BACKGROUND: Type 2 diabetes (T2D) is a complex disorder that is affected by multiple genetic and environmental factors. Extensive efforts have been made to identify the disease-affecting genes to better understand the disease pathogenesis, find new targets for clinical therapy, and allow prediction of disease.

CONTENT: Our knowledge about the genes involved in disease pathogenesis has increased substantially in recent years, thanks to genomewide association studies and international collaborations joining efforts to collect the huge numbers of individuals needed to study complex diseases on a population level. We have summarized what we have learned so far about the genes that affect T2D risk and their functions. Although more than 40 loci associated with T2D or glycemic traits have been reported and reproduced, only a minor part of the genetic component of the disease has been explained, and the causative variants and affected genes are unknown for many of the loci.

SUMMARY: Great advances have recently occurred in our understanding of the genetics of T2D, but much remains to be learned about the disease etiology. The genetics of T2D has so far been driven by technology, and we now hope that next-generation sequencing will provide important information on rare variants with stronger effects. Even when variants are known, however, great effort will be required to discover how they affect disease risk.

Type 2 diabetes (T2D) is a common complex disorder with an increasing prevalence worldwide. In 2010 it was estimated that 6.6% of the world population of individuals 20–79 years old have diabetes, and that T2D constitutes approximately 90% of diabetes cases (1). This number is expected to increase epidemiologically as a consequence of an aging population and changes in lifestyle.

Prevalence and Heritability of T2D

The risk of developing T2D is determined by both genetic and environmental factors. The genetic component can be analyzed by comparing the risk of developing disease between relatives of patients with T2D and the background population, often referred to as sibling relative risk (λS), which is around 3 in most populations (2).

The lifetime risk of developing T2D is 40% for individuals who have 1 parent with T2D and almost 70% if both parents are affected (3, 4). Interestingly, the risk is higher if the mother, rather than the father, is affected (4). In addition, the concordance rate of T2D in monozygotic twins is about 70% whereas the concordance in dizygotic twins is only 20%–30% (5, 6). Large differences in prevalence between ethnic groups also exist and seem to depend on genetic factors. In Sweden, immigrants from the Middle East have a 2- to 3-fold increased risk of T2D compared to native Swedes. In addition, these immigrants seem to have a slightly different form of diabetes with an earlier onset and lower C-peptide concentrations than Swedish patients. It is also more common for patients from the Middle East to have first-degree relatives with T2D (7).

The Role of the Environment

Although genetic factors are important in T2D, it is obvious that environmental factors play a key role in the development of the disease. The change toward a more affluent Western lifestyle that has taken place during the last 50 years has started a worldwide epidemic increase in the prevalence of T2D and obesity defined by body mass index (BMI in kg/m²). It is estimated that nearly 80% of individuals older than 15 years in the US and more than 60% of the same popu-
studies of T2D, only 2 genes have been reported to have T2D. Although great efforts have been put into linkage identifying genes that cause complex diseases, such as known inheritance mode, it has been less useful for mapping genetic diseases with strong penetrance and a Although this strategy has been very successful for causing variant is in LD with the genotyped marker. 

cause it was inherited from the same parent, more often 
targets, disease loci can be mapped on a genomewide level. 
Finding that affected family members share a certain marker that is identical by descent, i.e., identical because it was inherited from the same parent, more often than expected by chance, is evidence that a disease-causing variant is in LD with the genotyped marker. Although this strategy has been very successful for mapping genetic diseases with strong penetrance and a known inheritance mode, it has been less useful for identifying genes that cause complex diseases, such as T2D. Although great efforts have been put into linkage studies of T2D, only 2 genes have been reported to have been identified by linkage: calpain 10 (CAPN10)\(^3\) and transcription factor 7-like 2 (T-cell specific, HMGB-box) (TCF7L2) (Table 1 and Table 1 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol57/issue2).

