Scientific Overview:

The Origin and Implications of the Human Genome Project

William F. Dietrich

The effort to sequence the human genome is an attempt to decode and understand the biological instructions (the genome) that are passed, in the form of DNA, from parent to child.[1] This project is the logical extension of research in genetics that has been taking place for the last 150 years. The results of this project will affect biological research, medicine, and anthropology in fundamental ways. It is crucial that the public gain a balanced and sophisticated understanding of the Human Genome Project.

Prelude to the Human Genome Project

In its essence[2], Genetics is the study of two questions: (1) How and why do individuals tend to resemble their relatives and members of their own species? (2) Given the tendency toward sameness noted in our first question, how and why do

I want to acknowledge Natalie Watson for assisting me in the literature research for this article. I am also grateful to Jennifer Lanni, Ph.D., and Rebecca Mosher, M.D. for helpful comments and suggestions on the manuscript.

[1] Throughout this essay, I will utilize certain specialized scientific terms. I try, to the extent possible within this short format, to define these terms clearly in the text. However, I can recommend an excellent on-line glossary of nearly all the terms I will use. This has been published by the National Human Genome Research Institute: http://www.nhgri.nih.gov/DIR/VIP/Glossary/.

[2] My summary of genetics has been influenced by my reading of an excellent elementary genetics text: Anthony J. F. Griffiths, Jeffrey H. Miller, David T. Suzuki, Richard C. Lewontin, and William M. Gelbart, *An Introduction to Genetics*, 7th ed. (New York: W. H. Freeman, 1999).

individual, heritable biological differences arise?[3] The questions of genetics have been considered by many during the course of history. Unfortunately, most of these thinkers were unable to provide any durable insight.[4] The science of genetics was placed on a solid theoretical foundation in the mid-nineteenth century by the work of an Augustinian monk named Gregor Mendel who pursued his love of plant breeding in a monastery situated in what is now Brno, Czechoslovakia.

Mendel was concerned with discretely differing physical characteristics (or phenotypes) of his pea plants, such as color and appearance of the seeds and seed pods, the height of the plant, and the position of the flowers on the plant (some may recall the round and wrinkled peas discussed in high school biology courses). He studied how these traits were passed on to offspring of crosses between plants having different phenotypes. Mendel's results, which were presented in a single article published in 1865, were nothing short of revolutionary.[5]

In his paper, Mendel clearly and correctly explained two fundamental principles of genetics. First, Mendel deduced that each individual plant has two copies of information (one copy contributed by its male parent and the other contributed by its female parent) that can act together to influence the development of any given trait. These transmitted pieces of information are now commonly referred to as genes.[6] Second, Mendel observed that genes sometimes have different forms (or alleles) that predispose a plant to have different versions of a given trait.[7]

---

[3] For example, as a species, why do all humans (even distantly related ones) have eyes placed on the front of their skull, as opposed to the sides? Alternately, why are there obvious physical differences (such as height, eye color, etc.) even between siblings?

[4] For example, we should not be surprised that Aristotle had something to say on this subject. In fact, Aristotle correctly deduced that what is inherited by offspring must be the *information* (he called it the "form-giving principle" or *eidos*) needed to create certain characteristics. His deduction was based on his observation that characteristics acquired in adulthood are not passed on to subsequent generations. However, he did make a fatal error because he did not recognize the full and equal contribution of the female parent to the characteristics of the offspring: Reviewed in Mayr, Ernst. *The Growth of Biological Thought: Diversity, Evolution, and Inheritance* (Cambridge, MA: Belknap, 1982).

[5] The complete German and English translation texts of Mendel's paper, "Versuche über Pflanzen-Hybriden," are presented at a website called MendelWeb (Edition 97.1 1997), edited by Roger B. Blumberg (http://www.netspace.org/MendelWeb/). I believe this website will be informative, even to an audience of non-scientists. In addition to the text of Mendel's paper, there are links to helpful annotations and commentary about Mendel's work.

[6] I will caution the reader against referring to any gene as "the gene for" a given trait (diabetes, alcoholism, cancer, etc.). It is absolutely clear that *many* different genes collaborate to influence the form of any given trait. It is also clear that a single gene can affect the form of many different traits.

[7] For example, Mendel described an allele or form of a gene that could predispose to the development of yellow seeds, instead of the usual green ones. He described many other such variations in the course of his work.

