounds great, because our readers can afford ten dollars a month to find out if there is new information about their genetic makeup. You believe that early intervention is very important in diseases, such as diabetes and bipolar disorder. How soon can we hope that you will be able to make such a service available?

VENTER: Those were examples we gave when talking about what would be coming in the future. It's not ready yet. But we have almost finished the human genetic code. We announced a short while ago that we finished the sequencing phase. We have sequenced about 20-or-so million pieces—it is the world's largest jigsaw puzzle—and we are now using our supercomputer to assemble those pieces to get the exact genetic code of all the chromosomes.

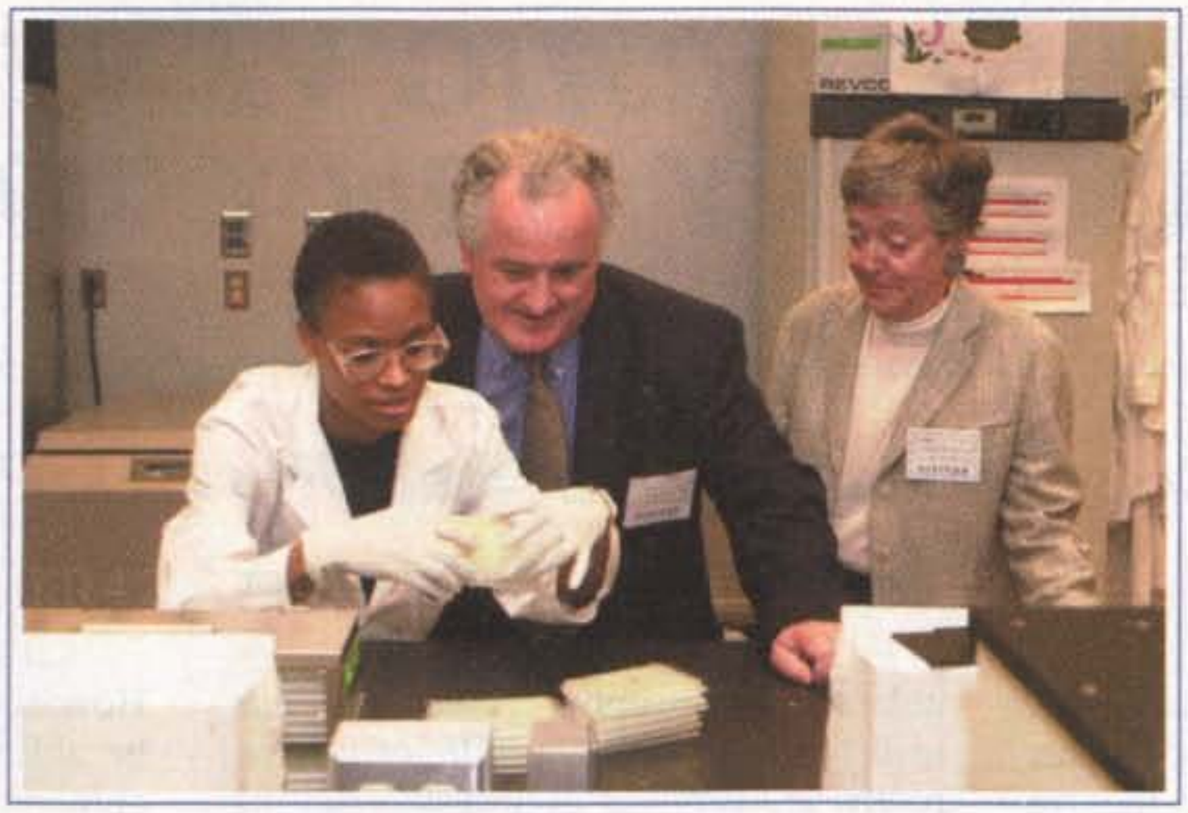
This is the beginning of the new phase of medicine and understanding. It is hard to imagine that even ten years ago, we knew only a small percentage of the human genes. Everything that we thought we knew about medicine was without understanding most of the basic science of our own physiology. We are going forward on the assumption that not only does speed matter, but knowledge matters.

But speed does very much matter. I am not getting younger each day. And look at the millions of people in this country with cancer right now. The changes in science, for the most part, have not yielded the tremendous breakthroughs that we had hoped for. It is because human and cellular physiology is extremely complex. We have 100 trillion cells in our bodies. That's not what we see when we look at each other. It is hard to imagine that complexity.

In each one of our cells, we have 80,000 to 100,000 genes. The same chromosomes are in all those 100 trillion cells, and each cell is slightly different. They are expressing different

genes and different amounts dynamically, so when something goes wrong with cellular growth, such as in cancer, it's hard to understand right now because we haven't had the complete instruction book. We haven't even known all the components. If you don't know all the components, how can you possibly know what goes on, other than just by finding things by luck? There have been some amazing breakthroughs. But to comprehensively understand and change things, we need to understand all this information.

Getting down to personalized medicine, you and I differ in around 3 million letters of genetic code out of 3 billion. Only about one in a thousand letters of genetic code is different between any of us, and only a small percentage of those actually lead to the differences between us. As a species, we are much more alike than we are different, yet we always focus on the differences. Those minor differences in the genetic code, however, determine all human traits at some level—not in a causal sense, not in an absolute deterministic sense, but in an influencing sense depending on the environment. Some of those differences will determine increased risk for cancer. Some will determine increased longevity—the wellness genes. Our goal is to empower individuals with having their own information about their own lives.



