DOE/ER/61577--T2

Origins of the Human Genome Project*

The human genome project was borne of technology, grew into a science bureaucracy in the United States and throughout the world, and is now being transformed into a hybrid academic and commercial enterprise. The next phase of the project promises to veer more sharply toward commercial application, harnessing both the technical prowess of molecular biology and the rapidly growing body of knowledge about DNA structure to the pursuit of practical benefits.

Faith that the systematic analysis of DNA structure will prove to be a powerful research tool underlies the rationale behind the genome project. The notion that most genetic information is embedded in the sequence of DNA base pairs comprising chromosomes is a central tenet. A rough analogy is to liken an organism's genetic code to computer code. The goal of the genome project, in this parlance, is to identify and catalog the 75,000 or more files (genes) in the software that directs construction of a self-modifying and self-replicating system — a living organism. The main scientific justification for the genome project is not that it will explain all of biology. By the software analogy, studying the structure of DNA cannot directly approach problems of hardware — cells and organs — or of networks — social and environmental interactions. Biology has from its inception made clear the importance of adaptability. The complexity of the brain and its connections, with tens of billions of cells and trillions of connections, or the immense adaptability of the immune system, responding to countless external threats (including infectious organisms) and internal disruptions (including cancer), make clear that the human body is more than the simple expression of tens of thousands, or even hundreds of thousands, of genes.

The genome project is premised on the claim that genetic maps and new technologies will be among the most useful scientific approaches to highly complex biological phenomena, not that these maps will be the end of biology. The genome project is a biological infrastructure

* by Robert Mullan Cook-Deegan, M.D., Institute of Medicine, National Academy of Sciences, 2101 Constitution Avenue, NW, Washington, DC 20418; 202-334-2328; 202-334-1385 fax; bcd@nas.edu (Internet)

initiative, deriving from the fact that with so many investigators using genetic approaches to explore the biological wilderness, it is time to build some roads. The study of DNA structure does unapologetically promise reductionist explanations of some biological phenomena, tracing causes of disease, for example, to mutations in identified genes — that is, identifiable changes in DNA structure that affect biological function. This should not be confused, however, with a simplistic genetic determinism, with all its historical and political baggage. Indeed, the study of a wider variety of genes, diseases, and biological functions will surely dispel the simple-minded renditions of gene function, overwhelming it with myriad concrete examples of biological complexity that defy explanation by linear causal chains. Genes will nonetheless be nodes in many of the causal networks of interesting biological phenomena, and determining DNA structure is one of the surest and fastest ways to probe those networks. Gene maps are essential to this process; the genome project is aimed at providing those maps.

The earliest and most obvious applications of genome research are tests for genetic disorders, but less obvious diagnostic uses may prove at least as important, such as forensic uses to establish identity (to determine paternity, to link suspects of physical evidence of rape or murder, or as a molecular "dog-tag" in the military). Genome research also promises to find genes expeditiously, making the genetic approach attractive as a first step in the study not only of complex diseases, but also of normal biological function. Each new gene is a potential target for drug development — to fix it when broken, to shut it down, to attenuate or amplify its expression, or to change its product, usually a protein. Finding a gene gives investigators a molecular handle on problems that have proven intractable before.

Science administrators and members of Congress who shepherded the budgets for genome research (and their counterparts in other nations and international organizations) supported the project not only because of its medical benefits, but also because they saw it as a vehicle for technological advance and creation of jobs and wealth. The main policy rationale for genome

DISCLAIMER

research was the pursuit of gene maps as scientific tools to conquer disease, but economic development was an explicit, if subsidiary, goal.

The genome project results from the confluence of tributaries that course through many provinces. The technical conception of the genome project derives mainly from precedents in molecular biology, but the story contains other major elements — the advance and dissemination of information technology, restructuring of the science bureaucracy, and increasing participation by commercial organizations. One way to trace these origins is to recount phases in the development of the genome project: how it got started, how it was redefined, and how it is now progressing. The history can be roughly divided into four stages: origins of the idea for a human genome project (the genesis), redefinition of its goals (a period of ideological conflict never completely resolved), emergence into a bureaucracy in the United States and several other nations (the Watson era), and transformation into a government-industry enterprise (still in progress).

Origins of the Idea

The genome project now embraces three main technical goals: (1) genetic linkage maps to trace the inheritance of chromosome regions through pedigrees; (2) physical maps of large chromosome regions, to enable the direct study of DNA structure in search of genes; and (3) substantial DNA sequence information, enabling the correlation of DNA changes with alterations in biological function. If history were logical, then the genome project would have grown from a discussion of each in turn, and how to bring them together into a coherent plan. History is not logical, however, and it was DNA sequencing technology rather than genetic linkage mapping that gave rise to the idea of a human genome project.

Three individuals independently came upon the idea of sequencing the human genome, that is, deriving the order of DNA bases comprising all human chromosomes. (Actually, this will, like other biological maps, be a composite or reference genome, as there is inherent variation

among individuals. While the order of genes and chromosome segments is generally quite stable, it is individual variations that are often of greatest interest. Gene maps help by laying out the overall structure, while much interesting biology comes from understanding how variations come about and what they cause.)

