Summary of Research Activities by Key Approach and Resource

Genomics

In the early 1950s, the race to discover the structure of DNA was on. At Cambridge University, James Watson and Francis Crick made physical models to narrow the possible DNA structure. At King's College in London, Maurice Wilkins and Rosalind Franklin took an experimental approach, looking at x-ray diffraction images of DNA. Based partially on Rosalind Franklin's data, Watson and Crick built a model in which each strand of the DNA molecule was a template for the other, allowing DNA to make identical copies of itself at each cell division. The structure so perfectly fit the experimental data that it was almost immediately accepted. Elucidating the structure of DNA has been called the most important biological work of the last 100 years, and the field it opened may be the scientific frontier for the next 100.

Introduction

Genomics is the study of an organism's entire genome—the complete assembly of DNA (deoxyribonucleic acid), or in some cases RNA (ribonucleic acid)—which transmits the instructions for developing and operating a living creature. It focuses not just on individual genes but also on the functioning of the genome as an interrelated network, and it is a new, rapidly expanding field of biological and medical research.

DNA is made up of four chemical compounds called "nucleotides"—adenine, thymine, guanine, and cytosine—denoted by the letters A, T, G, and C. These nucleotides are assembled in two parallel strands that are connected in the form of a double helix, and each nucleotide in one strand always links to the same partner on the other strand: A always pairs with T; C always pairs with G. Each of these pairings is referred to as a "base pair." The human genome consists of about 3 billion base pairs, packaged into 23 sets of chromosomes, in virtually every cell in the body. Identifying the base pairs—and thus the letters—and the order in which they appear on any stretch of DNA is called "sequencing" that segment. DNA's double helical structure was discovered in 1953, and the human genome was fully sequenced just less than 50 years later, in 2003, after a 13-year, U.S.-led international effort called the Human Genome Project.

The sequencing of the human genome generated immense scientific excitement. It provided a new means of analyzing the functions of cells, tissues, and systems in the body and understanding and attacking the causes of disease. It enabled broad new scientific disciplines such as proteomics, the study of the structure and function of all the proteins produced by the body in response to instructions carried by the genes. It also gave many people the impression that all the questions of biology had been answered and that the genome had been fully decoded. This is not so. Sequencing the genome indicated the order of the letters; the question now is how precisely the words are written and what they mean.

Every human disease or disorder has a genetic component. Some heritable diseases, such as cystic fibrosis or Huntington's disease, result from mutations to single genes—changes that disrupt their proper functioning. The role of genes is more complicated in most other diseases. Some diseases arise as a result of spontaneous gene mutations that occur during a person's lifetime; others are caused by complex cascades of changes in gene expression triggered by environmental factors. Differences as small as one letter in a stretch of DNA can cause disease directly or make people respond differently to particular pathogens or drugs. A single DNA base change in the "spelling" of the genome sequence—called a single nucleotide polymorphism, or SNP—also can help

researchers track down genes involved in disease. Heart disease, asthma, and myriad other diseases appear to have multiple genetic factors, although all the genes involved have not been identified. Many types of cancer are caused by damage to one or more genes that leads to further mutations as cells divide.

Scope of NIH Activity in Genomics Research

Virtually every NIH IC engages in some genome-related research. NCI sponsors an array of gene-oriented projects, including an effort to compile The Cancer Genome Atlas, a catalogue of the many genetic changes that occur in cancer cells. NHLBI supports a major epidemiological project, the Framingham Genetic Research Study, to search for genetic links to disease in 9,000 study subjects across three generations. NIAID's Microbial Genome Sequencing Centers program is sequencing the genomes of many disease-causing microorganisms, including the fast-mutating RNA virus that causes influenza, seeking information that may help design vaccines or therapies to avert worldwide pandemics.

NIH researchers and grant recipients also are sequencing other nonhuman genomes, and not just the genomes of our mammalian relatives such as the chimpanzee, to highlight stretches of DNA that have remained similar across species for millions of years. Such similarities—or small differences in otherwise similar stretches of DNA—can help determine the roles and importance of particular sequences, and also may point the way toward therapies for diseases that affect humans. AIDS, caused by the human immunodeficiency virus, is one such disease.

An international consortium led by NHGRI has begun an effort to identify every functional element in the human genome, called the Encyclopedia of DNA Elements (ENCODE) project. Initial results reveal that genes do not operate independently but are part of a complex network, and that most of the genome's "noncoding" DNA, that is, sequences that are not part of a gene, is not "junk" but appears to have important, heretofore unknown, functions.

Toward an Era of "Personalized Medicine"

ENCODE and other NIH programs also aim to develop new technologies to reduce the cost of genome sequencing and otherwise aid in understanding the human genome. This includes the development of computer techniques and software to organize and analyze immense amounts of data, which are made available free of charge to all qualified researchers via public databases. When the Human Genome Project began in 1990, DNA sequencing cost about $10 for each base pair. By 2007, that had been reduced to less than 1 cent, or less than $20 million for sequencing a full human-sized genome.

Ultimately, NIH would like to reduce the cost of sequencing an entire human genome—all 3 billion base pairs—to $1,000 or less, making possible a new era of "personalized medicine." When costs are reduced to the point that sequencing an individual patient's genome is feasible, and when the impact of small genetic changes on disease progression and therapy is better understood, clinicians will have powerful new methods with which to defend their patients' health.

Summary of NIH Activities

In FYs 2006 and 2007, NIH made significant progress toward exploiting the raw data of the human genome sequence and translating it into advances in human health. NIH-funded researchers and other scientists have laid the foundation for a scientific revolution—a truly new paradigm that will soon change medical research and the practice of medicine itself, moving beyond a one-size-fits-all approach. Most of the changes in practice and research that will matter for human health and our understanding of basic human traits have not yet happened. However, the next decade will yield the fruits of this foundational work, leading scientists increasingly closer to better means for preventing, diagnosing, and treating disease.

Among NIH's key accomplishments in the field of genomics in the FY 2006-2007 period were:

- Collaborating in the completion of the haplotype map of the human genome, known as the "HapMap": An

international effort, the HapMap identifies the location of more than 3.1 million SNPs along the 3 billion bases of human DNA. SNPs are relatively common variations that serve as markers for whole neighborhoods of gene-carrying DNA. As such, they are signposts by which researchers can compare individuals' genomes and hunt for genetic mutations that may be involved in disease.

- Confirmation that the genome is not a simple string of independent genes, but rather a complex network, for which the elements and functions are still incompletely understood: In a program that is still ongoing, the international ENCODE project (the acronym stands for " **ENC**yclopedia **O**f **DNA E**lements") conducted multiple analyses of carefully selected DNA segments totaling approximately 1 percent of the human genome—about 30 million base pairs—in an attempt to identify every functional element and to figure out which methods worked best for identifying functional elements. ENCODE's next phase is to determine the functions of the other 99 percent of the genome. NIH has launched a similar project, dubbed "modENCODE," to apply the same strict scrutiny to the genomes of two common laboratory model animals, the fruit fly *Drosophila melanogaster and the round worm Caenorhabditis elegans.*

- Full sequencing of additional vertebrate and nonvertebrate animal genomes: Completed vertebrate animal genomes include those of the dog, the horse, the cow, the opossum, the honeybee, and two nonhuman primates—the rhesus macaque and the chimpanzee. By 2007, NIH and NIH-funded centers also had sequenced thousands of different viruses, hundreds of bacteria, and many unicellular parasites, including two that cause malaria—not to mention two mosquito species, one a vector for human malaria, the other for avian malaria. Such data enable scientists to compare the genomes of different organisms and identify elements that are similar in many species. Scientists suspect that genetic elements that have remained similar in different species over millions of years of evolution have important functions; thus, similarities between different species' genomes may provide clues about human disease processes. Sequencing of other nonhuman genomes also is a major ongoing NIH program.

- Development of new laboratory tools and methods, and new computer algorithms for analyzing immense quantities of data, in order to reduce the cost of genome sequencing: A major goal of NIH sequencing programs is to reduce costs so that in time, physicians will be able to collect and use genomic data from their own patients—moving sequencing from blue-sky science to bedside therapy.

- Confirmation that genetic differences underlie much of an individual's response to medications, and that those genetic differences can be detected and potentially used to develop personalized treatment approaches: For example, in recent research, patients with two copies of a particular version of the serotonin 2A receptor gene responded significantly better to the antidepressant drug citalopram, a selective serotonin reuptake inhibitor, than did patients with different versions of the gene. Some day, such analyses could allow physicians to choose drugs tailored to individual patients rather than by a one-size-fits-all approach.

- New tests for diagnosing once-puzzling diseases and potential new therapies to treat them: Identifying the gene or genes involved in a disease can help scientists understand how the defect results in malfunction and thus point the way toward treatments. This approach is still new, but shows promise. For example, in 2003, NIH researchers identified the gene responsible for Hutchinson-Gilford progeria syndrome, which causes premature aging and heart disease in children and usually causes death by the teen years. They discovered that a single point mutation—a one-letter misspelling—in the gene known as *LMNA* produces a defective structural protein, which in turn causes misshapen nuclei in the patient's cells. Two years later, scientists following up on the discovery showed that an existing anticancer drug might correct the damage. Now, a 3-year clinical trial of this potential therapy for a devastating childhood disease is under way in the NIH Clinical Center.

