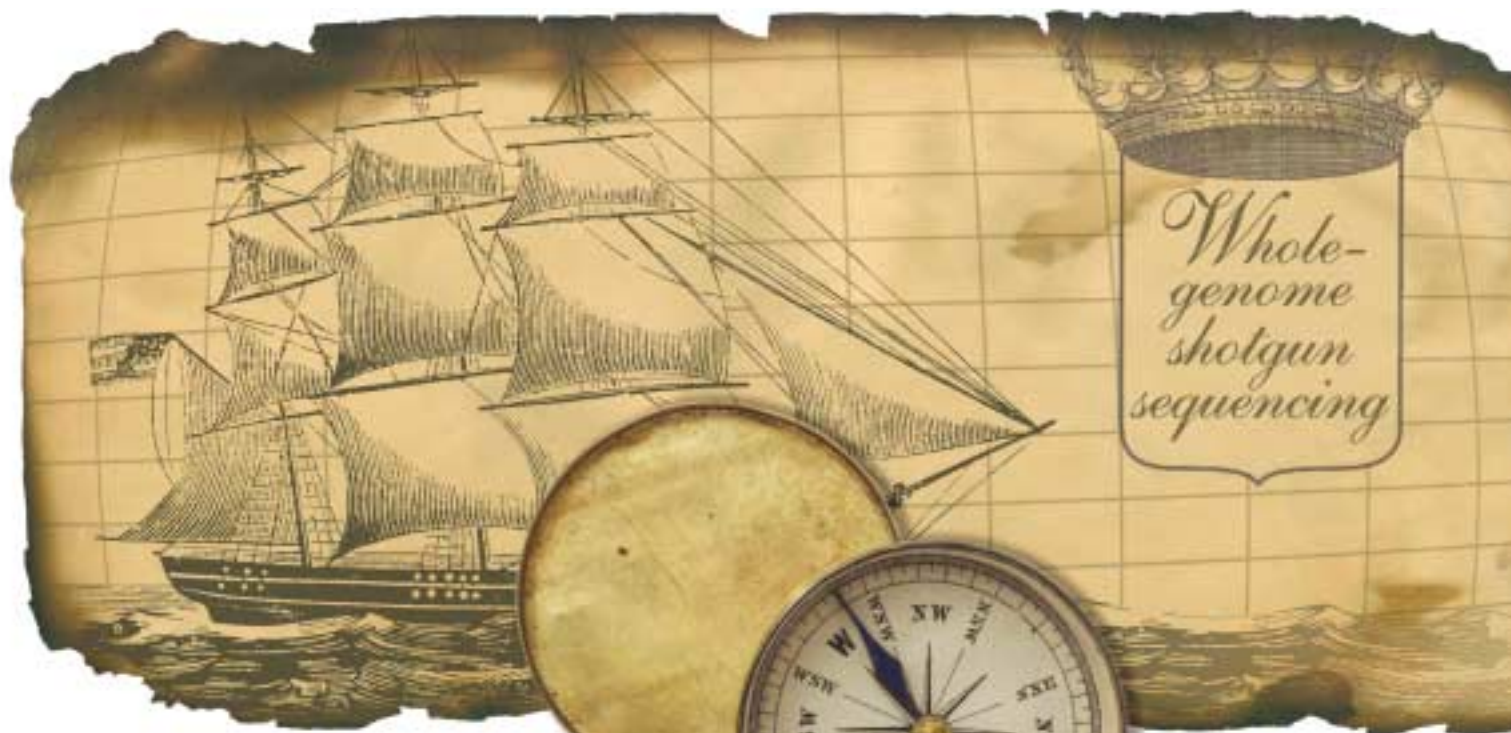






Charting the Map of Life

Explorers have traditionally set out from familiar ports toward the Unknown, creating maps of where they have been for others to follow. With time, these rough drawings are refined by those who come in their wake so that coastlines are charted, mountains measured, and rivers delineated to the mile. Today, scientists are grappling with a new map—that of the human genome—but unlike the one you can barely fold in your car, this map describes a landscape that remains almost a complete mystery: the 3.2 billion letters of DNA that code for the creation of a human being. The human genome is the total amount of DNA spooled into a set of 23 chromosomes found in the nucleus of every human cell. It's formed when the chromosomes of the male's sperm fuse with those of the female's egg to form a single-celled embryo, and it contains the genetic "instructions" that allow



the embryo to grow into a fully formed person.

