Dana Landress

The Nomenclature Precedent: Mapping Origins and Scientific Standards Creation in the Human Genome Project

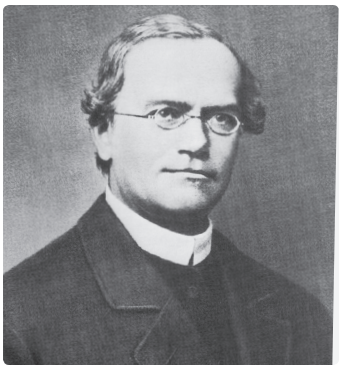
Science cannot be contained. In the age of mass information, news of scientific innovation is rapidly translated from laboratories to academic publications to newspapers, especially when scientific breakthroughs offer new information about diseases and improved quality of life. Perhaps the most publicized scientific endeavor of the past 50 years has been the Human Genome Project (HGP), a national effort to locate and map every gene in the human body. For more than 50 years, the HGP was discussed and worked on in laboratories throughout the United States. The project was not a strictly scientific endeavor, however, but was integrated into congressional legislation, city council meetings, courtrooms, and pharmaceutical board meetings, where non-scientific communities responded to the HGP in disparate ways. As a result, the innovations and implementations of genetic research in the project were articulated and managed by people both inside and outside of the laboratory.

The HGP created a dynamic network of collaborations that involved not just scientists, but ethicists, community members, activists, and computer

scientists.[1] While the mid-century intellectual origins of the project revealed language to be an important organizational tool for gene mapping, the gene nomenclature system was as much imposed by demands from communities and courtrooms as it was implemented by scientists. Although the history of the HGP has often been told as a set of events and discoveries, it can also be articulated as a narrative about people. When the history of the HGP is re-imagined in this way, it becomes clear that the project had informal mid-century origins long before its formal establishment in 1990. Beginning with a set of lectures given by Erwin Schrödinger in 1943, the first section of this paper examines the importance of language as a periodization tool to explain the longer intellectual history of the HGP. The second section then traces the importance of language when non-scientific professionals and community members engaged with the large-scale implications of the HGP. The following is a story about language and about people—academics and ordinary citizens, geneticists and jurors—who influenced the organizational underpinnings of a deeply humanistic and scientific endeavor: a genetic map of the human body.

The HGP formally began in 1990, when the US Congress jointly funded the Department of Energy (DOE) and the National Institutes of Health (NIH) to map out the human genetic sequence. These institutions had different goals in mind when designing the HGP. While the DOE was interested in the project as an opportunity to study the long-term genetic effects of nuclear radiation, the NIH wanted to maintain its reputation as a leader of American scientific research. The HGP generated excitement within the scientific community. Indeed, it was completed in only 11 years when project architects anticipated a 15-year period to completion. The fact that the human genome was mapped nearly four years before its intended completion date is indicative of the large-scale intellectual, financial, technological, and cultural resources afforded the HGP. By 2001, the DOE and the NIH alone had

1 Also known as the material semiotic method, sociologist Bruno Latour's Actor Network Theory is a useful way to conceptualize the relationship between material objects and conceptual ideas in the Human Genome Project. The idea of a human gene map had many abstract qualities, including representing human disease cartographically and storing this information in a cyber database. There were also concerns over patients rights and consent. In response to these concerns, individual and institutional actors cooperated in networks to map the human genome. For further information on Actor Network Theory see Bruno Latour, *Reassembling the Social: An Introduction to Actor-Network-Theory* (Oxford: Oxford University Press, 2005).



Literature on the history of genetics often begins with Gregor Mendel's (pictured) pea plants or with the formal discovery of DNA's helical structure by either Rosalind Franklin or James Watson and Francis Crick in the 1950s. (*Photo courtesy of Wikimedia.*)

invested approximately 2.7 billion dollars in the project.[2]

A considerable amount of funding had been allocated to the development of new sequencing and database technologies to increase mapping efficiency and to reduce the cost of research. Still, scientists were inundated with a copious amount of information. Mass data yielded from the project raised questions about the storage, access, and ownership of genetic research. To address these concerns, scientists began to contemplate a coherent system of classification for the mapped genomic regions, which also incorporated standardized molecular nomenclature. Such a system was thought to benefit the HGP in several ways: communication standards could facilitate information exchange between laboratories, increase project efficiency and organization, and offer a solution for sharing the project's findings to non-scientific communities. The story of the nomenclature system is, in fact, a story of how scientists and non-scientists responded to both organizational and ethical questions raised throughout the course of the HGP.

Historians disagree about both the formal and informal origins of the Human Genome Project. Literature on the history of genetics often begins with one of two origin points: in the nineteenth century with Gregor Mendel's pea plants or with the formal discovery of DNA's helical structure by either Rosalind Franklin or James Watson and Francis Crick in the 1950s. Still, other scholarship charts the project's formal commencement later, when institutional sponsorship and early HGP mapping efforts began in the late 1970s. Historiography on the HGP tends to periodize its progress around major events and findings, such as the preliminary mapping meetings held at Cold Spring Harbor Laboratory in the late 1970s. However, a study of the project's intellectual momentum suggests that

2 "International Consortium Completes Human Genome Project," National Human Genome Research Institute, April 14, 2003, News Release Archives, National Institutes of Health.

scholars had an established vision of a human gene map several decades prior to the institutional coalescence of the HGP in 1990. As intellectual momentum gave way to project implementation, the organizers of the HGP realized that different laboratories utilized different organizational systems for mapping genes. The confusion created by these discrepancies led the proponents of the HGP to advocate for a standardized language system that would facilitate communication and consistency among dozens of laboratories.[3] In addition to the project's chronological or event-based history, the HGP's intellectual history can be traced through individual proponents of a language system for gene mapping. In order to understand the lengthy process of standardized language creation in the HGP, it is necessary to first turn to the intellectual origins of the project.

