Medical Applications of Mapping and Sequencing the Human Genome
(Human Genomics)

Statement by
Victor A. McKusick, M.D.,
University Professor of Medical Genetics
Johns Hopkins University School of Medicine,
before Subcommittee on Oversight and Investigations of the
Committee on Energy and Commerce,
Wednesday, April 27, 1988

I. Introduction:

The title of these hearings was given as "Sequencing and Mapping". I reversed the order because, as indicated by the report of the NRC/NAS committee, mapping should be done first, and sequencing later, although there will be considerable overlap of the two efforts, and although further technologic developments should proceed immediately, and even though mapping itself requires technologic development. The reason for "map now - sequence later" is twofold: the technology of mapping is adequate for immediate implementation, e.g., for the saturated RFLP map (although further technologic development is necessary for the contig map), whereas further technical development is necessary for most efficient mass sequencing; and 2, the maps, of both types just mentioned, will facilitate tremendously the complete sequencing.

See Figure 1: a. The RFLP map. b. The contig map.

Genomics is a relatively new term suggested by Dr. Thomas H. Roderick of the Jackson Laboratory, Bar Harbor, Maine for the field of mapping and sequencing and related analyses of complex genomes, such as that of the human.

As of this date, approximately 1300 expressed genes have been mapped to specific ones of the human chromosomes and for many of these genes, indeed most, we have at least some information on the precise regional location. In addition to this, approximately 2500 anonymous (function unknown) DNA segments have been mapped to specific sites on the chromosomes. Approximately a third of these show variation of a type that makes it possible to use them as RFLPs.

See Appendix I: The Human Gene Map newsletter March 25, 1988.

See Figure 2: The rate of growth of human gene mapping, 1968-1988.

II. <u>The medical applications of genome mapping</u>.

In the last 5 years great excitement has been expressed in both the lay and the professional press over the mapping, one after another, starting with Huntington disease, of more than a dozen genetic disorders:

Huntington disease - chromosome 4;
Familial adenomatous polyposis of the colon - chromosome 5;
Cystic fibrosis - chromosome 7;
Retinoblastoma - chromosome 13;
Polycystic kidney disease - chromosome 16;
von Recklinghausen neurofibromatosis - chromosome 17;
Myotonic dystrophy - chromosome 19;
Familial Alzheimer disease - chromosome 21;
Bilateral acoustic neuroma - chromosome 22;
Duchenne muscular dystrophy - X chromosome; and others.

The general enthusiasm that greeted these reports was entirely justified. All of these conditions shared in common the characteristic that at the time the mapping was achieved there was no clue as to the nature of the fundamental biochemical defect and therefore it was impossible to devise a specific diagnostic test for carrier detection, preclinical diagnosis or prenatal diagnosis and impossible to develop any methods of treatment that would serve to correct or counteract the ill effects of that biochemical defect. A leading and immediate consequence of the mapping of the disorders listed above and others is that it permits the elucidation of the fundamental fault and the devising of specific diagnostic methods.

Demonstration of the basic biochemical fault by going directly to the localization of the genes has been referred to as "reverse genetics". Usually we start from the clinical abnormality and identify an abnormality in some protein, often an enzyme, and work back to the gene that encodes that protein or enzyme. In all of the conditions listed above it had been impossible to identify a specific protein that was abnormal -- as stated earlier, the nature of the fundamental defect was unknown -- but by finding where in the DNA the gene is located, one can by the reverse genetics approach determine the normal function of that segment of DNA, i.e., the protein for which it codes, and proceed from that point to determine what the gene-determined derangement in that protein is in the disease and how it exerts its pathological effect. This information is critical to the development of methods of treatment that can correct or ameliorate the disorder at some step between the mutant gene and the clinical manifestations of the condition.

The diagnostic value of the mapping information lies in the realm of "DNA diagnosis". In all simply inherited (Mendelian) disorders such as the ones listed earlier, a specific lesion in the chromosomes, i.e., in the DNA, is causative and the diagnosis can be made by going directly to the DNA for demonstration of that lesion. This can be used for detecting the carrier status in a disorder such as cystic fibrosis, muscular dystrophy or hemophilia. It can be used for preclinical (i.e., premorbid) diagnosis in a condition such as Huntington disease, familial adenomatous polyposis of the colon, or bilateral acoustic neuroma and it can be used for prenatal diagnosis of any of these conditions. Such diagnosis is, of course, a tremendous aid to genetic counseling in disorders such as cystic fibrosis and in all of these disorders can permit the early institution of therapy and measures of secondary prevention in disorders such as adenomatous polyposis of the colon.

Gene diagnosis involves what might be called "diagnostic biopsy of the genome". DNA is obtained, for example, from the white cells in a sample of blood taken from the arm vein and is subjected to the appropriate studies to determine if the lesion of this or that genetic disorder is present. The presence of the lesion can be determined either by direct methods, including sequencing, or by indirect linkage methods (from study of the DNA in other family members, does the association with specific closely mapped markers indicate that the abnormal gene is present?).

