Genetic epidemiology of diabetes

M. Alan Permutt, 1 Jonathon Wasson, 1 and Nancy Cox 2

1Department of Endocrinology, Metabolism and Lipid Research, Washington University School of Medicine, St. Louis, Missouri, USA.
2Department of Human Genetics, The University of Chicago, Chicago, Illinois, USA.

Conventional genetic analysis focuses on the genes that account for specific phenotypes, while traditional epidemiology is more concerned with the environmental causes and risk factors related to traits. Genetic epidemiology is an alliance of the 2 fields that focuses on both genetics, including allelic variants in different populations, and environment, in order to explain exactly how genes convey effects in different environmental contexts and to arrive at a more complete comprehension of the etiology of complex traits. In this review, we discuss the epidemiology of diabetes and the current understanding of the genetic bases of obesity and diabetes and provide suggestions for accelerated accumulation of clinically useful genetic information.

Definition of the problem

Diabetes is a metabolic condition in which the body fails to produce enough insulin. Type 1 diabetes (T1D) results from autoimmune destruction of insulin-producing β cells, which leaves the patient dependent on insulin injections for survival (1) T2D, formerly known as adult-onset diabetes, occurs when impaired insulin effectiveness (insulin resistance) is accompanied by the failure to produce sufficient β cell insulin. Patients can be placed on regimens to reduce weight or manage diet or treated with medication and, less often, insulin injections. This latter form of diabetes accounts for as much as 95% of cases. Gestational diabetes is another form of diabetes, defined as a state of glucose intolerance during pregnancy that usually subsides after delivery but has major implications for subsequent risk of T2D, as pregnancy serves as an “environmental” stressor that reveals a genetic predisposition. Other less common forms of diabetes include the rare, genetically determined disease maturity onset diabetes of the young (MODY), diabetes resulting from surgery, and other illnesses that constitute only 1–5% of cases. Based on plasma glucose measurements, 2 conditions have been identified with increased risk of the disease (2): (a) impaired glucose tolerance (IGT) is defined as hyperglycemia intermediate between normal and diabetic levels following a glucose load; (b) impaired fasting glucose (IFG), like IGT, is associated with increased cardiovascular disease (CVD) and future diabetes. Because complications of diabetes may develop years before overt disease, many consider T2D may now occur more commonly than T1D (10). The estimated lifetime risk of developing diabetes for individuals born in the United States in 2000 is 33% for males and 39% for females (7). It is highest among Hispanic females, at 53%. Diabetes is associated with large reductions in life expectancy, on the order of 11 years in males diagnosed at age 40. While an estimated 18.2 million persons had diabetes in the United States in 2002 (11), diabetes worldwide has been estimated to affect 151 million persons, and that number projected to increase to 324 million by 2025 (2).

Prevalence

The prevalence of diabetes in the United States has risen 40%, from 4.9% in 1990 to 6.9% in 1999 (7). A breakdown of the prevalence of diabetes by state from 1990 through 2001 and of obesity by state from 1991 through 2003 is shown in Figure 1, A and B, respectively (8, 9). The disease affects various groups differently, occurring 10 times more commonly in those older than 65 years compared with those younger than 45 years. Minority racial groups including Hispanics, African Americans, and Native Americans are generally affected at a rate 2–4 times that for white individuals. The recent increased prevalence has also been noted in children and adolescents, where T2D may now occur more commonly than T1D (10).

The medical burden of diabetes

The burden of diabetes is to a large extent the consequence of macrovascular and microvascular complications of the disease, which result in large increases in morbidity and mortality. For example, the prevalence of ischemic heart disease is 2–14 times the rate in age-matched nondiabetics (12). Diabetic retinopathy is the chief cause of blindness in the US. In 2000, diabetic renal disease accounted for 40% of new cases of end-stage renal disease, and diabetics are the largest group receiving dialysis (more than 50% of all cases) and renal transplants (approximately 25%). Lower extremity disease resulting from a combination of peripheral vascular disease and neuropathy causes an increase in lower extremity amputations. While improved glycemic control has been shown to reduce the incidence of microvascular complications, episodes of severe symptomatic hypoglycemia were 3 times higher in those receiving intensive insulin management therapy (13). Along with experiencing physical and cognitive disabilities, adults with diabetes have an age-adjusted mortality rate estimated to be twice that of nondiabetics (12). Risk factors for CVD, including systolic hypertension, elevated cholesterol levels, and cigarette smoking, independently predict CVD mortality, and any 1 risk factor affects outcomes more in persons with diabetes (14).