In the case of TCF7L2, a T2D locus was mapped to chromosome 10q in both an Icelandic and a Mexican-American population (10, 11). This region was later fine-mapped in the Icelandic population by use of 228 microsatellite markers covering a 10.5-Mbp region, pinpointing the locus to intron 3 of the TCF7L2 gene (12). The association between T2D and a number of single-nucleotide polymorphisms (SNPs) in the TCF7L2 gene has since been confirmed in numerous studies in different ethnic groups. The risk allele confers a relative risk of approximately 1.4 compared to homozygous carriers of the nonrisk allele, making this the strongest association with T2D by far (13). The mechanisms by which TCF7L2 affects diabetes susceptibility are still not completely understood. The transcription factor-7-like 2, encoded by TCF7L2, is part of a bipartite transcription factor, which also included β-catenin, that is a key effector in the Wnt-signaling pathway. In a study of Scandinavian individuals, we showed that the T-allele of SNP rs7903146 is associated with risk of T2D, impaired insulin secretion, incretin effects, and an enhanced rate of hepatic glucose production (14). We also found that homozygous carriers of the risk allele had a 5-fold increase in TCF7L2 mRNA expression in their islets (14). In addition, TCF7L2 has been suggested to regulate progluca-

\(^3\) Human genes: CAPN10, calpain 10; TCF7L2, transcription factor 7-like 2 (T-cell specific, HMGB-box); PPARG, peroxisome proliferator-activator receptor-gamma gene; IRS1, insulin receptor substrate 1; KCNJ11, potassium inwardly-rectifying channel, sulfonylurea 1, member 11; WFS1, Wolfram syndrome 1 ( wolframin); HNF1A, HNF1 homeobox A; and HNF1B, HNF1 homeobox B; HHEX, hematoepoietically expressed homeobox; SLC30A8, solute carrier family 30 (zinc transporter), member b; CDK2A2/CDK2NB1, cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)/cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4); IGF2BP2, insulin-like growth factor 2 mRNA binding protein 2; CDK4A1, CDK5 regulatory subunit alpha associated protein 1-like 1; FTO, fat mass and obesity associated; KCNQ1, potassium voltage-gated channel, KQT-like subfamily, member 1; PPARD, protein tyrosine phosphatase, receptor type, D; SRR, serine racemase; YY1AP1, YY1 associated protein 1; JAZF1, JAZF zinc finger 1; CDC123, cell division cycle 123 homolog; CAMK1D, calcium/calmodulin-dependent protein kinase II; TSPAN4-LGR5, tetraspanin 8–leucine-rich-repeat-containing G protein-coupled receptor 5; THADA, thyroid adenoma associated; ADAMT59, ADAM metallopeptidase with thrombospondin type 1 motif, 9; NOTCH2, notch 2; RBM5, RNA binding motif, single stranded interacting protein 1; MTRNR1, melatonin receptor 1B; BCL11A, B-cell CLL/lymphoma 11A (zinc finger protein); ZBED3, zinc finger, BED domain containing 3; KLF14, Krueppel-like factor 14; TPS3NIP1, tumor protein p53 inducible nuclear protein 1; TLE4, transducin-like enhancer of split 4; CHIC2D9, coiled-coil-helix-coiled-helix domain containing 9; ARAP1, ArfGAP with RH domain, ankryin repeat and PH domain 1; HMGAA2, high mobility group A1-hook 2; ZFAND6, zinc finger, AN1-type domain 6; PRK1, protein regulator of cytokinesis 1; DUSP9, dual specificity phosphatase 9; ADCYS, adenylate cyclase 5; MAD2L, MAP-kinase activating death domain; CRY2, cryptochrome 2 (photolyase-like); ADRA2A, adrenergic, alpha-2A-, receptor; FADS1, fatty acid desaturase 1; PROX1, prosper homeobox 1; SLC2A2, solute carrier family 2 (facilitated glucose transporter), member 2; GLIS3, GLIS family zinc finger 3; CDC40B2, C2 calcium-dependent domain containing protein 48; IGFI, insulin-like growth factor 1 (somatomedin C); GCK, glucokinase (hexokinase 4); DGKB-TMEM195, diacylglycerol kinase, beta 90kDa–transmembrane protein 195; GCKR, glucokinase (hexokinase 4) regulator; GPIR, gastric inhibitory polypeptide receptor; FADS, MACR, melanocortin 4 receptor; TME18, transmembrane protein 18; GNPDA2, glucosamine-6-phosphate deaminase 2; SH2B1, SH2B adaptor protein 1; MTH2, mitochondrial carrier homolog 2; NEG1, neuronal growth regulator 1; AIF1, allograft inflammatory factor 1; NCR3, natural cytotoxicity triggering receptor 3; BCL2ND3, BCL2ND3 domain containing: FAIM2, Fas apoptotic inducing molecule 2; PPARA, peroxisome proliferator-activator receptor-alpha; LEP, leptin; PDX1, pancreatic and duodenal homeobox 1; ADAM30, ADAM metallopeptidase domain 30; TRA2B, transformer 2 beta homolog; ETFV, ets variant 5; DGKG, diacylglycerol kinase, gamma 90kDa.
Table 1. Genetic variants associated with T2D.

