

# A Celebration of the Genome, Part IV



In this last of this series of genome anniversary vignettes, you will continue to find provocative thoughts—ranging from the impacts of this “big science” project on the rest of the research community to the commonalities between art and science. Exciting as all of the vignettes have been, our space was finite, and we know that there are many more opinions waiting to be expressed. We challenge you to reflect on the implications of this landmark event and look forward to your letters and e-letters!

—Barbara R. Jasny and Laura M. Zahn

## The Accelerator

**Eric S. Lander**

*President, Broad Institute of Harvard and MIT, Cambridge, MA, USA, and Cochair, President's Council of Advisors on Science and Technology*  
When the Human Genome Project (HGP) was proposed some 25 years ago, the notion was so foreign to biology that commentators had to resort to metaphors from physics. The HGP was biology's Manhattan Project, biology's Moon Shot, biology's Superconducting Supercollider particle accelerator.

To some, the project seemed like mindless drudgework aimed at a dubious goal. (Sydney Brenner waggishly suggested that the HGP be conducted in penal institutions, by “inmates sentenced to 20 megabases—with time off for accuracy.”) In reality, of course, it was the human genetics research of that early era that was marked by tedium (10 years to clone the gene for Huntington's disease!). What smart young student would want to be an ant in an army scaling a linkage peak?

Ultimately, the HGP yielded discoveries as remarkable as any atom-smasher or deep-space telescope. It revealed that the spectrum of protein-coding genes is far smaller than imagined, that physiology depends on a vast universe of regulatory controls and noncoding RNAs, that diseases arise from many unsuspected genes and pathways, that so-called junk DNA may be the mother of much invention.

In the end, though, the HGP might indeed best be viewed as a “high-energy accelerator”—not of particles, but of scientific work and scientific imagination. Individual investigators, the drivers of biomedical progress, can today carry out projects that once required legions: They can readily assay thousands of genes, millions of genetic markers, billions of nucleotides; they can interpret their findings in the context of public data sets representing tens of thousands of experiments worldwide and billions of years of evolutionary information. These capabilities have liberated them to think creatively and boldly about important biomedical challenges. Once seen as “big science,” the HGP has proved to be the most powerful enabler of “small science.”

Robert F. Kennedy famously said, “Some men see things as they are and say ‘Why?’ I dream things

that never were and say ‘Why not?’” Increasingly, young biomedical scientists are bringing this same attitude to their work. They are impatient with technological limits that stand in the way of knowledge. They transcend disciplinary boundaries, fusing experimental, computational, and clinical science into a new biology. They roll up their sleeves to create vast data sets, comprehensive reagent collections, and powerful new methods—and they share them freely. They are not afraid to work in teams, if by working together they can change the world.

In the past week, I have attended three scientific meetings, where I heard young scientists brimming with vision about predicting all the ways in which tumors can become resistant to a therapy, unraveling the molecular basis of psychiatric diseases, characterizing the entire human immune response to stimuli, mapping the complete genomic landscape of all transcription factors through development, creating a comprehensive catalog of all cellular circuitry, and devising general methods to speed the development of new therapeutics.

If I had been asked to pick out a present to celebrate the 10th anniversary of the HGP, I could not have chosen better.

## Making Sense of the Data

**Peter Donnelly**

*Director of the Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK*

There is an old aphorism that you wait forever for a London bus, and then two come along together. After the huge international efforts to sequence the human genome, the past 10 years have seen an explosion of data documenting human genetic variation, and we are at the tipping point for what will be an avalanche of high-quality human genomes. Within a few years, tens of thousands of human genomes will have been sequenced.

A major strand in research since the human genome sequence has been the systematic attempt to identify, annotate, and understand the functional elements in the genome. For much of this time, the collection and analysis of human variation data might have seemed to be running in parallel to the functional work, with no direct connection. But that is illusory: Data on human genetic diversity have proved to be a valuable tool in the quest to understand human biology. One example is recombination. We have learned a great deal recently about this fundamental biological process through the study of humans and, in particular, of human variation. Another is population genetics—as others have noted, human variation data offer an unparalleled opportunity to understand the forces shaping patterns of genetic diversity and to identify genomic elements under selection. But perhaps most significantly, the identification of DNA sequence variation associated with phenotypes of interest, particularly with human disease, provides the starting point



on another route to understanding biological function. The growing number of examples where the path from sequence variant to function has been elucidated offers an encouraging pointer to coming progress.

Although the collection of whole-genome data from large numbers of individuals in disease studies is set to become routine, making sense of that data is not yet straightforward. There are formidable analytical challenges ahead, even in the research context. Moving this kind of information into the clinic takes the challenge to another level. Like others, I am an enthusiast for personalized medicine, but one of the biggest obstacles to the use of individual genomic information in health care will be the need for robust analytical tools for its interpretation.

