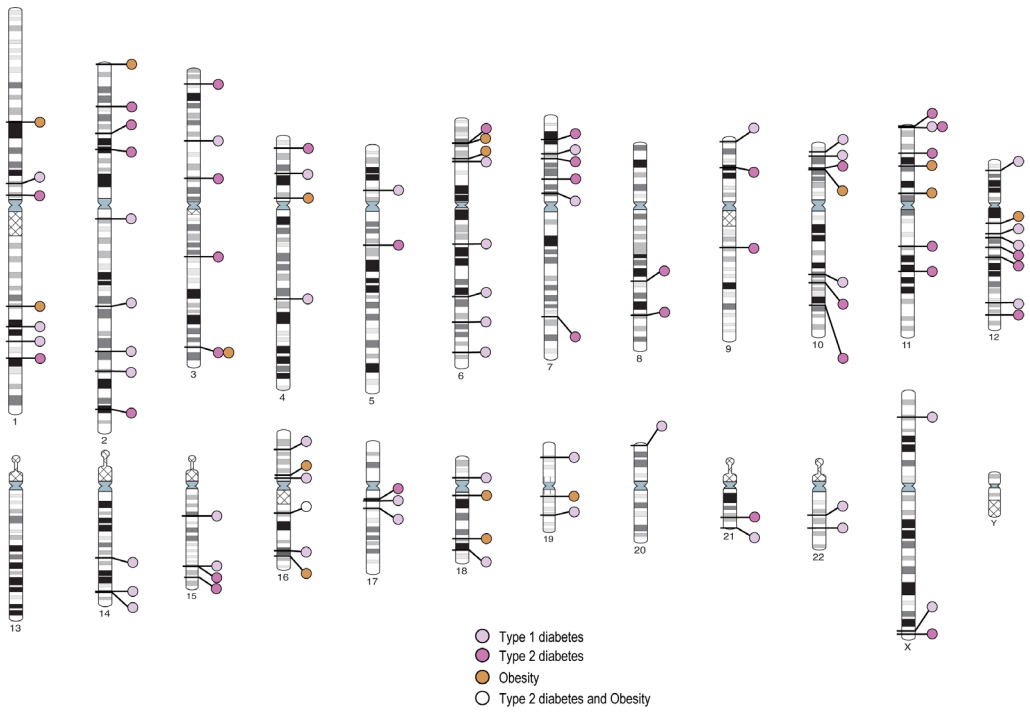


Numerous candidate genes and genetic regions associated with type 1 diabetes, type 2 diabetes, and obesity have been revealed through genome-wide association studies. (*Courtesy of Darryl Leja and Dr. Teri Manolio, www.genome.gov/gwastudies, with modifications.*)

GENETIC BASIS OF TYPE 1 DIABETES, TYPE 2 DIABETES, OBESITY, AND THEIR COMPLICATIONS

contents:

INTRODUCTION

Type 1 diabetes, type 2 diabetes, obesity, and their complications are caused by the joint effects of many genetic, environmental, and behavioral risk factors. Genetic risk variants are present throughout the lifespan, and amenable to study at all times—before, during, and after the development of a disease or disorder. As a result, knowledge of genetic factors offers the potential for prediction, patient stratification, and insights into early precursors of these conditions when preventive therapies might be applied. Such studies contribute to the understanding of disease mechanisms and could also point to potentially modifiable environmental (encompassing behavioral) factors that might affect the initiation and progression of diabetes, obesity, and their complications.

Complex human diseases and disorders, such as diabetes and obesity, are associated with a spectrum of rare and common genetic variants with hundreds of contributing loci in the human genome. Researchers use a technique known as linkage studies to identify causal variants for diseases associated with a single gene—or, as in the case of diabetes and obesity, variants that are relatively rare in the general population but that might have large effects in families that are strongly affected by these serious health conditions (Figure 1). To find more common causal variants that have smaller individual effects on disease susceptibility, researchers turn to association studies (Figure 1). Previously, such studies were limited to the analysis of candidate genes that had been suggested by the biological functions of their gene products. Advances in genotyping and DNA sequencing technologies now permit investigators to search for disease genes throughout the entire human genome in an unbiased manner. Genome-wide association (GWA) studies can be used to search for rare or common susceptibility genes in affected families or in large groups of individuals who do not have a family history of diabetes or obesity. Using these approaches, researchers have identified, to date, nearly 50 regions in the human genome that harbor risk variants for type 1 diabetes (Table 1), 38 for type 2 diabetes (Table 2), and 17 for obesity (Table 3). Much less is known about the genetic loci that influence susceptibility to diabetic complications.

Genetic factors—both known and those yet to be discovered—do not fully explain an individual's susceptibility to type 1 or type 2 diabetes, complications, or obesity, nor can genetics account for the increasing prevalence of these conditions in the last decade. Indeed, the risk of developing these complex medical problems is substantially influenced by environmental factors. Identifying such factors has been difficult due to the complexity and expense of long-term or cross-sectional studies in at-risk populations and the need to collect and store biosamples for future research to retroactively identify biomarkers of disease development. Nonetheless, researchers are exploring multiple, diverse factors that might play a role in diabetes and obesity, including viral infections and early exposure to cow's milk for type 1 diabetes, and reduced physical activity, increased consumption of manufactured and high-calorie foods, and stress for type 2 diabetes and obesity.

A critical unmet research need is the lack of genetic and environmental risk factor data in populations other than those of European origin, particularly those that are disproportionately affected by diabetes and obesity. The majority of candidate genes and causal variants for

diabetes and obesity have been identified in European-origin populations. The few studies of these genes in African American and Hispanic populations suggest that there may be differences in the importance of specific genes or causal variants from those identified in individuals of European origin. Targeted genomic and epidemiologic research is needed in order to better define the genetic regions of association and to reduce the rate of diabetes, obesity, and complications among high-risk minority populations.

Genetic and environmental factors interact in complex ways. For example, some individuals with high-risk variants in a particular gene might be at low overall risk for a disease (equivalent to the risk of the general population) because they carry variants in other loci that confer protection or because they live in a protective or low-risk environment. Environmental factors can also have a direct impact on genetic risk through direct modification of DNA, without change to the sequence, or through modification of DNA-associated proteins, in ways that confer persistent effects on gene expression—a phenomenon known as epigenetics. DNA marking by methylation or other biochemical reactions can occur during development or at any time throughout the lifespan and affect the way the marked gene is expressed even though the inherited gene sequence is not changed.

Exploring how epigenetic modification influences susceptibility to, and progression of, diabetes and obesity opens up the potential for new paradigms for predicting and treating these conditions.

The identification of genetic and environmental factors can lead to insights about disease pathogenesis. Based on the current collection of implicated genetic risk loci, it is obvious that multiple biochemical pathways are involved in diabetes and obesity, but not all of the pathways are likely to affect risk in the same way. Some may be associated with earlier or later ages of onset, slower or faster rates of beta cell mass loss, carbohydrate metabolism, apoptosis, or different patterns of antigen recognition in the autoimmune destruction of islets. The association may not be with diabetes and obesity per se, but rather with an intermediate phenotype or biomarker (autoantibody, metabolic profile, or related measurement), which, if under control of confirmed disease-associated variants, could be an early precursor of the disorder. By revealing which pathways are involved at different stages of diabetes or obesity, these data can help researchers identify new targets for therapeutic development and determine which people are most likely to benefit from specific interventions—an important step toward personalized medicine.