***Basic scientists, such as Gjange Pitman (left, shown here with Drs. Conneally and SerVaas), prepare plates of amplified DNA ready to be sequenced for analysis. Only 18 months old, Celera has built one of the world's largest computing centers and assembled more DNA sequencing power under one roof than anywhere else on earth.***

In the middle of the last decade, for example, Dr. Bert Vogelstein of Johns Hopkins University and I discovered three new genes that cause colon cancer. They are mismatched DNA repair enzymes that were discovered by comparing the human sequences to those from *E. coli*. Right now, colon cancer is totally treatable and curable if caught early enough. But what do we tell people? We advise them at age 50 to go in for their first colonoscopy, to look for symptoms, or to look for blood in the stool. But if, in looking at the genetic code, we know that somebody has a greatly increased risk of getting colon cancer—not an absolute determination but an increased risk—they would be smart not to wait until 50 or until symptoms appear. We hope in the near future from our efforts to have new, simple blood assays that would tell whether there is an increased risk for getting colon cancer and whether symptoms are starting to appear before any of their earlier warnings, so it could give you power over your life. If you know you have an increased risk, you can do something about it, instead of just waiting for symptoms to appear. This is part of the personalized medical paradigm that we are trying to build and go forward with. Mapping the human genetic code is just the first step in this whole process.

SERVAAS: If you are going to have diabetes because your mother or your grandfather had diabetes, you could find out early and do things to keep from getting complications from diabetes, such as neuropathy, amputations, and macular degeneration. You could, for example, take care of your feet.

VENTER: Certainly. Not all diabetes is absolutely genetic. Everything we have is a combination of what is in our genes and what is in the environ-

## Venterisms

* Our goal is to empower individuals with having their own information about their own lives.

* If you know you have an increased risk, you can do something about it instead of just waiting for symptoms to appear.

* The way medicine is practiced now, most drugs work on 30 to 60 percent of the population, yet we give the same drugs to the whole population.

* That is individualized medicine going forward. The goal is to deliver drugs that will treat your disease and not kill you in the process.

* When something goes wrong with cellular growths, such as in cancer, it's hard to understand right now because we haven't had the complete instruction book.

* We have 100 trillion cells in our bodies. That's not what we see when we look at each other. It is hard to imagine that complexity.

* As a species, we are much more alike than we are different, yet we always focus on the differences.

* The future of medicine will be preventive medicine.

* Medicine has to become individualized, based on the individual's genetic code.

* Instead of scientists spending decades of their lives trying to find genes, they can get them in a ten-second search of Celera's database, then start studying how they work together to change our physiology and cause disease.

* This will be the dawn of a true, new era in biology and medicine. It will happen much sooner than it would have without Celera.

* Our goal is to educate physicians and, more important, to educate individuals because then with that knowledge, we have power over our own lives.

ment. A lot of people get diabetes from overeating: the key is dietary control. But if you know that you have an increased risk of getting diabetes from your family history, we can also tell that from beginning to understand the genetic code better; then from early in your life, you will have a clearer direction to go.

That is just one example of how knowing this information at a very early stage can help in prevention. I predict that at some time in the next 5 to 20 years, as babies leave the hospital they will have their genetic code determined, and their parents will have that code on whatever the equivalent is of a DVD disk at that time. Parents can then go to the Celera database to understand what that information means in terms of potential for different human traits, the potential for disease, and the potential for wellness in areas that we can do something about because we know something about them or how to prevent them. The future of medicine will be preventive medicine.

SERVAAS: Newborn screening labs will then be able to send their DNA to Celera?

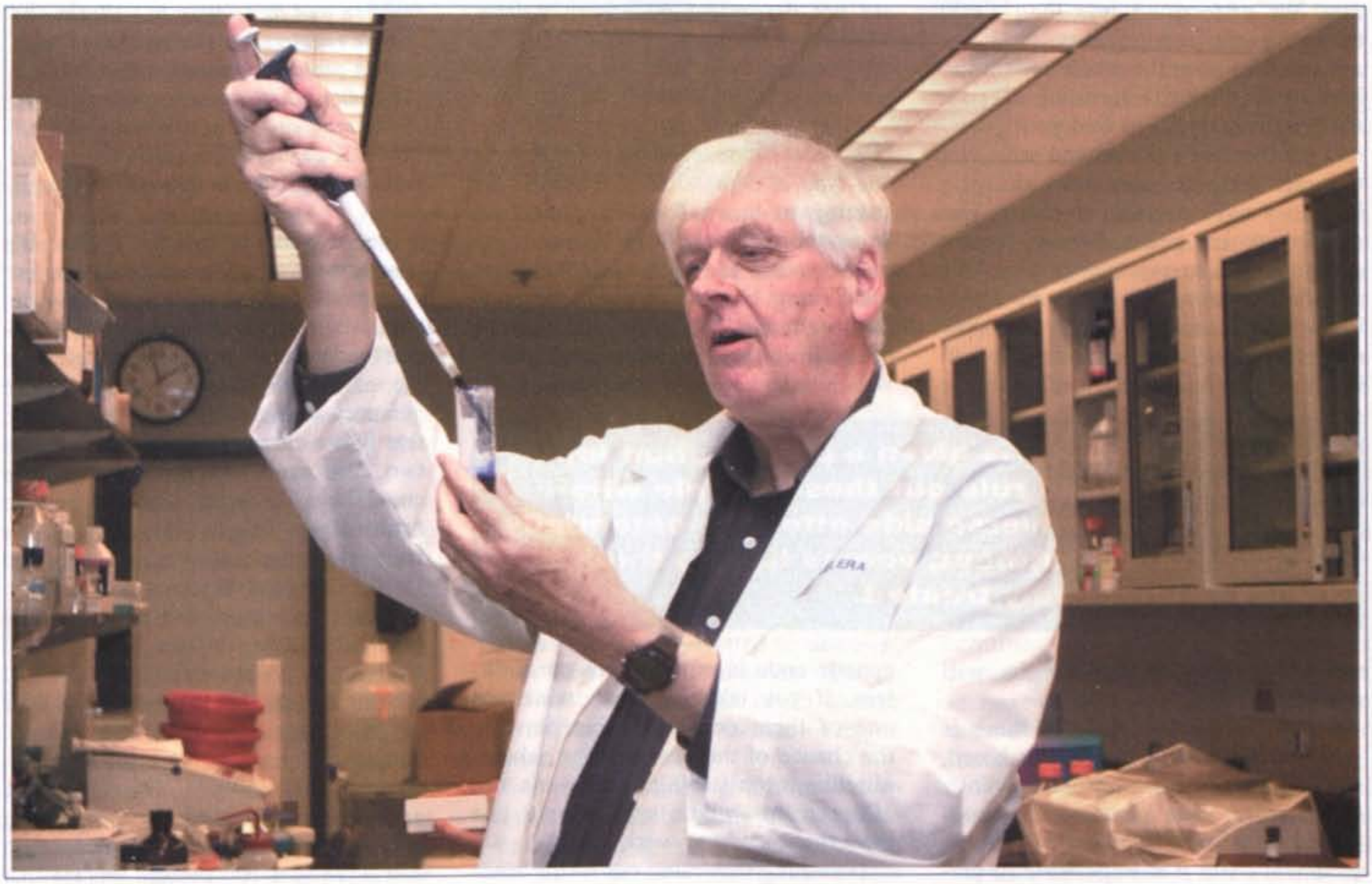
VENTER: Celera or others. It will be a very new technology, and it is hard to predict how fast this new technology will be coming. But it has been pretty amazing to me how fast this technology has come along already.

