XVIII.
National Human Genome Research Institute

INTRODUCTION

The National Human Genome Research Institute (NHGRI) was established in 1989 to lead the effort of the National Institutes of Health (NIH) in the Human Genome Project (HGP). NHGRI's Division of Extramural Research funds HGP research in laboratories throughout the country. Research on genetic and physical mapping, DNA sequencing, database development, and technology development for genome research, as well as studies of the ethical, legal, and social implications of genetics research, are supported by the extramural arm of NHGRI. In February 1993, the Institute expanded its role at the NIH by establishing the Division of Intramural Research, a cutting-edge program that (a) translates the tools of the HGP into knowledge about human genetic disease and its diagnosis and treatment and (b) serves as the hub for human genetics research at the NIH. NHGRI manages or co-manages two NIH-wide scientific service centers that are supported by various NIH Institutes and Centers. The Center for Inherited Disease Research provides high-throughput genotyping services, advice on study design, sophisticated technologies for data warehousing, and database assistance for research efforts attempting to identify gene variants involved in human disease. The NIH Intramural Sequencing Center provides NIH - intramural investigators with access to large-scale DNA sequencing and sequence analysis.

NHGRI, through its extramural and intramural research programs, contributes to identification of genes involved in human disease and study of the functions of these genes and their products. The HGP provides data, material resources, and technology that will improve the ability of scientists to conduct biological research rapidly, efficiently, and cost-effectively. This infrastructure has already dramatically accelerated the study of human inherited disease. In the laboratories of the Division of Intramural Research, with the tools produced by the HGP, scientists are developing and using the most advanced techniques to study the fundamental mechanisms of inherited and acquired genetic disorders.

HIGHLIGHTS OF RECENT SCIENTIFIC ADVANCES RESULTING FROM INTERNATIONAL ACTIVITIES

Progress on Human Genome Project

The HGP is an international research effort to characterize the human genome and the genomes of selected model organisms through complete mapping and sequencing of the DNA. The goals of the project are to develop technologies for genomic analysis; examine the ethical, legal, and social implications of research in human genetics; and train scientists to use the tools and resources developed through the HGP to pursue biological studies that will improve human health. Begun in October 1990, the HGP is funded in the United States by NHGRI and the U.S. Department of Energy. After the success of the pilot phase in March 1999, an international consortium launched the full-scale effort to sequence the estimated 3 billion base pairs that make up the human genetic instruction book. With the United States taking the lead and with important participation by China, France, Germany, Japan, and the United Kingdom, the consortium expects to produce at least 90% of the human genome sequence in a "working draft" form by the spring of 2000, years earlier than initially expected.

On November 17, 1999, the consortium deposited the 1 billionth base pair of the human genome sequence into the public databases such as GenBank. Achieving this important milestone marks the success of the transition from the pilot phase to full-scale production sequencing. This international effort is distinguished from private sector efforts in its commitment to free public access to sequence data, without restrictions on use, and in its assurance of a finished product that has exhausted all available technologies for sequencing difficult regions of the genome. The updated status of the HGP can be monitored at the Web site http://www.ncbi.nlm.nih.gov/genome/seq/.

First Chapter of Human Genetic Instruction Book Deciphered

An international team of researchers achieved a historic scientific milestone by unraveling, for the first time, the genetic code of an entire human chromosome. The sequence of the 33.5 million base pairs that make up the DNA of chromosome 22 was deciphered by researchers at the Sanger Centre, near Cambridge, England; the University of Oklahoma, Norman; Washington University, St. Louis, Missouri; and Keio University, Tokyo, Japan. The sequencing of the DNA of chromosome 22 was conducted as part of the international HGP. All of these extremely high-quality data are freely available in public databases for scientists to use without the constraints of secrecy agreements, patents, or fees. Chromosome 22 was the first of 23 human chromosome pairs to be deciphered, partly because of its relatively small size and its association with several diseases. Research now will focus on determining what information can be deduced from these data. Sequencing and mapping efforts have already revealed that chromosome 22 is implicated in the workings of the immune system, congenital heart disease, schizophrenia, mental retardation, birth defects, and several cancers including leukemia, but many more secrets will be discovered in this decoded text. The results of this work will give scientists insights into the way genes are arranged along the DNA molecule and will pave the way for major advances in the diagnosis and treatment of disease.