When scientists announced they had mapped a rough draft of the human genome—identified 85–90% of the ordered sequence of DNA on each chromosome—the event was lauded around the world as the greatest scientific achievement since the Apollo 11 moon landing. Some scientists called the accomplishment nothing less than the beginning of a true understanding of biology. At a 27 June 2000 White House ceremony featuring the heads of the two teams responsible for the feat—J. Craig Venter, president and chief scientific officer of the Rockville, Maryland-based Celera Genomics, a privately funded biotechnology company, and Francis S. Collins, director of the publicly funded National Human Genome Research Institute—President Clinton said, “Today we are learning the language in which God created life.”

The map of the human genome points to a vast uncharted territory, much of it a wasteland. Only 3–5% of the genome—corresponding to between 30,000 and 100,000 functional genes—is thought to be biologically functional. The remainder is so-called junk DNA that may someday be shown to have biologic merit, but that for now is largely seen as filler that remains in the genome for unknown reasons. Scientists expect that mapping the genome will lead to a host of innovations in biology and research. For example, DNA microarrays, devices that analyze the level of expression

of thousands of genes at a time, could be used to accurately diagnose cancer and infectious disease subtypes and to predict clinical outcomes. Scientists will also use the genome to look at the interactions of the environment, genetic makeup, and toxic exposures, including the ability of certain genes to detoxify the body and promote disease resistance.

The genome will provide tremendous resources for understanding human diversity and evolution. All humans on the planet are roughly 99.9% genomically identical, not surprising considering a common ancestry thought to date back 150,000 years to a tiny band of people in Africa. Within the remaining 0.1% of the genome are the 3 million letters of DNA that govern our physical differences. Many portions of the human genome, particularly those coding for metabolic processes, are identical to those of other species. Comparative genomics studies will provide insight into how metabolic and other physiologic systems evolved in different species.

But despite the great potential of genomics, scientists caution that public expectations need to be tempered with reality. Decrying what she calls the “media inflation of genetic technology,” Lily Kay, an associate of the museum of comparative zoology at Harvard University in Cambridge and author of *Who Wrote the Book of Life? A History of the Genetic Code*,

says, “We are bombarded daily by media reports of the genetic revolution. And the usual approach is to absorb uncritically these scientific forecasts as *fait accompli*.” The fact is, she says, people are as much a product of their environment as they are of their genes. And to suggest that genetics is the sole determinant that defines us as individuals, writes Eric Lander, director of the Whitehead Center for Genome Research in Cambridge, in a 12 June 2000 editorial in *The New York Times*, “stretches the science far beyond the data.”

The Nuts and Bolts

Mapping the human genome is the most recent event in a genetic time line dating back to Gregor Mendel, who discovered the basic principles of heredity in the mid-1800s. Mendel introduced the concept of the gene as a unit of information through which hereditary information is passed from one generation to the next. Later, the concept became less abstract with the discovery that genes are made of the substance DNA. The three-dimensional structure of DNA and its method of replication was discovered by James Watson, an American postdoctoral fellow, and British graduate student Francis Crick at Cambridge University in England. In their classic paper titled “Molecular Structure of Nucleic Acids. A Structure for Deoxyribose Nucleic Acid,” published in *Nature* in 1953, the

two scientists described the double-helical form of the molecule, shaped like a twisted ladder, in which each rung is made up of four nucleotides: adenine, thymine, cytosine, and guanine (typically abbreviated as A, T, C, and G). The nucleotides are arranged in a series of base pairs, in which A bonds with T, and C with G. In the years that followed, researchers discovered that genes code for amino acid sequences, which themselves comprise the proteins that make life possible.

A genome map is essentially a representation of where genes are located on the chromosome. At the coarsest level of resolution are genetic linkage maps, which describe gene locations based on their patterns of inheritance, for instance as observed in mating experiments with *Drosophila melanogaster* (fruit fly). The first such map was developed in 1913 by Alfred Sturtevant while he was an undergraduate in zoology at New York's Columbia University. Working with embryologist Thomas Hunt Morgan in the legendary "fly room" at Columbia, Sturtevant arranged the ordered sequence of genes for eye color, wing shape, body size, and other characteristics based on their appearance in consecutive generations of *Drosophila*. Morgan himself was the first to associate a specific gene with a specific chromosome; again experimenting with fruit flies, Morgan found that the mutant gene for white eyes (most fruit flies have red eyes) is found only in male progeny and is located exclusively on the X chromosome—a discovery for which he won the Nobel Prize for Medicine in 1933.

