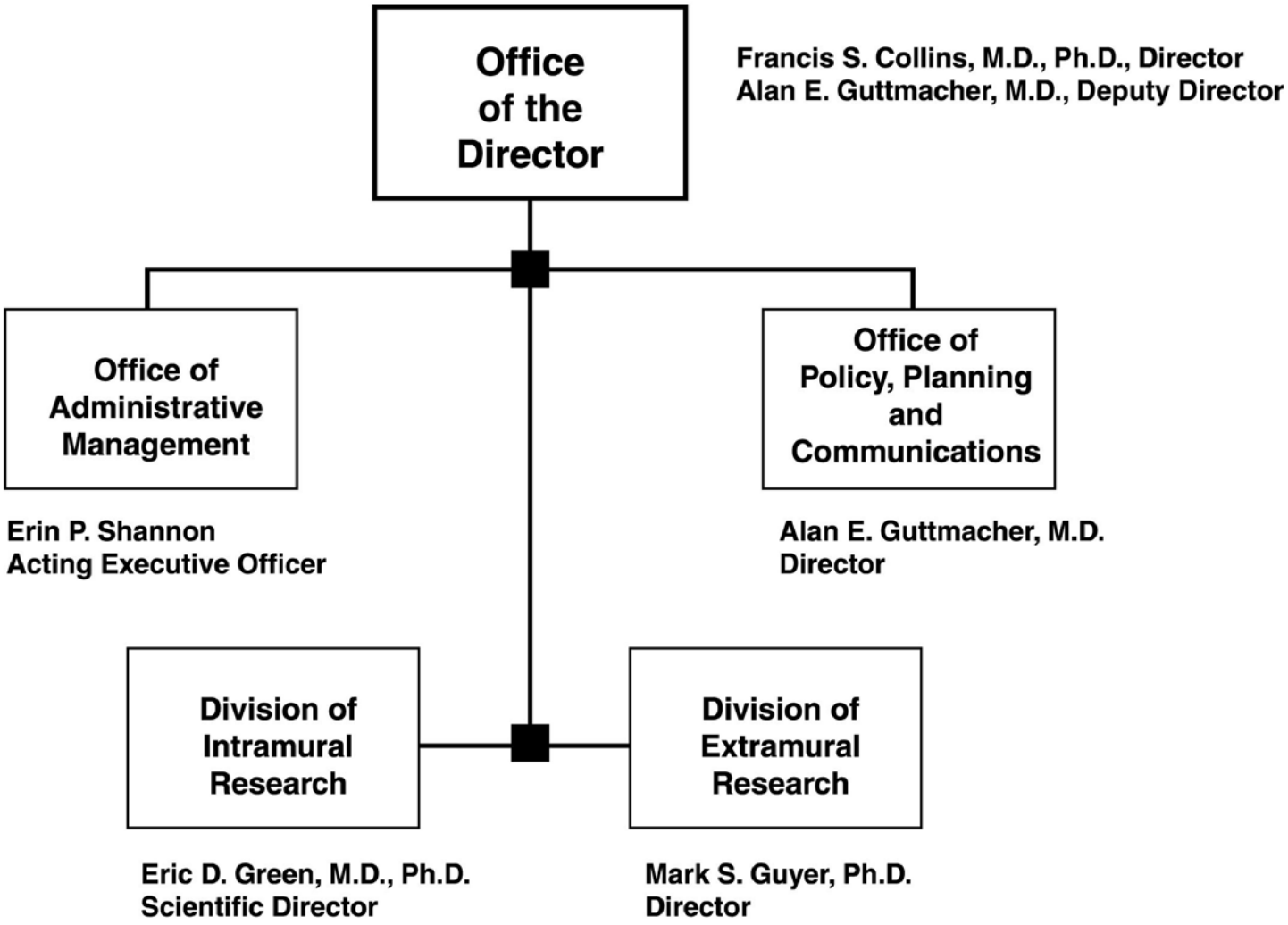
DEPARTMENT OF HEALTH AND HUMAN SERVICES

NATIONAL INSTITUTES OF HEALTH

National Human Genome Research Institute

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Organizational Structure



NATIONAL INSTITUTES OF HEALTH

National Human Genome Research Institute

For carrying out section 301 and title IV of the Public Health Service Act with respect to human genome research, $478,072,000.

National Institutes of Health
National Human Genome Research Institute

Amounts Available for Obligation 1/

Source of Funding	FY 2002 Actual	FY 2003 Amended President's Budget	FY 2004 Estimate
Appropriation	$429,515,000	$458,182,000	$478,072,000
Enacted Rescissions	(757,000)	(0)	---
Subtotal, Adjusted Appropriation	428,758,000	458,182,000	478,072,000
Real transfer to:			
Other HHS Agencies through Secretary's one-percent transfer authority	(464,000)	(0)	(0)
Comparative transfer from:			
Fogarty International Center for International Services Branch	25,000	25,000	0
Comparative transfer to:			
Office of the Director for program changes	(384,000)	(415,000)	(0)
National Institute of Biomedical Imaging and Bioengineering	(0)	(0)	(0)
Subtotal, adjusted budget authority	427,935,000	457,792,000	478,072,000
Unobligated Balance, start of year	0	0	0
Unobligated Balance, end of year	0	0	0
Subtotal, adjusted budget authority	427,935,000	457,792,000	478,072,000
Unobligated balance lapsing	(46,000)	---	---
Total obligations	427,889,000	457,792,000	478,072,000

1/ Excludes the following amounts for reimbursable activities carried out by this account:
FY 2002 - $32,581,000 FY 2003 - $34,210,000 FY 2004 - $35,222,000
Excludes $42,831 in FY 2002 and $168,705 in FY 2003 for royalties.

Justification

National Human Genome Research Institute

Authorizing Legislation: Section 301 of the Public Health Service Act, as amended. Reauthorizing legislation will be submitted.

Budget Authority:

FY 2002 Actual		FY 2003 Amended President's Budget		FY 2004 Estimate		Increase or Decrease	
FTEs	BA	FTEs	BA	FTEs	BA	FTEs	BA
285	$427,935,000	287	$457,792,000	282	$478,072,000	-5	+$20,280,000

This document provides justification for the Fiscal Year 2004 activities of the National Human Genome Research Institute (NHGRI), including HIV/AIDS activities. A more detailed description of NIH-wide Fiscal Year 2004 HIV/AIDS activities can be found in the NIH section entitled "Office of AIDS Research" (OAR).

INTRODUCTION

The National Human Genome Research Institute (NHGRI) leads the National Institutes of Health's (NIH) contribution to the International Human Genome Project (HGP), which is now very close to the historic goal of completing the sequence of the human genome. The NHGRI's ongoing mission also encompasses furthering understanding of the structure and function of the human genome, its role in health and disease, and the ethical, legal, and social implications of advances in genetics.

A series of five-year plans that set ambitious goals to achieve the objectives of the HGP and to advance understanding of the human genome have guided the institute since its inception. By clearly enumerating the program objectives of its U.S. component to the scientific community and the public and by providing measurable objectives to guide the work and gauge the progress of the HGP, these plans have been instrumental to the success of the HGP. The 1998-2003 plan, published in *Science* in 1998, emphasized the sequencing of the human genome. The plan accelerated the previously scheduled completion of the human genome sequence by two years, to the end of 2003. It also re-emphasized enhanced training of scientists and clinicians in genomics, technology development for DNA sequencing, and research in bioinformatics, computational biology, and the ethical, legal, and social implications (ELSI) of genomics as NHGRI objectives, while adding goals for technology development in functional genomics and for research in human sequence variation and comparative genomics. Virtually all the goals of the 1998-2003 plan are either accomplished or on a clear path towards completion.

Because that plan will soon expire and the looming essential completion of the human genome sequence provides powerful tools that will revolutionize genomics (and biosciences in general), the NHGRI is actively engaged in a more than year-long planning process to explore the future of genomics and to chart a course of leadership in the new era of genomics research that will follow the completion of all of the original goals of the HGP in the spring of 2003. This is an opportunity to think boldly and broadly about how to facilitate the translation of genomics to improvements in human health and how the NHGRI could best catalyze this transition. Unlike the previous five-year plans, the present effort incorporates all components of the institute: the extramural research program, the intramural research program, and the activities of the Office of the Director.

To launch this planning process, the institute invited a group of noted experts in genomics, its applications in biology and medicine, and its ethical, legal, and policy implications to a conference to identify the predominant themes that are likely to emerge in genomics research over the next ten to 20 years. A number of specific promising areas of genomic research were identified for further exploration. Over the course of 2002, a series of workshops further explored the ambitious, high-risk, and high-payoff ideas that resulted from this conference, and began to detail the NHGRI's role in achieving them.

Each workshop addressed the following questions: (1) Within the broad landscape of the workshop theme, what are the important priorities?; (2) What important priorities, challenges, and opportunities should the NHGRI tackle, and why? Which should it not tackle, and why not?; and (3) What specific activities should the NHGRI undertake to fulfill these priorities and opportunities?

In November 2002, the NHGRI convened a second conference to summarize and synthesize the findings and recommendations coming out of these workshops. Informed by these workshops and conferences, the institute is writing a new and ambitious plan outlining its vision for the future of genomics research and of the NHGRI. Publication of this new plan is anticipated for April 2003, a month that will also witness the essential completion of the finished sequence of the human genome and the 50th anniversary of James Watson and Francis Crick's seminal publication of the structure of DNA.

For more information about the NHGRI planning process or to review the workshop summaries, please visit http://www.genome.gov/Pages/About/Planning/.

A Celebration of The Genome - 50 Years of DNA: From Double Helix to Health

April 2003 will include three signal events in the history of genetics:

- The 50th anniversary of a short scientific report in *Nature* magazine, in which James D. Watson and Francis H.C. Crick described the double-helix structure of DNA;
- The essential completion by the International HGP of the finished reference sequence of the human genome. After a decade of technological advances and hard work, the international research teams that comprise the HGP will have made freely available in the public domain the entire sequence of genetic letters that make up the human genome, giving biomedical researchers an unparalleled tool to understand the workings of the human body;
- The publication in a leading scientific journal of a new plan for genomics, and for the NHGRI, that establishes priorities for the future of genetic and genomic research, predicting where human curiosity and creativity will take us next in understanding heredity and in applying our knowledge to improve health.