SERVAAS: But again, speed matters. If you are Katie Couric—whose husband, Jay Monahan, died at age 42 from colon cancer—wouldn't you want your daughters to have this genetic information for only ten dollars a month?

VENTER: Those are approximate numbers, but yes. Right now, Johns Hopkins University does colon cancer gene screening, so that is something that is available now. With certain increased risk of genes associated with breast and ovarian cancers, there are some tests that you can begin to do now if there is a clear-cut family history.

SERVAAS: But would most people know about that yet? And is it expensive?

VENTER: It's not so much that it is expensive; it's probably more a situation that, as you said, many people aren't aware that these tests are even available. A lot of physicians are not yet aware. So another key goal of what Celera is going to do, aside from helping pharmaceutical and biotech companies and universities come up with new understanding and treatments for disease, is to help educate physicians about what is available and how to use this information. That is one thing that we are going to do through our Web site. If you go to celera.com now, you will see that we have a newsmagazine that begins to educate people about genomics by providing



*Nobelist Hamilton O. Smith, M.D., director of DNA Resources at Celera, worked with Dr. Venter to complete the first genome in history in 1995—Haemophilus influenzae. Shown here in his lab, Dr. Smith performs the very first step in the sequencing process, taking the chromosomal DNA and breaking it down into millions of tiny pieces. "It is a real art form," says Craig Venter, "and Dr. Smith does this with his own hands."*

information about it. So we are starting an education process. Our company is only 18 months old, so speed does matter to us. We have changed a lot in an 18-month period of time. Our goal is to educate physicians and, more important, to educate individuals because, then, with that knowledge, we have power over our own lives.

SERVAAS: Once again since speed matters, how soon do you think it will be before this information is available and people can access this valuable knowledge through a subscription to your database?

VENTER: Speed, again, does matter. This information is going to cause a revolution in science across the board. We are going to have more studies on every gene and every disease. I spent ten years of my career just trying to find one gene—the protein that adrenaline interacts with to control our heart rate and other functions. That is how biology was practiced a decade ago. We understood less than 2,000 human genes and proteins. Now we are going to have the whole repertoire. Instead of scientists spending decades of their lives trying to find genes, they can get them in a ten-second search of Celera's database, then start studying how they work together to change our physiology and cause disease. We are going to see a tremendous growth. This will be the dawn of a true, new era in biology and medicine. But it is impossible to predict whether it is going to be breast cancer, heart disease, or

**"...Before anybody is given a prescription for that drug, we can rule out those people who would have an adverse side effect or potentially die from this treatment, versus having their diabetes effectively treated."**

where the first breakthroughs will come in a predictable fashion.

I feel very confident that there is going to be a shift across the board, and the knowledge of human biology is going to lead to a better understanding of the disease process and better treatments. But it is impossible to predict a time course, because these are discoveries waiting to be made. It will happen much sooner than it would have without Celera.

SERVAAS: Indiana University is a leader in the study of bipolar disease. We have scientists there who are quite good at researching the genetics of the disease and trying to find the right genes that may predict onset. It is very important for a youngster who inherited a gene for bipolar disease and is going through his first depression to be placed on the right medications. Many adolescents with bipolar disorder end their lives because they don't understand that the depression is a temporary thing. A physician might prescribe the wrong medication, such as Prozac, instead of lithium or a lithium-type drug. Would you be able to use your genetic databases to help these families through a subscription to your database so that they could determine whether a child may have inherited bipolar genes?

VENTER: The one thing that is missing is an understanding of the genes linked to bipolar disease. The trouble is that most complex human functions are not dependent on a single gene. Multiple genes may be involved in depression and these different areas. There are some exciting studies coming in and a potentially new gene diagnostic test from one pharmaceutical company for one type of depression. The studies are starting to move forward. Once we know the clear-cut links between sets of genes and the increased likelihood for getting bipolar disease, someone has predicted that that test can be done. To make it clear particularly with human behavior and diseases, the genetic code is only part of the problem. If you take identical twins and one of them develops schizophrenia, the chance of the other getting schizophrenia is only 50 percent—it's not 100 percent—which says it's not just the genetic code at work here.