Gene mapping (and indeed, the whole field of molecular biology) hit its stride in the late 1960s with the discovery that restriction enzymes could be used to cut DNA into specific sequences. In nature, restriction enzymes protect bacteria by slicing up invading viral DNA. But in the laboratory, they can be used as molecular "scissors" that recognize a highly specific DNA sequence, or type of sequence, and then cut the DNA at the same site in the sequence. Over 3,000 restriction enzymes have been identified to date, affording scientists great specificity when chopping DNA into isolated fragments. The fragments can be cloned (usually in bacteria) or duplicated using a variety of biochemical techniques to provide the unlimited genetic material needed for experimental studies.

These techniques have been used to generate higher-resolution physical maps that describe the biochemical structure of DNA and the ordered sequence of the genes themselves. By the early 1990s, scientists were constructing physical maps of model organisms using a procedure called

map-based sequencing. The process involves cutting DNA into fragments of about 200,000 base pairs called bacterial artificial chromosomes (BACs), taking note to record the position of the BACs on the genome, cloning the BACs in bacteria (such as *Escherichia coli*), determining the sequence of the base pairs, and then reassembling the BACs in their original order using a computer.

Map-based sequencing is the technique of choice for the Human Genome Project (HGP), a consortium of research centers funded by the National Institutes of Health, the U.S. Department of Energy, and the Wellcome Trust, a medical philanthropy based in London. The HGP was formed in 1990, with the goal of mapping the human genome by 2005, a date shortened by five years by politics, competition, and a variety of technical innovations. In addition to mapping the human genome, the project aims to store the information in databases, address the ethical, legal, and social issues raised by the project, and develop tools for data analysis and better sequencing technologies.

Scientists had still not completed a genome map for any organism by 1994, the year Venter and Nobel Prize-winning molecular biologist Hamilton O. Smith proposed speeding up the process with an alternative method they called whole-genome shotgun sequencing. In contrast to the HGP's method, in which the order of the BACs is known before they are each sequenced individually and then reassembled, the shotgun method involves cutting the DNA into small, random, overlapping pieces that are then sequenced and reassembled using a computer that compares all the pieces and matches the overlaps, thus assembling the whole genome. In 1995, with this technique, Venter and Smith mapped the genome of the disease-causing bacterium *Haemophilus influenzae*—the first completed genome of any single organism.

The period from 1995 to 2000 gave rise to a stunning series of technologic advancements, including computer automation and robotics, that accelerated the rate of genome mapping. At the forefront was a machine called the ABI Prism 3700 DNA Analyzer, introduced in 1998 by the Perkin-Elmer Corporation. This machine is involved in the last step of the sequencing pipeline. Its job is to separate fluorescently labeled DNA fragments by size to determine the sequence of nucleotide bases found on a strand of DNA. Now used by major DNA sequencing laboratories around the world, the Prism 3700 increased the rate of genome sequencing by approximately 20-fold.

The equipment used by both Celera and the HGP to draft the map of the human genome was virtually identical. The only difference between the two organizations was their basic methodology: the HGP used map-based sequencing while Celera used shotgun sequencing. The physical map they produced, the "language in which God created life," is an eye-numbing series of As, Ts, Cs, and Gs stretching into the billions.

Scientists acknowledge that much work on the human genome remains to be done. Since announcing the completion of the rough draft in June, Celera has moved on to other more lucrative pursuits befitting a biotechnology company, such as identifying and patenting gene sequences. Meanwhile, the HGP advances toward a final map, expected to be completed by 2003. At the present time, the BACs covering the two smallest chromosomes, numbers 21 and 22, are essentially complete. Chromosome 22 is particularly noteworthy: it's packed with over 545 known genes (at least 300 more are suspected) ranging in size from 1,000 to 583,000 bases of DNA. Gene variations on chromosome 22 are thought to be associated with at least 27 human disorders including brain cancers, schizophrenia, and multiple birth defects. Other BACs on other chromosomes are still in various states of disassembly, and there are still significant gaps to be filled. Nevertheless, the working draft is considered to be of great value for researchers looking for genes, and it represents a major accomplishment. Says Collins, "The completion of the human genome sequence will have a profound effect on understanding genetic contributions to human disease and the development of strategies for minimizing and preventing disease altogether."

