

# The Human Genome Project: Lessons from Large-Scale Biology

Francis S. Collins,<sup>1\*</sup> Michael Morgan,<sup>2</sup> Aristides Patrinos<sup>3</sup>

The Human Genome Project has been the first major foray of the biological and medical research communities into “big science.” In this Viewpoint, we present some of our experiences in organizing and managing such a complicated, publicly funded, international effort. We believe that many of the lessons we learned will be applicable to future large-scale projects in biology.

“It is essentially immoral not to get it [the human genome sequence] done as fast as possible,”

James D. Watson (1)

Thinking big comes naturally to many biologists. Pursuing biological research on a monumental scale traditionally has not. Given the depth of that dichotomy, it is amazing that any established scientist would consider signing on to a biological research endeavor of the magnitude of the Human Genome Project (HGP), let alone agree to help steer the ship. Yet, each of us did.

Now that biology’s first large-scale project has been completed, there are many outcomes to celebrate: the development of an array of new technologies; the generation of highly useful genetic, physical, and transcript maps of the genomes of several organisms; the coupling of a scientific research program with a parallel research program in bioethics; and, now, the highly polished sequence of the human genome, free and readily accessible to all. Not only has the HGP essentially completed all of its initial goals and more (Table 1), it has done so more than two years ahead of schedule and at a cost less than originally expected.

The experience of overseeing the work of the HGP provides a number of lessons about the organization and conduct of large, international collaborative projects (2). As such endeavors promise to become even more prevalent in the future, it is our hope these observations will prove useful to others venturing into the world of large-scale biology.

## The Early Days of the HGP

Some of the most significant lessons date to the HGP’s formative days in the mid-1980s,

when a handful of visionaries dared to break ranks with the prevailing view that biological research must always be conducted as a hypothesis-driven enterprise. The first serious discussion of the possibility of sequencing the human genome was convened in 1985 by Robert Sinsheimer, then chancellor of the University of California at Santa Cruz. Many thought the idea was crazy or, at best, premature. But in 1986, Charles DeLisi of the U.S. Department of Energy (DOE) decided to begin funding research into genome mapping and sequencing. In 1988, a special committee of the U.S. National Research Council of the U.S. National Academy of Sciences recommended the initiation of the Human Genome Project, calling for a 15-year project with funding of about \$200 million a year. Though considered gargantuan in biological circles, even that price tag was actually modest when compared with the costs of other big science proposals of that era (Table 2). It was an even better deal when the “useful life” of projects was considered, as the shelf life of the scientific tool produced by the HGP is, effectively, forever.

The genome project received a significant boost in late 1988 when Nobel Laureate James Watson stepped forward to lead a new National Institutes of Health (NIH) component of the effort, which had become a joint NIH-DOE project. Watson’s enthusiasm for the effort was captured by his comment that “only once would I have the opportunity to let my scientific life encompass the path from the double helix to the 3 billion steps of the human genome.” (3). At the same time he was charming Congress with his wit and forthright style, Watson proceeded to rattle the cages of many biological researchers with his blunt dismissal of critics. Motivating Watson was a desire to get the HGP done as swiftly as possible to realize the promise it offered for improving human health. Yet, when Watson departed in 1992, some feared that absent his strong leadership, the tasks of mapping and sequencing the genome would not proceed at the pace needed to finish the project on deadline and within budget.

We each joined the ranks of HGP management during the challenging period of the early to mid-1990s, with Francis Collins assuming the lead role at the NIH in 1993, Michael Morgan at The Wellcome Trust in 1992, and Aristides Patrinos at the DOE in 1995. The next several years were turbulent, as we learned “on the job,” made lots of mistakes, and experienced more than a few moments of great anxiety that the whole enterprise might fail; but ultimately, we watched the creativity, talent, and dedication of those involved in the public genome project surmount every obstacle and beat every deadline. We also realized that what we were learning had implications that extended beyond the human sequencing effort itself to the management of large-scale biology in general.

