

Extraordinary Research Opportunities

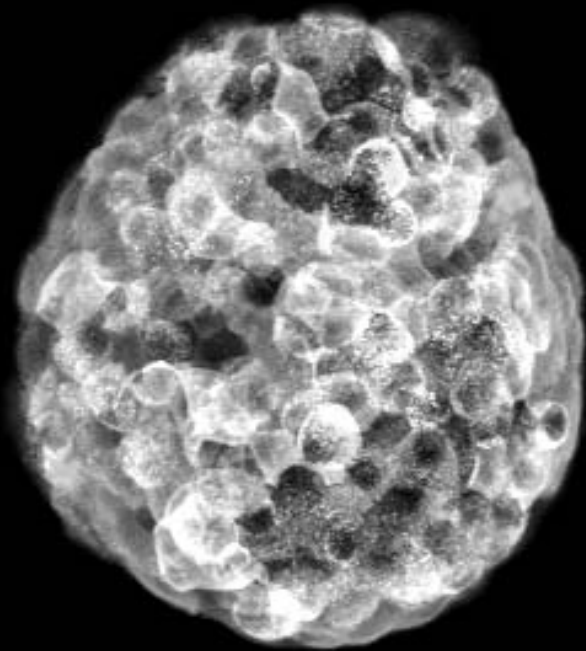


Photo: Dr. Todd C. Brelje and Dr. Robert L. Sorenson, University of Minnesota.

Extraordinary Opportunity:

Genetics of Diabetes

Both type 1 and type 2 diabetes are complex genetic diseases that result from interactions between multiple genes and environmental factors. The identification and understanding of the many genetic determinants of both forms of diabetes and the complications that they share are of critical importance to conquering this disease. With greater understanding of the genetic causes of diabetes will come the ability to identify those individuals who are at highest risk, and to identify new targets for action to prevent the disease or at least stop it in its tracks.

Identification of genetic variations that predispose an individual to diabetes and its complications may reveal new molecular pathways that will provide targets for the development of candidate drugs. Once researchers know the series of cascading events that result in full-blown diabetes or its complications, they will be able to devise and deliver interventions at various points along this continuum. In the future, knowledge of an individual's genetic makeup may also help physicians tailor therapies or prevention measures with great specificity to individual patients in order to maximize benefits and minimize unwanted side effects. While the absolute prevention of diabetes is the ultimate goal, the ability to delay the onset of diabetes by even a few years would have enormous benefits because the severity of diabetes complications correlates with duration of disease. Thus, the potential of genetic tests to identify high risk patients, who could benefit most from intervention, assumes added importance with the recent demonstration of success in delaying onset of type 2 diabetes in clinical trials. Finally, because diabetes appears to result from an interaction of genetic susceptibility with environmental factors, a better understanding of the genetics of diabetes could help to isolate the environmental influences and point the way to avoiding or minimizing their effects.

Since the issuance of the DRWG's Strategic Plan, the NIH has capitalized on a diversity of new knowledge, technologies, and scientific approaches to speed the identification of diabetes genes and garner insights into how they work. Among the most important of these is the new knowledge emanating from publication of the first draft sequence of the entire human genome in 2001. This tremendous accomplishment was the result of both an NIH-funded public effort and an industry-funded private effort. Although still being refined, these draft sequences have already provided researchers with invaluable information for identifying and localizing genes.

The NIH has undertaken multiple approaches to elucidate the genetics of diabetes. Particularly noteworthy are the establishment and utilization of:

- ◆ *Research consortia* to bring the collective talents and shared data resources of many investigators and institutions to bear on the search for diabetes genes.
- ◆ *Large-scale genome-wide scans* to hunt for genes in a more comprehensive and efficient manner.
- ◆ *Microarray technology* to determine patterns in gene activity and to “profile” how genes are expressed in different tissues and how they are altered in diabetes.

- ◆ *“Knockout” mice* to study the biologic effects of inactivating genes suspected of contributing to development of diabetes or its complications.
- ◆ *Animal models of diabetes* to identify genes involved in diabetes and its complications.

The NIH has been particularly responsive to the DRWG's emphasis on genetics consortia as a means of applying the full power of investigative talent and research resources to the hunt for diabetes susceptibility genes. Such consortia are crucial to advances in diseases involving multiple genes, such as type 1 and type 2 diabetes, because genetic clues are more likely to emerge when large amounts of data from many sources are combined and analyzed by a wide range of statistical approaches. Through these consortia, the human genome resources can be fully utilized and we are more likely to detect patterns that might otherwise escape observation in smaller studies. The consortia also provide expertise and mechanisms for continually refining approaches to the genetics of diabetes in order to incorporate the latest technology as soon as it becomes available.