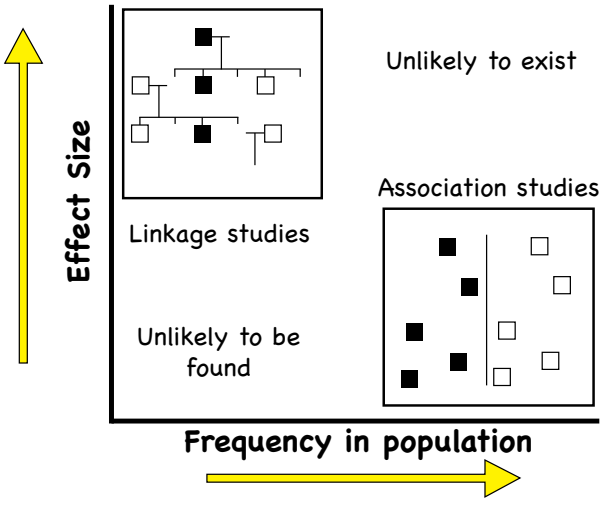


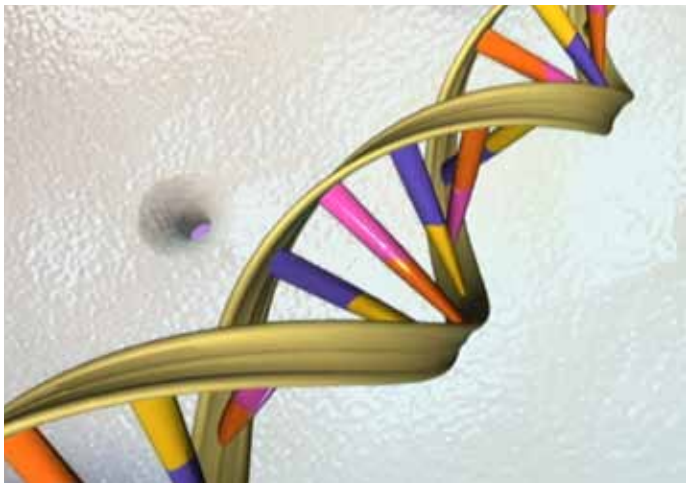
Figure 1. Hypothetical Characterization of Common and Rare Variants Contributing to Common Human Disease and Study Designs for their Discovery.

Low frequency variants with small effect size are unlikely to be found, while there also may be no high frequency variants with large effect size. Association studies are useful for detecting high frequency variants with small effect size, while linkage studies can detect lower frequency variants with large effect size.

Table 1. Type 1 Diabetes Candidate Susceptibility Genes and Putative Function. Genes and gene regions associated with type 1 diabetes that have been identified and/or confirmed through GWA studies* are listed here.

Chromosome	Gene (suggested)	Gene Name	Possible Disease Mechanism
1	*PTPN22*	protein tyrosine phosphatase, non-receptor type 22	Negative regulator of T cell activation
1	*RGS1*	regulator of G-protein signaling 1	Regulator of G protein signaling
1	*IL10*	interleukin 10	Pleiotropic effects in immunoregulation and inflammation
2	*IL18RAP*	interleukin 18 receptor accessory protein	Enhances IL18 binding activity of IL18R1
2	*IFIH1*	interferon-induced helicase	Receptor for dsRNA from viral infections
2	*CTLA4*	cytotoxic T lymphocyte antigen 4	Inhibitory signal to T cells
2	*STAT4*	signal transducer and activator of transcription 4	Transcription factor for IL12, IL23, and IFN
3	*CCR5*	chemokine (C-C motif) receptor 5	Chemokine receptor
4	-		Unknown
4	*IL2*	interleukin 2	Proliferation of T and B lymphocytes
5	*IL7R*	interleukin 7 receptor	Regulation of lymphopoiesis
6	*HLA-A,* *HLA-B,* *HLA-DRB1,* *HLA-DQB1*	major histocompatibility complex, class I, A major histocompatibility complex, class I, B major histocompatibility complex, class II, DR beta 1 major histocompatibility complex, class II, DQ beta 1	Antigen presentation
6	*BACH2*	BTB and CNC homology 1, basic leucine zipper transcription factor 2	Maintain IL-2 production
6	*C6orf173*	centromere protein W	Transcription regulator family
6	*TNFAIP3*	tumor necrosis factor, alpha-induced protein 3	Inhibit NF-kappa B activation as well as TNF-mediated apoptosis
6	*TAGAP*	T-cell activation RhoGTPase activating protein	T cell activation
7	*SKAP2*	src kinase associated phosphoprotein 2	Essential role in the src signaling pathway
7	*COBL*	cordon-bleu homolog (mouse)	Central nervous system development
9	*GLIS3*	GLIS family zinc finger 3	Development of pancreatic beta cells
10	*IL2RA*	interleukin-2 receptor alpha chain	T cell activation
10	*PRKCQ*	protein kinase C, theta	T cell activation
10	*RNLS*	renalase, FAD-dependent amine oxidase	Unknown
11	*INS*	insulin gene	Involved in glucose uptake
12	*CD69*	CD69 molecule	Role in proliferation, transmit signals in natural killer cells and platelets
12	*ERBB3*	v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)	Role in cell proliferation or differentiation
12	*KIF5A*	kinesin family member 5A	Intracellular organelle transport
12	*SH2B3*	SH2B adaptor protein 3	T cell activation
14	*DLK1*	delta-like 1 homolog (*Drosophila*)	Unknown
14	*C14orf181*	chromosome 14 open reading frame 181	Unknown
14	-		Unknown
15	*CTSH*	cathepsin H	Degradation of lysosomal proteins
15	*RASGRP1*	RAS guanyl releasing protein 1 (calcium and DAG-regulated)	T cell receptor signalling
16	*CLEC16A*	C-type lectin domain family 16, member A	Possible effects on T cell helper function
16	*IL27*	interleukin 27	CD4+ T cell differentiation
16	*CTRB2*	chymotrypsinogen B2	Involved in digestion and proteolysis
17	*ORMDL3*	ORM1-like 3 (*S. cerevisiae*)	Involved in protein folding
17	*SMARCE1*	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily e, member 1	Critical modulator of the androgen receptor
18	*PTPN2*	protein tyrosine phosphatase, non-receptor type 2	Regulates cell growth, differentiation, mitotic cycle, and oncogenic transformation
18	*CD226*	CD226 molecule	Positive regulation of Ig-mediated immune response, mast cell activation, NK T cell-mediated cytotoxicity
19	*ICAM1, 4, 5*	intercellular adhesion molecule 1, intercellular adhesion molecule 4 (Landsteiner-Wiener blood group), intercellular adhesion molecule 5, telencephalin	Involved in adhesion
19	*PRKD2*	protein kinase D2	Regulates protein kinase C activity
20	*SIRPG*	signal-regulatory protein gamma	Negative regulation of receptor tyrosine kinase-coupled signaling processes
21	*UBASH3A*	ubiquitin associated and SH3 domain containing A	Negative regulator of T cell activation
22	-		Unknown
22	*C1QTNF6*	C1q and tumor necrosis factor related protein 6	Unknown
X	*TLR8*	toll-like receptor 8	Single-stranded RNA recognition
X	*GAB3*	GRB2-associated binding protein 3	Facilitate macrophage differentiation

*As of June 2010



DNA double helix. The four basic units of DNA, or bases, termed "A," "T," "C," and "G," form sequences of base pairs indicated in blue-yellow and pink-orange. The human genome contains 3 billion base pairs. Variations in the sequences of these base pairs is at the foundation of human diversity—including differences in susceptibility to diabetes, obesity, and their complications. *(Image credit: National Human Genome Research Institute.)*

Knowledge of gene variants and their interactions with environmental risk factors can also inform strategies to predict which individuals are most likely to develop disease over time. In diseases such as type 1 diabetes for which there are no proven prevention methods, identifying genetically at-risk children can lead to earlier detection of disease onset and reduction of the risk of adverse events like diabetic ketoacidosis, which can be fatal if undiagnosed and untreated (see the "Special Needs for Special Populations" chapter). Identification of at-risk children is also essential for research studies testing approaches to prevent disease or to identify environmental triggers. In type 2 diabetes and obesity, the development of prediction models based on genetic, biochemical, and environmental factors can be used to identify high-risk individuals who would most benefit from lifestyle or pharmacologic interventions that have been shown to prevent or delay their onset. With several proven prevention approaches for type 2 diabetes, genetic risk factors may allow individual tailoring of preventive treatments.