The Scientific Community and HGP Origins

At the end of the nineteenth century, Austrian abbot Gregor Mendel pioneered the study of genetics in plants. By the mid-twentieth century, the field of molecular biology was increasingly concerned with human, rather than plant, genes. The study of human genetics interested biologists because genes revealed information about the inheritance of human traits. The medical community was especially interested in creating a gene map as a means to better understand human propensity for disease. Collaboration between scientific and medical research communities promised new insights into gene expression and the likelihood of inheriting diseases such as cancer, cystic fibrosis, sickle cell anemia, and Huntington's disease. In order to standardize either scientific or medical understandings of the human gene, however, some sort of unified vision of what was being discussed would be required.

One of the earliest descriptions of a human gene map was set forth by Austrian physicist Erwin Schrödinger in a series of published lectures

3 Prior to the development of a single nomenclature system, geneticists had several options for naming genes. Gene names could be based on primary function, location on the chromosome, numerical assignment, sequence tag site (STS), known genetic trait, disease associated with the gene, enzyme associated with the gene, etc. For further reading, see M.D. Zorn and C.R. Cantor, "Nomenclature Issues in the Human Genome Project," in *The Terminology of Biotechnology: A Multidisciplinary Problem* (University of California, Berkeley, 1990).

at Trinity College Dublin in 1943.[4] According to Schrödinger, the human cell, like other biological mechanisms, required order. Order, he stated, was governed by physical laws that possessed a high degree of accuracy. Interactions between separate but related mechanisms were brought into harmony through order, from the microscopic order of cellular processes to the larger whole being governed by the human body. Whether microscopic or systemic, physical interactions of the body depended upon a distinct order with identifiable patterns, Schrödinger posited.

Schrödinger was particularly concerned with the patterns that governed genetic structures. He understood that even the nearly invisible interactions of the quanta (the smallest amount of energy required for a physical interaction) informed and even defined visible biological interactions. Speaking specifically about the visibility of cellular material, Schrödinger noted, "Within every group [of genes] a linear map can be drawn up which accounts quantitatively for the degree of linkages between any two of that group, so that there is little doubt that they actually are located, and located along a line, as the rod-like shape of the chromosome suggests."[5] He was certain that it was possible to construct a linear map to organize and catalog a genetic "code-script," and he aimed to study the relationships of these microscopic life-giving entities. For this, Schrödinger also had a solution: mapping.

Schrödinger did not live to see the human gene map to completion, but he was convinced that understanding genetic order was critical to its materialization. Although applied physics dominated early twentieth-century research, Schrödinger believed that "the second half of the century will belong to molecular biology and genetics." Schrödinger insisted that scientists had "reached a point of dramatic change in our views of life and ourselves," and that "great discoveries [are] imminent." He concluded by writing that the implications of genetic research "will change our culture."[6] Schrödinger's vision not only advanced intellectual discourse regarding

4 Erwin Schrödinger, *What Is Life: The Physical Aspect of the Living Cell; with, Mind and Matter; & Autobiographical Sketches* (Cambridge; New York: Cambridge University Press, 2012).

5 Erwin Schrödinger, "What Is Life? The Physical Aspect of the Living Cell," *The American Naturalist, University of Chicago Press for The American Society of Naturalists* 79, no. 785 (1945): 554.

6 Johann Götschl, *Erwin Schrödinger's World View: The Dynamics of Knowledge and Reality*, Theory and Decision Library, vol. 16, Dordrecht (Boston: Kluwer Academic, 1992).

Vannevar Bush, who predicted the shift from wartime applied science to postwar basic research. Bush was an early advocate for genetics research in the second half of the twentieth century. (*Photo courtesy of Wikimedia.*)

the nature and structure of genes, it also brought the concept of organized genomic mapping out of the realm of individual imagination to that of legitimate scientific possibility.

Despite Schrödinger's prescience, the advancement of scientific understandings of genetic structure, function, and even location was modest during the 1940s. Scientific instrumentation remained insufficient for genetic visibility, and funding for the biological sciences was significantly diminished during wartime. However, this changed following World War II, when American scientific communities at large transitioned from a focus on applied science (science developed to enhance specific wartime projects, such as weapons manufacturing) to an emphasis on basic research (research that focused on creating or advancing new scientific knowledge). In July 1945, Vannevar Bush, Director of the United States Office of Scientific Research and Development, paved the way for this with his publication of *Science: The Endless Frontier*, a manifesto about the future of science written at the request of President Roosevelt. Bush insisted that basic research would be the "new frontier" of the scientific community.[7] In the report, Bush suggested that science should support new

7 Vannevar Bush, *Science—the Endless Frontier: A Report to the President on a Program for Postwar Scientific Research* (Washington, D.C.: National Science Foundation, reprint, 1980), Section 1. Especially relevant to basic scientific research is a set of norms introduced in 1973 by sociologist Robert K. Merton in his work *The Sociology of Science: Theoretical and Empirical Investigations.* This landmark publication outlined ethical principles to guide modern scientific research.

efforts to understand human disease and infection, "yet we find that the traditional sources of support for medical research... are diminishing and there is no immediate prospect of a change in this trend."[12] Bush called for a move to support basic research, but he understood that most academic institutions could not independently support the large-scale research he had in mind. The genetics community benefitted from Bush's proposed basic research funding because the field of genetics coalesced medicine and biology around disease prevention, a topic that Bush believed was crucial to the future of American scientific research. Indeed, Bush's letter to Franklin D. Roosevelt did much to sway federal support in favor of the enterprise of biological research.

