To Reveal the Genomes

The Human Genome Project, originating 20 years ago, and its many sequelae are certainly one of the greatest scientific successes of recent times. It is often said that success has a thousand fathers, and varied groups have claimed the project's origin. One of these was the workshop held at the University of California–Santa Cruz (UCSC) in May 1985 with the topic "Can We Sequence the Human Genome?" This workshop followed a letter I wrote in November 1984 as Chancellor of UCSC to University of California President David Gardner, in which I proposed a use for a considerable sum of money that the university had but seemed likely to lose.

> Dear David:
>
> Let me expand a bit on our brief discussion at the regents meeting on Friday. If the "_____" firmly intend to withdraw from the TMT project, then I have another project that we might propose to them. It is an opportunity to play a major role in a historically unique event—the sequencing of the human genome.
>
> A genome is the complete set of DNA instructions for the making of a species. The human genome is the complete set of instructions for a human being. We know that the haploid human genome is composed of some three billion nucleotide pairs (3×10^9).
>
> Clearly the human genome will be sequenced. It will be done, once and for all time, providing a permanent and priceless addition to our knowledge.
>
> In addition to satisfying our scientific curiosity, this knowledge will provide deep insight into other questions of interest. It will have major medical implications; we know that literally thousands of human ailments have genetic bases, in whole or in part.
>
> This knowledge will also have highly significant evolutionary implications. The biological differences between *Homo sapiens* and the chimpanzee are certainly due to the changes and rearrangements in the genomes of each as they have diverged from that of our common ancestor. To understand these changes will surely illuminate the ancient human quest to know what we are and where we came from.

This effort to attract funding for this purpose was unsuccessful, so I sought other sources. To do so, I believed I needed a stronger and broader validation that this project was both feasible and worthwhile and, with the assistance of three members of the Biology faculty—Harry Noller, Bob Ludwig, and Bob Edgar—convened the workshop. It is useful to recall the status of DNA sequencing at that time. The first genome ever sequenced, in 1977 by Fred Sanger, was the small, single-stranded DNA bacteriophage ØX174, ~5,300 nt. Since then, other viruses had been sequenced, including T7 bacteriophage at ~40,000 nt and lambda at ~49,000 nt. Work was under way on varicella zoster, the ~125,000-nt chickenpox virus, and conversations had begun among daring folk about sequencing a bacterium, *Escherichia coli*.

We knew we would ultimately want the complete sequences of many organisms (e.g., *Drosophila, Caenorhabditis,* and mouse), but *H. sapiens* would inevitably be the ultimate grail of such projects, both for the light it would shed on us, our origins, and our functions and for its evident utility in medicine. And surely, once we had the technology to sequence humans, the other organisms could be sequenced with relative ease. I had no doubt that this knowledge, if achievable, would unlock the mysteries of cells and organisms and of how they came to be what they are.

As a former virologist and the Chancellor of UCSC, I had come to this concept as the confluence of several ideas. As Chancellor, I had been involved in the conception of several large-scale scientific enterprises—involving telescopes (the TMT project) and accelerators—which were "Big Science," scientific projects requiring, in some instances, billions of dollars and the joint efforts of many scientists and engineers. It was thus evident to me that physicists and astronomers were not hesitant to ask for large sums of money to support programs they believed to be essential to advance their science. Biology was still very much a cottage industry, which was fine, but I wondered if we were missing some possibilities of major advances because we did not think on a large enough scale. Most laboratories were working on one gene or one enzyme at a time, yet it had become evident that many human disorders had, as any geneticist would expect, complex genetic components, and, plausibly, so must many human traits.

We needed much greater knowledge of the human genome, and, for that, we needed the technology to obtain and handle large amounts of data on genomes, on the myriad of interactions within a cell, and on the interactions of medicines within the body. We did not even have a grasp of the number of genes in these genomes or of the variety and complexity of proteins within their cells.

As Chancellor of UCSC, I had a major concern for the future of this young, growing campus. I wanted it to be-

This article is based on a presentation given as part of a special historical plenary session entitled "Origins of the Human Genome Project" that took place at the 55th Annual Meeting of the American Society of Human Genetics in Salt Lake City, October 2005.

Address for correspondence and reprints: Dr. Robert L. Sinsheimer, Department of Molecular, Cellular, and Developmental Biology, University of California–Santa Barbara, Santa Barbara, CA 93106-9610. E-mail: Sinsheim@lifesci.ucsb.edu

Am. J. Hum. Genet. 2006;79:194–196. 0002-9297/2006/7902-0003$15.00

come the peer of other University of California campuses such as Berkeley and Los Angeles and to become similarly recognized in the world of biology. Now, through a complex of circumstances, there seemed to be an opportunity to obtain funds to launch at least a modest human genome project if we could obtain validation that it was a feasible project.

