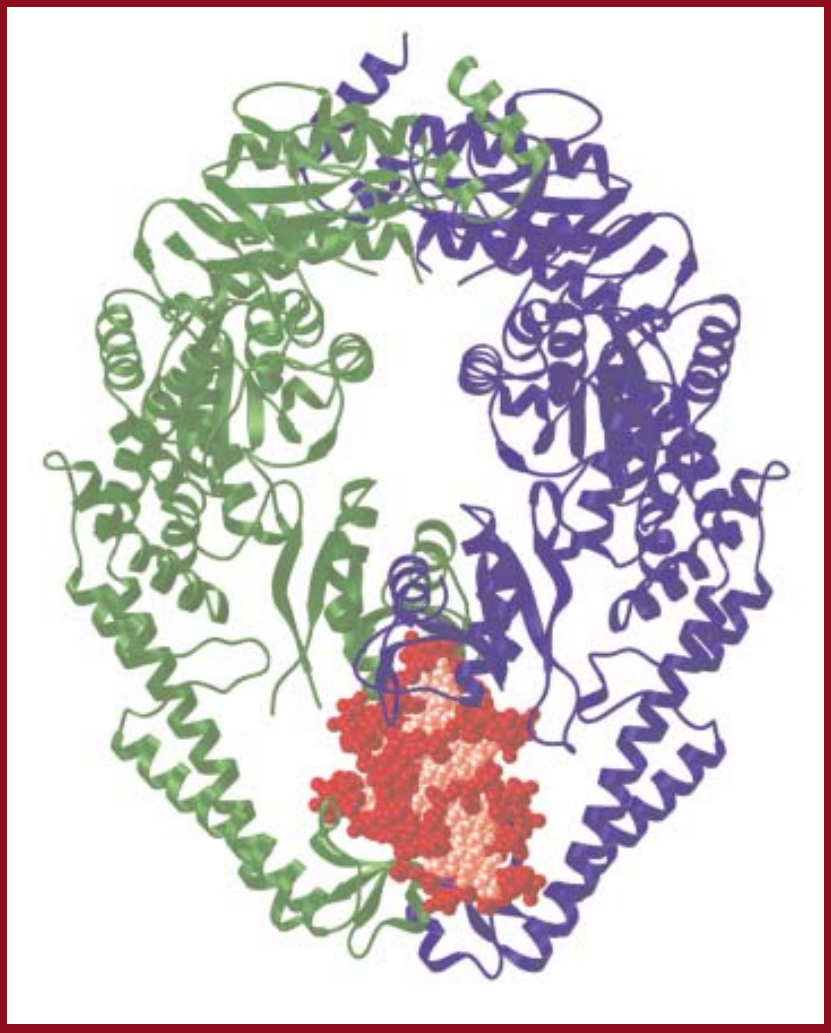


This three dimensional structure shows how the cellular protein MutS (blue and green ribbons) contacts a DNA double helix (red and pink) to repair mismatches in the two strands and prevent mutations. Maintaining the accuracy of a DNA sequence is crucial for all organisms, from bacteria to humans. Photo: Dr. Wei Yang. Reprinted with permission from Obmolova G *et al, Nature* 407: 703-10, 2000. © 2000 Nature Publishing Group *(http://www.nature.com)*

Cross-Cutting Science: Paving the Way to Discovery

Advances in medicine are largely dependent upon the accumulation of new knowledge about biologic processes, especially at the smallest levels of an organism–its genes, the proteins they encode, and the workings of cells. While the ultimate application of such basic research is not always obvious, major strides in fighting disease can be traced back to laboratory studies whose immediate relevance to health could not have been fully known or appreciated at the time they were conducted. Opportunities to make exciting new discoveries and advances are arising ever more rapidly with the development of new technologies, new approaches, and even new scientific disciplines. Described here are some recent studies of fundamental processes, ranging from the development of cells to the development of organisms, and new approaches and technologies that make such studies possible. The insights gained through this type of research can be expected to propel disease-oriented research, not only within the NIDDK mission, but also in many other fields. Investment in such cross-cutting scientific research today will have future applications that we cannot now describe with certainty, but which we know will surely be realized.

THE NIH ROADMAP

In October 2003, the NIH Director announced the first initiatives to be launched through a new, far-reaching strategic effort to identify opportunities and gaps in biomedical research that no single institute at NIH could address alone, but that the NIH as a whole should address in order to accelerate biomedical research. That effort is called the "NIH Roadmap for Medical Research." Designing the NIH Roadmap has been a collaborative process, which involved NIH senior scientific staff, with input from over 300 nationally recognized leaders in academia, industry, government, and the public.

The NIH Roadmap is a new vision for research that provides a framework for making strategic NIH investments to optimize the agency's entire research portfolio. The NIH Roadmap is divided into three major themes: (1) New Pathways to Discovery; (2) Research Teams of the Future; and (3) Re-engineering the Clinical Research Enterprise. Through the Roadmap effort, scientific initiatives were developed to propel research under these three themes. To be part of the NIH Roadmap, initiatives had to be deemed of high potential impact, had to enhance the disease and mission-specific activities of all of the NIH Institutes and Centers, and had to respond to the needs and concerns of the public. These initiatives were refined and are being carried out by nine focused "Implementation Groups" chaired by Directors of the NIH Institutes and Centers. To facilitate implementation, each initiative is being "led" by an Institute whose interests and expertise are closely aligned with the stated goals.

The NIDDK Director co-chaired the "Building Blocks, Biological Pathways, and Networks" Implementation Group, which is under the "New Pathways to Discovery" theme. This group focused on "systems biology," an integrative study of how all biological components work together in an organism to promote normal development and to sustain health. Further research is needed in this

area to understand how biological pathways are integrated in humans and in other complex organisms, to determine how disturbances in these pathways may lead to disease, and to discover how to restore disturbed pathways to their normal functions. Many of the diseases within the NIDDK mission are chronic conditions that result from disturbances in biological systems either at the outset of life or over the course of life. Thus, initiatives in this area are particularly relevant to NIDDK-supported research.

Roadmap Initiatives–New Technology Development: In order to increase knowledge in systems biology, new experimental technologies are needed to study cellular components more accurately, quickly, and on a very small scale. The impact of new technologies is underscored by the recent completion of the Human Genome Project, in which an emerging technology–high-throughput DNA sequencing–enabled scientists to achieve this momentous accomplishment. The data generated from this project can be used to understand the underlying genetics of healthy and disease states. It is an ongoing challenge for scientists to take large amounts of data and apply them to achieve greater knowledge about biology. Under the "New Pathways to Discovery" theme, Roadmap initiatives are encouraging development of new technology to increase understanding about cellular pathways. Such general technology development is beneficial to the NIDDK because the technologies can ultimately be used to study specific diseases within the Institute's mission.