Though basic science was gaining prominence, it would still take time before the vision of a genomic map would have any concrete reality. Prior to gene mapping, geneticists focused on mapping the structure of chromosomes, which bound together smaller units of genetic information. Twelve years after Bush's report, Albert Levan and Joe Hin Tjio, geneticists at the University of Lund, identified and published the 46 human chromosomal structures. The findings of Levan and Tjio directly contradicted the prevailing belief that there were 48 human chromosomes. Prior to their publication, chromosomal research had been widely suspended, as chromosome number and structure were an essential basis for continued research. Whereas Schrödinger's writings described the *mapping* of genomic structures as a viable scientific pursuit, Levan and Tjio defined chromosomes—the building blocks of human genetics—as entities that were open to further investigation. The process of peering into chromosomal structures that contained genetic material enhanced the depth and scale of microbiology, providing the possibility to discern the nature, structure, and functions of genetic material contained within each chromosome. Once the number of chromosomes had been accurately determined, the boundaries of genetic study could be defined both spatially

and visually, turning Schrodinger's vision into reality.[8]

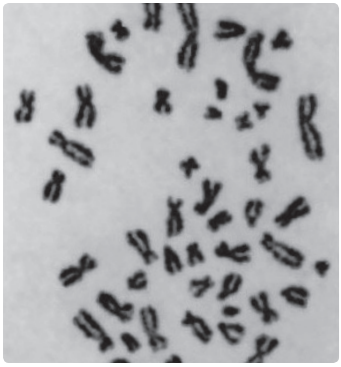
When Tjio and Levan discovered the structure and number of human chromosomes, their findings confirmed the viability of genetic mapping. During this revelatory moment, the scientific community turned to the question of language and nomenclature, which would be used to discuss a potential map of human chromosomes. By the mid-twentieth century, better microscopic lens resolution had led to more precise chromosome photographs, and geneticists speculated that genes on the chromosome might soon be studied in greater detail. If images of the chromosomes were to be translated into a map of chromosomes, the information had to be intelligible and navigable by scientists in different labs across the United States and around the world. Simply being able to see the same image was not enough.

Although Levan and Tjio's discovery made a common nomenclature necessary, American scientists had already discussed the possibility of a standardized nomenclature system. In 1953, three years prior to Levan and Tjio's publication, the NIH received a copy of a 76-page report titled, "A Proposal for Uniform Nomenclature in Bacterial Genetics."[9] A team of international geneticists had developed the proposal from a paper published at Brookhaven National Laboratory and sponsored by the US Atomic Energy Commission. It recommended a basic system for naming, referencing, and cataloging genetic information, thereby defending the importance of a nomenclature system from a number of perspectives.[10] First, it suggested that such a system was convenient and pragmatic since individual research groups would not need to devise their own classification system. Second, the proposal argued that a standardized nomenclature system would

8 H. Tjio and A. Levan, "The Chromosomes of Man," *Hereditas* (1956), 42. For a detailed account, see: M.A. Hulten, "Numbers, Bands and Recombination of Human Chromosomes: Historical Anecdotes from a Swedish Student," *Cytogenetic and Genome Research* 96, no. 1-4 (2002): 14-19; In their *Hereditas* publication, Tjio and Levan mention a study of chromosomes in embryonic liver mitosis conducted by Dr. Eva Hansen, who halted her research because her team was only able to locate forty-six of the presumed forty-eight human chromosomes. For further information, see: Daniel L. Hartl, *Essential Genetics: A Genomics Perspective*. (Burlington, MA: Jones & Bartlett Publishers, 2009); Albert Levan, "Chromosome Studies on Some Human Tumors and Tissues of Normal Origin, Grown in Vivo and in Vitro at the Sloan–Kettering Institute." *Cancer* 9 (1956): 648–663; Andrew J. Hogan, "The 'Morbid Anatomy' of the Human Genome: Tracing the Observational and Representational Approaches of Postwar Genetics and Biomedicine. The William Bynum Prize Essay," *Medical History* 58, no. 3 (2014).

9 M. Demerec, E.A. Adelberg, A.J. Clark, and Philip E. Hartman, "A Proposal for Uniform Nomenclature in Bacterial Genetics," *Cytogenetics and Cell Genetics* 54 (1966): 61-76.