The seminal technology that led to the genome project was a group of techniques for determining the actual sequence of base pairs in DNA. In 1954, just a year after Watson and Crick described the double helical structure of DNA, George Gamow speculated that DNA sequence was a four-letter code embedded in the order of base pairs [Gamow, 1954 #1017]. In 1975, Fredrick Sanger announced to a stunned audience that he had developed a way to determine the order of those base pairs efficiently [1-3]. Alan Maxam and Walter Gilbert at Harvard independently developed a completely different method that same year. This method was announced to molecular geneticists late in the summer of 1975 at scientific conferences, and circulated as recipes among molecular geneticists until formal publication in 1977 [4]. Half a decade later, many groups began successfully to automate the process, in North America, Europe, and Japan. The first practical prototype was produced by a team at the California Institute of Technology in 1986, under the direction of Lloyd Smith, as part of a large team under Leroy Hood[5]. This prototype was quickly converted to a commercially available instrument by Applied Biosystems, Inc., and reached the market in 1987.

The new technologies for DNA sequencing spread through the biomedical research community like wildfire. By 1978, it was becoming apparent that sequence information needed to be catalogued systematically to make it useful to the scientific community. The idea of a database to contain this information emerged as a priority from a meeting at Rockefeller University that year. After several years of often intense and acrimonious discussion, twin databases were established under the European Molecular Biology Laboratory in Heidelberg at as GenBank at Los Alamos National Laboratory[6]. These databases were established just as personal computers were beginning to prove their immense power in biology laboratories. The

explosion of minicomputers in the 1970s and microcomputers in the 1980s fueled the attention to DNA sequence information, because computational methods were obviously the only way to analyze the deluge of DNA sequence information produced by sequencing techniques[6-9]. The technologies were thus present, but it took the spark of an idea of using them as part of a large organized effort to ignite the fire, out of which rose the human genome project.

Robert Sinsheimer, then chancellor of the University of California, Santa Cruz, thought about sequencing the human genome as the core of a fund-raising opportunity in late 1984. He and others convened a group of eminent scientists to discuss the idea in May 1985[10]. This workshop planted the idea, although it did not succeed in attracting money for a genome research institute on the campus of UCSC. Without knowing about the Santa Cruz workshop, Renato Dulbecco of the Salk Institute conceived of sequencing the genome as a tool to understand the genetic origins of cancer. Dulbecco, a Nobel-Prize winning molecular biologist, laid out his ideas on Columbus Day, 1985, and subsequently in other public lectures and in a commentary for *Science* magazine[11; 12]. The commentary, published in March 1986, was the first widely public exposure of the idea, and gave impetus to the idea's third independent origin, already gathering steam.

Charles DeLisi, who did not initially know about either the Santa Cruz workshop or Dulbecco's public lectures, conceived of a concerted effort to sequence the human genome under the aegis of the Department of Energy (DOE). DeLisi had worked on mathematical biology at the National Cancer Institute, the largest component of the National Institutes of Health. How to interpret DNA sequences was one of the problems he had studied, working with the T-10 group at Los Alamos National Laboratory in New Mexico (a group of mathematicians and others interested in applying mathematics and computational techniques to biological questions). In 1985, DeLisi took the reins of DOE's Office of Health and Environmental Research, the program that supported most biology in the Department. The origins of DOE's biology program traced to

the Manhattan Project, the World War II program that produced the first atomic bombs, and concern about how radiation caused genetic damage.

In the fall of 1985, DeLisi was reading a draft government report on technologies to detect inherited mutations, a nagging problem in the study of children to those exposed to the Hiroshima and Nagasaki bombs, when he came up with the idea of a concerted program to sequence the human genome[13]. DeLisi was positioned to translate his idea into money and staff. While his was the third public airing of the idea, it was DeLisi's conception and his station in government science administration that launched the genome project.

Redefining the Technical Goals

Molecular biologists did not welcome the idea with open arms. While many, especially those who studied medical genetics and the inheritance of genetic diseases, were enthusiastic, the broader community of protein biochemists and even molecular geneticists were far more skeptical. The year 1986 was a time of setback and redefinition for the genome project. The nadir of the project's trajectory came at a meeting at Cold Spring Harbor Laboratory in June 1986. A rump session was called to discuss Dulbecco's editorial. Walter Gilbert, who had been infected with the Santa Cruz bug, laid out a rationale for the project and then began to describe its technical goals and price tag. The discussion quickly veered into the politics of biomedical research — the dangers that large projects posed for budgets to support small investigator-initiated research (the space shuttle used as the negative icon) and the questionable competence of DOE to run such a project. David Smith, as the DOE representative, faced a largely hostile audience, although he also got many private expressions of support.

The controversy provoked a number of events on the policy front, and the debate moved to Washington, DC. The Howard Hughes Medical Institute, which had begun to get interested in the genome project, held a well-attended international forum in July 1986. In October, NIH hosted a discussion in conjunction with a meeting of the NIH Director's Advisory Committee.

These two meetings exposed considerable rancor among the ranks of prominent molecular biologists, but they also began the search for common ground, and laid the groundwork for a two-year succession of countless meetings that redefined the human genome project. The redefinition took place most conspicuously in a committee of the National Research Council (NRC).

