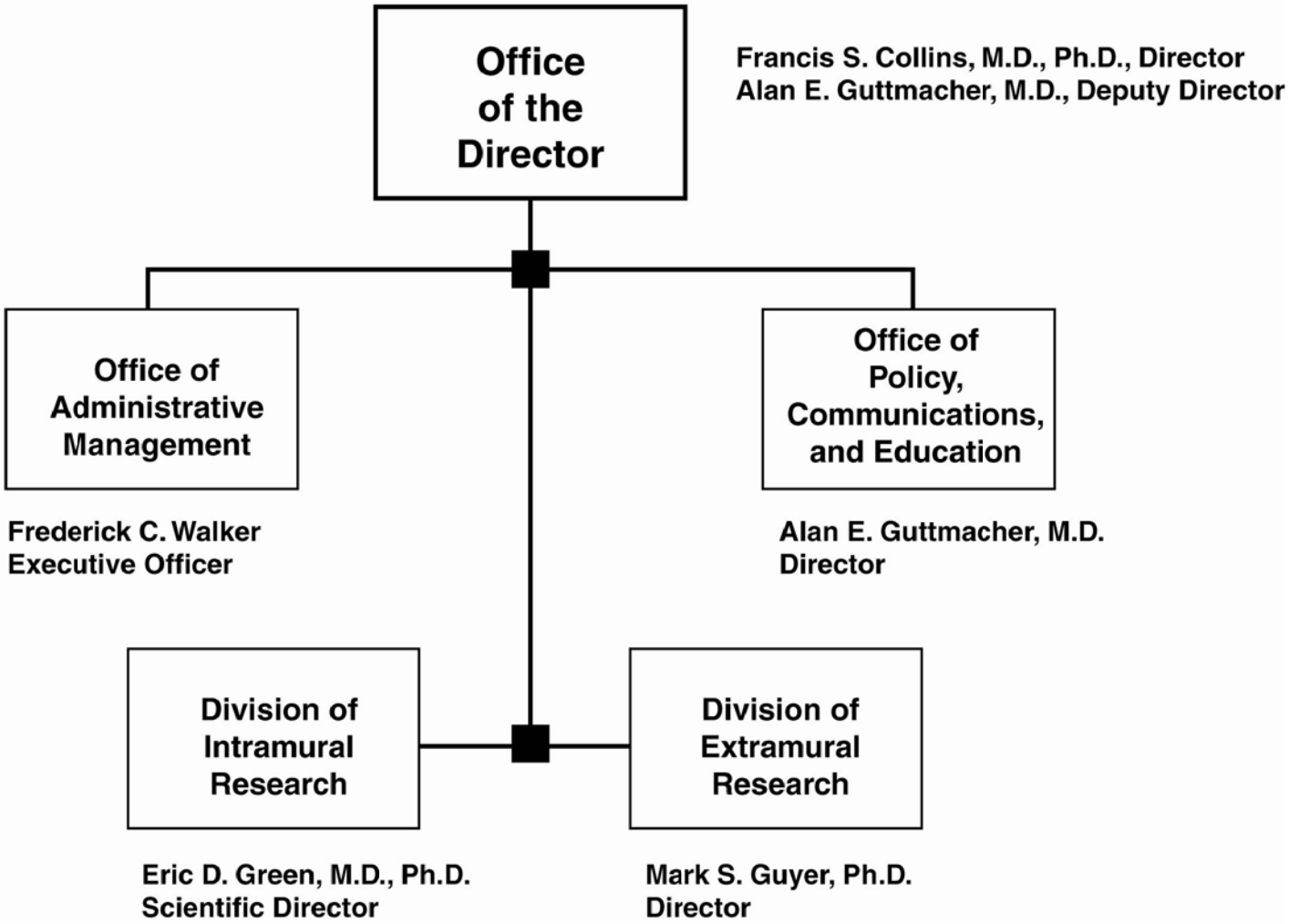
DEPARTMENT OF HEALTH AND HUMAN SERVICES

NATIONAL INSTITUTES OF HEALTH

National Human Genome Research Institute

FY 2006 Budget — Page No.

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, [$492,670,000] *$490,959,000*.

[Departments of Labor, Health and Human Services and Related Agencies Appropriations Act, as enacted by the Consolidated Appropriations Act for Fiscal Year 2005]

National Institutes of Health
National Human Genome Research Institute

Amounts Available for Obligation 1/

Source of Funding	FY 2004 Actual	FY 2005 Appropriation	FY 2006 Estimate
Appropriation	$482,222,000	$492,670,000	$490,959,000
Enacted Rescissions	(3,149,000)	(4,062,000)	0
Subtotal, Adjusted Appropriation	479,073,000	488,608,000	490,959,000
Real transfer under NIH Director's one-percent transfer authority to other ICs	11,523,000	0	0
Comparative transfer to NIBIB for Radiology Program	(62,000)	0	0
Comparative transfer to Buildings and Facilities	(183,000)	0	0
Comparative transfer to/from other NIH ICs for NIH Roadmap	(11,523,000)	0	0
Subtotal, adjusted budget authority	478,828,000	488,608,000	490,959,000
Unobligated Balance, start of year	0	0	0
Unobligated Balance, end of year	0	0	0
Subtotal, adjusted budget authority	478,828,000	488,608,000	490,959,000
Unobligated balance lapsing	(49,000)	---	---
Total obligations	478,779,000	488,608,000	490,959,000

1/ Excludes the following amounts for reimbursable activities carried out by this account:
FY 2004 - $4,342,000 FY 2005 -$4,350,000 FY 2006 - $4,350,000
Excludes $233,000 in FY 2004 and $134,000 in FY 2005 for royalties.

Justification

National Human Genome Research Institute

Authorizing Legislation: Section 301 of the Public Health Service Act, as amended.

Budget Authority:

FY 2004 Actual		FY 2005 Appropriation		FY 2006 Estimate		Increase or Decrease	
FTEs	BA	FTEs	BA	FTEs	BA	FTEs	BA
273	$478,828,000	285	$488,608,000	289	$490,959,000	+4	+$2,351,000

This document provides justification for the Fiscal Year 2006 activities of the National Human Genome Research Institute, including HIV/AIDS activities. A more detailed description of NIH-wide Fiscal Year 2006 HIV/AIDS activities can be found in the NIH section entitled "Office of AIDS Resaerch (OAR)."

INTRODUCTION

Since the completion of the Human Genome Project in April 2003, the National Human Genome Research Institute (NHGRI) has moved forward into the genomic era with a wide range of exciting new research initiatives aimed at improving human health. These scientific projects include The International HapMap Consortium, ENCODE (the ENCyclopedia Of DNA Elements), and a chemical genomics initiative. At the same time, the NHGRI has enhanced its Ethical Legal and Social Implications (ELSI) program, including the establishment of an extramural Centers of Excellence in ELSI Research (CEER) program and has created a new branch, the Social and Behavioral Research Branch, within its intramural program. The NHGRI also continues to play an active role in the public policy debate surrounding issues of genomics and genetics, and to provide informational and educational resources to health care professionals, those with genetic diseases, students, and the general public.

Although the original goals of the Human Genome Project are complete, much work is still needed to make genomic medicine a reality for the American public and the world. The NHGRI is keenly aware that, whether regarding science, medicine, policy, or ELSI, it cannot effectively do its part of that work alone. We rely heavily on partnerships with many federal agencies, as well as organizations outside the government. Through the science we fund and the work we do in partnership with others, the NHGRI serves as an active catalyst for improving health for all through the use of genetics and genomics.

Stories of Discovery: Completed Human Genome Sequence

In October 2004, the International Human Genome Sequencing Consortium, led in the United States by the NHGRI and the Department of Energy, published a description of the finished human genome sequence in the journal *Nature*. An international team of collaborators on the Human Genome Project worked to convert the draft genome into a highly accurate form, finishing its work, and thus the Human Genome Project, ahead of schedule and under budget, in April 2003. The paper reports the results of that completion process. This analysis reduces the estimate of the number of human protein-coding genes from 35,000 to only 20,000-25,000 - a surprisingly low number for our species, considering that only a decade ago most scientists thought there would be over 100,000 genes.

The *Nature* publication provides rigorous scientific evidence that the genome sequence produced by the Human Genome Project has both the high coverage and accuracy needed to perform sensitive analyses, such as identifying the number of genes, the segmental duplications involved in disease, and the "birth" and "death" of genes over the course of recent evolution. The assessment confirms that the finished sequence now covers more than 99 percent of the euchromatic (or gene-containing) portion of the human genome and was sequenced to an accuracy of 99.999 percent, which translates to an error rate of only 1 base per 100,000 base pairs — 10 times more accurate than the original goal.

The contiguity of the sequence is also massively improved in the finished sequence. The average DNA letter now sits on a stretch of 38.5 million base pairs of uninterrupted, high-quality sequence — about 475 times longer than the 81,500 base-pair stretch that was available in the working draft just three years earlier. Access to uninterrupted stretches of sequenced DNA can greatly assist researchers hunting for genes and the neighboring DNA sequences that may regulate their activity, dramatically cutting the effort and expense required to find regions of the human genome that may contain subtle variants involved in disease.

"Finished" doesn't mean that the human genome sequence is perfect. There still remain 341 gaps in the finished human genome sequence, in contrast to the 150,000 gaps in the working draft announced in June 2000. The technology now available cannot readily close these remaining gaps in the human genome sequence. Closing those gaps will require more research and new technologies, rather than industrial-scale efforts like those employed by the Human Genome Project.

More than 2,800 researchers took part in the International Human Genome Sequencing Consortium. Even more detailed annotations and analyses have already been published for chromosomes 5, 6, 7, 9, 10, 13, 14, 19, 20, 21, 22 and Y. Publications describing the remaining 12 chromosomes are forthcoming. The genome sequence will serve as a firm foundation for biomedical research into health and disease for centuries to come.

International Human Genome Sequencing Consortium, Finishing the Euchromatic Sequence of the Human Genome. Nature 431: 931–945, 2004.