As causal genes, their risk variants, and environmental triggers of diabetes, diabetic complications, and obesity are identified and characterized, their functional utility will become clearer. This chapter on the "Genetic Basis of Type 1 Diabetes, Type 2 Diabetes, Obesity, and Their Complications" describes how state-of-the-art technologies and resources can be applied to accelerate research on prediction, prevention, and treatment of these complex conditions.

Table 2. Type 2 Diabetes Candidate Susceptibility Genes and Putative Function. Genes and gene regions associated with type 2 diabetes that have been identified and/or confirmed through GWA studies* are listed here.

Chromosome	Gene (suggested)	Gene Name	Possible Disease Mechanism
1	*NOTCH2*	Notch 2	Beta cell development
1	*PROX1*	prospero homeobox 1	Beta cell development
2	*THADA*	thyroid adenoma associated	Unknown
2	*IRS1*	insulin receptor substrate 1	Insulin signaling
2	*GCKR*	glucokinase (hexokinase 4) regulator	Glucose metabolism
2	*BCL11A*	B-cell CLL/lymphoma 11A (zinc finger protein)	DNA binding
3	*PPARG*	peroxisome proliferator-activated receptor gamma	Insulin sensitivity
3	*ADAMTS9*	ADAM metallopeptidase with thrombospondin type 1 motif, 9	Metalloprotease
3	*IGF2BP2*	insulin-like growth factor 2 mRNA binding protein 2	Growth factor regulation
3	*ADCY5*	adenylate cyclase 5	Insulin secretion
4	*WFS1*	Wolfram syndrome 1 (wolframin)	Beta cell function
5	*ZBED3*	zinc finger, BED-type containing 3	DNA-binding
6	*CDKAL1*	CDK5 regulatory subunit associated protein 1-like 1	Cell cycle regulation
7	*JAZF1*	JAZF zinc finger 1	Transcription factor
7	*KLF14*	Kruppel-like factor 14	Transcription factor
7	*GCK*	glucokinase (hexokinase 4)	Glucose metabolism
7	*DGKB/TMEM195*	diacylglycerol kinase, beta 90kDa/ transmembrane protein 195	Insulin signaling
8	*SLC30A8*	solute carrier family 30 (zinc transporter), member 8	Beta cell zinc transporter
8	*TP53INP1*	tumor protein p53 inducible nuclear protein 1	Unknown
9	*CDKN2A/2B*	cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)/ cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)	Cell cycle regulation
9	*CHCHD9*	coiled-coil-helix-coiled-coil-helix domain containing 9	Unknown
10	*CDC123/CAMK1D*	cell division cycle 123 homolog (*S. cerevisiae*)/ calcium/calmodulin-dependent protein kinase ID	Cell cycle regulation
10	*HHEX/IDE*	hematopoietically expressed homeobox/ insulin-degrading enzyme	Transcription factor/insulin regulation
10	*TCF7L2*	transcription factor 7-like 2	Beta cell transcription factor
11	*KCNQ1*	potassium voltage-gated channel, KQT-like subfamily, member 1	Beta cell potassium channel
11	*KCNJ11*	potassium inwardly-rectifying channel, subfamily J, member 11	Beta cell potassium channel
11	*MTNR1B*	melatonin receptor 1B	Circadian regulation/insulin secretion
11	*CTCF-binding site*	CCCTC-binding factor (zinc finger protein) binding site	Imprinting
11	*CENTD2/ARAP1*	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 1	Unknown
12	*TSPAN8/LGR5*	tetraspanin 8/ leucine-rich repeat-containing G protein-coupled receptor 5	Unknown
12	*HMGA2*	high mobility group AT-hook 2	Unknown
12	*HNF1A*	HNF1 homeobox A	Transcription factor
15	*ZFAND6*	zinc finger, AN1-type domain 6	DNA binding
15	*PRC1*	protein regulator of cytokinesis 1	Cell cycle
16	*FTO*	fat mass and obesity associated	Altered BMI
17	*TCF2/HNF1B*	HNF1 homeobox B	Beta cell transcription factor
21	*KCNJ15*	potassium inwardly-rectifying channel, subfamily J, member 15	Beta cell potassium channel
X	*DUSP9*	dual specificity phosphatase 9	Signaling

*As of June 2010

RECENT RESEARCH ADVANCES

Recent advances in molecular, statistical, and computational technologies have rapidly increased knowledge of the genetic contributions to common multi-factorial diseases and disorders, including type 1 diabetes, type 2 diabetes, and obesity. Major examples of these advances are described below.

Human Genome Sequence, HapMap, and Genome Structure: The Human Genome Project, a 13-year effort coordinated by the U.S. Department of Energy, the NIH, and the Wellcome Trust, was completed in 2003. The Project goals were to identify all genes in human DNA, determine the sequences of the 3 billion base pairs (bp) that make up human DNA, store this information in databases, improve tools for data analysis, transfer related technologies to the private sector, and address the ethical, legal, and social issues that might arise from the project.

Sequencing the human genome permitted the construction of a more detailed and user-friendly map of genetic variation. The resulting HapMap, a catalog of common genetic variants that occur in humans, describes the type and location of the variants, their size and distribution among and within populations, and their frequency in different populations. The HapMap is based on the fact that genetic sequences between individuals differ at about one in every 1,200 bases, resulting in a genetic variant (also referred to as a marker or polymorphism). The HapMap assists researchers in mapping genetic variants for diseases or phenotypes in a way that can provide increased genomic coverage and pinpoint more discrete areas for follow-up study. Because genetic variants that are near each other tend to be inherited together (forming haplotypes), a small number of haplotypes can account for the majority of the common genetic variation. The International HapMap Project has identified common haplotypes in four populations as well as "tag" single nucleotide polymorphisms (SNPs) that uniquely identify these haplotypes. The methods now employed for GWA studies use these tag SNPs, which represent a subset of less than 1 million—far fewer than the 10 million common SNPs in the genome—thus avoiding redundant genotyping of millions of common SNPs in genetic studies.

In addition to single nucleotide polymorphisms, every individual's genome has regions that are inserted, deleted, inverted, and duplicated. These rearrangements are in the class of structural variants called copy number variants (CNV). It is estimated that approximately 8 percent of individuals have a 500,000 bp deletion in their genome, a span that is large enough to contain multiple genes. Specific regions in the

Genes play a central role in risk for developing diabetes, obesity, or complications—and can also affect how well a person responds to therapy. Solving the puzzle of how genetic factors interact within cells and with the environment in ways that promote or protect against disease is a critical goal for diabetes research. *(Image credit: Erwin Solbach, The Scientific Consulting Group, Inc.)*

genome contain duplicated sequences that are prone to recombination and are polymorphic in the population. These regions have been cataloged so they can be studied for their contribution to disease.

Structural variation and its effects on human disease represent a recent development in human genetics research, although large structural changes have long been recognized cytogenetically in rare syndromes. Recent studies suggest that CNV loci are widespread in the human genome and highly variable in size and frequency, with smaller CNVs being more frequent than larger ones. Current estimates, based on available re-sequenced genomes, suggest that each diploid human genome harbors approximately 3,000 CNVs greater than 100 bp, but these estimates are being refined. Two ongoing projects are expanding knowledge of human genome structural variation. The Genome Structural Variation (GSV) consortium (*www.sanger.ac.uk/humgen/cnv/42mio/*) has designed a high-density tiling array across the entire genome to type HapMap samples from 20 participants of Caucasian ancestry from Utah and 20 participants of African ancestry from the Yoruba of West Africa. In parallel, the 1000 Genomes Project (*www.1000genomes.org*) is rapidly expanding the catalog of common CNVs, thus providing another source for CNV discovery and assessment of their linkage disequilibrium (non-random association) with flanking SNPs and small insertion/deletions (indels). Current estimates suggest that dense DNA re-sequencing allows calling of approximately 2,000 deletions greater than 150 bp per densely sequenced diploid genome, in addition to 500-1,000 duplications.

