Diabetes constitutes a major public health problem. Although substantial progress has been made in defining the genetic risk for specific subtypes of diabetes (e.g., maturity-onset diabetes of the young), the majority of genetic risk of diabetes (for type 1 and type 2) remain unresolved. This review focuses on the current knowledge of the genetic basis of diabetes and its complications, specifically diabetic nephropathy (DN). Ultimately, identification of genes that contribute to risk (or protection) of diabetes and its complications will allow identification of patients who have diabetes and are at risk and targeted treatment/interventional strategies.

Genetics of Diabetes

Type 1 diabetes is the third most prevalent chronic disease of childhood, affecting up to 0.4% of children in some populations by age 30 yr, with an overall lifetime risk of nearly 1% (1,2). It is believed that a large proportion of cases of type 1 diabetes result from the autoimmune destruction of the pancreatic β cells, leading to complete dependence on exogenous insulin to regulate blood glucose levels (3). Type 1 diabetes is strongly clustered in families with an overall genetic risk ratio (the prevalence in siblings of a proband relative to the population prevalence, \( \lambda_s \)) of approximately 15 (4). (This compares with the less familial but more prevalent type 2 diabetes with \( \lambda_s \) of approximately 2). At least one locus that contributes strongly to this familial clustering resides within the MHC on chromosome 6p21, which accounts for nearly 40% of the observed familial clustering of type 1 diabetes, with a locus-specific genetic risk ratio (\( \lambda_s \)) of approximately 3 (5). In a recent analysis of data from three previous genomewide scans (United States, United Kingdom, and Scandinavia) as well as new families collected for the Type 1 Diabetes Genetics Consortium (http://www.t1dgc.org), 1435 multiplex families provided evidence for linkage of type 1 diabetes to the MHC (IDDM1), insulin (INS, IDDM2), a region that contains several genes, including CTLA4 (2q31-q33 [IDDM12 and IDDM7]) and seven other chromosome regions (6).

The genetic basis for type 2 diabetes has been difficult to resolve. Unlike type 1 diabetes, in which there seems to be an autoimmune process, type 2 diabetes is a disease of relative rather than absolute insulin deficiency. In type 2 diabetes, the pancreatic β cells become progressively less able to secrete sufficient insulin to maintain normal carbohydrate and lipid homeostasis (7). Furthermore, there is insulin resistance coupled with effects of aging, obesity, and reduced exercise (8). In addition, the strength of the genetic contribution to the cause of type 2 diabetes seems to be less than that for type 1 diabetes, with overall genetic risk ratio ranging from 2 to 4 (9). Despite the relatively low genetic risk ratio for type 2 diabetes, results of several genomewide scans have reported a number of chromosomal regions that may harbor genes that are involved in type 2 diabetes, with the most promising, replicated findings on chromosomes 1q21-q24, 2q37, 12q24 and 20 (reviewed in ref. [10]). The finding of linkage to type 2 diabetes in Mexican Americans to chromosome 2q37 (11) led to the characterization of susceptibility variants within the CAPN10 gene (12). Other type 2 diabetes linkage regions are being searched by development of consortia (the International 1q Consortium), and novel type 2 diabetes susceptibility genes may be identified in the future using the same approach. Several type 2 diabetes genes have been identified using investigations of candidate genes. Common variants in peroxisome proliferator–activated receptor-γ (13) and KCNJ11 (14) may predispose to type 2 diabetes with relatively low odds ratios (approximately 1.2). Variants that affect insulin signaling through IRS-1 (15,16) and glucose homeostasis PTPN1 (17,18) seem to be less common in frequency, suggesting that combinations of rare and common variants contribute to risk for type 2 diabetes in different pathways in different populations.