Genetics of Type 1 Diabetes

Type 1 diabetes usually strikes in childhood or young adulthood but can affect an individual at any age. It is an “autoimmune disease,” in which the body attacks its own tissue. Type 1 diabetes is characterized by the immune system’s destruction of insulin-producing beta cells, which are contained in clusters, called islets, within the pancreas. Without insulin, the body cannot regulate glucose metabolism—a condition that can rapidly cause death.

Researchers have yet to determine the precise factors that cause the immune system, which normally protects the body from harmful bacteria and viruses, to initiate this misguided attack on its own insulin-producing cells. However, many studies have suggested that an environmental exposure of some sort may trigger this autoimmune disease process in individuals who have an underlying genetic susceptibility to develop type 1 diabetes. Multiple genes are believed to be involved in this susceptibility and to interact with each other and with the environment to initiate the cascade that leads to the development of type 1 diabetes.

Scientists have already identified variations within genes of the major histocompatibility complex (MHC) that are major contributors to type 1 diabetes. The MHC genes are key to the immune system’s ability to differentiate between what is “foreign” to the body compared to what is “self.” Research has shown that the strongest genetic predisposition to type 1 diabetes is conferred by two common genetic variations of MHC molecules known as human leukocyte antigens (HLA, a subset of the human MHC genes). About 85 to 90 percent of individuals with type 1 diabetes are

positive for HLA types known as DR3-DQ2 or DR4-DQ8. People who have those HLA types are also much more likely to develop diabetes than are persons with other forms of HLA. Conversely, some HLA types have been identified that appear to reduce the risk of developing type 1 diabetes. With this knowledge, it is already possible to identify individuals at increased risk for type 1 diabetes, and these genetic tests have already been incorporated into ongoing clinical trials for diabetes prevention and clinical research to identify environmental triggers of the disease. The identification of other genes contributing to disease risk would greatly increase our ability to predict who will develop type 1 diabetes. It may eventually enable the introduction of prevention strategies in those at risk before an autoimmune attack is launched, while it can effectively be arrested or mitigated.

In response to the recommendations of the DRWG, several comprehensive new initiatives are under way to identify genes conferring risk for type 1 diabetes:

UNDERSTANDING THE ROLE OF THE HISTOCOMPATIBILITY GENES

With NIH support, the International Histocompatibility Working Group (IHWG)—a large, multi-project endeavor to define the role of HLA genotypes in many diseases—has expanded its scope to include additional studies on diabetes. One component of this project is studying the role of HLA in type 1 diabetes and determining whether there are variations in other genes within the HLA region that contribute to disease susceptibility. Through the IHWG, current efforts will be expanded to establish a central repository of genetic data relevant to type 1 diabetes and to provide an Internet-based information service for researchers. Supported by the NIH and the Juvenile Diabetes Research Foundation International (JDRF), the IHWG will also launch a project to discover genetic sequence variations, called single nucleotide polymorphisms (SNPs), within immune system genes that appear to be related to type 1 diabetes.

NEW GENETICS CONSORTIUM PROPELS SEARCH FOR TYPE 1 DIABETES GENES

The NIH has established an “International Type 1 Diabetes Genetics Consortium,” building upon recommendations from a pivotal meeting of international experts held in November 2000. This consortium is striving to identify type 1 diabetes susceptibility genes by “scanning” human genome sequences in families from the U.S., Europe, and Australia. Thus far, the data from over one thousand diabetic sibling pairs have been combined and analyzed. These analyses have revealed several chromosomal regions that may be harboring diabetes susceptibility genes. Because large sample sizes are required to identify susceptibility genes, especially for complex diseases such as diabetes, the consortium proposes to collect data on an additional 2,000 families for genetic analysis.

SEQUENCING CHROMOSOMAL REGIONS IMPLICATED IN A TYPE 1 DIABETES MOUSE MODEL

Researchers have found evidence that up to 20 chromosomal regions in the non-obese diabetic (NOD) mouse, which spontaneously develops type 1 diabetes, may contain genes that determine development of diabetes. In the NOD mouse, several of these regions have been narrowed down to an area that is amenable to sequencing. Several well-defined diabetes susceptibility regions are being sequenced in the NOD “Mouse Sequencing Project” to try to identify the sequence differences between the NOD mouse strain and non-diabetic mouse strains that account for increased susceptibility to diabetes. These sequences will be deposited in the Mouse Genome Database.