Genome Architecture: In addition to providing a new understanding of gene sequences, the Human Genome Project has elucidated many other features of the genome that regulate gene expression. The genome contains sequence elements that serve as regulatory regions and bind both enhancers and repressors of gene expression, many times in overlapping binding sites so that the binding of an enhancer will interfere with the binding of a repressor. Regions have been identified that insulate genes from the regulatory signals of neighboring genes. Other regions direct the binding of histones that form higher-order chromatin structure. One of the most significant recent findings was the discovery of microRNAs (miRNAs) that are encoded in the genome and are important for coordinating the expression of genes for cellular developmental programs.

High-Throughput Genotyping: The Human Genome Project and the International HapMap Project provided a mechanism for characterizing the variation and structure in the human genome that could be used for detecting association between genotype and disease or risk phenotype. Realizing the full potential of these new resources required the development of high-throughput genotyping, a major technological advance. Prior to the HapMap, most genetic studies were performed by genotyping a single (or few) variant on individual samples that would often cost $1/genotype/sample. A robustly powered study involved thousands of samples and at least 100,000 SNPs had to be genotyped to provide genome-wide coverage. The development of high-throughput genotyping technologies not only dramatically reduced the cost of genotyping (to under $0.01/genotype/sample) but also reduced the time for completing the experiment to a few weeks. Thus, the same project that previously would cost $500 million and require a year to complete could be accomplished (with 1 million SNPs) for $2.5 million in 3 months time.

Bioinformatics and Sharing Resources: The volume of data from the Human Genome Project, the International HapMap Project, the 1000 Genomes Project, and individual laboratories has been growing at almost exponential rates over the past 5 years.

The National Center for Biotechnology Information established the Single Nucleotide Polymorphism database (dbSNP) in 1998 (*www.ncbi.nlm.nih.gov/projects/SNP/*) to provide an easily accessible, curated catalog of human genetic variation for use in ongoing research. dbSNP serves as a central public repository for genetic variation and classifies SNPs, insertion/deletions, invariant regions, microsatellite repeats, named variants, and other uncharacterized heterozygous assays. The entries in dbSNP include disease-causing clinical mutations, as well as non-functional (neutral) polymorphisms. The database of Genotypes and Phenotypes (dbGaP) was developed to archive and distribute the results of studies that have investigated the interaction of genotype and phenotype (*www.ncbi.nlm.nih.gov/gap*). dbGaP provides a repository of genetic results from published studies, including GWA studies, medical sequencing, molecular diagnostic assays, and studies of association between genotype and non-clinical traits.

The NIH also capitalized on major clinical studies to develop shared resources over the past several years that could accelerate genetic discovery in diabetes and obesity. For example, to expand the usefulness of clinical studies by providing access to the biosamples and data to a wider research community, the NIDDK established three Central Repositories for biosamples, data, and genetic information collected in these studies. Materials are available or expected to be reposited from the Diabetes Control and Complications Trial (DCCT)/Epidemiology of Diabetes Interventions and Complications Study (EDIC), Genetics of Kidneys in Diabetes Study (GoKinD), Family Investigation of Nephropathy in Diabetes (FIND), Type 1 Diabetes Genetics Consortium (T1DGC) (which also provides summary association results for type 1 diabetes at *www.t1dbase.org*), and other studies. Other NIH clinical trials and studies, such as the NHLBI-led Framingham Heart Study, have also provided shared resources important for genetic studies.

Resequencing, the Human Exome Project, and the 1000 Genomes Project: The International HapMap Project has catalogued more than 1 million SNPs, most of which do not directly influence gene function, in the genome sequences of 269 people drawn from four diverse human populations. Once a candidate gene has been identified, a comprehensive approach to discovery of disease-causing variants is "population resequencing." In this approach, candidate genes are totally or partially sequenced, and the frequencies of variants are compared between those with and without disease and controls to identify genetic variations that may contribute to disease. One aspect of medical resequencing is developing a catalog of common and rare variation within genes in selected disease entities, such as diabetes and obesity. The Human Exome Project is expanding the content of the catalog by discovering genetic variation in all genes in the general population—not just patients and controls—that can potentially affect gene function directly. The 1000 Genomes Project represents an international consortium to create the most detailed map yet of human genetic variation by sequencing the entire genomes of approximately 2,000 individuals. Interrogation of the 1000 Genomes Project database will provide a means for identifying the full list of SNPs and short insertions/deletions in regions associated with complex human disease, whether they are coding, intronic, or intergenic. Together, the data from the Medical Resequencing Project, the Human Exome Project, and the 1000 Genomes Project are providing resources previously unimagined for genetic research in diabetes, obesity, and related phenotypes.

GWA Studies and Statistical Developments:

Analysis of the large volume of genetic marker data being generated required improvements in computing and analytic methods. With the number of SNPs approaching 1 million on each individual, strict criteria were required to achieve genome-wide levels of statistical significance. In addition, the increased depth of genetic information permitted more direct evaluation of a primary concern of GWA studies—the effect of bias due to unrecognized population stratification. With the advent of the HapMap and high-throughput genotyping, estimates of admixture and population structure could be made. The results of these advances in detecting and correcting for population heterogeneity improved disease gene detection at an analytic level over the genome.

Several approaches to analysis of genome-wide information have been used. The analysis of single SNPs and structural variants in a single population has been extended in order to pool information across multiple studies or across multiple strata within a single study. Although 1 million SNPs provides excellent coverage of the human genome, gaps remain that could be important. In order to assess the evidence of a candidate region existing in "un-genotyped" regions, an alternative approach to genotyping was developed in recognition that the HapMap contains the relevant information to permit "imputation" on the basis of the candidate region being near a SNP that is in the same haplotype block. Thus, the analytic approach of imputation permitted a statistical approach to provide probabilistic genotypes in regions that are not genotyped in the study, often increasing the information in the genome from less than 500,000 genotyped markers to over 2 million genotyped and imputed markers.

The genetic advances outlined in this chapter greatly increased understanding of the genome sequence, human variation, genotyping, and analytic approaches that have culminated in the discovery of a large number of candidate loci for type 1 diabetes, type 2 diabetes, obesity, and complications as detailed below:

Type 1 Diabetes: Ten years ago, only three gene loci had been identified for type 1 diabetes and replicated in many studies: *HLA*, *INS*, and *CTLA4**. Approximately 40 to 50 percent of the risk for type 1 diabetes can be attributed to alleles of the *HLA* class II loci in the major histocompatibility locus (MHC). The VNTR (variable number of tandem repeats) in the promoter of the insulin gene (*INS*) had also been shown to be associated with type 1 diabetes. This common variant is thought to influence the expression of insulin in the thymus, thereby affecting the ability of T cells to recognize this protein. The third locus, *CTLA4*, which encodes an inhibitor of T cell signaling, was identified using a functional candidate approach.

