LBL-29704

Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

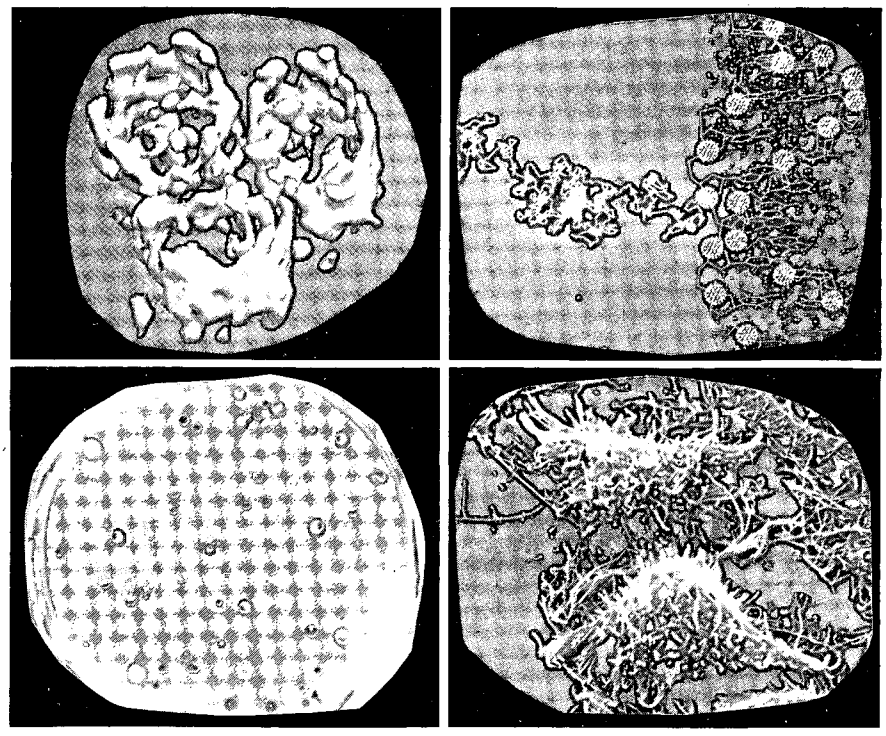
CELL & MOLECULAR BIOLOGY DIVISION

To be published as a chapter in **Clinical Neurosurgery,** Vol. 37, Williams and Wilkins, Publisher, Baltimore, MD, 1991

Implications of the Human Genome Project for Clinical Neurosurgery

C.R. Cantor

January 1990



LOAN COPY
Circulates
for 2 weeks
Bldg. 50 Library.
LBL-29704
Copy 2

Prepared for the U.S. Department of Energy under Contract Number DE-AC03-76SF00098

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

IMPLICATIONS OF THE HUMAN GENOME PROJECT FOR CLINICAL NEUROSURGERY

Charles R. Cantor
Human Genome Center
Lawrence Berkeley Laboratory
Berkeley, CA 94720

THE HUMAN GENOME PROJECT

The human genome project is the first large, organized, targeted biology program. It will span fifteen years, cost at least three billion dollars, and when finished will provide an immensely powerful resource of information and technology that will change the practice of both biological research and clinical medicine enormously. In addition, the project may serve as a model for other large biology efforts that will impact further on complex biological and clinical issues, especially those involved in the functions of the central nervous system.

The human genome is represented by all of the DNA molecules in a single cell. As far as we know, these molecules are the same in all cells except lymphocytes. The diploid DNA complement in a non-dividing cell is estimated to consist of six billion base pairs of DNA, distributed among 46 chromosomes, each of which contains a single, linear DNA molecule. Stretched end-to-end, these DNA's would extend about six feet. Except for the sex chromosomes in a male, each DNA molecule occurs in two very similar copies, one inherited from each parent. Thus the minimal information set needed to serve as a reference point is only about three billion base pairs, one copy of each type of chromosomal DNA.

The ultimate goals of the human genome project are to determine a three billion base pair reference DNA sequence and to locate the estimated 100,000 genes this sequence codes for. Modulated by environmental factors, these genes determine all that makes us human, from our physiognomy to our intellect and

This work was supported by the Director, Office of Energy Research, Office of Health and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

personality. Along the way to these goals there are a number of intermediate objectives, including the completion of several types of genome maps. The genetic map is constructed by observing the pattern of inheritance of multiple sets of markers which may include disease phenotypes and anonymous DNA polymorphisms. It is the coarsest map, with a resolution between pairs of adjacent markers in the 1 to 10 million base pair (Mb) range. At higher resolution is a physical map such as a restriction map. These maps provide a sketch of the genome. However, to determine the DNA sequence or to access genes directly, one must either have an ordered library, which is a set of cloned DNA fragments that spans the entire genome, or one must have sufficient amounts of known DNA sequence for each region so that amplification methods such as the polymerase chain reaction can be used, *in vitro*, to produce all the DNA from the region.