ENCyclopedia Of DNA Elements (ENCODE)

While the completion of the Human Genome Project in April 2003 and the publication of the finished human genome sequence in the October 2004 issue of *Nature* marked significant scientific achievements, this is only the end of the beginning toward the ultimate goal of using information about the human genome sequence to diagnose, treat, and prevent disease. The next logical step is to characterize the genes and the genetic elements that regulate gene expression, DNA replication, and chromosome structure. There is strong evidence that little understood parts of the genome have important functions, but scant information exists about where these other "functional elements" are located and how they work. With the goal of identifying the precise location and function of all sequence-based functional elements in the genome, the NHGRI launched the ENCyclopedia Of DNA Elements (ENCODE) project in the fall of 2003. Grants were awarded for the pilot phase, calling for application of existing technologies to a carefully chosen 30 megabases (roughly 1%) of the human genome. The pilot phase of the project is organized as an international consortium of computational and laboratory-based scientists and is open to all investigators who agree to abide by the project's criteria and guidelines for participation.

Many opportunities exist to encourage innovation in this area. Therefore, at the same time that high-throughput efforts were initiated using well-developed technologies, NHGRI launched a parallel effort to develop new technologies. The first set of grants to expand the repertoire of tools that can be applied to ENCODE or similar future projects was awarded in 2003. A second set of technology development grants was awarded in September 2004. It is envisioned that when an extensive "tool box" of technologies is available, it will be possible to annotate the entire human genome with information that will serve as a platform for more in-depth, detailed studies of biological function.

A manuscript describing the ENCODE project appeared in the October 22 issue of *Science*, detailing the scientific rationale and strategy behind the quest to produce a comprehensive catalog of all parts of the human genome crucial to biological function. The strategy, as laid out in the paper, calls for the creation of a genome "parts list" of all sequence-based functional elements in the human DNA sequence, including all protein-coding genes, non-protein-coding genes, regulatory elements involved in the control of gene transcription, and DNA sequences that mediate chromosomal structure and dynamics.

Many functional elements are only active in certain types of cells or at certain stages of development, which means it may be necessary to analyze many different types of human cells. In addition, to create a truly comprehensive inventory, more work needs to be done to learn about functional elements not surveyed in the pilot project, including centromeres (the central construction of each chromosome, which is critical for replication) and telomeres (the ends of chromosomes). In their *Science* article, ENCODE researchers set forth their plans for addressing these and other challenges.

NHGRI has designated the ENCODE project as a community resource project, which means that all data generated for this project will be deposited in free, public databases as soon as they are experimentally verified.

The ENCODE Project Consortium, The ENCODE (ENCyclopedia Of DNA Elements) Project. Science 306: 636–640, 2004.

Use of Comparative Genomics to Understand the Human Genome

The availability of the genome sequences of the human, the mouse, the rat and a wide variety of other organisms – from yeast to dogs to chimpanzees – is driving the development of an exciting new field of biological research, comparative genomics. By comparing the finished reference sequence of the human genome with genomes of other organisms, researchers can identify regions of similarity and difference, and thus better understand the structure and function of human genes in order to develop new strategies to combat human disease.

Comparative genomics can point researchers toward the signals that control gene function, which in turn should lead to innovative approaches for treating human disease and improving human health. In addition, the comparison to other animals may prove extremely helpful in understanding disease susceptibility. For example, chimpanzees do not suffer from some of the diseases that strike humans, such as malaria and AIDS, even though the chimp DNA sequence is 98.8 percent identical to ours. A comparison of the sequences of genes involved in disease susceptibility may reveal the reasons for this species barrier, thereby suggesting new pathways for prevention of human disease.

The successful sequencing of the human genome and the recent draft assemblies of the mouse and rat genomes demonstrate that large-scale sequencing projects can generate high-quality data at a reasonable cost. As a result, interest in sequencing the genomes of many other organisms has risen dramatically. Other organisms whose genomes have been sequenced include: two fruit flies (*Drosophila melanogaster* and *D. pseudoobscura*); two roundworms (*Caenorhabditis elegans* and *C. briggsae*); yeast (*Saccharomyces cerevisiae*) and several other fungi; a malaria-carrying mosquito (*Anopheles gambiae*), along with a malaria-causing parasite (*Plasmodium falciparum*); two sea squirts (*Ciona savignyi* and *C. intestinalis*); a long list of microbes; and a couple of plants, including mustard weed (*Arabidopsis thaliana*) and rice (*Oryza sativa*). Recently, the genomes of the honeybee (*Apis mellifera*), the chimpanzee (*Pan troglodytes*), the cow (*Bos taurus*), the dog (*Canis familiaris*), and the chicken (*Gallus gallus*) have also been sequenced. All of these genomes were chosen for sequencing because of the insights they can provide for human health.

Sequencing of the Rat, Bovine, Cow, Chicken and Honey Bee Genomes

Rat

In landmark a paper published in the April 1 2004 issue of the journal *Nature*, the Rat Genome Sequencing Project Consortium describe its efforts to produce and analyze a draft sequence of the Brown Norway strain of the laboratory rat (*Rattus norvegicus*). The project, led by the Human Genome Sequencing Center at Baylor College of Medicine in Houston, was funded primarily by the National Heart, Lung, and Blood Institute (NHLBI), which contributed $58.5 million dollars of support, and the NHGRI, which contributed $60 million.

Areas in which rat models have already advanced medical research include: cardiovascular diseases (hypertension); psychiatric disorders (studies of behavioral intervention and addiction); neural regeneration; diabetes; surgery; transplantation; autoimmune disorders (rheumatoid arthritis); cancer; wound and bone healing; and space motion sickness. In drug development, the rat is routinely employed to assess therapeutic efficacy and toxicity of drug compounds prior to human clinical trials. The rat genome sequence will facilitate all of these studies.

Rat Genome Sequencing Project Consortium, Genome Sequence of the Brown Norway Rat Yields Insights into Mammalian Evolution. Nature 428: 493–521, 2004.

Bovine

Sequencing of the bovine genome began in December 2003. In addition to helping medical researchers learn more about the human genome and thereby develop better ways of treating and preventing disease, the bovine genome sequence will serve as a tool for agricultural researchers striving to improve health and disease management of cattle and enhance the nutritional value of beef and dairy products.

The Hereford breed, which is used in beef production, was selected for the bulk of the sequencing project. Researchers plan to have a six-fold draft of the bovine genome completed in the first half of 2005. They are also comparing the bovine genome sequence with those of the human and other organisms that have already been sequenced. Results of these analyses will begin to be placed in public databases in the next several months.

Dog

Sequencing of the dog genome began in June 2003. NHGRI provided about $30 million in funding for the project to the Broad Institute, which is part of NHGRI's Large-Scale Sequencing Research Network. The first draft of the dog genome sequence has been deposited into free public databases for use by biomedical and veterinary researchers around the globe. The breed sequenced was the boxer, chosen after analyses of 60 breeds found it was one of those with the least variation in its genome and thus likely to provide the most reliable reference genome sequence. The initial assembly is based on seven-fold coverage of the dog genome. The NHGRI-supported researchers are currently comparing the dog and human genome sequences and plan to publish results of their analysis in the next several months.

Chicken

Sequencing of the chicken genome, the first avian genome to be sequenced, began in March 2003. NHGRI provided about $13 million for the project. The assembled genome of the red jungle fowl, *Gallus gallus*, which is the ancestor of domestic chickens, was deposited into public databases in 2004. The assembly is based on seven-fold sequence coverage of the chicken genome. To facilitate comparative genomic analysis, researchers also have aligned the draft version of the chicken sequence with the human sequence.

Recent outbreaks of avian flu have accelerated scientists' interest in learning more about the chicken genome and how genetic variation plays a role in the susceptibility of different strains to the disease. In addition to its tremendous economic value as a source of eggs and meat, the chicken is widely used in biomedical research. It is an important model in the study of embryology and development, as well as in research into the connection between viruses and some types of cancer. The chicken also is well positioned evolutionarily to provide an intermediate perspective between mammals, such as humans, and lower vertebrates, such as fish.

Honey Bee

Sequencing of the honey bee genome began in early 2003. NHGRI provided about $6.9 million in funding for the project and the U.S. Department of Agriculture contributed $750,000. Farmers value the honey bee for its ability to produce honey and pollinate crops. Besides its importance in agriculture, the honey bee serves as a model organism for studying human health issues, including immunity, allergic reaction, antibiotic resistance, development, mental health, longevity, and diseases of the X chromosome. Biologists also are interested in the honey bee's social instincts and behavioral traits.

Scientists are particularly interested in comparing the honey bee's genome with previously sequenced insect genomes, such as those of the fruit fly and mosquito, as well as with sequences from Africanized bee strains that have invaded many areas of the southern United States.

SCIENCE ADVANCES

HapMap

Background – All diseases have a hereditary component, but for most common diseases like diabetes, heart disease, and mental illness, the gene variants responsible for the increased risk have been difficult to identify. What is needed to solve this posing problem is a way of scanning across large regimes of chromosomes to find the spelling variants (which scientists call SNPs, or single nucleotides polymorphisms) that increase risk. Searching for these variants one-by-one is currently prohibitively difficult and expensive. But recent data indicates that SNPs are organized into "neighborhoods", referred to by geneticists as haplotypes. Understanding how genetic variation is inherited in DNA haplotypes can provide considerable savings in time, effort, and cost in uncovering hereditary factors in disease. NHGRI has taken a leadership role in the development of the HapMap (haplotype map), a catalog of haplotype blocks and the SNPs that tag them. Researchers can use the HapMap to find the genes and variants that contribute to many diseases; in addition, it will be a powerful resource for studying the genetic factors contributing to variation in individual response to disease, to drugs, and to vaccines.

Advance – To build the HapMap, a deep catalog of SNPs was needed. The HapMap consortium had initially planned to identify an additional 3 million new SNPs to fill in areas where the density of SNPs in public databases was insufficient, but due to advances in technology the project has now identified more than 6 million new SNPs, for a total of 9 million. Now these SNPs can be typed on carefully chosen DNA samples to define the haplotype "neighborhoods". The consortium has completed the collection of samples and consent from 270 individuals from four populations (CEPH, [U.S. residents with ancestry from Western and Northern Europe] Yoruba in Nigeria, Han Chinese in Beijing, and Japanese in Tokyo). Eight research groups have performed genotyping for 890,916 SNPs in these samples, as of September 2004. A newly issued grant to Perlegen Sciences Inc. will lead to a nearly fivefold increase in SNP density across the genome on the HapMap samples by the summer of 2005. The consortium is also developing innovative scientific strategies to choose which SNPs to study, to assess the quality of the data, and to derive haplotypes from the SNP data. The goals of the HapMap project are expected to be reached by the end of 2005. More information about HapMap is available at www.hapmap.org.