## Building the Best Teams

Good science can only happen with good scientists. Yet some early critics had predicted that the mind-numbing scale and need for carrying out many repetitive tasks would make the HGP unappealing to the brightest and best minds in the scientific community. But those pessimists had failed to account for the compelling vision represented by the HGP—a project that would only be done once in human history. They had also failed to appreciate that the very scale of the problem, along with the need to develop new technologies, new approaches to automation, and new computational strategies, would represent truly exciting scientific challenges. Finally, they underestimated the ability of Watson and other HGP leaders to spread their enthusiasm about the promise of the HGP, infecting many other scientists with a vision of what might be possible. And so, in the first few years of the HGP, a remarkable collection of scientists, representing many countries, many different disciplines, and many levels of seniority, coalesced around these shared goals.

It took most centers a while, however, to learn how to organize the most effective teams to tackle a big science project. John Sulston, director of the UK’s Sanger Centre (now the Sanger Institute) from 1993 to 2000, recalls that “at first everyone did everything,” following the tradition of manual sequencing groups (4). However, it soon became apparent to Sulston and others that, for the sake of efficiency and accuracy, it was best to recruit staff of varying skills—from sequencing technology to computer analysis—and to allocate the work accordingly.

<sup>1</sup>National Human Genome Research Institute, National Institutes of Health, Building 31, Room 4B09, 9000 Rockville Pike, Bethesda, MD 20892, USA. <sup>2</sup>The Wellcome Trust, 183 Easton Road, London NW1 2BE, England. <sup>3</sup>U.S. Department of Energy, 1000 Independence Avenue, S.W., Washington, DC 20585, USA.

\*To whom correspondence should be addressed. E-mail: fc23a@nih.gov

### The Process Must Be Science-Driven

Historically, there have been two distinct avenues for “managing” the vast pool of scientific talent taking part in megascience projects. There’s the top-down strategy, exemplified by The Manhattan Project’s charge to build an atomic bomb; and there’s the bottom-up approach, illustrated by astronomy teams advocating for varying instrument arrangements and competing for research time on the Hubble Space Telescope.

For the HGP, the decision-making process has been intentionally bottom-up, involving input from leading scientists at international laboratories funded through the peer-review process, as well as advisory councils of distinguished experts and numerous topic-specific workshops that collectively sought opinions from hundreds of scientists with a wide variety of expertise. Such an approach required a managerial leap of faith that all the project’s diverse participants shared a common vision and would pull together to turn that vision into reality. Our faith was justified.

Some might argue that a top-down approach is needed at times for instituting major changes in direction. We found, however, that such decisions, though requiring managerial leadership, still must be grounded on a solid scientific foundation. One such turning point occurred at a meeting in Houston, Texas, in February 1999. At that point, less than 15% of the genome had been sequenced. But the fundamentals of all of the required advances—increases in automation, sophisticated base-calling software, sequencing assemblers, laboratory management systems, and more advanced sequencing instruments—had been achieved. Scientists at some of the large

sequencing centers (especially Sulston at Sanger, Robert Waterston at Washington University, and Eric Lander at the Massachusetts Institute of Technology) had been calling for a dramatic acceleration of the enterprise since the mid-1990s, but there were stark differences in their proposed strategies. Evidence had emerged that a “working draft” of the human genome sequence would be extremely useful for finding disease genes and beginning to explore fundamental features of genome organization. In addition, the formation in May 1998 of a company, Celera Genomics, devoted to sequencing the human genome, created new uncertainties about the future course of the HGP and whether the human genome sequence might ultimately be available only by paid subscription (5).

So, there was “bottom-up” enthusiasm within the public consortium for some sort of major scale-up, but there were also many uncertainties heading into the Houston meeting. When one of us (F. Collins) proposed a specific strategy to move the timetable for the completion of the working draft up by 18 months to the spring of 2000, we were not at all confident about how it would be received by the already stressed sequencing centers. Center directors voiced concerns about the broadened scope and tight deadlines of a dramatically accelerated project, not to mention its odds of success. Everything we had learned in preceding years indicated that attempts to increase sequence production by more than two- or threefold in a year nearly always failed, and this proposal required a scale-up in sequence production by an order of magnitude in a matter of a few months. In the end, however, based on solid scientific

judgment, the center directors rose to the challenge that placed a working draft of the human sequence in the public domain by 26 June 2000.

### Meeting Managerial Challenges

At NIH, the familiar peer-review method of funding biomedical research, scaled up for the HGP, served both to reassure the academic research community of the quality of the work and as an excellent managerial tool for large-scale biology. The NIH approach to resource allocation enabled us to nurture centers that were doing well but also provided the clout necessary to phase out the centers that were failing to reach the most ambitious levels of production and cost efficiency. Just meeting goals was not enough; the centers that succeeded were those that constantly innovated and stretched beyond original expectations. Though difficult, every scientific manager must learn how to optimize production in order to deliver a project of this magnitude.