In summary, there is ample evidence that type 1 and type 2 diabetes are, in part, genetically determined. The magnitude of the genetic contribution to type 1 and type 2 diabetes differs dramatically (as shown by the genetic risk ratio \( \lambda_s \) for type 1 diabetes of 15, but \( \lambda_s \) is approximately 2 for type 2 diabetes). Furthermore, there is evidence that the genes that may contribute to risk for type 1 diabetes may differ from those that contribute to risk for type 2 diabetes. As we show below, the genetic risk for diabetic complications may be defined by yet another set of genes than those that modify risk for diabetes alone.
Indirect Evidence that Favors a Genetic Basis of DN

The complications of diabetes involve several organ systems and are often categorized as microvascular (retinopathy, nephropathy, and neuropathy) or macrovascular (cardiovascular and cerebrovascular). There is mounting evidence for the role of genetic factors in several diabetic complications, particularly DN, and cardiovascular and cerebrovascular complications of diabetes (19). DN remains the most common and rapidly increasing cause of ESRD in developed nations. Approximately 35% of European Americans with diabetes are at risk for the development of overt proteinuria, chronic renal insufficiency, and ESRD, with higher proportions observed among diabetic African Americans, Hispanic Americans, and Native Americans (20). Genetic predisposition or protection may be the most important DN risk determinants in type 1 diabetes. Only approximately one half of patients with poor glycemic control develop DN, whereas some patients do so despite relatively good control (21). Familial clustering of DN has been demonstrated by several investigators (22–24).

The majority of individuals with diabetes will develop background diabetic retinopathy; however, only one third of individuals with diabetes will develop overt kidney disease (25). The factors that initiate overt nephropathy are unclear, but consistent familial aggregation of kidney disease (albuminuria, creatinine clearance, and ESRD) and renal histologic changes have been observed frequently. The familial clustering of overt DN and diabetic ESRD has been observed widely in multiple racial and ethnic groups, with the earliest reports of familial aggregation of diabetic kidney disease in patients with type 1 diabetes (22,24). Family members with diabetes, even in the absence of clinical nephropathy, demonstrate similar patterns of glomerular involvement. Renal biopsy material from sibling pairs with type 1 diabetes were evaluated, and evidence of familial concordance in the percentage of the glomerulus occupied by mesangial matrix and the patterns of injury (presence of thickened glomerular basement membrane) were observed (26). In addition, consistent evidence across multiple studies also supports familial aggregation of albuminuria (27,28), GFR, and creatinine clearance (29–31).

Approaches to Identifying DN Susceptibility Genes

Multiple strategies exist for exploring the human genome for evidence of genetic factors that contribute to risk for complex disease (e.g., not determined by a defect in a single gene, often termed “Mendelian”). DN is an example of a complex human disease, with many genes and environmental risk factors contributing to susceptibility. These risk factors may be of differing magnitude, and not all may be required for initiation or progression of disease. Furthermore, the constellation of factors may vary from one population to another.

All approaches to detect disease susceptibility genes are based on demonstrating segregation of a trait (DN) with a measure of genetic variation. These approaches are dependent on the unobservable “true” genetic architecture of the genetic basis of disease as well as the available source of subjects (Figure 1). Susceptibility genes that are prevalent with large effect are unlikely to exist (single Mendelian genes with large effect are usually rare). Similarly, genes that are rare with small effect are unlikely to be detectable with current analytical strategies. In contrast, prevalent genes with small effect (common variant) or rare with large effect (rare variant) may be identified (32–34). Under a common variant hypothesis, association approaches (case-control, cohort studies) using a dense marker map have been proposed (whole genome association scans). Under a rare variant hypothesis, linkage approaches using extended families or sibling pairs (whole genome linkage scans) often have been performed (Figure 2).

Within either the common variant or rare variant hypotheses, both candidate gene evaluations and genome scans have been conducted. The characterization of the contribution of variants within candidate genes has evolved into two (often overlapping) strategies. The first strategy involves the study of polymorphisms within individual genes that should (based on the underlying biologic knowledge) have an effect on risk. The second is to explore a cluster of genes that may be involved in a pathophysiologic pathway, recognizing that multiple genes in the pathway may have recognized effects with each other and unrecognized effects on genes in the pathway and other genes (35–37). In each scenario, issues related to candidate gene selection, polymorphism evaluation within candidate genes, and analysis of multiple genes in pathways remain a significant and unresolved area of research.