Besides the DNA sequence itself, the human genome project seeks to provide the tools to use it. These tools are of three types. Our current ability to interpret DNA sequence directly is still very limited. We can distinguish most coding regions from the majority, 90-95%, of non-coding regions. However, our ability to provide estimates of gene function or protein structure simply from DNA sequence alone is still non-existent. To find genes involved in human disease we will need to compare the DNA sequences of many individuals. This will require extensive DNA sequencing beyond the single reference DNA sequence. To establish gene function in many if not most cases will require experimental approaches such as modification or interruption of the gene. Such experiments are impractical in humans, and ethical considerations would preclude them anyway. Thus the human genome project will have a significant parallel component in genome studies on model organisms such as yeasts, the fruit fly, and the mouse. It remains to be seen whether mice will be close enough to humans to provide a suitable experimental system to explore the function of genes in the central nervous system.

This is a serious concern, because as many as half of all our genes may be restricted in function to the brain. If mice do not suffice as a model organism, we may have to face the difficult prospects of primate work to explore the functions of these genes.

Compared to these lofty goals, the current accomplishments in genomic research are rather limited. The largest contiguous DNA sequence is less than 250,000 base pairs. The largest overlapping libraries are less than 5 Mb. The largest genomic restriction maps are less than 15 Mb. It is clear, then, that massive advances in technology will be needed to realize the goals of the human genome project. Fostering such technological improvements is a major goal of the first five years of human genome project funding. Indeed, improvements in technology are likely to consume a significant fraction of the first decade of support for the human genome project.

GENETIC DISEASES

A genetic disease is an inherited DNA difference. Current estimates suggest that when any two chromosomes are compared, whether the two DNA copies of a chromosome within one individual or two chromosomes from different individuals, there will be differences on average every 1000 base pairs. This amounts to three million differences per genome. Most of these are presumably phenotypically silent. Many of the remainder are presumably responsible for the individual variations which make us interesting, but variations that fall within what we consider to be the scope of normal human biology. The remainder can lead to inherited diseases. Some diseases, like Huntington's, cystic fibrosis and Duchenne muscular dystrophy, are caused, as far as we know, by defects in only a single gene. In such cases genetic studies will provide the approximate location of this gene. Physical maps will let us focus in on the region. If any disease phenotypes involve massive alterations in the gene, such as deletions or chromosome translocations, these will signal the exact position of the gene. If the

only available disease alleles are single base changes or other microscopic DNA variations, extensive DNA sequence analysis may be needed to distinguish the particular DNA variations that produce the disease, and therefore lie in the gene, from others nearby which are innocuous.

Many of the most significant inherited diseases or disease predispositions are likely to be caused not by single genes, but by alterations in any one of a set of genes. An example would be a metabolic pathway in which interruption of any step might lead to similar or even indistinguishable disease phenotypes. Much more completed cases can be imagined. These multigenic diseases include various mental illnesses like schizophrenia and bipolar manic-depressive disease, inherited cancers, diabetes, hypertension, cardiac disease, alcoholism, drug addiction and obesity. Today, our ability to find the genes involved in these diseases is very limited. If the clinical phenotype does not offer a logical way to subdivide affected individuals into distinct groups that might correlate with the involvement of different genes, genetic methods will fail in their attempt to provide the location of any of the genes. Thus while genes involved in single gene defects can be found now by existing methods, the more complex cases may have to wait until more powerful technology is developed, through the genome project, that allows many alternate gene locations to be tested for simultaneously.

IMPROVED DIAGNOSTICS

As the human genome project unfolds, its impact on medicine will be felt mainly in diagnosis. The process of DNA mapping and sequence is very rapid at first, and partially complete maps of many chromosomes will appear in just a few years; extensive patches of DNA sequence will appear in less than a decade. Long before the remaining gaps are filled in, the information acquired will be available for studies of normal and abnormal human biology. For diagnosis of an inherited disease one does not actually have to possess the affected gene itself. Innocuous

DNA polymorphisms in the region of a disease allele will be inherited along with that allele virtually all the time. Meiotic DNA recombination, a process which could separate a disease gene from these molecular fellow-travelers, occurs on average only once in every 100 million base pairs. Thus a marker and a gene spaced 1 Mb away will be co-inherited 99% of the time. Once we are in the neighborhood of a disease gene, accurate antenatal diagnosis is virtually assured, if sufficient access to family members segregating the disease is available. Carrier testing or testing of potentially affected individuals without access to other relatives (including those affected with the disease) requires prior identification of the disease gene itself as well as knowledge of specific DNA disease alleles.