The HapMap and Genetic Variation

Completion of the first phase of the HapMap in October 2005 by an international consortium of hundreds of researchers in six countries was one of the most significant developments in genomic research since the sequencing of the human genome in 2003.

The HapMap is the basic platform upon which most current genomic studies of human diversity are now built. It details the location of millions of relatively common single-letter variations in the human genome, that is,

variations that occur in at least 5 percent of people. The HapMap achieved two important goals: (1) it discovered most of the common variants in the genome and (2) it determined how these variants travel in "neighborhoods," or haplotypes, making it possible to track only a small percentage of all of the variants directly, allowing the rest to be inferred. It enables researchers to conduct studies that were simply impossible just a few years ago. When the HapMap was published, a commentary in the journal Nature noted that it had "succeeded in a spectacular way."[7]

In the early trailblazing years of genetic research, scientists largely were limited to seeking the single genes involved in classic, Mendelian-inherited diseases. A disease caused by a single damaged or inactive gene—such as cystic fibrosis or sickle cell anemia—could be traced in family history and then laboriously hunted down by trial-and-error comparisons of genetic variation across hundreds of families. However, diseases that involve several genes, where no single gene has a very large effect, have eluded such analysis, and most, if not all, human diseases involve a complex interaction of multiple genes. This is further complicated by the interactions of genes with environmental factors such as exercise, stress, and exposures.

The HapMap, together with advanced sequencing technology, now enables researchers to seek out the genetic roots of common, complex diseases by comparing and contrasting hundreds of thousands of points of variation among people. Thus, NIH-funded researchers have pioneered a whole new approach to genetic studies, called genome-wide association studies (GWAS, pronounced "gee-was").

The Big Picture: Genome-Wide Association Studies

GWAS examine not just a single stretch of DNA or the expression of a protein in a laboratory dish, but rather points of similarity and difference in the entire DNA sequences of people with or without particular diseases. In a typical GWA study, the genomes of 1,000 or more people with a particular disease are compared with the genomes of a similar number who are free of the disease. (Samples from many thousands of people are better, of course; the greater the number of individuals, the more accurate the study.) Theoretically, the "big picture" comparison of peoples' genomes will signal the presence of blocks of DNA that carry a gene or genes involved in the disease in question.

In the short time since they were devised, GWAS conducted by NIH or NIH-funded researchers have, among other discoveries:

- Identified a common genetic variation that significantly raises the risk of age-related macular degeneration. The finding strengthened our understanding of the link between the inflammation pathway and a devastating eye disease that often leads to blindness, and suggested a new treatment that is now under clinical study.
- Uncovered several genes that appear to play a role in bipolar disorder. One, which is active in the pathway through which lithium operates on the disorder, suggests a new treatment approach—seeking ways to regulate the enzyme involved, known as DGKH. Others may point scientists toward new directions for research.
- Located at least 10 sites of gene variants associated with type 2 diabetes—most of them never before identified. One of the sites includes two genes that had been studied in cancer, but never before associated with diabetes.
- Discovered three gene variants that may affect the ability of a person infected with HIV to control viral load and prevent or delay progression to AIDS. In addition to offering new approaches to anti-AIDS therapy, the apparent involvement of an immune system gene, *HLA-C*, may suggest a new avenue for research aimed at developing an HIV vaccine.
- Identified five new potential sites for breast cancer susceptibility genes. At least three of the five have been implicated in cell growth or cell signaling, rather than DNA repair or hormone metabolism, pointing the way toward new areas for basic research.

---

[7] Goldstein DB, Cavalleri GL. *Nature* 2005;437:1241-2, PMID: 16251937

- Found a major site associated with prostate cancer risk on chromosome 8, with several different haplotypes that confer risk, and which may explain a substantial fraction of the increased risk in African Americans.
- Discovered additional variants of genes that increase the risk for colon cancer, Crohn's disease, rheumatoid arthritis, multiple sclerosis, Alzheimer's disease, gallstones, celiac disease, atrial fibrillation, glaucoma, lupus, coronary artery disease, and type 1 diabetes, among others.

With support from NIH and other sources, scientists will follow up on these discoveries through further genomic research to confirm and refine findings and, through nongenomic investigations, to discover preventions, diagnostics, and treatments.

A new, large-scale GWA study of cardiovascular and other chronic diseases is now under way in Framingham, Massachusetts. In collaboration with the Boston University School of Medicine, NIH is screening DNA from subjects enrolled in the long-running Framingham Heart Study—up to 500,000 analyses of DNA from 9,000 people who have been followed over three generations since 1948. The Framingham study has been a key source of knowledge about heart disease, stroke, and other chronic diseases; the new genome-wide association analyses will add immensely to understanding the genetic factors involved.

The genome-wide association approach also is at the heart of a major effort to explore the relationship between genes and the environment in many common diseases. The trans-NIH Genes, Environment and Health Initiative (GEI), will add an additional step to GWAS: It will monitor the differing environmental factors to which people in the study are exposed, as well as genomic differences, to determine not only which genes may be involved in particular diseases, but also what specific environmental influences trigger disease in susceptible individuals. NIH awarded its first GEI research grants in 2007; in the program's first year, NIH plans to sponsor eight GWAS, two genotyping centers and more than 30 environmental technology projects—including efforts to develop small environmental sensors that people can wear or carry, like cell phones or iPods, to measure environmental exposures. The environment includes not only the chemical environment but also exposure to the behavioral environments of dietary intake, physical activity, psychosocial stress, and addictive substances.

In 2006, NIH also launched a 3-year series of GWAS seeking genes that raise the risk of prostate and breast cancer, known as the Cancer Genetic Markers of Susceptibility project.

Supplementing NIH's research efforts, a unique public-private partnership known as the Genetic Association Information Network (GAIN) has begun funding additional GWAS analyses of common diseases, beginning in late 2006 with studies of schizophrenia, bipolar disorder, diabetic nephropathy, attention deficit hyperactivity disorder (ADHD), major depression, and psoriasis. Managed by the nonprofit Foundation for the National Institutes of Health, GAIN is funded by private-sector partners, including Pfizer, Affymetrix, Perlegen Sciences, Abbott, and the Broad Institute of Massachusetts Institute of Technology and Harvard University.

As with other genetic data produced by NIH or NIH-funded researchers, all data from GWAS—including data resulting from the public-private GAIN studies—are made freely available to biomedical researchers worldwide through databases maintained by NIH. The trans-NIH GWAS Policy, released in August 2007, includes establishment of a central data repository of de-identified genetic (genotypic and phenotypic) data, and creates a more uniform approach to expanding investigators' access to GWA study data. Implementation guidance was released to intramural and extramural scientists in November 2007, and the policy became effective on January 25, 2008. Under the new guidelines, information is deposited into databases immediately, rather than being held back for months until it is published in scientific journals. This accelerates data availability, thereby facilitating the development of better diagnostic tools and the design of new, safe, and effective treatments.

Decoding Cancer

Understanding and developing new treatments for human cancer has long been a major goal of genetic research. Since the 1990s, a growing number of individual genes that predispose an individual to cancer have been

identified, such as the breast cancer genes *BRCA1 and BRCA2*. But it has become clear that cancer is not a disease caused by a single gene. Instead, cancer is known to involve many different forms of out-of-control cell growth and to be influenced by many different genes. A few of these mutations are inherited from a person's parents, but most occur during a lifetime of cell division, or, in some instances, are caused by some external environmental factor. (In some cases, the external factor is known, such as cigarette smoking in lung cancer; however, even smoking does not explain all cases of lung cancer, nor do all smokers get lung cancer.)

In its continuing effort to unravel human cancers, in 2006 NIH launched The Cancer Genome Atlas. In a 3-year pilot project, scientists at more than a dozen institutions will sequence and analyze genetic changes in tissue samples donated by thousands of brain, lung, and ovarian cancer patients. They will try to identify the specific alterations in genes associated with cancer and determine the genetic signatures of different cancer subtypes. Some cancers develop slowly; others are aggressive. Some respond to a particular chemotherapy; others do not. If the effort succeeds, The Cancer Genome Atlas will be expanded to cover other types of cancer (see also the section on *Cancer* in Chapter 2).

NIH already assembles—and makes available to medical researchers worldwide—a vast collection of genomic data resources and computer tools for accessing and analyzing that data, through such efforts as its Cancer Genome Anatomy Project and the Mammalian Gene Collection.

Nonhuman Genomes

NIH also continues to fund sequencing of the genomes of nonhuman organisms. Sequencing projects under way include the orangutan, the gorilla, and the gibbon genomes. In addition, NIH sponsors an ongoing program of sequencing the genomes of microorganisms that prey on humans. These efforts provide insights not only into potential approaches to controlling these organisms, but also into basic understanding of DNA, genes, and genomes. For example, studies of fruit flies and the round worm C. *elegans* have, for decades, been a source of basic knowledge about genes and their function that have enlightened studies in humans. Rats and mice are also key laboratory model animals and are hardly irrelevant to human genetics; more than 99 percent of human genes have analogs in the mouse. Studies of other mammals also can cast light on human disease. For example, a study of the dog genome suggested a possible new connection between human cancer and a gene that had never before been considered as a cancer suspect. The 2007 study revealed that a single gene is the major determinant of a dog's size, from Chihuahua to Great Dane. That gene, *IGF-1*, which codes for the hormone insulin-like growth factor-1, is similar to a gene in humans. If *IGF-1* is so important to size regulation in dogs, researchers say, it also may be involved in cell proliferation, and possibly cancer, in humans.