The NIDDK will be the lead Institute in implementing an initiative on "Metabolomics Technology Development." The "metabolome" is the complete set of metabolites in an organism; examples of metabolites include amino acids, peptides, and lipids. "Metabolomics" is the study of these low-molecular weight molecules. The purpose of this initiative is to promote the development of highly innovative and sensitive tools for studying metabolomics. The development of novel technologies can directly benefit the study of diseases within the NIDDK mission. For example, such technologies can be used to develop surrogate markers to predict risk, aid in diagnosis, and assess progression of the complications of diabetes. Metabolomics technologies can also be applied to advance understanding of the metabolic changes that occur with obesity and its co-morbidities. The ability to study metabolites at the single-cell level would also aid in characterizing tissues and organs that contain a variety of specialized cell types, such as the kidney.

Another initiative will develop a network of research Centers to create new tools for "proteomics," a high-throughput approach to studying the dynamics of protein interactions. The Centers will develop instruments, methods, and reagents for quantitative measurements of proteins at sub-cellular resolution and with very short timescales. The NIDDK plans to complement this Roadmap initiative with proteomics studies specifically geared toward diseases within the NIDDK mission. As the technology advances, NIDDK researchers can apply these new technologies to examine mechanisms regulating the pathology of many different diseases and conditions within the NIDDK mission.

Roadmap Initiatives–Molecular Libraries: Also under the "New Pathways to Discovery" theme, this initiative focuses on the use of "small molecules" to study biological processes. Small molecules have proved to be important tools to manipulate biological processes, many times leading toward important therapeutic advances. For example, small molecules can act to regulate the natural activities of important proteins to achieve a desired biological effect. With the emergence of large amounts of genomic information, it has become possible to use libraries of small molecules to screen for and identify compounds useful for study of the regulation of important

proteins. In order to find an appropriate molecule to interact with the protein of interest, scientists may screen a large number of small molecules from a "molecular library" of thousands of molecules. Because a very small number of molecules will interact with a target protein in the precise way needed to observe a biological effect, a molecular library screening approach increases the chances that an appropriate molecule will be discovered.

The importance and biological significance of using a molecular library approach was demonstrated in recent studies by NIDDK-supported researchers in the areas of diabetes and cystic fibrosis (see next chapter on "Diabetes, Endocrinology, and Metabolic Diseases"). NIDDK-supported researchers also used this approach to study a protein called "FXR." FXR is a key protein that functions as a receptor to regulate bile acid and cholesterol metabolism. Researchers hypothesized that it might be possible to test the biological function of FXR through the use of small molecules. Therefore, they used a molecular library of over 10,000 small molecules in a high throughput screening assay to identify any that would turn on, or activate, FXR. The result was identification of small molecules that had specific effects on FXR. This enabled more refined screening to finally arrive at a class of small molecules with better effects on FXR than natural compounds. These optimized molecules potently activated FXR, but, importantly, did not activate most other proteins that were similar to FXR, and facilitated studies to understand the underlying mechanism of action of FXR. Because disorders of bile and cholesterol metabolism are often interconnected, these findings now open the door for future research to develop additional small molecules that can have a bearing on diseases that these processes may regulate.

The Molecular Libraries Roadmap initiative will give researchers tools to study a wide range of diseases within the NIDDK mission. In addition, a new NIDDK initiative will develop assays for small molecule screening for potential mission-specific therapeutics. The development and application of molecular libraries can provide insights into important biological and disease-related processes, and serve as the basis for the generation of new drugs.

The Future: The Roadmap initiatives will address areas of major scientific opportunity and need for all of NIH, including the NIDDK. Exploiting new pathways to discovery, fostering interdisciplinary research, and re-engineering the clinical research enterprise will speed discoveries that can ultimately lead to improved prevention and treatment of obesity, diabetes, and digestive, kidney, urologic, and blood diseases. The Roadmap initiatives tackle areas that the NIDDK is not able to do on its own, even though advances in these areas would greatly help the study of the Institute's categorical disease. The NIDDK also plans to issue Institute-specific initiatives that will complement and enhance Roadmap initiatives. This approach will take full advantage of the Roadmap process to ensure that the NIDDK reaps maximum benefits from this exciting and innovative process for years to come.

Nicolaou KC, Evans RM, Roecker AJ, Hughes R, Downes M and Pfefferkorn JA: Discovery and optimization of non-steroidal FXR agonists from natural product-like libraries. *Org Biomol Chem* 1: 908-20, 2003.

VISUALIZING LIFE'S PROOFREADERS: STRUCTURAL STUDIES OF DNA REPAIR PROTEINS

Cell division is a basic biological process in which one cell becomes two. The new cell must contain all of the material in the original parent cell, in particular the genetic blueprint known as DNA, in order for the new cell to be functional and continue to divide. Making a copy of the DNA requires a process called "DNA replication," which uses the parent cell's DNA as a template to

make a new copy. DNA consists of "coding letters," also called "bases," which are used as a blueprint to make proteins that carry out biological processes. The DNA code for making a protein is analogous to using letters to make sentences. Specifically, there are four DNA bases (letters); three letters in combination code for an "amino acid" (word), and many amino acids string together to form a "polypeptide chain" (sentence). A functional protein may contain single or multiple polypeptide chains.

The process of DNA replication is very efficient; however, there are approximately 3,200 million bases that must be copied, so it is inevitable that some errors will occur. Approximately 30-3,000 errors are made each time a cell divides. This is analogous to typing on a computer–some words will contain typing errors. Luckily, a convenient tool–the spellchecker–finds the mistakes so they can be corrected. Likewise, the cell has different types of tools to "proofread" the DNA for mistakes so they can be repaired before the cell divides. Interestingly, these tools are found in all species, from bacteria to humans. Their presence demonstrates that maintaining the accuracy of this fundamental process is extremely important and has been conserved throughout evolution.

Why is it so important to correct errors in copying the DNA? When there is a mistake in the DNA, there can subsequently be a mistake in the protein whose production it controls. Many times, even one mistake prevents the protein from functioning properly and can lead to serious health problems. For example, certain types of cancer are caused because of mistakes in proteins that are important in regulating cell division. In addition, an increase in overall instability of DNA has been implicated in the aging process. Thus, maintaining the integrity of the DNA is crucial to the cell.

DNA Mismatch Repair: A type of mechanism, called "mismatch repair" (MMR), is one component of the cell's repair system to correct errors in DNA that may arise during replication or damage (such as damage to DNA bases by chemical agents). The importance of this process is underscored by the fact that people with defective MMR proteins are at increased risk for developing a type of colorectal cancer, called "hereditary nonpolyposis colorectal cancer" (HNPCC). Errors in MMR proteins are also found in some sporadic cancers, which are cancers that do not appear to be inherited.