10 Ibid.



A low resolution karyotype of forty-six human chromosomes discovered by Tijo and Levan in 1956. At that time, chromosome bands were not visible and a lack of organization made a nomenclature classification system difficult. (*Image reprinted with permission from BioMed Central, acquisitioned from Hereditas.*)

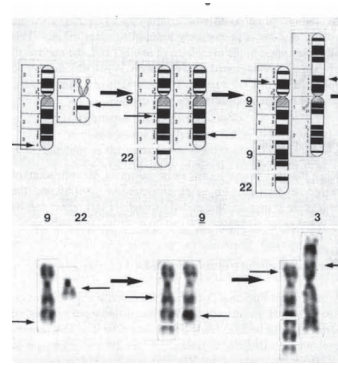
facilitate understanding and communication among scientists. Finally, it argued that the proposed nomenclature was beneficial because it was malleable enough to incorporate new genetic information. The authors affirmed that their "aims" were "uniformity, a unique designation for each strain, convenience for typing, editing, printing, record-keeping, and information retrieval." Furthermore, they argued that a standardized nomenclature system would provide "adaptability, simplicity, and clarity" as well as "comprehension by workers in all areas of biology." All of this would lead to "adaptability to new developments in the foreseeable future."[11] The report concluded with an example of the system, a list of proposed standard symbols based on known gene function. Despite this well-organized proposal, seven years passed before the nomenclature question was again discussed in mapping meetings.[12]

Even so, the 1960s were a decade defined by advancements in the field of genetics, when much of the groundwork was laid for the creation of a genetic mapping project. A number of discoveries in microbiology further expanded opportunities for new research about the structure, function, and utilities of Deoxyribonucleic Acid (DNA). In 1960, just seven years after Rosalind Franklin, James Watson, and Francis Crick discovered the binary helical structure of DNA, Sydney Brenner, Francis Crick, François Jacob, and Jacques Monod solved a problem that had eluded geneticists for nearly a decade. Their research team, sponsored by the California Institute of Technology, had discovered Messenger Ribosomalnucleaic Acid (mRNA), a set of molecules that are responsible for transferring genetic information

11 Ibid.

12 There is no clear answer as to why so much time would pass before any action was taken. One suggestion is that no single sponsoring institution or committee was assigned to undertake this project until 1974, when efforts to produce a human gene map were already well underway.

to the cytoplasm, where genetic information is expressed.[13] The next year, 34-year-old Marshall Nirenberg discovered that the genetic code was comprised of chemical units of DNA that specify how protein molecules are constructed. By 1966, Dr. Nirenberg had identified the first 63 sequences of human DNA.[14] Two years later, in 1968, Nirenberg and his colleagues were awarded the Nobel Prize in Physiology and Medicine for "their interpretation of the genetic code and its function in protein synthesis."[15] Collectively, the discoveries of the 1960s prompted new opportunities for discovering and assembling a map of the human genome. As individual research groups, laboratories, universities, and private institutions independently sponsored continued genetic research, the issues of intelligibility and nomenclature norms were still part of an unfinished discussion among geneticists.



A higher-resolution photograph of chromosomes 3, 9, and 22, published by the Standing Committee on Cytogenetic Nomenclature in Stockholm in 1978. Note the visible difference in organization in contrast to the original 1956 karyotype on the previous page. (*Image reprinted with permissions from S. Karger AG, Medical and Scientific Publishers.*)

In 1960, four years after Tjio and Levan's discovery prompted a discussion of a genetic nomenclature, British cytogeneticist Charles E. Ford convened a meeting on chromosomal nomenclature—a related, though not identical, subfield of microbiology. Later referred to as the Denver Conference for its location, the meeting was attended by 14 lead scientific investigators, all of whom had previously published human karyotypes (chromosome images). The group decided to number the chromosomes they had been researching and pair corresponding sets, to demarcate the sex chromosomes with "X" or "Y," and to further categorize all chromosomes by size groups.[16] The nomenclature system developed at the Denver Conference remained in use for decades following its

13 F.H. Crick, L. Barnett, S. Brenner, R.J. Watts-Tobin, "General nature of the genetic code for proteins," *Nature* 192, no. 4809 (1961): 1227–32.

14 Nicholas Wade, "Marshall Nirenberg, Biologist Who Untangles Genetic Code, Dies at 82." *New York Times*, January 21, 2010.

15 Franklin H. Portugal, *The Least Likely Man: Marshall Nirenberg and the Discovery of the Genetic Code* (Cambridge, Massachusetts: The MIT Press, 2015).

16 For further reading on the X and Y chromosomes, see Sarah Richardson, *Sex Itself: The Search for Male and Female in the Human Genome* (Chicago: University of Chicago Press, 2013).

establishment, having emerged in response to rapid breakthroughs in genetic research that created a need for a common scientific and medical vocabulary to further communication and organization in cytogenetics.[17]

The methodology devised at the Denver Conference for mapping chromosomes was highly intelligible, as reported in *The Lancet*:

> "In designating a particular band, four items [were] required: (1) the chromosome number, (2) the arm symbol (i.e., the short or long arm of the chromosome), (3) the region number, and (4) the band number within that region. These items [were] given in order without spacing or punctuation. For example, 1p33 indicated chromosome 1, short arm, region 3, band 3."[18]

Cytogenetic chromosome maps became progressively more detailed with higher resolution photographic images and added their discoveries onto this framework. Initial depictions of chromosomes evolved into regions and bands, which served as landmarks for genetic diseases. Increased photographic resolution and visibility led to more complex mapping spaces, but the standardized nomenclature established at the Denver Conference facilitated chromosome mapping as a precursor to gene mapping. With a clear vocabulary and a consistent methodology for identifying and categorizing new information, human chromosome research prospered. Following the Denver Conference's success in establishing a standardized nomenclature for chromosome mapping, geneticists later advocated for a similar organizational system for gene mapping.