Nonstandard abbreviations used: ASP, affected sibling pair; CVD, cardiovascular disease; DZ, dizygotic; IGT, impaired glucose tolerance; IRAS, Insulin Resistance Atherosclerosis Study; MODY, maturity onset diabetes of the young; MZ, monozygotic; SNP, single nucleotide polymorphism; T1D, type 1 diabetes; T2D, type 2 diabetes; VNTR, variable number of tandem repeats.

Conflict of interest: The authors have declared that no conflict of interest exists.

Although increase in diabetes prevalence occurs mostly in middle-aged and older adults, there is strong evidence of an increase in the prevalence of T2D in children (10). For example, in Japan the incidence in school children (6–15 years old) has doubled over a 20-year period, such that T2D is now more common than T1D (15). In the US, up to 45% of the newly diagnosed diabetics in the pediatric age group have T2D (10). This rise in diabetes rates in children reflects, at least in part, the growing prevalence of obesity in this age group (16).

Direct medical expenditures and lost productivity due to diabetes were estimated to cost the US $132 billion in 2002 (17). The per capita expenditures were twice those for individuals without the disease. While the prevalence of diagnosed diabetes is less than 5% of the population, almost $1 of every $5 spent on health care in the US is for patients with diabetes. As the prevalence of diabetes increases with age, and because of the increasing diabetes-prone populations, it has been estimated that the number of diagnosed cases will increase. Thus the projected total cost in 2002 dollars could be as high as $192 billion by 2020.

**Etiology of the diabetes epidemic**

The sudden increase in diabetes in the last few years is due not to genetic factors but rather to the increase in obesity. This phenomenon is currently being documented in Africa, where the incidence of diabetes is rising with urbanization. The incidence is also rising among Africans who have immigrated to the US (18, 19). Epidemiological studies have regularly shown the relationship between diabetes and obesity, mediated in part by nutritional and lifestyle factors (20, 21). The most common measure of obesity, body mass index (BMI), combines measurements of height and weight. People with a BMI greater than 25 are said to be overweight, while those with a BMI greater than 30 are defined as obese (22). The Nurses Health Study showed that the risk for developing diabetes increased sharply for individuals observed as having a BMI greater than 23 for 16 years and was increased 20-fold for those with a BMI greater than 30 (18). In a recent study of measures of obesity and CVD risk factors in Australian adults, the prevalence of T2D rose from 5% in normal-weight to 16% in obese males; of hypertension, from 20% to 49%; and of dyslipidemia, from 18% to 61%, with even higher prevalence in females (23).

The molecular and physiological relationships between obesity and diabetes are not fully understood, and this subject is an area of intense investigation (see ref. 22 for review). The “thrifty genotype” hypothesis was proposed to account for a genetic advantage of accelerated fat deposition during times of restricted availability of calories, which leaves individuals faced with harmful consequences given the abundant food supply and reduced levels of physical activity in developed countries today (24). Noting an association between low birth weight and increased incidence of diabetes in later life, Hales and Barker have hypothesized that intrauterine malnutrition results in reduced birth weight and to subsequent changes leading to disease in adults (25, 26). This phenomenon, also known as the “thrifty phenotype” hypothesis, proposes that fetal malnutrition results in impaired pancreatic β cell development and insulin resistance. Offspring are subsequently more prone to diabetes and the metabolic syndrome when exposed to abundant nutrition later in life. In this regard, the increased prevalence of T2D in offspring of diabetic mothers may be a consequence of environmental factors operating on a genetic background, i.e., an altered intrauterine environment superimposed on a genetic predisposition in the fetus. While epidemiological studies have confirmed these observations, virtually nothing is known of their mechanisms, and this is an active area of investigation (27). If the relationship between obesity and diabetes could be understood, or obesity effectively prevented with treatment, then therapies directed at these mechanisms might curtail the increasing incidence of the disease.