Pathways and Candidate Genes Involved in DN

Current knowledge of the pathophysiology of DN suggests a complex interaction of multiple pathways associated with alterations in the balance between extracellular matrix production and removal, with interactions between genetic and environmental factors. These pathophysiologic concepts have been reviewed previously (38,39). The genes that contribute to variation and expression in the pathways are just now being explored in the context of complex metabolic networks. Complex-
Search for DN Susceptibility Genes

The search for genes that influence risk for DN can take place at many levels. The two classical approaches are the case-control and family-based designs. Within the case-control design, the vast majority of studies have used a candidate gene approach, whether interrogating individual candidate genes or genes related to specific biologic pathways. The case-control design determines the evidence of association between a genetic variant and the outcome of DN or an intermediate, quantitative phenotype (quantitative trait locus [QTL]). Within the family-based approach, the unit of analysis is often a pair of siblings, as parents are often unavailable for study (except in the case of type 1 DN). Several general analyses of family data have been used: The genome scan for linkage of DN, the candidate gene evaluation of linkage, and the family-based association approach. Recently, interest in mapping QTL for complex disease has been increasing. For example, the clinical phenotype (DN) is the result of multiple genetic and environmental factors; however, the underlying physiologic (quantitative) trait (e.g., urinary albumin excretion) may have stronger genetic effects (in the absence of those environmental factors that lead to overt disease). Thus, the genetic predisposition to nephropathy may be reflected in the level of urinary albumin excretion as a quantitative trait, and, with exposure to diabetes, high-normal levels of urinary albumin excretion may worsen to microalbuminuria and overt proteinuria. An implication of this hypothesis is that the genetic variation in the quantity of albumin excreted in the urine is determined by the same QTL as those genes that predispose to nephropathy but may not be dependent on the diabetic state. Results of mapping genes for DN or urinary albumin excretion, under this model, would lead to the same subset of genes in a common pathway.

In a genome screen, no a priori hypothesis is made as to the nature or the location of the specific genetic susceptibility genes for DN or its intermediate phenotypes. Instead, the genetic determinants of these phenotypes are tested on a large number of polymorphic DNA markers scattered at roughly equal intervals throughout the human genome. This approach has been applied successfully in a number of Mendelian disorders, with investigations proceeding from positive linkage results to linkage disequilibrium mapping to gene identification. Once linkage has identified a region in the genome that may contain a susceptibility gene, use of markers in the narrow region in cases and controls (for example) could find markers that are associated (rather than linked) to the disease. Current strategy uses haplotype-tagging single-nucleotide polymorphisms (SNP) derived from HapMap Project, which allows the definition of causative “haplotypes” that segregate with disease in the families.

It is, however, a much greater challenge to apply this approach to genetically complex disorders. For most complex diseases, it is likely that multiple genes influence disease ex-

Figure 2. Methods for genetic dissection of complex human traits. Solid symbols represent patients (diabetic nephropathy); open symbols represent control subjects (individuals with diabetes of long duration without nephropathy).
pression, making it difficult to isolate and characterize the effects of each disease-determining locus. Genes may interact (epistasis) or induce susceptibility independently (heterogeneity). In addition, complex diseases such as DN are influenced by environmental factors.

Despite these complicating factors, the genome screen strategy has been successful for type 1 diabetes (6), type 2 diabetes (12), Crohn’s disease (41), asthma (42), and schizophrenia (43). The likelihood of success has increased with advances in high-throughput genotyping, development of semiparametric genetic analysis methods, the potential to accumulate sufficient patient and family materials, and more refined definition and measurement of phenotypes. Although the most complex to execute, genome scans of appropriate family collections with significant prevalence of DN offer hope for identifying the primary genetic components of DN. For complex genetic entities such as DN, analysis of 1000 DNA (or more) from many different families is necessary. This corresponds to >400,000 genotypes. The microsatellite or SNP markers are not specifically associated with genes but are markers for the inheritance of each small segment of the human genome within which they lie. When the genome scan is completed, the result is a data set that follows the inheritance of every part of every chromosome in each individual in the study. The resulting statistical analyses would identify those chromosomal regions with the strongest evidence of linkage (assumed to contain the susceptibility genes). The advantage of the genome screen strategy is the comprehensive evaluation of the entire genome.