More recently, candidate gene studies identified two additional loci, *PTPN22* and *IL2RA*, which also modulate T cell activity. The *CTLA4*, *IL2RA*, and *PTPN22* genes have also been shown to be associated with other autoimmune diseases. The application of genome-wide SNP typing technology to large sample sets and comparisons with results from other diseases have identified new loci for type 1 diabetes, including *IFIH1*. The Wellcome Trust Case Control Consortium (WTCCC) studied 2,000 cases from the JDRF/WT British case collection and 2,500 controls from the British 1958 Birth Cohort (B58BC). In addition, a GWA study conducted on the type 1 diabetes cases from the GoKinD study of diabetic nephropathy added power to

* Names of genes/loci associated with type 1 diabetes, type 2 diabetes, and obesity are provided in the Tables.

detect additional loci. The T1DGC studied an additional 4,000 cases from the JDRF/WT British case collection and 2,500 controls from the B58BC. Combining results from these studies and replicating these in further case-control sample sets from Great Britain, Denmark, and the T1DGC families has confirmed nearly 50 genes/loci for type 1 diabetes at genome-wide significance (Table 1). Most of the newly-identified genes seem to affect T cell function and the immune system.

Type 2 Diabetes: Although many genes have been identified for rare Mendelian forms of diabetes, no genes were known for the most common form of the disease until 10 years ago. Association studies looking at candidate genes identified two genes associated with type 2 diabetes: *PPARG* and *KCNJ11* (*kir 6.2*). Both genes were selected as candidates because they were targets for diabetes drugs. A SNP genotyping project following linkage mapping studies found an association with *TCF7L2*, a gene coding for a transcription factor. Although variants in this gene confer a 40 percent increased risk of developing diabetes, it is not clear how this factor contributes to the pathophysiology of diabetes. Five groups around the world conducting GWA studies in European-origin populations identified an additional 12 genes for type 2 diabetes. A similar GWA study in Japan identified yet another gene, *KCNQ1*, which encodes a potassium channel subunit. Replication of the top associations from these studies in additional individuals, plus meta-analysis of data from multiple GWA studies, has increased the number of confirmed type 2 diabetes genes/regions to nearly 40 (Table 2). The candidate genes identified to date have modest effects on the disease, and many of them seem to affect the ability of the beta cell to secrete insulin.

Obesity: Several genes have been reported to be associated with obesity. GWA studies of individuals with type 2 diabetes identified the gene *FTO*, but its effect on diabetes was mediated by body mass index (BMI). In addition, the gene *MC4R*, which causes a monogenic form of obesity, has also been associated with common obesity. The Genomewide Investigation of ANThropometric measures (GIANT) consortium combined results from 15 GWA studies for BMI associations in over 32,000 individuals and identified 17 genes that affect BMI (Table 3). Many of the genes, including *MC4R*, are known to act in the central nervous system, consistent with a role in regulating eating behavior and body weight.

Complications: Familial clustering of diabetic kidney disease has been identified and replicated in multiple studies using sib-pair and cohort designs. GWA studies have been completed on type 1 diabetes cohorts from the EDIC and GoKinD studies. In addition, the FIND study has used an admixture approach and a GWA study to identify genes predisposing to kidney complications. Several candidate genes have been suggested by these association studies, including those encoding angiotensin-converting enzyme (*ACE*), FERM domain containing 3 (*FRMD3*), cysteinyl-tRNA synthetase (*CAR*S), Carnosinase, engulfment and cell motility protein 1 (*ELMO1*), superoxide dismutase 1, soluble (*SOD1*), and vascular endothelial growth factor (*VEGFA*), but no single locus has been convincingly established.

Gene Expression Using Microarrays and Expression Quantitative Trait Loci Analysis:

DNA microarrays can be used to measure changes in gene expression levels, to detect SNPs, or to sequence genomes. In typical gene mapping approaches, a gene (locus) is detected that contributes to variation in a quantitative trait, such as fasting glucose or BMI—thus termed a quantitative trait locus (QTL). In contrast to traditional QTL mapping, expression QTL (eQTL)

Table 3. Obesity Candidate Susceptibility Genes and Putative Function. Genes and gene regions associated with obesity that have been identified and/or confirmed through GWA studies* are listed here.

Chromosome	Gene (suggested)	Gene Name	Possible Disease Mechanism
1	*NEGR1*	neuronal growth regulator 1	Neuronal outgrowth
1	*SEC16B/RASAL2*	SEC16 homolog B (*S. cerevisiae*)/ RAS protein activator like 2	Unknown
2	*TMEM18*	transmembrane protein 18	Neuronal development
3	*ETV5*	ets variant 5	Unknown
4	*GNPDA2*	glucosamine-6-phosphate deaminase 2	Unknown
6	*PRL*	prolactin	Hormone regulation
6	*NCR3, AIF1, BAT2*	natural cytotoxicity triggering receptor 3, allograft inflammatory factor 1, HLA-B associated transcript 2	Unknown
10	*PTER*	phosphotriesterase related	Unknown
11	*BDNF*	brain-derived neurotrophic factor	Regulated by nutritional state
11	*MTCH2*	mitochondrial carrier homolog 2 (*C. elegans*)	Cellular apoptosis
12	*FAIM2*	Fas apoptotic inhibitory molecule 2	Adipocyte apoptosis
16	*SH2B1*	SH2B adaptor protein 1	Neuronal role in energy homeostasis
16	*MAF*	v-maf musculoaponeurotic fibrosarcoma oncogene homolog (avian)	Adipogenesis and insulin-glucagon regulation
16	*FTO*	fat mass and obesity associated	Altered BMI
18	*NPC1*	Niemann-Pick disease, type C1	Intracellular lipid transport
18	*MC4R*	melanocortin 4 receptor	Hypothalamic signaling
19	*KCTD15*	potassium channel tetramerisation domain containing 15	Unknown

*As of June 2010

Effect of epigenetic changes. This image shows mice that have an identical DNA sequence in a gene that both determines the color of their fur and, when not properly regulated, also promotes obesity. Their different coat colors and body size arise from variation in the chemical modification (methylation) of this gene. Therefore, even though the mice have the same DNA sequence, these epigenetic marks have a dramatic effect on their physical appearance. Moreover, the degree of this chemical modification can be influenced during development by factors such as maternal diet. Regulation of gene activity by epigenetics plays an important role in diabetes, obesity, and their complications. *(Image courtesy of Dr. Robert A. Waterland and reprinted from Journal of Pediatrics, 149, Waterland RA, Epigenetic mechanisms and gastrointestinal development, S137-S142, Copyright 2006, with permission from Elsevier.)*

mapping identifies regions of the genome that contribute to variation in gene expression. In eQTL studies, thousands of expression phenotypes are characterized on microarrays and, as a result, thousands of QTLs will be proposed. A major feature of eQTL analysis, therefore, is to study the relationship between the variation in the genome contributing to the quantitative trait and the expression of a gene (the transcriptome). In principle, a massive amount of genome annotation (QTL analysis) can be superimposed onto eQTL information for detection of genetic contribution to expression that may be near (cis) or distant (trans) to the functional gene. Gene expression microarrays have been used in numerous applications, including identifying novel genes associated with certain cancers, classifying tumors, and predicting patient outcome, although much more work needs to be conducted before these techniques can be used clinically. These techniques have begun to be applied to diabetes and obesity and, in the future, could yield clinically relevant information.

Identification and Treatment of Neonatal Diabetes Mellitus: Neonatal diabetes mellitus (NDM) is a rare genetic condition that usually occurs in the first 6 months of life. Most people with NDM are mistakenly identified as having type 1 diabetes and treated with insulin injections. Research has allowed attribution of NDM in about 50 percent of individuals to dominant mutations in *SUR1*, *KCNJ11*, or *INS*, or recessive mutations at the gene encoding the enzyme glucokinase, *GCK*. The *SUR1* and *KCNJ11* genes encode the two protein subunits of a potassium ion channel that regulates insulin secretion; the mutations prevent the normal release of insulin from pancreatic beta cells. Genetic testing can be used to identify people who have these mutations and is now recommended for all diabetes diagnosed before the age of 6 months. Recent studies have shown that many of these infants' diabetes can be managed with the oral drug sulfonylurea. Not only is this drug less burdensome than insulin therapy, which requires injections and monitoring of blood glucose, but it also results in improved control of diabetes and less hypoglycemia in children with NDM.