To celebrate these accomplishments the NHGRI is planning a number of events for next April:

- A two-day scientific symposium at the National Institutes of Health that will be web cast to institutions around the world. Participants—including James Watson, Francis Collins, and members of the International Human Genome Sequencing Consortium—will describe the science and history of the HGP. In addition, the symposium will explore the future of science and medicine made possible by breakthroughs in genomic science, and will include the unveiling of the NHGRI's new plan for the future of genomics;
- A half-day public symposium at the Smithsonian's National Museum of Natural History, featuring such scientific leaders as Watson and Collins, as well as consumer and other perspectives. The symposium will focus on how genomics influences science, health, and society;
- A national "DNA Day," on which high schools throughout the country mark the 50th anniversary of the description of the DNA double helix. Schools will utilize such tools as a videotape of an educational event that James Watson and Francis Collins will have with high school biology students on April 15th, a multimedia education kit and web-based lesson plans produced by the NHGRI, and a mentor network that the institute has created with the American Society of Human Genetics. High schools will be encouraged to make this the culmination of a month-long focus on genetics, to involve numerous other activities;
- A Congressional proclamation that April 2003 is "Genome Month" and April 25, 2003, is "DNA Day;"
- The installment of a genomic exhibit "Genome: How Life Works" at the Smithsonian Institution's Arts and Industries Building, by Clear Channel Communications, with support from Pfizer, Inc.;
- Activities at science museums across the country. Items available to museums will include a program guide of genomics-related events, a training workshop for museum staff and a kit of materials and equipment.

STORIES OF DISCOVERY

First Analysis of the Mouse Genome Sequence

A concerted effort to sequence the entire mouse genome began in October 2000. Three private companies, six NIH institutes, and the Wellcome Trust supported the initial phase of the project, which achieved its first goal, three-fold coverage of the genome, in May 2001. Members of the Mouse Genome Sequencing Consortium (MGSC), with support from the NHGRI and the Wellcome Trust, have continued working towards achieving a fully finished sequence by 2005.

In April 2002, the MGSC completed the mouse draft sequence and made available the first public assembly of the entire mouse genome. This marked the end of the whole genome shotgun phase of mouse sequencing, which generated enough sequence to represent the genome seven times over. In the December 5 issue of *Nature*, the MGSC published a paper reporting production of a high quality, highly ordered, representative assembly of the genome and an initial analysis of its content.

The sequence data and assemblies have been freely available throughout the course of the project. The next step of the project, already underway, is to convert the draft sequence into a finished sequence to provide a permanent foundation for biomedical research in the 21st century.

Previous comparisons of mice and humans showed that they share gene sequences to a high degree and share virtually the same overall set of genes. The availability of the assembly of the mouse genome allows for a closer analysis of these similarities, and of differences, on a genome-wide scale. Already, this comparison has produced a number of interesting observations. Perhaps the most striking observation is that fully two-thirds of that part of the human genome which is most similar to the mouse, implying that it serves some critical function, falls in parts of the genome not previously known to be important.

The increase in the amount of genomic sequence data for many organisms, including multiple mammals, has spawned a new generation of gene finding computer programs that use genome comparisons. These programs were tested using the mouse and human data and appear promising, making approximately 1,500 gene predictions additional to those from previous methods. Experimental validation suggests that these programs are more than 85% accurate.

Some striking examples of gene families (sets of genes whose members are structurally and functionally related) illuminate differences between the two species. In several instances, the number of genes in a gene family for a mouse is greater than in the human. It has been shown that the mouse genome has many more genes for olfactory receptors, genes that affect taste and smell, than the human genome. The new analysis shows that several different gene families controlling various aspects of reproduction in mouse have also expanded, including genes controlling pregnancy and selection of mates.

An annotated, highly continuous mouse sequence provides a powerful tool for finding genes involved in simple and complex diseases and for identifying other sequence elements that

regulate those genes. One can move directly from genetic mapping to identification of candidate genes, using only a few deft clicks of a computer mouse applied to the public databases.

The availability of an advanced draft of the mouse genome sequence will accelerate progress in biomedical research by improving the utility of the mouse as a model research system to study and understand human disease, and to develop and test new treatments in ways not possible with humans.

The Creation of a Haplotype Map

The NIH and researchers in the United States, Canada, China, Japan, and the United Kingdom are preparing to develop a "haplotype map" (HapMap) of the human genome by identifying how variation in the genome is organized along all of our chromosomes. The HapMap should be a key resource for finding genes affecting health, disease, and response to drugs and environmental factors.

Biomedical researchers have previously developed highly successful positional cloning methods to find the genetic basis of rare diseases that single genes strongly affect. However, multiple genetic and environmental factors influence many common diseases, such as diabetes, cancer, stroke, psychiatric disorders, heart disease, arthritis, and asthma. Linkage strategies that work well for single-gene "Mendelian" disorders lack power to map such multi-gene disorders, and, far too often, have yielded only weak linkages that follow-up studies fail to confirm. Thus, relatively little is known about the genetic basis of these common diseases, or of the factors that determine individual risk of disease clinical course or response to treatment. Discovering the particular DNA sequence variants that contribute to common disease risk and response offers a unique opportunity for illuminating pathways of disease causation in humans.

Sites in the genome where individuals differ in their DNA spelling by a single "letter" are called single nucleotide polymorphisms (SNPs). Recent work has shown that about 10 million SNPs are common in human populations. SNPs are not inherited independently; rather, sets of adjacent SNPs are inherited in blocks. The specific pattern of particular SNP spellings in a block is called a haplotype. Recent studies have also shown that although a region of DNA may contain many SNPs, it takes only a few SNPs to uniquely identify or "tag" each of the haplotypes in the region. This presents the possibility of a major shortcut in identifying hereditary factors in disease – instead of testing 10 million SNPs, a carefully chosen subset of about 400,000 SNPs would essentially provide all of the critical information. This should make approaches to finding regions with genes that affect diseases much more efficient and comprehensive, since effort will not be wasted typing more SNPs than necessary and all regions of the genome can be covered.

Most common haplotypes occur in all human populations, although their frequencies may vary considerably. Initial studies also indicate that the boundaries between the blocks are remarkably similar among populations in Europe, Asia, and Africa. These data provide strong support for the idea that a human haplotype map built with samples from these three geographic areas would apply to most populations in the world, although further testing of this conclusion is needed.

This new initiative will support the development of the HapMap, a catalog of the haplotype blocks and the SNPs that tag them. In addition to its use in studying genetic associations with disease, the HapMap will be a powerful resource for studying the genetic factors contributing to variation in response to environmental factors, in susceptibility to infection, in host immune responses, and in the efficacy of, and adverse responses to, drugs and vaccines.

One should note that a set of haplotype tag SNPs generally will not include the actual causative variant that influences disease risk or drug response. The responsible variant will travel on a particular haplotype, whose overrepresentation in individuals with the particular phenotype will point to that small segment of the genome.

The international consortium of countries committed to conduct and fund the HapMap Project will work in partnership with the SNP Consortium, consisting of the Wellcome Trust and several pharmaceutical companies. A steering committee will include members from the funded groups, the funding agencies, and appropriate others. The NIH began making grant awards for the HapMap project on September 30, 2002.

The NIH plans to require rapid data release of the SNPs, genotypes, quality scores, haplotypes, haplotype blocks, and associated information that the HapMap project generates. A protocol for the quickest release of data that allows adequate checking of their quality will be developed.

An interdisciplinary advisory group has been formed to provide advice on the scientific and ethical issues the HapMap Project raises, especially with regard to sampling populations. Ultimately, the development of this powerful tool will allow the biomedical research community to understand complex genetic diseases much more fully and will lead to improved treatments and, ultimately, cures for many of these disorders.

SCIENTIFIC ADVANCES

Regulation of a Critical Cell Cycle Checkpoint by the Breast Cancer Gene *BRCA1*

Studies in the mid 1990s identified mutations in the *BRCA1* gene that are the major cause of hereditary breast and ovarian cancer, and also play a role in causing other cancers. Women inheriting these mutations have a significant lifetime risk of developing breast cancer and/or ovarian cancer. This discovery has led to a great deal of work aimed at better understanding the function of the *BRCA1* gene.

New work by NHGRI researchers has helped establish that the *BRCA1* gene likely plays a role in the development of proteins that govern cell proliferation. When DNA, the genetic material of a cell, suffers damage the cell normally stops reproducing and activates a complex of proteins that can find and repair the damaged sections of DNA. This “pause and repair” step prevents the replication of damaged cells. Unless the cell carries out this repair step, the damaged DNA can be propagated to daughter cells. If the DNA mistakes are located in critical places, this can lead to cell growth that causes cancer. Studies indicate that mutations in *BRCA1*, which normally acts as a quality control, can disrupt the ability of cells to recognize DNA damage and halt division.

Such studies narrow where in the DNA damage pathway *BRCA1* functions. Research on *BRCA1* will also lead to greater understanding of the basic mechanisms of many cancers and of cell proliferation. More experiments are needed in order to pinpoint the *BRCA1* target.

Hereditary Form of Prostate Cancer Linked to Gene on Chromosome 1

Prostate cancer is one of the most common cancers in American men – an estimated 189,000 new cases occurred in 2001. Family history is the strongest risk factor. The NHGRI is working diligently to understand the hereditary factors – the genes – that contribute to risk for developing prostate cancer.

This line of inquiry promises to unlock secrets of prostate cancer; secrets whose opening will lead to new methods both to prevent and to treat prostate cancer much more effectively. For instance, one current study supported by the NHGRI suggests that while most men inherit two normal copies of a gene called "*RNASEL*"– one from their mother and one from their father – some men with hereditary prostate cancer inherit a defective copy of this gene from one of their parents. The normal *RNASEL* gene makes a specific enzyme that prostate cells need to function normally. While a prostate cell with one normal and one abnormal copy of the gene makes less of this enzyme, it still produces enough to allow the cell to be healthy. However, in such cells, the normal copy of the gene occasionally happens to mutate – to change – so that it, too, is abnormal. If this happens, the cell, and all the prostate cells that derive from it later, produce so little, if any, of this necessary enzyme, that the cell and its progeny are not healthy – indeed, in a manner that we do not yet understand, they grow into a cancer.

As these NHGRI-supported scientists search for genes that cause prostate cancer, they hope to develop better tests to predict and detect the disease, and to develop better treatments based on an understanding of the molecular causes of disease. There is strong evidence that understanding the mechanism of a disorder – for instance, how a mutation in *RNASEL* changes the working of the gene - will speed up the development of more effective treatments, and ultimately, cures and effective preventive strategies.