In September, 1986, two projects were initiated to study the idea. The NRC, the largest operational arm of the National Academy of Sciences, approved a study. The NRC appointed a committee of extremely prestigious researchers chaired by Bruce Alberts of the University of California at San Francisco. This study committee vigorously debated the merits of a concerted scientific program, carrying out in microcosm the debate transpiring more broadly in the scientific community.

The NRC committee took a commonsense approach, looking at the scientific and technical steps that would be necessary to construct comprehensive maps of the human genome and to make sense of the resulting information. They started by bringing together those constructing various kinds of genetic maps in different organisms. The idea of a human genetic linkage map grew out of work in viruses, bacteria, yeast, and other organisms. The key insight grew from a 1978 inspiration shared between David Botstein, then at the Massachusetts Institute of Technology, and Ronald Davis of Stanford. In a discussion at Alta, Utah, they speculated that researchers could find natural DNA differences among individuals in families, most of which would not necessarily lead to clinically detected differences, to trace the inheritance of chromosome regions through those families.

Each person has a pair of each of the 22 nonsex chromosomes. (Women also have a pair of X chromosomes, while men have an X and a Y.) Botstein and Davis suggested that if detectable differences could be found for discrete chromosome regions, then one could figure out which of each parent's chromosome pair was inherited by each child. A map of such differences would enable geneticists to determine the approximate location of disease-associated and other genes,

even if they had no prior clues about the gene's function [14]. By late 1979, the first such DNA marker was found by Arlene Wyman and Raymond White, working in Worcester, Massachusetts[15].

These heterogeneous DNA markers were quickly used to hunt for disease genes, demonstrating the utility of the gene mapping idea. Suppose, for example, that some adult progeny of the same mother (or father) Huntington's disease also developed it, while other children did not. If the affected children all inherited DNA from the same region of chromosome 4, while those unaffected inherited the other copy of that DNA, this would be strong statistical evidence that DNA in that chromosome 4 region contained the Huntington's disease. This is exactly what James Gusella and others discovered in 1983, when they linked Huntington's disease to the tip of chromosome 4[16]. The DNA marker they used to track the passage of chromosome 4 in families was not the gene itself, but a nearby region that just happened to differ among family members so that the investigators could tell the chromosomes apart. Finding the gene itself took another decade of arduous work, but it was ultimately successful, made possible only because genetic linkage narrowed the zone of DNA to scan for the offending mutation[17].

The second cluster of mapping techniques centered on structural catalogs of DNA fragments, rather than markers to track inheritance through pedigrees. The general idea was to take native chromosomal DNA, break it into fragments that could be copied by various cloning techniques, and then put the DNA fragments (now plentiful enough to study in the laboratory) back in order. If this could be done for all the chromosomes, then once a gene's location were narrowed to a particular region by genetic linkage, then the DNA from that region would already be available in a test tube for direct analysis.

The techniques for physical mapping were again derived from work on viruses and bacteria, and by the mid-1980s, pioneering groups had moved into constructing physical maps of larger and more complex organisms. Maynard Olson and his colleagues at Washington University were working on a physical map of yeast, which was a very powerful model for the

genetics of organisms with nucleated cells[18]. In Cambridge, UK, Alan Coulson, John Sulston and their colleagues were working on a physical map of the nematode *Caenorhabditis elegans*[19]. *C. elegans* had been identified by Sydney Brenner as a powerful model to apply genetic techniques to study development and behavior of organisms containing differentiated organs, including a primitive nervous system[20]. John Sulston had mapped the lineage of every cell in the body of one developmental stage[21; 22], and others at Cambridge had traced the connections of the entire nervous system[23]. While the entire genomes of yeast and nematode were only the size of a singe human chromosome, many believed that similar techniques would prove applicable for the entire human genome, more than an order of magnitude larger. The prospects for physical mapping brightened in 1987, when David Burke and Georges Carle, working with Maynard Olson, developed a technique to clone DNA fragments hundreds of thousands of base pairs in length[24], considerably reducing the complexity of constructing large-scale physical maps.

The NRC committee ultimately redefined the project to embrace the entire set of genetic maps, giving much greater prominence to genetic linkage mapping and physical mapping than to sequencing. The committee also underscored the importance of organisms other than the human[25]. The committee recommended an annual budget of $200 million for 15 years, supporting the budget recommendations of a previous DOE advisory committee[26]. The budget recommendations of the two reports were quite similar, but where the DOE advisors urged DOE to take the lead, the NRC committee recommended only that there be a lead agency, and proffered NIH, DOE and NSF as the three options.

The congressional Office of Technology Assessment (OTA) project on the human genome initiative was approved in the same hour of the same day as the NRC study. While the NRC committee crafted a scientific strategy and made specific recommendations, the OTA report focused more on its policy rationale (why Congress should or should not support it) and the attendant policy issues. OTA surveyed international activity, and dwelt far more on issues of technology transfer, ethical and social implications of genome research, and research

management[27]. OTA's only substantive difference with the NRC report centered on the notion of a "lead agency." OTA warned that if a lead agency meant control of all funding, then picking one would invite internecine warfare between NIH and DOE, the most likely result of which would be death of the project. OTA did not offer specific recommendations, but in congressional testimony, it clearly favored a truly collaborative effort worked out between the two agencies, with a congressionally mandated task force as the backup option if the agencies failed to produce an acceptable agreement[28].