The DOE brought to the table its own strengths in the managerial arena: an impressive network of national laboratories, each with its own area of scientific expertise. DOE leaders’ experience in managing large-scale science and technology projects provided critical input to the HGP, starting during the formative years and continuing through today. Like NIH, DOE faced its share of tough decisions. One of the most difficult moves, given the sensitivities inherent in the national lab system, was centralizing DOE’s HGP effort in the Joint Genome Institute.

As for The Wellcome Trust, this charitable organization wielded its considerable influence as a relatively small but unfettered

Table 1. HGP goals and dates of achievement.

Area	Goal	Achieved	Date
Genetic map	2- to 5-cM resolution map (600 to 1,500 markers)	1-cM resolution map (3,000 markers)	September 1994
Physical map	30,000 sequence-tagged sites (STSs)	52,000 STSs	October 1998
DNA sequence	95% of gene-containing part of human sequence finished to 99.99% accuracy	>98% of gene-containing part of human sequence finished to 99.99% accuracy	April 2003
Capacity and cost of finished sequence	Sequence 500 Mb/year at <\$0.25 per finished base	Sequence >1,400 Mb/year at <\$0.09 per finished base	November 2002
Human sequence variation	100,000 mapped human SNPs	3.7 million mapped human SNPs	February 2003
Gene identification	Full-length human cDNAs	15,000 full-length human cDNAs	March 2003
Model organisms	Complete sequences of <i>E. coli</i> , <i>S. cerevisiae</i> , <i>C. elegans</i> , <i>D. melanogaster</i>	Finished sequences of <i>E. coli</i> , <i>S. cerevisiae</i> , <i>C. elegans</i> , <i>D. melanogaster</i> , plus whole-genome drafts of several others, including <i>C. briggsae</i> , <i>D. pseudoobscura</i> , mouse, and rat	April 2003
Functional analysis	Develop genomic-scale technologies	High-throughput oligonucleotide synthesis	1994
		DNA microarrays	1996
		Normalized and subtracted cDNA libraries	1996
		Eukaryotic, whole-genome knockouts (yeast)	1999
		Scale-up of two-hybrid mapping	2002

funder to make commitments without the concerns or time constraints of those operating within a political environment. The Trust's international outlook also helped to catalyze efforts to reach worldwide scientific accord on many vital issues, including pre-publication data release.

Another challenge proved to be that managers of "big science" endeavors also have their own bosses, replete with the power to make or break a project. The HGP provided us with unavoidable training in the art of keeping a long-term scientific project—and its budget—on course in an ever-changing

ton, Washington University School of Medicine in St. Louis, Missouri, and the Whitehead Institute, Cambridge, Massachusetts. Means for coordination among the G5 were established in 1998 with the use of a weekly conference call. Initially somewhat prickly, these calls served to share technical and experimental advances within this group whose members had, only a few months before, been competing for the same pool of funds. One useful innovation was to spend part of each call on a "lab meeting" format, in which each center in rotation would present some new advance in automation, experimental protocol, or computational analysis.

established before the publication of the first completed human chromosome (9). Project management at NHGRI, with the aid of weekly reports provided by NCBI, tracked the accumulation of deposited finished bases for each sequencing center and determined whether projected goals were likely to be met, and provisions were made to reassign finishing responsibilities in the event of problems. Fortunately, no major problems arose that required this remedy.

In addition to finished base pairs, many other measures of finishing were tracked for each chromosome. These efforts were assisted by a group of eight scientists drawn from the international sequencing centers, each of whom accepted the final responsibility for ensuring that each chromosome sequence was completed to the pre-agreed standard. These chromosome coordinators submitted regular reports based on clone tiling paths of the individual chromosomes under their responsibility. This enabled both the project managers and the sequencers themselves to track progress on gap closure, an essential process in finishing chromosomes. The reports provided the status of the individual clones and also allowed the managers and sequencers to assess each of the approximately 27,000 joins between clones to be certain each link was rigorously justified.