Low Vitamin C Levels May Be Linked to Massive Brain Bleeding and Lung Failure in Premature Newborns

The only proven human requirement for vitamin C, or ascorbic acid, is to prevent scurvy, a disease characterized by bleeding gums, anemia, skin hemorrhages, and death. The recommended daily allowance for vitamin C for women is 75 mg daily, which is increased by 10 mg during pregnancy – totaling a little more than the amount in the average orange. However, vitamin C intake varies greatly in the general public, ranging from 20 mg to 10,000 mg per day.

Scientists at the NHGRI's Genetic Disease Research Branch, the National Institute of Digestive and Kidney Diseases (NIDDK), and the division of Neonatology at the University of Pennsylvania School of Medicine and Children's Hospital have discovered a possible link between reduced vitamin C availability during pregnancy and the devastating respiratory failure and massive cerebral bleeding that can occur in a newborn immediately following premature birth. These scientists created a mouse model with a defective *Slc23a1* gene, which encodes the

protein that transports vitamin C into cells. They discovered that the mouse model could not deliver vitamin C from the blood to many fetal tissues or get it across the placental border. Therefore, the *Slc23a1*gene-deficient mice had noticeably reduced levels of ascorbic acid in their blood and very low or undetectable levels in their brains and other organs.

The study demonstrated that mice deprived of vitamin C during pregnancy died almost immediately after birth from bleeding in the brain and respiratory failure. In humans, intracerebral hemorrhage and respiratory failure are frequent causes of serious morbidity and death in premature infants. Since 20% of the population consumes less than the recommended dietary allowance for vitamin C intake, the study has important implications for what may happen with partial vitamin C deficiency, as well as for the devastating effect of a total absence of vitamin C. Scientists at the NHGRI and the NIDDK are investigating whether there are genetic differences in vitamin C absorption or transport, which could render some pregnant women and their fetuses more susceptible to partial dietary deficiency of the vitamin.

Genetic Defect Responsible for Brain Disorder Among Amish Babies

Over the past 40 years, 61 babies with a birth defect marked by profoundly small head and brain size have been born to 23 nuclear families in the Old Order Amish community in Lancaster County, Pennsylvania. None of the children has lived beyond the age of 14 months, and most die at four to six months of age.

An international team of researchers, led by the NHGRI, has discovered the genetic cause for a rare form of microcephaly, a devastating brain disorder that has stricken infants among the Old Order Amish for at least nine generations. In their study, the NHGRI team found that a defect in the gene causes developing cells to lose their normal ability to transport the building blocks of DNA across the inner membrane walls of the mitochondria, tiny structures that function as the cell's metabolic powerhouses. Researchers believe that without this carrying ability, called mitochondrial deoxynucleotide transport, the cell's mitochondria cannot make DNA properly, causing the brain of the fetus to develop abnormally. The NHGRI data also indicate that mitochondrial deoxynucleotide transport may play a crucial role in normal prenatal brain growth.

This study led to a significant finding for all prenatal brain development, making a tie between energy metabolism and brain development. Scientists will look at how this abnormality ties into other genes that are known to cause microcephaly, figure out how the genes interact with each other, and then look for other connections between energy metabolism and brain development.

Sequencing Genomes: Progress in Model Organisms and Selection of New Target Genomes

From the HGP's outset, the NHGRI and its partners have included among their research goals the mapping and sequencing of the genomes of several other organisms. Complete genomic sequences of many bacteria (e.g., *E. coli)*, several fungi (e.g., *S. cerevisiae)*, and two invertebrates (the roundworm *C. elegans*, and the fruit fly *D. melanogaster)* were generated in the past few years. Those sequencing efforts provided an enormous amount of information, including cost-effective ways to sequence entire genomes and greater understanding of the biological data encoded in genomic DNA sequence. Following the draft sequence of the human genome, interest in sequencing the genomes of many other organisms rose dramatically.

Genomic sequences for a number of important organisms, beyond those initially identified by the HGP, have been determined.

- Primary among these is the laboratory mouse. In May 2002, a draft sequence with 96 % of the mouse genome in long, continuous stretches of DNA was placed in public databases. This major milestone provides a key tool for interpreting the human sequence;
- Since February 2001, the Rat Genome Sequencing Project has achieved about 90% of its overall goal and will complete a draft sequence by the end of 2002. The rat is important in physiological studies and the availability of its genomic sequence will greatly facilitate studies using the rat and accelerate the availability of new therapies;
- The zebrafish (*D. rerio*) has become an important model organism and is currently used widely in developmental biology. Sequencing of its genome began in 2001, and by July 2002 an initial assembly of the zebrafish genome was released.

A peer review process for selecting which genomes to sequence has been established. Scientists who want to champion an organism for genomic sequence must write a "white paper" that clearly presents the arguments for putting their choice at the head of the line. After two rounds of white paper submissions, this process has determined the highest priority as: chicken, chimpanzee, cow, dog, a set of fifteen fungi, honey bee, sea urchin, and two protozoans (*Tetrahymena* and *Oxytricha*). Sequencing of the chicken and the chimpanzee has already begun. New cycles of this review process occur every four months, and it is likely that other genomes will be added to the high priority list soon.

Availability of genomic sequence and assemblies for multiple organisms provides crucial data for identifying functionally important human DNA sequences and, ultimately, understanding human disease. By comparing the sequences of different organisms, researchers gain valuable clues to understanding the human genome's structure and function.

Model Organism Databases

With the magnitude and diversity of new scientific data, research communities face challenges sharing data across disciplines, especially in genomics. Need exists to develop and maintain databases of genomic data to support the production of robust, exportable software that can be widely shared among different databases. To address these issues, Model Organism Databases (MODs), funded by the NHGRI, capture and organize key information, integrate data from disparate sources, and facilitate the formulation of new hypotheses and new perspectives.

The MODs have recently faced these data challenges by developing new database programs to help the research community. Each MOD has solved problems that are unique to its database. In addition, through the Generic Model Organism Database project, the NHGRI has encouraged development of shared database software to increase functionality of the databases.

Integration of the diverse data generated as part of the HGP is critical for understanding human disease and developing treatments. The MODs are developing new methods to keep pace with the increasing amount of data and to assist further in their analysis. Many of these methods will be useful in future human genome databases that will convert findings in model organisms to

new approaches to improving human health.

Bacterial Artificial Chromosome Resource Network

In the past several years, one system for recovering and purifying large DNA segments emerged as the method of choice for the construction of chromosomal DNA libraries - the bacterial artificial chromosome (BAC). BAC "libraries" that contain complete collections of DNA fragments from a given genome propagated in bacteria are important tools in the study of DNA sequence at a whole genome level. With increasing interest in genomic approaches to biological research, demand for new BAC libraries will increase rapidly in the next few years.

To date, the complete sequencing of all large eukaryotic genomes (including *C. elegans*, *D. melanogaster*, *A. thaliana*, *M. musculus*, and *H. sapiens*) has involved BAC clones. Recent experience in sequencing the human genome shows that determining the sequence of a large, repeat-rich genome requires an ordered map of clones, such as BAC clones, to assign stretches of DNA sequence to their location in the genome and to assemble regions containing repeated DNA sequences.

To increase the number of available BAC libraries and to generate a large number of BAC libraries from many organisms, the NHGRI, the National Center for Research Resources, the National Institute of Mental Health, and the National Institute for Child Health and Human Development awarded a set of cooperative agreements to form the NIH BAC Resource Network. The network will enable researchers to sequence specific genomic regions and to find and characterize genes. It will produce at least 15 BAC libraries at ten-fold coverage of "mammalian-size" genomes or the equivalent. A panel of six scientists oversees the network, regularly evaluates the program's progress, and makes recommendations to the participating institutes about the program. A peer review committee assesses requests for organism candidates for BAC libraries on the basis of scientific interest and strategic feasibility. For each organism proposed, the committee recommends whether the NIH BAC Resource Network should accept the request and, if so, whether it should be assigned to a high or to a standard priority pool.

Because BAC clones are relatively large and appear to represent an organism's genome faithfully, the BAC system will also be the choice for isolating targeted regions of genomic DNA from additional organisms used in specific biological studies, including a variety of mouse strains, and even from individual humans. Although worldwide genomic sequencing capacity has increased considerably in the last few years, certain research organisms will not be completely sequenced for some time. Availability of BAC libraries for many organisms will enable genomic studies to advance in the absence of complete genomic sequence.

Yeast Gene Deletion Mutant Set - A Powerful Tool for Understanding Genes and Proteins

Sequencing its genome and identifying all its genes is just the beginning of understanding the biology of an organism. Scientists want to know the function of all the proteins the genes

contained in that genome encode. They also want to know how the genes interact to form genetic networks and how the proteins work together in the cell to carry out biological processes. Deciphering this information is technically challenging, and much of the work to develop large-scale approaches to do so has occurred in model organisms, such as the yeast, *S. cerevisiae*. Yeast has served as an important model organism for both developing new genomic technologies and understanding the genomes of higher organisms, including humans.

One approach used to study gene function and genetic regulatory networks is to make and analyze genetic mutations that alter protein expression. While the principle behind this approach is as old as classical genetics, the availability of a full genome sequence affords a new opportunity to assess systematically and rapidly the function of every gene in a cell. Following the completion of the sequence of *S. cerevisiae* in 1996, the NHGRI supported the U.S. component of an international consortium formed to construct systematically a complete collection of gene-deletion mutants. To allow researchers to keep track of each deletion strain and to assess the effects of the mutation, short DNA sequences, "molecular bar codes," were inserted into each gene deletion. The bar codes can track mutant strains even within populations. Use of these molecular bar codes avoids the need to study the fitness of each deletion strain individually and provides an efficient, high-throughput approach. Construction of the set of deletion strains was recently completed, with 96% of genes in the yeast genome disrupted. These strains are available through distributors in both the U.S. and Europe. A number of investigators now use this set of deletions in a variety of studies aimed at understanding gene function.

The use of gene deletion strains provides an opportunity to employ a large-scale approach to determine the function of genes and their protein products and to start to determine how genes and their expressed proteins interact with each other.