▶ USING ANIMAL MODELS TO IDENTIFY GENES FOR TYPE 1 DIABETES

The DRWG underscored the immense value of animal models as tools for finding genes that influence or cause diabetes. The NOD mouse and BioBreeding (BB) rat are spontaneous models of type 1 diabetes that are being intensively investigated to identify genes that affect the risk of autoimmune diabetes. One approach for mapping the genes responsible for diabetes in these models has been the introduction of genetic material from a diabetes-resistant mouse or rat strain onto the NOD or BB backgrounds, producing “congenic” strains. These congenic strains can be used to identify crucial genetic variations between the susceptible and resistant strains. For example, this approach has recently been used to identify a gene in the BB rat that causes diabetes. Valuable congenic mouse models are being preserved and distributed in a newly established “NOD Mouse Repository” at the Jackson Laboratory in Maine. The availability of these congenic models should aid in the discovery of the genes that determine diabetes in the NOD mouse.

Since the DRWG report in 1999, important new information has emerged about the genetic basis of type 1 diabetes:

▶ NEW CHROMOSOMAL REGIONS IMPLICATED IN HUMAN STUDIES

Over the past two years, researchers have been using family studies to identify additional genes that contribute to type 1 diabetes. Two independent studies conducted in the U.S. and Great Britain have analyzed genome-wide scans in pairs of siblings with diabetes (sib pairs). The data from these two studies have been incorporated with data from a new U.S. cohort and the combined results on 831 sib pairs have been analyzed. A second effort in Scandinavian patients studied 464 sib pairs. Both of these studies identified chromosomal regions that show evidence of harboring genes that contribute to diabetes; however, all of these regions were not reproduced in both studies. Interestingly, some of these regions have been independently linked to other autoimmune diseases, such as rheumatoid arthritis, celiac disease, Crohn’s Disease, and thyroid disease. Efforts are under way, through the International Type 1 Diabetes Genetics Consortium, to combine the data from all of these studies as well as to collect and analyze additional samples to unambiguously localize diabetes genes relevant to individuals of different ethnic and racial backgrounds.

► IDENTIFICATION OF A NEW GENE CAUSING DIABETES IN THE RAT

The BB rat model of type 1 diabetes is used to study non-MHC genes contributing to the disease. One such mutant gene on rat chromosome 4 is responsible for the trait of lymphopenia, which is characterized by a reduced number of white blood cells. This mutant gene is essential for the development of diabetes. Positional cloning was used to show that lymphopenia is due to a frameshift mutation in a novel member of the immune-

associated-nucleotide (IAN) related gene family, resulting in a significant truncation of the protein. Although this is a newly-identified member of this gene family and its function is not known, other members of this family are expressed in T cells and are thought to be involved in T cell development. A challenge for the future is to investigate the role of this new family of genes in the development of diabetes to provide new insights into the disease process.

Genetics of Type 2 Diabetes and Obesity

As with type 1 diabetes, susceptibility to type 2 diabetes and obesity is determined by both genetic and environmental factors. For type 1 diabetes, we have identified a major disease gene (HLA) but we know relatively little about environmental factors involved in development of disease. In contrast, for type 2 diabetes, we know that environmental factors such as diet and activity are risk factors but we know much less about the specific genes involved in disease susceptibility.

We have gathered clues on the genes that may be involved by studying rare syndromes that have diabetes as a key feature of the disease (see table) but which develop because of defects in a single gene. Over ten years ago, the first gene to be directly linked to diabetes, the insulin receptor gene, was identified. Mutations in this gene were responsible for the development of several syndromes: leprechaunism, type A insulin resistance and Rabson-Mendenhall syndrome. More recently, five genes have been identified that, when mutated, can lead to the development of Maturity Onset Diabetes of the Young (MODY). MODY occurs when there is defective glucose-stimulated insulin secretion from the beta cell. Four of the MODY genes,

HNF-4 alpha (*MODY1*), HNF-1alpha (*MODY3*), IPF-1 (*MODY4*) and HNF-1beta (*MODY5*), are transcription factors that are important for the development of the beta cell. MODY2 is the glucokinase gene, which is the beta cell's glucose sensor. Mutation of a novel gene required for beta cell function has been identified as the cause of another diabetic syndrome, Wolfram Syndrome. Several different mutations in mitochondrial DNA have been shown to lead to diabetes that is often associated with deafness. Finally, in the past year, three mutant genes responsible for lipotrophic diabetes have been identified; these genes may help to elucidate the role of the fat cell in the development of diabetes.