The genome project rose like the Phoenix from the ashes of Cold Spring Harbor. A vigorous two-year debate culminated in a pair of reports that smiled on, indeed pointed out the inevitability of, systematic gene mapping on the scale of the entire human genome. The next step was to translate the scientific strategy into a funded set of coordinated programs.

Establishment of Government Programs with Process Goals

The first move toward a genome bureaucracy came in the fiscal year 1987 DOE budget. DeLisi set aside $5.5 million of discretionary funds already appropriated, reprogramming them for his newly conceived genome research program. The first congressional action came with the fiscal year 1988 budgets, during hearings in the Spring and summer of 1987. DeLisi cleared a several-year program of genome research funding through the Department and then with the White House Office of Management and Budget. This was incorporated into the President's budget, and duly appropriated, with earmarked spending authority beginning in October 1987. On the NIH side, no request for genome research funding went into the President's budget request, but in response to questions from the House Appropriations subcommittee, Wyngaarden indicated that NIH could use $30 million for gene mapping if Congress chose to appropriate $500 million or more in excess of the Presidential request. Nobel laureates James D. Watson and David Baltimore met with Members and staff from both House and Senate Appropriations Committees in May 1987, primarily to seek additional funding for AIDS research, but Watson also asked for $30 million in genome research funds. The House duly earmarked $30 million,

but the Senate only earmarked $6 million, and a compromise between the two houses split the difference.

The genome project was thus established by congressional action at both NIH and DOE, beginning with the 1988 budget. DOE had long before established a genome program office; in October 1988, Wyngaarden appointed Watson an associate director for NIH in charge of genome research coordination. The newly appropriated funds were to be spent through the National Institute of General Medical Sciences in fiscal years 1988 and 1989, but Watson's office was to coordinate these funds with over $300 million being spent on genome research throughout the NIH institutes. In October 1989, the Department of Health and Human Services established the National Center for Human Genome Research at NIH, giving it authority to expend federal research funds directly, beginning with the 1990 fiscal year, rather than channeling them through the National Institute of General Medical Sciences.

The National Science Foundation had a major instrumentation program, substantial interests in plant and animal genome research, and considerable strength in computational biology, but it did not earmark funding or create a new management structure.

Outside the United States, an Italian genome program began in May 1987[29], tracing its roots to Renato Dulbecco's talk for the Italian Embassy in Washington, DC on Columbus Day 1985. In the USSR, Alexander Bayev and Andrei Mirzabekov presented the idea for a genome program to government officials in December 1987, and secured support for a new program after Bayev addressed the General Assembly of the USSR Academy of Sciences in March 1988, and subsequently obtained approval from the USSR Council of Ministers in December 1988[30]. When the USSR dissolved, the genome project survived, as a component of the Russian science program.

The United Kingdom launched its genome program in February 1989, combining forces between the government's Medical Research Council and the private Imperial Cancer Research

Fund in London[31; 32]. British molecular biologist Sydney Brenner wrote a letter to the European Commission in February 1986 to urge creation of an EC program aimed at a "Map of Man"[33]. Genome research programs on bacteria, yeast, and other organisms developed at EC over the next year. The human genome research program elicited concern in the European Parliament about its social and ethical implications. The EC program ultimately set aside over 7 percent of its budget to scrutinize these impacts, changed its name from "predictive medicine" to "human genome analysis" to address concerns among the German Green Party. With these changes and some other minor stipulations, the EC human genome program began in June 1989[34-37].

The process in Japan was complex. Japan was the first nation to have a government program dedicated to automating the process of DNA sequencing. Akiyoshi Wada was appointed director of a program that began in April 1981 for this purpose, sponsored by Japan's Science and Technology Agency and carried out at the RIKEN Institute in Tsukuba City. (By contrast, the first government funds for automation of DNA sequencing came in a 1984 grant to Caltech.)

When debate about the genome project began in North America and Europe in 1985, and especially when it picked up in 1986 and 1987, Japan's Ministry of Education, Science, and Culture (Monbusho), which supports the vast majority of university-based scientific research, appointed an advisory committee chaired by Osaka University professor Kenichi Matsubara. Monbusho began a modest genome research effort in April 1989, and the Science and Technology Agency expanded its genome research efforts that same year. The Ministry of Health and Welfare initiated an intensified effort to support hunts for disease-associated genes, and the Ministry of International Trade and Industry began planning for its own genome initiative in 1990, although its initiation was delayed by competition for funds. Japan's agriculture ministry began an effort to map the rice genome, funded largely by private funds gathered at sporting events.

France announced plans to mount a government-supported genome research effort in June 1990, and set aside funding beginning in October that year[38]. This augmented a relatively small grants program for genome research commenced in 1988. Canada joined the chorus in 1992[39]. Several European nations also augmented their funding for human genetics during this period,[36; 40] contributing to an accelerating pace of gene discovery. In June 1990, Latin American scientists formed a regional network to encourage collaboration on genome research with laboratories in North America and Europe and among themselves[41; 42].

The genome project thus grew rapidly into an international effort supported by many governments and the EC. There was strong consensus on the need for complete genetic linkage and physical maps, and general agreement about the need for new sequencing technologies. There was disagreement, however, about the degree to which large-scale DNA sequencing should be initiated and outright controversy about the best scientific strategy to pursue in large-scale sequencing efforts.