In bacteria, an important protein that is involved in MMR is called "MutS." MutS can be called a "sensor" that sends signals to many other cellular proteins when there is an error in the DNA. MutS can recognize DNA bases that are improperly matched, as well as small loops of unpaired bases; MutS can also recognize a limited repertoire of chemically altered bases. When MutS finds an error, it can communicate to other MMR proteins to begin the DNA repair process. In some instances, MutS can initiate a cell death pathway that kills the cell when the errors are irreparable. The role of MMR proteins, such as MutS, in regulating cell death pathways comes into play in the treatment of cancer patients. Cisplatin, a chemotherapy drug, is not capable of "killing" the cancer cell on its own. Cisplatin damages DNA and then requires MutS to "sense" the damage and activate the cell death pathway. However, many cancer patients become resistant to the chemotherapy because the cancer cells have found a way to inactivate the MMR pathway. The resistance to chemotherapy is a serious barrier that prevents successful treatment of cancer patients.

Because the DNA replication process itself is so efficient, there are generally not many mistakes. Thus, how does MutS find the few errors embedded among the correctly formed DNA? How does MutS recognize different types of errors? These

questions are important to begin to unravel the underlying molecular mechanisms of this fundamental biological process. Recent studies by NIDDK scientists have begun to provide many answers.

Determining the Three-Dimensional Structure of Proteins: Proteins are three-dimensional structures that have characteristics such as loops, curves, and grooves. It is impossible to determine the three-dimensional structure of a protein by only knowing the sequence of individual building blocks (amino acids) that make up the protein. Therefore, researchers use special tools and techniques to determine this structure. One experimental method used by scientists is called "X-ray crystallography," in which proteins in liquid solution are subjected to conditions to make them form crystals. Once high-quality crystals are formed, they diffract X-rays and the patterns of diffraction are analyzed to determine the three-dimensional structure of the protein.

Solving a structure of a protein gives scientists considerable insight into how it may function in the cell. This approach permits scientists to actually "see" what the protein looks like and how it may be interacting with other molecules. This visualization can bridge the gap between genetic studies and understanding disease states. For example, genetic studies can identify mutations in proteins that cause disease; however, it is often not known how the mutations may alter the protein's function. With a protein's three-dimensional structure, scientists may be able to see how the specific mutations that cause disease may be detrimental to the protein's function–such as changing the structure of a protein or preventing the protein from interacting with other proteins or DNA. This knowledge can greatly improve the ability of scientists to hypothesize about ways to "fix" the protein–instead of making blind guesses about how the mutation may be harmful.

Three-Dimensional Structural Studies of MutS: NIDDK scientists have recently performed important structural and biological studies on a bacterial MutS protein that have given insight into the role of MutS in MMR. MutS was very difficult to crystallize because a portion of the protein was mobile. To overcome this barrier, the scientists removed the mobile portion, which did not affect the protein's function, but did enable the researchers to determine its structure.

The NIDDK researchers determined that it is necessary for two MutS polypeptide chains to come together–in an asymmetric manner–to bind the DNA. Interestingly, they determined that MutS binding caused the DNA to severely bend. This observation helps to explain why MutS is able to recognize many types of DNA damage. DNA is similar to a wooden ladder that has many rungs. When the rungs are constructed properly, a person is able to step on them and be supported. However, if there is a crack in the wood of a single rung, that rung is much weaker and will easily bend or break when a person steps on it. The structural studies suggest that MutS recognizes and then can help to fix errors in DNA by finding the "weak rung," or the area of DNA that can bend.

Once MutS finds the DNA error, how does it communicate with other proteins to activate a correction through mismatch repair (MMR)? When the researchers compared the three-dimensional structures of MutS alone to MutS bound to DNA, they discovered that MutS itself undergoes a change in structure when it binds DNA. It is in this DNA-bound conformation that MutS interacts with other MMR proteins to begin repair. Thus, researchers hypothesize that MutS must be bound to a DNA mismatch in order to initiate MMR.

An important portion of MutS is able to convert a high energy metabolite, called "ATP," into a lower energy form called "ADP." Previous studies had shown that this conversion was necessary for

MMR. To further examine the importance of this portion of MutS, the NIDDK researchers analyzed the three-dimensional structure of MutS bound simultaneously to DNA and ADP. The structure, as well as insights from additional experiments, led them to conclude that the overall role of ATP to ADP conversion was to increase the sensitivity and specificity of MutS in finding the error in the DNA. ATP was necessary for the MutS to recognize the error in the DNA; it was also necessary to "authorize" the repair process to start, thus preventing repair from being performed on a "correct" DNA molecule. For example, if MutS bound to a piece of DNA that had no errors, the ATP would cause the MutS to dissociate from the DNA so that it did not repair something that was already correct. If MutS were bound to DNA containing an error, the ATP was the signal that gave the "green light" for repair to occur. Additional genetic and biochemical studies showed that MutS must bind both ATP and the incorrect DNA in order for repair to begin.

Researchers would also like to study the structure of MutS bound to both DNA and ATP. However, this MutS-DNA-ATP complex is very unstable, because the ATP is rapidly converted to ADP. Thus, to date, it has not been possible to perform structural studies of MutS with ATP. NIDDK researchers therefore looked for other molecules that were similar to ATP to see if they could use an alternate molecule to determine the three-dimensional structure. They found a molecule, called "ADP-berillium fluoride," or "ABF," which they have used in subsequent X-ray crystallography studies. ABF is similar to ATP and binds to MutS, but is not converted to ADP. In addition to making it useful for three-dimensional studies of MutS, these properties make ABF an invaluable tool to further study the DNA repair process.

The structural biology studies have also enabled the researchers to make predictions about the regions of MutS that may be critical for its biological function. This capability has led to biochemical studies of the protein to understand the importance of different regions in regulating repair. For example, the researchers have determined that a region of MutS, called the "HuH" motif, is necessary for function. These types of studies demonstrate that a combination of structural biology, biochemical, and genetic approaches is an efficient and productive way to begin to understand a very complex process.

Implications: Repairing damaged or incorrectly replicated DNA is a critical process for all living things. When this process is not faithfully maintained, subsequent mistakes in proteins can occur, which may lead to serious health problems, such as cancer. The previously described studies using bacterial proteins can be expanded to studying mismatch repair in other organisms, including humans. Because the processes are similar, much of the knowledge gained from the bacterial studies can be directly applied to higher organisms.

Understanding the molecular basis and regulation of DNA repair processes can help researchers determine what may go wrong biologically that can lead to a disease state. For example, the structure of MutS can be used to map mutations that cause diseases. It is known that errors in the MMR proteins can cause an inherited form of colorectal cancer. NIDDK researchers have used MutS structural studies to determine that many of the mutations involved in this type of cancer are located in regions of MutS that are critical for its function. This may help researchers understand why the mutated protein does not function properly in disease. Information about a fundamental biological process such as DNA repair can also be used to increase knowledge about cancer in general. This work can therefore be applied to other cancers within the NIDDK and NIH research mission, such as prostate and pancreatic cancers.