Once Mendel's descriptions of inheritance in plants were confirmed by others to be true for many different organisms, it became clear that his work answered, in bare outline form, the two primary questions of genetics that I posed earlier. (1) The genes that are passed from parent to offspring help to direct the growth, development, and functioning of the newly created organism in a way that assures uniformity within a species and phenotypic similarity to close relatives—why all pea plants have similarly shaped flowers, for example. (2) The inheritance of various combinations of the distinct versions of these genes (the alleles) can help to explain the differences in physical form and function between and within species—why some pea plants produce only red flowers, and others, only white, for example.

As important as Mendel's results were, they still left many questions and nuances unexplored. One important question was: What is the physical nature of the gene? This mystery was solved in the middle of the past century when experiments in bacteria showed that a component of cells called DNA was necessary and sufficient to transmit hereditary information. DNA is a long chemical polymer consisting of multiple repeating subunits referred to as A, G, C, and T. The DNA of an organism is packaged into pieces called chromosomes.[8] Eventually, evidence emerged from a broad variety of experiments that demonstrated that genes were simply portions of chromosomes having specific contiguous stretches of DNA subunits. Each gene has its own unique spelling of A, G, C, and T subunits, and the exact order of these subunits specifies a code of information.

Another important question left unanswered by Mendel was: How is the information contained in the genes utilized to affect the phenotypes of the organism? This is a difficult question that is still being studied today. However, we do have the beginning of an answer. By and large, the codes contained within genes provide blueprints for the production of bodily components called proteins, which, like DNA, are also polymers. In contrast to DNA, which serves solely to hold information, proteins carry out a vast array of specific functions within the body of an organism—to help eye cells detect color, to enable the heart muscle to contract, to assist in the digestion of the food we eat, etc. Thus, different proteins undertake a vast variety of molecular functions that are required for the life, development, and functioning of a given organism. From a structural standpoint, proteins are more complex than DNA because they are constructed of different combinations of twenty, rather than four, distinct subunits (these protein subunits are called amino acids). All living things contain a large number of proteins having different amino acid orders and compositions, and the exact spelling (or sequence) of amino acids in a given protein is specified by the code within a given gene.[9] In short, the genes encode for

[8] Each different species has different amounts of DNA packaged into different numbers of chromosomes. For example, humans have twenty-three chromosome sets and a DNA content of some six billion DNA subunit pairs, whereas fruit flies have approximately two hundred million DNA subunit pairs packaged into four chromosome sets.

[9] While it is true that genes most often encode for protein products, there are genes that encode for other biological polymers. I try to acknowledge this ambiguity in the rest of the essay by referring generically to the things encoded by genes as "gene products."

components that are necessary[10] to create and maintain the form and function of an individual organism.

With this information established, several important lines of investigation culminated in the following observations: (1) The alternate forms (or alleles) of genes are simply alternate spellings of the genetic code, which result in production of a slightly different gene product, or, in some cases, altered amounts of the original gene product. These differences have been shown to be at the root of the predispositions for different phenotypic characters. (2) In some laboratory organisms, such as a bacterium called *E. coli*, and the common baker's yeast *S. cerevisiae*, it became feasible to routinely isolate the gene sequence differences that could cause observed phenotype differences.[11] (3) The ability to predict the specific product encoded by a gene established a concrete link between the study of phenotype (genetics) and the study of molecular function (biochemistry and pharmacology). This connection made it possible to use genetics to identify the sequence of genes with effects on traits of interest, and then to use biochemistry and pharmacology to design drugs or other interventions which could alter the molecular behavior of the gene products for some beneficial purpose.

Origins of the Human Genome Project

In general, biologists share the conviction that a deeper understanding of the mechanical aspects of life will ultimately lead to an increased ability to treat and avert human physical suffering. The Human Genome Project came into being because the arguments for its impact on understanding human biology were compelling, and its potential for helping human society were perceived as very great. Given its highly publicized and celebrated status today, it is strange to recall that the project faced initial opposition from some members of the research community.[12]

---

[10] I want to caution the reader against interpreting this statement as an advocacy for biological determinism. Genes are not necessary *and* sufficient. The interactions between the information encoded by the genes and the circumstances of the individual's life will determine the phenotypic outcome.

