The Human Genome Project:
Revealing the Shared Inheritance of All Humankind

Francis S. Collins, M.D., Ph.D.
Monique K. Mansoura, Ph.D.

National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland.

The information derived from the Human Genome Project, an international effort to decode the information embedded in the human genome, will revolutionize the practice of medicine in the 21st century by providing the tools to determine the hereditary component of virtually all diseases. This will lead to improved approaches to predict increased risk, provide early detection, and promote more effective treatment strategies. To be ultimately successful, these improvements in research and health care must reach everyone. This success will depend on participation from a broad spectrum of the population, such as scientists, clinicians, research participants, and active discussants, in deliberations of ethics and public policy. The Human Genome Project has helped to inform us about how remarkably similar all human beings are—99.9% at the DNA level. Those who wish to draw precise racial boundaries around certain groups will not be able to use science as a legitimate justification. However, studying the 0.1% of human genetic variations, particularly the distribution of single nucleotide polymorphisms, between affected and nonaffected individuals will significantly inform biomedical researchers about the genetic contributions to complex diseases such as cancer, diabetes, and mental illness. We must all work together to ensure that the risks of such research are considered carefully and that the medical benefits are made available to all.

© 2001 American Cancer Society

KEYWORDS: hereditary component, human genome, medical ethics, social issues.
affected a family member. In fact, virtually every disease has a genetic component, including such common diseases as cancer. The objective of the HGP is to unravel some of these mysteries of disease by unraveling the thread of DNA present in nearly every cell in our bodies. The genetic code within DNA holds many potential insights for individual susceptibilities and resistances to disease.

The HGP is distinct from most other biomedical research in that it has been defined by a series of very specific and quantifiable goals. We are happy to report that essentially all of these milestones have been achieved ahead of schedule and under budget. Early accomplishments included the construction of genetic and physical maps. These maps are research tools that have proven invaluable in the identification of more than 100 genes involved in diseases such as Huntington disease, achondroplasia, colon cancer, and breast cancer. These maps also provided a framework on which to proceed with the most visible goal of the HGP: sequencing the approximately 3 billion base pairs, or letters, that constitute the human genome.

The genetic code in humans as well as in all animal and plant species is spelled out in an exquisitely simple four-letter alphabet. The four chemical constituents of each DNA molecule are abbreviated A (adenine), C (cytosine), G (guanine), and T (thymine). What the genetic alphabet lacks in variety, it makes up for in volume. The DNA contained in a human cell is packaged in 23 pairs of chromosomes, each one containing millions of letters. If all 3 billion letters of the human genome were printed out on standard paper and stacked up, then the pile of paper would be as high as the Washington Monument.

Improvements in technology and the efforts of thousands of scientists throughout the world have resulted in an accelerated timetable for sequencing the human genome. A working draft of 90% of the human genome is on the brink of completion, and a finished, polished product in which gaps are closed and ambiguities resolved is anticipated by 2003. Progress can be monitored at a web site of the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/genome/seq), where a public data base, GenBank, holds all of the genetic information deposited by the international contributors to the HGP. Since 1996, participants in the public effort have held to the principle that newly generated genetic sequence data will be deposited into public data bases such as GenBank on a daily basis.

The study of the human genome and variations in it may shed light on how we are all different and, just as importantly, how we are all the same. What does the study of genetics tell us about concepts of race and ethnicity? Let us start with this observation from historian Evelyn Brooks Higginbotham: “When we talk about the concept of race, most people believe that they know it when they see it but arrive at nothing short of confusion when pressed to define it.”

When we look around at the people who surround us in our multiethnic society, we tend to focus on differences rather than similarities. We see superficial variations in skin and hair color and facial features, yet all externally visible traits represent only a tiny fraction of the genetic endowment of individuals. It is a much more profound revelation to realize how similar we are at the most fundamental molecular level.