GENES CAUSING MONOGENIC FORMS OF DIABETES

Gene/Protein	Syndrome	Year
Insulin Receptor	Leprechaunism	1988
	Type A Insulin Resistance	1989
	Rabson-Mendenhall	1990
tRNA ^{Leu}	MELAS Syndrome	1991
HNF-4alpha (MODY 1)	MODY	1996
Glucokinase (MODY 2)	MODY	1992
HNF-1alpha (MODY 3)	MODY	1996
IPF-1 (PDX) (MODY 4)	MODY	1999
HNF-1beta (MODY 5)	MODY	1999
Wolframin	Wolfram Syndrome	1998
Lamin A/C	Dunnigan Lipodystrophy	2000
Seipin	Berardinelli-Seip	2001
Scurfin	X-linked Neonatal Diabetes	2001
AGPAT2	Berardinelli-Seip	2002
ALMS1	Alstrom Syndrome	2002

As demonstrated by these monogenic syndromes, there are many different pathways that lead to the development of diabetes. The spectrum of clinical presentations of type 2 diabetes—and the close association of the disease with other conditions such as obesity, high blood pressure, and lipid abnormalities—add to the complexity of defining the causative genes. Characterization of specific subgroups of people with the disease is important for genetic analysis. For example, if a gene were predominantly responsible for the early onset of

type 2 diabetes, then the effects of that gene would be easier to observe in a study that focused on type 2 diabetes in children. Its effects could be diluted in a study that looked at all ages. While these considerations make it particularly challenging to identify the genetic variations or mutations that predispose to type 2 diabetes, it is nonetheless critically important to meet this challenge, given the huge and increasing health burden type 2 diabetes imposes on the American people and its disproportionately heavy burden on minority populations.

In response to the recommendations of the DRWG, and taking advantage of new technologies for genetic analysis, ongoing efforts to identify genes for type 2 diabetes have been expanded and new efforts have begun:

BUILDING CONSORTIA TO IDENTIFY GENES FOR TYPE 2 DIABETES

The NIH has bolstered the research efforts of the “International Type 2 Diabetes Genetic Linkage Analysis Consortium,” which was formed in 1998 to accelerate the search for genes. The pooling of data from the many groups in the consortium increases the probability of identifying genes that influence this genetically complex disease. This combined effort also means that more samples are available for analysis of individual ethnic groups than is possible within a single study. With support from the NIH, the groups in the consortium are currently undertaking more precise mapping of a region of a potential diabetes susceptibility gene suspected to be on chromosome 1q as well as investigating potential susceptibility genes on other chromosomes. Because African Americans are disproportionately affected by type 2 diabetes and are underrepresented among the data in the consortium, the NIH has supported the collection of additional DNA samples from this population. These samples have been genotyped by the NIH-funded Center for Inherited Disease Research (CIDR) and will be combined with the other African American samples in the consortium for analysis of genetic variation that may be specific to this population. Biannual meetings of consortia participants are convened to monitor progress attained to date and to identify new research avenues and strategies.

UNRAVELING THE GENETICS OF OBESITY

Identification of genetic factors that influence obesity may provide new tools to combat this condition, beyond exercise and modification of diet. To this end, the NIH is considering establishing a network of collaborations to further the understanding of the genetics of obesity. This should complement our studies on the genetics of type 2 diabetes and may lead to an understanding of the genetic relationships of these two conditions. Meetings to discuss this potential network have been held and a demonstration project is planned.

FOCUSING ON DISPROPORTIONATELY AFFECTED MINORITY POPULATIONS

Two NIH intramural efforts have been launched to investigate the genetics of diabetes and obesity in minority populations that are disproportionately affected by type 2 diabetes. The African American Diabetes Mellitus Study (AADM) is a collaboration between the NIH and Howard University to study genetic risk factors that predispose people of African descent to diabetes. Because diabetes is also influenced by environmental risk factors such as diet, one potentially powerful approach is to study genetic risk factors in West Africans because they have fewer dietary and nutritional confounding variables than do African Americans. The goal of recruiting 400 pairs of siblings affected with diabetes was met in the fall of 2000 and samples from West Africa have been genotyped by CIDR.

Another effort is an ongoing study of genetic factors that lead to both type 2 diabetes and obesity in the Pima Indian population of Phoenix, Arizona. This population has one of the highest known rates of diabetes in the world. Working closely with the Pima Tribal Council, the NIH is also studying the development of complications and the genes that predispose to diabetic kidney disease, which is highly prevalent among the Pimas. Because the Pima population is relatively homogeneous, studies could identify factors that would not only shed light on the genetics of diabetes in the Pimas, but also on genetic propensities for developing the disease that may exist in other populations.