Implications - The HapMap should reduce the number of SNPs required to examine the entire genome for association with a disease risk from the 10 million SNPs that exist to roughly 500,000 "gold standard" or "tag" SNPs. This will make genome scan approaches to finding regions with genes that affect diseases much more efficient and comprehensive, since funds and effort will not be wasted typing more SNPs than necessary, and all regions of the genome can be included.

In addition to its use in studying genetic associations with disease, the HapMap should be a powerful resource for studying the genetic factors contributing to variation in response to environmental factors, in susceptibility to infection, and in the effectiveness of and adverse responses to drugs and vaccines. All such studies will be based on the expectation that there will be higher frequencies of the contributing genetic components in a group of people with a disease

or a particular response to a drug, vaccine, pathogen, or environmental factor than in a group of similar people without the disease or response. Using just the tag SNPs, researchers should be able to find chromosome neighborhoods that have different haplotype distributions in the two groups of people. Each region would then be studied in more detail to discover which variants in which genes in the region contribute to the disease or response, leading to more effective interventions. This should also allow the development of tests to predict which drugs or vaccines would be most effective in specific individuals.

Citation - The International HapMap Consortium, Integrating Ethics and Science in the International HapMap Project. Nature Reviews Genetics 5: 467-475, 2004.

Gene Variants May Increase Susceptibility to Type 2 Diabetes

Background - Type 2 diabetes usually begins after age 40, and is more common in overweight, inactive people and those with a family history of diabetes. In the United States, type 2 diabetes disproportionately affects African Americans, Hispanic/Latino Americans, and American Indians. Affecting about 17 million people nationwide, this form of diabetes accounts for 90 to 95 percent of all diabetes cases in the United States. Its prevalence has risen steadily in the past 30 years, and it is increasingly being seen in younger people, even in children. Hallmarks of the disease are insulin resistance - the inability of target tissues to respond to insulin - and a gradual failure of pancreatic beta cells to produce enough insulin.[1]

Advance - International research teams studying two distinct populations have both found variants in a gene that may predispose people to type 2 diabetes, the most common form of the disease.

Homing in on a wide stretch of chromosome 20, flagged by earlier studies as a likely location for a type 2 diabetes susceptibility gene, the teams identified four genetic variants (SNPs) that are strongly associated with type 2 diabetes in Finnish and Ashkenazi Jewish populations. All four SNPs cluster in the regulatory region of a single gene, hepatocyte nuclear factor 4 alpha (*HNF4A*), a transcription factor that acts as a "master switch" regulating the expression of hundreds of other genes. *HNF4A* turns genes on and off in many tissues, including the liver and pancreas. In the beta cells of the pancreas, it influences the secretion of insulin in response to glucose. Recently other research groups have confirmed the result in several other populations, including Mexican-Americans.

Implications - If an individual has the HNF4A variation, it appears to raise their risk of type 2 diabetes about 20 to 30 percent. The variation does not cause diabetes unless an individual has it in combination with other, yet-to-be-identified, genetic susceptibility factors, together with certain environmental influences such as obesity and/or lack of physical exercise. Translating the discovery into a treatment that benefits people with diabetes or those at risk is still years away, since much more still needs to be learned about this gene and how to modulate its function.

[1] http://www.diabetes.org/diabetes-statistics/national-diabetes-fact-sheet.jsp

Scientists have come a long way in understanding the basis for diseases arising from single-gene mutations. Understanding the genetic basis of the more common, polygenic diseases such as diabetes has been much more difficult. But the tools of genomics are beginning to reveal many details about risk of common disease that had previously been unapproachable.

Citation - Silander K, Mohlke KL, Scott LJ, Peck EC, Hollstein P, Skol AD, Jackson AU, Deloukas P, Hunt S, Stavrides G, Chines PS, Erdos MR, Narisu N, Conneely KN, Li C, Fingerlin TE, Dhanjal SK, Valle TT, Bergman RN, Tuomilehto J, Watanabe RM, Boehnke M, Collins FS, Genetic Variation Near The Hepatocyte Nuclear Factor-4 Alpha Gene Predicts Susceptibility To Type 2 Diabetes. Diabetes 53:1141-9, 2004.

Location of Potential Familial Lung Cancer Gene Discovered

Background - Lung cancer is by far the leading cause of cancer death in the United States (over 160,000 deaths expected in 2004), and its five-year survival rate is only 15 percent. Such a high mortality, combined with the very significant role of smoking as a cause of any cancer makes finding histories of familial lung cancer or collecting genetic samples extremely difficult and time consuming. But it is critical to discover why some people are likely to develop lung cancer after only second hand smoke exposure, while others are apparently quite resistant.

Advance - An interdisciplinary consortium consisting of 12 research institutions and universities, including the National Cancer Institute (NCI) and NHGRI, identified a major lung cancer susceptibility region on a segment of chromosome 6. The findings appeared in the September 2004 issue of *American Journal of Human Genetics*.

The group found strong evidence that a lung cancer susceptibility gene(s) is co-inherited with a genetic marker on chromosome 6. Markers on chromosomes 12, 14, and 20 also indicated possible linkage to lung cancer susceptibility, although the results were not as strong. Identifying the chromosomal region was a critical first step, but more work needs to be done.

Implications - The next goal for these researchers is to examine this 20 million base pair segment on the long arm of chromosome 6 more closely, with the aim of locating the exact gene or genes that cause lung cancer susceptibility. The tools provided by the HapMap (see above) will speed this analysis. This region contains numerous genes that are likely candidates for the susceptibility gene, including four suspected tumor suppressor genes.

Another interesting discovery the team made involved the effects of smoking on cancer risk for carriers and non-carriers of the predicted familial lung cancer gene. They found that in non-carriers, the more they smoked, the greater their risk of cancer. In carriers, on the other hand, any amount of smoking increased lung cancer risk. These findings suggest that smoking even a small amount can lead to cancer for individuals with inherited susceptibility.

Citation - Bailey-Wilson JE, Amos CI, Pinney SM, Petersen GM, de Andrade M, et al., A Major Lung Cancer Susceptibility Locus Maps To Chromosome 6q23-25. Am J Hum Genet. 75: 460–474, 2004.

Transgenic Animals Produced Using Cultured Sperm

Background - Past efforts to modify sperm cells genetically in animals prior to fertilization have been stymied by the failure of these cells to mature under *in vitro* growth conditions.

Advance - A Japanese-U.S. team reported last February in the *Proceedings of the National Academy of Sciences* the successful creation of transgenic fish using sperm genetically modified and grown in a laboratory dish, an achievement with implications for a wide range of research from developmental biology to gene therapy.

In their study, researchers from Fukui Prefectural University in Obama, Japan, and the NHGRI, developed innovative techniques to produce genetically modified zebrafish using sperm cells grown under laboratory, or *in vitro,* conditions. Drawing upon cell culture expertise, the Japanese-U.S. team developed a system that enables immature sperm cells, or spermatagonia, taken from male zebrafish to survive long enough outside of the fish to receive foreign genes inserted by a retrovirus. Those cells go on to develop into mature, functional sperm. The genetically modified, cultured sperm are then used to fertilize zebrafish eggs in a laboratory dish — a process known as fertilization — resulting in the production of transgenic embryos and, ultimately, transgenic zebrafish.

Implications - Although further refinement and testing is needed, these new techniques have the potential to speed the production of many different types of transgenic animal models that will shed new light on human development and disease. The sperm culturing system will allow researchers to explore further the basic biology of sperm production in vertebrates. The findings also may prove helpful to researchers exploring pre-fertilization strategies for human gene therapy, thus allowing preventive treatment for certain genetic disorders.

One of the biggest advantages of the cultured sperm approach is that transgenic zebrafish created in this way carry the inserted, foreign gene in every cell of their bodies, including their germ cells. This means the fish will transmit the foreign gene along to their offspring in a pattern identical to their natural genes.

The zebrafish, *Danio rerio,* is a small, transparent aquarium fish used as a model system for studying the biology of vertebrates. The fish, which share many of the same genes as humans, are ideal for genetic studies because of their rapid rate of reproduction and because their genes can be readily mutated.

Kurita K, Burgess SM, Sakai N, Transgenic Zebrafish Produced By Retroviral Infection of In Vitro-Cultured Sperm. Proc Natl Acad Sci U S A. 101:1263-7, 2004.

Mouse Reference Transcriptome

Background - The completion of the sequencing of the human genome in April 2003 represented an enormous scientific achievement and the start of a new era in biomedical research. However, it also emphasized the importance of the next steps – to determine the function and therapeutic potential for all human genes.

One of the most universally asked questions about genes is tissue localization, i.e., where is each gene expressed. Every cell in the body has the same instruction book, but not every paragraph is read in every cell. Determining which genes are expressed in which tissues is often invaluable in determining candidate functions for a given unknown gene, and candidate genes for a given physiological process. The gathering of such information is often the first step a laboratory scientist will take when confronted with a gene of unknown function.

Advance – Creation of a catalog of genes expressed in mice. Using newly developed and exacting protocols, 88 tissues, representing a full survey of tissues, have been isolated from male and from female mice. RNA, representing the genes that are tuned on, has been extracted and has undergone a digital analysis of the level of expression of all of the genes in the genome. Bioinformatic analysis is being performed currently, and by the end of calendar year 2004 all results will be fully available to the scientific community in public databases at the National Center for Biotechnology Information.

Implications - A searchable database of the tissue expression of every gene will be widely used by researchers, will decrease duplication of effort, and will accelerate the transition of genome information from sequence to biological and disease-related function.

Citation - The Mouse Reference Transcriptome Consortium, A Reference Transcriptome of the Mouse by Massively Parallel Signature Sequencing. *Manuscript in preparation*, 2005.

NIH ROADMAP

Molecular Libraries

A class of organic chemical compounds, commonly referred to as "small molecules," (to distinguish them from large molecules like proteins) has proven extremely important to researchers exploring cellular functions at the molecular level. Such molecules have also been valuable for treating diseases. Most medicines marketed today are small molecules. Small molecules interact with a particular cellular target to block or enhance a particular function.