Gene-Environment Interactions

One of the most exciting applications for genomics is in the area of gene-environment interactions. Now light-years beyond the theories of evil spirits and "bad blood" espoused by our ancestors, modern science views illness as the outcome of three related factors: genetics, environmental exposures, and aging. Like death and taxes, aging is a certainty. But within the complex interplay of genes and the environment lies a range of potential targets for disease prevention and treatment—particularly for cancer, pulmonary diseases, neurodegenerative disorders, developmental disorders, birth defects, reproductive function, and autoimmune diseases, all of which have been shown to be influenced by environmental agents. In fact, in a paper published in the 13 July 2000 issue of the *New England Journal of*

Medicine, Paul Lichtenstein of the Karolinska Institute in Stockholm, Sweden, and colleagues state, “Inherited genetic factors make a minor contribution to susceptibility to most types of [cancers]. This finding indicates that the environment has the principal role in causing sporadic cancer.”

One group spearheading the effort to elucidate gene–environment interactions is the Environmental Genome Project (EGP), headquartered at the NIEHS. Maynard Olson, director of the University of Washington Genome Center in Seattle, where much EGP research is conducted, describes gene mapping as an “extremely powerful tool for furthering investigations into human biology and its interactions with the environment.” Many of the enzymes and proteins involved in toxicity have already been identified in classical cell biology studies, he says. What genomic information allows researchers to do is build on this knowledge by identifying how many types of these proteins are expressed, and in what specific tissues they are found.

The mission of the EGP is to identify genes already mapped by other programs (the organization is continually soliciting candidate genes), and then to resequence the genes in an exhaustive search for the variations that augment resistance or susceptibility to environmental exposures. Most variants of interest are called single nucleotide polymorphisms, or SNPs, which are genes whose ordered DNA sequence is mixed up in ways that alter protein expression. If these proteins are involved in metabolizing or detoxifying chemical agents, it’s likely that exposure will cause greater harm. For example, a protein called p53 participates in cell signaling processes related to DNA repair, which is important because DNA is continually bombarded by carcinogens from inside and outside the body. If p53 detects chromosomal damage, it signals the cell to stop DNA synthesis or even undergo cell death. A person with a genetically inherited mutation that alters the function of the *p53* gene may be highly susceptible to cancer. Not surprisingly, the EGP has focused its initial efforts largely on finding SNPs among genes that, like *p53*, participate in DNA repair. Other processes of interest include cell cycle control, xenobiotic metabolism, and immune and inflammatory reactions.

The most recent addition to environmental genomics at the NIEHS is a newly formed National Center for Toxicogenomics, announced in December 2000 [see NIEHS News, p. A22]. Jose Velazquez, a program administrator for the NIEHS Division of Extramural Research and Training, says the center will complement

the activities of the EGP by investigating patterns of gene expression and protein function in response to chemicals. “It’s not enough just to understand the expression of the gene,” he says. “You also have to understand the function of the protein.” The problem is that protein function—particularly in the context of environmental exposure—is extremely difficult to characterize. The way the genome reacts to chemicals is highly dependent on timing and dose. Either of these parameters can exert a major influence on gene expression. Furthermore, chemicals trigger biochemical cascades within cells in which some proteins are turned off, others are turned on, and some aren’t affected at all. It’s really more a question of genome expression than gene expression. And indeed, many genomic scientists say the key to the future of genomics is actually found in functional genomics, or proteomics, which seeks to understand the global activity of proteins in a cell at any given time. Understanding how all the pieces fit together is as much a computational challenge as it is a biochemical mystery. For this reason, the center has also proposed the improvement of mathematical paradigms for the study of protein function as a major agenda item for the future.

The EGP’s ultimate goal is to sponsor and support epidemiologic studies of gene–environment interactions, both at the NIEHS and by outside researchers funded through the institute’s extramural grants program. Empowered by databases with useful information about genes and gene functions, researchers in the genomic era will be able to more clearly determine how specific populations respond to their environment. The activities now under way at the EGP constitute a critical step in that direction.