Held in Rotterdam, the 1974 Human Gene Mapping Conference was the first formally recognized collective call for a nomenclature system for genes. Although participants established no formal guidelines at the meeting, a committee was formed to discuss the possibility of standardized terminology, as had been devised for chromosome mapping at the Denver Conference. The advantages were clear: increased organization in mapping, intelligibility and access to information for all scientists, and a decreased

17 Standing Committee on Human Cytogenetic Nomenclature, *An International System for Human Cytogenetic Nomenclature: Report of the Standing Committee on Human Cytogenetic Nomenclature,* Hässelby Castle, Stockholm, Sweden, September 4-9, 1977.

18 J. Lejeune, et al., The Denver Conference Proceedings, "A Proposed Standard System of Nomenclature of Human Mitotic Chromosomes," *The Lancet,* vol. 275, no. 7133 (1978): 1063-1065.

The Tower of Babel by Dutch painter Pieter Brueghel the Elder (1563). Scientists often referenced this myth, which claims that language variation was a divine intervention to thwart humanity's hubris. (*Image courtesy of Wikimedia.)*

likelihood that geneticists in different labs would replicate research. The committee in charge of creating a nomenclature system consisted of Dr. Harry Harris from the University of Pennsylvania School of Medicine, Dr. Meera Khan of the Netherlands Department of Human Genetics, Tom Shows, microbiologist and editor of *Cytogenetics and Cell Genetics*, and Dr. Victor McKusick of Johns Hopkins School of Medicine. Deemed the Committee on Terminology, the group determined that "guidelines need[ed] to be established for naming the human genetic markers, including the terms to be used for loci, genes, phenotypes, and polypeptide chains."[19] The Committee met again the following year, 1975, to discuss further the possibility of a nomenclature system, although a draft was not presented until the Human Gene Mapping Conference in Edinburgh in 1979.

Concerns about nomenclature in the Human Genome Project represented larger intellectual quandaries about the utility and limitations of language structures as an organizational system for processing scientific phenomena. Debates over gene nomenclature often referenced the biblical tower of Babel, highlighting the importance of devising a singular language

19 "Report of the Committee on the Genetic Constitution of the X Chromosome," from the International Workshop on Human Gene Mapping, *Cytogenetics and Cell Genetics* vol. 14, no. 3 (Rotterdam, Netherlands, 1974): 190-195.

system when dozens of gene mapping languages were already in use. The question of which institution or lab had the authority to impose one language system over another loomed large. However, unlike the Denver Conference, where a common language system for chromosome mapping was devised within the scientific community, the standards for gene mapping were imposed from the outside, and not by scientists alone. In the five years that passed between 1974 and 1979, several events had already shaped the future prospects of the project, revealing the complexity of mapping without a single system of nomenclature and underscoring the centrality of public engagement in scientific standards creation in the early years of the HGP.

Public Influence on Nomenclature Standards

The year of 1977 was critical in advancing the prospect of a Human Genome Project. The previous June, researchers at Harvard and the Massachusetts Institute of Technology had developed a technique known as gene splicing. An early predecessor to recombinant DNA techniques, gene splicing allowed geneticists to insert a gene sequence into a pre-existing sequence and then replicate this new segment widely.[20]

This new technique did not stay in the laboratory long. By 1977, gene splicing caused controversy in both the media and scientific circles. *The New York Times* reported that Boston residents feared "that new, particularly durable viruses could escape from a laboratory." Bostonians also expressed "commercial concerns" because the private universities were "not subject to the Government regulations that control gene splicing research at federally financed universities or hospitals."[21] In response to what later became known as the "Cambridge Gene Scare," the Cambridge Public Health Department and Cambridge Town Council issued a set of ordinances regulating activities related to human genetic research. In the months prior to this regulation, the town council had threatened a moratorium on gene splicing research within the city.[22] The moratorium was implemented after a Harvard newspaper published an article on genetic research titled,

20 One of the earliest applications of gene splicing technology was a method to replicate human insulin.

21 "Gene-Splicing Concern in Boston," *The New York Times,* May 31, 1981.

22 Nicholas Wade, "Gene-splicing: Cambridge Citizens OK Research but Want More Safety," *Science* vol. 195, no. 4275 (1977): 268–69.

"Gene Splicing Controversy: Visions of Great Benefits and Grave Perils."[23] Prior to this publication, which was highly circulated among the public, citizens were unaware of Harvard's effort to study gene splicing and thus were ill-informed about the real or imagined dangers of such work. Since little was known about gene splicing and recombinant DNA efforts, the article produced legitimate concerns that "disease-producing bacteria like *streptococci* could, as a result of genetic engineering, accidentally be made immune to antibiotics and other drugs used to treat them," or that "a bacterium that now inhabits the human body without doing harm might receive a genetic transplant that would cause it to begin manufacturing a deadly toxin."[24] As concerns circulated among the public, elected officials threatened a two-year moratorium on all genetic research in Cambridge. The prospect of discontinued research in the midst of exciting genetic breakthroughs further encouraged geneticists to establish laboratory research standards.

The February 1977 issue of *Science* outlined the various local protocols developed in response to growing concerns over gene splicing technology. In a matter of months, Massachusetts, New York, California, Michigan, New Jersey, and Wisconsin had each issued protocols for gene splicing practices in regional laboratories. The New York attorney general's environmental health bureau proposed a bill stipulating that all scientists engaged in gene splicing research had to be certified, trained, and monitored by the state health commissioner.[25] Just one month earlier, in January of 1977, the NIH had revised their guidelines for recombinant DNA research.[26] The updated guidelines, later given federal authority as national legislation, included regulations on genetic splicing and replication.[27]

After the public outcry, the NIH authorized cohesive standards for gene splicing, which solved the temporary concerns over recombinant DNA

23 Sandra Stencel, "Controversy over Gene Splicing: Visions of Great Benefits and Great Perils," *Genetic Research*, (Cambridge: Cambridge University Press, 1977).