As is the case with humans, scientists can learn even more when they have data from many representative microbes of the same kind. For example, NIH has collected and sequenced the whole genomes of more than 2,500 human and avian influenza samples. The data from this ongoing project may help researchers anticipate the frequent evolutionary mutations in the virus that make designing a vaccine so difficult. It also may enable them to predict whether, and when, the A/H5N1 avian flu virus will mutate into a form that can easily infect humans, and to design a vaccine to counteract it. The possibility of an avian flu breakout into humans raises fears of a disaster similar to the 1918 Spanish flu pandemic, which is estimated to have killed 1 to 2 percent of the total world population. In 2007, an NIH research team developed a strategy for predicting the mutations that would permit the avian flu virus to adapt to humans—as few as two mutations could do it—and it is now possible to monitor newly isolated viruses to assess whether this possibility is occurring.

Genome Sequencing and Technology

Virtually all NIH sequencing programs have a dual purpose. Their aim is not just to answer a conventional research question, such as what is the DNA sequence of this organism or that gene, but also to reduce the cost of sequencing itself, and to increase the speed and efficiency of the task of analyzing DNA sequences.

For example, a consortium of 11 teams of investigators known as the ENDGAME consortium—the acronym stands for Enhancing Development of Genome-wide Association Methods—is seeking new approaches to conduct GWAS,

aimed specifically at lowering their cost and enhancing their usefulness. The Large-Scale Sequencing Program, which involves several sequencing centers throughout the United States, not only produces sequence data on a wide range of organisms to answer research questions, but also seeks ways to cut sequencing costs.

NIH's Genome Technology Program focuses directly on the development of new methods for transcribing DNA sequences, comparing sequences to identify variations, and determining the effects of such variations on genetic function and thus human health. Such analyses require significant computer backup. Because the human genome comprises more than 3 billion DNA base pairs, there are more than 3 billion possible points of difference between the genomes of any two individuals, and a genome-wide association study may involve several thousand individuals. Without such analytic efforts—which DNA researchers call "annotating," and could not be accomplished without sophisticated and innovative computer programming—DNA sequences are simply disconnected strings of letters in an alien language.

Currently, the field is undergoing a revolution in sequencing technology. The cost of sequencing the entire genome of an individual human being has been reduced from several billion dollars to between $100,000 and $1 million. NIH's goal is to bring that cost down to $1,000—and to truly bring genomic science to the bedside. That era of personalized medicine may be only a few years away.

Notable Examples of NIH Activity

Key for Bulleted Items:

E = Supported through **E**xtramural research
I = Supported through **I**ntramural research
O = **O**ther (e.g., policy, planning, and communication)
COE = Supported through a congressionally mandated **C**enter **o**f **E**xcellence program
GPRA Goal = Concerns progress tracked under the **G**overnment **P**erformance and **R**esults **A**ct

The Big Picture: Genome-Wide Association Studies

Genome-Wide Association Studies (GWAS) and Database of Genotype and Phenotype (dbGaP): In December 2006, NIH released the initial dbGaP dataset using genome-wide association study data from the Age-Related Eye Diseases Study (AREDS), a landmark study of the clinical course of Age-related Macular Degeneration (AMD) and cataracts. AREDS documents, protocols, and aggregated data are made available with no restrictions. In order to protect patient confidentiality, de-identified individual-level patient characteristics and family data are accessible only by authorized investigators. Correlating phenotype and genotype data provides information about the genetic and environmental interactions involved in a disease process or condition, which is critical for better understanding complex diseases and developing new diagnostic methods and treatments. Using these data, recent studies have linked two genes with progression to advanced AMD. After controlling for other factors, certain forms of the genes increased risk of AMD progression 2.6- to 4.1-fold; smoking and body weight further increased risk with these gene variants.

- Seddon JM, et al. *JAMA* 2007;297:1793-800, PMID: 17456821
- For more information, see http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gap
- This example also appears in Chapter 2: *Chronic Diseases and Organ Systems* and Chapter 3: *Disease Registries, Databases, and Biomedical Information Systems*.

- (E) (**NEI**, NIA, NLM)

Genetic Association Information Network (GAIN): GAIN is a public-private partnership initiative that will elucidate the genetic factors influencing risk for many complex diseases. The resulting data will be made available in a central database managed by NIH for no-cost access by the scientific community. Of the six initial studies receiving funding through GAIN, four will target mental disorders: schizophrenia, bipolar disorder, major depression, and attention deficit hyperactivity disorder.

- For more information, see http://www.fnih.org/GAIN2/home_new.shtml
- For more information, see http://grants1.nih.gov/grants/guide/rfa-files/RFA-MH-07-060.html
- (O) (NHGRI)

Genome-Wide Genotyping in Parkinson's Disease (PD): NIH researchers ecently conducted genome-wide genotyping of publicly available samples from a cohort of 267 Parkinson's disease patients and 270 neurologically normal controls to identify any common genetic variability with significant effect on the risk for PD. The project has produced around 220 million data points in the 537 subjects, the largest collection of publicly available genotypes in a case-control cohort. The release of these data facilitates research on PD and other neurodegenerative disorders, and the genotypes from neurologically normal controls can be used as a comparison cohort for other studies, dramatically reducing the cost of future research.

- Fung HC, et al. *Lancet Neurology* 5:911-6, PMID: 17052657
- For more information, see http://www.nia.nih.gov/NewsandEvents/PressReleases/PR20060927parkinsons.htm
- This example also appears in Chapter 2: *Neuroscience and Disorders of the Nervous System*
- (E/I) (**NIA**, NINDS)

Enhancing Development of Genome-Wide Association Methods (ENDGAME): The ENDGAME consortium, which comprises 11 interactive teams of investigators, has been initiated to explore new approaches for designing and conducting GWAS of complex diseases. ENDGAME investigators are developing and testing innovative, informative, and cost-effective study designs and analytical strategies and tools for performing the studies. All strategies and tools developed will be made available to the scientific community. Results from ENDGAME are expected to enhance greatly the utility of GWAS for increasing understanding about genetic variations and their role in health and disease.

- This example also appears in Chapter 2: *Chronic Diseases and Organ Systems*.
- (E) (**NHLBI**, NCI, NHGRI, NIEHS, NIGMS)

Population Genomics, GAIN, and GEI: In February 2006, HHS announced the creation of two related groundbreaking initiatives in which NIH is playing a leading role. The Genetic Association Information Network (GAIN) and the Genes, Environment, and Health Initiative (GEI) will accelerate research on the causes of common diseases. GAIN is a public-private partnership among NIH, the Foundation for the NIH, Pfizer, Affymetrix, Perlegen, the Broad Institute, and Abbott. GEI is a trans-NIH effort combining comprehensive genetic analysis and environmental technology development to understand the causes of common diseases. Both GAIN and GEI are powered by completion of the "HapMap," a detailed map of the 0.1 percent variation in the spelling of our DNA that is responsible for individual predispositions for health and disease. Data from GAIN will narrow the hunt for genes involved in six common diseases. In June 2007, the first GAIN dataset, on attention deficit hyperactivity disorder, was released. GEI will provide data for approximately another 15 disorders and will develop enhanced technologies and tools to measure environmental toxins, dietary intake, and physical activity, as well as an individual's biological response to those influences.

- For more information, see http://www.genome.gov/19518664
- For more information, see http://www.genome.gov/19518663

- For more information, see http://genesandenvironment.nih.gov/
- For more information, see http://www.genome.gov/11511175
- This example also appears in Chapter 2: *Chronic Diseases and Organ Systems* and Chapter 3: *Technology Development*.
- (E/I) (NHGRI)

Genetic Roots of Bipolar Disorder Revealed by First Genome-Wide Study of Illness: According to NIH-funded research, the likelihood of developing bipolar disorder depends in part on the combination of small effects of variations in many different genes in the brain, none of which is powerful enough to cause the disease by itself.

- For more information, see http://www.nimh.nih.gov/press/mcmahon-bipolar-genetics.cfm
- This example also appears in Chapter 2: *Neuroscience and Disorders of the Nervous System*.
- (E) (NIMH)

Gene Expression Changes in Facioscapulohumeral Muscular Dystrophy (FSHD): Results from a genome-wide scan of skeletal muscle biopsies suggest a link between eye blood vessel defects and muscle defects that characterize FSHD. Patient participants were recruited from the National Registry for Myotonic Dystrophy and FSHD Patients and Family Members.