One thing that came out of the studies that we have done in which we tried to find the minimal set of genes involved in cellular life, for cell-replicating life, is that the definition of life

***Completing the map of the human genome is just the first step, according to Dr. Venter. Understanding gene function and its relation to wellness and disease requires knowledge of the protein-protein interactions that carry out bodily functions—an emerging field of research called "proteomics," and Celera is now building its own proteomics facility to accomplish this.***

is not dependent on the genes. It's context-sensitive; it's what is in the environment. Instead of this ongoing debate where some people say it is nature versus nurture, it is very clear to me, as the one who has decoded more genomes than anybody else in history, that it is a clear-cut combination of the two. You can't separate the two. You could have the same identical genetic repertoire—which identical twins have—with the same genetic code, and yet they develop different personalities, different lifestyles, and different diseases, depending on the environment. Nothing is absolute. People who are looking for absolute answers from their genetic code will be disappointed. People will have to understand statistics and probabilities. If you have a greatly increased chance of getting Huntington's disease, does that mean that there is a 100 percent chance that you will get Huntington's disease, or does that mean you have ten times the average of getting Huntington's, compared to the rest of the population?

The other challenge is to help physicians and the general population understand what this information means. We all want oversimplistic definitions. Science has been looking for these for a long time. There was a so-called cystic fibrosis gene. It turns out that the chloride ion channel that is linked in some cases to cystic fibrosis is also linked to many other diseases. Genetic changes in that one gene can cause cystic fibrosis, but they can also cause chronic pancreatitis, male sterility, or no disease whatsoever. Until we understand that complete repertoire, it's important how people get this information and what backup information they get with it.

The research has to go a lot further. You and I both want answers. I would love to be able to predict my own future and understand what it means. But we don't exist independent of our environment, so we have to understand the relationship between the genetic code and diet. What we put into our bodies very much affects the outcome. We are very early on in the science of this. Hopefully, having this new information will move it forward much more rapidly. But we are not ready to do these tests because we don't have the answers yet. We don't know the right tests to do, and we don't have the right answers to give people.

SERVAAS: How are you working with university centers or pharmaceutical companies to get the right answers? You mentioned Huntington's, and we are proud of what Dr. Mike Conneally, professor of genetics at the Indiana University School of Medicine, did for Huntington's. I thought that 50 percent of the offspring of someone with Huntington's disease would have Huntington's. I didn't know environment had anything to do with that.

VENTER: There is a whole range of degrees of symptoms, and even ages where disease occurs. We don't understand all of that, yet. Huntington's disease is probably more the exception than the rule. It was thought that there were many diseases linked to single genes. It turns out that that is probably not the case. Nothing in our bodies works on its own. That is the importance of having this broad range of information.

SERVAAS: Is cancer the next step for you? With so many facets to your career, are you working on a lot of things all at once?

VENTER: Celera is trying to provide this information across the board. We are trying to push the revolution at every front. We have some of the biggest and best pharmaceutical companies in the world using our data right now through an Internet link into our computers as we speak. They are using this information to try to find the targets—or genes—related to diseases, where having a drug interfere with the function of that protein could change the outcome of the disease. That is happening right now.

We are providing the information to universities and to biotech companies. The research is going forward right now with this new information across the board.

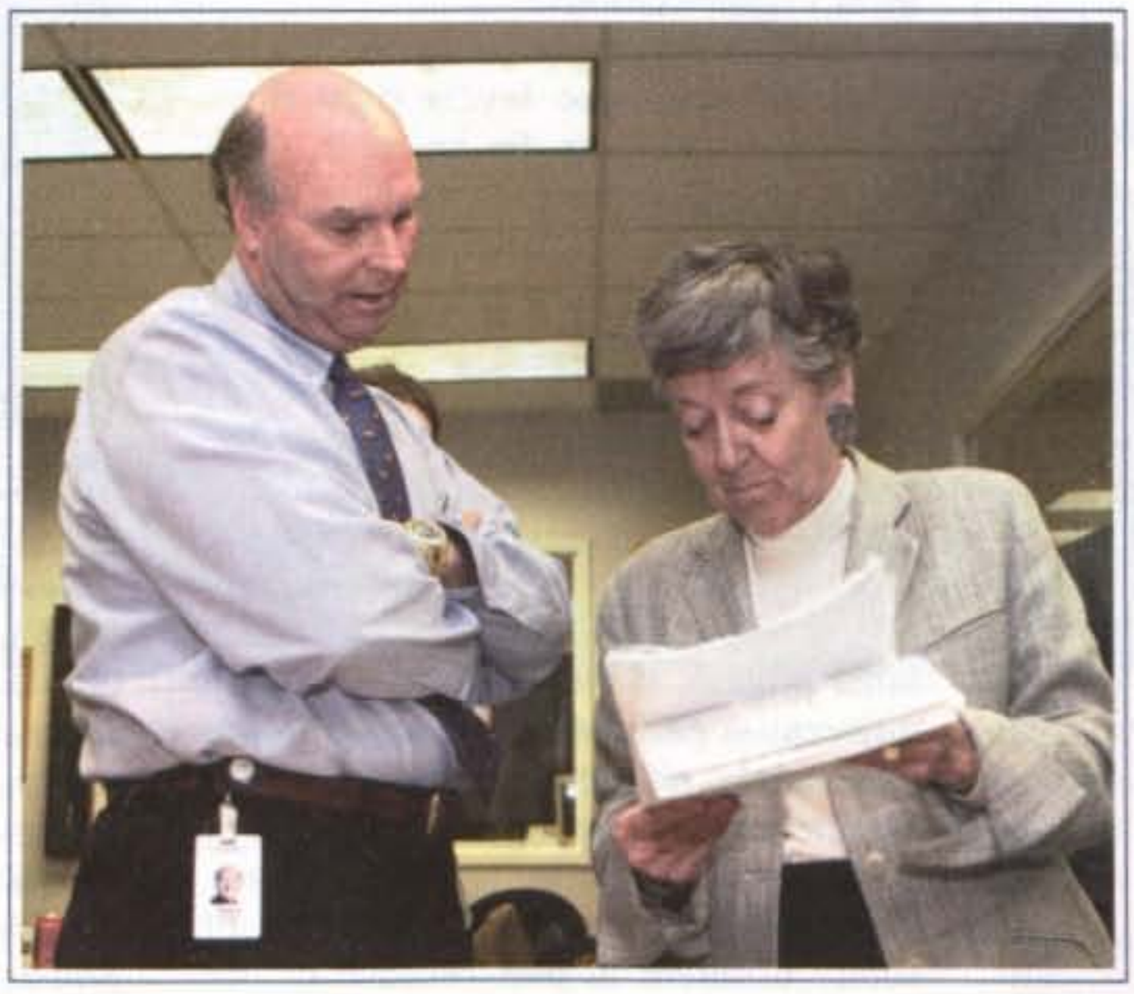
We are starting a new initiative where we are going to sequence all the proteins in the body, because we can't predict all the structures just from the genetic code. An example is the gene for insulin. Everyone thinks that insulin just results in a single product. Actually, there are four proteins made from the insulin gene. Two of those come back together to form the insulin molecule; the other two were thought to be just waste products. It turns out they are hormones on their own. From one gene, we get three different hormones. We don't yet understand the complexity with genes and proteins that have been studied for decades. So we have a very long way to go with the others.