III. Elucidation and specific diagnosis of cancers - for more accurate prognosis and more effective prevention and cure.

Work in the last decade, particularly the work related to mapping the human genome, establishes beyond all doubt the chromosome theory of cancer which was advanced by Theodore Boveri in 1914 and by some others before him. The work has furthermore demonstrated a high order of specificity for the type of DNA change responsible for a given cancer. This means that all cancers are acquired genetic disorders, somatic cell genetic disorders.

Classically, disease is divided into three categories according to the role of genetic factors:

1. Single gene disorders which show Mendelian pedigree patterns. Most of these conditions are individually rare, but in the aggregate, since there are many of them, they represent a substantial body of disease. (It is these disorders that are cataloged in Mendelian Inheritance in Man.)

2. Multifactorial disorders, i.e., disorders in which multiple environmental and genetic factors collaborate in causation. Examples are common conditions such as hypertension, mental illness, coronary heart disease, and frequent varieties of congenital malformations. These are discussed in Section C. inasmuch as these are likely to be elucidated greatly by genomic information.

3. Abnormalities of chromosome number and structure as in trisomy of chromosome 21, which leads to Down syndrome.

Although in many ways arbitrary, this classification has usefulness and must now be expanded to include a fourth category of genetic disease - somatic cell genetic disease - of which cancers is the main component. In addition, much autoimmune disease probably has its basis in somatic cell mutation and some congenital malformation are probably the consequence of localized somatic cell mutations. A somatic mutation theory of aging has been around for quite some time.

Mapping information of two different types (and a correlation of the two) has been responsible for the establishment of the somatic mutation basis of cancers: the demonstration of specific visible chromosomal changes in specific neoplasm (translocation, inversions or deletions) and the mapping of specific oncogenes to the same regions. Deletions are often demonstrable with the molecular markers even when they are not visible microscopically ("microdeletions"). Specific changes in DNA underlie all malignant growths, even those that have a demonstrable environmental factor in their induction; for example, small cell cancer of the lung in cigarette smokers has a consistent change in the short arm of chromosome 3.

The lesion in DNA characteristic of each malignancy and/or stage of malignancy can be used for the specific diagnosis of cancers. Complete mapping and sequencing of the human genome will accelerate the definition of the abnormality that characterizes each malignancy. Demonstration of the specific DNA lesions will be the means for diagnosis, replacing the presently used, relatively crude morphologic changes observed under the microscope. More accurate prognosis and more effective prevention and cure can be expected from the refinement of diagnosis provided by the study of DNA.

C. <u>Elucidation and detection of susceptibilities to common disorders</u>.

Hypertension, mental illness, coronary disease and some congenital malformations such a neural tube defects and cleft palate are examples of common disorders with a multifactorial basis. Methods for identifying the individual genes involved in these multifactorial disorders have been for the most part imperfect up to this point; indeed, they have been almost nonexistent. Complete mapping and sequencing of the human genome can be expected to reveal such genes or to open up the methods by which they can be identified.

D. <u>The medical applications of genome sequencing</u>

a. There are many types of maps of the human genome. The banded chromosomes represent one type of map. The RFLP map and the contig map (see Figure 1) are others. The cDNA map, which is a map of expressed genes because the cDNAs which are mapped are prepared from mRNAs, is another useful form of map. But the nucleotype sequence of the human genome is the ultimate map. Complete sequencing will be necessary to identify all the genes. Of the estimated 50,000 to 100,000 genes of man, only about 4400 are now identified, and the existence of many of these has only indirectly been deduced (see Figure 3). The sequence of the individual genes and of the DNA that separates them, i.e., the complete sequence of the human genome, will be the basis for the definition of the precise derangement in each clinical disorder, a requirement for diagnosis and treatment. Although about 3,000 separate Mendelian disorders are identified on the basis of clinical and family studies, for less than 400 of these disorders is the gene that is defective known and for only a few dozen of these is the precise defect in the gene known. (See Appendix II for a listing of known molecular defects in Mendelian disorders.)

IV. Concluding comments

From the advances of the last few years, the usefulness of gene mapping and the sequencing to clinical medicine has been extensively demonstrated even though the surface has scarcely been scratched. The applications already in hand adumbrate tremendous usefulness of the complete map and sequence. The complete map and sequence would be an invaluable tool to medicine and biology for many years to come.

See Appendix III for a reprint of the series of 4 articles which were published in MEDICINE (January 1986, January 1987, July 1987, January 1988) entitled "The Morbid Anatomy of the Human Genome: a review of gene mapping and clinical medicine".

Medical historians tell us that the anatomical treatise of Vesalius, published in 1543, had a powerful influence on the development of medicine. It turned physicians from fruitless speculation to a empiricism. It gave the profession of medicine a body of basic knowledge that was uniquely its own. It provided the basis for Harvey's physiology (1628) and Morgagni's morbid anatomy (1761).

The map and sequence of the human genome are the new human anatomy. They provide a neo-Vesalian basis for medicine in the future.