In fact, the study of human genetic variation has enlightened our understanding of just how similar we are. It is estimated that the DNA sequence between any two individuals is 99.9% identical. Although genetic variations do exist, they seldom segregate in a manner that conforms to the racial boundaries constructed by sociopolitical means. The distribution of this 0.1% of differences among us is revealing. Studies have proven that the vast majority of these genetic variations are found within and not between populations, indicating that these variations were present in our shared ancient human founder group. This reflects the relatively young age and historically small size of our species. Research supports the hypothesis that modern humans originated from a founder population of about 10,000 individuals in Africa and that there was an expansion and outward migration 40,000–100,000 years ago. Furthermore, the evolution of our species really cannot be drawn correctly as a tree, with branches that never intersect. Instead, because gene flow has occurred in many directions, the history of the human population is more like a trellis with multiple interconnections. It is increasingly clear that there is no scientific basis for defining precise ethnic or racial boundaries. Those who wish to draw such exact boundaries cannot use science as a legitimate justification.

Although it will not provide scientific support for sharply defined racial categorization, cataloging the 0.1% variation in the human genome is currently a focus of intense research, because the data will provide information about increased susceptibilities or resistances to disease. The most common variations are called single nucleotide polymorphisms (SNPs). In collaboration with a private-sector consortium, the NHGRI is supporting research to identify SNPs from a common pool of 450 samples representing individuals whose geographic origins are Africa, Asia, Europe, and the Americas before colonization. These samples were obtained from individuals who provided full, informed consent and whose individual background
and identity will remain unknown, because all identifiers have been removed from the samples. A vast, dense catalog of SNPs will enable researchers to perform association studies that compare affected and unaffected individuals with common, complex diseases, such as cancer, heart disease, and mental illness. These case-control studies will look for correlations or “associations” between particular SNPs and diseases, most of which will apply to many different populations.

However, what about diseases such as sickle cell anemia that have a higher prevalence in the African-American population? Although sharp racial boundaries are meaningless at a molecular level, we are all part of historic extended families. The sickle variant arose in a “founding” individual thousands of years ago. (There actually appear to be at least three independent origins of the sickle mutation.) Positive selective factors, in this case, resistance to malaria afforded to carriers of a single copy of this variant, caused this variant to reach high frequency in descendants of the founder. A gene variant like this that appears at high frequency in offspring of an original common ancestor often suggests selective advantage but also can occur by a more random process known as “genetic drift.”

What about other diseases, such as some cancers, that disproportionately affect certain racial and ethnic minorities? For example, African-American men are 32% more likely to develop prostate cancer than white men. Does this imply some environmental cause or a genetic cause, such as a founder effect? The answer is not yet known, although a number of studies are beginning to address this question. Studies of hundreds of families, most of them white, with multiple males affected with prostate cancer have identified regions of chromosome 1 and chromosome X that are likely to harbor variations in genes that lead to increased susceptibility to the disease, although the precise genes involved have not yet been identified. However, most all studies looking for genetic contributions to common, complex diseases have been conducted primarily on white populations. Historically, minority communities have been hesitant to participate in genetic research. Members of minority communities will not easily forget the misuses and abuses of the not-so-distant past, such as the Tuskegee experiment or the debacle in the 1970s with sickle cell screening. The biomedical community has an obligation to work to regain the trust of communities that have undergone such troubling experiences. Individuals must be provided with sufficient information and be given the opportunity to balance risks with the possible benefits of participation in present day research. If minority communities do not fully participate in this research, then they may not experience the expected benefits. Knowing the factors, both genetic and environmental, that contribute to susceptibilities to disease will be essential to designing effective prevention, screening, and treatment strategies.

Broader inclusion of minorities as biomedical researchers also will facilitate the inclusion of minorities as participants in research trials. The recruitment and retention of individuals from populations that traditionally have not been involved in the health research enterprise has been an ongoing concern of many in the scientific and medical research communities. This is of particular importance for the NHGRI, given the far-reaching implications of genetic information and technologies for both individuals and groups. In addition to participating in traditional NIH recruitment mechanisms, such as the minority supplement and the predoctoral fellowship programs, the NHGRI has initiated a number of new and innovative training programs. These include a new career award for individuals from disadvantaged backgrounds and the ongoing short course in genomics for faculty at minority institutions.