It remains difficult to predict which small molecules will be most effective at modulating a given biological process or disease state. Therefore, researchers must systematically screen tens or hundreds of thousands of small molecules to find a successful match between a chemical and its target. Over the last ten years the pharmaceutical and biotechnology sectors have built the capacity for such screening, but similar resources have not existed in the public sector. The Molecular Libraries Roadmap will offer public sector biomedical researchers access to small organic molecules that can be used as chemical probes to study the functions of genes, cells, and biochemical pathways. It will provide new ways to explore the functions of major components of cells in health and disease.

While the primary goal of this initiative is to enable researchers to develop small molecules for use as research probes, NIH anticipates that these projects will also facilitate the downstream development of new drugs, by providing early stage chemical compounds that will be considered attractive by researchers in the private sector into the drug-development pipeline. NIH may even

consider pursuing this possibility with public funds for compounds that show exceptional promise for rare diseases, since these may not be attractive for development by the private sector.

Three key technological advances drive NIH's current effort to provide access to small molecule technology for academic researchers. First, the successful completion of the Human Genome Project has provided an enormous cache of human biology to be studied and potential new targets to be explored. Second, developments in chemistry have given researchers in the public sector the ability to synthesize large numbers of related molecules, a capability previously available only to researchers in pharmaceutical and biotechnology companies. Third, advances in robotic technology and informatics now allow scientists to screen hundreds of thousands of compounds in a single day, an order of magnitude greater capacity than was available a decade ago.

In June 2004, NHGRI announced the establishment of the NIH Chemical Genomics Center - the first component of the nationwide Molecular Libraries Screening Centers Network. This intramural center, which will have a staff of about 50 scientists, will begin high-throughput screening of small molecules by early 2005. Up to 8 pilot extramural centers will be funded at academic institutions and other locations across the country in the spring of 2005. All of these will function as an integrated network, showing a common publicly available database (PubChem, already activated in September 2004) which will display the results of all screens.

Roadmap initiatives outside of Molecular Libraries in which NHGRI is involved

As an institute focused on advances in genetics and genomics issues that affect nearly all areas of research, NHGRI has participated extensively in many initiatives of the NIH Roadmap.

Nanomedicine, an offshoot of nanotechnology, refers to highly specific medical interventions at the molecular scale for curing disease or repairing damaged tissues.
NHGRI co-chairs the nanomedicine roadmap initiative In September 2004, the initiative funded a first round of 20 nanomedicine center planning grants. Awards for nanomedicine research centers will be made in FY 2005.

NHGRI is a member of the implementation group of the Building Blocks, Biological Pathways and Networks initiative, in which researchers will focus on the development of new technologies to accelerate discovery and facilitate comprehensive study of biological pathways and networks.

NHGRI is a member of the implementation group of the Structural Biology Roadmap initiative. This is a strategic effort to create a "picture" gallery of the molecular shapes of proteins in the body. This research investment will involve the development of rapid, efficient, and dependable methods to produce protein samples that scientists can use to determine the three-dimensional structure, or shape, of a protein.

NHGRI also participates in the Bioinformatics and Computational Biology roadmap initiative. A central focus of the effort will be a set of National Centers for Biomedical Computing.

And finally, NHGRI participates in the Interdisciplinary Research roadmap initiative which includes funding for: training of scientists in interdisciplinary strategies; specialized centers to

help scientists forge new and more advanced disciplines from existing ones; and planning of forward-looking conferences to catalyze collaboration among the life and physical sciences, important areas of research that historically have had limited interaction.

NEW INITIATIVES

Knockout Mouse Project

The technology to knock out genes, turning genes off in order to test their function, in mouse embryonic stem cells was developed in the late 1980's. Use of this technology has already led to many insights into human biology and disease. However, knockout mice have been made available to the research community for only about 10% of the estimated 25,000 mouse genes. In recognition of the wealth of information that mouse knockouts can provide, NHGRI and other partners coordinated a meeting at Cold Spring Harbor's Banbury Conference Center in the fall of 2003 to enable members of the genomics community to discuss the feasibility of a dedicated project to produce knockout mice for every mouse gene and make them available as a community resource for the scientific community. These discussions resulted in a coordinated, cooperative plan by scientists in the research community. A Commentary was drafted by the meeting attendees and published in the September 2004 issue of *Nature Genetics*. Discussions with the Wellcome Trust, researchers in the European Union, and the private sector have shown considerable enthusiasm for the project, and planning for its implementation is now underway.

The U.S. Surgeon General's Family History Initiative

Anticipating the changes that sophisticated genetic testing and other genomics advances will have on the U.S. health care system, the NHGRI is continuing to work to prepare the American public for the integration of such tools into their own health care. An important but often neglected tool is the family history. Moreover, while advances arising out of the Human Genome Project will soon add important new genomics tools for these tasks, family history will remain highly relevant for years to come. Yet, many individuals are unaware of the medical histories of their relatives, and many health professionals underutilize this information in advising patients about how to maintain good health.

Led by the Office of the Surgeon General, the NHGRI, the Office of the Director of the National Institutes of Health, the Centers for Disease Control and Prevention, the Health Resources and Services Administration, and the Agency for Healthcare Research and Quality are working closely to coordinate efforts to increase the American public's awareness of the importance of family history, to provide accessible methods for easily obtaining an accurate family history, and to increase health professionals' use of the family history in disease prevention and health promotion. These organizations are teaming up to make Thanksgiving Day 2004, the day that American families traditionally gather together, the first annual National Family History Day.

The goals of National Family History Day include:

- Increasing the American public's awareness of the importance of family history in health;
- Giving the American public a free web-based family history tool to gather, understand, evaluate, and use family history to improve their health;

- Increasing health professionals' awareness of the importance of family history;
- Giving health professionals tools to gather, evaluate, and use family history information and to communicate with their patients about family history;
- Increasing genomics and health literacy;
- Preparing both the American public and their health professionals for the coming era in which genomics will be an integral part of regular health care.

Thanksgiving Day 2004 served as the inaugural National Family History Day. In a phased approach, this year's initiative focused on increasing awareness of the importance of family history and on laying a foundation to make this an effective annual national campaign. U.S. Surgeon General Richard Carmona was the public face of this year's initiative, and NHGRI Director, Francis Collins also served as a prominent spokesperson. The involved agencies worked closely with federal, state, and local governmental agencies, voluntary health organizations, health professional organizations, community organizations, federal grantees, and commercial sector partners, and others to reach out to the American people and health professional workforce. A coordinated national communications effort was planned.

For more information see www.hhs.gov/familyhistory.

Minority Outreach

Our understanding of human genetic variation is in its infancy. It is also clear that the use of race and ethnicity in biomedical, and particularly genomics, research is the subject of heated debate. The NHGRI held a roundtable discussion about this issue on March 8–10, 2004. That meeting addressed the historical, political, societal, and clinical elements of race, ethnicity, and genetics, with the goal of developing a better understanding of the relationships between genomics, race, and ethnicity and of the consequences of uncovering these relationships. In addition, Dr. Collins participated in a meeting in May 2003 entitled "Human Genome Variation and 'Race': The State of the Science at Howard University." Dr. Collins also contributed to a published summary of the meeting in *Nature Genetics* just prior to the American Society of Human Genetics meeting in Toronto in October 2004.

Collins FS, What We Do And Don't Know About 'Race', 'Ethnicity', Genetics And Health at The Dawn of The Genome Era. Nat Genet. 36 Suppl: S13-5, 2004.

Centers for Excellence in ELSI Research

On August 31, 2004, the NHGRI's ELSI research program announced the funding, with contributions from the Department of Energy and the National Institute of Child Health and Human Development, of four interdisciplinary centers as part of its *Centers for Excellence in ELSI Research (*CEER) program, a new initiative to address some of the most pressing ethical, legal and social questions facing individuals, families and communities in the genome era. Each of the new centers will assemble a team of experts in several disciplines, such as bioethics, law, behavioral and social sciences, clinical research, theology, public policy, and genetic and genomic research. The interdisciplinary nature of these teams will allow the centers to develop innovative research approaches focused on specific sets of issues that relate to the numerous

applications of genomic research. The centers' output will be critical in formulating and implementing effective and equitable health and social policies related to genomic research. In addition, the centers will create new environments to support the growth of the next generation of researchers devoted to exploring the ethical, legal, and social implications of genomic research. Special efforts will be made to recruit potential researchers from currently under-represented groups.

The first four Centers will be at Case Western Reserve University's Center for Genetic Research Ethics and Law, The Duke Center for the Study of Public Genomics, Stanford University School of Medicine's Center for Integration of Research on Genetics and Ethics, and the University of Washington's Center for Genomic Health Care and the Medically Underserved. In addition, the NHGRI has awarded three exploratory grants that provide two to three years of support to investigators for planning and developing potential new centers at their institutions. These were awarded to Georgetown University, Howard University, and the University of North Carolina, Chapel Hill.

Intellectual Property Initiatives

NAS Study on Intellectual Property in Genomic and Protein Research and Innovation

In 2004, the NHGRI, along with three other NIH institutes and the office of the NIH Director, funded the National Academies of Sciences' Board on Science Technology and Economic Policy (STEP) and its Science, Technology, and Law Program to convene a Committee on Intellectual Property in Genomic and Protein Research and Innovation. The Committee's charge is to examine the patenting and licensing of human genetic material and proteins and their implications for biomedical research, therapeutic and diagnostic products, and medical practice. They will also compare the U.S. patent system, as it relates to genetics and proteomics, with those of Europe and Japan. The Committee, chaired by Princeton University President Shirley Tilghman, will meet five times and is expected to release its report in the summer of 2005.

NHGRI Intellectual Property Rights in Genetics and Genomics RFA

In June 2004, the NHGRI released a Request for *Applications (RFA)* on "Intellectual Property Rights in Genetics and Genomics" (see http://www.genome.gov/10001618). The purpose of this RFA is to encourage study of the role of laws and policies regarding intellectual property rights in genetics and genomics research and development, and the effect of such laws and policies on progress in these fields and on commercialization, drug development, health care delivery, and the public health.