Although determination of a protein's three-dimensional structure may be a difficult process, it gives scientists a way to actually "see" what a protein looks like. The structural studies per-

formed by NIDDK researchers complement biochemical and genetic studies, and have directly led to a new model for the mechanism by which MutS regulates mismatch repair of defective DNA. This knowledge builds a framework for future studies on this process as it relates to cancer, aging, and resistance to chemotherapy.

Alani E, Lee JY, Schofield MJ, Kijas AW, Hsieh P, Yang W: Crystal structure and biochemical analysis of the MutS•ADP•beryllium fluoride complex suggests a conserved mechanism for ATP interactions in mismatch repair. *J Biol Chem.* 278: 16088-94, 2003.

Junop MS, Obmolova G, Rausch K, Hsieh P, Yang W: Composite active site of an ABC ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair. *Molecular Cell* 7:1-12, 2001.

Biswas I, Obmolova G, Takahashi M, Herr A, Newman MA, Yang W, Hsieh P: Disruption of the helix-u-turn-helix motif of MutS protein: Loss of subunit dimerization, mismatch binding and ATP hydrolysis. *J Mol Biol.* 305: 805-16, 2001.

Obmolova G, Ban C, Hsieh P, Yang W: Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature* 407: 703-10, 2000.

POTENTIAL AND PLASTICITY OF STEM CELLS–AN EVOLVING PICTURE

Scientists are striving to understand the processes that occur during normal development, when a vast number of different cell types are generated from a single fertilized egg. If they can understand normal development, scientists will have a better chance of determining how to recapitulate development in an adult in order to replace cells damaged by disease. Even within adults, special cells known as stem cells retain the ability to divide, and the divisions can give rise either to more stem cells or to cells that will differentiate into specific cell types.

Currently, scientists are determining the usefulness of different types of stem cells for treating human disease. Until now, replacement of cells has only been possible *via* organ or cell transplantation. However, doctors are unable to treat every needy patient with transplantation, because there are limited supplies of donor cells and organs. Stem cells are heralded as a possible means for overcoming this treatment barrier. Furthermore, by studying how stem cells become specialized cells within organs and tissues, researchers are reaching a greater understanding of the processes underlying normal and abnormal tissue regeneration and repair. This, in turn, is enabling them to develop and test treatments that could augment the body's reparative processes and foil disease.

The various types of stem cells are believed to differ mainly in the limits of their "potential"–their ability to differentiate into other cell types. Embryonic stem (ES) cells arise early in development. Because they must give rise to all the different cell types and tissues of the body, embryonic stem cells are thought to have almost unlimited potential.* Adult stem cells, on the other hand, reside within a mature tissue or organ and have been traditionally viewed as able to differentiate into a more limited number of cell types, although numerous investigators have reported that some of these cells actually may have a much greater differentiation potential than previously thought–a phenomenon also known as "plasticity." The NIDDK is supporting research on both types of stem cells, within established policies for NIH funding.

Bone Marrow Adult Stem Cells Take Different Paths to Become Liver Cells and Islets: There may be several means whereby stem cells derived from one mature, adult tissue or organ can apparently turn into differentiated cell types characteristic of another organ upon transplantation–including the presence of highly potent adult stem cells in donor tissues, the presence of several types of adult stem cells in donor tissues, and the fusion of donor cells with recipient cells. Cell fusion occurs

* Consistent with the policy announced by President George W. Bush on August 9, 2001, NIH funding of research involving human stem cells is in accordance with specific criteria established by the Administration.

during normal mammalian development in the formation of bone and muscle, is a feature of certain types of cancer, and can occur in cultured stem cells. Recently, in a major discovery, researchers showed that adult bone marrow-derived cells were able to repair damaged liver tissue in mice by actually fusing with the host liver cells. Previously, it was shown that mice with a fatal metabolic liver disease, tyrosinemia type I, which is caused by loss of a vital enzyme, regain normal liver function through transplantation of hematopoietic stem cells purified from bone marrow. It turns out that these stem cells do not alone directly generate new liver cells, but instead spontaneously fuse with the diseased liver cells to form a healthy-functioning hybrid cell containing both stem cell and liver DNA, with the liver cell molecules dominating the hematopoietic factors. It also appears that the contribution of the hybrid stem-liver cells depends on both the normal regenerative capacity of the liver, and the space in the organ that is created by the degeneration of the diseased host liver. Generally, when cell fusion occurs, the cell that results contains a greater than normal number of chromosomes. Remarkably, however, some of the new, hybrid cells contained the normal two copies of each chromosome. It is possible that in these hybrid cells, "reduction division" occurs, whereby chromosome pairs are lost. These new findings are an example of research on the mechanisms underlying putative adult stem cell plasticity, and also have implications for treating certain genetic metabolic disorders, with the possibility of gene transfer through cell fusion.

Researchers are also testing the potential of adult bone marrow-derived stem cells to replace the insulin-producing beta cells of the pancreas, which are lost to autoimmune destruction in type 1 diabetes. A new report from studies in mice has revealed that these stem cells may be a potential source of beta cells. Unlike the stem-liver cell fusion hybrid described previously, however, researchers systematically ruled out cell fusion as the mechanism for producing beta cells. In this research, mouse bone marrow-derived cells that express the "enhanced green fluorescent protein" (EGFP) if the insulin gene is actively making insulin, were transplanted into mice whose bone marrow was destroyed. Four-to-six weeks later, EGFP-positive (insulin-producing) cells appeared in the host pancreatic islets. Detailed studies revealed these cells to have several beta cell markers–they expressed the glucose transporter 2 gene and transcription factors that are indicators of beta cell differentiation, and were responsive to factors that induce insulin secretion. Thus, bone marrow-derived cells represent a possible new source of cells for beta cell replacement therapy for type 1 diabetes, and some forms of type 2 diabetes. This therapeutic approach would not require co-administration of immunosuppressant drugs to keep the recipient's immune system from rejecting the cells because the bone marrow could be taken directly from the diabetic patient.

"Niche" Control of Adult Stem Cell Fate:
Specialized microenvironments (niches) must exert critical influences on stem cell fate, but the nature of these influences remains poorly understood. Researchers studying this phenomenon in mouse bone marrow cells have discovered that bone-forming cells, called osteoblasts, are a major regulatory component of the bone marrow microenvironment. A key ingredient is parathyroid hormone (PTH), which is an important regulator of bone growth and metabolism. Bone marrow from a mouse strain altered so that the receptor for PTH is continuously activated in osteoblasts produced twice the number of hematopoietic stem cells as normal. A molecular signaling system between the osteoblasts and stem cells was found to be responsible for governing the formation of stem cells. PTH both boosted the number of osteoblasts and supercharged the signaling system. Normal mice injected with PTH also showed increased stem cell production, as well as marked improvement in survival following bone marrow transplantation. These findings may

be important for lessening the health risk for bone marrow transplant recipients, who have a limited supply of stem cells. The identification of PTH, osteoblasts, and signaling pathways as important factors in the stem cell microenvironment provides pharmacological targets with therapeutic potential for stem cell-based therapies.