As a result of these efforts, new information has emerged about the genetic basis of type 2 diabetes:

▶ GENES IMPLICATED IN TYPE 2 DIABETES

Several genes have been implicated in the pathogenesis of type 2 diabetes. Researchers have recently identified variants of a gene that contribute to diabetes in Mexican Americans. This gene has been named *calpain 10* because it is a member of a calcium-activated neutral protease family. Proteases regulate cell processes by catalyzing the cleavage of proteins within the cell, which may have the effect of either activating or degrading the target protein. An important goal is to identify the protein targets of *calpain 10* to better understand its role in the cell and in diabetes. Association studies have also suggested that another gene, *PPAR gamma*, may contribute to type 2 diabetes.

The *PPAR gamma* gene plays an important role in the differentiation of adipocytes affecting the distribution of lipids. Its protein product is also the target of the thiazolidinediones, a class of insulin-sensitizing drugs used to treat type 2 diabetes. A common variant or allele of *PPAR gamma*, *Pro12Ala*, has been studied. By combining multiple studies, researchers showed that patients with the more common *Pro12* allele have an increased incidence of diabetes. Additional research will be undertaken to explore these preliminary findings with regard to the *calpain 10* and *PPAR gamma* genes.

▶ GENES INVOLVED IN PANCREATIC ISLET DEVELOPMENT MAY CONTRIBUTE TO DIABETES RISK

In another genetic study, NIH and French investigators have demonstrated that mutations in the insulin promoter factor (*IPF-1*) gene may be related to the development of type 2 diabetes. The *IPF-1* gene is critical for embryonic development of the pancreas and for regulation of endocrine pancreas-specific genes in adults. As previously described, some individuals with particular mutations in *IPF-1* develop a rare form of diabetes, Maturity Onset Diabetes of the Young (MODY). Three novel *IPF-1* mutations were found when screening over 60 unrelated individuals of French ancestry with type 2 diabetes. This is the first evidence that *IPF-1* may represent a diabetes-predisposing gene in a portion of individuals with the common form of type 2 diabetes.

continued

In the quest for gene variations and mutations that influence susceptibility to diabetes, one group of investigators recently described two mutations in the *NEUROD1* gene that are associated with the development of type 2 diabetes. The protein encoded by this gene functions as a regulatory switch for the development of that portion of the pancreas that secretes insulin and glucagon—hormones responsive to blood glucose levels. In a mouse model with two disrupted copies of this gene, pancreatic islet development is abnormal and overt diabetes develops. In order to determine if *NEUROD1* plays a causative role in type 2 diabetes in humans, investigators screened over 90 individuals with the disease for mutations in this gene and identified two families with mutations. It is possible that the development of type 2 diabetes in carriers of either of these two mutations may result from a disruption of gene activity in the islets, but it is unlikely that *NEUROD1* is a major factor in most forms of diabetes.

► IDENTIFYING GENES FOR OBESITY

In humans, over 27 studies have employed genome-wide scans to identify variations in genes that predispose to the development of obesity. Several candidate chromosomal regions have been identified in multiple studies (for example, chromosomes 10p and 2p). These regions are being mapped to identify the genes contributing to obesity. A new opportunity in diabetes research is to identify a link between these obesity genes and a risk for type 2 diabetes.

In addition, genetic clues about the ominous connection between type 2 diabetes and obesity

are being obtained through study of the lipodystrophies—a group of disorders characterized by selective loss of fat from various parts of the body. These conditions are often accompanied by insulin resistance or diabetes, elevated levels of blood lipids, and vascular disease. Patients with one of the rare forms of lipodystrophy, familial partial lipodystrophy, lose fat at puberty from the extremities, trunk, and gluteal regions of the body, but excess fat becomes deposited in the face, neck, and back. Investigators have found that mutations in the *lamin A/C* gene, which is known to cause a form of muscular dystrophy, also cause this syndrome. Characteristics of this gene and the protein it produces suggest that it may be involved in one or more activities required by fat cells in specific tissue beds. Researchers have hypothesized that the loss of fat cells affects insulin sensitivity through reduced levels of “adipocyte-derived circulating factors.” One such factor is leptin, a hormone produced by fat cells that regulates food intake and energy metabolism. Thus, this research may provide insights into complex metabolic problems, such as those seen in type 2 diabetes and obesity, by defining new pathways that regulate fat cell mass and insulin sensitivity.