I am telling you that speed matters but it is going to be a long time before we have all the answers. But that is why speed matters so much, because we need to change that equation.

SERVAAS: How does this incredibly complex and intricate genetic information affect your feelings about how we got here? In Craig Venter's mind, how did DNA come to be?

VENTER: One of the biggest experiments—and we will know the answer by the end of this year—is whether there are any human specific genes or not. It is thought that human development is due to changing the minute instructions associated with each gene. Instructions which direct that one gene should be turned on in the brain at this time of development to this level, instead of another part of the brain at a different time of development at a different level. It is the changing of the rheostats or dimmer switches that determines how much of a protein is turned on or turned off that is thought to have changed the human brain to what it is. We will know that in a few years.

I have sequenced viral genomes, bacterial genomes, plant genomes, the fruit fly—the first insect, and now the human genome, and it is clear that the sequence is largely identical across all life forms that we have looked at. We found the colon cancer gene that I mentioned by comparing the human



*Dr. Venter and his team are in final assembly of what he terms "the biggest jigsaw puzzle in the world." If it takes this long to decipher our genetic "book of life," we asked, "How do you think DNA originated?" We suggested that the Master Planner must have had some great intelligence to put it all together. He replied, "They had a lot more than human intelligence to do it."*

genes to those from bacteria where the function had been studied for DNA repair. So there is no doubt in my mind that we evolved by changing the genetic code over billions of years on this planet. That still begs the question of how life originated on earth.

SERVAAS: The Master Planner had some intelligence to put it all together?

VENTER: They had a lot more than human intelligence to do it.

SERVAAS: [laughing] That's a good answer.

MARCH: In cardiology, as well as other areas, one factor that has become very important in determining how patients do is not only their predilection for disease, but also how they compensate for having that disease. Some folks with diabetes, for example, experience severe problems in their peripheral vasculature, such as in their legs, while others don't. Some people who have serious blockages in their blood vessels supplying the heart, which could lead to a heart attack when completely blocked, will grow very nice collateral vessels. Others will not do that at all. Is this genetically determined? Why do some people do much better while others don't, and where are you in your genetic understanding of those second-level compensations?

VENTER: We can only begin to speculate on those levels right now. With blood pressure, for example, there are over 300 genes that we now know of involved in the regulation of blood pressure. In each of us, these genes play a slightly different role, so the genetic changes in those genes cause different interactions. So we change the spelling of the gene that leads to the change in the spelling of the protein. But, in fact, what is important is how that affects the interactions with another protein. If that spelling is slightly different in each of us, we are going to have subtle, different protein-protein interactions. Envision all six billion people on the planet, and we interact. Our interactions with each other are all subtly different in part because of our genetic code.

MARCH: Everybody's "palm pilot" is different.