Two years ago, the NHGRI began a productive collaboration with Howard University investigator, Dr. Georgia Dunston, on hereditary factors in African-American prostate cancer. With significant support from Dr. John Ruffin and the Office of Research on Minority Health at the NIH, the objective of this project is to uncover genetic variations in African-American men that are associated with increased susceptibility to prostate cancer. Family histories and tissue samples are being collected at seven sites throughout the United States, including Detroit, Chicago, Washington DC, New York, Houston, Atlanta, and Columbia. At most locations, African-American urologists serve as principal investigators who are involved in all aspects of the research project, including sample acquisition, data analysis and interpretation, and the publication of results. We currently have enrolled over 40 families in the study, and twice as many are expected to participate by the end of 2001. Preliminary data suggest involvement of genomic regions other than the previously identified chromosome 1 and chromosome X linkages. From this study, we hope to learn whether specific hereditary factors make prostate cancer so common and sometimes fatal in African-American men; then, that information will be used to develop better diagnostic tests and therapeutic interventions.

Such scientific advances were envisioned over a decade ago by the planners of the HGP. They recognized that the information gained from mapping and sequencing the human genome would have profound
implications for the health of individuals, families, and society. In addition to the potential for this research to dramatically improve human health, they realized that it would also raise a number of complex ethical, legal, and social issues. How should this new genetic information be interpreted and used? Who should have access to it? How can people be protected from the harm that may result from its improper disclosure or use? To address these issues, the Ethical, Legal and Social Implications (ELSI) Program was established as an integral part of the HGP. The ELSI Program was designed to provide a new approach to scientific research by identifying, analyzing, and addressing the ethical, legal, and social implications of human genetics research at the same time that the basic scientific issues were being studied.

The ELSI Program is viewed as essential to the success of the genome project in the United States and is supported with federal HGP funds. From its onset, the NHGRI has committed 5% of its annual research budget to study ELSI issues. One of the ELSI research goals for the current 5-year plan is to “explore how socioeconomic factors, gender, and concepts of race and ethnicity influence the use, understanding, and interpretation of genetic information, the utilization of genetic services, and the development of policy.” To begin addressing this goal, NHGRI recently issued a Request for Applications for grant proposals that examine ELSI issues surrounding the study of sequence variation research, with a particular focus on racial, ethnic, and socioeconomic issues. Such research will be vital to developing resources, including properly trained health care professionals in minority and underserved communities and culturally sensitive educational materials. These efforts, combined with an informed and involved community, will be an important step toward reducing barriers to access and avoiding unequal benefits of new genetic technologies as they become increasingly integrated into health care.

The NHGRI is expanding efforts to establish and maintain working relations with voluntary health organizations, such as the Intercultural Cancer Council, and to offer opportunities for dialogue with the public. In November 1999, the NHGRI hosted the first annual Consumer Day to inform participants about the NHGRI, the HGP, and how this research into “genetic medicine” may affect the lives of present and future generations (http://www.nhgri.nih.gov/consumer_day99/). In addition to presenting the latest advances in research and technology, the program offered sessions on genetic testing, genetic counseling, and genetic information resources and an update on protections for the privacy and fair use of genetic information. We were encouraged by the very favorable response to last year’s program and look forward to repeating this event later this year and for years to come.

The NHGRI also has contributed to important gains in the policy arena. The NHGRI has followed a model of cosponsoring workshops on topics like the fair use and privacy of genetic information in the workplace that have been attended by members of voluntary health organizations, professional societies, and other interested parties. The findings and recommendations of the workshop participants have been published and offer guidance for state and federal policy makers to protect individuals against genetic discrimination.16–20 In 1996, Congress passed the Health Insurance Portability and Accountability Act (HIPAA).21 HIPAA was the first step toward implementation of a number of the recommendations for protecting against health insurance discrimination; however, gaps remain. More recently, President Clinton took an important step toward protecting federal employees from genetic discrimination in the workplace when he signed the first Executive Order of the 21st century last February (http://www.nhgri.nih.gov/NEWS/Executive_order/index.html). This Executive Order prevents federal employers from accessing or using genetic information in hiring or promotion decisions among the 2.8 million federal employee workforce. It is important for individuals to participate in the policy process to work toward extending these protections against genetic discrimination to everyone.

The NHGRI will continue to work to extend opportunities that will ensure broad inclusion in all aspects of the HGP. Participation in research, training, ethical discussions, and policy development by individuals representing diverse communities will be vital to its success. We stand at the dawn of the 21st century, and our generation is the first to witness the remarkable identity of the universal thread of life encompassed within the cells of all people. We hold great hope that, by working together, our shared inheritance at the molecular level will translate into shared benefits in the application of the revolutionary medical discoveries that the HGP will enable.

REFERENCES