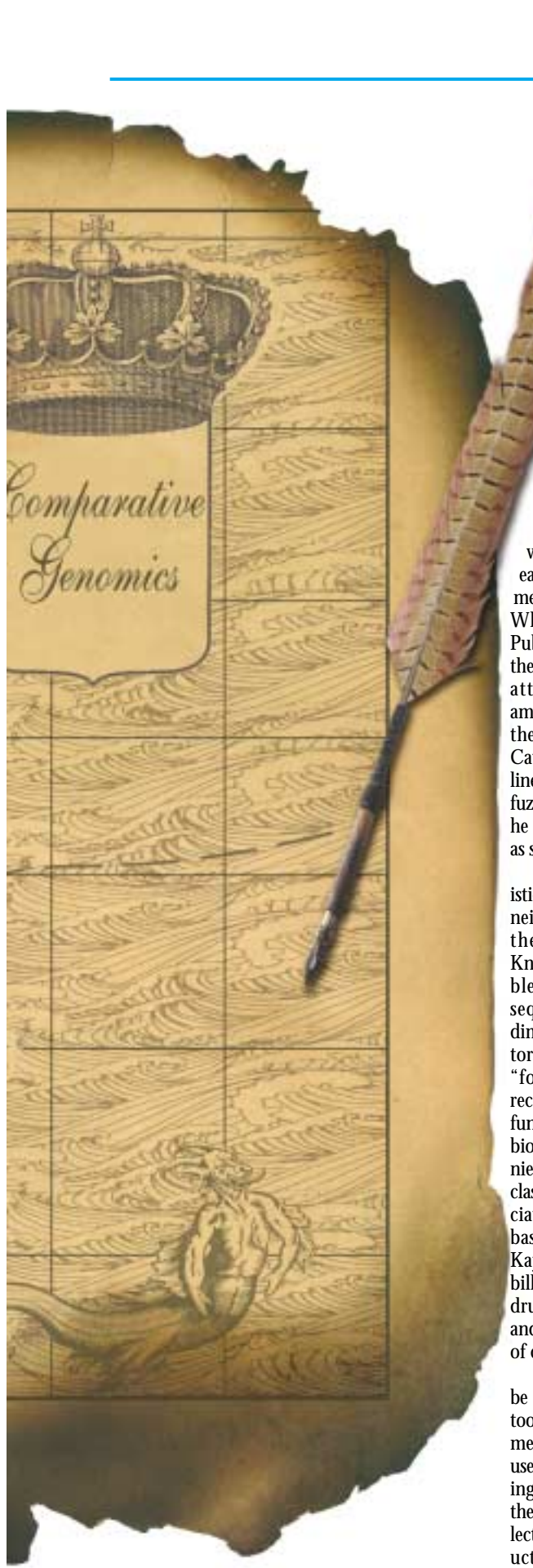
Comparative Genomics

Although the public’s attention has been captured by human genomics, mapping and studying the genomes of other organisms—or comparative genomics—is also important. Jonathan King, a professor of molecular biology at the Massachusetts Institute of Technology in Cambridge and a board member of the Council for Responsible Genetics, says, “Most genes evolved long ago, and there is enormous homology for essential functions like metabolism.” Therefore, he says, the amino acid sequences of many proteins are identical throughout all the higher species, a phenomenon known as genetic conservation. Comparative genomics is generally undertaken for two reasons: to provide road maps that help researchers locate genes for inherited characteristics and behavior in humans

and other creatures, and to advance the study of genomic evolution. As the maps become available, blocks of human DNA will be compared to other species to understand not only sequencing gaps but also genetic conservation. To date, physical maps of yeast, the soil worm *Caenorhabditis elegans*, and the fruit fly (in addition to numerous bacteria and viruses) have been completed. Scientists are also mapping the genomes of the mouse, rat, cat, dog, pig, cow, goat, zebrafish, rainbow trout, tilapia fish, medaka fish, rabbit, chicken, sheep, and horse.

Researchers at the Laboratory of Genomic Diversity (LGD) at the National Cancer Institute’s Frederick Cancer Research and Development Center in Frederick, Maryland, are mapping the cat genome, which apparently shares more order homology (meaning similar order of genes on the chromosome) with the human than any other nonprimate species that has been studied. According to William Murphy, a geneticist with the LGD, this means that genomes of both cats and humans have changed relatively little during the 90 million years since mammals diverged from their parent ancestor. Comparing the genome to a deck of cards, he says the genes of other laboratory species including mice, rats, and dogs have been substantially reshuffled over time, making them more challenging than the cat for studying genomic evolutionary history.