The relationship between obesity and diabetes has been extensively studied in inbred strains of mice (28). Mice from a single inbred strain fed a high-fat diet all became insulin resistant, yet only about half became both obese and diabetic. Interestingly, 10% became diabetic but resisted obesity, and 10% became obese but not diabetic. The mechanisms responsible are unlikely to be purely genetic, and the results are consistent with the hypothesis that epigenetic changes and stochastic factors contribute to the phenotypic diversity. More recently, ER stress was shown to be the etiology of obesity-induced insulin resistance and diabetes in experimental mouse models, and this mechanism promises to be a rewarding area of investigation in the near future (29). If the degree of ER stress varies among mice, then perhaps this could explain the phenotypic differences in mice that are genetically identical.
As it has long been noted that levels of FFAs are increased in obese individuals, their accumulation in skeletal muscle has been proposed to compete with circulating glucose resulting in hyperglycemia, hyperinsulinemia, and ultimately insulin resistance (30). Recently, using magnetic resonance spectroscopy in patients with obesity and/or T2D, Shulman et al. have shown a reduction in the rate of insulin-stimulated glucose metabolism secondary to reduced muscle glycogen synthesis, associated with a blunted increase in intramuscular glucose 6 phosphate relative to concentration in insulin-resistant offspring of diabetic parents compared to control individuals (31, 32). Increased intracellular fatty acid metabolites were found to result in decreased insulin signaling and impaired glucose transport. Other factors contributing to the insulin resistance of obesity include the tendency to store fat in the abdominal region as opposed to the extremities and defects in adipocyte fatty acid metabolism and mitochondrial fatty acid oxidation. While impaired mitochondrial activity in insulin-resistant offspring of patients with T2D has been observed (33), the genetic basis for reduced mitochondrial biogenesis has not been elucidated. Reduced mitochondrial activity results in reduced energy expenditure, obesity, increased intramuscular fatty acid accumulation, and insulin resistance and has therefore been incorporated into the thrifty gene hypothesis (32).

Heritability of diabetes
The recent increased prevalence of obesity and diabetes must be largely attributable to changes in nongenetic risk factors. Yet environmental aspects must certainly accelerate the disease in those with genetic predisposition. There is a clear need to understand the genetic basis for the regulation of food intake, energy expenditure, and variations in energy balance in various individuals. In the long run, it may be more beneficial to develop treatments based on these genetic mechanisms than to rely on the use of will power to modify lifestyle. Moreover, different aspects of environment may be more critical in different subsets of individuals. We know from the natural history of diabetes complications that when patients are first diagnosed, there may already be marked progression of microvascular and macrovascular complications (34). The overwhelming majority of obese individuals have insulin resistance, yet only 5–10% develop pancreatic β cell failure and diabetes. Discovering the genetic risk factors for the disease will likely have many positive consequences.

The familial occurrences of both T1D and T2D have been long noted. A sibling’s risk of developing T1D (5–10%) is perhaps 12- to 100-fold greater than the risk in the general population (0.1–0.4%) (35). Concordance in monozygotic (MZ) twins has been consistently shown to be greater than that in dizygotic (DZ) twins (36). For T2D, the concordance among MZ twins has been observed to be 50–92%, higher than the 37% concordance in DZ twins (36). Thus while the relative risk to a sibling, a measure of the genetic contribution, is considerably greater for T1D than for T2D, the concordance and absolute risk are substantially greater for T2D, which perhaps underscores the importance of the environmental contribution to the latter (37).

Quantitative phenotypes related to glucose homeostasis are also known to be heritable (38). In families with an increased genetic susceptibility to T2D, heritability estimates for β cell function and features of the insulin resistance syndrome of 72% and 78%, respectively, were calculated (39). The heritability of other features of the insulin resistance syndrome, including BMI, blood pressure, and serum lipid and insulin sensitivity levels, was also estimated to be high. Evidence for heritability of these metabolic phenotypes was reported in studies of Pima Indians (40) and non diabetic Japanese Americans (41); in the Insulin Resistance Atherosclerosis Study (IRAS) among family members of African American and Hispanic heritage (42); and in a study of the familial aggregation of the amount and distribution of subcutaneous fat and responses to exercise training in the HERITAGE Family Study (43). These studies strongly support the role of both genetic and environmental factors in the etiology of diabetes and the insulin resistance syndrome.

