Health Matrix: The Journal of Law-Medicine

Volume 12 | Issue 2

2002

The Human Genome Project: A Public Good

Kathy Hudson

THE HUMAN GENOME PROJECT: A PUBLIC GOOD

Kathy Hudson[†]

THIS DISCUSSION WILL COVER the genome project and what we do with technology transfer at the National Human Genome Research Institute. The last couple of years have been really remarkable both scientifically and sociologically.

It has been my observation that the Genome Project has almost become a household word. My evidence for this is threefold. First of all, it is the appearance of cartoons about it in newspapers and magazines that we pick up every day.

My second line of evidence is my mother, who lost track of what I do when I went to graduate school in the early 80s—though she was very proud of my accomplishments, she did not understand them. On February 13th of this year, my mother called me to congratulate me and my colleagues on having rolled out the genome sequence, and she proceeded to engage me in a very sophisticated discussion about genomics. I was so puzzled. What happened?

It turns out that my mother had watched the press conference where we unveiled the sequence of the human genome and its analysis not once, not twice, but three times on C-Span. I think she learned quite a bit.

My third line of evidence comes from an experience I had getting together for a drink with an old colleague of mine. We were talking about the Genome Project and how things have evolved; and the bartender and another guest jumped into our conversation and they knew everything. It was remarkable.

So, in fact, I think we have become a household word. While I think people's understanding is somewhat superficial, I am hoping that that will provide the entree for a deeper understanding as we move forward and translate the Genome Project into beneficial drugs and interventions. So the reason that we

---

[†] Ph.D., Assistant Director, National Human Genome Research Institute, National Institutes of Health.

are doing the Human Genome Project is that all disease has some genetic component. That genetic component can range from a very substantial one, as in so-called single gene disorders, such as cystic fibrosis, to a very minor contribution, as in AIDS, where there are genes that can render an individual resistant to infection from HIV.

Most diseases fall in the intermediate category where there are a large number of genes and a large number of complex environmental factors interacting with one another to increase disease risk. We need to understand that as a fundamental level in order to allow the discovery of new drugs and interventions.

I will say in a sentence or two: What is the genome? The genome is the sum of all the hereditary material within our cells. The DNA is packaged into twenty-three packages called chromosomes, imaginatively named "chromosome 1," "chromosome 2," "chromosome 3," and so forth.

All of the chromosomes are made up of DNA. If you took the DNA from a single cell and stretched and laid it end to end, it would stretch about six feet long. The DNA itself is made up of four simple subunits. The DNA alphabet is four letters, A, C, T and G. There are about 3.1 billion letters of DNA in the human genome. These letters contain instructions or genes that control all of the fundamental biological processes of life. There are the genes that make a liver cell a liver cell, a brain cell a brain cell, and a single cell fertilized egg become a human child. It is pretty remarkable stuff.

In the spelling of the DNA, the precise order of the letters is also incredibly important. Like a computer code where the order of the zeroes and ones is critical to the functioning of the program, so is the order of the A's, C's, T's, then G's. So if you have a G where there should be an A, that may create a protein that misbehaves or does not function at all, and causes disease or increases risk of disease.

So, the goal of the Human Genome Project is to map and sequence the human genome and to identify and understand the genes within it. Our best-known undertaking is having recently deciphered the 3.1 billion base pairs in the human genetic code. It was a pretty remarkable effort. In June of 1998, we had very little of the human genome sequence. This was the international effort involving sequencing centers around the world in what was by far the largest international basic science collaboration ever.

The remarkable thing sociologically—and I hope somebody studies this at some point—was taking a group of people who had by their very nature been competitors ("I'm not going to reveal my secrets because it will allow the other guy to compete more effectively for grants") and to see a transition from competition, to them coming together and working in a very organized and systematic effort to get the working draft of the human genome completed.

In the steepest part of this slope here the genome centers were producing 1,000 letters of DNA sequence every second of every day, seven days a week.

We first started this ramp up in March of 1999 and in the following fifteen months a lot was accomplished. In June 2000, with that working draft in hand, there was a celebration. It was sort of an odometer moment. We did not know much about what the genome did, but we had all the letters or most of the letters. So, there was a celebration, and a lot of fanfare, and headlines across the country. And there was a celebration jointly with Celera Genomics, who also independently developed their working draft of the human genome.

So, in the intervening time, from June until February, really the interesting part started. Now we had all of the A's, T's, C's and G's, and we wanted to figure out what the words were, what the sentences were, and to make meaning of them. So for me as a molecular biologist and a geneticist, this is really the most exciting time.

Using the best brains in biology and in computational biology from around the world, six months were spent really scouring that sequence to figure out what we could learn about age-old mysteries in biology; and, in fact, we discovered new mysteries about biology.