**Table 2. Comparison of prices of large government projects circa 1990 with their projected useful life-span.**

Proposed project	Projected cost (\$ billion)	Target completion date	Estimated life-span (years)
Space Station Freedom	30.0	1999	30
Earth Observing System	17.0	2000	15
Superconducting Super Collider	11.0	1999	30
<b>Human Genome Project</b>	<b>3.0</b>	<b>2005</b>	<b>Perpetual</b>
Hubble Space Telescope	1.5	1990	15 to 20

sea of political masters with many different agendas. Though there were moments of puzzlement, even hilarity, when politics and science collided (such as a conversation with a lawmaker who thought the genome was only found in the gonads), for the most part we found elected leaders of our countries interested and inspired by this project.

### The Importance of International Participation

The Human Genome Project has been an international endeavor from the start. Initially, the international interactions were set up on a scientist-to-scientist level. Had we at first tried to organize this project through heads of state and ministers of health, the HGP would not have worked so easily and efficiently. The sequence of the human genome was obtained ultimately by 20 centers in six countries: China, France, Germany, Great Britain, Japan, and the United States. All the centers, large and small, played a critical role in the overall effort. The involvement of scientists from diverse nations, working shoulder-to-shoulder despite the lack of any project-wide centralized financial authority, provided a wonderful global sense to this investigation of our shared inheritance. The entire group of 20 centers met face-to-face on a regular basis, providing the opportunity to visit each other's genome centers.

A particularly important leadership role was played by the five largest centers (informally known as the "G5"), which included the Sanger Institute; the DOE's Joint Genome Institute in Walnut Creek, California; and three NIH-funded centers at Baylor College of Medicine in Hous-

### The Value of Explicit Milestones and Quality Assessment

Production projects rise or fall on deliverables. The planners of the HGP included a set of interconnected goals as part of the original master plan that was pivotal in our constant effort to optimize outcomes. Regular revisitations of these goals, as evidenced in the NIH-DOE series of 5-year plans in 1990 (6), 1993 (7), and 1998 (8), were critical exercises in the rigorous analysis of progress and the establishment of ambitious milestones.

Not only were these goals explicit in a manner never before seen in biology, most of the goals were attached to a timetable and sets of intermediate milestones, as well as lists of objective methods to check data quality, enforce standards, and track project costs. However, all managers know that goals, deadlines, and quality checks are worthless if no one tracks whether they are being met. Each center instituted detailed methods of quality assurance and quality control. A "round robin" exercise in which randomly chosen sequence files from each center were checked by two other centers proved very illuminating and reassuring. Furthermore, each of our agencies, but especially the National Human Genome Research Institute (NHGRI) with the help of the National Center for Biotechnology Information (NCBI), had a sufficiently large number of qualified staff whose crucial job was to monitor the sequencing progress and make sure that the project was not slipping.

To ensure consistency among the various centers' efforts to finish their own regions, quality goals for finished chromosomes were

### Rapid Prepublication Data Release

From the beginning, one of the operating principles of the HGP has been that its data and resources should be made available rapidly to the entire scientific community. Release of this fundamental precompetitive information promotes the best interests of science and helps to maximize the public benefit to be gained from research. This has involved the release of data well before publication—far more rapidly than is standard in the scientific community.

In 1991, NHGRI and DOE developed a data release policy that called for the release of data and materials no later than 6 months after generation. In 1996, in a defining moment at a meeting convened by The Wellcome Trust in Bermuda, Sulston and Waterston led the International Human Genome Sequence Consortium to adopt the so-called Bermuda Principles, which expressly call for automatic, rapid release (in this case, within 24 hours) of sequence assemblies of 1 to 2 kilobase (kb) or greater to the public domain. These principles were publicly endorsed by U.S. President Bill Clinton and British Prime Minister Tony Blair in a joint statement issued in March 2000.

A January 2003 scientific gathering in Fort Lauderdale, Florida, "Sharing Data from Large-scale Biological Research Projects," sponsored by The Wellcome Trust, issued a report that

supported applying the Bermuda Principles to “all sequence data, including both the raw traces submitted to the trace repositories at NCBI and ENSEMBL, and whole-genome shotgun assemblies,” while also calling for the scientific community of users to appropriately acknowledge and respect the contributions of data producers (10).