As the human genome project proceeds, we will acquire the ability to diagnosis scores, perhaps hundreds, of inherited human diseases. However, we will have this ability long before we have a corresponding ability to do anything about these diseases at either the level of prevention or therapy. This will lead to some potentially serious societal problems. Most individuals will probably not want to be tested for potential diseases, if the outcome of that test cannot be translated into any better medical care. The right of individuals under such circumstances to refuse testing must be protected; the desire of employers or insurers to gain access to this information must be suppressed. It is not that current practices are inadequate. There are really no new ethical issues that transcend current concerns about privacy of medical records. However, as more and more major diseases become visible under the molecular magnifying glass, the temptation to invade individual privacy will become greater and greater.

The first broadly-felt impact of the human genome project on clinical medicine is likely to be in improved differential diagnosis of polygenic disorders. What we call one disease today, like schizophrenia, is likely , within a decade, to be subdividable into a series of different diseases, each caused by a different gene defect.

As our ability to diagnose becomes finer, it should be possible to start to rationalize variations in the responses to existing therapeutic modalities. In many cases this should lead, rapidly, to improved therapeutic regimens as variants of a disease are distinguished. To appreciate the impact of this, it is useful to reflect that we are talking about disease states that potentially affect a majority of the world's population.

Identifying a disease gene, however, by no means leads to an immediate cure for that disease. Witness the case of sickle-cell anemia, where the precise molecular defect has been known for decades, with no improved therapy resulting. However, there is a reason not to be too discouraged by this example. The pathophysiology of sickle-cell disease is extraordinarily complex, and its connection with the molecular defect in affected individuals is very subtle. In most cases the route from gene defect to improved therapy should be shorter and straighter, but it is not a trivial step. It has been said that to understand the function of a single gene is a lifetime of scientific research. The human genome project, then, will seed 100,000 research careers. A few percent of these are likely to culminate in cures for major inherited human diseases, but the clinical impact of much of this work may not be felt until the middle of the 21st century.

LONG-TERM IMPLICATIONS

Surgeons, I imagine, have to deal with a confounding array of individual variations. Most of these are genetic. While they may not intrinsically cause disease, some surely affect the details of surgical intervention. The dramatic advances in medical imaging that have occurred over the past decade presumably allow many of these phenotypic variations to be visualized prior to surgery. However, as the genes involved are identified and their function understood, the possibility of a molecular imaging capability, far finer than any spectroscoping imaging, will emerge. It may well be that in the future, prior to any surgery, we will

demand a panel of DNA diagnostics that will pinpoint key individual variations in anatomy, or in drug- or trauma-responsiveness that will affect alternative surgical choices.

As we learn more about our genome, an inevitable concern is human genetic manipulation. Can we improve on nature's means to provide a better potential either for an individual or for his progeny? These are far from trivial concerns, but the prospect of immediate eugenic nightmares seems far-fetched. Some of the characteristics we hold most dear, like facial features, intellect and athletic ability, are likely to be modulated by large numbers of genes in complex interaction with the environment. Attempts to manipulate these genes are likely to have a large downside risk, probably too great to justify by the potential benefits of such manipulation. This complexity, plus our long lifespan, makes it seem unlikely that anyone will seriously execute a program of systematic human genome "improvement." However, we have to remain on guard against any individuals who might contemplate such a program, since, at least at present, the risks to both society and individuals seem to far outweigh potential benefits.

Assuming the success of the human genome project, and there are no major obstacles that appear likely to stand in the way of its success, the store of information accumulated and the new tools developed to obtain it will change the face of biological research. Similar studies are sure to be mounted on other organisms, especially plants and animals of economic importance. Computers will be needed to handle all these data, and it will be absolutely essential that all biologists be much better trained in informatics than past or present generations of researchers. Such research mainstays as laboratory notebooks are sure to be replaced by computer files and databases. Other large biology projects may well follow in the wake of the human genome project. For example, it will be quite important to develop a cell lineage and fate map for most or all of the cells present during mammalian

development. It will also be interesting, one day, to have a complete map of all of the synaptic connections in a typical mammalian central nervous system. These projects are certainly comparable in magnitude to the human genome project and certainly ought to be able to compete favorably, in terms of interest and importance, with other large science endeavors. Their impact, in the long run, on clinical neurosurgery is likely to be no less than that of the genome project.

LAWRENCE BERKELEY LABORATORY
TECHNICAL INFORMATION DEPARTMENT
1 CYCLOTRON ROAD
BERKELEY, CALIFORNIA 94720