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 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 epigenetic states drive cell type specificity.
Fruits of Genome Sequences for Biology
David Botstein

Science 331 (6020), 1025.
DOI: 10.1126/science.1204038