Hemochromatosis and Iron Overload Study (HEIRS) Update

Iron overload is a common disorder estimated to affect between 1 in 200 and 1 in 400 individuals in the U.S. A major cause of iron overload is the genetic disorder, hereditary hemochromatosis (HH). It is estimated that one in every 400 Caucasians has HH due to variations in the gene, *HFE*. The genetic contributions to HH in non-Caucasian populations are not yet clear. When two specific *HFE* mutations were discovered in 1996, some observers suggested carrying out widespread screening for *HFE* mutations. In 1997, a panel of experts convened by the NHGRI and the Centers for Disease Control and Prevention concluded that it would be premature to consider widespread genetic screening for HH prior to answering important questions about prevalence, penetrance, genotype/phenotype correlations, and the psychosocial impact of such screening. Questions about prevalence, penetrance, and genotype/phenotype correlations were particularly problematic in diverse populations.

To answer these questions, the NHGRI is collaborating with the National Heart, Lung, and Blood Institute of the NIH to fund a large multi-center study designed to study iron overload and HH in diverse populations. Recruitment of participants began in February 2001, and continues at five field centers. As of September 2002, more than 80,000 participants had been recruited into the initial screening phase, almost 50% of who are from minority populations.

Comprehensive clinical exams of those screening positive are underway and a family study is beginning.

Analysis of data will begin in the coming year, and early findings from this study are expected within the next one to two years. The findings will inform health policy debates about possible widespread implementation of screening for iron overload and hereditary hemochromatosis.

User's Guide to the Human Genome

Genomic information, and the methods that allow useful access to it, have revolutionized biology. The NHGRI's commitment to rapid and open data release and to providing usable tools that ensure access by both advanced genomic researchers and lay biologists has made genomic information available throughout the scientific community and to other interested parties. The availability of the sequences of many genomes through the Internet has made an extraordinary amount of information freely accessible to anyone with a desktop computer and a link to the World Wide Web. However, the information itself is not enough to allow efficient use. Interested people from outside the centers of genomic study need to know where best to view the information in a form suitable for their purposes and how to take advantage of the software provided for retrieval and analysis.

In September 2002, researchers from the NHGRI published a User's Guide to the Human Genome. The electronic version of the guide appears on *Nature* magazine's website (www.nature.com/cgi-taf/DynaPage.taf?file=/ng/journal/v32/n1s/index.html) and subscribers to *Nature Genetics* received a free hard copy version. The guide offers help to those who otherwise might have trouble using the products of genomics. It is an additional resource for scientists who wish to make use (or better use) of both sequence data and the major tools that can be employed to view these data. The guide provides information in a highly inviting and understandable format. It is written in a practical, question-and-answer format, with step-by-step instructions on how to approach a representative set of problems using publicly available resources.

The User's Guide to the Human Genome, which is freely available on the World Wide Web, will help investigators, scientists, and students better understand and capitalize on the enormous potential of the data from the HGP. By making the HGP data more accessible, it will speed the pace of discovery, while encouraging young scientists to enter the field of genomics.

NEW INITIATIVES

The ENCODE Project: Encyclopedia of DNA Elements

In April 2003, the sequence of the human genome will be essentially complete. Although this is a significant achievement, much remains to be done to interpret our own instruction book. Before the information that the sequence contains can be used optimally, the identity and precise location of all of the protein-encoding and non-protein-encoding genes must be determined. The identity of other functional elements encoded in the DNA sequence, such as promoters and other transcriptional regulatory sequences, along with determinants of chromosome structure and

function, such as origins of replication, also remain largely unknown. A comprehensive encyclopedia of all of these features will greatly further utilization of the sequence to better understand human biology, to predict potential disease risks, and to stimulate the development of new therapies to prevent and treat these diseases.

To encourage rigorous comparison of existing computational and experimental approaches, and to stimulate the development of new ones, the NHGRI is creating a highly interactive public research consortium to carry out a pilot project for testing and comparing existing and new methods to identify functional sequences in DNA. Working in a cooperative effort to analyze rigorously a carefully chosen representative sample of the human genome sequence, investigators from both the public and private sectors will evaluate the relative merits of different techniques, technologies, and strategies in identifying all the functional elements in human genomic sequence, and in filling gaps in our ability to annotate genomic sequence. They will also assess how to scale up such methods to enable analysis of the entire human genome.

The ultimate goal of this project is to improve access to information, resources, ideas, expertise, and technology that are beyond the scope of any single group, and to affect the entire community of researchers interested in mining genomic sequence. The expected outcome is a clear path to determining all of the functional elements in the entire human genome sequence and integrating the information in a manner that will guide future basic and clinical research.

Development of a Compound Screening Resource for NIH-funded Investigators

The imminent essential completion of the human genome sequence in April 2003 presents both the opportunity and the pressing imperative to translate this unprecedented scientific accomplishment into tangible improvements in human health. For example, all currently marketed pharmaceuticals target the products of fewer than 500 human genes, or only about one percent of all human genes. A many-fold increase in usage is needed to take full advantage of the genome for therapeutic purposes. One of the ways that the NHGRI will pursue this is the identification and dissemination of research tools that allow better understanding and utilization of the genome. While some tools to do this exist currently, many of those focused on therapeutics are low-throughput and not yet scaled appropriately for whole-genome analysis. Besides powerful existing nucleic acid-based technologies, the NHGRI has identified another approach that has received less attention but is particularly suited to genome-wide utilization, the use of small organic compounds. These are the types of molecules used in most marketed pharmaceuticals, and are also referred to as "drug-like," or "small" molecules. Use of these chemical compounds to probe gene function will complement more conventional nucleic acid and antibody approaches. Such compounds have several attractive features for genome analysis: (a) their structural diversity is wide, mirroring the diversity of the genome; (b) unlike nucleic acid based methods, which tend to only inactivate targets, they may activate or inactivate them; (c) they can enter cells readily; (d) high throughput robotic methods analogous to those used in the HGP can identify them; and (e) for targets found to be therapeutically attractive, partnerships between the public and private sectors can use them as starting points for drug development.

This approach offers enormous potential. However, it is a new approach in genomics, and largely new to basic biomedical research as a whole. As a result, it would require substantial investments in physical and human capital. In collaboration with several other NIH institutes,

the NHGRI is currently planning for these needs, which would include: (a) large libraries of chemical compounds (500,000 – 1,000,000 total) of appropriate structural diversity and properties; (b) robotic assay capacity, also termed high throughput screening; (c) assay development capacity; (d) medicinal chemistry capacity to transform "hits" identified by high throughput screening into workable chemical probes; and (e) distribution capacity to disseminate the reagents to the biomedical research community efficiently.

While the subject of this effort will be new to the NHGRI, the approach is not, and follows directly on the NHGRI's past successes. The NHGRI pioneered high-throughput biology with the HGP and has subsequently exported this capacity successfully to the academic research community in the design and operation of such entities as the Center for Inherited Disease Research (CIDR). The same principles and vision will apply to this new screening effort, which should have a direct and powerful impact on the translation of genomic discoveries into deeper biological insight and new pathways to treatment of disease.

Studying the Genetic Basis of Health

The methods of analysis to find genes and genetic variants that contribute to disease can also help find the genes and genetic variants that contribute to health – that is, to resistance to disease. Studying disease resistance means studying people who do well under certain genetic or environmental conditions that cause many others to develop disease. The NHGRI plans to support development of new tools and analytical methods to discover the genetic components of disease resistance. By finding genes and genetic variants that are protective, researchers will better understand the disease process and how it can be slowed, or even prevented. This will assist in designing more efficient and effective disease screening strategies, establishing more productive prevention paradigms, and identifying better targets for drug development.

Researchers need to develop novel approaches to identify people who carry disease-resistant gene variants. People with a disease are easy to identify at clinics for that disease; those without the disease cannot be found in this way. One approach for finding people with protective variants is to look at those who are at high risk for a disease, but do not develop it. For instance, a variant form of the gene *CCR5* that protects against HIV infection was found by looking at highly exposed people who did not become infected. A similar approach could be used to study diabetes, for example. While obese people have a high risk of developing insulin resistance as a first stage in adult-onset diabetes, some obese individuals have normal insulin sensitivity, and they could be identified and studied. Another approach is to study relatives of people with a disease, who do not themselves have the disease. This would help to find genes and variants that decrease the risk of developing the disease or cause it to develop later or with milder symptoms. Researchers will be able to use the tools and analytic methods that the NHGRI supports to study a wide range of susceptibilities and resistances to diseases, disorders, toxins, and drug reactions.

Sequencing Individual Genomes for $1,000 or less

Development of new technologies and strategies for large-scale, high-throughput generation of biological data at relatively low cost has been a hallmark of the HGP. To understand the genome sequence that we have placed in the public domain, there is still an urgent need to improve and add to our existing toolkit of technologies. Over the past decade, the cost of genomic sequencing

has declined dramatically, falling from $10 to less than 9 cents per base pair. However, that cost must plummet further to maximize the impact of genomic sequencing on human health. Continued improvement in current technology is likely to yield more cost reduction, but not the magnitude of reduction needed to make large-scale DNA sequencing part of routine medical care. With that goal in mind, the NHGRI, along with many partners, will actively pursue a plan for developing new technology that would sequence an individual's genome for $1,000 or less.

The rationale for very inexpensive sequencing of whole genomes is strong. Other existing methods for detecting genetic variation among different individuals are powerful, but make assumptions of where the differences will be found. However, they are likely to miss small differences that do not appear in very many people at any particular location in the genome. Existing methods will play a part in discovering genetic contributions to common diseases, but will be limited in applying that knowledge to prevention and diagnosis for individuals.

Another important motivator for the $1,000 genome sequencing is the essential biological information it will yield. Comparison of mouse and rat genome sequences to human has shown that expanded comparative genomic analysis across species yields great insights into the genetics of human health and disease. Each time we obtain sequence for the genome of another animal, plant, or microbe, our understanding of the information encoded in the human genome increases. But the current cost of sequencing requires very careful selection of which genomes to sequence, and severely limits the opportunity to increase such understanding. DNA sequencing offers an opportunity to discover many microorganisms that we cannot now even isolate. Accurate, rapid sequencing may also be the best way to quickly detect and understand bioterrorism threats.

Achieving the $1000 genome will require the development of technology that exponentially increases the speeds at which DNA is processed, sequenced, and analyzed. This new technology would enable the 3 billion letters in a person's genome to be analyzed in a few hours, so doctors could quickly use the genetic information to understand the individual's disease susceptibility and tailor therapy to the individual's genetic profile. This is an audacious goal: it imagines sequencing a genome about 50,000 times cheaper, and a year faster, than we can today.