Similarly, genes involved in a congenital form of complete lipodystrophy called Berardinelli-Seip may also have relevance to diabetes and obesity. Mutations in two genes cause this syndrome: one linked to chromosome 9q and one linked to chromosome 11q. The gene on chromosome 11 has now been isolated and named *seipin*. This gene is expressed in brain and testes and its function is currently being sought. It codes for a unique protein that probably crosses the membrane at least two times. The gene on chromosome 9 has been identified as *AGPAT2*, which codes for an enzyme required for the synthesis of triglycerides.

Genetics of the Complications of Diabetes

Genes are a critical factor not only in the onset of type 1 and type 2 diabetes but also in the onset and progression of the complications that result from both forms of the disease. For example, previous studies have suggested a familial clustering of diabetic kidney disease.

Because the complications of diabetes do not appear in all patients, nor with the same severity, researchers believe that some diabetes patients carry gene variants that make them more susceptible to certain complications than other patients. For example, it is known that diabetes patients from minority groups are more prone to certain complications than non-minorities with diabetes. To solve this conundrum, the NIH has undertaken several initiatives to pinpoint genes that predispose individuals to develop complications from diabetes.



Photo: Richard T. Nowitz, for NIDDK.

FOLLOW-UP OF PATIENTS WHO PARTICIPATED IN THE DIABETES CONTROL AND COMPLICATIONS TRIAL (DCCT)

An important genetics component has been added to the follow-up study of type 1 diabetes patients who participated in the landmark DCCT, which demonstrated that patients who maintain their blood glucose levels as close to normal as possible throughout the day can prevent or delay eye, kidney, and nerve complications of the disease. The DCCT follow-up study is the Epidemiology of Diabetes Interventions and Complications (EDIC), described in further detail in “Clinical Trials and Clinical Research of Critical Importance.” In total, these patients with diabetes have been followed clinically for up to 15 years. Recently, the EDIC was expanded to include a genetic study that will acquire samples from nearly 1,400 patients studied in the DCCT and their parents. Data will be used initially to study candidate genes that may predispose to diabetic complications.

FINDING SUSCEPTIBILITY GENES FOR DIABETIC KIDNEY DISEASE

Two studies have recently shown that diabetic kidney disease is a highly heritable trait. In response to this new information, the NIH established the Family Investigation of Nephropathy and Diabetes (FIND) study. The FIND study consortium has been formed to uncover candidate genes associated with development of kidney complications in patients with type 1 or type 2 diabetes. FIND has also incorporated a retinopathy (eye disease) study as part of its subject evaluation. A specific objective is to search for susceptibility genes in subpopulations of Caucasians, African Americans, Hispanic Americans, and Native Americans across the U.S. Two methods are being used to try to identify the genes that confer susceptibility to kidney disease and retinopathy in diabetes patients. Investigators will use both traditional family-based analyses (i.e., affected and discordant sib pair and relative pair analyses) in all families, and Mapping by Admixture Linkage Disequilibrium (MALD) in African American and Hispanic American cases and controls.

FINDING SUSCEPTIBILITY GENES FOR DIABETIC EYE DISEASE (RETINOPATHY)

In September 2000, the NIH convened a multi-disciplinary group of clinicians, epidemiologists, statistical geneticists and molecular geneticists to discuss recent developments in human molecular genetics and the opportunities created for studying the genetic predisposition underlying the development and progression of diabetic retinopathy. The group generated several recommendations, including a need to support multi-disciplinary studies to track down genes for diabetes and its complications, to support development and distribution of animal models, to develop novel methods of retinopathy assessment, and to attract genetic epidemiologists and statistical geneticists to work in vision research. Based on these recommendations, the NIH issued a solicitation for additional research projects. It is hoped that increased understanding of the genetics underlying the development or progression of retinopathy will have an impact on future diagnostic, prevention, and intervention strategies.

CONTRIBUTIONS OF MOUSE MODELS TO DISCOVERIES ABOUT DIABETES COMPLICATIONS

The NIH has launched an initiative to establish a cross-disciplinary Mouse Models of Diabetic Complications Consortium. By developing mouse models of diabetes complications that mimic human disease, this consortium will facilitate the study of disease prevention and treatment and the testing of candidate genes that emerge from human genetic studies. Mice developed through the consortium will be distributed to the research community. (See also “Mouse Models of Diabetes.”)