These studies have all advanced progress towards developing alternative sources of cells for transplantation, as well as other possible therapeutic approaches, to treat human diseases. The NIDDK is supporting several efforts designed to capitalize on and extend previous stem cell and developmental biology discoveries and to stimulate new discoveries.

Beta Cell Biology Consortium (BCBC): The BCBC initiative was established to facilitate interdisciplinary approaches that will advance the understanding of pancreatic islet development and function. The long-term scientific goal is to develop a cell-based therapy to restore normal insulin production and action to diabetic patients. To accomplish this goal, the BCBC supports a host of stem cell-related research efforts, including development of key reagents to permit the prospective isolation of stem/progenitor cells in both neonatal and adult animals, development of assays to test the efficacy of pancreatic stem/progenitor cells as a cure for type 1 diabetes, and development of protocols to efficiently differentiate embryonic and adult stem cells into pancreatic islet tissue. Important resources of the BCBC include (1) EPConDB, a searchable database that contains information about genes expressed in the mouse and human pancreas during development (*http://www.cbil.upenn.edu/EPConDB/*), (2) a Microarray Core that is producing and distributing mouse and human cDNA microarrays, (3) an Antibody Core that is developing antibodies to human cell surface antigens and important markers of progenitor cells, (4) a Mouse ES Core that is producing important transgenic animal models that will be essential in identifying and characterizing pancreatic progenitor cells, (5) a Human Stem Cell Subcommittee that is establishing protocols for expanding and isolating human pancreatic progenitor cells, (6) a BCBC Coordinating Center (Vanderbilt University) which oversees the BCBC website (*www.betacell.org*), and coordinates all activities of the BCBC including scientific cores, investigator retreats, and the "pilot and feasibility studies" program, and (7) a Stem Cell Training Core which will train members of the BCBC to work with both adult and embryonic human stem cells.

Progenitor Cell Genome Anatomy Projects (GAPs): The successful treatment of many chronic and debilitating diseases afflicting Americans today will depend on the ability to replace organs or to stimulate regeneration and recovery of damaged organs. To build upon the achievements of the Human Genome Project, the NIDDK and other NIH Institutes have established a range of Genome Anatomy Projects (GAPs) to map the complex network of cellular interactions in normal and diseased tissues. One NIDDK initiative is the Progenitor Cell Genome Anatomy Project, which is studying how progenitor cells develop into different types of cells that form the organs and tissues of the body. It is important to understand these developmental processes, and how progenitor cells maintain and regenerate tissues and organs in health and disease. Research approaches include developing biomarkers to detect and classify stem cells and progenitor cells; profiling the cells to catalogue genes that are active; and creating tools for characterizing the functions of these genes. This research will capitalize on the sequence data from the Human Genome Project. It is also a goal to distribute to the broad research community well-characterized cells, DNA, and specific tools for progenitor cell analysis developed by the GAPs. Furthermore, the development of bioinformatics systems, including databases, will ensure that data produced are available to researchers worldwide

soon after being generated in the laboratories. Multiple Progenitor Cell GAPs are supporting research to identify and describe stem cells located within specific tissues of the gastrointestinal lining, liver, pancreas, kidney, urinary tract, prostate, and bladder. Hematopoietic Cell Lineage GAPs are supporting work to describe gene expression in bone marrow-derived stem cells. Another NIDDK effort will sponsor studies to describe normal development and stem cells of the gastrointestinal tract, liver, and exocrine pancreas. The NIDDK hopes that knowledge gained from these studies will enable medical doctors to use stem cells and developmentally-regulated genes to repair and replace damaged and diseased tissue.

Calvi LM, Adams GB, Weibrecht KW, Weber JM, Olson DP, Knight MC, Martin RP, Schipani E, Divieti P, Bringhurst FR, Milner LA, Kronenberg HM and Scadden DT: Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature* 425: 841-6, 2003.

Ianus A, Holz GG, Theise ND, and Hussain MA: In vivo derivation of glucose-competent pancreatic endocrine cells from bone marrow without evidence of cell fusion. *J Clin Invest* 111: 843-50, 2003.

Vassilopoulos G, Wang PR, and Russell DW: Transplanted bone marrow regenerates liver by cell fusion. *Nature* 422: 901-4, 2003.

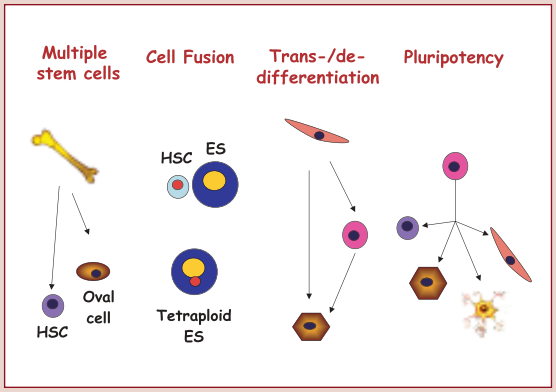
Wang X, Willenbring H, Akkari Y, Torimaru Y, Foster M, Al-Dhalimy M, Lagasse E, Finegold M, Olson S, and Grompe M: Cell fusion is the principal source of bone-marrow-derived hepatocytes. *Nature* 422: 897-901, 2003.

Stem Cells: Promise and Reality

Dr. Catherine Verfaillie

The NIDDK National Advisory Council meets three times annually to provide advice to the Institute regarding its research portfolio and broad issues of science policy. These meetings are also an opportunity for the Council members to learn about recent scientific advances in different fields through presentations from NIDDK-supported extramural scientists. In 2003, the Council and NIDDK staff were privileged to hear from two leading scientists, Dr. Catherine Verfaillie and Dr. David Altshuler, who are conducting research studies in stem cell biology and human genetic variation in disease, respectively. The "Scientific Presentations" in this chapter are meant to capture the essence of their talks.

Catherine Verfaillie, M.D., is Professor of Medicine and Director of the Stem Cell Institute at the University of Minnesota. She received her M.D. from the Catholic University of Leuven in Belgium in 1982, and came to the University of Minnesota 1987 after an internal medicine residency and fellowship in hematology in Belgium. Dr. Verfaillie's recent research has focused on the plasticity of stem cells, the mechanisms controlling their differentiation, and evaluation of their therapeutic potential.



The recently observed capacity of adult stem cells from one tissue or organ to become cells from an unrelated tissue or organ may depend upon one or more different, biologically relevant pathways. Scientists are pursuing research to sort out the basis for this characteristic of stem cells, which is known as "plasticity."