### Technology Matters

Without the development of new technologies and strategies for large-scale, high-throughput generation of biological data at low cost, we would be nowhere near completion of the human genome sequence today (Fig. 1). From the beginning, the project emphasized the development and pilot testing of new technologies. Spurred by pilot project funding at a dozen centers, a series of creative innovations chipped away at rate-limiting steps of large-scale, gel-based Sanger dideoxy sequencing. Pivotal to the project’s success was the genome centers’ implementation of major improvements in library production, template preparation, and laboratory information management, so that less and less human intervention was required in the main production pipelines. The advent of capillary sequencing machines from Amersham and ABI provided a much-needed boost in efficiency, enhancing the gains already being made due to the use of better enzymes and dyes.

Of course, as is usually the case with new technologies, there were plenty of problems. Feeding the sequencing pipeline with pre-mapped bacterial artificial chromosomes (BACs) posed a major challenge, especially within the context of our new accelerated schedule. The capillary sequencing machines, on which our ramped-up effort depended, were fresh off instrument production lines in 1999 and performed poorly in a large-scale production environment for the first few months. But the technology-focused approach eventually began to yield the desired results. By January 2000, our 20 sequencing centers were collectively sequencing 1000 base pairs a second, 24 hours a day, 7 days a week. In the mid-1980s, in contrast, even the best-equipped laboratories could produce only about 1000 base pairs a day.

### Perspectives on the Private Sector

The HGP has joined hands often with the private sector. We have worked closely with instrumentation companies to develop genome-scale technologies. For example, NHGRI’s funding of

DNA microarray research laid the groundwork for the establishment of Affymetrix; similarly, both NHGRI and DOE funding contributed significantly to the development of the application of capillary electrophoresis to DNA sequencing. We have also blazed new paths for partnering with private industry to advance basic biological research, including creation of public domain data on expressed sequence tag (EST) databases, sequencing the genomes of model organisms, and discovering human single nucleotide polymorphisms (SNPs). And we, along with other leaders in the field, anticipate such public-private interactions in basic biological research will become even more common in the future (5).

However, when it comes to science, there is nothing the public—or the media—seems to love better than a heated competition, whether it was the space race between the United States and the Soviet Union or the long-running feud over discovery of the AIDS virus. Whether real or perceived, the competition between the public project and Celera’s private venture was certainly a first of its kind. Not surprisingly, widely varying opinions were expressed on how best to deal with the situation.

From our perspective, it was the commitment to free and rapid data release, not techno-

logical issues, that lay at the heart of the divide between public sequencers and Celera’s private venture. Each group held a profoundly different view of the best way to release genomic data to achieve public benefit. The Celera model was based on a belief that release of genomic data could and should be done by the private sector, and it envisioned varying degrees of data access to scientists depending on their affiliations. However, for those of us guiding the public project, the idea of restricting access to pre-competitive data of such importance to biology and medicine raised the specter of a delay in utilization by the broad scientific community, and ultimately a delay in public benefit.

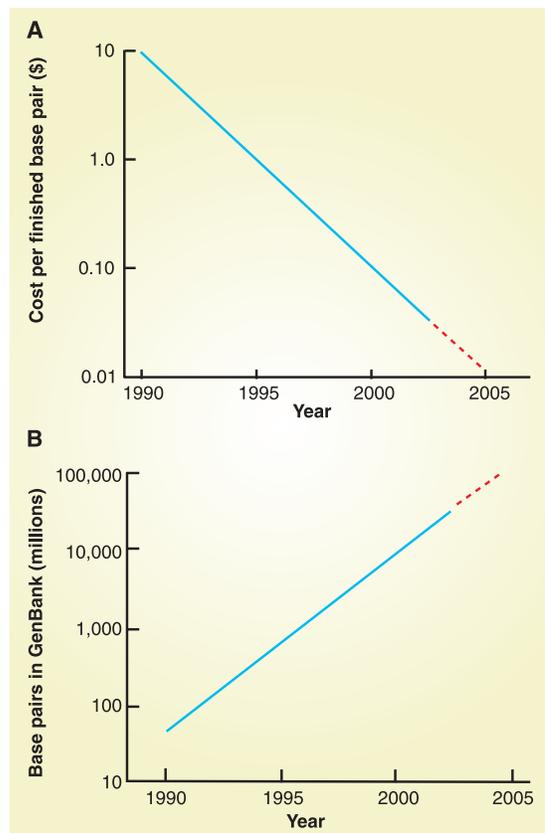
Many outside observers, concerned that the contentious aspects of the relationship between the HGP and Celera were damaging the credibility of all parties, urged that a collaborative approach be sought. Many serious attempts at such models were undertaken, but ultimately all failed. Nonetheless, though it proved impossible to merge the two enterprises, an agreement was brokered, with the help of the negotiating skills of one of us (A. Patrinos), leading to the joint announcements of the working draft in June 2000 and simultaneous publications in February 2001.