Opportunities to Harness New Tools in the Search for Diabetes Genes

Research undertaken since the DRWG issued its Strategic Plan has opened up many new avenues of scientific opportunity. Some of the most compelling opportunities relate to harnessing the tools of modern molecular genetics, new techniques for physiologic characterization of study populations, and bioinformatics. Highlights of the opportunities for application of these new technologies to diabetes follow:

CAPITALIZING ON THE HUMAN GENOME SEQUENCE AND NEW TOOLS OF GENETIC ANALYSIS

As the DRWG pointed out, with the historic achievements of the Human Genome Project, an explosion in the amount of raw human genetic data has occurred. The first drafts of the human and mouse genomes are now available. Comparison of the genomes of multiple species, including the human and mouse sequences, will enhance efforts to identify genes in mouse models of diabetes; this, in turn, will aid in the characterization of human gene variants that influence the development of diabetes. While the publication of the first draft sequence of the entire human genome in 2001 is a landmark scientific achievement, annotation of this sequence must proceed at a rapid pace in order to localize all disease-causing genes and use that knowledge to benefit patients and their families. Such efforts at annotation will help to accelerate progress in finding diabetes genes in regions identified by genome scans.

IDENTIFYING ASSOCIATIONS BETWEEN DNA VARIATIONS AND DISEASE

In the search for genetic variations associated with diabetes, it would be prohibitively difficult, at the current cost of DNA sequencing, to sequence the entire genome of every individual in a study. Therefore, scientists have developed methods to test for sequence differences called single nucleotide polymorphisms, or SNPs. Some of these SNPs are relatively rare, but many are relatively common. Already, through the SNP Consortium and the human genome sequencing effort, over 2,500,000 SNPs have been identified and deposited in a public database. These are now available for fine mapping of chromosomal regions of interest and localizing disease-associated polymorphisms. To pursue opportunities to identify diabetes genes efficiently, researchers will require access to high-throughput, cost-effective SNP genotyping services.

continued

Recent studies suggest that genetic variation across the genome is organized so that most SNPs are in DNA “blocks” that have been transmitted for many generations without recombination. Each “block” has a distinct set of SNPs, called a “haplotype,” which defines it. Because each “block” has only a few common haplotypes, a few SNPs can be chosen that uniquely identify or tag each of those haplotypes, even though the block may contain many more SNPs. Use of this approach would greatly increase the efficiency with which researchers can narrow the search for DNA variations associated with diabetes. There is sufficient evidence to begin construction of a large-scale haplotype mapping effort for European (Centre d’Etude du Polymorphisme Humain, CEPH), Asian (Japan and China), and African (Nigeria) populations. The goal of the envisioned initiative for a haplotype map is to create a comprehensive resource for human biomedical research, capturing the complete catalogue of the common genome ancestral segments (haplotype blocks) observed in these major human populations. Plans are under way to frame a new NIH initiative in this area based on expert external advice garnered at a recent scientific meeting.

▶ ENHANCING ACCESS TO GENETIC SAMPLES

Many large clinical studies have collected DNA, cell lines, and other samples that could be extremely valuable for genetic studies. To make these samples widely available to investigators, the NIH is establishing central facilities for archival storage of biosamples and data collected in large, multi-site studies and for processing of genetic samples collected in these studies. These

facilities will: (1) meet the storage and genetic sample handling needs of large, multi-site studies funded by the NIH; (2) enhance the value of large studies by increasing access to the biosamples and data that have been collected; and (3) serve as a standard of quality and bioethics as this field evolves. The sample, genetics, and data repositories will enable resources to be shared by the research community and ensure that research participants make a maximal contribution to scientific discovery, while protecting the rights of research participants. The availability of shared resources from repositories will expedite discovery of susceptibility genes for diabetes and its complications and encourage work by new investigators, those with novel approaches, and others not included in current collaborations.

▶ DEVELOPING RESOURCES FOR GENETIC STUDIES OF DIABETES

In order to apply the new genetic information to the study of diabetes and its complications, researchers have suggested the establishment of a National Diabetes Case-Control Collection for collecting, annotating, and making data available to the medical research community. A sample of 5,000 extensively phenotyped diabetes patients and carefully matched controls could be expected to provide sufficient power to detect genes responsible for diabetes. Of particular scientific utility would be a common database to store the genotypic data developed by individual investigators to aid in the identification of interacting genes.