Comparative genomics is also useful for identifying models of human hereditary disease. There are a multitude of genetic diseases that show up in humans and other animals. Cats, for example, carry genes for hemophilia, polycystic kidney disease, and hypertrophic cardiomyopathy. According to an article by Stephen J. O’Brien and colleagues from the LGD published in the 15 October 1999 issue of *Science*, nearly every human gene has a mouse homologue, making this traditional laboratory animal highly amenable to the study of genetic illnesses. Studies in mice have already identified mutated genes for multiple disorders that are also present in humans. In the best known example, a mutated gene involved in metabolism is present in both overweight mice and morbidly obese humans. Similarly, comparative mapping studies of hypertension in rats have uncovered candidate genes for the same disease in humans. Within the year, scientists with the HGP expect to complete working drafts of the mouse and rat genomes, which will expand the potential for comparative inference relating to human health. As this occurs, scientists will use rodent models, and those of other species as well, to identify candidate



genes for analogous functions in humans and to define their interactions with other genes in the context of mutation, environmental exposure, infectious disease, sex, aging, and more.

The Myth of Determinism

Ironically, even as gene–environment interactions grow in scientific prominence, scientists worry that the recent media hype is helping to resurrect an outdated theory of “genetic determinism,” which suggests that individual phenotypes are governed wholly by genetic make-up. The danger, they say, is that people will take a fatalistic attitude toward disease and discount the effects of environment and lifestyle. David Page, chair of the Whitehead Task Force on Genetics and Public Policy and a professor of biology at the Massachusetts Institute of Technology, attributes the problem to a tendency among scientists to simplify genomics for the press, which have adopted “Gene X Causes Disease Y” as their standard headline. Without efforts to dispel some of the fuzziness surrounding genetic mechanisms, he says, the public will believe the truth is as simple as the headlines would suggest.

And just as the genome isn’t deterministic in the context of the entire organism, neither is it deterministic with respect to the structure of individual proteins. Knowledge of the genome makes it possible to describe a protein’s amino acid sequence, but not to predict its three-dimensional shape. A host of internal factors within the cell contribute to protein “folding,” which governs the protein’s receptor chemistry and therefore all of its functionality. This is a major problem for biotechnology and pharmaceutical companies hoping to use genomics to create new classes of drugs—a fact that is rarely appreciated by the public. “This is why gene-based biotechnology is in trouble,” says Kay. “Gene-based biotechnology lost \$5 billion last year. One hundred seventy-six drugs were proposed; one was approved, and the others faltered in the second stage of clinical trials.”

Ultimately, some suggest, genomics will be more useful as a research and diagnostic tool than as an agent that lifts medical treatment to new heights. However, efforts to use genomics for these purposes are increasingly hamstrung by a new application for the law in biology: gene patents and intellectual property protection for gene products [see Spheres of Influence, p. A30]. Private-sector companies are increasingly

patenting gene sequences, cell lines, genetically modified organisms, and even natural species—a practice that has angered many prominent scientists. Leveling sharp criticism at what he calls a “radical extension of patent law,” King says, “Genes are the products of millions of years of evolution and are in the deepest sense products of nature. They are not the inventions of individuals, corporations, or institutions.” King predicts that gene patents will retard research because investigators won’t share information for fear of undermining the ability to file patents later, and will impede health care delivery because providers will need to pay licensing fees to use gene products. In addition, he says, such patents provide a legal mechanism for companies to charge excessive sums for genetic screening (in the event they own the patent to the gene sequences of interest to the patient) and even more for gene therapy (in the event the screen shows the patient’s own gene is defective).

But, illustrating the contentious nature of the debate, these concerns were rejected by Philip Reilly, the chairman of the board and chief executive officer of Interleukin Genetics, a biotechnology company located in Waltham, Massachusetts. Reilly says that without legal protection for intellectual property, the \$20 billion in research dollars spent annually by biotechnology and pharmaceutical companies would dry up overnight. “I’m the CEO of a very small biotech company,” he says, “and I have never seen a more Darwinian process than trying to bring a biotech product forward. The vast majority of these companies fail, and the return on investment is very low.” He also charges that it is false to assume that biotechnology companies are looking to establish a “captive population for gene therapy.” Responding to negative publicity following the September 1999 death of 18-year-old Jesse Gelsinger, who died following experimental gene therapy for a rare metabolic disorder, he says, “Venture capital companies are running scared from gene therapy.”

And so, even as genomics promises to play a vital role in the future of biologic research, the overriding question facing society is how to use the information in ways that benefit the common good. Clearly, the answers to these questions remain unknown. What seems apparent is that genomics must be placed in the context of the whole human experience, including the environment in which we live. “Poverty is the main source of disease in the world,” says Olson. “It’s not genetic variation [that degrades public health] as much as it is economic variation. And we shouldn’t forget that.”

Charles W. Schmidt