To date, only a few full descriptions of genome scans for DN have been published. The first genome scan that searched for nephropathy and retinopathy loci (microvascular complications) was performed in the Pima Indians (44). The strongest evidence for linkage was observed on chromosome 7q. Additional evidence for linkage was found on chromosomes 3 and 9 (DN and retinopathy) and 20 (DN only). A similar approach was used in families with type 1 diabetes (45), in which linkage was found to a region that contains the angiotensin II type 1 receptor gene (AGTR1). Other genome scans were completed recently. Members of a large Turkish kindred that contains multiple individuals who have type 2 DN (with multiple cousin–cousin marriages and significant inbreeding) were subjected to a genome scan, and a strong linkage peak was observed on chromosome 18 (46). Evaluation of these loci in the Pima Indians also showed evidence of linkage, although this region was not linked in the original Pima genome scan. The first genome scan of DN in African American families has only now been completed (47). Evidence for a nephropathy locus on chromosome 3 was detected in approximately 30% of the families, on chromosome 7p in approximately 40% of the families (those with longer duration of diabetes before diagnosis of ESRD), and on chromosome 18q in approximately 65% of the families (those with early age of diabetes diagnosis). Although these

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
<th>Position</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHE-1 (SLC9A1)</td>
<td>Na+/H + antiport-1</td>
<td>1p36.1-p35</td>
<td>NHE-1 activity is increased in type 1 diabetes with nephropathy (51,52)</td>
</tr>
<tr>
<td>TGF-β (TGF-β1, TGF-β2, TGF-β3)</td>
<td>Transforming growth factor-β (1, 2, 3)</td>
<td>19q12-q13.31, 1q41, 14q24</td>
<td>TGF-β mRNA, proteins (and TGF-β-receptor mRNA) identified in rodent glomerular cells (53–55)</td>
</tr>
<tr>
<td>GH1</td>
<td>Growth hormone</td>
<td>17q24.2</td>
<td>Diabetes causes decreased hepatic production of IGF-1; this decrease leads to an increased in growth hormone secretion (56–60)</td>
</tr>
<tr>
<td>IGF-1, IGF-1R</td>
<td>Insulin-like growth factor-1, insulin-like growth factor-1 receptor</td>
<td>12q22-q23, 15q26.3</td>
<td>VEGF expression increased in glomeruli of diabetic animals; diabetic rats that are treated with anti-VEGF antibody do not have hyperfiltration and reduced AER (61–63)</td>
</tr>
<tr>
<td>VEGF</td>
<td>Vascular endothelial growth factor</td>
<td>6p12</td>
<td></td>
</tr>
<tr>
<td>RAAS</td>
<td>ACE, AGT, AT1 (AGTR1), AT2 (AGTR2)</td>
<td>17q23, 1q42-q43, 3q21-q25, Xq22-q23</td>
<td>BP regulation and sodium homeostasis are crucial to progression of nephropathy (64–66)</td>
</tr>
</tbody>
</table>

aAER, albumin excretion ratio; RAAS, renin-angiotensin-aldosterone system; ACE, angiotensin-converting enzyme inhibitor; AGT, angiotensin-1; AT1, angiotensin II receptor type 1; AT2, angiotensin II receptor type 2.
genome scan studies have been performed in diverse populations with different study designs and analytical methods, investigators can take some encouragement from the results. Several of these linkage locations are consistent between studies. These complementary data increase the likelihood that several chromosomal locations mapped by linkage analysis reflect the presence of true DN genes.