### Social Consequences

Becoming the first large scientific undertaking to dedicate a portion of its budget for research into social, legal, and ethical implications, the HGP under Watson’s guidance set aside 3 to 5% of its budget to study how our exponential increase in knowledge about humans’ genetic makeup may affect humankind. The ethical, legal, and social implications (ELSI) program at NHGRI and DOE has provided an effective base from which to assess the implications of genome research. An example of how ELSI research has helped to inform public policy is the fact that more than 40 states in the United States have passed genetic nondiscrimination bills, many based on model language that grew out of such research. Another example is the training of more than 3000 judges through 20 workshops on the fundamentals of genetics that are becoming increasingly important in their courtrooms. The HGP’s forward-looking approach to ethical, legal, and social implications is now being used as a model in other research endeavors.

### Beyond the Human Genome Project

While the HGP is drawing to its conclusion, we fully expect its heirs, the new disciplines of genomics and genomics-based medicine, to carry on its tradition of pushing the envelope of biological thinking. At least for the foreseeable future, there is a pressing need for similar large-scale, coordinated efforts that provide early and unrestricted release of key biological data to the



**Fig. 1.** (A) Decrease in sequencing costs, 1990–2005. (B) Increase in DNA sequence in GenBank, 1990–2005.

scientific community. Among the enterprises that have already headed down this road are comparative genomics, with the genomic sequencing and analysis of key model organisms being done by some of the centers that worked on the human genome. Large-scale genomic enterprises now extend well beyond straightforward sequencing, as witnessed by the recent launch from our own funding agencies of other large-scale biology initiatives involving functional genomics (11), structural biology (12), microbial genomics and proteomics (13), and haplotype mapping focusing on human populations (14). Other candidates for the “big science” approach to biology include interagency and international approaches to biological database creation and maintenance, the construction of public small-molecule libraries for use by basic scientists in their efforts to chart biological pathways, and the large-scale application of microarray technologies with the potential for applications in a wide range of biological research settings.

The millions of people around the world who supported our quest to sequence the human genome did so with the expectation that it would benefit humankind. Now, at the dawning of the genome era, it is critical that we encourage the same intensity toward deriving medical benefits from the genome that has characterized the historic effort to obtain the sequence. If research support continues at vigorous levels, we imagine that genome science will soon begin revealing the mysteries of hereditary factors in heart

disease, cancer, diabetes, schizophrenia, and a host of other conditions. Genomics holds the promise of “individualized medicine,” tailoring prescribing practices and management of patients to each person’s genetic profile. These revelations should also lead us to develop and target drugs in a rational fashion to genes, protein pathways, and networks shown to be involved in primary disease pathogenesis. Furthermore, a better understanding of the genetic factors that influence susceptibility and/or response to various infectious diseases could have an enormous impact on health in the developing world. Genomics also has the potential to help agricultural researchers develop better crops and livestock, environmental scientists create better methods of cleaning up toxic material, production experts streamline industrial processes, and energy researchers work toward sustainable, nonpolluting energy sources.

For this grand vision (15) to come true, however, we in the biologic research community need to pursue the next generation of research projects with the same determination and creativity that the dedicated scientists of the HGP used to spell out the human genetic code. This Herculean challenge is by no means limited to biologists. We call on leaders across science and society, across academia and industry, and across political and geographic boundaries to join us on this exciting voyage to understanding ourselves. If the past 50 years of biology is any indication of the future, the best is certainly yet to come.