The value of monogenic subtypes of diabetes
Recent reviews provide detailed appraisals of both linkage and association studies in diabetes (38, 44). We focus here on the patterns that have emerged in the search for genetic risk factors for diabetes. Linkage mapping, positional cloning, and candidate gene studies have been most successful in the forms of diabetes with the simplest genetic models. Early studies characterizing diabetes resulting from mutations in insulin (INS) and insulin receptor (45) genes as well as the mitochondrial genome (46) provided important insights into glucose homeostasis, but it is the studies on MODY that provide the classic example of the successful application of genetics to diabetes (47, 48). Linkage mapping studies on MODY were quite successful, and within a few years, a combination of positional cloning and studies of positional candidate genes led to the identification of glucokinase (GCK) and hepatocyte nuclear factors 4α and 1α (HNF4A and TCF1) as genes in which a single mutation could lead to the complex metabolic phenotype of diabetes (see ref. 49 for a recent review). GCK is the rate-limiting enzyme in glucose metabolism; HNF4A and TCF1 are transcription factors expressed in a variety of tissues, including the liver and pancreas where they work in a transcription regulatory network to regulate the expression of genes involved in glucose transport and metabolism (49).

Once these genes had been implicated in MODY, a number of other transcription factors within the same regulatory network were also found to be MODY genes. As noted above, mutations in MODY genes, at the INS and INS receptor genes, or in the mitochondrial genome collectively account for only a small proportion of diabetes — about 1–5% (50). However, the monogenic forms of diabetes provide important insights into how we should be thinking about the genetic components of a phenotype as broad and metabolically complex as that of diabetes. The genes in which a single mutation is sufficient to generate this phenotype play an absolutely central role in glucose homeostasis (e.g., INS, the INS receptor, GCK) or are capable of affecting the regulation of many genes that act within this overall pathway (e.g., transcription factors). From these studies we learned that glucose homeostasis is a balance between insulin production, determined by β cell mass and/or function, and insulin action. Some of the genes that have been shown to affect these processes either in humans or experimental animals are illustrated in Figure 2.

The general observations on the genes implicated in monogenic forms of diabetes fit well within the emerging theories for how biological networks might be expected to perform and suggest that a systematic application of network theory to the metabolic and regulatory pathways underlying glucose homeostasis may provide a fruitful avenue for prioritizing genes for future studies. Recent research suggests that a scale-free topology is a nearly universal feature of networks, whether we are considering air traffic patterns,
network theory with even the information we already have about the pathways implicated in glucose homeostasis, we could predict that genes/proteins located at hubs that are the most highly interconnected are those most vulnerable to degrading the overall system. Thus, genetic variation at hub genes may be more likely to lead to detectable perturbations in glucose homeostasis. While we might use network theory to prioritize genes for study simply based on the relative connectedness of the genes at the hubs in the networks, it might also be useful to overlay such an analysis on existing information we have about potential genetic risk factors for diabetes via linkage mapping or linkage disequilibrium mapping studies. A similar strategy was recently applied in studies of Alzheimer disease, with promising results (54).

Genetic studies on T1D

Genetic linkage studies of T1D and T2D have been quite variable, spanning the spectrum of results generally observed for disorders with complex inheritance. T1D is unique among complex disorders in the magnitude of the familial risk attributable to a single locus, HLA. While HLA was originally implicated through association studies as a candidate gene (55), the magnitude of the evidence for linkage at HLA in T1D is larger than has been observed for a linkage in any other complex disorder, although some other autoimmune disorders also have strong evidence for linkage in the HLA region (56). Because of the major contribution of HLA to the familial risk of T1D, identification of the other genetic risk factors may be more akin to identifying modifier loci for monogenic disorders than identifying primary susceptibility loci for complex disorders. Indeed, the non-HLA genes that have been reproducibly characterized as T1D susceptibility loci have been identified largely through candidate gene studies. Evidence for linkage at the non-HLA loci implicated in T1D has been uneven at best. For example, the very common class I alleles at INS variable number of tandem repeats (VNTR) are significantly overtransmitted from parents heterozygous for this allele to offspring affected with T1D (57). There was no evidence for linkage of the INS region to T1D in 100 affected sibling pair (ASP) families (58) and only modest evidence in more than 200 families (59). In a