DISCOVERING HOW GENES FUNCTION IN TISSUES AFFECTED BY DIABETES — GENOME ANATOMY PROJECTS

In order to identify all of the genes that may play an important role in diabetes, the NIH has initiated a Diabetes Genome Anatomy Project (DGAP). This project seeks to enable researchers to take a systematic approach to the complex problem of diabetes by cataloging all genes “expressed” or “turned on” in tissues that are relevant to the disease. The initial step in this multi-faceted program was an initiative on the Functional Genomics of the Developing Pancreas. That start-up project brought together scientific experts in pancreas development, DNA sequencing, and bioinformatics to: (1) share research tools and reagents in an accelerated effort to identify all genes involved in beta cell development and function; (2) explore possible roles for these genes in the development of diabetes; and (3) exploit this information for the development of cell-based therapy for diabetes. This project has already identified 500 novel genes that are expressed primarily in the pancreas. These genes are being evaluated using new microarray technology to understand their function in the pancreas and how their expression is altered in diabetes. Microarrays use fluorescent markers to show which cellular genes are actively expressed and making proteins, and permit the simultaneous analysis of thousands of genes. However, this technique requires specialized facilities, hardware, and software. In addition, methods need to be standardized to capture the data and to report it in a standardized format that can be stored and shared among many investigators. Finally, data need to be collected in a common public database to allow comparisons in different experimental systems. Efforts are ongoing to develop a database for the set of pancreas genes that can be expanded to incorporate similar data from other tissues.

Based on the success of this approach, the NIH now plans to expand this effort to other tissues affected by diabetes. The program will provide support for cutting-edge technologies, such as gene expression microarrays and proteomics approaches. Proteomics is the study of how all the proteins of an organism work together and coordinate to perform all the functions necessary for life. The program will also encourage the development of new bioinformatics tools to spur rapid, accurate, and meaningful data analysis. To make comprehensive gene expression technologies widely available to researchers, the NIH also established Biotechnology Centers. These centers are a critical means of providing gene profiling resources to investigators working on diabetes and other diseases.

DISCERNING SIGNATURES OF GENETIC SUBTYPES

What we call type 2 diabetes is most likely a collection of diseases with different underlying genetic causes. Thus, not everyone who develops type 2 diabetes will have the same diabetes susceptibility genes. In order to identify genes that influence the development of diabetes and to understand how they work, it will be necessary to find ways to subclassify patients with diabetes into relatively homogeneous subgroups. If the genes for diabetes differ among subgroups, analyses that do not incorporate subclassification may miss genes because their effects are diluted out by samples from patients with other causes of diabetes. Researchers can now apply many new technologies to discern the “signatures” of different diabetes subtypes. These include microarray studies, nuclear magnetic resonance (NMR)

continued

technology, or other techniques to subdivide patients according to how diabetes manifests itself in specific tissues or organs. One method to subdivide patients is to study the unique group of patients who develop type 2 diabetes in adolescence. Patients who develop diseases earlier are likely to have a higher burden from disease genes; therefore, studies to identify the causative genes in these patients are likely to be more efficient to perform and more successful.

For studies of complications genetics, an important goal is to identify at the earliest possible point those individuals who are developing diabetic complications so that they can be correctly classified with regard to complications risk before the development of overt symptoms. Revolutionary technologies, such as new imaging techniques, can contribute greatly to the development of surrogate markers for complications, which could help researchers attain this goal of early identification. A high priority of the diabetes research community, which resonates at scientific meetings and conferences, is the development of new statistical and bioinformatics approaches to incorporate multiple physiologic (phenotypic) and genetic (genotypic) characteristics of patients for genetic analysis into a single database that would be widely available to investigators.

▶ CAPTURING AND INTEGRATING GENETIC DATA THROUGH BIOINFORMATICS

As advanced genetic technologies generate more extensive amounts of data, a new science is emerging for capturing and integrating these data in accurate and usable ways. For diabetes and obesity, the integration of information on genetic loci in humans with data from studies in mice will be essential. It will also be important to be able to relate these findings with clinical data and to

cross-reference studies in the scientific literature. As additional microarray studies are undertaken, a key scientific objective will be the integration of data on gene expression with data on gene location and with other biologic and clinical parameters. Although some of these types of information sharing and analysis problems have been addressed in business applications, adapting these approaches to biological data will be a major challenge for the next few years. Major opportunities include:

Development of new statistical and computational methods to analyze and evaluate the large amount of genetic data that will be developed using the SNP genome-wide association approach and their application to studies of diabetes and its complications.

Pilot studies to identify methodological and operational issues surrounding the conduct of genome-wide genetic association studies.

Dissemination of information on the ethical and legal issues related to genetic data to aid in its appropriate use in advancing research endeavors.

In summary, since issuance of the DRWG's Strategic Plan, NIH program efforts and research advances have provided many opportunities to understand more fully the underlying genetic factors that make certain individuals more susceptible to type 1 or type 2 diabetes and to diabetic complications. We are now in a position to translate many of these genetic advances into clinical practice by identifying patients at high risk of developing type 1 diabetes and by screening type 2 diabetes patients for genes that already have been implicated in the development of this disease. This knowledge is critical to the development of improved prevention and treatment strategies that could put a halt to or mitigate the effects of these devastating diseases.