[11] Let me emphasize that research using non-mammalian organisms is an extremely important avenue for discovering genes whose human versions could represent important therapeutic targets for human ailments. This is due to the fact that even distantly related organisms utilize surprisingly similar forms of information (genes) to specify similar molecular functions (proteins and other gene products) to perform similar necessary functions. Therefore, information obtained about a gene in one organism can often be used to predict the existence of a related gene in other organisms. Nevertheless, mammalian (e.g.: human) development and physiology have some distinct aspects that can only be uncovered by studying mammalian systems, making it important to conduct mammalian genetic research also.

[12] The idea of a large, multicenter project devoted to the sequencing of the human genome was treated with considerable skepticism because many biologists were concerned that the vast majority of government funding for research would be funneled off into the project, rather than to a diversity of projects pursued by many laboratories. In the end, the fears were unfounded, and the fruits of the Genome Project are now important sources of ideas for investigator initiated research.

The notion of determining the sequence of the human genome began to emerge in the late 1980s. One of the important motivations for considering this idea was, as we have seen, the backdrop of evidence that suggested that efficiently finding genes that influence human diseases would foster the development of new diagnostic and therapeutic tools. Unfortunately, the usual genetic methods for identifying genes of interest in experimental organisms were either not applicable or did not scale well to the extremely large human genome. Therefore, the idea was put forth that the sequence of the human genome was a critical piece of infrastructure for finding important genes.

Human geneticists were beginning to use the inheritance patterns of disease traits in human families to roughly localize the responsible misspelled genes to small portions of chromosomes. However, getting any closer proved to be prohibitively difficult and expensive. It was suggested that the prior existence of the human genome sequence would allow investigators to quickly search through the DNA of a portion of a chromosome for all the genes that reside in that area. This high-resolution gene map of the region would greatly facilitate the discovery of the crucial misspellings that predispose family members to a particular inherited disease.[13]

Even if there are no common misspellings in a gene that result in disease, simply identifying new genes and studying their expression can provide fruitful areas for research. With the DNA sequence of a gene and modern molecular biology techniques, one can ask when and where in the body a particular gene product is made. This expression information often provides critical clues about the biologic process that underlies disease traits. For example, one can ask whether people afflicted with immune disorders make the same gene products, in the right amounts, at the right time, and in the same parts of the body. Information about gene expression can be extremely useful to pharmaceutical companies in making guesses about likely targets of therapy. For example, the discovery of a gene that makes a product that is found exclusively (or largely so) in cancer cells could suggest an important target for the development of a targeted cancer therapy having great efficacy and minimal side effect.

The Post-Genome World

The nearly complete sequence of the human genome[14], is a tremendous technical achievement, but, because it is essentially a piece of infrastructure, its worth is

[13] Since the NIH was supporting the identification of human disease genes in several laboratories already, the powerful argument was made that funding the Human Genome Project (in the face of the budget trimming efforts then underway) would facilitate these gene discovery projects, and, in the long run, save money. See the testimony of Dr. Eric S. Lander before the House Committee on Appropriations, April 23, 1990, for a particularly cogent delivery of this argument.

[14] E.S. Lander, et al . "Initial sequencing and analysis of the human genome." *Nature* 409 (2001): 860–921. Venter, J. C., et al. "The sequence of the human genome." *Science* 291 (2001): 1304–1351. Both of these journal issues also contain excellent commentary articles for the interested reader. The current incomplete status of the sequence is a direct consequence of the conflict of goals for the public and private consortia. The pri-

best evaluated by considering what studies have been enabled by its completion. Therefore, I will close this essay by reflecting on likely lines of scientific endeavor that will be facilitated by the sequence data.

The identification of genes which, when misspelled, predispose individuals to disease ("risk factors") has been helped tremendously by the efforts of the Genome Project to create better gene cloning infrastructure. This point can be clearly illustrated with a simple statistic. Prior to the development of easily used and comprehensive human genome maps (a process that began in earnest in about 1992), the rate of human disease gene discovery was about two per year. After 1992, that rate increased by about six-fold to more than twelve per year.[15] It is reasonable to expect that the genome sequence, because it is one of the best genome maps yet produced, will spark even more increases in disease gene discovery. Already, the discovery of disease genes has contributed to our understanding of the biological processes that underlie all sorts of diseases, and, in some cases, has allowed the development of diagnostic tests.