Dr. Verfaillie explained that many researchers share the hope of one day treating diseases through cell-based approaches. To this end, scientists are exploring the capability of undifferentiated progenitor cells, called stem cells, to be induced experimentally to become specialized cells of the body. While stem cells are in the spotlight of research today, as recently as six years ago, scientists who worked on stem cells toiled in relative obscurity. This changed radically in 1998, when human embryonic stem cells were first identified. Furthermore, evidence accumulated since that time, in over 300 scientific publications, suggests that adult stem cells may have greater differentiation potential than previously thought, and that hematopoietic cells of the bone marrow might have the ability to differentiate into other cell types than blood cells. The field of stem cell biology moved to the forefront of scientific discussions, as well as discussions in the lay press. Although promising, this field requires extensive fundamental studies to understand how stem cells function and react to their environment, an exciting pursuit that will enlighten scientists about the steps stem cells take as they mature and commit themselves to different developmental fates whereby they become the myriad tissues and organs in the body.

SCIENTIFIC PRESENTATION

Stem Cell Basics

Stem cells have three fundamental properties: (1) they undergo self-renewing cell divisions; (2) they can differentiate into multiple different functional cell types; and (3) when administered to a human or an animal, they can functionally reconstitute an organ or tissue that has been destroyed.

Scientists are studying several types of stem cells. Adult stem cells are rare populations of undifferentiated cells found in the tissues of adult animals and humans. Research has shown that adult bone marrow cells, which feed the body's circulatory and immune systems, may be a good source of these cells. Another type of stem cells, called embryonic stem (ES) cells, may also be derived from either animal or human tissue.*

Because research on all of these types of cells is still in its early stages, scientists cannot now predict which of them may prove most effective and appropriate for therapeutic purposes. ES cells may turn out to be a better cell source for certain differentiated cell types, while adult stem cells might be a better source for other cell types. Research on both cell types, however, will have a synergistic effect in advancing knowledge in the field, and scientists are working diligently to characterize these cells and understand their workings.

There has been recent excitement about ES cells because they can differentiate into all of the body's cell types and maintain differentiation capacity when grown in the laboratory for long periods of time. Thus, ES cells might constitute an unlimited source of cells for a wide array of therapies. However, several difficulties are associated with potential clinical use of human ES cells. These cells would be obtained from a donor and might thus be rejected as foreign by the patient's immune system; the cells could form tumors (teratomas); and the generation of these cells from blastocysts (early embryos) is a controversial procedure.

Adult stem cells are already being used for therapies such as bone marrow transplantation. There are several positive aspects to using adult stem cells. One is that they can be obtained directly from the patient, eliminating, for some potential therapeutic applications, the problems associated with unrelated-donor-cell rejection. Another is that, because adults, not embryos, are the source of the cells, the research is not controversial. There also is growing evidence that adult stem cells appear to be more potent, or have greater "plasticity," than was previously believed. Numerous recent reports indicate that adult stem cells from one tissue can differentiate into specific cell types normally found in other tissues. For example, when bone marrow cells are injected into the hearts of animals damaged by a heart attack, these cells appear to differentiate into cells with the characteristics of heart muscle cells. Finally, findings suggest that bone marrow stem cells may be found in the pancreas, liver, and nervous system, kidneys, and other organs.

Potency of Adult Stem Cells

Dr. Verfaillie's research interest in certain components of the bone marrow led her to initiate studies on cells called mesenchymal stem cells. At the beginning, the goal was to purify such cells from patients with one of the inborn errors of metabolism–muccopolysaccharoidosis type 1, a lysosomal storage disease–in order to genetically-modify these cells for use in therapy for this disorder. However, in the course of these studies, her group identified so-called multipotent adult progenitor cells–opening up a new research avenue. Initially, the question was asked: Could adult mesenchymal cells within the bone marrow differentiate into other cell types? It was found that mesenchymal cells indeed had a much greater differentiation potential than one might expect. In addition to forming bone, cartilage, fat, and muscle cells, mesenchymal cells grown in the laboratory under the right conditions could be directed to form both liver cells and nerve cells. This discovery is significant because liver

and nerve cells normally arise from two different tissue layers during embryonic development, the embryonic endoderm and ectoderm, respectively, while mesenchymal cells arise from a third distinct layer, the mesoderm. Moreover, when adult mouse mesenchymal stem cells were placed into mouse blastocyst embryos, they were found to contribute to the subsequent development of every organ in the body. Because these adult mesenchymal stem cells were shown to be highly potent, possibly even as potent as ES cells (which are "pluripotent"), they were termed multipotent adult progenitor cells (MAPCs).

Many interesting questions arise from these studies, including, how can adult stem cells be endowed with properties of early embryonic cells, i.e., with the plasticity of becoming cells that normally would derive from endoderm, mesoderm, or ectoderm? Possible answers to questions regarding plasticity are that multiple tissue-specific stem cells are present in different organs; that this phenomenon is actually the result of fusion of the donor cell with resident cells in an organ; that cells undergo de-differentiation and re-differentiation; and, finally, that true multi- or pluripotent stem cells do persist in postnatal life. There is at least some scientific evidence supporting each of these theories. Thus, scientists need to continue carefully characterizing the mechanisms underlying apparent adult stem cell contributions to multiple or unexpected cell types in their studies, in order to truly understand the nature of plasticity.

Uses of Stem Cells–Basic Knowledge, Therapy, and Bioengineering

Dr Verfaillie stressed that one of the most important things that can be learned from stem cells is to understand the basic processes of self-renewal and differentiation of these cells in different tissues. This knowledge ultimately could lead to drug discovery, much in the same way that understanding stem cell biology and the hematopoietic system formed the basis for development of blood cell growth factors for therapeutic purposes. When stem cells in other tissues are better understood, pharmacologic agents–instead of cell replacement–may be used to therapeutically repair damage in the body.

For example, MAPCs that were experimentally induced to form cartilage and bone cells were studied using gene array technology. As expected, the lineage of these cells types was found to be very similar. However, there were a number of factors controlling gene expression (transcription factors) that were expressed differently by these two cell populations. These differences might help point to the gene switches that direct cells to become bone or cartilage. Thus, stem cells and their differentiated progeny may be used to define genetic programs that need to be activated or inactivated for cell differentiation to occur.

Clinically, ongoing studies are further characterizing MAPCs and developing them for use in therapy for a wide array of conditions–including genetic diseases, such as muscular dystrophy and lysosomal storage diseases, and degenerative diseases and disorders in which tissue repair is required, such as diabetes, Parkinson's disease, arthritis, heart muscle damage from a heart attack, and liver failure. Progress is already being made with regard to central nervous system abnormalities, diabetes, and liver disease. For instance, one of the best examples of what embryonic stem cells can do therapeutically has come out of research on brain disease at the NIH. Researchers created a rodent model with a Parkinson's disease-like deficit by destroying dopamine-producing (dopaminergic) neurons in rat brains. The researchers then injected mouse embryonic stem cells that had been induced to differentiate into dopaminergic-like neurons *in vitro* into the brains of these rats. They observed functional improvement of the rats, and showed by structural and electrophysiological analysis that the implanted ES-derived cells had repaired the brain defect.