Figure 2

Diabetes results from an imbalance between the insulin-producing capacity of the islet β cell and the requirement for insulin action in insulin target tissues such as liver, adipose tissue, and skeletal muscle. Some of the many genes that have been shown or could possibly contribute to the imbalance are illustrated.
large sample (767 families) including ASP families from both the US and the United Kingdom, there was significant evidence for linkage near INS (60). Similarly, initial evidence for linkage near CTLA4, another locus with support for affecting susceptibility to T1D through association studies, was detected in some individual samples, but little evidence for linkage in the region of CTLA4 was found in larger, combined samples (60). Finally, PTPN22, a locus recently implicated as a candidate gene in T1D (61), rheumatoid arthritis (62), and systemic lupus erythematosus (63), shows little evidence for linkage to T1D even in relatively large samples (60).

**Genetic studies on T2D**

The success of linkage mapping for T2D has been similar to that observed for other complex disorders, which has been, regrettably, quite limited. More than 25 genome-wide screens have been conducted on samples from all over the world (for review, see ref. 44). Despite the number of studies, there are only a few regions with replicated evidence for linkage: 1q, 3q, 8p, 10q, 12q, and 20q. Even in these regions, however, evidence for linkage is far from universal (3–7 of the more than 25 studies show nominally significant evidence for linkage), peaks are broad, and it is unlikely that all studies with linkage signal in a given region reflect the contribution of the same susceptibility genes.

**Factors contributing to the complexity of analysis**

Why has linkage mapping been relatively unsuccessful in localizing susceptibility genes for T1D and T2D in even relatively large, combined data sets? The complexity of the underlying genetic model is clearly a contributing factor. We have almost certainly underestimated the number of different genetic risk factors for both disorders and overestimated the magnitude of effect that might be expected for any one of these loci, excepting, perhaps, HLA in T1D. It might be argued that the initial success with HLA in T1D, the first susceptibility gene successfully linked to a complex disorder, encouraged us to establish a series of unrealistic expectations for how genetic risk factors contribute to diabetes and other complex disorders. Among the complexities of the genetic models for complex disorders that are likely to contribute to the difficulties in linkage mapping are gene-gene and gene-environment interactions. Such interactions are difficult to accommodate in primary linkage mapping studies and yet are a requisite part of the definition of a complex trait. It is not clear whether these problems could be solved by increasing either the sample sizes for linkage studies or the number of phenotypes examined. In particular, the failure to adequately measure and account for nongenetic factors affecting risk of diabetes almost certainly has reduced our ability to successfully map genetic risk factors.

The diagnosis of diabetes has long been standardized and is both reliably and inexpensively achieved with a simple blood test. These factors contributed substantially to making diabetes the first complex disorder to be widely studied using genetic tools. But the diagnosis of diabetes is designed to focus on the clinical consequences of elevated blood glucose levels rather than the underlying genetic liability to this very complex metabolic disease. The simple dichotomous diagnosis masks a tremendous amount of clinical heterogeneity, and it is likely that the genetic heterogeneity of diabetes is at least as great as the clinical heterogeneity. Thus, efforts to specify more genetically homogeneous samples according to clinical characteristics might also be fruitful. For example, stratifying T1D families for linkage analyses according to antibody positivity, or patients with T2D by BMI, might lead to more consistent and reproducible results in linkage mapping studies. Additionally, analysis of quantitative traits that may be related to the primary dichotomous trait of T2D, such as insulin resistance, β cell mass and performance, BMI and other features of the metabolic syndrome, may lead to the identification of genes contributing to risk of T2D (64). Several such studies have been already been initiated (see ref. 65 for review), including a recent IRAS study of quantitative traits in African American and Hispanic families (66), and the results point to promising genomic regions, though no causative genes have yet been identified. In this regard, use of animal models could help in gene identification, as syntenic regions are being evaluated in congenic strains in order to narrow regions conveying genetic risk for T2D, as, for example, in obese mice (67). We must also recognize, however, that the quantitative phenotypes that we now know how to measure easily are not necessarily the phenotypes best able to characterize the genetic liability to T2D.