<table>
<thead>
<tr>
<th>Variants</th>
<th>Nearest gene</th>
<th>Full gene name</th>
<th>OR T2D</th>
<th>Study type*</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4607103,rs6795735</td>
<td>ADAMTS9</td>
<td>ADAM metallopeptidase with thrombospondin type 1 motif, 9</td>
<td>1.09–1.1</td>
<td>MA</td>
<td>Voight et al. (41), Zeggini et al. (53)</td>
</tr>
<tr>
<td>rs1552224</td>
<td>ARAP1 (CENTD2)</td>
<td>ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 1</td>
<td>1.13–1.14</td>
<td>MA</td>
<td>Voight et al. (41)</td>
</tr>
<tr>
<td>rs243021</td>
<td>BCL11A</td>
<td>B-cell CLL/lymphoma 11A (zinc finger protein)</td>
<td>1.08–1.09</td>
<td>MA</td>
<td>Voight et al. (41)</td>
</tr>
<tr>
<td>rs2975760,rs3792267</td>
<td>CAPN10</td>
<td>Calpain 10</td>
<td>1.17</td>
<td>LA</td>
<td>Weedon et al. (20)</td>
</tr>
<tr>
<td>rs12779790</td>
<td>CDC123/CAMK1D</td>
<td>Cell division cycle 123 homolog–calcium/calmodulin-dependent protein kinase ID</td>
<td>1.09</td>
<td>MA</td>
<td>Voight et al. (41), Zeggini et al. (53)</td>
</tr>
<tr>
<td>rs7754840,rs1044033</td>
<td>CDKAL1</td>
<td>CDK5 regulatory subunit associated protein 1-like 1</td>
<td>1.12–1.25</td>
<td>GWAS</td>
<td>Saxena et al. (32), Scott et al. (33), Voight et al. (41), Steinhorsdottir et al. (44)</td>
</tr>
<tr>
<td>rs13292136</td>
<td>CHCHD9/TLE4</td>
<td>Coiled-coil-helix-coiled-helix domain containing 9/transducin-like enhancer of split 4</td>
<td>1.11–1.20</td>
<td>MA</td>
<td>Voight et al. (41)</td>
</tr>
<tr>
<td>rs10811661,rs10965250</td>
<td>CDKN2A/2B</td>
<td>Cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)–cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)</td>
<td>1.20</td>
<td>GWAS</td>
<td>Saxena et al. (32), Scott et al. (33), Voight et al. (41)</td>
</tr>
<tr>
<td>rs5945326</td>
<td>DUSP9</td>
<td>Dual specificity phosphatase 9</td>
<td>1.27</td>
<td>MA</td>
<td>Voight et al. (41)</td>
</tr>
<tr>
<td>rs1111875,rs5015480</td>
<td>HHEX</td>
<td>Hematopoietically expressed homebox</td>
<td>1.13–1.18</td>
<td>GWAS</td>
<td>Saxena et al. (32), Voight et al. (41), Sladek et al. (43)</td>
</tr>
<tr>