Technological paths to achieve the $1000 genome are apparent, but many hurdles must be overcome. The research and development, followed by commercialization, needed to turn this far-reaching vision into reality is likely to take at least 15 years. The NHGRI's vision is that, by setting this ambitious goal, the institute will spur greater creativity in the search for technological solutions that could completely change the face of biomedical research and medical practice.

The key technological constraints on achieving the $1,000 genome are speed, integration, and accuracy, each of which has hidden costs. Furthermore, to address the largest possible market's needs, the sequencing infrastructure must cost much less than today's sequencing factories (where a single machine costs $350,000) and be of a size and complexity manageable in the typical research lab or clinic setting. Another component necessary for the NHGRI's research portfolio will be integration of sequencing technology with new tools for computational analysis, as well with innovative large-scale, functional assays to attach meaning to the sequence data.

It will take many new ideas, paired with powerful engineering, to achieve the $1,000 genome. The NHGRI has successfully risen to daunting challenges before, as evinced in the soon-to-be

completed HGP, and is eager to do so again. To meet the goal of the $1,000 genome, the NHGRI will partner with the best minds in academia and industry to develop new technologies, with the aim of providing all Americans the opportunity to benefit from the HGP on an individual basis.

Completely Revised NHGRI Web site

Even before the release of the Web browser, Netscape, the NHGRI used the growing network of computers that would become the Internet to share scientific information. In early 1993, institute scientists produced a report card showing how extensively newly created genetic markers covered human chromosomes and shared that information with the research community through a workstation sitting on a staff member's desk. In 1994, the NHGRI launched its first full-fledged Web site. On June 21, 2002, the NHGRI launched a totally redesigned Web site with a new Internet address - genome.gov. The streamlined Web site address makes it easy for users to access a comprehensive and authoritative government site focused on genomic research, including the international HGP. The new Web site supports ongoing scientific studies by researchers inside and outside of the institute and provides a reliable source of genomic information for a wide range of audiences, including individuals with genetic disorders and their families, teachers and students, and the general public. Links on the site connect users to other federal government sources for information about genomics.

The architecture of the new site organizes the institute's information into seven major categories. More than one-third of the pages are in the research section of the site, providing an overview of the institute's scientific activities, including an overview of the HGP. The research section also includes a complete description of the research program of the institute's intramural scientists and an extensive collection of online resources, ranging from scientific databases to software tools, to links to other genome research centers around the world. The site also includes sections on health, policy and ethics, educational resources, careers and training, grants, and news.

The new site demonstrates the NHGRI's commitment to ensuring that all users, including those with special needs, have easy access to the institute's information. Additional services will be added to the new Web site, including a visual media database that will allow users to download high-resolution scientific images and illustrations and multimedia and video-streaming services.

OTHER AREAS OF INTEREST

Reducing Health Disparities

The NHGRI recognizes the critical importance of empowering people from minority communities and of engaging them in genomics research to ensure that they benefit from its applications. Genome research offers tremendous challenges and opportunities for improving human health, as well as for exploring some of the most profound ethical, legal, and social issues of our time. As the institute's research mission has grown, the need for inclusion of minority communities has become even more imperative.

The rewards of genomic research will be realized only with active participation of all racial and ethnic groups. An important application of genomic research will be to investigate how DNA sequence variation affects phenotypic differences, including differing susceptibility to disease

among various groups. The significant societal ramifications of this research also need attention. Such research must involve individuals with diverse perspectives. Genome research affects all; thus, all populations need to set the research agenda and examine the broader issues it raises.

The NHGRI has intensified its efforts to address issues of health disparities by developing a strategic plan that identifies goals in areas such as research projects, information sharing, development of partnerships, and increasing diversity of the research workforce. The plan also addresses the ethical, legal, and social implications of genetics research, training to build research capacity, and education and outreach. One focus of the plan is furthering research collaborations between the NHGRI, the National Center for Minority Health and Health Disparities of the NIH, and minority serving institutions. This has included working with the National Human Genome Center at Howard University to identify the genetic and environmental factors that contribute to the disproportionately high incidence of certain complex diseases in African Americans (such as prostate cancer and diabetes), expanding minority training opportunities in the NHGRI's Centers of Excellence in Genomic Science (CEGS) program and in its genome production centers, and conducting a comprehensive epidemiology/genetic study to determine environmental and familial/genetic risk factors for prostate and breast cancer in Barbados. The NHGRI has also developed an action plan to increase the number of individuals from underrepresented minority populations trained to pursue research in genomics and/or in the ethical, legal, and social implications of genomic research.

The increased visibility of the HGP since its publication of the draft human genome sequence in February 2001 has made it important to intensify education and outreach efforts to underserved minorities. For example, the NHGRI, along with the NIH Office of the Director, welcomed members of the National Medical Association (NMA), the nation's leading professional organization representing African-American physicians, to the NIH in June 2002 to increase knowledge of genomics and to allow the NIH, the NHGRI, and the NMA to work together to ensure that this important new biomedical tool benefits minority populations.

In August, 2002, to shepherd and coordinate these and other such efforts, the NHGRI selected a senior level consultant, Vence Bonham, J.D., who is an authority regarding issues of race, ethnicity, and genetics and health disparities, to join its Office of the Director.

Effects of Gene Patents on Genetic Testing and Research

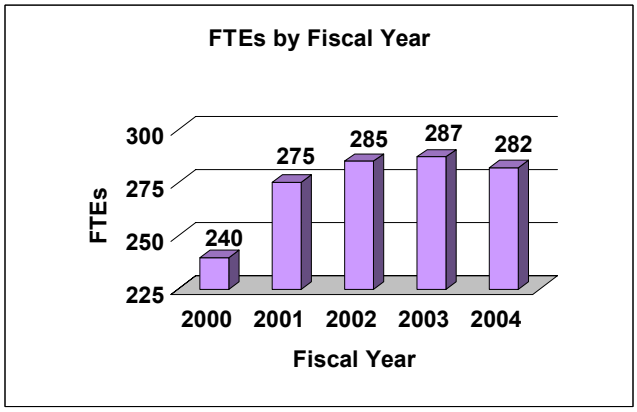
The NHGRI continues to be concerned about the issue of gene patenting and licensing. In order to gain a better understanding of these issues, the NHGRI has funded case studies and surveys, combining qualitative and quantitative approaches, to describe and analyze the effects of patents that award proprietary claims to the use of DNA sequences in genetic tests. NHGRI-supported researchers have also interviewed technology transfer and licensing officers at the NIH, academic institutions, and private firms and gathered available written policies. In addition, they have collected and described existing licensing terms for a small sample of genetic sequence patents, focusing initially on genetic testing claims and expanding to related claims and technologies.

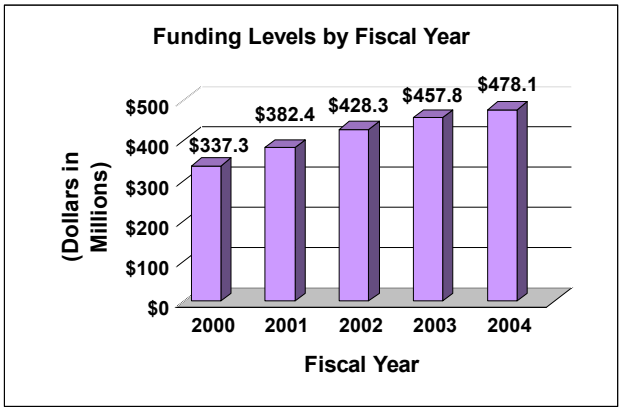
The NHGRI also held a round table discussion in December 2002 with outside experts in gene patenting. The purpose of the roundtable was to explore the ramifications of patenting and

licensing genetic sequence data and SNPs on healthcare delivery and research in the current climate and in the future. The NHGRI will be able to utilize this information to help inform policy development and to define further research that would inform the policy process.

BUDGET POLICY

The Fiscal Year 2004 budget request for the NHGRI is $478,072,000, including AIDS, an increase of $20,280,000 and 4.4 percent over the FY 2003 amended President's Budget Request. A five year history of FTEs and Funding Levels for NHGRI are shown in the graphs below. Note that Fiscal Years 2001 and 2000 FTEs are not comparable for the NIH Human Resources functional consolidation.