In another series of experiments studying the potential of stem cells to correct central nervous system abnormalities, human MAPCs were used to repair brain damage in an animal model of stroke. In this case, the stem cells themselves were not directly responsible for the repair, but had acted by an as-yet unknown mechanism to minimize stroke damage, repair the blood vessel structures in the brain, or possibly recruit stem cells from the brain to go to the damage zone.

For type 1 diabetes, there is some evidence in animals to suggest that ES cells may be able to differentiate into insulin-producing beta cells. Other cells that might be sources for beta cells come from the pancreas itself, and from liver cells that have had transcription factors introduced that are important for beta cell development. In liver disease, a recent example of research on the potential of stem cells for correcting defects comes from studies in which hematopoietic stem cells have been shown to correct the liver disease, hereditary tyrosinemia, in animals. Notably, correction of this defect appears to have resulted from fusion of the healthy donor stem cells with diseased liver cells to produce functional hybrid cells–a phenomenon which requires further study to understand its implications for therapy development.

Finally, researchers foresee the use of stem cells to produce pieces of organs or whole organs–first relatively simple structures such as arteries or heart valves, then more complex organs such as the liver and kidney. This future is moving closer. At the University of Minnesota, muscle cells created from MAPCs have been used to engineer smooth muscle layers that have mechanical characteristics similarto newborn rat aortic smooth muscle cells. By co-culturing the engineered and aortic cells, a structure is produced that looks and functions like a normal artery.

Future Research Directions

Dr. Verfaillie emphasized that stem cell research has made significant progress over the past 6 years in expanding knowledge about developmental cell pathways and cell plasticity, and inspiring new concepts for treating disease. While stem cells hold the promise of new and improved therapies, she believes that clinical trials testing these approaches are at least 5 to 10 years away. Much additional research is still needed.

The NIDDK is continuing progress toward the twin goals of understanding the fundamental processes underlying stem cell development and applying that knowledge for clinical use and benefit. Large-scale research efforts include the Beta Cell Biology Consortium, which is supporting a number of stem cell research efforts with the ultimate goal of developing a cell-based therapy for insulin delivery to cure diabetes. The Institute is also supporting a number of Progenitor Cell Genome Anatomy Projects. These GAPs are conducting research on organ- and tissue-specific stem cells that includes their identification and isolation, characterization of their gene expression patterns during tissue development, and development of better tools for characterizing the functions of these genes. The NIDDK is also participating in the NIH-wide Stem Cell Task Force that is, in consultation with external scientists, developing strategies and initiatives to enhance scientific research and resources in this exciting and promising field.

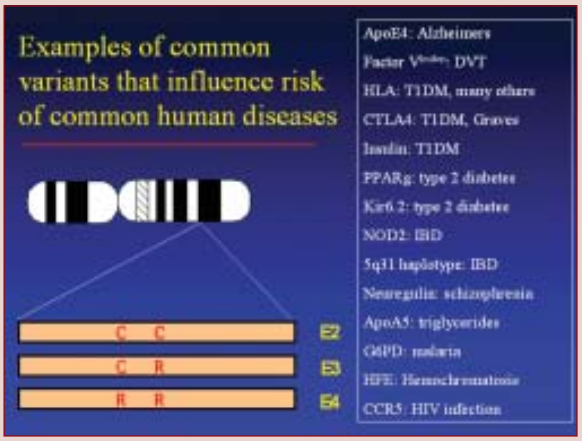
---

* Consistent with the policy announced by President George W. Bush on August 9, 2001, NIH funding of research involving human embryonic stem cells is in accordance with specific criteria established by the Administration.

Human Genome Variation and the Genetics of Common Human Disease

Dr. David Altshuler

David M. Altshuler, M.D., Ph.D., is Assistant Professor of Genetics and of Medicine at Harvard Medical School, and a member of the Department of Molecular Biology and Attending Physician in the Diabetes Unit at Massachusetts General Hospital (MGH). He is also Director of Medical and Population Genetics at the Whitehead Institute/MIT Center for Genome Research, and an Affiliate Member of the Whitehead Institute for Biomedical Research. He received his M.D. and Ph.D. from Harvard Medical School, and completed clinical training in Internal Medicine and Endocrinology at Massachusetts General Hospital. Dr. Altshuler's research focuses on the genetic basis of common diseases.



In the human population, genetic variation contributes to individual differences in susceptibility to common, complex diseases.

Identical twins excepted, everyone looks different. The scientific term for an organism's physical appearance is "phenotype," and there is a great deal of phenotypic variation among the human population. Some people are tall, some short, some have black hair, some red, and so on. These differences in phenotype are caused by subtle variations in people's genetic makeup–their "genotype." And, just as genetic differences can give rise to different appearances, they can also predispose some people to disease. From a scientific perspective, a central question is, to what extent do underlying genetic differences contribute to the common diseases?

Consider two people: next-door neighbors who live in the same environment. If one develops type 2 diabetes, what are the chances that the other will, too? If they are truly unrelated, that chance is around 5 to 10 percent. But, if they are siblings, the risk rises to about 30 percent. If they are identical twins–two people who share the same genotype–the risk rises to 80 to 90 percent. Clearly, then, genetic factors play an important role in the development of this disease. The challenge facing researchers is to identify which genetic elements are important.

For rare genetic diseases, especially those caused by a mutation in a single gene, researchers study families in which the disease regularly occurs. By correlating the appearance of the disease with the inheritance of specific chromosomal regions, scientists can narrow down the stretch of DNA that is likely to carry the responsible mutation. For common diseases that have a genetic component, this approach has not worked particularly well, because these diseases

result from the inheritance of multiple genetic variants, each of which may have only a subtle influence on the overall risk. It is much harder to find these subtle effects than the dramatic influences seen in diseases such as cystic fibrosis and Huntington's disease. For common conditions, scientists have to look at a large population of affected individuals and compare them with a large population of unaffected ones, and try to determine which genetic elements are associated in a more subtle way with development of the disease. Such studies are called "association studies," and are much more difficult to carry out and interpret than studies of diseases caused by a mutation in a single gene.