From the standpoint of the individual, the data gleaned from the Human Genome Project promises a better understanding of personal risk for specific diseases and the possibility for a healthier life, as more risk factors are identified and diagnostic tests are developed. This information will provide physicians and patients with the possibility for more advanced preventative medicine. The person who knows he is at risk for heart disease can plan his diet and exercise program accordingly, and see a cardiologist before he is ever ill. Likewise, many women with familial breast and ovarian cancer have already taken advantage of diagnostic tests to determine whether they carry specific versions of genes that predispose them to these cancers.[16]

Many biotechnology and pharmaceutical companies have devoted large efforts to use human gene sequence to speed up the development of treatments for

---

vately funded group utilized a more rapid (almost "strip-mining") approach that would yield an incomplete sequence, since they were most interested in identifying a majority of the genes in the genome. This strategy was intended to rapidly enable the use of the gene sequences in screens for use by the pharmaceutical industry. This forced the public consortium to abandon a more thorough and difficult approach in order to ensure that the genome sequence remained as a property of the public as a whole. Now that the "cream" of the sequence has been skimmed, it will fall primarily to the public sequencing project to complete the sequence. This task will take several more years, but will provide important additional benefits to researchers in terms of completeness, speed, and ease of use of the data.

[15] NHGRI Fact Sheet: Positional Cloning. http://www.nhgri.nih.gov/Policy_and_public_affairs/Communications/Fact_sheets/positional_cloning.html.

[16] It should be noted that this information can lead to difficult decisions and must be accompanied by medical counseling. In the example above, some women carrying an allele which puts them at high risk for breast cancer choose to undergo prophylactic bilateral mastectomy rather than wait to see if they develop the disease. But for other women, the knowledge that they do not carry a risk-predisposing allele has freed them and their daughters from a lifetime of worry.

diseases. This strategy is still in its infancy and has yet to show significant results, as the development of drugs often takes a decade; therefore, the impact the human genome sequence has had on the pharmaceutical industry will not be felt for some time to come. However, it seems likely that industrial scientists will ultimately be successful in turning at least some of the scientific observations which are now being gathered into useful drugs for human therapy.

Aside from the discovery of disease-related genes, the genome sequence has facilitated research in other fields, such as human population genetics and evolutionary biology. The analysis of small spelling differences in the DNA of many individuals from various populations around the world has shown that there are very few consistent sequence differences that would provide a biological basis for distinctions made between races. The data analyzed from the human genome project will help erase long-standing stereotypes that have divided cultures worldwide. Evolutionary biologists have also begun to take advantage of human genome data to document our close genomic relatedness to other species.[17]

Perhaps the most interesting direction prompted by the human genome sequence will be the analysis of very complicated genetic phenomena. Throughout this essay, I have tried to indicate that various gene forms can *predispose* to certain phenotypic outcomes. As a technical matter, it has been much easier to identify alleles with very potent effects on the phenotype of interest. However, these variant forms with strong effects are relatively rare. Much more common are alternate gene spellings that have very small effects on an individual's overall chance of getting a particular disease. It has been speculated that many of our most common ailments, such as cancer and diabetes, are in fact caused by a collection of gene alleles with mild effects, rather than by a single gene allele of large effect. The genome sequence is facilitating the development of extremely dense genetic maps that may allow geneticists to detect the effect of genes of this type.

Epilogue

The Human Genome Project is like any other major technological advance—it can be used properly or abused. However, the primary motivation behind the Human Genome Project is for the good of humankind. Now, more than ever, an open dialogue is needed between scientific researchers, ethicists, and the general public to ensure that the promise of the Human Genome Project is realized.

[17] These points are reviewed by S. Paabo, "The human genome and our view of ourselves, *Science* 291 (2001): 1219–1220. I realize that some may object to the inferences of the evolutionary origins of man that are supported by the observations of genomic relatedness between, say, chimpanzees and man. While I have little to say on matters of anthropology, I would like to mention two additional related points. First, as a practical note, the similarities in the information contained in the genomes of different organisms permits us to do complete studies of gene function in experimental organisms and still be able to use that information to predict the biological behavior of the human organism. This is a tremendous advantage that makes biomedical research a viable enterprise. Second, as a biologist, I am astounded at the wondrous beauty of life as revealed by the commonalities among organisms. I would hope that the provocation of such feelings of awe in others would inspire greater respect for all the creatures of the earth.