KEY QUESTIONS AND FUTURE DIRECTIONS FOR RESEARCH

The last decade has seen major advances in knowledge of the human genome, the development of technologies to probe the genome in a rapid, unbiased, and detailed manner, and the identification of many, though by no means all, causal genes and environmental determinants of diabetes, complications, and obesity. This section builds on these findings by defining key questions and future research directions that aim to identify all causal genes/variants and environmental determinants, understand how these factors interact to cause onset or progression of these conditions, and use this knowledge to better predict, prevent, and treat diabetes, diabetic complications, and obesity.

Genes and Pathways

Although great progress has been made in discovery of genomic regions that contribute to risk of type 1 diabetes, type 2 diabetes, and obesity, relatively few causal genes have been definitively identified. Moreover, a catalog of causal variants in those genes has not been developed, particularly in populations not of European origin. As researchers continue to discover the candidate genes and causal variants in the regions of interest, it is becoming clear that evaluation of single genes is unlikely to explain the complex biology of diabetes. To address these issues, an integrated approach is needed to identify the sets of genes and gene products that are involved in the biochemical pathways related to diabetes. These pathways could include "known" pathways, such as those involved in immune response, apoptosis, insulin signaling, and insulin action, or might require discovery of previously unrecognized pathways. The resulting information can be used to create a highly annotated and interactive public database that is specific to diabetes- and obesity-related phenotypes. In this fashion, the genetic architecture of diabetes and obesity at the DNA level can be correlated with the function of genes/variants at the expression and protein levels. This information would provide an "anatomy" of diabetes and obesity genetic susceptibility.

Key Questions

- **What are the causal genes and variants influencing or residing within each candidate susceptibility locus?**
- **Are the candidate genes/regions identified in European-origin populations (where most of the studies have been performed) also operative in other, ethnically diverse populations?**
- **Do candidate genes/risk variants interact to modify risk, and how is the penetrance of disease alleles affected by environmental factors?**
- **Are there subsets of genes that, taken together, represent a causal pathway that could define a therapeutic target?**
- **What are the effects of identified genetic variants and the integration of genomic, expression, and proteomic profiling on disease risk?**
- **Can genetic variation be coupled with gene expression profiles at the RNA and protein levels to catalog target tissues at the**

population, individual, and cellular levels for both humans and animal models?

- **Can model organisms be utilized to advance research from human genetic studies, and can results from model organisms direct targeted human studies?**

Future Directions

- ➤ **Develop standardized and emergent protocols for assessing phenotypic characteristics of populations, both clinical and epidemiologic, for use in genetic studies.**
- ➤ **Understand how candidate genes contribute to disease risk.**
- ➤ **Elucidate the interactions among genes at the cellular level and discover common pathways of risk.**

Numerous attempts have been made to link various genes and environmental factors to diabetes and obesity risk, in order to identify a genetic network or pathway that could be specifically involved in susceptibility. Whether specific subsets of genes and causal variants or pathways of genetic risk factors are important and whether those interact with specific environmental factors in a population have yet to be determined. Addressing these questions requires new and expanded resources and research tools. Studies of large samples will enable researchers to conduct better-powered genomic analysis of targeted candidate genes. Similarly, analysis of diverse populations from multiple ethnic groups with extensive biospecimens and DNA collections will help determine whether genes and pathways identified largely in populations of European origin are broadly applicable. This line of research would be accelerated by ready access to state-of-the-art technologies, such as high-throughput, low-cost DNA sequencing, expression profiling using microarray technologies and direct DNA sequencing, highly sensitive immunoassays, improved mass spectrometry approaches, such as stable isotope labeling with amino acids in cell culture (SILAC), and the development of accurate animal models with tissue-specific manipulation of variants in candidate genes. High-performance computing and mathematical modeling approaches are needed to conduct pathway analyses that incorporate multiple sources of data. (See also the chapter on "Resource and Infrastructure Needs for Diabetes Research.")

Detection of Rare Variants

To fully characterize genes associated with diabetes and obesity, large-scale DNA sequencing technologies need to be employed to make human sequencing a tool for both research and medical practice. Human DNA sequencing in large populations would enable identification of both common and rare variants that reside in genes and in intragenic regions. Identifying variation and determining how it influences risk of diabetes and obesity must be developed at both the molecular and the analytic level. The normal range of human variation present in populations may be extensive, and disease may manifest only in the presence of specific environmental or behavioral factors. Thus, characterization of human sequence variation provides a basis for understanding how specific environmental exposures can result in disease. Sequence-based strategies need to be efficiently scaled with developing technology to be complementary with efforts to obtain increasingly precise and reproducible phenotypic data. As more is learned from sequencing about the genomic contribution to diabetes, obesity, and complications, and as the cost of obtaining sequence information decreases, these data will become increasingly important for estimating future disease risk, improving prevention

and diagnostic tools, and treating disease, including prevention of complications.

Key Questions

- **How can sequence variation that is rare in populations, yet accounts for familial risk of disease, be identified?**
- **Can genomic sequence data from many individuals with known phenotypes provide insight into the effect that natural variation in genome structure has on susceptibility to diabetes and obesity?**
- **What is "normal" sequence variation compared to "risk" variation in the context of environmental triggers that lead to diabetes and obesity?**
- **Can population-specific DNA sequences be identified that are associated with disease risk and that are predictive of response to therapies?**

Future Directions

- ➤ **Perform DNA sequencing in tens of thousands of participants with type 1 diabetes, type 2 diabetes, and obesity to detect all sequence variants that may be associated with risk of these conditions.**
- ➤ **Correlate sequence variants with the level of risk for development of diabetes, obesity, and their complications.**

Distinguishing normal and risk sequence variation demands high-throughput, cost effective resources and accessible tools, including: catalogs of sequence variation at all disease-related gene sites for use in comparison across samples and studies; detailed biosample and DNA collections from large numbers of individuals for the characterization of sequence variation in multi-ethnic populations; access to high-throughput, low-cost DNA sequencing using "next generation" technologies; and Web-based data repositories and bioinformatics and biostatistics tools that can incorporate common and rare sequence variants with existing epidemiological data to model disease risk. The current volume of DNA sequence data has been limited and restricted to "normal variation."

Gene-Environment Interactions

Diabetes and obesity are complex human traits that result from both genetic and environmental factors. One factor in type 1 diabetes is thought to be viral infections that, in genetically susceptible individuals, trigger an innate autoimmune process, leading to destruction of the insulin-producing pancreatic beta cells. In type 2 diabetes and obesity, behavioral factors contribute to weight gain, resulting in progressively dysfunctional metabolism of glucose or fat in genetically susceptible individuals. Early functional studies of the *FTO* gene indicate that mutations/variations in this gene alter behavior in a way that can lead to obesity. Thus, the interaction of genetic factors with environmental factors likely explains much of the risk for diabetes and obesity. Diabetic complications also have strong genetic and environmental components. For example, while predisposition to diabetic kidney disease is clearly heritable, poor glycemic control is known to increase the risk of this and other complications. Importantly, the risk factors that trigger disease could be context-specific, meaning that an individual with a genetic variant would also need to be exposed to a specific environmental trigger for the greatest risk to occur. Identification of these interacting partners should provide important insights for risk prevention and early detection of disease.