As the genome project was transformed from a series of meetings and policy reports into an actual scientific program, it added several process goals The technical goals for gene mapping remained, but several policy goals were added. One distinctive aspect of the genome project was its explicit attention to technology development in addition to science. Attaining the technical goals depended on new technologies, and developing new biological methods, instruments, automata and robots, and other new technologies became an explicit objective.

An unprecedented commitment to support research on social, legal, and ethical implications of genome research became the second process goal. Discussion about the social implications of human genetics had attended the genome debate from its earliest phases in Washington, and the history of eugenics cast a long shadow over the genome debate, particularly in German-speaking Europe. Both the NRC and OTA reports explicitly acknowledged the importance of social and ethical issues, and the need to address them head-on as the genome project progressed.

James Watson announced that the NIH program would include a budget set-aside for such research when he was announced as associate director for human genome research in September 1988. Other programs throughout the world, except the UK program, followed suit. (In the UK, such discussion was generally delegated to the private Nuffield Council, established to mediate a national debate on matters of bioethics.) This development deserves a separate treatment, but one particular aspect of this program deserves special mention here — a renewed commitment to technology transfer.

Ensuring that the fruits of genome research were quickly translated into useful applications (and thence into jobs and wealth) became a second process goal for the human genome project. Even as the various government programs noted above began to take shape, private interests also began to mount genome research programs, some of them more significant than publicly funded programs in their nations. In the United States, the Howard Hughes Medical Institute focused on issues not drawing sufficient attention from government, concentrating on databases and helping support the initiation of the Human Genome Organization to coordinate international efforts. In the UK, the Imperial Cancer Research Fund was an equal partner with the government Medical Research Council early on, and the private Wellcome Trust made even larger investments in new genome research and informatics centers in 1992 and 1993. In France, the most vigorous genome research effort was supported by the Centre d'Etude du Polymorphism Humain (CEPH), which formed a partnership with the private French Muscular Dystrophy Association to establish the Genethon, a highly automated genome research facility outside Paris. This effort was started quickly, and dwarfed the government genome research program. In Japan, the Saitama Research Center, the Chiba prefectural government, and other private groups began genome research efforts separate from the various government-sponsored programs.

The international efforts were united in a desire to share map and DNA sequence data widely. The idea behind gene maps was to use them as tools to speed research, and to reduce the need for multiple laboratories throughout the world to develop maps of the same regions when

hunting for different genes. Maps would only be as useful insofar as they were complete, and completeness depended on sharing data freely and rapidly. CEPH was formed in 1984 to forge an international collaboration for genetic linkage maps of human chromosomes[43]. The groups searching for various genes also formed international collaborations, intended to speed sharing of data and materials. This international ethic of sharing, however, had to contend with a growing set of commercial attachments that seemed likely to alter the rules governing collaboration within and across national borders.

Commercial Pursuits

Most of the initial efforts were funded by nonprofit groups hoping to further research. Beginning in 1992, however, a new wave of genome research centers began to take shape, only these were often supported by venture capital or private corporate funds. Existing genome research centers also developed ties to industry. In mid-1992, J. Craig Venter announced his intention to form The Institute for Genomic Research (TIGR). (His work formed the basis for the patent application for expressed sequence tags, which is discussed below.). This new institute was then the largest private investment, and its work was linked through agreements on intellectual property rights to a somewhat larger for-profit unit, Human Genome Sciences, Inc. Human Genome Sciences, Inc., in turn, announced an agreement in excess of $130 million with Smith-Kline-Beecham in May 1993, and William Haseltine was selected as Chief Executive Officer. Another company, InCyte, began a major program in genome research during 1992 and into 1993. Several private firms, including Mercator, Darwin Molecular, Genomyx, and others, pursued plans to develop instruments or pursue pharmaceutical development strategies that involved some mix of genome research.

Corporate funds were not attracted merely by hot science, but also by the prospects of diagnostic applications and more expeditious drug discovery. In every nation where the genome project was presented to its government, including the USSR, promoters pointed to the potential for genome research to create jobs and wealth through new technology. The true potential for

wealth, however, lay not in the new technologies, but in applying them to practical uses. There would doubtless be a spate of new instruments and reagents that could be sold, but this would be a relatively small research market in comparison to medical diagnostics, and smaller still in comparison to therapeutic pharmaceuticals or agriculture. In the medical arena, the most compelling rationale for corporate investment was not in technologies being pursued, but in the terrain being mapped, that is, genes embedded in the human genome. Private investments presumed a means to stake claims on that territory. Those claims would necessarily change the complexion of research, altering the rules by which materials and data were exchanged. The claims being staked were in the form of patents or trade secrets.

Each national government had thus been encouraged a genome research program not only to expedite biomedical research, but also to promote national economic development. These goals could not both be pursued to their logical ends without conflict, as national economic development would by definition mean winning an international economic competition, which was not entirely compatible with unfettered international sharing of data, information, and technology.