The Human Genome Project championed free and open access to genetic and genomic data. NIH policy recognizes the appropriateness of intellectual property protections for discoveries that are associated with useful products, but promotes the free dissemination of research tools whenever possible, especially when the prospect of commercial gain is remote. Over the past three decades, however, many patents have been granted on gene sequences and other types of basic information derived from genetic sequence. This has generated apprehension that gene patents are granted too broadly or freely, especially for foundational tools. The concern is that too liberal issuance of patents, especially when coupled with exclusive licensing practices, will

result in reach-through restrictions or excessive fees, and inhibit investigators from conducting research with these tools. This, it is feared, will ultimately deter medical research and public health advances. The NHGRI hopes to fund research through this RFA that will address some of these issues, which we know are of great interest to many basic biomedical researchers.

Social and Behavioral Research Branch

Last year the NHGRI formed a Social and Behavioral Research Branch within its intramural program. This new branch has the overarching and broad objective to investigate social and behavioral factors that facilitate translation of genomic discoveries for health promotion, disease prevention, and health care improvements. The main focus of the branch is to conduct research on the social and behavioral aspects of translating genomic discoveries into improved health. The branch is currently focused on four conceptual domains:

- Testing risk communications;
- Developing and evaluating behavioral interventions;
- Using genomic discoveries in clinical practice; and
- Understanding social, ethical and policy implications of genomic research.

This new branch is already forming relationships with other components of the NIH who work on behavioral research, in an attempt to form productive collaborations in this area. The plan is for this new branch to assist in the translation of basic biomedical research discoveries into practical behavioral interventions related to genetic disorders.

OTHER AREAS OF INTEREST

Education of Health Care Professionals

The NHGRI has developed several educational programs to prepare health care professionals for the integration of genomics into primary health care. In 1996 the NHGRI, along with the American Medical Association and the American Nurses Association, formed the National Coalition for Health Professional Education in Genetics (NCHPEG). NCHPEG serves as an "organization of organizations" committed to a national effort to promote health professional education and access to information about advances in human genetics. NCHPEG members are an interdisciplinary group of leaders from over 120 diverse health professional organizations, consumer and volunteer groups, government agencies, private industry, managed care organizations, and genetics professional societies. It draws on the collective expertise and experience of its members to accomplish its mission.

In 2004, NCHPEG launched the Genetics Resources On the Web (GROW) website. The mission of the GROW site and search engine is to optimize the use of the web to provide health professionals and the public with high quality information related to human genetics, with a particular focus on genetic medicine and health. The GROW web site search engine effectively searches for human genetics content across over two dozen GROW members' web sites. Communication among GROW members is encouraged via meetings and activities.

The NHGRI has also worked closely with the American Academy of Family Physicians (AAFP), an organization that has taken a keen interest in genomics. In 2005 AAFP chose to make Genomic Medicine the theme of its Annual Clinical Focus, in collaboration with the NHGRI and other organizations. NHGRI Director Francis Collins gave the keynote address at the AAFP annual meeting, where he described how genomics will become an integral part of medicine in the coming years and urged the group to prepare itself for use of its new tools. Beginning in January 2005, AAFP members will be able to access online Continuing Medical Education (CME) programs about genomics and its relationship to a variety of clinical topics including: family history; breast cancer; colon cancer; bipolar disorder; and Alzheimer's disease.

Sickle Cell Disease Research

The NHGRI is deeply interested in applying genomics research to further understanding, prevention, and treatment not only of common, complex diseases, but also of less common, single gene disorders. The NHGRI cannot, of course, tackle each of the several thousand Mendelian diseases, but it is interested in developing paradigms, or "proof of principle" examples of effective application of genomics approaches to single gene disorders.

Sickle cell disease (SCD) continues to be a significant cause of mortality, morbidity, and health disparities in the US and globally, despite the fact that it was the first disease whose genetic basis was defined. For these and other reasons, sickle cell disease appeared to be a suitable focus for an effort to apply genomics approaches to a single gene disorder. The NHGRI approached the National Heart, Lung, and Blood Institute (NHLBI), which has chief responsibility at the NIH for SCD research, about developing a trans-NIH effort to apply genomics to SCD. The NHGRI and NHLBI, in conjunction with the National Institute of Diabetes and Digestive and Kidney Diseases, the Office of Rare Diseases, the Fogarty International Center, and the Foundation for the National Institutes of Health organized a conference entitled "New Directions for Sickle Cell Therapy in the Genome Era," in November 2003. Over 120 individuals from the US and abroad took part in the conference. (See http://www.genome.gov/11509561 for the conference report)

A trans-NIH staff working group, co-chaired by NHGRI and NHLBI, was formed, including representation from eight institutes and centers, to follow up on the recommendations from the conference. That group launched the Sickle Cell Disease Research Listserv (SCDR-L) to facilitate easy communication between those involved in SCD research. In addition, a Clinical Research Network and Registry RFA is planned, to create an international prospective registry of patients with SCD to facilitate cooperative clinical research. Establishing "Novel Idea Grants" for new therapies to treat SCD is also under consideration. It is likely that a training program for application of genomics and proteomics to hemoglobin disorders as well as a program in chemical genomics approaches in SCD research, will be funded in 2005 or 2006. In the summer of 2005, two working groups will be convened, one to investigate how to improve access for investigators to expertise and resources for moving potential new drugs for SCD through required clinical and regulatory processes and into the marketplace, and the other to focus on refinement of safety features of lentiviral hemoglobin gene transfer vectors as a prelude to *ex vivo* clinical trials of hemoglobin gene transfer for treatment of SCD and Cooley's anemia. Finally, efforts to identify genes that modify the phenotypic effects of mutations in hemoglobin genes, and to study the mechanisms and effects of such modifications, are envisioned to follow the formation of the clinical research network and registry.

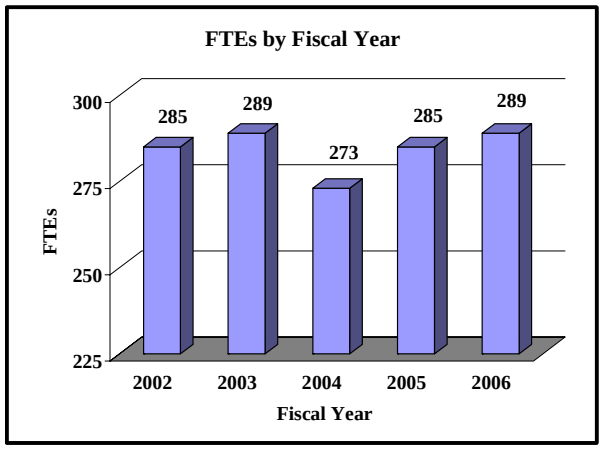
Direct to Consumer Marketing of Genetic Tests

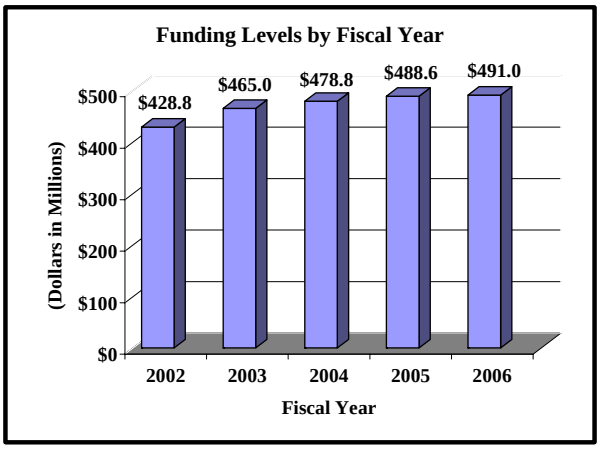
The marketing of products or services that promise to provide consumers with genetic insights into personal health has recently proliferated dramatically. Unfortunately, some advertisements exaggerate the scientific basis of the claims made, and/or fail to communicate effectively, if at all, the limitations of the specific genetic information offered. On March 23, 2004 NHGRI held a workshop, including many governmental agencies and nongovernmental organizations, on issues surrounding the direct to consumer marketing of genetic tests. Following a series of productive discussions, the attendees proposed three major recommendations for next steps in this area: (1) facilitate the development of a stakeholder consensus document outlining "best practices" for direct to consumer (DTC) advertising in the realm of genetic tests and, ideally, services; (2) facilitate the development of a formal petition to the Federal Trade Commission outlining concerns with current DTC advertising practices for genetic tests; and (3) develop a research agenda able to inform future advertising practices and any policy development. The NHGRI is continuing to work with various partners on this issue.

Budget Policy

The Fiscal Year 2006 budget request for the NHGRI is $490,959,000, an increase of $2,351,000 and 0.5 percent over the FY 2005 Final Conference Level. Also included in the FY 2006 request, is NHGRI's support for the trans-NIH Roadmap initiatives, estimated at 0.89 percent of the FY 2006 budget request. This Roadmap funding is distributed through the mechanisms of support, consistent with the anticipated funding for the Roadmap initiatives. A full description of this trans-NIH program may be found in the NIH Overview.

A five year history of FTEs and Funding Levels for NHGRI are shown in the graphs below.