- Osborne RJ, et al. *Neurology* 2007;68:569-77, PMID: 17151338
- For more information, see http://www.niams.nih.gov/Funding/Funded_Research/registries.asp#dystrophy
- This example also appears in Chapter 3: *Disease Registries, Databases, and Biomedical Information Systems* and Chapter 2: *Neuroscience and Disorders of the Nervous System*.
- (E) (**NIAMS**, NCRR, NINDS)

Decoding Cancer

The Cancer Genome Anatomy Project (CGAP): The goal is to determine the gene expression profiles of normal, precancer, and cancer cells to improve detection, diagnosis, and treatment for the patient. The CGAP Web site makes various tools for genomic analysis available to researchers. Through worldwide collaborations, CGAP seeks to increase its scientific expertise and expand its databases for the benefit of all cancer researchers.

- For more information, see http://cgap.nci.nih.gov
- This example also appears in Chapter 2: *Cancer*.
- (E/I) (NCI)

Genome-Wide Association Studies of Cancer Risk: Beginning with the Cancer Genetic Markers of Susceptibility (CGEMS) initiative for breast and prostate cancer, NIH has capitalized on its long-term investment in intramural/extramural consortia by creating strategic partnerships to accelerate knowledge about the genetic and environmental components of cancer induction and progression. Using powerful new technology capable of scanning the entire human genome, these efforts have recently identified unsuspected genetic variants associated with increased risk for developing cancers of the prostate, breast, and colon. Additional scans, either planned or under way, will be directed at cancers of the pancreas, bladder, lung, and other organs. The results of these genome-wide studies, together with the follow-on studies planned to narrow the search for causal gene variants, promise to provide novel clinical strategies for early detection, prevention, and therapy. To expand upon these emerging opportunities, a new Laboratory of Translational Genomics (LTG) has been established to further characterize genetic regions associated with cancer susceptibility, and to identify gene-gene and gene-environment interactions. LTG will create opportunities for collaboration and data sharing in order to accelerate the translation of genomic findings into clinical interventions.

- For more information, see http://epi.grants.cancer.gov/Consortia/tablelist.html
- For more information, see http://epi.grants.cancer.gov/BPC3/cohorts.html
- For more information, see http://cgems.cancer.gov/index.asp

- For more information, see http://epi.grants.cancer.gov/PanScan/
- This example also appears in Chapter 2: *Cancer* and Chapter 3: *Epidemiological and Longitudinal Studies*.
- (E/I) (NCI)

The Cancer Genome Atlas (TCGA): TCGA is a comprehensive and coordinated effort to accelerate our understanding of the molecular basis of cancer through the application of genome analysis technologies, including large-scale genome sequencing. The goal of TCGA is to develop a free, rapidly available, publicly accessible, comprehensive catalogue, or atlas, of the many genetic changes that occur in cancers, from chromosome rearrangements to DNA mutations to epigenetic changes—the chemical modifications of DNA that can turn genes on or off without altering the DNA sequence. The overarching goal of TCGA is to improve our ability to diagnose, treat, and prevent cancer.

- For more information, see http://cancergenome.nih.gov/index.asp
- This example also appears in Chapter 2: *Cancer* and Chapter 3: *Technology Development.*
- (E/I) (**NCI**, NHGRI)

Nonhuman Genomes

The Dog Genome and Human Cancer: Cancer is the number one killer of dogs, and studying the major cancers in dogs provides a remarkably valuable approach for developing a better understanding of the development of cancer in humans. The clinical presentation, histology, and biology of many canine cancers very closely parallel those of human malignancies, so comparative studies of canine and human cancer genetics should be of significant clinical benefit to both. Furthermore, information gained from studying the genetic variant involved in dog size can provide important information for studying cell growth in humans and has the potential to be a useful tool in cancer research. A 2007 article by NIH researchers reported a genetic variant that is a major contributor to small size in dogs, followed by a second study finding that a mutation in a gene that codes for a muscle protein can increase muscle mass and enhance racing performance in dogs.

- Sutter NB, et al. *Science* 2007;316:112-5, PMID: 17412960
- Mosher DS, et al. *PLoS Genet* 2007;3:e79, PMID: 17530926s
- For more information, see http://www.genome.gov/25520294
- This example also appears in Chapter 2: *Cancer* and Chapter 3: *Molecular Biology and Basic Sciences*.
- (I) (NHGRI)

Microbial Genomics: NIH has made significant investments in two large-scale programs to sequence microbes and genomes over the last decade. Sequenced pathogens include hundreds of bacteria, fungi, parasites, invertebrate vectors of diseases, and viruses (including those pathogens that cause anthrax, influenza, aspergillosis, tuberculosis, gonorrhea, chlamydia, and cholera, and many that are potential agents of bioterrorism). NIH also provides comprehensive genomic, bioinformatic, and proteomic resources and reagents to the scientific community. These include (1) Microbial Genome Sequencing Centers, which rapidly produce high-quality genome sequences of human pathogens and invertebrate vectors of diseases, (2) The Pathogen Functional Genomics Resource Center, which provides functional genomic resources, (3) Bioinformatics Resource Centers, which provide access to genomic and related data in a user-friendly format, and (4) Proteomics Research Centers, which support research on the full set of proteins encoded in a microbial genome. The NIH Influenza Genome Sequencing Project has sequenced over 2,800 human and avian influenza isolates (as of November 28, 2007). NIH scientists recently exploited these data to explain the global spread of resistance to adamantanes, a first-generation class of anti-influenza drug.

- For more information, see http://www3.niaid.nih.gov/research/topics/pathogen/default.htm
- This example also appears in Chapter 2: *Infectious Diseases and Biodefense*.
- (E/I) (NIAID) (GPRA Goal)

Tools for Genetic and Genomic Studies in Emerging Model Organisms: In FYs 2006 and 2007, NIH funded eight grants that create genetic and genomic resources for model organisms whose genomes have been recently sequenced. These organisms include fish, invertebrates, and microbes used to understand human health, development, and disease. The resources include reagents and mutant lines, a center for high-throughput mutagenesis, genetic maps, databases, and stock centers.

- For more information, see http://grants.nih.gov/grants/guide/pa-files/PA-04-135.html
- (E) (NIGMS)

Human Microbiome Project: The human microbiome is the set of microbes that naturally inhabit the human nose, mouth, gut, vagina, and skin. The interactions between human hosts and these microbial communities at multiple body sites are known to be important for health, yet relatively little is known about them. The concept and plan for the NIH Roadmap Human Microbiome Project (HMP) was approved in 2007. By leveraging both the traditional approach to genomic DNA sequencing and the metagenomic approach (which allows the genomic sequencing of all microbes contained in a single sample), the HMP will lay the foundation for further longitudinal studies of human-associated microbial communities. Program initiatives are to characterize the genomes of the indigenous microbes of the human nose, mouth, gut, vagina, and skin, referred to as the "human microbiome," and determine whether individuals share a core human microbiome; to understand the relationship between the human microbiome and changes in human health; to develop novel technological and analytic tools needed to support these goals; to establish a data analysis and coordinating center and a resource repository; and to address the ethical, legal, and social implications raised by human microbiome research.

- For more information, see http://nihroadmap.nih.gov/hmp
- For more information, see http://www.genome.gov/26524200
- (E) (Roadmap–all ICs participate)

Scientists Complete Full Sequence of Opportunistic Oral Bacterium: Over the last decade, scientists have assembled the complete DNA sequences of several important oral bacteria. Now NIH-funded investigators have decoded and added another important bacterium, *Streptococcus sanguinis*, a key player in the formation of the oral biofilm, to the list. Although not regarded as a pathogen in the mouth, *S. sanguinis* is known to enter the bloodstream where it can colonize heart valves and contribute to bacterial endocarditis, a condition that kills an estimated 2,000 Americans each year. With the bacterium's genetic blueprint now publicly available online, scientists can better study the dynamics of biofilm formation and possibly tease out new leads to prevent tooth decay and periodontal disease. They also now can systematically identify and target sequences within the DNA of *S. sanguinis* that are critical to the infectious process, providing invaluable information in designing more effective treatments for endocarditis.

- Xu P, et al. *J Bacteriol* 2007;189:3166-75, PMID: 17277061
- This example also appears in Chapter 2: *Infectious Diseases and Biodefense*.
- (E) (NIDCR)

Genome Sequencing and Technology

Genome Technology and the $1,000 and $100,000 Genome Initiatives: DNA sequencing spells out the order in which our chemical building blocks are arranged, making DNA sequencing a powerful resource for biomedical research. Although DNA sequencing costs have dropped by more than three orders of magnitude since the start of the Human Genome Project, sequencing an individual's complete genome for medical purposes is still prohibitively expensive. Developing technology to make whole-genome sequencing more affordable would enable the sequencing of individual genomes to become part of routine medical care. The Genome Technology program supports research to develop new methods, technologies, and instruments to rapidly, and at low cost:

- Transcribe DNA sequences

- Check sequences for genetic variations (SNP genotyping)
- Aid research to understand the effects of genetic variations on genomic function

Additionally, NHGRI supports two types of sequencing grants: (1) "Near-Term Development for Genome Sequencing" grants support research aimed at sequencing a human-sized genome at 100 times lower cost than is possible today ($100,000) and (2) "Revolutionary Genome Sequencing Technologies" grants aim to develop breakthrough technologies that will enable a human-sized genome to be sequenced for $1,000 or less. Currently, only analyses of ~ 500,000 Single Nucleotide Polymorphisms (SNPs) are being performed commercially at this cost; an individual's complete genome sequence (~ 3 billion base pairs) would offer vastly more information.