A major limitation in gene identification has been adequate sample size with carefully and consistently phenotyped individuals. Several major multicenter studies are examining subjects and establishing renewable resources for genetic studies, particularly in the area of DN. These initiatives include GoKinD (Genetics of Kidney in Diabetes), FIND (Family Investigation of Nephropathy and Diabetes), and EDIC (Epidemiology of Diabetes In Complications) Genetics. The purpose of the GoKinD Study is to establish a repository of DNA and clinical information from adults with long-term type 1 diabetes, with or without kidney disease, along with their parents (trios). The GoKinD study has screened >5500 participants and obtained DNA to generate lymphoblast cell lines that are available for researchers. The FIND study also has established immortalized cell lines from approximately 6000 individuals in families that are enriched for the presence of overt DN (mainly from type 2 diabetes). A data from a sibling without diabetes has also been collected for mapping of type 2 diabetes genes. African American, European American, Mexican American, and American Indian families are being recruited in nearly equal numbers. The EDIC Genetics study has recruited nearly 1500 individuals with type 1 diabetes (including measures of nephropathy and retinopathy) and 3000 relatives. Thus, in the near future, a significant set of resources will become available for investigators of the genetic basis of type 1 and 2 diabetes and DN.

Summary

For effective reduction of the effect of diabetes and its complications and numerous public and personal health consequences, every effort must be made to minimize the development of diabetes and early progression to diabetic complications. A current limitation of progress in this regard is uncertainty on who is “at risk.” Although there are no recognized preventive measures for type 1 diabetes, numerous studies have demonstrated the relative efficacy of weight loss, diet, and exercise on normalization of glucose homeostasis, even in individuals with no clinical manifestations of type 2 diabetes (48,49). Unfortunately, despite significant relative risk reduction of these procedures, the absolute risk reduction remains low, resulting in the need to treat several hundred “healthy” (without diabetes) individuals to prevent a single person from progressing to overt disease event. The low absolute risk reduction in any single individual may also contribute to the underutilization of these procedures by patients and would require substantial increases in health care cost if it were administered to all patients who are at risk. It is also clear that many people develop diabetes and (ultimately) diabetic complications despite the absence of these risk factors. These facts underscore the need to develop more effective means to identify individuals who are most likely to develop diabetes and complications for targeted intervention and to continue to explore underlying mechanisms that contribute to their greater risk for disease. Genetic characterization is one important facet of an individual’s risk profile that has yet to be proved as an efficacious strategy, yet it has the potential to allow a targeted and effective approach to identify a smaller subset of patients for whom specific interventions/treatments should be most effective.

It is apparent from genetic studies of DN and studies of other diseases such as type 1 diabetes and type 2 diabetes that confirmation of results in different populations are of great value in trying to assess the validity of individual findings. Evidence of confirmation from multiple studies is the most widely used metric for accepting evidence of the importance of a gene in contribution to variation to risk for diabetes and its complications. Although this outcome would be comforting, there are (at least) two situations that limit the utility of population replication as a means for determining success of susceptibility gene identification. One limitation is if the “truth” in the genetics of diabetes or complications is in a series of population-specific private mutations within a common susceptibility framework. In this case, each population would be consistent with evidence for linkage at multiple sites, but the specific pattern of mutations could be different. A second limitation is that the same constellation of genetic factors contributes to susceptibility in all populations, but the subset of factors and effect sizes may differ. Thus, simple comparison of results across populations may provide erroneous “failure to replicate.”

The technology for carrying out large-scale molecular genetic studies is rapidly improving. In the near future, it will be possible to combine the power and the ease of association studies with comprehensive examination of the genome (whole genome association). One intriguing study that used elements of this approach has been reported (50) with 55,000 gene-based SNP loci. A single gene encoding solute carrier family 12 member 3 (SLC12A3), the gene associated with Gitelman’s syndrome, exhibited consistent evidence for association with DN susceptibility with a single SNP in intron 24 having statistically significant association. The results suggested that the Arg913Gin substitution in the SLC12A3 gene provided protection from development of DN. With increasing coverage of the genome (approximately 500,000 SNP), these studies may become a valuable tool in identification of common variants that enhance the search for diabetes and complication susceptibility genes.

References