<td>rs531343</td>
<td>HMGA2</td>
<td>High mobility group AT-hook 2</td>
<td>1.10–1.20</td>
<td>MA</td>
<td>Voight et al. (41)</td>
</tr>
<tr>
<td>rs7957197</td>
<td>HNF1A</td>
<td>HNF1 homeobox A</td>
<td>1.07–1.14</td>
<td>MA</td>
<td>Voight et al. (41)</td>
</tr>
<tr>
<td>rs1920792,rs7501939,rs757210,rs4430796</td>
<td>HNF1B</td>
<td>HNF1 homeobox B</td>
<td>1.1–1.17</td>
<td>CS</td>
<td>Winckler et al. (115), Voight et al. (41), Gudmundsson et al. (42)</td>
</tr>
<tr>
<td>rs2334499</td>
<td>HCCA2</td>
<td>YY1 associated protein 1</td>
<td>1.35</td>
<td>GWAS</td>
<td>Kong et al. (52)</td>
</tr>
<tr>
<td>rs4402960,rs1470579</td>
<td>IGF2BP2</td>
<td>Insulin-like growth factor 2 mRNA binding protein 2</td>
<td>1.14</td>
<td>GWAS</td>
<td>Saxena et al. (32), Scott et al. (33), Voight et al. (41)</td>
</tr>
<tr>
<td>rs2943641,rs7578326</td>
<td>IRS1</td>
<td>insulin receptor substrate 1</td>
<td>1.09–1.12</td>
<td>CS</td>
<td>Almind et al. (25), Voight et al. (41)</td>
</tr>
<tr>
<td>rs864745,rs849134</td>
<td>JAZF1</td>
<td>JAZF zinc finger 1</td>
<td>1.12–1.13</td>
<td>MA</td>
<td>Voight et al. (41), Zeggini et al. (53)</td>
</tr>
<tr>
<td>rs5219</td>
<td>KCNJ11</td>
<td>Potassium inwardly-rectifying channel, subfamily J, member 11</td>
<td>1.09–1.14</td>
<td>CS</td>
<td>Hani et al. (28), Saxena et al. (32), Voight et al. (41)</td>
</tr>
<tr>
<td>rs2237892,rs163184,rs231362</td>
<td>KCNQ1</td>
<td>Potassium voltage-gated channel, KQT-like subfamily, member 1</td>
<td>1.08–1.23</td>
<td>GWAS</td>
<td>Voight et al. (41), Unoki et al. (48), Yasuda et al. (49)</td>
</tr>
<tr>
<td>rs972283</td>
<td>KLF14</td>
<td>Kruppel-like factor 14</td>
<td>1.07–1.1</td>
<td>MA</td>
<td>Voight et al. (41)</td>
</tr>
<tr>
<td>rs10923931</td>
<td>NOTCH2/ADAM30</td>
<td>Notch 2–ADAM metallopeptidase domain 30</td>
<td>1.14</td>
<td>MA</td>
<td>Voight et al. (41), Zeggini et al. (53)</td>
</tr>
<tr>
<td>rs1801282,rs13081389</td>
<td>PPARG</td>
<td>Peroxisome proliferator-activated receptor gamma</td>
<td>1.14–1.24</td>
<td>CS</td>
<td>Deeb et al. (21), Saxena et al. (32), Voight et al. (41)</td>
</tr>
</tbody>
</table>