Those results were published in *Nature* magazine in February. We were particularly pleased with the cover, which shows a double helix section of DNA, but it is actually made up of a mosaic of people's faces, people from around the world. That seemed very fitting to us because the DNA sequence—the human genome—is about all of us. It is about the human species. So we were very pleased with the cover.

But scientists do have a sense of humor, so as the analysis was going forward and the paper was in review and all sorts of stuff was happening, there were more e-mails about whose faces

we wanted to insert on the cover than about the analysis of the genome, at least for a couple of days. So we did include the faces of Watson and Crick, who discovered the double helix structure of DNA.

What did we find when we looked at the genome? Well, we found a lot, and I will just mention a couple of things. The first is that humans have fewer genes than were expected. The number that I have been using and we have been using at the Institute since 1995 was that humans have between 80,000 and 100,000 genes. It turns out that we have far fewer than that, something in the range of 30,000 to 35,000 genes. That number was independently arrived at by Celera Genomics, who feels very confident that while that number may change by a little bit up or down, it is not going to change dramatically. There are those in the private sector who have invested quite heavily in actually selling access, who are saying there are more than 35,000 genes, who are disputing this claim, but we feel quite confident that it is solid.

So, what can we now do with all of these genes in hand? One of the really amazing things that we can now do is look at whether or not genes are turned on or turned off in any particular cell type or at any stage in pathology, in the development of a tumor, in the development of an organ system, in people affected by diabetes, and those who are unaffected by diabetes. We can look at the expression of those genes in a single experiment.

It used to take us ten years to find a gene and to look at its expression. Now we can do it in a single experiment. In fact, the roadblock is informatics—developing the computational power to be able to analyze this data. So, on a little chip or microarray you can look at the expression of 12,000 genes in a single experiment, where a red spot indicates that that gene at that position is on and a green spot indicates that that particular gene at that position is off. It is an incredibly powerful technology; one that was actually developed by a company that was founded with a small business grant from NIH.

So, we are not going to see drugs pouring into the marketplace as a direct consequence, within a year, because of the Human Genome Project. There is a long development process, and the NIH, and our institute in particular, are only involved in the very beginnings of that pathway.

So, being able to get to the end of this flow chart to develop the preventive interventions, the new drug therapies, pharmacogenomic approaches in gene therapies, will require the private sector to get involved, and it will take an indeterminate amount of time for any given condition that you want to approach.

I want to give a quick example of the end points here. Pharmacogenomics is the relatively new field of being able to ascertain genetic profiles that dictate whether or not I will respond to a particular drug; whether I will respond really well to a particular drug, or whether or not, in fact, I will have an adverse consequence to a particular pharmaceutical. That is going to be very important to the future of the pharmaceutical industry as we tailor drugs to individuals' genetic profiles.

The other end point here is the development of new drugs. In sequencing the human genome, we identified a huge number of new drug targets, previously unknown drug targets. So now, as has been mentioned many times, not only is the biotechnology industry heavily investing in this area, but also every major pharmaceutical company has a major genomics component.

There are actually things starting to dribble out of the end of this pipeline now, which is tremendously gratifying. One example is a drug called Gleevec that is being produced by Novartis to treat chronic myelogenous leukemia. Chronic myelogenous leukemia is the result of a reciprocal translocation, where a bit of one chromosome switches places with a bit of another chromosome. When those bits of chromosomes change places, parts of two genes at that junction point are fused. That fusion protein misbehaves and causes the disease.

Novartis looked very carefully at that fusion protein and developed a small molecule drug that would block its action. Gleevec is in clinical trials now and is showing enormous positive effect, not only on leukemia, but also it is now being used on solid tumors to great effect. In fact, a colleague of mine has recently started taking Gleevec, and he is now back at work as a result of treating a solid tumor with this drug.

There are a number of other such drugs that are entering various stages of clinical trials, and gradually their number will increase. I think we do have to counteract the hype but I think there is great promise here and we should not lose sight of that.

So, the Genome Project goals are to provide the raw materials not only for biotechnology but also for the pharmaceutical

industry. We have developed the genome sequence, the raw fundamental information to be mined and used by industry, and we have developed a catalog of sequence variants, places where your genome varies from my genome, which are going to give us information about disease risk and responsiveness to drugs.

We also develop bioinformatics tools and we study the genomes of other organisms, because they can tell us a lot about our own. But, in fact, it is going to be the private sector that is going to use this wealth of basic knowledge to develop new medicines and bring them to market.

There has been discussion about how far into applied research NIH and government funding reaches, and how far down into basic research private sector funding reaches. NIH applied research funding is relatively modest. The NIH spends a relatively modest fraction of its budget on clinical research. At the same time, genomics companies and pharmaceutical companies are spending an increasing amount on basic, fundamental biological research. This intersection of research missions offers all sorts of interesting opportunities for collaboration and coordination. But the private sector, the government, and academics trying to work together must be mindful of the different cultures and the different constraints on how they operate.