24 Ibid.

25 Nicholas Wade, "Gene-splicing: At Grass-roots Level a Hundred Flowers Bloom," *Science* 195, no. 4278 (1977): 558–60.

26 0ffice of the Director, NIH, "National Institutes of Health Environmental Impact Statement on NIH Guidelines for Research Involving Recombinant DNA Molecules, Part One and Part Two" (Bethesda: NIH Press, 1977), 147-438.

27 For a comprehensive history of Recombinant DNA Guidelines issued by the NIH, see D.S. Fredrickson, "A History of the Recombinant DNA Guidelines in the United States," Department of Health in Collaboration with the National Institutes of Health, (Bethesda: NIH Press, 2005).

research, but did not address the larger question of nomenclature for gene identification and mapping, which had now gained broad public attention. The gene splicing ordinances issued by the NIH involved the consolidation of many local practices for gene splicing. The Cambridge Gene Scare also roused the attention of several national organizations, including the Occupational Safety and Health Administration, the Environmental Protection Agency, and the Food and Drug Administration. The creation of a single standardized system of uniform guidelines about gene splicing safety occurred only after public concern prompted changes in scientific regulation of genetic research, despite the fact that scientists had been discussing the practicality of creating some sort of guidelines for this research for years. Following the Cambridge Gene Scare, the benefits of standardization were increasingly evident: standardized research protocols would allow geneticists from various labs to clearly articulate their research findings, and standardized nomenclature would help scientists explain these findings in order to promote greater public understanding. Moreover, a standard nomenclature would prevent uninvited public distrust about new genetic research.

Nomenclature and Mass Information Management

Influenced by public debate and discussion within the scientific community, the nomenclature question also was tied to a number of practical concerns about information management related to the HGP. Many prominent institutions—including the NIH, the Department of Energy, and the National Science Foundation—considered sponsoring the standardization project, yet it was unclear which institution would be given primary responsibility for creating an intelligible nomenclature system. The continued use of multiple language systems in genetics threatened both clear communication among research sites and mapping efficiency, slowing down research and diluting its potential impact.

One way the HGP increased efficiency was through multi-site research, whereby each lab was assigned a specific portion of the human genome to map and catalog. Inter-institutional collaboration became especially crucial as continued discoveries of genetic regions in the 1980s produced massive amounts of information about the location of nucleotide sequences within genes. While new computer database technologies were implemented

to catalogue and map genetic data, the challenges of mass information management remained unresolved. The sheer quantity of information that had to be classified and named made this research challenging. Whereas the Denver naming system classified 46 chromosomes, each marked with bands and subregions, gene mapping produced much greater quantities of information: approximately 3.2 billion base pairs needed to be mapped.[28] As Vannevar Bush had so aptly anticipated, genetics appeared to be an endless frontier of new information. Debates ensued over how to best organize genetic data. Many believed that the search for information should be localized and comprehensive rather than general and expansive.[29] However, the approach most common to those who worked on the HGP was to first collect as many sequences as possible and presume that functional information about disease, inheritance, and mutation of these sequences would be added at a later date. This method prioritized information collection and was aligned with the HGP's larger goals of mapping efficiency and timely project completion.

In 1980, approximately a decade before the formal commencement of the HGP, American advocates of the project began to voice their concerns about nomenclature and data organization, adding to debates previously held about genetic nomenclature more broadly. Proponents of a standardized nomenclature system, such as Dr. Donald Lindberg, director of the US National Library of Medicine, argued that a nomenclature system would significantly reduce the time and resources required to complete the human gene map.[30] In 1984, Yale Professor and HGP architect Frank Ruddle insisted that gene mapping could not progress efficiently without a standardized mapping nomenclature.[31] In a 1989 article, key HGP advocates argued that a consistent and comprehensive language system for genome mapping needed to replace the multitude of different mapping linguistics

28 "Inside Life Science: Genetics by the Numbers," *NIH News in Health* (Bethesda: NIH Press, 1977).

29 H. Wain, J. White, and S. Povey, "The Changing Challenges of Nomenclature," *Cytogenetics and Cell Genetics* vol. 86, no. 2 (1999): 162-164.

30 Donald Lindberg, *The Growth of Medical Information Systems in the United States* (Lexington: Lexington Books Press, 1979), 194.

31 R.L. Miller, C. Partridge, W. Kidd, F. H. Ruddle, "The Yale Human Gene Mapping Library," *Cytogenetic Genome* vol. 37, no. 4 (1984): 394-397. The first genome-related methodology patent was granted in 1980 to Stanley Norman Cohen and Herbert Boyer for cloning the gene that codes for insulin. The licensing royalties for this patent exceeded 300 million dollars. See Gerald Karp, *Cell and Molecular Biology: Concepts and Experiments* (Wiley, 2009), 976-977.

already in place at partner research sites. Concerned with the lack of a common language system, central HGP architects admonished: "Lest we replay the failed effort to build the tower of Babel, it would be wise to move decisively toward adoption."[32] In the absence of a standardized language system, geneticists had devised nomenclature systems specific to each research site, which often conformed to the organizational practices of each individual laboratory instead of adhering to a set of standards established to maintain consistency among the HGP's varied research sites.