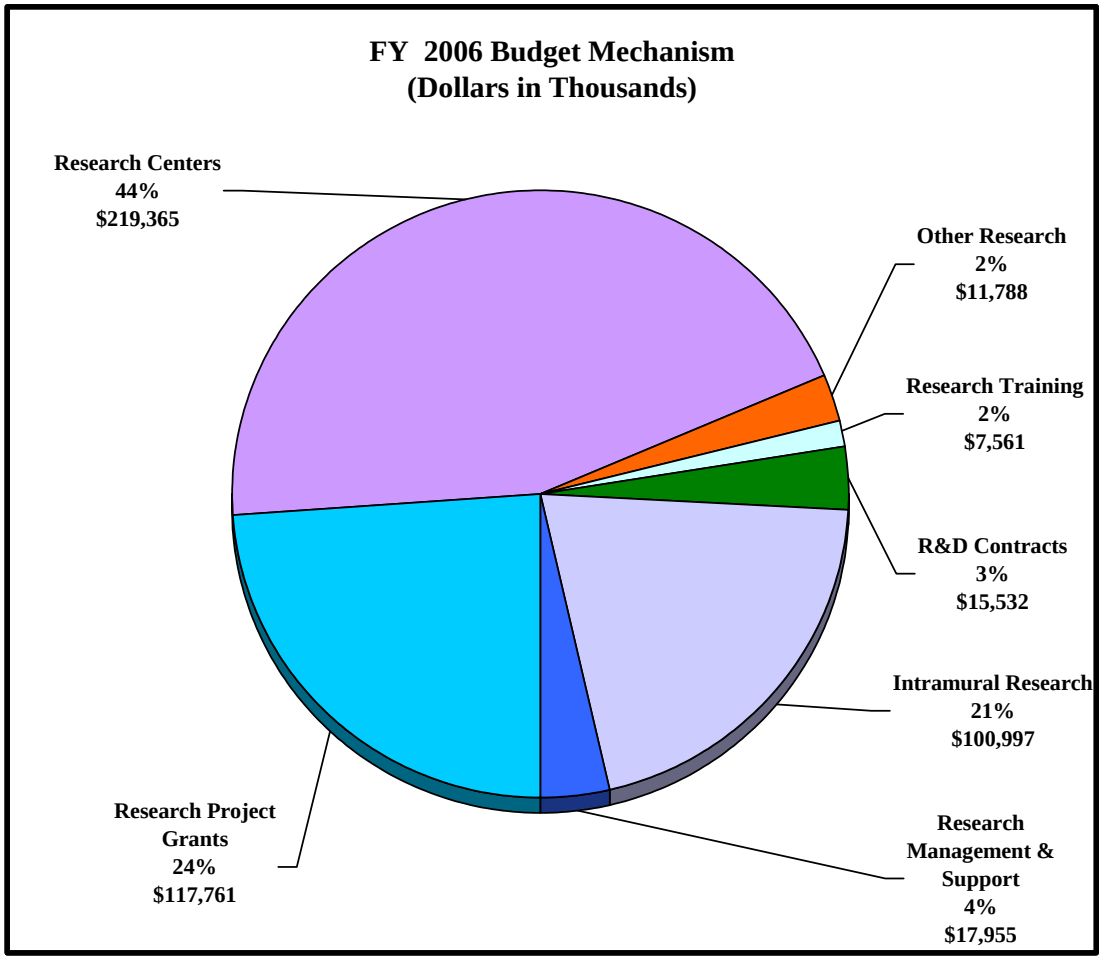


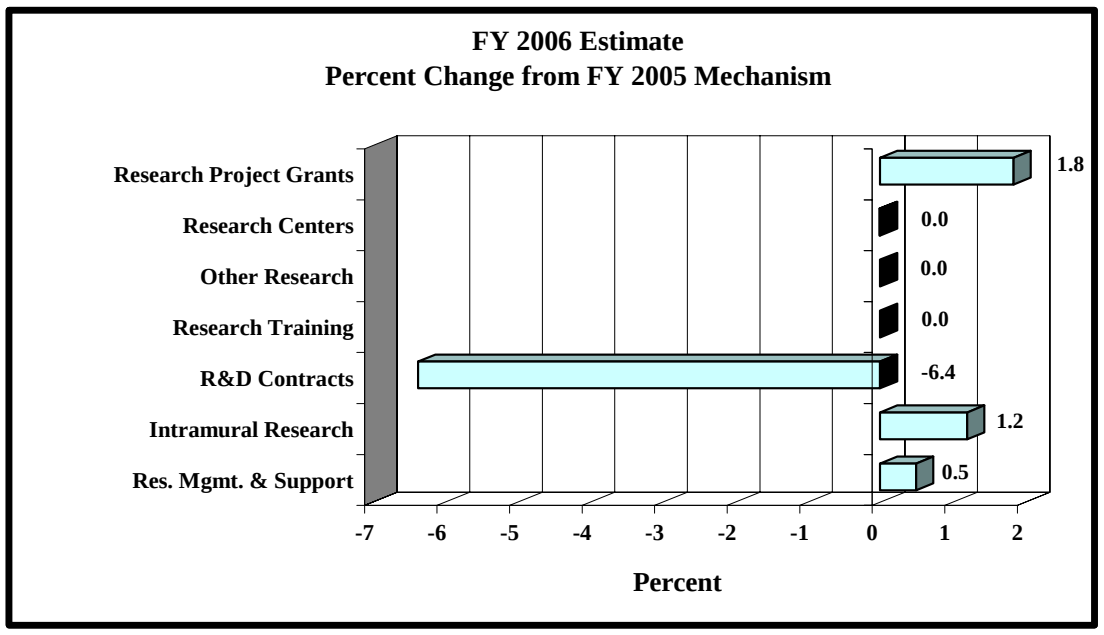
NIH's highest priority is the funding of medical research through research project grants (RPGs). Support for RPGs allows NIH to sustain the scientific momentum of investigator-initiated research while pursuing new research opportunities. We estimate that the average cost of competing RPGs will be $447,800 in FY 2006. While no inflationary increases are provided for direct, recurring costs in non-competing RPG's, where the NHGRI has committed to a programmatic increase in an award, such increases will be provided.

Advancement in medical research is dependent on attracting, training, and retaining the best and the brightest individuals to pursue careers in biomedical and behavioral research. In the FY 2006 request, most stipend levels for individuals supported by the Ruth L. Kirschstein National Research Service Awards are maintained at the FY 2005 levels. To help prevent the potential attrition of our next generation of highly trained post-doctoral trainees, stipend levels for post-docs with 1-2 years of experience are increased by 4.0 percent. This will bring these stipends closer to the goal NIH established for post-doc stipends in March 2000. In addition, individual post-doctoral fellows will receive an increase of $500 in their institutional allowance for rising health benefit costs. The need for increased health benefits is particularly acute for these post-doctoral trainees, who, because of their age and stage of life are more likely to have family responsibilities. The increases in stipends and health insurance are financed within the FY 2006 request by reducing the number of Full-Time Training Positions. The NIH believes that it is important to properly support and adequately compensate those who are participating in these training programs so that the programs can continue to attract and retain the trainees most likely to pursue careers in biomedical, behavioral and clinical research.

The Fiscal Year 2006 request includes funding for 32 research centers, 56 other research grants, including 33 clinical career awards, and 19 R&D contracts. Intramural Research receives an increase of 1.2 percent, which includes funds for 4 additional Roadmap FTEs, and Research Management and Support receives an increase of 0.5 percent, the same as the NIH total increase.

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 2004 Actual		FY 2005 Appropriation		FY 2006 Estimate	
Research Grants:	No.	Amount	No.	Amount	No.	Amount
Research Projects:						
Noncompeting	106	$69,451,000	136	$73,143,000	137	$62,264,000
Administrative supplements	(17)	2,915,000	(18)	3,200,000	(18)	3,200,000
Competing:						
Renewal	12	4,254,000	12	5,374,000	12	5,374,000
New	58	26,033,000	53	23,734,000	82	36,719,000
Supplements	0	0	0	0	0	0
Subtotal, competing	70	30,287,000	65	29,108,000	94	42,093,000
Subtotal, RPGs	176	102,653,000	201	105,451,000	231	107,557,000
SBIR/STTR	39	10,228,000	39	10,181,000	40	10,204,000
Subtotal, RPGs	215	112,881,000	240	115,632,000	271	117,761,000
Research Centers:						
Specialized/comprehensive	24	191,819,000	28	190,849,000	23	194,488,000
Clinical research	0	0	0	0	0	0
Biotechnology	15	27,459,000	13	28,526,000	9	24,877,000
Comparative medicine	0	0	0	0	0	0
Research Centers in Minority Institutions	0	0	0	0	0	0
Subtotal, Centers	39	219,278,000	41	219,375,000	32	219,365,000
Other Research:						
Research careers	33	6,134,000	35	6,362,000	33	6,362,000
Cancer education	0	0	0	0	0	0
Cooperative clinical research	0	0	0	0	0	0
Biomedical research support	0	9,000	0	0	0	0
Minority biomedical research support	0	0	0	0	0	0
Other	24	5,205,000	20	5,426,000	23	5,426,000
Subtotal, Other Research	57	11,348,000	55	11,788,000	56	11,788,000
Total Research Grants	311	343,507,000	336	346,795,000	359	348,914,000
Research Training:	FTTPs		FTTPs		FTTPs	
Individual awards	13	549,000	12	561,000	11	561,000
Institutional awards	157	7,493,000	139	7,000,000	137	7,000,000
Total, Training	170	8,042,000	151	7,561,000	148	7,561,000
Research & development contracts	25	17,517,000	19	16,588,000	19	15,532,000
(SBIR/STTR)	(0)	(20)	(0)	(0)	(0)	(0)
	FTEs		FTEs		FTEs	
Intramural research	219	94,617,000	221	99,798,000	225	100,997,000
Research management and support	54	15,145,000	64	17,866,000	64	17,955,000
Total, NHGRI	273	478,828,000	285	488,608,000	289	490,959,000
(RoadMap Support)		(1,645,000)		(3,089,000)		(4,390,000)
(Clinical Trials)		(8,645,000)		(8,968,000)		(9,048,000)

NATIONAL INSTITUTES OF HEALTH
National Human Genome Research Institute

Budget Authority by Activity
(dollars in thousands)

ACTIVITY	FY 2004 Actual		FY 2005 Appropriation		FY 2006 Estimate		Change	
	FTEs	Amount	FTEs	Amount	FTEs	Amount	FTEs	Amount
Extramural Research:								
Human Genome Research		$369,066		$370,944		$372,007		$1,063
Subtotal, Extramural research		369,066		370,944		372,007		1,063
Intramural research	219	94,617	221	99,798	225	100,997	4	1,199
Res. management & support	54	15,145	64	17,866	64	17,955	0	89
Total	273	478,828	285	488,608	289	490,959	4	2,351

NATIONAL INSTITUTES OF HEALTH
National Human Genome Research Institute

Summary of Changes

FY 2005 Estimate				$488,608,000
FY 2006 Estimated Budget Authority				490,959,000
Net change				2,351,000
	FY 2005 Appropriaton		Change from Base	
CHANGES	FTEs	Budget Authority	FTEs	Budget Authority
A. Built-in:				
1. Intramural research:				
a. Within grade increase		$27,109,000		$323,000
b. Annualization of January 2005 pay increase		27,109,000		251,000
c. January 2006 pay increase		27,109,000		468,000
d. One less day of pay		27,109,000		(104,000)
e. Payment for centrally furnished services		15,524,000		78,000
f. Increased cost of laboratory supplies, materials, and other expenses		57,165,000		1,158,000
Subtotal				2,174,000
2. Research Management and Support:				
a. Within grade increase		7,120,000		99,000
b. Annualization of January 2005 pay increase		7,120,000		66,000
c. January 2006 pay increase		7,120,000		123,000
d. One less day of pay		7,120,000		(27,000)
e. Payment for centrally furnished services		1,398,000		7,000
f. Increased cost of laboratory supplies, materials, and other expenses		9,348,000		189,000
Subtotal				457,000
Subtotal, Built-in				2,631,000