- For more information, see http://www.genome.gov/10000368
- For more information, see http://www.genome.gov/19518500
- This example also appears in Chapter 3: *Technology Development*.
- (E) (NHGRI)

Large-Scale Sequencing Program: NIH's Large-Scale Sequencing Program funds three major research centers in the United States to conduct genetic sequencing. During and since the completion of the Human Genome Project, NIH-funded centers have used their industrial-scale enterprises to improve DNA sequencing methods, thereby substantially decreasing costs and increasing capacity. For many years, the Program has achieved twofold decreases in cost approximately every 20 months. One of the main projects now under way is the sequencing of the genomes of other primates, such as orangutan, baboon, gibbon, and marmoset (in addition to chimpanzee and macaque, which are complete). By comparing the human genome to that of other primates, researchers can find important information about both health and abilities that are uniquely human and those shared with other species. The Program also supports the genomic sequencing of human pathogens (organisms that cause disease in humans) and their vectors (the organisms that carry those pathogens). For other relevant NIH programs see previous section, Microbial Genomics. Also, many mammals are being sequenced to identify elements that are functionally important to human biology. These studies will undoubtedly unveil new biological insights to increase our understanding of how the human genome works.

- Rhesus Macaque Genome Sequencing and Analysis Consortium, et al. *Science*. 2007;316:222-34, PMID: 17431167
- For more information, see http://www.genome.gov/10001691
- This example also appears in Chapter 3: *Molecular Biology and Basic Sciences* and Chapter 3: *Technology Development*.
- (E) (NHGRI)

How Fast Is Evolution? Traditionally, scientists thought that evolution happened very slowly. They believed that it is quite rare to have major DNA changes (also called radical mutations) that benefit organisms and are passed on to future generations. Recently, NIH-funded researchers learned that in some cases, evolution can happen very quickly. By analyzing how DNA varies from person to person, and comparing human and chimpanzee DNA, the researchers discovered that radical mutations undergo a two-step selection process. Most mutations never make it past the first step, and slip out of the gene pool without being passed on to subsequent generations. But the rare mutations that survive this first cut spread rapidly throughout the species. These observations have relevance for our own species because, even though radical mutations represent only 10-12 percent of the differences between human and chimpanzee DNA, they may be responsible for some of the most significant differences between the two species.

- Gojobori J, et al. *Proc Natl Acad Sci U S A* 2007;104:3907-12, PMID: 17360451
- (E) (NIGMS)

Functional Genomics of Disease

Longevity Assurance Gene (LAG) Initiative and Interactive Network: The identification and functional characterization of genes and biological pathways controlling longevity and lifespan have advanced significantly, in

large part as a result of the efforts of scientists participating in the NIH-supported LAG Initiative and Network. The LAG Initiative has led to the identification of over 100 new longevity- associated genes, along with many other conserved biological processes and pathways that regulate longevity in a host of divergent species, including humans. These and similar discoveries are helping to illuminate disease processes, identify new predictive biomarkers, and facilitate identification of targets for preemptive drug therapy.

- (E) (NIA)

Women's Health Initiative: In January 2007, NIH awarded support for a dozen 2-year research projects to apply genomics, proteomics, and other innovative technologies to improve understanding of several major diseases that commonly affect postmenopausal women. The new endeavor builds on results of the long-running Women's Health Initiative, which conducted several clinical trials and an observational study to examine strategies for preventing heart disease, breast and colorectal cancers, and osteoporosis in a cohort of over 160,000 subjects. Investigators will use stored blood, DNA, and other biological samples and associated clinical data to analyze genetic factors and biological markers that may be useful in predicting disease outcomes or the effects of therapeutic and preventive regimens in postmenopausal women.

- For more information, see http://www.whiscience.org/baa/2006.php
- This example also appears in Chapter 2: *Chronic Diseases and Organ Systems* and Chapter 3: *Epidemiological and Longitudinal Studies*.
- (E) (NHLBI)

Inflammatory Bowel Disease Genetics Consortium: This consortium of researchers in the United States and Canada applies knowledge from the Human Genome Project to the identification of genetic factors influencing the development of inflammatory bowel diseases (IBD). A genome-wide screen of samples collected recently identified three IBD susceptibility genes. The identification of such genetic factors can provide key insights into disease development and targets for designing more effective therapies for IBD.

- Rioux JD, et al. *Nat Genet* 2007;39:596-604, PMID: 17435756
- For more information, see http://grants.nih.gov/grants/guide/rfa-files/RFA-DK-02-011.html
- This example also appears in Chapter 2: *Chronic Diseases and Organ Systems*.
- (E) (NIDDK)

A Multidisciplinary Approach to Nicotine Addiction: Nicotine addiction is the number one preventable public health threat, with enormous associated morbidity, mortality, and economic costs. NIH-supported research has generated new knowledge to support the development of more effective prevention messages and treatment approaches. Several notable examples characterize NIH's multidisciplinary approach to targeting the best treatment (or combination of treatments) for nicotine addiction. Genomic studies have recently uncovered a series of genes associated with nicotine addiction that could provide new targets for medications development and for the optimization of treatment selection. Pharmacologic studies, critical to understanding the basis of nicotine's mode of action, have recently revealed that its addictiveness may hinge upon its ability to slowly shut down or desensitize the brain's response to nicotine. A recent imaging study indicated that a part of the brain called the insula may play an important role in regulating conscious craving. This exciting finding provides a new target for research into the neurobiology of drug craving and for development of potentially more effective smoking cessation and other addiction treatments. Results of a Phase II clinical trial strongly suggest that a nicotine vaccine, which works by preventing nicotine from ever reaching the brain, may be a particularly useful tool for cessation programs in the not-too-distant future.

- For more information, see http://www.drugabuse.gov/ResearchReports/Nicotine/Nicotine.html
- This example also appears in Chapter 3: *Clinical and Translational Research*, Chapter 2: *Cancer*, and Chapter 2: *Neuroscience and Disorders of the Nervous System*.

- (E) (**NIDA**, NCI) (GPRA Goal)

The Collaborative Study on the Genetics of Alcoholism (COGA): In its 18th year, COGA is a multisite, multidisciplinary family study with the overall goal of identifying and characterizing genes that contribute to the risk for alcohol dependence and related phenotypes. COGA investigators have collected data from more than 300 extended families (consisting of more than 3,000 individuals) who are densely affected by alcoholism. Several genes have been identified including GABRA2, ADH4, ADH5, and CHRM2, which influence the risk for alcoholism and related behaviors such as anxiety, depression, and other drug dependence. In addition to genetic data, extensive clinical neuropsychological, electrophysiological, and biochemical data have been collected and a repository of immortalized cell lines from these individuals has been established to serve as a permanent source of DNA for genetic studies. These data and biomaterials are distributed to qualified investigators in the greater scientific community to accelerate the identification of genes influencing vulnerability to alcoholism. COGA will continue to identify genes and variations within the genes that are associated with an increased risk for alcohol dependence and will perform functional studies of the identified genes to examine the mechanisms by which the identified genetic variations influence risk.

- For more information, see http://zork.wustl.edu/niaaa/
- This example also appears in Chapter 2: *Chronic Diseases and Organ Systems*, Chapter 3: *Molecular Biology and Basic Sciences*, and Chapter 2: *Neuroscience and Disorders of the Nervous System*.
- (E) (NIAAA) (GPRA Goal)

New Genetics Tools Shed Light on Addiction: NIH-supported research is taking full advantage of the massive databases and rapid technologies now available to study how genetic variations influence disease, health, and behavior. Such genetic studies are critical to teasing apart the molecular mechanisms and the genetic predispositions underlying diseases like addiction. Investigators studying various neurological and psychiatric illnesses have already linked certain genes with specific diseases using custom screening tools known as "gene chips" (e.g., the neurexin gene has been found to play a role in drug addiction). A next-generation "neurochip" is being developed with 24,000 gene variants related to substance use and other psychiatric disorders. Applying this tool to addiction and other brain disorders will advance our understanding not only of vulnerability to addiction and its frequent comorbidities, but also of ways to target treatments based on a patient's genetic profile (i.e., a "pharmacogenetic" approach). To complement these efforts, NIH is investing heavily in the emerging field of epigenetics, which focuses on the lasting modifications to the DNA structure and function that result from exposure to various stimuli. Attention to epigenetic phenomena is crucial to understanding the interactions between genes and the environment, including the deleterious long-term changes to brain circuits from drug abuse. A focus on gene-environment interactions has recently been expanded to incorporate developmental processes, now known to also affect the outcome of these interactions. The resulting Genes, Environment, and Development Initiative (GEDI) seeks to investigate how interactions among these factors contribute to the etiology of substance abuse and related phenotypes in humans.