The seriousness of the conflict was brought to the surface by an international controversy provoked by a US patent application filed by NIH in June 1991. This patent application will be discussed at greater length and with greater authority by others in this conference, but several points should be made clear here. First, much of the public controversy was poorly framed in ethical terms. Sanctimonious claims were made about direct links between human genes and human dignity. DNA is a universal genetic code, and it will be difficult if not impossible to distinguish human genes from those derived from other organisms. This argument cannot be taken too far, as it is obvious that the human genome in aggregate contains the plans for a human instead of a monkey or nematode or yeast, but it is equally clear that very few, if any, genes will be exclusively human in origin. A classic 1975 paper by King and Wilson showed that the average protein sequence differed only one percent between humans and pygmy chimps, and the

difference at the DNA level was only slightly greater [King, 1975 #915]. The obvious implication was that humans differed more in the timing and quantity of gene expression, rather than which genes there were.

It is far from clear what a proscription on patenting "human" genes would entail, how it could be made meaningful in the law, and whether it would do any good. In most cases, patenting an animal gene and then slightly modifying it for another patent would cover the same material as a human gene. A simple genetic determinism would seem to lie at the root of this equation of DNA with dignity. The factors that distinguish humans from other organisms seem more likely to be nuances of gene expression, development, and environmental response than the collection of genes in the human genome. The brain, for example, is an organ seemingly adapted to be able to change its structure and function in response to environmental stimuli, even more than other organs. No CD-ROM containing Lincoln's DNA sequence could tell us much we would care to know about why he became an historically important figure.

The NIH patent dispute did surface a true international policy dilemma nonetheless, but it was not in patenting policy *per se* but in conflicts between the goal of quickly constructing comprehensive maps and databases as a worldwide scientific effort, and the goal of linking genome research to each nation's domestic economic development. It was not a simple conflict with data-sharing, since investigators in each company could release data as soon as patents were filed. Rather, it was the incentive for each nation to structure its science effort so as to secure its intellectual property rights before the others. Data could be shared only after stakes were claimed, and this could theoretically provoke an international genome gold rush.

If one of the purposes of an international effort was to reduce the duplication of effort that necessarily follows from a purely competitive strategy, then this efficiency was at risk. Taken to an absurd extreme, each nation might choose to attempt to patent the pathways to all human genes before making its data available to others. In this case, all nations would have to map the entire genome. Every nation would be aiming at the same goal, expending its resources to win

the race, but only the winning effort would secure the intellectual property rights. This is a recipe for inefficiency, a true multi-player prisoner's dilemma.

A final point about the NIH patent application is that the policy dilemma was sure to surface. If NIH had not filed a multi-gene patent application, private firms surely would have. The terms of the debate might have been different, and it might have been long delayed and less conspicuous, as the patent application need not have been publicly known for some time, but the debate was nonetheless inevitable. Whether a quieter and later debate might have been better or worse is a matter about which we can surely speculate, but will never be certain.

One of the most interesting aspects of technology transfer related to the genome project is how the project is caught in a changing of the rules. To make this point more starkly, we can perhaps discuss what might have been different if the techniques for DNA sequencing had been patented, as surely they could have been. These techniques are at least as central to research as the polymerase chain reaction that was patented. In the long list of citations to technical orgins of the human genome project, some items have been patented, and others not. The Cohen-Boyer patent for recombinant DNA was a centrally important technique of molecular biology. It was patented, but then licensed for relatively low fees. The polymerase chain reaction, discovered at Cetus Corp. in 1983 and then sold to Hoffmann-La Roche in 1991, was patented and then controlled through a complex set of relatively high-fee licenses for various applications and reagents. The two main techniques for DNA sequencing itself developed in 1975, however, were surely patentable but were never patented. Laboratory instruments, such as DNA sequenators and DNA synthesizers, were sold, with the price of the instrument and its reagents covering patent fees. These disparate ways of handling research methods and tools clearly affected who could use them, and perhaps also the pace of discovery and application, but how and to what degree was a matter of speculation and ideology more than empirical analysis.

It is far from clear what can explain these differences, aside from historical happenstance and the changing norms of biomedical research between the 1970s and the 1990s. It is even

more evident that there is not analytical answer to the question: is it good for science to patent discoveries? Or the question: is it good for the nation to patent research tools? Or even the question: is it good for technology tranfer to patent discoveries? Answers to these questions will no doubt differ from case to case, but analysis of the factors that distinguish cases might well lead to more sophisticated, and more successful, national policies and international agreements regarding intellectual property and the sharing of data, materials, and technologies.

Those grounded in the pharmaceutical industry often take the benefits of patenting as an article of faith, as well they might since the entire industry truly rests on a foundation of patent protection for chemical entities. There is nonetheless a disturbing dearth of literature on the transaction costs of patenting, or the untoward effects on the research enterprise from a need for complex cross-licensing and constraints on sharing of data and materials, especially in the domain of research tools. Those grounded in the ethos of science, in contrast, take the benefits of free exchange as an article of faith, but there is here a dearth of data about the therapeutic innovations foregone for lack of private investment.

Patent law has historically proven to be a flexible instrument, and a powerful engine for innovation, but it is equally clear that much of the debate about patent policy and technology transfer takes place in the absence of empirical data about outcomes, let alone analysis of long-term social impacts. The permissive interpretation of biotechnology patent law of the 1980s combined with a series of "technology transfer" statutes and executive orders to make a volatile mix. These trends moved policy strongly toward heavier reliance on patents, but with little analysis of their impact on the pace of discovery or on international science. Where facts are sparse, ideology fills the void. Even a cursory inspection of technology transfer policies relating to genome research leads to one obvious conclusion: all nations will be better off if the contending ideologies are disciplined by carefully designed empirical research.