NATIONAL INSTITUTES OF HEALTH
National Human Genome Research Institute

Summary of Changes--continued

CHANGES	2005 Current Estimate Base		Change from Base	
	No.	Amount	No.	Amount
B. Program:				
1. Research project grants:				
a. Noncompeting	136	$76,343,000	1	($10,879,000)
b. Competing	65	29,108,000	29	12,985,000
c. SBIR/STTR	39	10,181,000	1	23,000
Total	240	115,632,000	31	2,129,000
2. Research centers	41	219,375,000	(9)	(10,000)
3. Other research	55	11,788,000	1	0
4. Research training	151	7,561,000	(3)	0
5. Research and development contracts	19	16,588,000	0	(1,056,000)
Subtotal, extramural				1,063,000
	FTEs		FTEs	
6. Intramural research	221	99,798,000	4	(975,000)
7. Research management and support	64	17,866,000	0	(368,000)
Subtotal, Program		488,608,000		(280,000)
Total changes	285		4	2,351,000

NATIONAL INSTITUTES OF HEALTH
National Human Genome Research Institute

Budget Authority by Object

		FY 2005 Appropriation	FY 2006 Estimate	Increase or Decrease
Total compensable workyears:				
	Full-time employment	285	289	4
	Full-time equivalent of overtime & holiday hours	1	1	0
	Average ES salary	$138,874	$145,818	$6,944
	Average GM/GS grade	11.6	11.6	0.0
	Average GM/GS salary	$73,158	$74,420	$1,262
	Average salary, grade established by act of July 1, 1944 (42 U.S.C. 207)	$70,029	$71,237	$1,208
	Average salary of ungraded positions	94,334	95,961	1,627
	OBJECT CLASSES	FY 2005 Appropriation	FY 2006 Estimate	Increase or Decrease
	Personnel Compensation:			
11.1	Full-Time Permanent	$11,513,000	$11,911,000	$398,000
11.3	Other than Full-Time Permanent	11,941,000	12,359,000	418,000
11.5	Other Personnel Compensation	312,000	323,000	11,000
11.7	Military Personnel	232,000	240,000	8,000
11.8	Special Personnel Services Payments	3,512,000	3,646,000	134,000
	Total, Personnel Compensation	27,510,000	28,479,000	969,000
12.0	Personnel Benefits	6,517,000	6,740,000	223,000
12.1	Military Personnel Benefits	178,000	184,000	6,000
13.0	Benefits for Former Personnel	24,000	25,000	1,000
	Subtotal, Pay Costs	**34,229,000**	**35,428,000**	**1,199,000**
21.0	Travel & Transportation of Persons	1,358,000	1,319,000	(39,000)
22.0	Transportation of Things	185,000	185,000	0
23.1	Rental Payments to GSA	5,000	5,000	0
23.2	Rental Payments to Others	289,000	289,000	0
23.3	Communications, Utilities & Miscellaneous Charges	498,000	498,000	0
24.0	Printing & Reproduction	143,000	143,000	0
25.1	Consulting Services	611,000	611,000	0
25.2	Other Services	22,019,000	21,644,000	(375,000)
25.3	Purchase of Goods & Services from Government Accounts	47,004,000	46,869,000	(135,000)
25.4	Operation & Maintenance of Facilities	5,339,000	5,259,000	(80,000)
25.5	Research & Development Contracts	5,364,000	5,264,000	(100,000)
25.6	Medical Care	797,000	797,000	0
25.7	Operation & Maintenance of Equipment	1,585,000	1,585,000	0
25.8	Subsistence & Support of Persons	0	0	0
25.0	**Subtotal, Other Contractual Services**	**82,719,000**	**82,029,000**	**(690,000)**
26.0	Supplies & Materials	8,983,000	8,868,000	(115,000)
31.0	Equipment	5,843,000	5,720,000	(123,000)
32.0	Land and Structures	0	0	0
33.0	Investments & Loans	0	0	0
41.0	Grants, Subsidies & Contributions	354,356,000	356,475,000	2,119,000
42.0	Insurance Claims & Indemnities	0	0	0
43.0	Interest & Dividends	0	0	0
44.0	Refunds	0	0	0
	Subtotal, Non-Pay Costs	**454,379,000**	**455,531,000**	**1,152,000**
	Total Budget Authority by Object	**488,608,000**	**490,959,000**	**2,351,000**

NATIONAL INSTITUTES OF HEALTH
National Human Genome Research Institute

Salaries and Expenses

OBJECT CLASSES	FY 2005 Appropriation	FY 2006 Estimate	Increase or Decrease
Personnel Compensation:			
Full-Time Permanent (11.1)	$11,513,000	$11,911,000	$398,000
Other Than Full-Time Permanent (11.3)	11,941,000	12,359,000	418,000
Other Personnel Compensation (11.5)	312,000	323,000	11,000
Military Personnel (11.7)	232,000	240,000	8,000
Special Personnel Services Payments (11.8)	3,512,000	3,646,000	134,000
Total Personnel Compensation (11.9)	**27,510,000**	**28,479,000**	**969,000**
Civilian Personnel Benefits (12.1)	6,517,000	6,740,000	223,000
Military Personnel Benefits (12.2)	178,000	184,000	6,000
Benefits to Former Personnel (13.0)	24,000	25,000	1,000
Subtotal, Pay Costs	**34,229,000**	**35,428,000**	**1,199,000**
Travel (21.0)	1,358,000	1,319,000	(39,000)
Transportation of Things (22.0)	185,000	185,000	0
Rental Payments to Others (23.2)	289,000	289,000	0
Communications, Utilities and Miscellaneous Charges (23.3)	498,000	498,000	0
Printing and Reproduction (24.0)	143,000	143,000	0
Other Contractual Services:			
Advisory and Assistance Services (25.1)	611,000	611,000	0
Other Services (25.2)	22,019,000	21,644,000	(375,000)
Purchases from Govt. Accounts (25.3)	27,329,000	27,223,000	(106,000)
Operation & Maintenance of Facilities (25.4)	5,339,000	5,259,000	(80,000)
Operation & Maintenance of Equipment (25.7)	1,585,000	1,585,000	0
Subsistence & Support of Persons (25.8)	0	0	0
Subtotal Other Contractual Services	**56,883,000**	**56,322,000**	**(561,000)**
Supplies and Materials (26.0)	8,983,000	8,868,000	(115,000)
Subtotal, Non-Pay Costs	68,339,000	67,624,000	(715,000)
Total, Administrative Costs	**102,568,000**	**103,052,000**	**484,000**

NATIONAL INSTITUTES OF HEALTH
National Human Genome Research Institute

Significant Items in House and Senate Appropriations Committee Reports

FY 2005 House Appropriations Committee Report Language (H. Rpt. 108-636)

Item

Targeting disease prevention – The Committee commends NHGRI for creating the Encyclopedia of DNA Elements (ENCODE) project, which has a long-term goal of creating a comprehensive catalogue of the functional components encoded in the human genome. This project will help scientists mine and fully utilize the human sequence, gain a deeper understanding of human biology, predict potential disease risk, and stimulate the development of new genome technologies as well as strategies for the prevention and treatment of disease. The Committee is also pleased with the progress of the international haplotype mapping (HapMap) project and looks forward to its completion in 2005. In each of these areas, the Committee recognizes that advances in the use of new genomic technologies, including different types of microarrays, can enhance NHGRI's global leadership in meeting these goals. (p. 98)

Action taken or to be taken

Following completion of the Human Genome Project, the next logical step was to characterize both the genes and intergenic elements that regulate gene expression, DNA replication, and chromosome structure. With the goal of identifying the precise location and function of all sequence-based functional elements in the genome, the NHGRI launched the ENCyclopedia Of DNA Elements (ENCODE) project. Grants were awarded in 2003 for the pilot phase, calling for application of multiple innovative technologies to 30 megabases (roughly 1%), of the human genome. The pilot phase of the project is organized as an international consortium of computational and laboratory-based scientists, including those working with microarray technology; it is open to all investigators who agree to abide by the project's criteria and guidelines for participation.

Pre-existing technologies would not have been able to achieve all of the aims of even the initial phase of the ENCODE project. Therefore, at the same time that high-throughput efforts were being initiated using well-developed technologies, NHGRI launched a parallel effort to develop new technologies. The first set of grants to expand the repertoire of tools that can be applied to ENCODE or similar future projects was awarded in 2003. A second set of technology development grants was awarded in September 2004. It is envisioned that when an extensive "tool box" of technologies is available, it will be possible to annotate the entire human genome with information that will serve as a platform for more in-depth, detailed studies of biological function.

The ENCODE Consortium is already generating a remarkable amount of data and many collaborations are occurring between investigators within and outside of the Consortium. A manuscript describing the ENCODE project is included in a recent issue of *Science*[2]. For more information also see http://www.genome.gov/10005107.

Understanding how genetic variation is inherited in blocks of DNA known as haplotypes can provide considerable savings in time, effort, and cost in uncovering hereditary factors in disease. NHGRI has taken a leadership role in the development of the HapMap, a catalog of haplotype blocks and the Single Nucleotide Polymorphisms (SNPs) that tag them. Researchers can use the HapMap to find the genes and variants that contribute to many diseases and, in addition, it will be a powerful resource for studying the genetic factors contributing to variation in individual response to disease, to drugs, and to vaccines.