**Emerging patterns and implications for study design**

The patterns emerging from the linkage and association studies that have identified genetic risk factors for diabetes offer intriguing insights into the challenges we face in improving our study designs. Some of the factors are common but so low risk that they would be quite difficult to detect in linkage mapping studies. For example, the allele increasing risk of T2D at PPARG has a frequency of 0.85–0.95 in most of the world’s populations (68) and is associated with very modest increase in risk. Similarly, the class I alleles (or polymorphisms in linkage disequilibrium with them) increasing risk for T1D at the INS VNTR are found at a very high frequency in populations of European and Asian descent (0.70–0.85), but they increase risk only modestly (69). There are rarer amino acid polymorphisms that have been reliably associated with diabetes. For example, the allele increasing risk of T2D at the T504A polymorphism at CAPN10 ranges in frequency from 0.04 to 0.16 (70), and the allele increasing risk of T1D at the R620W polymorphism at PTPN22 ranges in frequency from 0.08 to 0.14 (69).

But many of the polymorphisms associated with increased risk of diabetes identified to date are not amino acid polymorphisms. The variation at CTLA4 implicated in T1D appears to affect splicing (71), while variation at the INS VNTR (69), at CAPN10 (72), and at HNF4A (73, 74) (as discussed in more detail below) may affect gene expression. Linkage mapping, even with very large samples, will miss many of these risk factors. Similarly, genome-wide association mapping focused exclusively on common haplotypes will miss many of the rarer risk alleles. Strategies targeting known amino acid polymorphisms will miss rare, unknown susceptibility variants and may not detect the effects of the common non-coding sequence polymorphisms either. Until it becomes clearer whether there will be a predominant frequency spectrum or polymorphism type in the genetic variation affecting susceptibility to diabetes (whether type 1 or 2), it seems prudent to adopt strategies that enable detection of susceptibility alleles with a wide range of frequencies and effects. Some known genes associated or linked with diabetes are listed in Table 1 (61, 70, 73, 74, 90–117).

**Sample size for low-risk genes associations**

Identification of susceptibility alleles for T1D, outside the HLA locus, and T2D, whether through positional cloning or in the context of studies on functional candidates, has been challenging. Initial positive results are usually only inconsistently replicated. For
example, conflicting results of different studies on the Pro12Ala polymorphism in the PPARG2 gene were resolved by analysis of large family and case control samples; and a meta-analysis of all published studies further demonstrated that this polymorphism does affect risk of T2D, but only to a small degree (68). Similarly, the E23K polymorphism in the Kir6.2 subunit of the ATP-regulated potassium channel has been shown by meta-analysis to contribute a small but significant risk to the disease in the populations studied (75, 76). Except for HLA in T1D, the susceptibility alleles for T1D and T2D quite modestly affect risk of disease, which mandates the study of large sample sizes. This argues for large sample sizes. This argues for large sample sizes for T1D and T2D quite modestly affect risk of disease, which mandates the study of large sample sizes. This argues for large sample sizes for T1D and T2D quite modestly affect risk of disease, which mandates the study of large sample sizes. This argues for large sample sizes for T1D and T2D quite modestly affect risk of disease, which mandates the study of large sample sizes. This argues for large sample sizes for T1D and T2D quite modestly affect risk of disease, which mandates the study of large sample sizes. This argues for large sample sizes for T1D and T2D quite modestly affect risk of disease, which mandates the study of large sample sizes. This argues for large sample sizes for T1D and T2D quite modestly affect risk of disease, which mandates the study of large sample sizes. This argues for large sample sizes.