US Genome Research Budgets at NIH and DOE

Based on budget documents prepared for the House and Senate Appropriations Committees 1987-1993, and projections by the Department of Energy and National Center for Human Genome Research.

Fiscal Year	DOE ($ million)	NIH ($ million)
1987	5.5*	-
1988	10.7	17.2
1989	18.5	28.2
1990	27.2	59.5
1991	47.2	87.4
1992	61.4	104.8
1993	63.1	106.1

* The first year's funding at DOE came from funds that Charles DeLisi reprogrammed from research budgets within the Department, and did not require congressional action. The first congressionally earmarked funding for both NIH and DOE came in Fiscal Year 1988.

References

1. Judson, HF. (1987). *Mapping the Human Genome: Historical Background* (Mapping Our Genes contractor reports, Vol. 1, NTIS Order No. PB 88-160-783/AS). Office of Technology Assessment, US Congress. September 1987.

2. Sanger, F, and Coulson, AR. (1975). Rapid Method for Determining Sequences in DNA by Primed Synthesis with DNA-Polymerase. *Journal of Molecular Biology* **94** : 441-448.

3. Sanger, F. (1975). The Croonian Lecture, 1975: Nucleotide Sequences in DNA. *Proceedings of the Royal Society of London* **B 191** : 317-333.

4. Maxam, AM, and Gilbert, W. (1977). A New Method for Sequencing DNA. *Proceedings of the National Academy of Sciences (USA)* 74 (February): 560-564.

5. Smith, LM, Sanders, JZ, Kaiser, RJ, Hughes, P, Dodd, C, Connell, CR, Heiner, C, Kent, SBH, and Hood, LE. (1986). Fluorescence Detection in Automated DNA Sequence Analysis. *Nature* **321** (12 June): 674-679.

6. Smith, TF. (1990). The History of the Genetic Sequence Databases. *Genomics* **6** (April): 701-707.

7. Bishop, MJ, and Rawlings, CJ. (1987). *Nucleic Acid and Protein Sequence Analysis: A Practical Approach*. IRL Press, Oxford, UK.

8. Waterman, MS. (1989). *Mathematical Methods for DNA Sequences*. CRC Press, Boca Raton, FL.

9. Waterman, MS. (1990). Genomic Sequence Databases. *Genomics* **6** (April): 700-701.

10. Sinsheimer, R. (1989). The Santa Cruz Workshop, May 1985. *Genomics* **5** : 954-956.

11. Dulbecco, R. (1986). A Turning Point in Cancer Research: Sequencing the Human Genome. *Science* **231** : 1055-1056.

12. Dulbecco, R. (1987). A Turning Point in Cancer Research: Sequencing the Human Genome. *In Viruses and Human Cancer*, Ed., pp. 1-14. Alan R. Liss, Philadelphia.

13. DeLisi, C. (1988). The Human Genome Project. *American Scientist* **76** : 488-493.

14. Botstein, D, White, RL, Skolnick, M, and Davis, RW. (1980). Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. *American Journal of Human Genetics* **32** : 314-331.

15. Wyman, AR, and White, RL. (1980). A Highly Polymorphic Locus in Human DNA. *Proceedings of the National Academy of Science (USA)* **77** : 6754-58.

16. Gusella, JF, Wexler, NS, Conneally, PM, Naylor, SL, Anderson, MA, Tanzi, RE, Watkind, PC, Ottina, K, Wallace, MR, Sakaguchi, AY, Young, AM, Shoulson, I, Bonilla, E, and Martin, JB. (1983). A Polymorphic DNA Marker Genetically Linked to Huntington's Disease. *Nature* **306** : 234-238.

17. Huntington's Disease Collaborative Research Group, MacDonald, ME, Ambrose, CM, Duyao, MP, Myers, RH, Lin, C, Srinidhi, L, Barnes, G, Taylor, SA, James, M, Groot, N,

MacFarlane, H, Jenkins, B, Anderson, MA, Wexler, NS, Gusella, JF, Bates, GP, Baxendale, S, Hummerich, H, Kirby, S, North, M, Youngman, S, Mott, R, Zehetner, G, Sedlacek, Z, Poustka, A, Frischauf, A-M, Lehrach, H, Buckler, AJ, Church, D, Doucette-Stamm, L, O'Donovan, MC, Riba-Ramirez, L, Shah, M, Stanton, VP, Strobel, SA, Draths, KM, Wales, JL, Dervan, P, Housman, DE, Altherr, M, Shiang, R, Thompson, L, Fielder, T, Wasmuth, JJ, Tagle, D, Valdex, J, Elmer, L, Allard, M, Castilla, L, Swaroop, M, Blanchard, K, Collins, FS, Snell, R, Holloway, T, Gillespie, K, Datson, N, Shaw, D, and Harper, PS. (1993). A Novel Gene Containing a Trinucleotide Repeat That Is Expanded and Unstable on Huntington's Disease Chromosomes. *Cell* **72** (26 March): **971-983.**

18. Olson, MV, Dutchik, JE, Graham, MY, Brodeur, GM, Helms, C, Frank, M, MacCollin, M, Scheinman, R, and Frank, T. (1986). Random-Clone Strategy for Genomic Restriction Mapping in Yeast. *Proceedings of the National Academy of Sciences* **83** (October): 7826-7830.