The HapMap project is on schedule to be completed by the end of 2005. While the consortium had initially planned to identify an additional 3 million new SNPs to fill in areas where the density of SNPs in public databases was insufficient, the project has been able to identify more than 6 million new SNPs, for a total of 9 million. The consortium has completed the collection of samples and consent from 270 individuals from four populations (CEPH, [U.S. residents with ancestry from Western and Northern Europe] Yoruba in Ibadan, Nigeria, Han Chinese in Beijing, and Japanese in Tokyo). A newly issued grant to Perlegen Sciences, Inc. will lead to a significant increase in SNP density across the genome on the HapMap samples. The consortium is also developing scientific strategies to choose which SNPs to study, to assess the quality of the data, and to derive haplotypes from the SNP data. More information about the HapMap project is available at www.hapmap.org.

[2] The ENCODE (ENCyclopedia Of DNA Elements) Project. (2004) ENCODE Project Consortium. Science. Oct 22;306(5696):636-40.

Item

Neurogenomics and neuroinformatics – The Committee encourages NHGRI to assist private efforts to collect and integrate the various repositories of neuroscience information, including brain imaging, with genome information and cross-referencing this with human disease to advance knowledge along a broad front of neurologic, psychiatric and neurodegenerative diseases, their causes and cures. (p. 98)

Action taken or to be taken

The NIH recognizes the need to collect and integrate disparate types of neuroscience data (i.e., genetic, anatomical, and physiological), resulting from different types of experiments, conducted on different subjects and in different laboratories, in order to make sense of the incredibly complex nervous system. NIH is supporting several major efforts to this end.

The NHGRI's focus has been on collecting and providing genomic data that can be used by anyone in the research community including those from the private sector. The Human Genome Project, along with several other genome sequencing projects of animal models, will enable many advances in the area of neuroscience. NHGRI supports development of novel views of these data, such as the ones provided by the UCSC Genome Browser, to help investigators working in neuroscience and other fields to extract meaning from the comparative genome sequence data sets. With support from more than a dozen NIH Institutes, the NHGRI is now working to complete a haplotype map via the HapMap project, which will provide the other disease-specific NIH institutes the data they need to integrate information on disease phenotypes with genotype information. For instance, the HapMap will provide new insights into complex diseases (with many genes) such as autism, multiple sclerosis, stroke, depression and schizophrenia.

The National Center for Research Resources (NCRR) has created the Biomedical Informatics Research Network (BIRN), a nationwide, high-performance computer network that enables biomedical investigators across the country to share and mine large datasets. Currently, researchers using BIRN are concentrating on neuroimaging studies of mice and humans on several neurodegenerative diseases including schizophrenia, mild Alzheimer's disease, and Parkinson's disease. Because of the network, researchers can combine images of research subjects and data from multiple laboratories. Scientists in the network are also developing the tools to compare neuroimaging data with cellular- and molecular-level data in both mice and humans.

Informatics tools are essential at all stages of neuroimaging, allowing scientists to control highly sophisticated imaging instruments and to make sense of the vast amounts of complex data that are generated. The NINDS and NIMH developed the Neuroimaging Informatics Technology Initiative (NIfTI) to provide coordinated and targeted service, training, and research to speed the development and enhance the utility of informatics tools related to neuroimaging.

The Gene Expression Nervous System Atlas (GENSAT) initiative, supported by NINDS, NIDA, and NIMH, is creating an "atlas" of the mouse brain, which links gene expression and neuroanatomy data. As of the summer of 2004, data on 1,000 genes have been entered into

publicly accessible databases. The completion of the mouse genome project, sponsored by NHGRI, paved the way for GENSAT and other gene expression studies of mouse models. The NHGRI also supports the Mouse Genome informatics database.

The NINDS and NIMH support a Microarray Consortium that offers gene expression profiling services to all NINDS and NIMH extramural investigators and makes all of the resulting data publicly available through web-based databases. Current projects focus on neurodegenerative disorders, psychiatric conditions, epilepsy, brain tumor, traumatic brain injury, autism, stroke, development, cognition, and sleep and circadian rhythms. Because the experiments at the microarray centers in the consortium are conducted under the same conditions, it is possible to combine data from independent experiments for analysis of many research projects simultaneously. The intramural programs of NINDS, and NIMH and NHGRI also share a Microarray Core Facility aimed at providing intramural investigators with access to state-of-the-art technologies for understanding patterns of gene expression. The NHGRI supported the development of most of the infrastructure of for this work including the original technology development for microarrays and set the standard for rapid data sharing.

FY 2005 Senate Appropriations Committee Report Language (S. Rpt. 108-345)

Item

Gene-Environment Interactions – NHGRI is commended for its partnerships with other Institutes and the Office of Behavioral and Social Sciences Research to push the frontier of genetics research forward by examining gene-environment interactions. Research on the environmental stimuli, such as behaviors, experience of stress, or exposure to certain physical conditions, that lead to the expression of genes is critical if science is to reap the benefits of the mapped genome. NHGRI is encouraged to work with OBSSR on multidisciplinary training programs to increase the number of skilled scientists who can bridge the behavioral and genetic realms. (p. 154)

Action taken or to be taken

The NHGRI believes integrated research that advances our understanding of how environmental factors affect gene expression in health and disease is an area of great importance. A recent paper published in the journal *Nature* in October 2004 describing the near-complete human genome sequence will greatly assist the scientific community in better understanding genetic contributions to disease and gene-environment interactions.[3] Over the past few years, several Institute of Medicine expert panels commissioned by NIH have concurred in suggesting that understanding social and behavioral phenomenon will be essential to maintain and further promote public health improvements. In 2001 the NHGRI and the NIH Office of Behavioral and Social Sciences Research (OBSSR) formed the trans-NIH Working Group on Interactions among Genetic, Behavioral, and Social Factors. The working group includes representatives of numerous NIH institutes and centers and enables the NIH to coordinate its efforts on these matters. The working group has now asked the Institute of Medicine (IOM) to draft a new report on "Assessing Interactions Among Social, Behavioral, and Genetic Factors in Health," which will help to guide the NIH's efforts in this regard. The IOM study also included consideration of NIH training needs in the area of genetics and behavioral research. In its intramural program, NHGRI has recently created a new Branch on Behavioral and Social Science Research and recruited a highly regarded senior scientist as Branch Chief. The NHGRI is also working with OBSSR and others to identify intramural training opportunities related to behavioral genetics. Finally, the NHGRI has been working with OBSSR on a new NIH "Roadmap" initiative: "Interdisciplinary Research Teams of the Future." The Roadmap has recently reissued an "Interdisciplinary Health Research Training" RFA that has attracted proposals that create new ways to provide individuals trained in one discipline with formal coursework and laboratory training in a second discipline. This training initiative is ideal for the cross-disciplinary approach needed to address the complex issues of genetics and environmental interaction. The NHGRI is working to help attract genome scientists as potential applicants for this RFA.

[3] International Human Genome Sequencing Consortium. (2004) Finishing the euchromatic sequence of the human genome.
Nature. Oct 21;431(7011):931-45.

NATIONAL INSTITUTES OF HEALTH
National Human Genome Research Institute

Authorizing Legislation

	PHS Act/ Other Citation	U.S. Code Citation	2005 Amount Authorized	FY 2005 Appropriation	2006 Amount Authorized	2006 Budget Estimate
Research and Investigation	Section 301	42§241	Indefinite	$481,047,000	Indefinite	$483,398,000
National Human Genome Research Institute	Section 401	42§285b	Indefinite		Indefinite	
National Research Service Awards	Section 487(d)	42§288	a/	7,561,000		7,561,000
Total, Budget Authority				488,608,000		490,959,000

a/ Amounts authorized by Section 301 and Title IV of the Public Health Act.

NATIONAL INSTITUTES OF HEALTH
National Human Genome Research Institute

Appropriations History

Fiscal Year	Budget Estimate to Congress	House Allowance	Senate Allowance	Appropriation 1/
1997	$177,788,000 2/	$189,267,000	$180,807,000 2/	$189,657,000 3/
1998	202,197,000 2/	211,772,000	218,851,000	217,704,000
1999	236,275,000 2/4/	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	458,182,000	468,037,000	468,037,000
Rescission				(3,042,000)
2004	478,072,000	478,072,000	482,372,000	482,222,000
Rescission				(3,149,000)
2005	492,670,000	492,670,000	496,400,000	492,670,000
Rescission				(4,062,000)
2006	490,959,000			

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 $128,000
4/ 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 2004 Actual	FY 2005 Appropriation	FY 2006 Estimate
Office of the Director	6	6	6
Office of Administrative Management	16	16	16
Office of Policy, Communications and Educati	14	14	14
Division of Intramural Research	219	221	225
Division of Extramural Research	18	28	28
Total	273	285	289
FTEs supported by funds from Cooperative Research and Development Agreements	(1)	(1)	(1)

FISCAL YEAR	Average GM/GS Grade
2002	10.9
2003	10.8
2004	11.6
2005	11.6
2006	11.6

NATIONAL INSTITUTES OF HEALTH
National Human Genome Research Institute

Detail of Positions

GRADE	FY 2004 Actual	FY 2005 Appropriation	FY 2006 Estimate
Total - ES Positions	1	1	1
Total - ES Salary	$132,261	$138,874	$145,818
GM/GS-15	23	23	23
GM/GS-14	11	11	11
GM/GS-13	31	33	33
GS-12	42	42	42
GS-11	19	18	18
GS-10	3	3	3
GS-9	12	14	14
GS-8	8	12	12
GS-7	7	10	10
GS-6	0	3	3
GS-5	0	0	0
GS-4	0	0	0
GS-3	1	1	1
GS-2	0	0	0
GS-1	0	0	0
Subtotal	157	170	170
Grades established by Act of July 1, 1944 (42 U.S.C. 207):			
Assistant Surgeon General			
Director Grade	1	1	1
Senior Grade	1	1	1
Full Grade	1	1	1
Senior Assistant Grade			
Assistant Grade			
Subtotal	3	3	3
Ungraded	115	115	119
Total permanent positions	158	161	165
Total positions, end of year	276	289	293
Total full-time equivalent (FTE) employment,end of year	273	285	289
Average ES salary	$132,261	$138,874	$145,818
Average GM/GS grade	11.6	11.6	11.6
Average GM/GS salary	$72,580	$73,158	$74,420

NATIONAL INSTITUTES OF HEALTH
National Human Genome Research Institute

New Positions Requested

	FY 2006		
	Grade	Number	Annual Salary
Roadmap Staff Scientist	Title 42	4	$100,000
Total Requested		4	