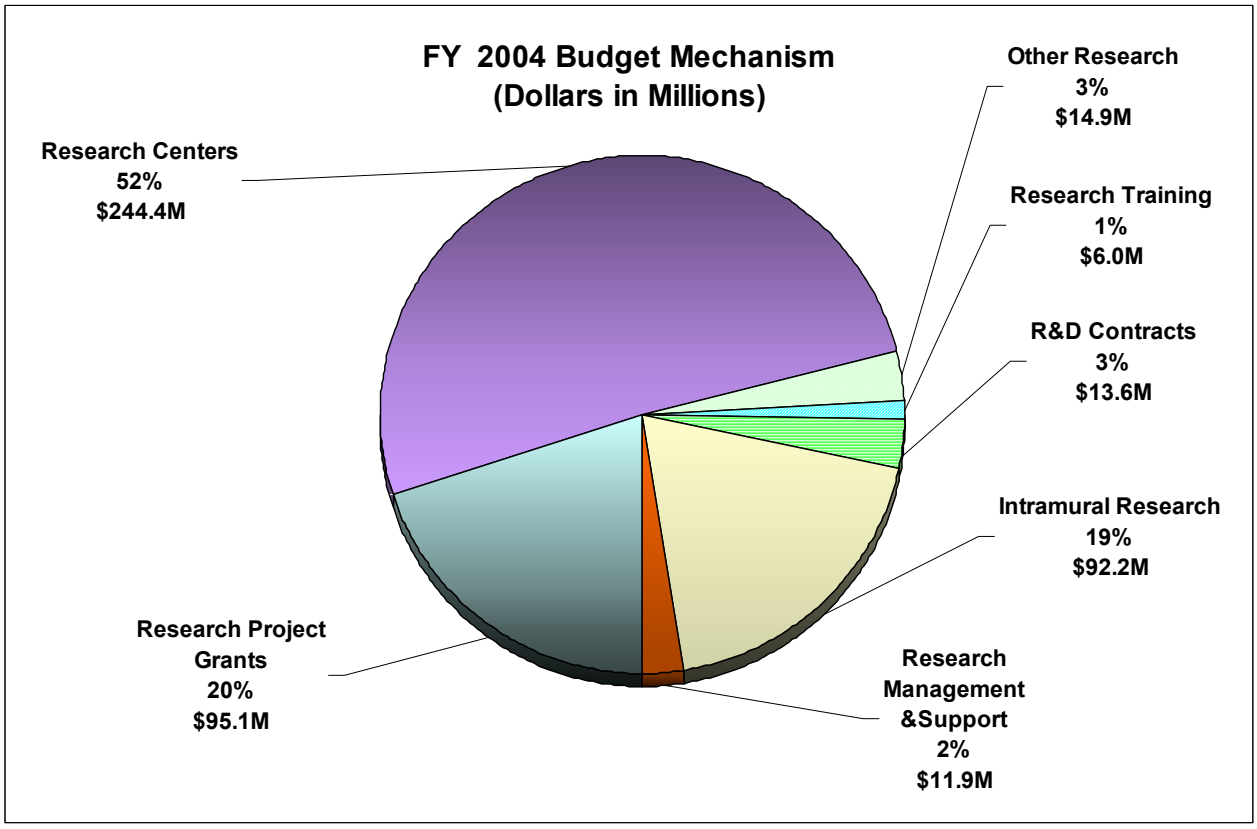


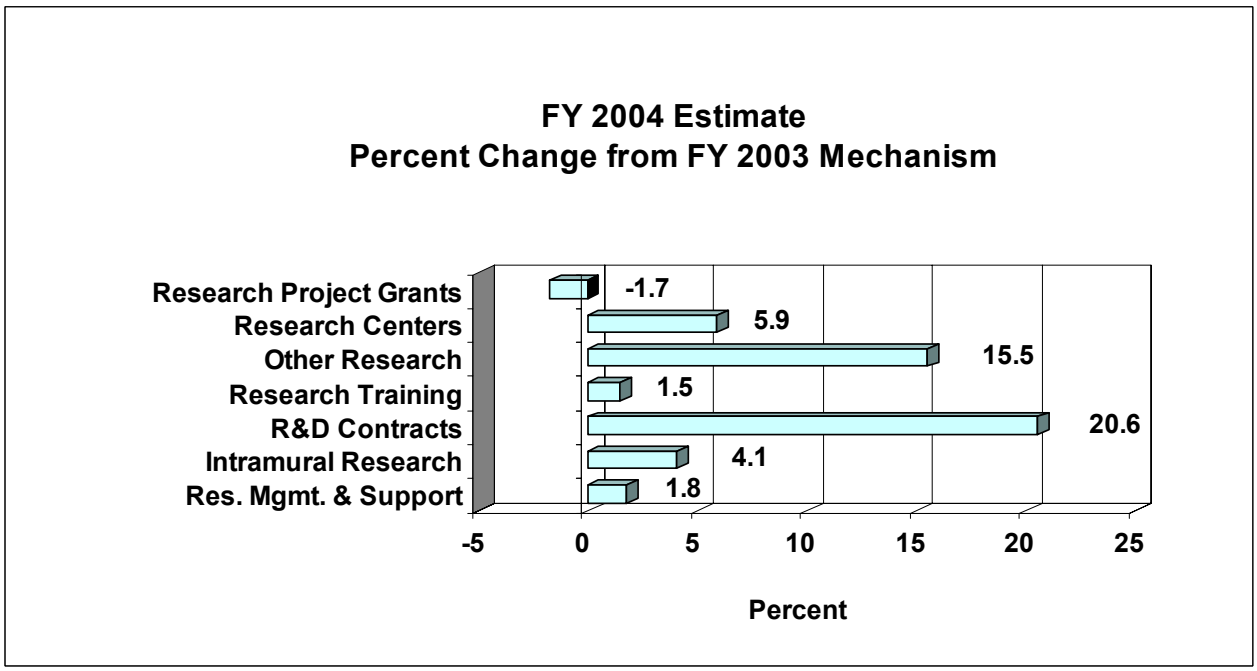
NIH's highest priority is the funding of medical research through research project grants (RPGs). Support for RPGs allows NIH to sustain the scientific momentum of investigator-initiated research while providing new research opportunities. The Fiscal Year 2004 request provides an aggregate average cost increase of -0.9 percent, although average cost increases for competing RPGs are provided at 2.6 percent. Forty-one new and competing grant awards will be made. Also in Fiscal Year 2004, NHGRI will fully fund one Academic Research Enhancement Award (AREA) at a total cost of $155,000.

Promises for advancement in medical research are dependent on maintaining the supply of new investigators with new ideas. In the Fiscal Year 2004 request, NHGRI will support 131 pre- and postdoctoral trainees in full-time training positions, the same number as in Fiscal Year 2003. Stipend levels for NRSA trainees will increase by 4 percent over Fiscal Year 2003 levels for predoctoral fellows, and from 1 to 4 percent, based on years of experience, for postdoctoral fellows.

The Fiscal Year 2004 request includes funding for 37 research centers, 60 other research grants, including 30 research career awards, and 18 R&D contracts. Also included in the request is $150,000 for Best Pharmaceuticals for Children Act studies. Intramural Research and Research Management and Support receive increases of 4.1 and 1.8 percent respectively, over Fiscal Year 2003. These increases will support pay and other related costs. In addition, the Intramural Research program currently has several nationwide recruitment efforts in the areas of cancer genetics, computational biology, and behavioral and social sciences. These new and revised programs will be at the forefront of biomedical research and will fully exploit the knowledge and resources gained from the completion of the human DNA sequence, as well as the other resources provided by the Human Genome Project. These programs will facilitate the translation of genomics to improvements in human health and well-being.

The mechanism distribution by dollars and percent change are displayed below:





NATIONAL INSTITUTES OF HEALTH

National Human Genome Research Institute

Budget Mechanism - Total

MECHANISM	FY 2002 Actual		FY 2003 Amended President's Budget		FY 2004 Estimate	
Research Grants:	No.	Amount	No.	Amount	No.	Amount
Research Projects:						
Noncompeting	72	$51,439,000	75	$60,356,000	70	$55,502,000
Administrative supplements	(21)	5,979,000	(26)	5,533,000	(26)	5,567,000
Competing:						
Full funded	0	0	0	0	1	155,000
Single year	44	24,279,000	38	21,791,000	40	23,523,000
Renewal	5	4,070,000	4	2,294,000	4	2,352,000
New	39	20,209,000	34	19,497,000	36	21,171,000
Supplements	0	0	0	0	0	0
Subtotal, competing	44	24,279,000	38	21,791,000	41	23,678,000
Subtotal, RPGs	116	81,697,000	113	87,680,000	111	84,747,000
SBIR/STTR	36	8,851,000	38	9,079,000	40	10,322,000
Subtotal, RPGs	152	90,548,000	151	96,759,000	151	95,069,000
Research Centers:						
Specialized/comprehensive	22	194,602,000	24	209,867,000	25	220,570,000
Clinical research	0	0	0	0	0	0
Biotechnology	10	19,965,000	11	20,798,000	12	23,800,000
Comparative medicine	0	0	0	0	0	0
Research Centers in Minority Institutions	0	0	0	0	0	0
Subtotal, Centers	32	214,567,000	35	230,665,000	37	244,370,000
Other Research:						
Research careers	27	4,645,000	28	5,126,000	30	5,783,000
Cancer education	0	0	0	0	0	0
Cooperative clinical research	0	0	0	0	0	0
Biomedical research support	0	0	0	0	0	0
Minority biomedical research support	0	0	0	0	0	0
Other	21	9,456,000	26	7,810,000	30	9,161,000
Subtotal, Other Research	48	14,101,000	54	12,936,000	60	14,944,000
Total Research Grants	232	319,216,000	240	340,360,000	248	354,383,000
Research Training:	FTTPs		FTTPs		FTTPs	
Individual awards	17	768,000	10	411,000	14	590,000
Institutional awards	114	5,014,000	121	5,486,000	117	5,394,000
Total, Training	131	5,782,000	131	5,897,000	131	5,984,000
Research & development contracts	17	9,462,000	17	11,307,000	18	13,632,000
(SBIR/STTR)	(0)	(0)	(0)	(0)	(0)	(0)
	FTEs		FTEs		FTEs	
Intramural research	224	82,754,000	222	88,542,000	221	92,176,000
Research management and support	61	10,721,000	65	11,686,000	61	11,897,000
Cancer prevention & control	0	0	0	0	0	0
Construction		0		0		0
Total, NHGRI	285	427,935,000	287	457,792,000	282	478,072,000
(Clinical Trials)		(6,382,000)		(6,829,000)		(7,119,000)

NATIONAL INSTITUTES OF HEALTH
National Human Genome Research Institute

Budget Authority by Activity
(dollars in thousands)

ACTIVITY	FY 2002 Actual		FY 2003 Amended President's Budget		FY 2004 Estimate		Change	
	FTEs	Amount	FTEs	Amount	FTEs	Amount	FTEs	Amount
Extramural Research:								
Human Genome Research		$334,460		$357,564		$373,999		$16,435
Subtotal, Extramural research		334,460		357,564		373,999		16,435
Intramural research	224	82,754	222	88,542	221	92,176	(1)	3,634
Res. management & support	61	10,721	65	11,686	61	11,897	(4)	211
Total	285	427,935	287	457,792	282	478,072	(5)	20,280

NATIONAL INSTITUTES OF HEALTH
National Human Genome Research Institute

Summary of Changes

2003 Amended President's Budget				$457,792,000
2004 Estimated Budget Authority				478,072,000
Net change				20,280,000
	2003 Amended President's Budget Base		Change from Base	
CHANGES	FTEs	Budget Authority	FTEs	Budget Authority
A. Built-in:				
1. Intramural research:				
a. Within grade increase		$23,218,000		346,000
b. Annualization of January 2003 pay increase		23,218,000		180,000
c. January 2004 pay increase		23,218,000		356,000
d. One extra day of pay		23,218,000		90,000
e. Payment for centrally furnished services		15,080,000		302,000
f. Increased cost of laboratory supplies, materials, and other expenses		50,244,000		848,000
Subtotal				2,122,000
2. Research Management and Support:				
a. Within grade increase		6,371,000		110,000
b. Annualization of January 2003 pay increase		6,371,000		49,000
c. January 2004 pay increase		6,371,000		98,000
d. One extra day of pay		6,371,000		24,000
e. Payment for centrally furnished services		727,000		15,000
f. Increased cost of laboratory supplies, materials, and other expenses		4,588,000		59,000
Subtotal				355,000
Subtotal, Built-in				2,477,000