#### References and Notes

1. J. D. Watson, *The New York Times*, 5 June 1990, p. C1.
2. W. H. Lambricht, *Managing “Big Science”: A Case Study of the Human Genome Project* (The Price Waterhouse Coopers Endowment for the Business of Government, Arlington, VA 2002). Available at [www.endowment.pwcglobal.com/pdfs/LambrichtReport2.pdf](http://www.endowment.pwcglobal.com/pdfs/LambrichtReport2.pdf).
3. J. D. Watson, *Science* **248**, 44 (1990).
4. J. Sulston, G. Ferry, *The Common Thread: A Story of Science, Politics, Ethics and the Human Genome* (Joseph Henry Press, Washington, DC, 2002), pp. 73–76.
5. M. V. Olson, *J. Mol. Biol.* **319**, 931 (2002).
6. U.S. Department of Health and Human Services, U.S. Department of Energy, *Understanding Our Genetic Inheritance. The U.S. Human Genome Project: The First Five Years* (Government Printing Office, Washington, DC, 1990).
7. F. S. Collins, D. Galas, *Science* **262**, 43 (1993).
8. F. S. Collins et al., *Science* **282**, 682 (1998).
9. *NHGRI Standard for Quality of Human Genomic Sequence* [revised 10 February 2002]. Available at [www.genome.gov/page.cfm?pageID=10000923](http://www.genome.gov/page.cfm?pageID=10000923).
10. “Sharing Data from Large-scale Biological Research Projects,” meeting sponsored by The Wellcome Trust, Fort Lauderdale, FL, January 2003. Available at [www.wellcome.ac.uk/en/1/awtpubrepdat.html](http://www.wellcome.ac.uk/en/1/awtpubrepdat.html).
11. *The ENCODE Project: ENcyclopedia Of DNA Elements* [updated February 2003]. Available at [www.genome.gov/Pages/Research/ENCODE/](http://www.genome.gov/Pages/Research/ENCODE/).
12. Structural Biology Consortium [updated 19 April 2001]. Available at [www.wellcome.ac.uk/en/1/biosfgstr.html](http://www.wellcome.ac.uk/en/1/biosfgstr.html).
13. Genomes to Life [updated 26 February 2003]. Available at <http://doegenomestolive.org>.
14. NHGRI, *The Haplotype Map*. Available at [www.genome.gov/Pages/Research/HapMap](http://www.genome.gov/Pages/Research/HapMap).
15. F. S. Collins et al., *Nature*, in press.
16. We express our profound gratitude to leaders of all of the sequencing centers involved in the International Human Genome Sequencing Consortium, along with all other participants in the HGP, 1990 to 2003.

#### VIEWPOINT

## Realizing the Potential of the Genome Revolution: The Genomes to Life Program

Marvin E. Frazier,<sup>1</sup> Gary M. Johnson,<sup>2</sup> David G. Thomassen,<sup>1\*</sup> Carl E. Oliver,<sup>2</sup> Aristides Patrinos<sup>1</sup>

The systems biology revolution is proceeding along multiple pathways as different science agencies and the private sector have adopted strategies suited to their particular needs and cultures. To meet this challenge, the U.S. Department of Energy has developed the Genomes to Life (GTL) program. A central focus of GTL is environmental microbial biology as a way to approach global environmental problems, and its key goal is to achieve, over the next 10 to 20 years, a basic understanding of thousands of microbes and microbial systems in their native environments. This focus demands that we address huge gaps in knowledge, technology, computing, data storage and manipulation, and systems-level integration.

This month we celebrate progress in completing the human genome sequence as well as the 50th anniversary of James Watson and Francis Crick’s Nobel Prize-winning description of the DNA double helix (1). Spurred by achievements in genomic research, the 21st century will bring revolutionary solutions to some of our most urgent and

expensive challenges in health, energy, and the environment. In this new era, we will understand the workings of entire biological systems and how they interact with and respond to their environments. The Office of Science of the U.S. Department of Energy (DOE) will use that information to conduct basic research into biological solutions for

producing clean energy, removing excess carbon dioxide from the atmosphere, and remediating contaminated environments left as a legacy of the Cold War.

Genomic information is providing the starting point for understanding the instructions for the manufacture of all of life’s molecular machines and the systems needed to control and operate them. Through a labyrinth of pathways, networks, chemistry, and

<sup>1</sup>U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, SC-70, <sup>2</sup>Office of Advanced Scientific Computing Research, SC-30, Germantown Building, 1000 Independence Avenue, S.W., Washington, DC 20585–1290, USA.

\*To whom correspondence should be addressed. E-mail: david.thomassen@science.doe.gov