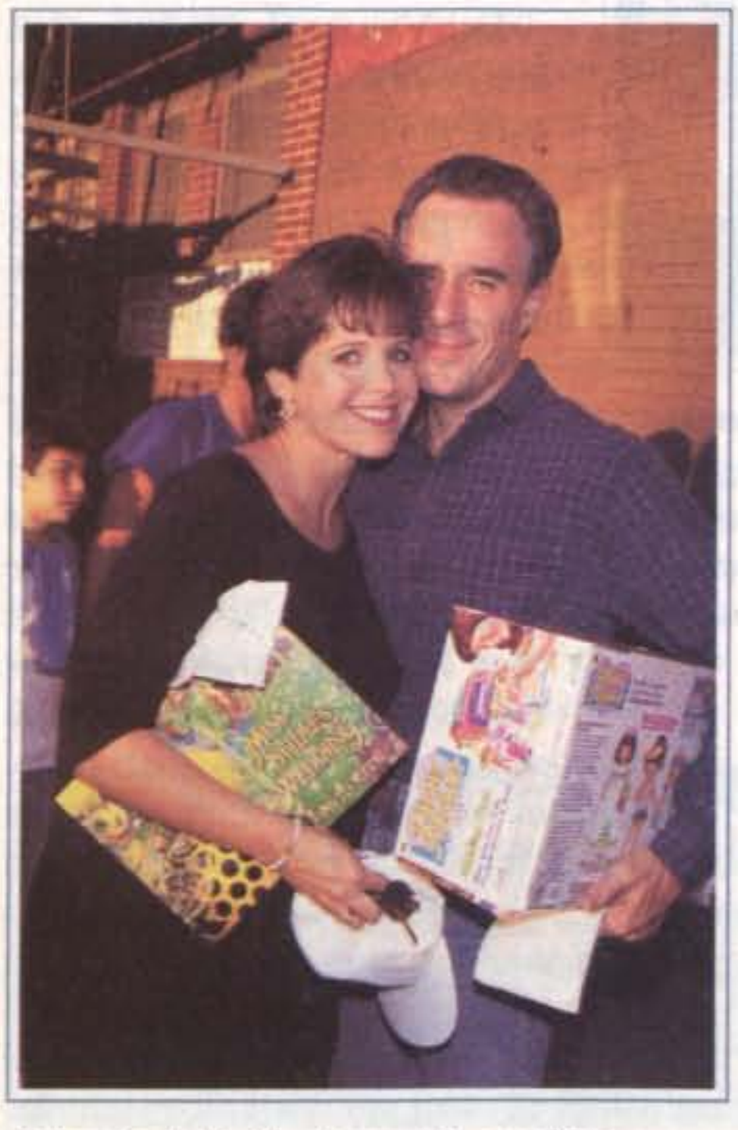
VENTER: That's right. That's very much the case at the protein level. It seems impossible to understand with the complexity of the biology right now, but that is where the supercomputing comes in. It is not just enough to have the genetic code. That is where proteomics comes in. We are setting up a facility here to sequence one million proteins a day to try to understand these protein-protein interactions. Then we can understand why the spelling of a protein is slightly different, and why the same individual has a different spelling in a particular gene, this protein that we know there are interactions with that lead to this different effect.

We are starting a study with a research group at La Jolla to look at hypertensive patients with well-characterized differences like you are talking about that have not been studied. We are going to see if we can find changes in genetic code linked to some of these subtle differences, because it is the subtle differences that really determine, in many cases, life or death outcomes.

That is a very key part of this effort. We are at the infancy of this science, but having the complete genetic code is going to drive this research forward very quickly. We are going to do more in the next two or three years than has probably happened in the past 30 years in understanding these kinds of interactions. That still means we don't know the answer to your question. What it does mean, however, is that there is hope to get the answer in the relatively near future.

MARCH: That is very impressive. With hypertension, there is a population of roughly 50 percent who are very salt-sensitive and shouldn't eat salt, while others are relatively insensitive; another percentage of hypertensives actually become worse when their so-

*In 1994, Dr. Bert Vogelstein, a leading cancer researcher at Johns Hopkins University Medical Institutions in Baltimore, and Dr. Venter collaborated in identifying three genes associated with colon cancer—"markers" now used to screen high-risk patients, potentially saving thousands of lives. "The completion of the human genome will help us considerably in our efforts to identify the genetic basis of colon and other cancers," Dr. Vogelstein told us.*



*Celera's individual genetic profiles, provided through its subscription database of genetic information, could help identify people at high risk for certain diseases. Because Katie Couric's husband died at 42 from colon cancer, early genetic screening of their daughters could provide potentially lifesaving information that would allow them to take steps to avoid the complications of the disease—one example of how knowing one's genetic code could empower the individual.*

*Leading researchers, such as cardiologist Dr. Keith March (above), eagerly await the completion of the human genome to better understand the genetic link in heart disease, high blood pressure, and other diseases. This knowledge will shed light not only on such fundamental questions as how humans grow and survive, but on why and how we develop disease, leading to better treatments and possible cures. By understanding genetic variations and protein function, physicians may soon tailor therapy to suit each individual's unique genetic profile.*

dium intake is restricted. With the La Jolla group, are you going to address that issue in a way that people will know earlier in the development of high blood pressure whether or not they should restrict salt?

VENTER: It is one of the potential benefits that will come out of the study, but those are great examples of how the subtle genetic variations in all of us, the three million letters of the genetic code in which we all differ from each other, determine our own sensitivities. The way medicine is practiced now, most drugs work on 30 to 60 percent of the population, yet we give the same drugs to the whole population. We give instructions to the whole population that you should, for example, take a baby aspirin a day because it will help you if you have a heart attack or a stroke. It turns out that is only true for one out of three people, and there is one letter change in the genetic code of one gene that seems to determine whether you are one of those three or not.

That is the type of individualized information that Celera hopes in the next few years to begin to provide to people so that we can understand our own genetic makeup and what that means in how we interact with the environment. Imagine how it is going to change drug delivery.

A major drug was just taken off the market for treating Type 2 diabetes, because one out of 10,000 people had adverse side effects, including about 60 people who died from it. That should be predictable at the level of the genetic code so that before anybody is given a prescription for that drug, we can rule out those people who would have an adverse side effect or potentially die from this treatment, versus having their diabetes effectively treated. That is individualized medicine going forward. The goal is to deliver drugs that will treat your disease and not kill you in the process.

MARCH: That is phenomenal. So as physicians, we need to look at the therapy we use for patients as part of their environment.

VENTER: That's right.

MARCH: Some people respond to a therapy the way we want them to, while other people will not. Responsiveness, lack of responsiveness, and possibly bad side effects from a specific drug then may just be due to individual genetic makeup. Is that correct?

VENTER: Certainly a major portion of it will be subtly changed by each of our environments. But the hope is that a lot of this will be absolutely predictable from either protein diagnostics or DNA diagnostic tests.

MARCH: This information will help us understand why some people respond to certain drugs while others don't?

VENTER: That's right.