Key Questions

- **What kinds of sample and data resources are needed for analyzing genetic variation in groups of participants with diabetes and obesity or in healthy populations before they develop complications, so that environmental triggers can be identified in those at high genetic risk?**
- **How and to what extent will information be collected on environmental triggers, especially unknown/potential triggers for diabetes, obesity, and their complications?**
- **Can research tools used in mouse models of disease be used to identify potential modifier effects in human genetic data?**
- **How can genetically determined epidemiologic risk factors be identified and monitored as biomarkers of exposures that interact with genetic risk variants?**
- **What recent changes in human exposures, diets, or social, cultural, and behavioral activities contribute to onset of disease in genetically predisposed individuals? Can any of these factors be modified to lower risk?**

Future Directions

➤ **Determine how candidate genes or sequence variants interact with environmental risk factors that can lead to disease outcome.**

The complexity of the environment is greater than that of the genome, and a personal environment can change regularly. Thus, for each candidate gene/disease variant, many potential interactions with environmental risk factors are possible. As diabetes and obesity develop over time, the coordinated tracking of exposures with subclinical markers of disease is necessary to characterize the interactions of genes with environments that increase risk. Biomarkers or surrogate environmental predictors are needed to identify agents that trigger disease progression. Environmental factors that might influence the risk of diabetes and obesity include nutrition, stress, physical activity, infections, and the human microbiome, which could influence susceptibility to both major forms of diabetes as well as obesity.

➤ **Develop resources and technologies to study gene-environment interactions.**

A critical component of genetics research is the development of novel resources and technologies to identify gene-environment interactions. Large, ethnically and geographically diverse cohorts of individuals need to be ascertained to establish research biobanks. Collection of specimens, genotyping for disease variants, and prospective follow-up of participants will help to identify those who are at risk for developing diabetes, complications of diabetes, and obesity, so that they can be asked to participate in detailed phenotypic and behavioral studies. Evaluation of the collected exposure data and biomarkers in the context of genetic risk should enable the identification of factors that are precursors of these complex medical problems and may help predict their outcome.

Genetics and Health Disparities

A significant level of disparity in the burden of diabetes, obesity, and diabetes complications exists among different ethnic and racial populations living in the United States. African Americans and Hispanics born in 2000 in the United States are estimated to have over a 40 percent lifetime risk for diabetes, a rate that is almost twice that of Americans of recent European origin (3).

They also have a younger age of onset, contributing to their greater risk of complications. From 1980 to 2006, the age-adjusted prevalence of diabetes almost doubled among African Americans, while the age-adjusted prevalence from 1997 to 2006 among Hispanics increased about 20 percent (8). American Indians have the highest prevalence of type 2 diabetes in the United States—after adjusting for population age differences, the rate of diagnosed diabetes in American Indians aged 20 and older is 16.1 percent (1). Obesity is much more prevalent in African Americans than those of European origin. Diabetic complications also disproportionately affect African Americans and Hispanics. In 2006, African Americans with diabetes were more than twice as likely to be diagnosed with end-stage kidney disease due to diabetes as those of European origin; similarly, there was almost a 70 percent greater likelihood of this diagnosis among Hispanics, after adjusting for population age differences (9). Ethnic differences, such as diet, life factors, and disease risk, appear to be present from conception through the entire lifespan.

Key Questions

- **Are genes and risk variants for diabetes and obesity in ethnic and racial minority groups in the United States, such as African American, Hispanic, Asian, and American Indian populations, the same as those found in populations of European origin?**
- **Are genetic variants identified in ethnic and racial minority populations for diabetes and obesity risk also predictive of pre-clinical disease, and do these variants interact with non-genetic risk factors similar to those identified in populations of European origin?**
- **Do genetic factors in minority populations predict outcome of treatment and complication risk?**

Future Directions

➤ **Identify the genetic and environmental bases of differences in diabetes onset, progression, and response to treatment in high-risk, minority populations.**

Research in genetics, social structures, cultural factors, behavior, and risk modification should focus on populations not of European origin that are at greatest risk to develop diabetes and obesity. Genetic research can aid in prediction and prevention of diabetes, obesity, and their complications in high-risk groups. Research is needed not only to screen minority populations for genes identified in populations of European origin, but also to search for evidence of novel genetic susceptibility factors in diverse populations. This knowledge could point to new methods for early screening and treatment of diabetes and obesity using genetic-risk-based approaches. Identification of genetic factors that predict outcome to treatment and complication risk could improve diabetes care and treatment of major complications. Finally, the ability to identify individuals with high genetic risk for developing type 2 diabetes and obesity would inform the design of clinical trials to prevent their onset in racial and ethnic minority populations in the United States. Working with minority communities in an integrated approach could be beneficial for advancing research on the genetic and environmental triggers of diabetes and obesity.

Epigenetic Contributions to Risk

In addition to genetic factors, disease susceptibility may be determined, in part, by environmental influences

that occur during development. It is possible that single nutrients, toxins, behaviors, cultural factors, environmental exposures, or combinations of these factors can alter the expression of genes. These genes may be already recognized as containing sequence variants that influence risk for diabetes, diabetic complications, and obesity, or they may be unknown but important once modified. This process, known as "metabolic imprinting," can affect the establishment of gene regulation during development, providing a potential biologic mechanism for disease susceptibility. Epigenetics is the study of persistent changes in gene expression that are not due to DNA sequence variants. Animal studies have suggested that epigenetic variation can contribute to obesity and, perhaps, diabetes. Epigenetic changes may help explain how gestational diabetes contributes to the long-term risk of diabetes and obesity in the offspring of affected pregnancies. Studying these changes may also contribute to understanding how a finite period of good glucose control early in the course of diabetes can slow the development of diabetes complications decades after the limited period of intensive management. Although the mechanism of this "metabolic memory" remains to be established, epigenetic changes provide a possible explanation. (See also the "Diabetes Complications" chapter for discussion of metabolic memory.) Changes in the environment (e.g., diet, exposures to various agents, stress) that induce epigenetic modification could account for much of the increased prevalence of diabetes and obesity in the United States.

Key Questions

- **Do DNA methylation and other aspects of epigenetic modification contribute to inter-individual variation in the risk of diabetes, obesity, and diabetes complications?**
- **Do epigenetic mechanisms correlate with risk and serve as therapeutic targets?**
- **What is the potential interaction between epigenetic modification and a pro-inflammatory environment and oxidative stress, and how does this interaction affect the risk of diabetes and obesity?**

Future Directions

➤ **Identify epigenetic markers that influence susceptibility to diabetes, obesity, and/or diabetes complications.**

Environmental exposures at any time from fetal development through adulthood may trigger a chemical change, such as a methyl group binding to a base in a gene or in a gene regulatory sequence, resulting in aberrant gene silencing or activation. In this manner, diet (i.e., nutrients), metabolic change (i.e., activity, glucose or lipid levels, inflammation), or exposures to pathogens (infection, toxins) can result in biochemical changes in specific target genes that may affect the risk of diabetes, obesity, or response to therapy. Research is needed to determine the role of specific exposures on embryonic or fetal development and correlate those exposures with clinical phenotype. The impact of social (stress), behavioral (maternal care), and environmental (pesticide, toxins, synthetics) factors on epigenetic changes and risk of diabetes, obesity, and diabetes complications must be considered with respect to health disparities. Existing biosamples and DNA can be used to detect epigenetic changes related to known diabetes and obesity genes or to environmental exposures. With enhanced knowledge of the impact of epigenetic changes on disease susceptibility, researchers can explore the potential for de-methylation therapies using dietary (e.g., vitamins, nutrients) and pharmacological

approaches to prevent or reverse risk for diabetes and obesity.

Translation of Genetic Research from Bench to Bedside

Diabetes and obesity are characterized by an extended pre-clinical period during which change in normal function escalates into subclinical and clinical disease. In the case of type 1 diabetes, autoimmune destruction depletes beta cells and reduces the capability to secrete insulin and maintain glucose homeostasis. In type 2 diabetes, insulin resistance and beta cell dysfunction play important roles in progression of the disease as well as in deterioration of glycemic control and risk of complications. Obesity is often marked by long-term weight gain with concomitant increase in caloric intake, decrease in physical activity, and linked risk of type 2 diabetes. Each biological process, therefore, exhibits a pre-clinical period that provides a window for interventions. These interventions could act to prevent overt disease by slowing or halting the progression of immune or metabolic dysfunction. Similarly, diabetic complications develop over the course of the disease, offering the opportunity to intervene to delay or avert their onset.