Until now, scientists were uncertain about whether an entire human chromosome could be sequenced in this manner. For ex-

ample, they did not know whether insurmountable problems would prevent assembling their sequencing data. The presence of a small number of unclonable gaps (10) was not unexpected, but the scientists adhered to the standards that a chromosome should not be considered "essentially complete" (1) until the DNA sequences of regions that can be cloned and sequenced with current technology have been determined to high accuracy (less than one error in 50,000 base pairs) and (2) until the sizes of any remaining gaps have been determined.

Single Nucleotide Polymorphisms: New Tools for Tracing Inherited Diseases

The DNA of any two individuals is 99.9% identical. The 0.1% difference represents genetic variation that can lead to differences in the risk of developing various diseases. Some diseases (e.g., cystic fibrosis and Huntington's disease) result from differences in the DNA sequence of single genes. However, many common diseases (e.g., diabetes, cancer, heart disease, psychiatric disorders, and asthma) are influenced by complex interactions among multiple genes and by nongenetic factors, such as diet, exercise, smoking, and exposure to toxins. A catalog of the sites in the genome where the DNA sequence differs among individuals will help in the effort to discern the genetic signals associated with a disease, amid the noise from other influences on the disease.

Led by NHGRI, the NIH organized the establishment of the DNA Polymorphism Discovery Resource, which consists of 450 DNA samples obtained under strict ethical guidelines from anonymous, unrelated U.S. residents of diverse ethnic backgrounds. This resource is now the major source in the search for DNA variants known as single nucleotide polymorphisms (SNPs). In the next 2 years, NIH-supported researchers expect to find about 200,000 SNPs. This effort is complemented by a similar effort in the private sector. The SNPs Consortium consists of 10 large pharmaceutical companies, Motorola, and the Wellcome Trust, which are collaborating to identify an additional 310,000 SNPs and to regularly deposit the information into the public SNP database.

Armed with a robust catalog of SNPs, researchers can then study persons with or without particular diseases to discover the variants related to differences in disease risk and response to therapy. The large number of genetic variants will help researchers to identify disease-related genes, especially those for common diseases, with the goal of understanding the causes of the diseases. Association of variants with a disease will facilitate development of diagnostic tests and provide targets for further study to understand the biological processes underlying health and disease. This understanding in turn will fuel development of improved prevention and treatment strategies. Because genetic variants contribute to individual differences in response to drugs, the identification and understanding of these variants will allow physicians to choose the most effective drug on the basis of a patient's particular variants.

SUMMARY OF INTERNATIONAL PROGRAMS AND ACTIVITIES

Extramural Programs

In addition to support for sequencing of the part of the human genome that is being explored by U.S. scientists, NHGRI continues to fund a portion of an international consortium involved in the *Saccharomyces* Genome Deletion Project. This project aims (1) to generate a complete set of yeast mutants resulting from deletion of genes and (2) to assign functions to the genes by studying the mutants. The strains with deletion(s) are being generated by a consortium of nine European and nine North American laboratories.

International Databases

The NIH provided the sole support for the following international databases: Mouse Genome Database; SacchDB (yeast, *Saccharomyces cerevisiae*); OMIM (Online Mendelian Inheritance in Man); and GenClinics (a medical genetics knowledge base). Flybase (*Drosophila melanogaster*) is supported by the NIH and the British Medical Research Council.

The Mouse Genome Database and Flybase are maintained locally by institutions in Australia, France, Japan, and the United Kingdom. A mirror Flybase site is maintained locally in Israel.