MARCH: For those people who aren't so familiar with the term, could you tell us what is proteomics?

VENTER: It's a terrible term, derived from genomics. The genome is our

*see Discovery Can't Wait on page 68*

## *Discovery Can't Wait*

*continued from page 45*

complete collection of chromosomes, or genetic material. The term *proteomics* was developed to try to encompass the whole protein world. It is harder because it is not as easily defined. Our proteins constantly change or are dynamically regulated. Genes can have alternate splicing, so in different cells the same gene leads to subtly different protein structures. That is how we regulate physiology. The field of proteomics is trying to get the structure of all these different proteins to understand which ones are expressed in what cells and what regulates them and to try and determine their three-dimensional structures so that we really understand how the proteins interact with each other. That's where biology happens. When we look at each other, we don't see DNA and RNA. We see protein, the results of protein action and somatic action. All diseases are treated and occur at the protein level. The DNA and RNA determine how much protein is there and in what cells. Proteins are key. Getting the genetic code was the way to get to the protein world orders of magnitude faster than the way science was conventionally going. The key thing with getting the genetic code was to understand the protein world, not to understand DNA per se.

PERRY: Are you collaborating with universities or planning on future university studies in your current project?

VENTER: That was a great question. Our goal is to have our database of the human genetic code in every university and every research center to be catalytic and move things forward. People listening and reading the press may naively assume from some of the government efforts that just having the genetic code on the Internet will lead to immediate treatments for disease. That is a very false assumption. This is such a large amount of information. If we were to print the human genetic code right now, it would form a stack of papers a hundred feet higher than the Washington Monument. We have the largest database in the entire medical world. It is over 60 terabytes [*one thousand billion bytes of computer storage*] of data; it's going to be a petabyte [*one thousand terabytes*] of data within a year. The computer facility that we have is a key part of understanding genetic code. We are trying to change how science is done from studying one gene, one protein at a time. I spent the early part of my career studying biology, and that is the integration of all this information. You can't do it without massive computers.

That is what Celera is providing as a service to researchers around the world. We are trying to do collaborations and just directly provide the information. There are no strings attached to the data when people have subscriptions to the databases. People make discoveries. They are encouraged to publish and patent their discoveries and do whatever they can to move medicine forward quickly.


*"Do you have any other collateral other than 100 free hours from your Internet provider?"*

We do have some clear-cut collaborations. We have one with City of Hope, understanding some of the changes in the estrogen receptor gene with breast cancer and trying to understand the links of certain haplotypes to breast cancer. We are collaborating with a company called Gemini that has an identical-twin database to use identical-twin information matched with their disease to investigate how we can use that information to understand the correlation between genetic changes and diseases where they occur, because it is far more complicated than just looking at single genes. We are trying to provide this information across the board as a tool to everyone to advance every field.

MARCH: That will be a phenomenal advance for science and medicine. You mentioned that you would sequence at least a million proteins a day?

VENTER: That is the goal.

MARCH: Will that make you the group with the most thorough pool of information on proteomics in the world?

VENTER: It will clearly be the largest database. Again with that amount of data every single day, the computers, the storage, the way that we retrieve this data and ask intelligent questions about this becomes the biggest challenge. We are working, for example, with Oracle in trying to design new databases that can actually deal with this much data; the databases don't exist right now. Banks, for example, have a big data-storage problem, but they don't have to relate your bank account to my bank account.

With genomics, biology, and proteomics, we have to understand the interactions of all the genes. The interactions are all crossed-linked, so we have to be able to access all the information simultaneously. That is a tremendous challenge for biology and computing.

MARCH: When you have high blood pressure with 300 different genes involved and each one playing a small role in the extent and manifestation of high blood pressure, do you think you can account for 80 percent or 90 percent of the cases?

VENTER: We also have to understand the environmental factors that influence high blood pressure. It is not just the genes. You and I might have the same genetic changes, but if you consume ten times as much salt as I do, you might have a different response in your physiology than I do in mine. One of the important messages is that life is context-sensitive. Genetics is context-sensitive. You can't just look at the genetic codes. We have to look at other aspects of peoples' lives with the same genetic code to understand any two people. That is why the twin studies are important. Just consuming twice as much salt or fat may totally trigger certain genetic pathways to be altered that wouldn't happen in me if

I am consuming less of those. That is why it is a big equation. This new understanding won't happen overnight. But we have to start collecting data on the scale of millions of proteins a day and understanding this in a broad set of individuals.

MARCH: One key cardiovascular risk factor that is not usually thought of as a genetic risk factor is tobacco abuse. Do you think it is possible that we might be able to identify a genetic basis for some folks becoming very easily addicted to tobacco, while others are more resistant? Perhaps we could then use that information as an early intervention to focus anti-tobacco education in genetic situations similar to what we already know about alcohol.

VENTER: Yes. In fact, not just the addiction potential but the disease-causing potential. Some of us have naturally occurring proteins and changes in our genetic codes that make us more resistant to the harmful effects of smoking. Some people have argued that we don't want to do those tests, because it would tell some people that it might be safer for them to smoke than they would think otherwise if they don't know that they are just all at high risk. Smoking is not good for anybody, but it is certainly less good for some people than it is for others. All of these tests have some social consequences, aside from just trying to deal with preventive medicine and disease. You might find that you are extremely susceptible to the harmful effects of smoking; I might find that I am less susceptible to these and, therefore, I could wrongly interpret that it's OK for me to smoke, but one shouldn't. In fact, smoking will shorten both of our lives.

MARCH: Some education is very critical to understanding how to deal with this information.

VENTER: That's right. It's critical.

PERRY: How does this education differ from what has been going on presently in preventive medicine? For some time, we have heard the messages: Don't smoke, eat less fat, and avoid alcohol. Will the genetic information offer more proof to the overall concept of prevention?