Scientific discoveries must be translated into practical applications to reduce risk of diabetes and obesity. This bench-to-bedside approach to translational research requires that basic scientists provide clinical researchers with new tools for use in patients and for assessment of their impact. Simultaneously, clinical researchers need to make novel observations about the nature and progression of treatments in order to stimulate new basic research. Translational research has proven to be a powerful process that drives the clinical research engine. However, a stronger research infrastructure could strengthen and accelerate this critical part of the clinical research enterprise with respect to diabetes and obesity in all populations.

Key Questions

- **How can the development of diabetes and obesity investigators who are well-trained, multidisciplinary and interdisciplinary, and able to form research teams be fostered?**
- **Can an incubator be created for innovative research tools and information technologies focused on translational and behavioral research in diabetes and obesity?**
- **Will current guidelines on human participants research permit synergism of multidisciplinary and interdisciplinary clinical and translational research to facilitate the application of new knowledge and techniques in clinical practice?**
- **Can opportunities be developed to bring physiologists (both animal and human) into a productive collaboration with geneticists to bridge research gaps?**
- **What methods can be developed to translate novel techniques of prediction, prevention, and treatment into the general community?**

Future Directions

➤ **Optimize the use of genetic and environmental risk factor data in the design of translational and clinical research programs for diabetes and obesity.**

The design of translational research paradigms and development of clinical trials requires very careful risk/benefit analysis that puts a premium on prediction of disease risk and potential outcome. As part of optimizing both the collection and use of data in these

studies, they should be designed to provide access to all individuals interested in participation and not restricted to any one group (defined by ethnicity, socioeconomic status, etc). Similarly, application of new technologies for the treatment, prevention, and prediction of diabetes and obesity should be open and transparent to all individuals. Innovative processes are needed for recruitment and retention of high-risk populations for diabetes and obesity. Therapies identified from research protocols with higher risk of adverse effects should be matched to participants with higher predicted risk of developing diabetes or obesity. Risks should be proportionate to potential benefit, and risks to children should be limited. The ability to provide novel therapies for diabetes and obesity would be advanced by the development of regional, centralized provider networks that can provide these therapies at high volume and distribute them rapidly to health care providers and researchers for application to all populations. Biomarkers that could more accurately predict response to specific types of therapies would increase the efficiency of trials and enable potential earlier intervention, at a stage where pre-clinical intervention may improve outcomes.

Pharmacogenetics/Pharmacogenomics

Pharmacogenomics, the branch of pharmacology and genetics that studies the influence of genetic variation on drug response, is performed by correlating SNP and/or gene expression data with a drug's efficacy or toxicity. Pharmacogenomics (on a genome-wide level) attempts to develop rational means to optimize drug therapy, with respect to a person's genotype, to ensure maximum efficacy with minimal adverse effects. As more genes and pathways that contribute to disease are identified, the same technologies can be used to personalize the prediction of response to treatment. One use of this growing genetic technology is to identify individuals whose response to particular drugs is determined, in part, by their genes. GWA studies and candidate pathway approaches have identified genes in populations that lead a person to respond to lower doses of drugs, experience adverse drug effects, or not respond to a particular drug at all. For example, the FDA has approved a genetic test to help assess warfarin sensitivity, as one-third of people metabolize this anti-coagulant medication more slowly than the general population and, therefore, experience a higher risk of bleeding. In this case, some of the previously unexpected response to warfarin depends on variants of two genes, the cytochrome P450 2C9 gene (*CYP2C9*) and the vitamin K epoxide reductase complex subunit 1 gene (*VKORC1*). Some of these genetic differences are responsible for individual differences in drug metabolism. Currently, studies are under way to examine genetic differences that lead to different responses to drugs in people with or at risk of type 2 diabetes. In the NIH-led Diabetes Prevention Program clinical trial, which tested interventions to prevent or delay type 2 diabetes in people at high risk, treatment with metformin reduced the risk of developing diabetes by 31 percent among participants in that arm of the study relative to the control group. However, further analysis revealed that metformin was not effective in study participants carrying a lysine at a particular position in the protein encoded by *KCNJ11*, while it was effective in participants with a glutamate at that position. One of the first uses of genetics for personalized medicine may be to test individuals to determine which drug regimen would be most effective.

Key Questions

- **What are the genes and variants that predict response to specific treatments of diabetes and obesity?**
- **Do genes that are identified for treatment response correlate with genes that predict risk of complications?**
- **Are genetic variants for response to treatment the same in different ethnic and racial populations, and do these variants interact with similar non-genetic risk factors?**
- **Is the use of genetic information in disease management cost effective? Does it lead to better patient outcomes?**

Future Directions

➤ **Identify the genetic and environmental bases of differential response to pharmacologic treatment of diabetes and obesity, as well as their relationship to progression and complications of disease.**

Research in pharmacogenomics should focus on the relationship between inherited predictors of response to pharmacologic agents used in the treatment of diabetes (type 1 and type 2), obesity, and diabetic complications. Differential response to drugs used in treatment of these diseases has been observed across all population and ethnic groups. While some of the variation in response can be attributed to factors related to health status, compliance, and other confounding factors, there is a clear role for genetic factors. Pharmacogenomics can be applied to the problems associated with treatment of disease and complications at several levels. Genes that are associated with successful treatment of diabetes and/or obesity may identify a molecular mechanism that provides insight on the etiologic framework of the disease. Variants in genes that predict treatment response can influence drug design in ways that could improve efficacy of the compound. Differential effects of genes across populations, including the interaction of genes with environmental risk, can be used to stratify populations in early efficacy studies or trials to improve the quality of clinical decision making and treatment options. Importantly, the identification of genes associated with adverse drug effects can greatly improve quality of life and reduce pharmacologically mediated morbidity and mortality. Thus, pharmacogenetic research will be of key importance to developing strategies to individualize diabetes prevention and therapy.

➤ **Evaluate the utility of genetic information from a public health perspective.**

Direct-to-consumer commercial companies are marketing genetic tests to the population; yet, whether these tests lead to better patient outcomes or improve health at the population level is unknown. Behavioral research into motivational strategies that improve patient behavior, outcomes research that validates the use of genetic results, and economic research that evaluates the feasibility of deploying point-of-care genetic testing must be carried out. As genetic information emerges, there is a need for trials that specifically address the question of whether *a priori* knowledge of genetic information 1) affects patient or practitioner behavior, 2) leads to better patient outcomes, and 3) is cost effective. Some of these studies can be initiated before a comprehensive set of genetic determinants associated with type 1 diabetes, type 2 diabetes, and obesity is compiled.

IMPORTANCE OF RESEARCH GOALS AND STRATEGIES: HOW TRANSLATING RESEARCH OUTCOMES MAY LEAD TO IMPROVEMENTS IN HEALTH

Both genetic and environmental exposure data can provide important leads into new therapeutic strategies, identify new treatment targets, and define the utility of many forms of treatment for diabetes, diabetic complications, and obesity. For example, therapy of children with a rare neonatal form of diabetes has been transformed by the identification of the genetic basis of the disorder and use of targeted therapy. Despite the complexity of the more common forms of diabetes, there has been substantial progress in the ability to identify individuals at increased genetic risk. Important new information about how genetic variation affects disease risk and response to therapy is emerging that will create opportunities to improve the health of people who are living with these diseases or at risk of developing them. Genetics research could be used to identify people who are at high risk of developing diabetes or obesity so that they can be enrolled in prevention programs, where available, or closely monitored for disease development so that appropriate treatment can begin immediately at disease onset. Genetics information could also help clinicians identify which people are most likely to benefit from specific therapies and which have an elevated risk for adverse events. Assembling interactive, multidisciplinary teams of investigators will facilitate greater translation of genetics research into the clinical arena and public health.