International Meetings

NHGRI provided support for the following international meetings in fiscal year 1999 (FY 99):

- 12th International Mouse Genome Meeting, in Garmisch-Partenkirchen, Germany, in October 1998;
- meetings of the five largest international genome-sequencing centers, in December 1998 and February, April, May, and July 1999;
- International Workshop on Chromosomes in Solid Tumors, in Tucson, Arizona, in January 1999;
- Advances in Clinical Gene Therapy Research, in Bethesda, Maryland, in April 1999;
- Genome Mapping, Sequencing, and Biology Meeting, at Cold Spring Harbor Laboratory, New York, in May 1999;
- Capturing and Recording Human Sequence Variation, in Victoria, Australia, in May 1999;
- International Bioiron '99—World Congress on Iron Metabolism, in Sorrento, Italy, in May 1999;
- 12th International *C.* (*Caenorhabditis*) *elegans* Meeting, in Madison, Wisconsin, in June 1999;
- 2nd Annual Meeting of the American Society of Gene Therapy, in Washington, D.C., in June 1999;
- Annual David W. Smith Workshop on Malformations and Morphogenesis, in Schlangenbad, Germany, in August 1999;
- Microarray Meeting: Technology, Application, and Analysis, in Scottsdale, Arizona, in September 1999; and
- 5th International Strategy Meeting on Human Genome Sequencing, in Hinxton, United Kingdom, in September 1999.

During FY 99, NHGRI staff met with representatives of government, industry, and academic institutions from various countries, including Brazil, China, Finland, France, Germany, Japan, and the United Kingdom.

Intramural Programs and Activities

West African Origins of Type 2 Diabetes Mellitus in African Americans

During the past several years, the NIH Office of Research on Minority Health has supported innovative joint research by investigators from Howard University, Washington, D.C., and scientists in the intramural research program of NHGRI. The collaboration involves

the study of genetic risk factors for type 2 diabetes mellitus in African Americans.

There is a high prevalence of environmental risk factors for diabetes in the African-American population. It is more productive to study genetic risk factors in West Africans, because they are thought by many anthropologists to be the founding population of modern African Americans and have fewer dietary and nutritional confounding variables. For this study of diabetes, five recruitment sites from a total of 24 applications were selected through a peer review process; two sites were in Ghana and three were in Nigeria. Due to the logistic challenges in an investigation of this type in West Africa, the study was planned in stages, to allow assessment of a site by the ability to recruit appropriate patients; to collect blood, urine, and other clinical data; and to send the samples and data to the coordinating center at Howard University. The 1-year pilot project met its goal of recruiting 15 affected sibling pairs per site. On the basis of this experience, a full-scale study was implemented in September 1998, with anticipation that 400 affected sibling pairs from West Africa would be enrolled by the end of the study period. The study has started to yield high-quality data, and several expert scientists have joined the research team at the coordinating center at Howard University.

Hereditary Prostate Cancer

In FY 99, NHGRI continued a collaborative effort with scientists in Tampere, Finland, to explore the genetics of hereditary prostate cancer (HPC). To facilitate the identification of a gene or genes involved in HPC, investigators are identifying Finnish families with multiple cases of prostate cancer. Because of the unique population history of Finland, Finnish patients are likely to share a common ancestral mutation predisposing them to HPC. A shared ancestral mutation would in turn make it possible to use studies of inheritance to pinpoint the location of the mutated gene in the human genome. A gene predisposing to HPC in U.S. families (HPC1) has been identified on chromosome 1. The Finnish study facilitated identification of a second locus on chromosome X. This HPC-X locus may cause up to 40% of HPC in the Finnish population. To increase power for refining linkage of these and other candidate loci to development of HPC, the scientists are ascertaining the genotype of additional pedigrees in a survey using a more densely spaced set of markers. They are also obtaining extended pedigrees of patients with prostate cancer in Finland, Iceland, and Sweden and are using results in Iceland and Sweden to confirm the findings in the Finnish population.