**Table 1**

<table>
<thead>
<tr>
<th>Type</th>
<th>Gene</th>
<th>Gene name</th>
<th>Function</th>
<th>SNP or allele or locus or marker</th>
<th>Refs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1D</td>
<td>HLA</td>
<td>Human leukocyte antigen</td>
<td>Involved in numerous aspects of metabolism</td>
<td>Variants in multiple genes</td>
<td>99, 100</td>
</tr>
<tr>
<td>T1D</td>
<td>INS</td>
<td>Insulin</td>
<td></td>
<td>VNTR</td>
<td>99, 101</td>
</tr>
<tr>
<td>T1D</td>
<td>CTLA4</td>
<td>Cytotoxic T-lymphocyte–associated protein 4</td>
<td>Immune system regulation</td>
<td>T17A</td>
<td>102, 103</td>
</tr>
<tr>
<td>T1D</td>
<td>PTPN22</td>
<td>Protein tyrosine phosphate, non-receptor type 22</td>
<td>Immune system regulation</td>
<td>SNP C1858T</td>
<td>61, 104</td>
</tr>
<tr>
<td>T2D</td>
<td>ABC8</td>
<td>ATP-binding cassette, subfamily C, sulfonlurea receptor</td>
<td>Regulator of potassium channels and insulin release</td>
<td>SNPs in various exons</td>
<td>105, 106</td>
</tr>
<tr>
<td>T2D</td>
<td>CAPN11</td>
<td>Calpain 10</td>
<td>Protease</td>
<td>Various intronic SNP haplotypes</td>
<td>70, 107, 108, 109</td>
</tr>
<tr>
<td>T2D</td>
<td>GCGR</td>
<td>Glucagon receptor</td>
<td>Controls hepatic glucose production and insulin secretion</td>
<td>G40S</td>
<td>108, 109</td>
</tr>
<tr>
<td>T2D</td>
<td>GCK</td>
<td>Glucokinase</td>
<td>Glucose metabolism</td>
<td>Microsatellite in 3' end of gene</td>
<td>110, 111</td>
</tr>
<tr>
<td>T2D</td>
<td>KCNJ11</td>
<td>Potassium inwardly-rectifying channel, subfamily J, member 11</td>
<td>Regulation of insulin secretion</td>
<td>E23K</td>
<td>112, 113</td>
</tr>
<tr>
<td>T2D</td>
<td>PPARγ</td>
<td>Peroxisome proliferator-activated receptorγ</td>
<td>Transcription factor</td>
<td>P12A</td>
<td>114, 115</td>
</tr>
<tr>
<td>T2D</td>
<td>HNF4A</td>
<td>Hepatocyte nuclear factor 4 α</td>
<td>Transcription factor</td>
<td>P2 promoter SNPs</td>
<td>73, 74</td>
</tr>
<tr>
<td>T2D</td>
<td>SLC2A1</td>
<td>Glut 1</td>
<td>Glucose transporter</td>
<td>XbaI(−) restriction site</td>
<td>116, 117</td>
</tr>
</tbody>
</table>

*with the exception of MODY 4 and MODY 6.

**Investigating potential regulatory regions of candidate genes through haplotype-tagged SNPs**

While every nonsynonymous coding single nucleotide polymorphism (SNP) in candidate genes should be tested for possible contribution to disease susceptibility, 2 recent studies highlight the importance of conducting association studies with markers in potential regulatory regions. Earlier studies with SNPs in or near the coding region of HNF4A, a gene previously shown to be mutated in rare cases of MODY (77), had yielded no association with T2D (78). More recently, it was discovered that a second promoter exists 40 kb upstream of the gene (79, 80) and that SNPs in the region of this second promoter and in other parts of the noncoding sequence of HNF4A were associated with T2D in Ashkenazi Jews (74) and in a sample in Finland (73). Results of follow-up studies in other populations may provide some confirmation of the association between T2D and noncoding SNPs at HNF4A and its regulatory regions (81), but we should not be surprised if some studies in even large replication samples fail to observe associations and should be equally prepared for the possibility that not all studies will identify the very same polymorphisms as showing association. The nature of regulatory variation virtually insures that effects attributable to one polymorphism might be attenuated by effects of a second polymorphism — thus, the cumulative effects of regulatory variants may be poorly predicted by the marginal effects measured for any individual variant.

**DNA diagnostics and pharmacogenetics in clinical trials**

The use of genomic tools provided by the Human Genome Project offers the opportunity to identify individuals at risk, classify subtypes of the disease, choose therapy based on more accurate diagnosis (82), more precisely delineate the environmental factors that contribute to the onset and progression of the disease and its complications, and monitor responses to therapy (83, 84). Recently, genetic information was applied to clinical diabetes management in a randomized crossover trial of gliclazide, an agent affecting insulin secretion, and metformin, an agent that enhances insulin action. Compared with patients with typical T2D, patients with diabetes caused by a particular MODY mutation in TCF1 (85) had a 4- to 5-fold greater response to gliclazide than to metformin. Another example was the recent finding of heterozygous mutations...
in the ATP-sensitive K+ channel subunit of the Kir6.2 gene in 7 of 11 patients with neonatal diabetes (86). This gene plays a critical role in glucose-stimulated insulin secretion. Remarkably, several patients who previously required insulin injections were taken off insulin and treated with oral medication, which again illustrates the efficacy of pharmacogenetics for treatment of some diabetics.