Furthermore, concerns over efficiency were also related to the diversion of significant funding toward the genome project during an "unprecedented" scientific funding crisis.[33] The HGP was in its planning stages during the global economic recession of the 1970s and 1980s. Efficient research offered the potential for economic stimulation in the fields of science, technology, and healthcare. However, the adoption of an organizational nomenclature system, though advocated by prominent geneticists, appeared secondary to concerns over funding. While a pre-determined nomenclature system might have increased the time to HGP completion, such a system might also have circumvented the role of courts in genetic standards creation.

Legal Interpretations of Nomenclature

Prescient debates over the establishment of genetic nomenclature foreshadowed one of the most contentious ethical dilemmas of human genome research: the possibility that human DNA was patent-eligible. While the human body had long been central to debates about a gene map, the potential for profitable pharmaceutical and biomedical applications of genes invited entirely new questions about the commercialization of scientific research. As genetic nomenclature debates continued to materialize in the late 1970s, the Supreme Court heard a landmark case related to a patent application for a gene that produced insulin. Courtroom intervention demonstrated continued public and federal interest in

32 C. Cantor, D. Bostien, L. Hood, M. Olson, "A Common Language for Physical Mapping of the Human Genome," *Science* vol. 245, no. 4925 (September 29, 1989), 1434-1435.

33 National Center for Human Genome Research, Complaints and Criticisms File, Box BCD7, Human Genome Archives National Reference Center for Bioethics at Georgetown. For further information, see M.A. Fortun, *Mapping and Making Genes and Histories: The Genomics Project in the United States, 1980-1990,* Ph.D. Thesis, Harvard University, (Cambridge: Harvard University Press, 1993).

scientific research on human genetics and contributed to the discussion of systematization that had been taking place since the 1950s. The Supreme Court's decision in this case revolved around the unanswered question of nomenclature.

On July 22, 1997, the US Federal Court of Appeals convened to review a suit brought against major pharmaceutical player Eli Lilly by the University of California. The lawsuit, later hallmarked as *Lilly*, concerned two patent applications filed by the respective litigants for licensing rights to a recombinant DNA method used to produce human insulin. John Shine, an Australian geneticist working at the University of California, was the first to file for a methods patent in May of 1977. After review from the US Patent and Trademark Office, Shine and his team of geneticists were encouraged to re-submit the application with the addition of a clear written description of the recombinant method they had developed to produce "vertebrae or mammalian" insulin.[34] Since the adoption of the original United States Patent Act of 1793, written description had served as a required patent mechanism to ensure that an inventor could not extend the claims or benefits of the invention beyond its actual scope. It was also a standardized proof to ensure that those seeking a patent could intelligibly explain the method for arriving at the end product. According to the US Patent and Trademark Office, the written description clause for patent applications required that an applicant "show possession of the claimed invention" through a description of "the invention with all of its limitations." Such applications were to be written in "clear, concise, and exact terms as to enable any person skilled in the art ... to make and use the same invention."[35] While geneticists at the University of California revised their patent application for recombinant insulin, the scientist's inability to provide a sufficient written description of the structure and function of insulin evidenced a lack of common genetic language standards, such that the recombinant method could not be articulated and replicated by other geneticists. The nomenclature question, now intricately linked to concerns about scientific standards and patent applications, had suddenly unraveled.

34 Janice M. Mueller, "The Evolving Application of the Written Description Requirement to Biotechnological Inventions," *Berkeley Technology Law Journal* vol. 13, no. 2 (1998): 629.

35 Christopher M. Homan, "Is Lilly Written Description a Paper Tiger? A Comprehensive Assessment of the Impact of Eli Lilly and its Progeny in the Courts and PTO," *Albany Law Journal of Science and Technology*, vol. 17 (2007). See also Regents University of California v. Eli Lilly Company, 35 USC. 112. 2000

In the meantime, Eli Lilly and Co. had produced a highly profitable drug, Humulin, based on the recombinant techniques licensed by the University of California. From 1977 to 1980, Eli Lilly invested over 60 million dollars on manufacturing facilities that were equipped to produce and distribute Humulin on a rapid and systematic scale.[36] The company projected that, once Humulin reached the market, worldwide pharmaceutical sales would peak at 1.1 billion dollars. The patent request filed by Lilly for synthetic insulin was the catalyst intended to propel the pharmaceutical industry into a new economic frontier.[37]

On October 24, 1997, the Court ruled that Eli Lilly and Co. had not infringed upon either of the two patents held by the University of California for the recombinant plasmids utilized in the production of human insulin. Although geneticists at the university had filed multiple patent applications to license their technique, the Court rejected the validity of the patents based upon "a lack of adequate written description."[38] Furthermore, the Court ruled that the DNA patent for human insulin required "a precise definition, such as by structure, formula, or physical properties," and that "an adequate written description of DNA require[d] more than a mere statement that it [was] part of the invention." What was required, the Court ruled, "was a description of the DNA itself."[39]

The ruling had both immediate and long-term consequences. In the short term, bioengineering patent applications were held to a more rigorous standard, which required a demonstration of the invention, the method, the product, and an intelligible nomenclature system to present these components. The *Lilly* decision also indicated the sober reality of scientific

36 Gary L. Nelson, *Pharmaceutical Company Histories* (Bismarck: Woodbine Publishers, 1983). See also: E.J. Khan Jr. *All in a Century, The First 100 Years of Eli Lilly and Company* (Indianapolis: Eli Lilly and Company, 1976).