Hereditary Breast Cancer

NHGRI also continued a joint effort with scientists in Tampere and Helsinki, Finland, and in Lund, Sweden, to explore the genetics of hereditary breast cancer. Genetic epidemiologic evidence indicates that the two known genes that predispose a person to developing breast cancer (BRCA1 and BRCA2) have a role in only about 15%–30% of families with hereditary breast cancer in these populations. This finding suggests that a third, unknown, susceptibility gene (BRCA3) may be involved. To facilitate the identification of a new gene or genes involved in hereditary breast cancer, the investigators have obtained data on and DNA samples from many families in Finland and Sweden with family members who have breast cancer but do not have the BRCA1 or BRCA2 gene. High-throughput fluorescent genotyping is performed at NHGRI to search for new loci involved in genetic predisposition to breast cancer.

Finnish-U.S. Investigation of Type 2 Diabetes Mellitus

The Finnish-U.S. Investigation of Type 2 Diabetes Mellitus aims to identify susceptibility genes for type 2 diabetes and for the related intermediate quantitative traits in a Finnish population. A genome scan using more than 400 genetic markers has been completed on DNA from approximately 2,400 persons. Analysis of affected sibling pairs has detected evidence for linkage to type 2 diabetes on both arms of chromosome 20 and suggestive evidence for linkage on the long arm of chromosome 11. Further statistical evaluations of the data set, including quantitative trait linkage and association analyses, have identified additional chromosomal locations that may contain susceptibility genes. A second set of approximately 2,900 DNA samples from other affected persons and members of extended families have been genotyped for markers on chromosome 20. Additional markers are being evaluated in regions of interest, and technologies are in place for high-throughput detection and analysis of densely spaced single nucleotide polymorphic markers. Further analyses of linkage and linkage disequilibrium will be used to narrow intervals for eventual positional cloning of the susceptibility genes.

Parkinson's Disease

NHGRI researchers also work with neurologists at the University of Patras, Greece, in a study of familial Parkinson's disease. Early-onset Parkinson's disease rarely occurs repeatedly in families. It usually occurs as an isolated case in families in which the hereditary contributions to susceptibility are numerous and complex and their effects are obscured by environmental factors. Scientists study the rare families with the strongly inherited early-onset form of disease, in an effort to identify additional genes that, when altered, can cause Parkinson's disease. They are also examining isolated populations in Greece that include multiple patients with Parkinson's disease. Isolated populations are likely to carry only a subset of the genes that predispose to the common forms of Parkinson's disease, making these factors easier to study and identify.

Tissue Microarray Technology

In addition, NHGRI continued a joint effort with scientists at the University of Switzerland, Basel, to develop technology for high-throughput molecular profiling of very large numbers of cancer specimens. The Pathology Institute, Basel, has more than 150,000 archival samples of tumor tissue and approximately 4,000 freshly frozen samples. Up to 3,000 specimens will be selected for analysis. The tumor tissue microarray technology, developed at NHGRI, enables a high-throughput analysis of molecular alterations in cancer cells at the DNA, RNA, or protein level. Up to 1,000 tumor specimens can be evaluated in a single experiment.

Oral Clefts

NHGRI collaborates on a study of the genetics of oral clefts (cleft lip, cleft palate, or both) with investigators in Syria. Researchers obtain family history, clinical data, and blood samples from persons whose families have several members affected with oral clefts, without the presence of a well-known genetic syndrome. DNA from the blood sam-

ples will be genotyped, and NHGRI scientists will perform statistical analyses by using these data, together with family history and clinical information, to determine whether there is evidence of genetic susceptibility to oral clefts. For any genetic areas tentatively identified, there will be additional investigation with the use of a dense map of genetic markers and further statistical analyses. The chromosomal regions most likely to contain an area that increases risk for oral clefts will be investigated with molecular genetic techniques designed to clone and sequence the gene in question. Several hundred persons have been studied in Syria, and genotyping of the first set of families is scheduled to begin in the near future.