A large number of clinical trials for both T1D and T2D are currently being conducted. The Type 1 Diabetes Genetics Consortium (T1DGC; http://www.t1dgc.org) will organize international efforts to identify genes that determine an individual’s risk of T1D through the identification of 2,500 new families with 2 or more affected siblings. To explore approaches to treatment of T2D in youth, the TODAY (Treatment Options for Type 2 Diabetes in Adolescents and Youth; http://www.TODAYstudy.org) study will enroll 750 children and teenagers that have recently been diagnosed with T2D. Participants will be assigned to groups for treatment aimed at weight reduction and increasing physical activity. The Look AHEAD (Action for Health in Diabetes) trial is a multicenter, randomized clinical trial that will examine the consequences of a lifestyle intervention designed to achieve and maintain weight loss over the long term through decreased caloric intake and increased exercise in 5,000 obese patients with T2D. The National Heart, Lung, and Blood Institute–led Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial is designed to test the effects of glycaemia and blood pressure control on major CVD events and the use of fibrate treatment to increase HDL cholesterol and lower triglycerides (http://www.accordtrial.org/public/index.cfm). The Bypass Angioplasty Revascularization Investigation 2 Diabetes (BARI 2D) trial (http://www.bari2d.org) addresses questions about therapy in adults with T2D and stable CAD who might be candidates for revascularization. These studies represent ideal opportunities to incorporate DNA diagnostic testing and assessment of variable responses to therapeutic interventions.

**Large prospective cohort study for the effects of genes and environment on diabetes**

While case-control studies have much to offer for the assessment of the interactions between genes and environmental factors, Francis Collins noted that clinically diagnosed cases may represent only the more severely affected individuals with the disease and highlighted the difficulties of selecting an unbiased control group (87). To more accurately quantify genetic contribution and population-wide risk, he proposed prospective, population-based cohort studies (88). In a case-control study, there would be no samples available from cases prior to the onset of disease to search for predictive markers. A prospective study of 200,000 people would likely yield more than 5,000 cases of diabetes, based on current prevalence, in addition to numerous other features of the metabolic syndrome such as obesity, hyperinsulinemia, dyslipidemia, hypertension, and CVD. This population, if also studied with high-throughput, low-cost genotyping, could represent a major resource for physician-scientists to accelerate the incorporation of genetics into clinical medicine.

**Unanswered questions and future opportunities**

Many of the questions raised here cannot yet be answered. Will genome-wide association studies work in a way that genome-wide linkage mapping studies did not? Will the identification of more homogeneous subsets of patients be the key to making any real headway in either association or linkage mapping studies? Are non-genetic risk factors largely uniform, or are they perhaps as variable as genetic risk factors? Will we really get improvement in understanding with a “better” version of more of the same, or do we need to move to something qualitatively different? Genomic technology is advancing rapidly, and larger, higher-powered studies will soon be possible — these studies should allow us to address these questions. The challenge now is for clinical scientists to provide well-characterized populations with carefully recorded phenotypic and environmental data. This challenge will extend to the acquisition of new organizational skills to collate these data from many centers and provide integration with the large volume of genetic data soon to be generated (89). The opportunities are great for future diabetes genetic epidemiology research to provide clinically useful information, the most vital goal of the Human Genome Project.

**Acknowledgments**

The authors wish to acknowledge NIH grants DK049583 (“Metabolic Basis of NIDDM: A Sib Pair Analysis”) and DK58026 (“International Type 2 Linkage Consortium”); and Ping An, Richard Bergman, Ernesto Bernal-Mizrachi, Rudy Leibel, and Mike Province for helpful discussions and suggestions regarding the manuscript. Corentin Cras-Meneure helped with illustrations as well as discussions.

Address correspondence to: M. Alan Permutt, Washington University School of Medicine, 660 South Euclid Avenue, Campus Box 8127, St. Louis, Missouri 63110-1010, USA. Phone: (314) 362-8680; Fax: (314) 747-2692; E-mail: apерmutt@im.wustl.edu.


**The Journal of Clinical Investigation**

http://www.jci.org

Volume 115 Number 6 June 2005

1437

**review series**
tion of the transcription factor to a varying degree. Nucleic Acids Res. 28:430–437.
family with a nonsense mutation (R154X) in the