19. Coulson, A, Sulston, J, Brenner, S, and Karn, J. (1986). Toward a Physical Map of the Genome of the Nematode *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences (USA)* **83** (October): 7821-7825.

20. Brenner, S. (1973). Genetics of Behavior. *British Medical Bulletin* **29** : 269-271.

21. Sulston, JE, Schierenberg, E, White, JG, and Thompson, JN. (1983). The Embryonic Cell Lineage of the Nematode *Caenorhabditis elegans*. *Developmental Biology* **100** : 64-119.

22. Sulston, JE. (1983). Neuronal Cell Lineages in the Nematode *Caenorhabditis elegans*. *Cold Spring Harbor Symposia on Quantitative Biology* **48** : 443-452.

23. White, JG, Southgate, E, Thompson, JN, and Brenner, S. (1986). The Structure of the Nervous System of the Nematode *Caenorhabditis elegans. Philosophical Transactions of the Royal Society of London* Series B, Vol. 314 : 1-340.

24. Burke, DT, Carle, GF, and Olson, MV. (1987). Cloning of Large Segments of of Exogenous DNA into Yeast Artificial-Chromosome Vectors. *Science* 236 : 806-808.

25. National Research Council. (1988). *Mapping and Sequencing the Human Genome*. National Academy Press, Washington, DC.

26. Health and Environmental Research Advisory Committee. (1987). *Report on the Human Genome Initiative* . Office of Energy Research, US Department of Energy.

27. US Congress. (1988). *Mapping Our Genes—Genome Projects: How Big? How Fast?* . Office of Technology Assessment, OTA-BA-373, Washington, DC: Government Printing Office; reprinted by Johns Hopkins University Press. April.

28. US House of Representatives. (1988). *Coordination of Genome Projects, in Committee report on H.R. 4502 and S. 1966, the Biotechnology Competitiveness Act* (Committee Print No. 138). Subcommittee on Natural Resources, Agriculture Research, and Environment and the Subcommittee on Science, Research, and Technology of the House Committee on Science, Space and Technology. 14 July.

29. Dulbecco, R. (1990). The Italian Genome Program. *Genomics* 7 (June): 294-297.

30. Bayev, AA. (1989). *The Human Genome, A General Overview (Genom Cheloveka, Obshchii Uzgliud)* . Scientific Council for the State Scientific-Technical Program "Human Genome," Moscow.

31. Alwen, J. (1990). United Kingdom Genome Mapping Project: Background, Development, Components, Coordination and Management, and International Links of the Project. *Genomics* 6 (January): 386-388.

32. Ferguson-Smith, MA. (1991). European Approach to the Human Gene Project. *FASEB Journal* 5 (January): 61-65.

33. Brenner, S. (1986). *Map of Man* (First draft communique to Bronwen Loder). Commission of the European Communities. 10 February 1986.

34. Commission of the European Communities. (1988). *Proposal for a Council Decision Adopting a Specific Research Programme in the Field of Health: Predictive Medicine: Human Genome Analysis (1989-1991)* (COM(88) 424 final — SYN 146). Commission of the European Communities. 20 July.

35. Commission of the European Communities. (1989). *Modified Proposal for a Council Decision, Adopting a Specific Research and Technological Development Programme in the Field of Health — Human Genome Analysis: (1990-1991)* (COM(89) final — SYN 146). Commission on the European Communities. 13 November.

36. McLaren, DJ. (1991). *Human Genome Research: A Review of European and International Contributions* . Medical Research Council, United Kingdom. January.

37. McLaren, DJ. (1991). The Human Genome — UK and International Research Initiatives. *MRC Annual Report, April 1990-March 1991*: 44-50.

38. Jordan, BR. (1991). The French Human Genome Program. *Genomics* 9 (March): 562-563.

39. Spurgeon, D. (1992). Canada Commits Money for Human Genome Research. *Nature* 357 (11 June): 428.

40. European Science Foundation. (1991). *Report on Genome Research 1991* . European Science Foundation, Strasbourg.

41. Allende, JE. (1991). A View from the South. *FASEB Journal* 5 (January): 6-7.

42. Allende, J. (1988). *Background on the Human Genome Project* . Red Latinoamericana de Ciencias Biologicas. 28 June-1 July.

43. Dausset, J, Cann, H, Cohen, D, Lathrop, M, Lalouel, J-M, and White, RL. (1990). Centre d'Etude du Polymorphisme Humain (CEPH): Collaborative Genetic Mapping of the Human Genome. *Genomics* 6 (March): 575-577.

The following two references were added in revision:

1. Gamow, G. (1954). Possible Relation between Deoxyribonucleic Acid and Protein Structures. *Nature* 173 (13 February): 318. I thank Maynard Olson for pointing out this reference.

2. King, M-C, and Wilson, AC. (1975). Evolution at Two Levels in Humans and Chimpanzees. *Science* 188 (11 April): 107-116.