NATIONAL INSTITUTES OF HEALTH
National Human Genome Research Institute

Summary of Changes--continued

CHANGES	2003 Amended President's Budget Base		Change from Base	
	No.	Amount	No.	Amount
B. Program:				
1. Research project grants:				
a. Noncompeting	75	$65,889,000	(5)	($4,820,000)
b. Competing	38	21,791,000	3	1,887,000
c. SBIR/STTR	38	9,079,000	2	1,243,000
Total	151	96,759,000	0	(1,690,000)
2. Research centers	35	230,665,000	2	13,705,000
3. Other research	54	12,936,000	6	2,008,000
4. Research training	131	5,897,000	0	87,000
5. Research and development contracts	17	11,307,000	18	2,325,000
Subtotal, extramural				16,435,000
	FTEs		FTEs	
6. Intramural research	222	88,542,000	(1)	1,512,000
7. Research management and support	65	11,686,000	(4)	(144,000)
8. Cancer control and prevention				
9. Construction				
Subtotal, program		446,106,000		17,803,000
Total changes	287		(5)	20,280,000

NATIONAL INSTITUTES OF HEALTH
National Human Genome Research Institute

Budget Authority by Object

	FY 2003 Amended Pres. Budget	FY 2004 Estimate	Increase or Decrease
Total compensable workyears:			
Full-time employment	287	282	(5)
Full-time equivalent of overtime & holiday hours	2	2	0
Average ES salary	$136,006	$138,726	$2,720
Average GM/GS grade	10.9	10.9	0.0
Average GM/GS salary	$62,774	$64,029	$1,255
Average salary, grade established by act of July 1, 1944 (42 U.S.C. 207)	$72,747	$74,202	$1,455
Average salary of ungraded positions	$80,596	$82,208	$1,612

OBJECT CLASSES		FY 2003 Amended Pres. Budget	FY 2004 Estimate	Increase or Decrease
	Personnel Compensation:			
11.1	Full-Time Permanent	$9,787,000	$9,871,000	$84,000
11.3	Other than Full-Time Permanent	10,611,000	10,995,000	384,000
11.5	Other Personnel Compensation	612,000	626,000	14,000
11.7	Military Personnel	208,000	212,000	4,000
11.8	Special Personnel Services Payments	2,785,000	2,785,000	0
	Total, Personnel Compensation	24,003,000	24,489,000	486,000
12.1	Personnel Benefits	5,492,000	5,637,000	145,000
12.2	Military Personnel Benefits	94,000	96,000	2,000
13.0	Benefits for Former Personnel	0	0	0
	Subtotal, Pay Costs	**29,589,000**	**30,222,000**	**633,000**
21.0	Travel & Transportation of Persons	1,391,000	1,412,000	21,000
22.0	Transportation of Things	142,000	144,000	2,000
23.1	Rental Payments to GSA	1,000	1,000	0
23.2	Rental Payments to Others	470,000	477,000	7,000
23.3	Communications, Utilities & Miscellaneous Charges	584,000	561,000	(23,000)
24.0	Printing & Reproduction	115,000	117,000	2,000
25.1	Consulting Services	222,000	225,000	3,000
25.2	Other Services	2,041,000	2,425,000	384,000
25.3	Purchase of Goods & Services from Government Accounts	50,067,000	55,787,000	5,720,000
25.4	Operation & Maintenance of Facilities	3,388,000	3,439,000	51,000
25.5	Research & Development Contracts	6,013,000	7,002,000	989,000
25.6	Medical Care	552,000	560,000	8,000
25.7	Operation & Maintenance of Equipment	1,608,000	1,632,000	24,000
25.8	Subsistence & Support of Persons	0	0	0
25.0	**Subtotal, Other Contractual Services**	**63,891,000**	**71,070,000**	**7,179,000**
26.0	Supplies & Materials	10,597,000	10,755,000	158,000
31.0	Equipment	7,699,000	7,052,000	(647,000)
32.0	Land and Structures	0	0	0
33.0	Investments & Loans	0	0	0
41.0	Grants, Subsidies & Contributions	343,311,000	356,259,000	12,948,000
42.0	Insurance Claims & Indemnities	0	0	0
43.0	Interest & Dividends	2,000	2,000	0
44.0	Refunds	0	0	0
	Subtotal, Non-Pay Costs	**428,203,000**	**447,850,000**	**19,647,000**
	Total Budget Authority by Object	**457,792,000**	**478,072,000**	**20,280,000**

NATIONAL INSTITUTES OF HEALTH
National Human Genome Research Institute

Salaries and Expenses

OBJECT CLASSES	FY 2003 Amended Pres. Budget	FY 2004 Estimate	Increase or Decrease
Personnel Compensation:			
Full-Time Permanent (11.1)	$9,787,000	$9,871,000	$84,000
Other Than Full-Time Permanent (11.3)	10,611,000	10,995,000	384,000
Other Personnel Compensation (11.5)	612,000	626,000	14,000
Military Personnel (11.7)	208,000	212,000	4,000
Special Personnel Services Payments (11.8)	2,785,000	2,785,000	0
Total Personnel Compensation (11.9)	**24,003,000**	**24,489,000**	**486,000**
Civilian Personnel Benefits (12.1)	5,492,000	5,637,000	145,000
Military Personnel Benefits (12.2)	94,000	96,000	2,000
Benefits to Former Personnel (13.0)	0	0	0
Subtotal, Pay Costs	**29,589,000**	**30,222,000**	**633,000**
Travel (21.0)	1,391,000	1,412,000	21,000
Transportation of Things (22.0)	142,000	144,000	2,000
Rental Payments to Others (23.2)	470,000	477,000	7,000
Communications, Utilities and Miscellaneous Charges (23.3)	584,000	561,000	(23,000)
Printing and Reproduction (24.0)	115,000	117,000	2,000
Other Contractual Services:			
Advisory and Assistance Services (25.1)	222,000	225,000	3,000
Other Services (25.2)	2,041,000	2,425,000	384,000
Purchases from Govt. Accounts (25.3)	20,737,000	23,922,000	3,185,000
Operation & Maintenance of Facilities (25.4)	3,388,000	3,439,000	51,000
Operation & Maintenance of Equipment (25.7)	1,608,000	1,632,000	24,000
Subsistence & Support of Persons (25.8)	0	0	0
Subtotal Other Contractual Services	**27,996,000**	**31,643,000**	**3,647,000**
Supplies and Materials (26.0)	3,815,000	3,884,000	69,000
Subtotal, Non-Pay Costs	34,513,000	38,238,000	3,725,000
Total, Administrative Costs	**64,102,000**	**68,460,000**	**4,358,000**

NATIONAL INSTITUTES OF HEALTH
National Human Genome Research Institute
Significant Items in Senate Appropriations Committee Reports

The following section represents FY 2003 Congressional requirements for reports and significant items derived from Senate Report 107-216. These actions discussed below are contingent on inclusion of similar language and funding in the final FY 2003 appropriation and related reports. Additional items may be transmitted at a later date as a result of the final Conference report.

Item

Behavioral research **–** Recent research has revealed that different genes can be turned on or turned off at different points in a person's life. Understanding what events or behaviors influence gene expression is an important frontier of scientific knowledge. The Committee encourages the NHGRI to develop collaborations with other Institutes and the Office of Behavioral and Social Sciences Research to support integrative research aimed at understanding the role of environmentally induced gene expression in the course of disease and in the promotion of health.

Action taken

The NHGRI sees integrative research that advances the understanding of how environmental factors affect gene expression in health and disease as an area of great importance. Because of recent advances in genomics that have resulted from the availability of the almost finished sequence of the human genome, this research has great promise. Along with the NIH's Office of Behavioral and Social Sciences Research (OBSSR), the NHGRI convened a workshop on October 31, 2001 titled "Studying Interactions among Social, Behavioral, and Genetic Factors in Health," and subsequently established the trans-NIH Working Group on Interactions among Genetic, Behavioral, and Social Factors. Indeed, both the OBSSR and NHGRI see this working group as so important to their programs that it is co-chaired by the Director of the OBSSR (Dr. Raynard Kington) and the Deputy Director of the NHGRI (Dr. Alan Guttmacher). The working group includes representatives from numerous NIH institutes.

Item

Epilepsy **–** The Committee encourages the Institute to continue to intensify its efforts to identify epilepsy genes for the more than 40 different types of epilepsy, and to assist the NINDS in the search for a genetic fingerprint diagnostic test aimed at improving drug therapy for epilepsy. The Committee further encourages the Institute to coordinate efforts with the NINDS to create a national consortium to identify new epilepsy susceptibility genes through a large-scale genotype: phenotype screen. The Committee urges the Institute to continue to make epilepsy research a priority and to coordinate research efforts with other Institutes through the Interagency Epilepsy Coordinating Committee.

Actions taken

The NHGRI is committed to understanding the causes of all forms of diseases, including epilepsy, and to developing effective therapies for them. The NHGRI coordinates its work on epilepsy research with the NINDS, the lead NIH institute working on epilepsy research, as well as with other institutes as needed.

The NHGRI funds two research projects related to epilepsy, through its Human Development Section in its Division of Intramural Research. The first has been open for several years; the second is in the developmental stage. The first project relates to the Pallister-Hall syndrome (PHS), a rare disorder that includes, among its features, a specific brain malformation that predisposes to epilepsy. The same brain malformation can also occur as an isolated malformation without the other signs of PHS. Individuals with PHS generally have seizures that respond to treatment, whereas in individuals with the same brain malformation but without the other features of PHS, the seizures are severe, generally not responsive to treatment, and lead to severe disability. The Human Development Section is performing clinical and molecular research to understand how this brain malformation causes epilepsy, to discover why its effects are so variable, and to find effective treatments.

The second project is under development in conjunction with the Epilepsy Research Branch of the NINDS intramural division. NHGRI scientists have identified a large kindred with an apparently novel form of inherited epilepsy. A clinical protocol has been written and approved to begin the study of this family. Discovering the cause of this inherited form of epilepsy should shed light on genetic pathways that can cause epilepsy and provide possible targets for drug discovery.

The Institute is also overseeing the US participation in the International HapMap Project along with 17 other NIH institutes. This project aims to develop the human haplotype map, the HapMap, to speed the genetic understanding of diseases such as epilepsy. The HapMap will be a set of about 500,000 genetic variants that describe the patterns of genetic variation (haplotypes) in the genome. Researchers will then use the HapMap to find regions of the genome that are associated with complex diseases. The HapMap will be a powerful resource for finding the genetic factors contributing to these diseases. Researchers will then be able to study the disease process and how the genetic factors interact with environmental factors such as injury or diet, as well as drugs.