## Fruits of Genome Sequences for Biology

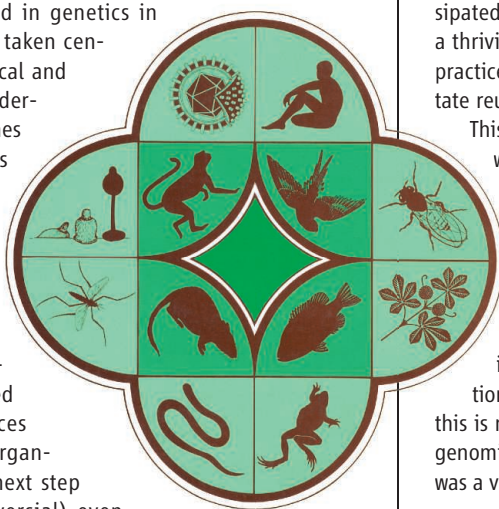
David Botstein

Director, Lewis-Sigler Institute for Integrative Genomics,  
Princeton University, Princeton, NJ, USA

When I first became interested in genetics in the early 1960s, DNA had just taken center stage. We studied its chemical and physical properties, and we understood that inferences about genes and genomes (yes, this word was used then) were actually about information encoded in DNA sequences, which we could not read or interpret. By the 1970s, we had learned how to use recombinant DNA technology to manipulate DNA in bacteria and viruses, which allowed us to recover (clone) sequences encoding proteins from any organism (including humans). The next step became obvious (and controversial) even before it had been reduced to reality: expressing these coding sequences in easily cultivated cells and producing pure recombinant proteins in quantity at reasonable cost. This enabled production of previously rare protein therapeutics and allowed biologists, biochemists, and structural biologists to study pure proteins.

One would have thought, after these developments, that when the possibility of sequencing entire genomes was first raised, it would be regarded as an obvious next step with great promise for science and medicine. Instead, it was met with much skepticism; in the beginning, I was among the skeptics. Unlike the controversy over recombinant DNA, which revolved around issues of safety, opposition to sequencing the human genome was driven by concerns about the extreme cost (estimated then at \$3 billion) and effort required. The opposition (including me) felt that diversion of these kinds of resources to “big science” might so distort the nature of our scientific community that the cost would outweigh the benefit. There was no consensus then around the benefits of the genomic sequences, for science or for society.

In 1988, a National Research Council study (on which I served), proposed a compromise whereby much smaller, and therefore cheaper, genomes of genetic model organisms would be sequenced first. The critical argument for me, and indeed for much of the scientific community, was that the sequences of the model organisms could be interpreted through experimental work, yet the homology among similar proteins in diverse organisms would allow us to transfer much of the biological interpretation to the human genome. Genomic sequences of many organisms, not just the human, would allow us to read and ultimately interpret the information in DNA in all of them. So it turned out. The benefits for science have been nothing short of revolutionary.



- We no longer need to theorize or speculate about evolution. In the genome sequences, we have data that fully and quantitatively document the evolution, from common ancestry, of all life on Earth.
- Insights about the functions of human genes and proteins continue to come fast, most often from studies of their homologs in model organisms. We now can study all the genes of an organism simultaneously via methods that were mostly invented to get the sequencing done in the first place.
- The cost of sequencing has fallen dramatically. It is now literally easier and cheaper to sequence the genome of a bacterial or yeast mutant than it is to isolate the gene and sequence only the relevant bits.
- As sequencing costs have fallen, it has become practical to follow sequence heterogeneity in populations, which may allow us to understand the inheritance of complex phenotypes and the basis of complex human diseases. Such studies have transformed our understanding of the origins and history of the human species.

The fears of big science around sequence technology have largely dissipated. Today, individual investigators outsource routine sequencing to a thriving service industry at an astonishingly modest cost. Data-release practices introduced during the human genome sequence project facilitate reuse of existing data in place of pointless and expensive repetition.

This has spread to the functional genomics community and beyond. As with all technology development, some issues remain, such as the cost of computational and sequencing infrastructure, which is still beyond the means of individual small laboratories. These can be dealt with well short of big science by modest increases in funding for shared facilities.

When I began my career, I never imagined that someday I could simply look up a gene's coding sequence; find its orthologs in other organisms; and order, from a service organization, a mutation to my specification for an experiment to reveal gene function. Yet this is now our world, the direct result of a collective agreement to make genomic sequencing a priority in the last decades of the 20th century. It was a very good decision.

## Presenting the Human Genome: Now in 3D!

Yijun Ruan

Senior Group Leader and Associate Director, Genome Institute of Singapore, Singapore, Republic of Singapore

The completion of the human genome sequence in 2001 was, to me, the most important accomplishment in biology. Since then, we have journeyed to the next frontier through significant improvements in our ability to analyze and map gene expression and transcription factor-binding sites in the human genome. We now understand that the genome is far more complex than linear information could explain. Therefore, to fully appreciate the rules by which the genome operates on an organismal level, we have to comprehend higher-order chromosomal organization. To reach that pinnacle, we need first to understand how the genome is spatially organized and how that organization affects basic nuclear and cellular processes. We also need to learn how transcriptional dynamics and epi-