This slide shows the very familiar technology transfer cycle. The National Human Genome Research Institute does have an intramural program, a research program much like a fine human genetics department. While it is one of the smallest of the NIH Institutes, it has the third largest, most robust technology transfer operation at the NIH, third only to the National Cancer Institute and the National Institute of Allergy and Infectious Diseases. We have an active cooperative research and development agreement and technology transfer program.

Technology transfer has had an impact and there have been some trends in the last few years. There certainly has been a dramatic increase in academic, university patenting and there has been an increase in industrial funding of basic research. Interestingly, the number of academic scientific publications that cite patents has gone up and the co-authorship between industrial and nonindustrial partners has also gone up. A by-product of all of this is that the number of intellectual property lawyers per dollar spent has also gone up.

There is an area of research in genomics which I think sort of falls out of the traditional technology transfer cycle. Tradi-

tionally in an academic environment scientists come up with a great idea, develop that idea, file for a patent, license that invention to the private sector, and do a pass-off. What we are increasingly finding in large-scale genomics is that there is a motivation for private sector and academic scientists to generate the basic resources together.

NIH could generate the sequence of the human genome or we could generate the sequence of cDNAs, copies of the genes that are actually expressed, and we could encourage our investigators to keep hold of that information and mine that information until they reach the utility standards established by the PTO before putting that out in the public domain.

We do not do that. In fact, as a condition of our grant awards, we say that this data must be made immediately publicly accessible. Harold Varmus, when he was director of the NIH, initiated a program across the NIH to sequence so-called cDNAs, copies of the expressed genes. Dr. Varmus sought and obtained a declaration of exceptional circumstances that said NIH could award grants and ask that the grantee institutions not file patent applications on their inventions and put that data into the public domain as an exception to Bayh-Dole.

Large public-private collaborations to generate basic fundamental raw sequence and genomic information are springing up at an incredible rate. The first was the SNPs Consortium. An SNP is a place where my genome varies from your genome, and we want to develop a catalog of these variants. Nearly two million of these places in the human genome that vary in sequence among individuals have already been identified. All of that data has been placed into the public domain.

The second public-private genomics collaboration is the Mouse Sequencing Consortium. There are half a dozen or so academic institutions involved in each of these consortia. Recently we created the Rat Sequencing Network, again a public-private-academic-government consortium to sequence the genome of the rat.

Then the last, which just recently filed as a 501(c)(3) organization, is the International Genomics Consortium, which is an effort to obtain gene expression data on thousands and thousands of cancer samples. People have raised questions about the privacy of clinical medical records and information about tissue samples. This consortium is attempting to balance those con-

cerns. And, again, there will be no intellectual property on the International Genome Consortium results.

I think that one of the critical things for both the academic institutions and the corporate entities involved in genomics and genetics is to be good citizens and to ensure that the public has confidence and will be willing to use these products and participate in this research. One of the things standing in the way of that confidence is fear of misuse of genetic information. A cartoon from the New Yorker in 1996 was very prescient. A guy is interviewing for a job and the employer says, "Very nice resume. Please leave a sample of your DNA with my secretary on your way out the door."

There was a story on "60 Minutes II" about the Burlington Northern, Santa Fe Railroad case, which is really a call to action for all of us to make sure that we have the right policies in place for handling genetic information in the workplace. Are we asking for genetic information? Are we using genetic information in the workplace?

The Brotherhood of the Maintenance of the Way Employees filed suit against Burlington Northern after employees who had filed Workers Compensation claims based on carpal tunnel syndrome were asked to come for a physical evaluation and provide blood samples. And without their knowledge or consent, those blood samples were subjected to genetic testing.

The clever wife of one of the workers learned of this, told the union and the Equal Employment Opportunity Commission, which then filed lawsuits. Burlington Northern said, "Oh, gee, we didn't know we were doing this. This is a big mistake." It is actually quite ironic. There is only a single publication in the scientific literature indicating any correlation between the tests that they were performing and the condition that they were seeking to detect.

Burlington Northern settled with the union last week. In their settlement agreement they promised that they would never do genetic testing again. They also said they would destroy all of the samples, which has raised a big brouhaha because some of the employees who were tested said: "Wait a minute; that is my evidence. If I want to file a case, my evidence is now gone." But Burlington Northern said that they would destroy the samples, destroy the records, and that they would never do genetic testing again unless it was agreed upon with the union.

In an incredibly clever public relations move, Burlington Northern said that as a part of the agreement, they now believe that genetic information should not be used in the workplace as a condition of hiring, firing, assignments, or any other employee-related decisionmaking, and that they will begin within thirty days of signing the agreement to lobby the Congress and the White House to enact legislation to prevent the misuse of genetic information in the workplace.