Dr. Leslie G. Biesecker, MD, a Senior Investigator in our Genetic Disease Research Branch and Head of the Human Development Section, serves as the NHGRI's representative to the Interagency Epilepsy Coordinating Committee.

Item

Privacy – The Committee remains concerned about the proper use of genetic information and encourages the NHGRI's ongoing efforts, though its ELSI program, to examine the privacy and fair use of genetic information. Other important issues related to human genetics research and its consequences should also be studied, including: the appropriate use of genetic tests; the protection of human subjects who participate in genetic research; the development of policies to

guide research into genetic variation; and complex social issues, such as how genetics informs concepts of race and ethnicity.

Actions taken

Since the early 1990s the ELSI program at the NHGRI has funded research in the areas of privacy and fair use of genetic information, and the Institute remains keenly interested in these issues from many perspectives. The Institute also funded several other research projects on the related topics mentioned.

1. The appropriate use of genetic tests is an ongoing priority for ELSI research and one of four program areas around which the portfolio of ELSI research is organized. The ELSI program has supported more than 60 research projects examining these issues. As an example, NHGRI is currently funding a project looking at the issues surrounding the use of genetic testing to determine possible susceptibility to the development of Alzheimer's disease. The data generated by this study will assist in the development of guidelines for clinicians for genetic testing, risk assessment and appropriate counseling scenarios.

2. The protection of human subjects who participate in genetic research is another ongoing priority for the ELSI research program and one of the four ELSI program areas. The ELSI program has supported close to 20 projects exploring these issues. For example, the NHGRI is funding an ongoing study to determine how the prospect of direct benefit to research subjects in gene transfer research (GTR) (usually called "gene therapy") is understood and discussed by research subjects, investigators, study coordinators, and IRBs, and explained in consent forms. Based on these findings, the researchers hope to develop an improved policy standard for the presentation of benefits in GTR specifically and clinical research generally.

3. The development of policies to guide research into genetic variation is a new area of emphasis for the ELSI program.

Seven applications to the 1999 RFA on Ethical, Legal, and Social Implications of Research into Human Genetic Variation were funded in 2000. A consortium of these studies has been organized in order to allow researchers to compare findings on common issues, reduce duplication of effort, and promote sharing of information.

Four peer reviewed papers detailing some of the preliminary findings of the studies were published in 2001. Also in FY 2001, two other related applications were funded and became part of the Genetic Variation Consortium. These projects include: (1) a study of African American community review of genetics research; and (2) a study on concepts of race and ethnicity in genetics research. The NIDCD, NIEHS, and NIGMS are participating in this RFA and have provided co-funding for four of the grants.

A new RFA, "Studies of the Ethical, Legal, and Social Implications (ELSI) of Human Genetic Variation Research for Individuals and Diverse Racial and Ethnic Groups," was released in November 2001. These applications will be reviewed in December 2002 and funded in the spring of 2003.

4. How genetics informs concepts of race and ethnicity is of continued importance for NHGRI. This was further emphasized in 1998 when it was identified as one of five goals for ELSI research.

The issues surrounding genetics, race, and ethnicity are currently being addressed through a number of the studies funded as part of the genetic variation consortium. This issue is also one of the specific research questions to be addressed by the applications that will be funded under the new genetic variation RFA. An example of a study currently being funded by NHGRI in this area that is not part of the variation research consortium is a project by Seymour Garte at the University of Medicine and Dentistry of New Jersey, looking at the implications of genomics research on racial definition. The specific aims of his project are to develop scientific and philosophical tools to address and clarify these issues, to better define the categorization of humans according to DNA sequence characteristics, and to provide more useful categories with respect to human genetic diversity as applied to biomedical research and clinical applications.

In August, 2002, to shepherd and coordinate the institute's efforts regarding race and genetics, the NHGRI selected a senior level consultant, Vence Bonham, J.D., who is an authority regarding issues of race, ethnicity, and genetics and health disparities, to join its Office of the Director as Senior Advisor for Minority Health and Health Disparities.

Item

Type 1 Diabetes Genetics Consortium – The Committee is aware of the development of a Type 1 and Type 2 Diabetes Genetics Consortium which will collect and share valuable DNA information from juvenile diabetes patients from studies around the world. The Committee encourages the NHGRI to collaborate with the NIDDK in efforts to determine the genetic origins of juvenile diabetes by directing resources towards this important initiative.

Actions taken or to be taken

The NHGRI continues to collaborate with the NIDDK especially in the area of research on diabetes. Indeed, the institute sees this work as so important that its Deputy Director serves as its liaison to the Type 1 Diabetes Genetics Consortium. The NHGRI also continues to support two major diabetes projects conducted by investigators in its Division of Intramural Research and their collaborators. Both studies aim to identify genetic variations that lead to increased susceptibility to Type 2 diabetes mellitus. The institute funds one extramural research project exploring genetic factors in the etiology of Type 1 diabetes. NHGRI has also recently funded a new Center of Excellence in Genomic Science (CEGS) at the University of Washington, which aims to: (1) develop tools for studying natural genetic variation; and (2) apply those tools to develop an improved understanding of the molecular basis of genetic susceptibilities to Type 1 diabetes as well as other complex genetic disorders.

Just as for many other complex genetic disorders, the Institute's HapMap Project will play a major role in helping to better understand Type 1 and Type 2 Diabetes. The HapMap will help assist the scientific community identify the genes that play a significant role in diabetes.

Another resource that the NHGRI supports, which will enable further understanding of Type 1

and Type 2 Diabetes, is the Mammalian Gene Collection (MGC). The MGC seeks to identify and sequence a representative full open reading frame (ORF) clone for each human and mouse gene. Over the past year the MGC has produced over 80 cDNA libraries enriched for full-length cDNAs derived from human tissue and cell lines, and mouse tissue. As of February 2001, a non-redundant set of more than 20,000 human and mouse putative full ORF clones has been assembled. These clones are highly accurate and all the resources are publicly accessible to the biomedical research community. This resource allows the scientific community to study particular genes, such as those associated with diabetes in great detail.

NATIONAL INSTITUTES OF HEALTH
National Human Genome Research Institute

Authorizing Legislation

	PHS Act/ Other Citation	U.S. Code Citation	2003 Amount Authorized	2003 Amended President's Budget	2004 Amount Authorized	2004 Budget Estimate
Research and Investigation	Section 301	42§241	Indefinite	$451,895,000	Indefinite	$472,088,000
Research Institute	Section 41B	42§285b	Indefinite		Indefinite	
National Research Service Awards	Section 487(d)	42§288	a/	5,897,000	b/	5,984,000
Total, Budget Authority				457,792,000		478,072,000

a/ Amounts authorized by Section 301 and Title IV of the Public Health Act.
b/ Reauthorizing legislation will be submitted.

NATIONAL INSTITUTES OF HEALTH
National Human Genome Research Institute

Appropriations History

Fiscal Year	Budget Estimate to Congress	House Allowance	Senate Allowance	Appropriation 1/
1995 2/	$152,010,000	$151,878,000	$151,878,000	$151,518,000 3/
Rescission				(331,000)
1996	166,678,000 2/	170,041,000	163,943,000 2/	169,041,000
Rescission				(266,000)
1997	177,788,000 2/	189,267,000	180,807,000 2/	189,657,000 4/
1998	202,197,000 2/	211,772,000	218,851,000	217,704,000
1999	236,275,000 2/5/	246,111,000	249,891,000	264,892,000
Rescission				(185,000)
2000	271,536,000 2/	308,012,000	337,322,000	337,322,000
Rescission				(1,795,000)
2001	353,427,000 2/	386,410,000	385,888,000	382,384,000
Rescission				(192,000)
2002	426,739,000 2/	423,454,000	440,448,000	429,515,000
Rescission				(757,000)
2003	458,182,000			
2004	478,072,000			

1/ Reflects enacted supplementals, rescissions, and reappropriations.
2/ Excludes funds for HIV/AIDS research activities consolidated in the NIH Office of AIDS Research.
3/ Excludes enacted administrative reductions of $ 161,000
4/ Excludes enacted administrative reductions of $ 128,000
5/ Excludes reductions of $721,000 for the budget amendment for bioterrorism.

NATIONAL INSTITUTES OF HEALTH
National Human Genome Research Institute

Detail of Full-Time Equivalent Employment (FTEs)

OFFICE/DIVISION	FY 2002 Actual	FY 2003 Amended Pres. Budget	FY 2004 Estimate
Office of the Director	7	7	6
Office of Administrative Management	15	16	14
Office of Policy, Planning and Communications	11	12	11
Division of Intramural Research	224	222	221
Division of Extramural Research	28	30	30
Total	285	287	282
FTEs supported by funds from Cooperative Research and Development Agreements	(2)	(2)	(2)

FISCAL YEAR	Average GM/GS Grade
2000	10.8
2001	10.7
2002	10.9
2003	10.9
2004	10.9

NATIONAL INSTITUTES OF HEALTH
National Human Genome Research Institute

Detail of Positions

GRADE	FY 2002 Actual	FY 2003 Amended Pres. Budget	FY 2004 Estimate
ES-6			
ES-5			
ES-4			
ES-3			
ES-2		1	1
ES-1			
Subtotal	0	1	1
Total - ES Salary	$0	$136,006	$138,726
GM/GS-15	20	18	18
GM/GS-14	14	15	15
GM/GS-13	26	26	26
GS-12	37	37	37
GS-11	23	23	21
GS-10	2	2	2
GS-9	25	28	26
GS-8	12	12	13
GS-7	16	15	15
GS-6	9	8	8
GS-5	4	4	4
GS-4	1	1	0
GS-3	1	1	0
GS-2			
GS-1			
Subtotal	190	190	185
Grades established by Act of July 1, 1944 (42 U.S.C. 207):			
Assistant Surgeon General			
Director Grade	1	1	1
Senior Grade	2	2	2
Full Grade			
Senior Assistant Grade			
Assistant Grade			
Subtotal	3	3	3
Ungraded	114	115	115
Total permanent positions	155	158	154
Total positions, end of year	307	309	304
Total full-time equivalent (FTE) employment,end of year	285	287	282
Average ES level	--	ES-2	ES-2
Average ES salary	$0	$136,006	$138,726
Average GM/GS grade	10.9	10.9	10.9
Average GM/GS salary	$60,887	$62,774	$64,029