Genetic Variation, Environment, and Disease

Dr. Altshuler pointed out that it may seem counterintuitive to think that a widespread, common set of genetic variations could predispose a large number of people to a common disease. After all, one would think that mutations that cause people to get sick would be eliminated from the population by natural selection over many generations. Natural selection is the process by which minor genetic variations cause organisms to be more or less fit for their environments. Those who are more fit survive, and pass on their genes to their offspring. Those who are less fit do not pass on their genes, and these mutations are eliminated from the population. So, how could mutations that cause relatively common diseases today have persisted in the human genome for so long? The answer lies in the observation that mutations can be either beneficial or harmful in the context of the organism's environment. Thus, mutations that conferred a benefit in the past may no longer be beneficial in the modern environment.

One example of this sort of phenomenon is the "thrifty gene" hypothesis. For most of human history, people struggled to find enough to eat. Scientists have hypothesized that individuals who were able to accumulate fat might have had an advantage, because they were more likely to stockpile energy when food was plentiful and were therefore more likely to survive when food was scarce. However, the environment has changed dramatically in the last few hundred years, much too quickly for our genes to evolve to match it. The ability to accumulate fat is no longer beneficial in an environment in which calorie-rich food is plentiful and life is often sedentary. Obesity is emerging as one of the most important public health issues, as more and more people become overweight and develop diseases associated with obesity, such as type 2 diabetes and cardiovascular disease.

Human Genetic Variation

Dr. Altshuler noted that, to understand what exactly is meant by the term "genetic variation," it is helpful to think of it at the level of DNA. Chromosomes are made up of long stretches of DNA, which is composed of a double-stranded, linear sequence of nucleotides: adenosine (A), cytidine (C), guanosine (G), and thymidine (T). Geneticists have traditionally focused on mutations that alter the protein encoded by a gene. Scientists now appreciate that DNA sequences that do not code for proteins–which make up about 99 percent of our DNA–are also a potentially important source of genetic variation.

For long stretches, a given DNA sequence may be identical in two different people. However, every so often, one (or more) of the nucleotides differs. Such a site is called a "single nucleotide polymorphism"

or SNP. SNPs appear in the human genome roughly once every 1,000 nucleotides and a given SNP may be quite widespread. For example, in 60 percent of the population, a given chromosome may have an A at a certain position, while the other 40 percent of the population may have a G instead. These two forms–A and G–are called alleles, or variants, of that SNP. At some point in the distant past, there was a single form, but a mutation altered the base, and this mutation was passed down and spread throughout the population. "Haplotype" is the term used to describe a set of SNP alleles along a region of a chromosome. Theoretically, there could be many haplotypes for a chromosomal region, but studies typically find relatively few, common haplotypes in the entire human population.

One of the ways scientists used haplotypes is as a kind of molecular "signature" for a specific stretch of DNA. To examine large stretches of DNA comprehensively, scientists use a technique called "haplotype mapping." As its name implies, haplotype mapping uses a series of SNPs as markers to identify a stretch of DNA. Because all people today are descended from the same small population of humans that lived long ago, it is often the case that the same haplotype will be seen at a high frequency in an apparently unrelated set of individuals. When such a particular set of SNPs is seen more frequently in people with a disease than in those without it, those SNPs and their alleles are said to be associated with the disease. This finding suggests that there may be genes in that chromosomal region that contribute to developing the disease. Thus, haplotype mapping is a powerful and useful tool for narrowing down regions of DNA that might contain genes associated with common diseases. Scientists have used haplotype mapping to identify candidate genes that contribute to diseases such as Crohn's disease (an inflammatory bowel disease) and type 2 diabetes.

Common Genetic Variation, Gene Expression, and Disease

Dr. Altshuler described a complementary approach to looking for genes that contribute to disease, which is to examine genes, not one at a time, but in large sets of genes known collectively to contribute to a given biological system–for example, the mitrochondrion, the cell's energy "factory." Studies such as these focus, not on genetic variation at the DNA level, but on changes in which genes are turned on or off at a particular time or under certain conditions. By looking at changes in the expression levels of many genes at once, so-called "expression profiling," researchers can get a better view of the "big picture," and detect large-scale changes that might be missed in studies that focus on changes in a single gene.

Dr. Altshuler and his colleagues have recently used expression profiling in a small study of people with type 2 diabetes, pre-diabetes, and a control group with neither condition. They used DNA microarrays to measure gene expression of over 22,000 genes in these three groups. The researchers found that people with diabetes had, in general, lower expression of genes involved in metabolism in the mitochondria, the cell's energy centers. Out of 106 genes previously known to be involved in the chemical process known as oxidative phosphorylation, expression of 94 was decreased people with diabetes. Interestingly, these changes were modest on a gene-by-gene basis, but overall, the result was unmistakable: this important metabolic pathway was less active in people with diabetes. When the researchers looked closely at the genes affected, many of them were found to be influenced by the regulatory protein peroxisome proliferator activator protein-gamma co-activator-1 alpha. Thus, it might be that a subtle alteration of this single gene has a more dramatic effect, as it cascades down through the series of genes it controls.

Although a small study, this investigation is significant because it demonstrates that it may be necessary to examine networks of genes in order to detect changes in activity. Studies of change in a single gene are practical when the change in that gene's expression is relatively large and overall expression patterns vary little from one person to another, but expression profiling permits researchers to examine subtle changes in gene expression patterns on a higher level. Such strategies ultimately may be necessary to detect small changes in pathways, which may ultimately turn out to be a defining feature of common, complex genetic disorders, such as type 2 diabetes.

The Future

One possible implication of the emerging appreciation for common genetic variation is how it could alter current approaches to disease treatment. For example, the knowledge that certain combinations of haplotypes might place a person a risk of developing a particular disease could influence treatment. Similarly, if people with one haplotype are less responsive to a particular treatment than others, it would be possible to give them alternative therapies from the outset. While such knowledge is years away, through studies of genetic variation and interaction with the environment, the possibility exists to "personalize" medicine through therapies tailored to a particular individual.

The NIDDK, along with multiple other NIH Institutes and Centers, is sponsoring an initiative entitled "Large-Scale Genotyping for the Haplotype Map of the Human Genome." This initiative will develop a map of the haplotype patterns and of the genetic variants that are most informative for detecting these patterns. The haplotype map is expected to be a key resource for finding genes affecting health, disease, and response to drugs and environmental factors, and for beginning to understand the pattern of human genetic variation.

Also, the NIDDK is sponsoring a Diabetes Genome Anatomy Project (DGAP). This overall goal of this project is to identify the genes and gene sets, as well as the proteins, involved in insulin action and the predisposition to type 2 diabetes. The DGAP is also examining secondary changes in gene expression that occur in response to the metabolic abnormalities present in diabetes. The DGAP will define the "normal" patterns of gene expression and response to insulin, the impact of diabetes on gene expression patterns and response to insulin, and the extent to which genetic variability might contribute to the alterations in expression or to the development of diabetes. This project, and the resultant database, will aid investigators as they strive to unravel the complexity of insulin action and its perturbation in diabetes, and ultimately will help develop more effective and specific modes for studying and treating the disease.