- For more information, see http://grants.nih.gov/grants/guide/rfa-files/rfa-da-07-012.html
- For more information, see http://nihroadmap.nih.gov/roadmap15update.asp
- This example also appears in Chapter 2: *Neuroscience and Disorders of the Nervous System* and Chapter 3: *Technology Development*.
- (E/I) (**NIDA**, NCI, NIAAA, NIMH) (GPRA Goal)

Clinical Proteomic Technologies Initiative for Cancer: The completion of the Human Genome Project in 2003 has been a major catalyst for proteomics research and NIH has taken a leading role in facilitating the translation of proteomics from research to clinical application through its Clinical Proteomic Technologies Initiative for Cancer. The overall objective of this Initiative is to build the foundation of technologies (assessment, optimization, and development), data, reagents and reference materials, computational analysis tools, and infrastructure needed to systematically advance our understanding of protein biology in cancer and accelerate discovery research and clinical applications.

- For more information, see http://proteomics.cancer.gov/
- This example also appears in Chapter 2: *Cancer* and Chapter 3: *Technology Development*.
- (E/I) (NCI)

Medical Sequencing: The completion of the human genome sequence as well as genomic sequences of numerous other organisms has already made a substantial impact on both biological and medical research. Public access to the raw data produced from these large-scale sequencing efforts has empowered many additional studies about the genomic contributions to disease. To expedite the transition from research data to medical practice, NIH supports initiatives that both drive technology that will make whole genome sequencing affordable and produce data useful to biomedical research. Making the sequencing of any individual's complete genome affordable will allow personalized estimates of future disease risk and improve prevention, diagnosis, and treatment of disease. NIH's medical sequencing program is utilizing DNA sequencing to identify the genes responsible for rare, single-gene diseases; sequence all of the genes on the X chromosome to identify the genes involved in sex-linked diseases; and survey the range of variants in genes known to contribute to common diseases.

- For more information, see http://www.genome.gov/15014882
- This example also appears in Chapter 2: *Chronic Diseases and Organ Systems*.
- (E/I) (NHGRI)

Systems Biology Approach to Salivary Gland Physiology: Previous research has catalogued the genes and proteins expressed in the salivary glands. This initiative puts those catalogues into context by defining when and where genes and proteins are expressed and how they function as parts of a fully integrated biological system. The initiative combines the power of mathematics, biology, genomics, computer science, and other disciplines to translate this highly detailed information into more precise and practical leads to treat Sjögren's syndrome, a debilitating autoimmune disorder that affects millions of Americans. The initiative also will help in learning to use saliva as a diagnostic fluid for a variety of conditions, from AIDS to cancer to diabetes.

- For more information, see
  http://www.nidcr.nih.gov/GrantsAndFunding/See_Funding_Opportunities_Sorted_By/ConceptClearance/CurrentCC/SysAppySal.htm
- For more information, see http://grants2.nih.gov/grants/guide/rfa-files/RFA-DE-08-001.html
- This example also appears in Chapter 2: *Autoimmune Diseases*, Chapter 2: *Chronic Diseases and Organ Systems*, and Chapter 2: *Infectious Diseases and Biodefense*.
- (E) (NIDCR)

Genetics of Kidneys in Diabetes (GoKinD): This program facilitates investigator-driven research into the genetic basis of diabetic kidney disease through a biospecimen repository. Individuals with type 1 diabetes were screened to identify two subsets, one with clear-cut kidney disease and another with normal kidney function despite long-term diabetes. Nearly 10,000 DNA, serum, plasma, and urine samples—plus genetic and clinical data—from more than 1,700 adults with diabetes have been collected. The entire GoKinD collection is being genotyped for whole genome association studies as part of the previously described Genetic Association Information Network (GAIN).

- Mueller PW, et al. *J Am Soc Nephrol* 2006;17:1782-90, PMID: 16775037
- For more information, see http://www.jdrf.org/index.cfm?fuseaction=home.viewPage&page_id=B9C33021-1321-C834-0382E079E7865807
- This example also appears in Chapter 2: *Chronic Diseases and Organ Systems* and Chapter 3: *Clinical and Translational Research*.
- (E) (NIDDK)

Environmental Genomics: NIH's Environmental Genome Project (EGP) was set up to catalogue all of the common variants, or single nucleotide polymorphisms (SNPs), in the coding and noncoding regions of the selected candidate

genes. These candidate genes were chosen to fall into eight categories: cell cycle, DNA repair, cell division, cell signaling, cell structure, gene expression, apoptosis (cell death), and metabolism. Since 2005, EGP has been expanded to include resequencing of factors controlling epigenetic modification of gene expression and nuclear receptors or other environmentally responsive genes. The newest NIH initiative on Environmental Genomics is supporting studies of the mechanisms of susceptibility to environmentally influenced diseases. This research is focusing on the critical common pathways through which environmental factors influence human health and the determinants of individual and population susceptibility to these stressors. Each application for this program was required to have a cross-stressor, cross-strain, and/or cross-species comparison depending on which comparative biology approach was most appropriate for the system of study. Two distinct approaches to utilizing comparative biology for understanding environmentally induced disease are used: (1) a genetically driven approach to define the genetic-environment interactions that contribute to the pathophysiologic responses and individual susceptibility or protection from disease and (2) a pathway and network-driven approach to defining molecular mechanisms that mediate the pathophysiological responses to toxins.

- For more information, see http://www.niehs.nih.gov/research/supported/programs/egp/
- (E) (NIEHS)

The NIH Pharmacogenetics Research Network (PGRN): NIH established the PGRN in 2000 to study how genes affect the way a person responses to medicines. The network includes 12 interdisciplinary research groups, each focused on a specific problem. Recently, one team (the Pharmacogenetics of Anticancer Agents Research Group) identified 63 genetic variants that regulate human responses to the anticancer drug etoposide. The drug can cause severe side effects, including leukemia. Knowing the genetic basis of these side effects will help scientists develop tests to identify which cancer patients can be treated safely with etoposide.

- Huang RS, et al. *Proc Natl Acad Sci U S A* 2007;104:9758-63, PMID: 17537913
- For more information, see http://www.nigms.nih.gov/Initiatives/PGRN
- (E) (NIGMS)

DNA Test for Charcot-Marie-Tooth Disease: Charcot-Marie-Tooth disease, one of the most common inherited neurological disorders, affects one in 2,500 people in the United States. Its symptoms start in early adulthood and include progressive arm and leg pain that leads to difficulty walking and manipulating objects. Using a special strain of mice, new genomic technologies, and information from the mouse and human genome sequences, NIH-funded researchers rapidly identified a mutation that causes a subtype of the disease. Knowledge of the specific gene defect will enable development of a DNA test to confirm the diagnosis in patients and predict risk for family members.

- Chow CY, et al. *Nature* 2007;448:68-72, PMID: 17572665
- For more information, see http://www.med.umich.edu/opm/newspage/2007/charcot.htm
- This example also appears in Chapter 2: *Neuroscience and Disorders of the Nervous System*.
- (E) (**NIGMS**, NINDS)

How the Genes in Cells Are Turned On and Off: In any cell, only a small fraction of the genes are activated. Scientists know that DNA is rolled around protein spools into structures called nucleosomes. They suspect that a gene's position on the nucleosome determines whether it is activated. Recently, NIH-funded investigators used state-of-the-art techniques to discover a DNA sequence that appears to mark the start of activated genes in yeast cells (a similar sequence is predicted to play the same role in human cells). The sequence appears at the same place on almost all of the thousands of nucleosomes in the study—a location that is accessible to the proteins that activate genes. Improper gene activation is linked to cancer and other diseases, therefore identification of a DNA sequence that regulates gene activation will help researchers prevent, detect, or correct problems with gene activation that are associated with these diseases.

- [Albert I, et al *Nature* 2007;446:572-6](#), PMID: 17392789
- (E) (NIGMS)

Gene Influences Antidepressant Response: Whether depressed patients will respond to an antidepressant depends, in part, on which version of a gene they inherit. Having two copies of one version of a gene that codes for a component of the brain's mood-regulating system increased the odds of a favorable response to an antidepressant by up to 18 percent, compared to having two copies of the other, more common version.

- For more information, see http://www.nimh.nih.gov/press/stardgene.cfm
- This example also appears in Chapter 2: *Neuroscience and Disorders of the Nervous System*.
- (E) (NIMH)

Potential Therapy for Children Afflicted With Progeria Syndrome: Hutchinson-Gilford progeria syndrome (HGPS) is a genetic disorder of accelerated aging. In addition to other symptoms of aging, HGPS patients suffer from accelerated cardiovascular disease and often die in their teen or even pre-teen years from heart-related illnesses. No treatments are currently available for HGPS; however, recent work led by NHGRI researchers indicates that farnesyltransferase inhibitors (FTIs), a class of drugs originally developed to treat cancer by blocking the growth of tumor cells, are capable of reversing the effects of the defective HGPS protein, lamin A. Ongoing studies in a mouse model have validated the results of preliminary experiments, and a clinical trial of FTIs in children with progeria began in 2007. In FY 2008, researchers plan on expanding the study to investigate whether FTIs are capable of reversing the detrimental effects after progression of the cardiovascular anomalies that are seen in the mouse model. The development of biological assays to assess the effects of FTI treatment on the patients' cells is in progress to monitor potential beneficial effects of the clinical trial. In addition, it has been demonstrated that the progerin protein is present in small amounts in normal aging tissues. The investigation of this phenomenon is being pursued as a contributory factor to the normal aging process.