37 In the 1980 landmark case, *Diamond v Chakrabarty*, the Supreme Court adjudicated the question of whether genetically modified organisms, in this instance synthetic bacteria created to decompose crude oil sediments, were patent eligible. The court agreed to hear the case because the bacteria was deemed "an improvement on the constitution of matter" hearkening back to Locke's argument that products of nature could be converted into property through a process of laborious improvement. The 5-4 ruling famously declared that anything "under the sun that is made, or perhaps improved upon, by man" was patent eligible. In 1980, the same year of *Diamond v Chakrabarty*, Bayh-Dole legislation permitted federally funded research to qualify as patent eligible. An important implication of the Bayh-Dole legislation was that revenue could be collected by the recipients (i.e, a company, university, or private R&D firm) rather than returned to the federal government.

38 *The Regents of the University of California v. Eli Lilly and Company*, 119 US 43 (1997).

39 Ibid.

The first printout of the human genome, presented as a series of books. (*Photo courtesy of Wikimedia.*)

commercialization—that inventors were not always the beneficiaries of their work. Whereas some legal scholars admonished *Lilly* as "an unmitigated disaster ... with the potential for causing untold havoc in the biotechnology field," others viewed the interpretation of written description as "a new challenge to validity" and a mechanism to ensure "sufficient disclosure of an invention."[40] Now often referenced in biotechnology and legal studies, the modified Lilly Written Description (LWD) doctrine set forth a highly contested and technically formidable standard for subsequent biomedical innovations. The decision highlighted the importance of using an intelligible nomenclature system that could be understood and verified by different scientists.

Some of the earliest questions posed by Human Genome map architects about the creation of a system of organizational nomenclature were paralleled by the US Federal Court as jurors considered the relevance of written description. A description of scientists' method for retrieving recombinant insulin was insufficient without the DNA sequence description. The description needed to substantiate both genetic

40 Christopher M. Homan, "Is Lilly Written Description a Paper Tiger? A Comprehensive Assessment of the Impact of Eli Lilly and its Progeny in the Courts and PTO," *Albany Law Journal of Science and Technology*, vol. 17 (2007).

structure and function, with sequence-level precision, which Eli Lilly had successfully authored, despite the University of California having submitted a patent application several years prior to Lilly's appropriation of the technique. The Court's decision was not merely a vote for Eli Lilly and Co., it was also a vote for standardization and nomenclature. Excellent scientific research meant little if it could not be translated, interpreted, and applied consistently.[41]

The Nomenclature Debate's Long-Term Effects

Tracing a history of the development of the HGP from 1945 to 1977 reveals that language catalyzed early intellectual momentum among the scientists and policy advisors who first envisioned a human gene map. As technological advances yielded new scientific discoveries about the structure and function of genes, talk of institutional sponsorship and information storage translated into practical concerns about the feasibility of one of the largest scientific undertakings of the twentieth century. Amidst this flurry and excitement, geneticists proposed the establishment of a nomenclature system, much like the one that had been successfully implemented to map human chromosomes at the Denver Conference in 1960. However, between the 1974 Rotterdam HGP Meeting and the 1979 Edinburgh HGP Meeting, where geneticists first devised a standardized system of nomenclature for the project, several events significantly altered the course of the HGP and its relationship to technology, society, and law. The Cambridge Gene Scare proved the need for a standardized nomenclature system by revealing how poor communication standards between scientists and laypersons created public distrust, and how the desperate need for clear communication about genetic research had allowed the public to intervene in scientific standards creation. Once realized, the need for nomenclature produced a number of organizational questions related to the description, classification, storage, and efficiency of newly discovered genes. When the nomenclature question remained unresolved in 1977, the year Shine and his team of geneticists at

41 The following is a brief, but not comprehensive, list of cases that relied upon the new standards of Lilly Written Description: *University of Rochester v. G.D. Searle & Co.*, 358 US 916 (2004); *Enzo Biochem Inc. v. Gen Probe Inc.*, 323 US 956 (2002); *Tronzo v. Biomet, Inc.*, 156 US 1154 (1998); *Gentry Gallery Inc. v. Berkline Corp.*, 134 US 1473 (1998); *Festo Corp v. Shokestu Kinzoku Kogyo Kabushiki Co.*, 535 US 722 (2002); *J.E.M. Ag. Supply Inc. v. Pioneer* Hi-*Bred Intl.,* 534 US (2001).

the University of California filed for patent applications on human insulin, the imprecise presentation of their findings failed to pass the legal written description requirement, thereby allowing a federal court to determine the scientific standards necessary to obtain a patent.

The nomenclature question reveals a set of structural inconsistencies between the scientific community, the legal sphere, and the private sector. Legal scholars have suggested that the debates over intellectual property in the Human Genome Project transcended the structural legal framework in place to respond to emerging debates over biological property and ownership. The alteration of written description, from a method description to a comprehensive product description, set a precedent for future cases involving genetic research and ultimately reconfigured the boundaries and limitations of scientific discovery at the outset of the project. By understanding long-term debates over the use of nomenclature related to research on human genetics, historians can trace the development of the Human Genome Project over the course of almost 50 years after the end of the Second World War. Moreover, the involvement of lawyers, journalists, scientists, policy advisors, and community activists in the creation of a particular form of genetic nomenclature demonstrates the complex processes involved in scientific research in the twentieth century—a complexity that is unlikely to lessen in the years to come.