The purpose of the workshop was to consider the feasibility of sequencing the human genome and to outline the likely strategy and essential stages for such a project: development of more-powerful methods for genetic mapping and cloning, provision of automated means of sequencing and of robotic agents, and the invention of improved means of data storage and access. We set out to invite key scientists in each of these areas. To our pleasure, most invitees expressed interest, although some indicated great skepticism. Present at this workshop were most of the leading people who could provide an intelligent answer to that question, including (16 in all) Bart Barrell (Medical Research Council [MRC]), David Botstein (MIT), George Church (University of California–San Francisco), Ron Davis (Stanford), Wally Gilbert (Harvard), Lee Hood (Caltech), Hans Lehrach (European Molecular Biology Laboratory), John Sulston (MRC), and Michael Waterman (University of Southern California).

We discussed in some detail the various problems to be encountered and how they might be resolved with the technology of the time and some likely advances. We contemplated how the chromosomes might be separated in adequate amounts. We considered development of a physical map of cloned fragments with contiguity identified by overlap and by denaturation mapping. Sequencing of the fragments would clearly require significant technical improvement over the current techniques. Human genetic polymorphism was recognized as a potential difficulty that suggested use of one or a few emblematic DNAs. We recognized the need for advanced computer programs to cope with storage and analysis of such large masses of data. Also, we projected that the enterprise would require 15 years at a cost of $1 per base, or $3 billion, which turned out to be a remarkably good estimate.

As the workshop progressed, as the various scientists described the status of their work, and as we analyzed the problems to be solved and the likelihood of progress toward their resolution, the mood of participants swung from extreme skepticism to confidence in the probable feasibility of such a program. Lee Hood was particularly confident in the development of fast and automated technologies. Several participants had significant doubts whether such a program *should* be initiated—whether it was a wise approach and whether the requisite funds could be justified—but feasibility was no longer an issue.

Sources of hesitation ranged from concerns over the introduction of Big Science into biology to arguments that most human DNA is "junk" (i.e., noncoding), so why sequence it? I disagreed with both objections. To be sure, Big Science, per se, is not a virtue. The common characteristic of Big Science projects in other fields has been that they provide a facility essential to advance the field further. Such a facility did not seem to be needed in biology. But, as I have indicated, what biology did need was a massive information base, a detailed knowledge of the genetic structure of at least several key organisms, including, for obvious reasons, humans. Also, we needed to establish a group or collective science more proportionate to the task—to understandings formed by many minds, linked electronic-memory banks and computing facilities, and research tools and centers for analysis. Knowledge of the genomes would delineate the scope of the problem and, at the same time, define the scale of the enterprise needed to elaborate the mechanisms of life and then to develop interventions for human purposes. I saw little merit in the second argument about junk DNA. Coding regions, control regions, putative nonspecific regions, and, as we now know, RNA-coding regions of DNA are all intermixed. It is only by examining sequence, together with functional studies, that one can actually decide what is or is not junk.

Even with the validation from the workshop, I was not successful in my efforts to obtain private funding for this project. Then, a year later, the Atomic Energy Commission and, subsequently, the National Institutes of Health became interested in this project, and the Human Genome Project was launched. As we know, the Human Genome Project has succeeded beyond our hopes, and now genome sequences are coming forth in abundance—bacteria, fungi, *Drosophila, Caenorhabditis, Fugu,* mice, rats, plants, and chimpanzees. From these efforts, we learn more and more about the paths of evolution and the history of our species on this planet. Of course, as might have been expected, the availability of these sequences has opened the doors to whole new avenues of research, such as the intricacies of gene control, which seem to grow ever more complex, and many more "omes"—the transcriptome, the proteome, the interactome, the metabolome, and the cancer genome—all interesting and valuable, which lead through informatics to the concept of the cell as an integrated array of information processors. Furthermore, the technology is useful for other questions about the diversity of bacteria in our gut, in the soil, or in the sea; the search for the origins of varied disorders; and the complexity of genome expression during development in varied cells.

Big Science projects, such as the large sequencing projects and the development of varied microarrays, seem to coexist well with the traditional smaller-scale professor-and-student projects. I think it is fair to say that the Human Genome Project has launched the field of biology into a new and greater era.

There is one other, more personal impetus that I would add. For this, I need to take you back many years. In 1939, I read the book "You and Heredity," by Amram Scheinfeld. By today's standards, it was rather naive. The principal references were to Mendel and Morgan. Beadle and Tatum,

Avery, McLeod and McCarty, and Watson and Crick were in the future. The genetic code, much less its near universality, was yet unknown. Scheinfeld could only mention such characteristics as eye color and hair color as genetically determined, but he speculated about the inheritance of behavioral and intellectual properties. This was my first inkling that our personal characteristics and individual natures were in part determined by our individual inheritances, through these mysterious factors called "genes."

How marvelous it would be, I thought, if we could have a complete roster of these genes in which must lie the origins of our distinctive human qualities—the gifts of language and poetry, of mathematics and logic, of curiosity and wonder—though I had no reason to believe that it would happen in my lifetime. To unravel and spell out in detail the human genome and to learn how these genes gradually play the intricate score that results in each one of us has been the impossible dream for me for >60 years. To see it actually happen is simply a great joy.

Robert L. Sinsheimer
Department of Molecular, Cellular,
and Developmental Biology
University of California–Santa Barbara
Santa Barbara