- [Cao K, et al. *Proc Natl Acad Sci U S A* 2007;104:4949-54](#), PMID: 17360355
- [Capell BC, et al., *Proc Natl Acad Sci U S A* 2005;102:12879-84](#), PMID: 16129833
- For more information, see http://www.genome.gov/15515061
- This example also appears in Chapter 3: *Clinical and Translational Research* and Chapter 2: *Life Stages, Human Development, and Rehabilitation*.
- (I) (NHGRI)

Genomic Studies of Autism: NIH has supported a number of studies that are pointing to potential genetic causes of autism.

- For more information, see http://www.nimh.nih.gov/science-news/2007/tiny-spontaneous-gene-mutations-may-boost-autism-risk.shtml
- For more information, see http://www.nimh.nih.gov/science-news/2007/largest-ever-search-for-autism-genes-reveals-new-clues.shtml
- For more information, see http://www.nimh.nih.gov/science-news/2006/gene-linked-to-autism-in-families-with-more-than-one-affected-child.shtml
- For more information, see http://www.nimh.nih.gov/science-news/2007/new-tests-may-help-researchers-detect-genetic-basis-for-autism.shtml
- This example also appears in Chapter 2: *Neuroscience and Disorders of the Nervous System*.
- (E) (**NIMH**, NCRR, NICHD, NINDS)

Resources

Rodent Model Resources for Translational Research: Mouse and rat models are the primary testbed for preclinical research and have played a vital role in most medical advances in the last century. Rodent models comprise about 90 percent of all animal studies enabling a wide range of genetic and physiological research on human disease. NIH

plays a major role in supporting the availability of normal and mutant mice and rats for translational research. Recent accomplishments include:

- Knockout Mouse Project (KOMP)—a Trans-NIH initiative to individually inactivate each protein-coding mouse gene to better understand the genetic functions of the estimated 22,000 mouse genes, which are, in many cases, very similar to human genes.
- KOMP Repository—established in FY 2007 to acquire and distribute the mouse models produced by the KOMP.
- Mutant Mouse Regional Resource Centers—distribution of genetically engineered mice increased by 50 percent in FY 2006 because of increased demand.
- Rat Resource and Research Center—acquisition and distribution of rat models increased by 50 percent in FY 2006 because of increased demand.

  - For more information, see http://www.genome.gov/17515708
  - For more information, see http://www.genome.gov/25521840
  - For more information, see http://www.mmrrc.org/
  - For more information, see http://www.nrrrc.missouri.edu/
  - For more information, see http://www.ncrr.nih.gov/comparative_medicine/resource_directory/rodents.asp
  - This example also appears in Chapter 3: *Clinical and Translational Research*.
  - (E) (NCRR)

NIMH Genetics Repository: Over the last 9 years, NIMH has built the infrastructure for large-scale genetics studies through the NIMH Human Genetics Initiative. Through this Initiative, NIMH established a repository of DNA, cell cultures, and clinical data, serving as a national resource for researchers studying the genetics of complex mental disorders.

  - For more information, see http://nimhgenetics.org/
  - This example also appears in Chapter 3: *Disease Registries, Databases, and Biomedical Information Systems* and Chapter 2: *Neuroscience and Disorders of the Nervous System*.
  - (E) (NIMH)

Database of Genotype and Phenotype (dbGaP): Research on the connection between genetics and human health and disease has grown exponentially since completion of the Human Genome Project in 2003, generating high volumes of data. Building on its established research resources in genetics, genomics, and other scientific data, NIH established dbGaP to house this growing body of information, particularly the results of GWAS, which examine genetic data of subjects with and without a disease or specific trait to identify potentially causative genes. By the end of 2007, dbGaP included results from more than a dozen GWAS, including genetic analyses added to the landmark Framingham Heart Study and trials conducted under the Genetic Association Information Network. dbGaP is to become the central repository for many NIH-funded GWAS in order to provide for rapid and widespread distribution of such data to researchers and accelerate the advance of personalized medicine.

  - For more information, see http://view.ncbi.nlm.nih.gov/dbgap
  - This example also appears in Chapter 3: *Disease Registries, Databases, and Biomedical Information Systems* and Chapter 3: *Epidemiological and Longitudinal Studies*.
  - (I) (NLM)

Candidate Gene-Association Resource: Over the years, NHLBI has supported a number of major population studies that have collected extensive data on cardiovascular disease and its risk factors and manifestations. To increase the utility of the data for conducting genetic association studies, NIH initiated the Candidate Gene Association Resource program in FY 2006. This new resource will have the capacity to perform high-throughput genotyping for up to 50,000 subjects in cohort studies that have stored samples and data available on a wide array of characteristics (phenotypes) associated with heart, lung, blood, and sleep disorders. The linked genotype-

phenotype data will form an invaluable resource for investigators seeking to identify genetic variants related to those disorders.

- For more information, see http://public.nhlbi.nih.gov/GeneticsGenomics/home/care.aspx
- This example also appears in Chapter 2: *Chronic Diseases and Organ Systems* and Chapter 3: *Disease Registries, Databases, and Biomedical Information Systems*.
- (E) (NHLBI)

Framingham SNP-Health Association Resource (SHARe): The Framingham SHARe is a comprehensive new effort by NIH and the Boston University School of Medicine to pinpoint genes underlying cardiovascular and other chronic diseases. The program builds on the Framingham Heart Study (FHS), which was begun in 1948 to identify factors that contribute to cardiovascular disease, and on other NIH-funded research demonstrating that common but minute variations in human DNA, called single nucleotide polymorphisms (SNPs), can be used to identify genetic contributors to common diseases. The initiative will examine over 500,000 genetic variants in 9,000 study subjects across three generations. NIH will develop a database to make the data available to researchers around the world. The database will help researchers integrate the wealth of information collected over the years in the FHS with the new genetic data, resulting in an increased understanding of genetic influences on disease risk, manifestation, and progression. Because of its uniqueness in including three generations of subjects with comparable data obtained from each generation at the same age, the FHS is the first study to be included in the SHARe initiative. NIH is currently considering expansion of SHARe to include other large longitudinal studies such as the Jackson Heart Study and the new Hispanic Community Health Study.

- For more information, see http://www.nhlbi.nih.gov/new/press/06-02-06.htm
- This example also appears in Chapter 2: *Chronic Diseases and Organ Systems* and Chapter 3: *Epidemiological and Longitudinal Studies*.
- (E) (**NHLBI**, NLM)

Conserved Domain Database and RefSeq: NIH's Conserved Domain Database (CDD) is a powerful means to deduce the function of newly discovered proteins. CDD is particularly valuable to researchers working on drug development and those requiring a synthesis of information on protein biological function, 3-D structure, and sequence conservation. In FY 2006 NIH met its GPRA goal of developing methods to classify at least 75 percent of proteins from sequenced genomes according to evolutionary origin and biological structure. NIH also met the FY 2006 GPRA goal of building a high-quality collection of reference sequences (the RefSeq database) to provide a unified view of the best available genetic information on organisms.

- Marchler-Bauer A, et al. *Nucleic Acids* Res 2007;35:D237-40, PMID: 17135202
- Pruitt KD, *et al Nucleic Acids* Res 2007;35:D61-5, PMID: 17130148
- For more information, see http://www.ncbi.nlm.nih.gov/RefSeq/
- For more information, see http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml
- (I) (NLM) (GPRA Goal)

ENCODE: The ENCyclopedia Of DNA Elements (ENCODE) is an international research consortium organized by NIH that seeks to identify all functional elements in the human genome. The initial 4-year pilot phase has just been completed, and the consortium has published a series of papers describing a complex network in which genes and other regulatory mechanisms interact in complex ways. Other insights include the discovery that the majority of DNA in the human genome is transcribed into functional molecules, called RNA, and that these transcripts extensively overlap one another. These findings challenge long-held beliefs that the genome has small sets of genes and vast amounts of "junk" DNA. Until now, most studies have concentrated on the functional elements of specific genes, and have not provided information about functional elements in the vast majority of the genome that does not contain genes. ENCODE's exciting discoveries may well reshape the way scientists think about the genome and pave the way for more effective approaches to both understanding and improving human health.

- [The ENCODE Project Consortium, et al. *Nature* 2007;447:799-816](#), PMID: 17571346
- For more information, see http://www.genome.gov/10005107
- This example also appears in Chapter 3: *Molecular Biology and Basic Sciences* and Chapter 3: *Technology Development*.
- (E) (NHGRI)

The Knockout Mouse Project (KOMP): The NIH Knockout Mouse Project (KOMP) is an NIH-wide effort to create a publicly available resource of knockout mouse mutations that can be used to study human disease. Knockout mice are strains of mice in which specific genes have been completely disrupted, or knocked out. By studying these mice, researchers can evaluate the effect of this systematic disruption of different genes on physiology and development. Understanding the effects of gene disruption in mice will provide powerful tools to develop better models of inherited human disease. NIH has awarded 5-year cooperative agreements for the creation of knockout mice lines to Regeneron Pharmaceuticals Inc. to a collaborative team from Children's Hospital Oakland Research Institute, and to the Wellcome Trust Sanger Institute in England. NIH has also recently awarded $4.8 million to the University of California, Davis, and the Children's Hospital of the Oakland Research Institute to establish and support a repository for the KOMP. The repository will enable many more researchers to have access to the knockout mice, and will ensure product quality for the 8,500 types of knockout mice currently available.