VENTER: In scientific literature, we have heard at least 20 fluctuations in whether drinking coffee is harmful or not. The nature of the equation was pretty complicated. We are trying to understand, at a fairly simplistic level, complex biology—looking at the outcome. If we had 100 identical twins to look at these effects in a well-controlled study, some of these answers might not be so varying. But it is hard to pin things down exactly in terms of an absolutely causal relationship, because it varies for each of us. Individual genetic variation will be a key. For some people, a high-fat diet is actually lethal at a very early stage, just as for some people smoking at all can cause lung cancer and very premature death much more rapidly than it can in others. There are tremendous differences in all of us. That is where we are trying to look across a population. That is the trouble with clinical trials and new drugs. The new drug can possibly cure a disease in a fantastic fashion, but it only works on 50 percent of the people you give it to.

It is very difficult when you are doing drug studies, then, to prove statistically that this treatment will cure this disease because, in fact, it may do it 100 percent of the time in the right people in 50 percent of the population; in the other half, it may actually cause toxic effects. I think the same information will be used to improve therapies. That is why I say that medicine has to become individualized, based on the individual's genetic code.

MARCH: The ideal environment then is matched to the person. Just like the ideal spouse is matched to the person.

VENTER: I think that is part of the environment. [Laughter]

PERRY: Does the slogan "Speed matters. Discovery can't wait." sum up Dr. Craig Venter?

VENTER: Yes, it is based on my personal philosophies.

CONNEALLY: Many scientists have major problems in getting bioinformatics people, and that is key in this area. You said earlier that you have already sequenced the human genome and that now you are trying to integrate the sequences and get all the interactions into your enormous computers. To do that, you need manpower. Are you having much difficulty in finding good people?

VENTER: We have been very lucky in this area. In fact, I think we have probably 40 percent of the top talent in the world here at Celera in terms of algorithm development. People like computer scientists Eugene Myers, Ph.D., and Granger Sutton helped define these new algorithms that allow the genome to be seen. Just 18 months ago, the top scientists in this field were saying what we were doing was impossible and that it wouldn't work. It worked because of the application to the new algorithms and the new mathematics and the supercomputing to this problem that was not obvious to many biologists working at a much smaller scale. But you are right; bioinformatics is the absolute key, and that is the one area that is hard to build fast enough. Building complex software to deal with this


*"He's mostly adjusted to retirement OK, except that he keeps sending me memos."*

mountain of information takes time. We have over 100 software engineers here, and we are buying other companies and trying to get more quickly, but it is the rate-limiting step. It's such a new discipline that there are not enough people trained on the planet. We could hire them all here.

CONNEALLY: Yes, and it still wouldn't be enough.

VENTER: And it wouldn't be enough. We are trying to help train people faster, building the tools and making the data interpretable. That is one of the keys we are offering, and that is why some people assume that this kind of business model is to keep the data a secret. In fact, it is just the opposite. We will go out of business if researchers like you and your colleagues don't use the information. But we are trying to create the tools and the computer infrastructure so that you can actually do new experiments that each of us can only dream about.

CONNEALLY: I was amazed at how rapidly you sequenced the first insect, Drosophila—the fruit fly—even though it is much smaller than the human genome. It happened much more rapidly than smaller organisms, so we know the pace is extremely rapid. We were very impressed with your facilities. We had a wonderful tour today. But you're serious that you feel you can sequence a million proteins a day, not a million amino acids?

VENTER: Yes, exactly, a million proteins a day. And that would not have been possible without having the complete genetic code. In aspect sequencing, you break the proteins down into smaller fragments. Right now when people do these studies, most of those fragments don't match anything in the genetic code because we only had a partial genetic code. Now, with the complete genetic code, it is a very big calculation to take all the protein fragments, based on their masses and the mass spectrometer, and link them back to what it is. But again, it is the application of mathematics and high-end computing to solve biology problems.

There are some exciting new instruments that are coming out that make this possible as well. All of the success has been due to not just new strategies but also to exciting new technology developed by our sister company, PE Biosystems, that has made the DNA sequencers, the mass-specs. And it has been due to the new mathematical algorithms and the application of the high-end computing.

CONNEALLY: This takes an enormous scale of computing power. In fact, I don't know what a petabyte is. It is so enormous going from megabytes to gigabytes to that level of activity.

VENTER: We have the largest computer devoted to biology in the world.

CONNEALLY: Are you or any of your subsidiaries going to be involved in actual genetic testing—taking a swab test of the cheek, for example?

VENTER: The plan is to get to that point. We want to make sure that the information is incredibly accurate and that we can give not only meaningful results in terms of reading the genetic code, but also to help you and your physician interpret the genetic information accurately as to what it really means, so that people don't take drastic steps with the information. We are early in that process, so we don't want to do it too early until we really understand the outcome. Based on your pioneering work, look at what has happened in the last decade from your research for the Huntington's disease gene. We are just now starting to get breakthroughs in understanding what that gene may actually do and how it may work. These studies really take time, even once we have a little breakthrough like your work has led to.

CONNEALLY: In Parkinson's disease, we realize that there is a major environmental cause, but there is also a genetic background that probably protects against environmental insults. That is where the genome sequencing is going to help us to narrow down in these areas.

It was intriguing what you said about schizophrenia, that identical twins have only a 50 percent chance of having the genetic predisposition for schizophrenia. If one of the twins has schizophrenia and the other doesn't, the risk to the children of both of the twins is identical. The schizophrenic twin does not pass on a higher risk for his or her children, so obviously environment is playing a major role.

VENTER: There is clearly a genetic and environmental component. We have so many variables that it makes it a real challenge to sort out the vast array of information.