- [Austin CP, et al *Nat Genet* 2004;36:921-4](#), PMID: 15340423
- For more information, see www.komp.org
- This example also appears in Chapter 3: *Technology Development*.
- (E/I) (NHGRI)

Genetics Home Reference: The Genetics Home Reference Web site provides basic information about genetic conditions and the genes and chromosomes related to those conditions. Created for the general public, the site was expanded to include summaries for more than 225 genetic conditions, more than 380 genes, all the human chromosomes, and information about disorders caused by mutations in mitochondrial DNA.

- For more information, see http://ghr.nlm.nih.gov
- This example also appears in Chapter 3: *Health Communication and Information Campaigns and Clearinghouses*.
- (I) (NLM)

The U.S. Surgeon General's Family History Initiative: Many people see most diseases as the result of interactions of multiple genes and environmental factors. Health care professionals have known for a long time that common diseases, such as heart disease, cancer, and diabetes, and rare diseases such as hemophilia, cystic fibrosis, and sickle cell anemia, can run in families. In a collaborative effort between the Office of the Surgeon General, NIH, the Centers for Disease Control and Prevention (CDC), the Agency for Healthcare Research and Quality (AHRQ), and the Health Resources and Services Administration (HRSA), the U.S Surgeon General's Family History tool was created. The U.S. Surgeon General's Family History tool (available in both English and Spanish) is free, and has proven to be an effective personalized tool for individualizing preventive care and disease prevention—in other words, maintaining good health. Recently updated, this tool allows individuals to record health conditions that have affected their relatives. It utilizes a three-generation pedigree to gather information on health conditions in one's family to help doctors take action to keep individuals and families healthy.

- [Guttmacher AE, et al.*N Engl J Med* 2004;351:2333-6](#), PMID: 15564550
- For more information, see http://www.hhs.gov/familyhistory
- For more information, see https://familyhistory.hhs.gov
- This example also appears in Chapter 2: *Chronic Diseases and Organ Systems*.
- (O) (**OD**, NHGRI)

Influenza Virus Resource: This database of more than 40,000 influenza virus sequences allows researchers around

the world to compare different virus strains, identify genetic factors that determine the virulence of virus strains, and look for new therapeutic, diagnostic, and vaccine targets. The resource was developed by NCBI using data obtained from NCBI's Influenza Virus Sequence Database and from NIAID's Influenza Genome Sequencing Project, which has contributed sequences of the complete genomes from over 2,500 influenza samples. In FY 2006 more than 11,000 influenza virus sequences were entered into the database, and new search and annotation tools were added to assist researchers in their analyses.

- Wolf YI, et al. *Biol Direct* 2006;1:34, PMID: 17067369
- Chang S, et al. *Nucleic Acids Res* 2007;35:D376-80, PMID: 17065465
- For more information, see http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html
- For more information, see http://www.niaid.nih.gov/dmid/genomes/mscs/influenza.htm
- This example also appears in Chapter 3: *Disease Registries, Databases, and Biomedical Information Systems*, Chapter 2: *Infectious Diseases and Biodefense*, and Chapter 3: *Molecular Biology and Basic Sciences*
- (I) (NLM)

Ethical, Legal, Social and Behavioral Issues

Genetic Factors in Health Disparities: A major concern in the era of genomic health care is to ensure that all racial, ethnic, and cultural groups benefit fully from genomic technology. One GPRA goal is to establish the role of genetic factors in three major diseases for which health disparities are noted. Building on the foundation of the Human Genome Project (HGP), NIH, as part of the International HapMap Consortium, has developed a way to scan large regions of chromosomes for variants (called SNPs, or single nucleotide polymorphisms) associated with increased risk of disease. Understanding the role of genetics in diseases characterized by health disparities will rely on such tools. As an example, the FUSION (Finland-United States Investigation of Non-Insulin-Dependent Diabetes Mellitus Genetics) study collected 820 million genotypes in 2006, which resulted in the identification of at least four new genetic variants associated with increased risk of diabetes and confirmed existence of another six. The findings boost to at least 10 the number of genetic variants confidently associated with increased susceptibility to type 2 diabetes—a disease that affects more than 200 million people worldwide, and a major cause of health disparities.

- Scott LJ, et al. *Science* 2007;316:1341-5, PMID: 17463248
- For more information, see http://nihperformance.nih.gov/
- (E/I) (NHGRI) (GPRA Goal)

Ethical, Legal and Social Implications (ELSI) Centers of Excellence for ELSI Research (CEERs): This center program has funded four full centers and three exploratory centers involving investigators in a wide range of disciplines to devise and employ interdisciplinary approaches to investigate ELSI issues such as:

- Intellectual property issues surrounding access to and use of genetic information
- Factors that influence the translation of genetic information to health care
- Conduct of genetic research that involves human subjects
- Use of genetic information and technologies in non-health care settings such as employment, insurance, education, criminal justice, or civil litigation
- Impact of genomics on concept of race, ethnicity, and individual/group identity
- Implications of uncovering genomic contributions to human traits and behaviors such as mental illness or aging for how we understand health and illness
- How different individuals, cultures, and religious traditions view the ethical boundaries for the uses of genomics

The use of CEERs resources and expertise to design and implement multifaceted and multidisciplinary investigations of particularly complex, persistent, or rapidly emerging ELSI issues is an important addition to ongoing genetic, genomic, and ELSI research efforts. Additionally, each CEER trains many young ELSI researchers each year.

- For more information, see http://www.genome.gov/10001618
- For more information, see Chapter 4: *NIH Centers of Excellence*.
- (E) (NHGRI)

Multiplex Initiative: With the completion of the sequence of the human genome, genetic susceptibility tests that give "personalized" information about risk for a variety of common health conditions are now being developed and marketed. This genetic information ultimately will improve primary care by enabling more personalized treatment decisions for common diseases such as diabetes and heart disease. This information also might motivate patients to change unhealthy behaviors. NIH investigators have teamed with the Group Health Cooperative in Seattle and the Henry Ford Health System in Detroit to launch a study to investigate the interest level of healthy, young adults in receiving genetic testing for eight common conditions. Called the Multiplex Initiative, the study will also look at how people who decide to have the tests interpret and use the results in making health care decisions. One thousand subjects who meet the study's eligibility requirements will be offered free multiplex genetic testing. The testing is designed to yield information about 15 different genes that play roles in common diseases such as type 2 diabetes and coronary heart disease. Trained research educators will make followup telephone calls to help subjects interpret and understand test results, and subjects will receive newsletters to update them on new developments about the tested genes. This research should provide insights into how best to utilize the powerful tools of genomic medicine to improve health.

- For more information, see http://www.genome.gov/25521052
- This example also appears in Chapter 2: *Chronic Diseases and Organ Systems* and Chapter 3: *Clinical and Translational Research*.
- (E/I) (NHGRI)

Genes, Behavior and the Social Environment: Moving Beyond the Nature/ Nurture Debate: This 2006 Institute of Medicine report was requested in order to examine the state of the science on gene-environment interactions as related to health, with a focus on the social environment. Report recommendations identified approaches and strategies to strengthen the integration of social, behavioral, and genomic research and training needs.

- For more information, see http://www.iom.edu/CMS/3740/24591/36574.aspx
- (E) (**OBSSR**, NHGRI, NIGMS)

NIH Revision Awards for Studying Interactions Among Social, Behavioral, and Genetic Factors in Health: These program announcements solicit applications for competitive supplements (revisions) to NIH grants to add a genetics/genomics component to a behavioral or social science project or the converse, i.e., to add a behavioral or social science component to a genetics/genomics project. This ultimate goal of this initiative is to elucidate how interactions among genetic/genomic, behavioral and social factors influence health and disease. The knowledge gained by such research will improve our understanding of the determinants of disease as well as inform efforts to reduce health risks and provide treatment.

- For more information, see http://grants.nih.gov/grants/guide/pa-files/PAR-08-065.html
- For more information, see http://grants.nih.gov/grants/guide/pa-files/PAR-08-066.html
- For more information, see http://grants.nih.gov/grants/guide/pa-files/PAR-08-067.html
- (E) (**OBSSR**, NCCAM, NCI, NEI, NHGRI, NIA, NIAAA, NIAMS, NIDA, NICHD, NIDCD, NIDCR, NIDDK, NIMH, NINR, NINDS, ODS)

Summer Training Institute in Genes, Environment and Behavior Research: This training institute scheduled for summer of 2009 will target behavioral and social scientists at various career levels. The activity is designed to instruct the subjects in the theoretical and practical foundations of genetics and genomics and to introduce them to research on gene-behavior-environment interactions. The institute will help train a cadre of behavioral and social scientists capable of working in interdisciplinary teams to improve our understanding of how interactions

among genes, behaviors, and environments contribute to health and disease.

- (E) (OBSSR)