Continued on page 244
gon gene expression, and thus glucagon-like peptide 1 (GLP-1) synthesis, from intestinal endocrine L cells (15). However, studies in humans have shown that risk-allele carriers have concentrations of GLP-1 that are within reference intervals but have impaired insulin secretion in response to GLP-1 infusion, indicating that the effect lies at the level of GLP-1 action on β-cells rather than on GLP-1 secretion (14, 16).

The other gene mapped by linkage analysis is a locus on chromosome 2 that was first mapped in 1996 by Hanis et al. (17). In 2000 the locus was fine mapped and the causative gene shown to be CAPN10, the gene for calpain 10, a cysteine protease with largely unknown functions in glucose metabolism (18). Despite a number of negative replication studies, several meta-analyses have shown consistent association of CAPN10 with T2D (19, 20). Nevertheless, none of the large genomewide association studies (GWAS) has identified CAPN10 as being associated with T2D.

### Candidate Gene Studies

Identification of disease genes can also be made on the basis of association testing in populations rather than in families. Because of the short LD stretches in unrelated individuals, a very large number of markers are needed to perform mapping on a genomewide level. Until recently, the only feasible strategy was to study candidate genes that had high probability of affecting the studied trait based on the known function of the gene. Although a number of studies have reported association of functional or positional candidates, only 6 genes have been consistently associated with T2D: peroxisome proliferator-activated receptor gamma (PPARG), insulin receptor substrate 1 (IRS1), potassium inwardly-rectifying channel, subfamily J, member 11 (KCNJ11), Wolfram syndrome 1 (wolframin) (WFS1), HNF1 homeobox A (HNF1A), and HNF1 homeobox B (HNF1B) (Table 1).

### Table 1. Genetic variants associated with T2D. (Continued from page 243)

<table>
<thead>
<tr>
<th>Variants</th>
<th>Nearest gene</th>
<th>Full gene name</th>
<th>OR T2D</th>
<th>Study typea</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs8042680</td>
<td>PRG1</td>
<td>Protein regulator of cytokinesis 1</td>
<td>1.07–1.1</td>
<td>MA</td>
<td>Voight et al. (41)</td>
</tr>
<tr>
<td>rs17584499</td>
<td>TPTRD</td>
<td>Protein tyrosine phosphatase, receptor type, D</td>
<td>1.57</td>
<td>GWAS</td>
<td>Tsai et al. (51)</td>
</tr>
<tr>
<td>rs7593730</td>
<td>RBMS1</td>
<td>RNA binding motif, single stranded interacting protein 1</td>
<td>1.11</td>
<td>MA</td>
<td>Qi et al. (54)</td>
</tr>
<tr>
<td>rs13266634, rs11558471, rs3802177</td>
<td>SLC30A8</td>
<td>Solute carrier family 30 (zinc transporter), member 8</td>
<td>1.12–1.15</td>
<td>GWAS</td>
<td>Saxena et al. (32), Voight et al. (41), Sladek et al. (43), Dupuis et al. (55)</td>
</tr>
<tr>
<td>rs391300</td>
<td>SRR</td>
<td>Serine racemase</td>
<td>1.28</td>
<td>GWAS</td>
<td>Tsai et al. (51)</td>
</tr>
<tr>
<td>rs7903146, rs12255372, rs4506565, rs11196218</td>
<td>TCF7L2</td>
<td>Transcription factor 7-like 2 (T-cell specific, HMG-box)</td>
<td>1.4</td>
<td>LA</td>
<td>Grant et al. (12), Saxena et al. (32), Voight et al. (41)</td>
</tr>
<tr>
<td>rs7578597, rs11899863</td>
<td>THADA</td>
<td>Thyroid adenoma associated</td>
<td>1.15–1.17</td>
<td>MA</td>
<td>Voight et al. (41), Zeggini et al. (53)</td>
</tr>
<tr>
<td>rs896854</td>
<td>TP53INP1</td>
<td>Tumor protein p53 inducible nuclear protein 1</td>
<td>1.06–1.1</td>
<td>MA</td>
<td>Voight et al. (41)</td>
</tr>
<tr>
<td>rs7961581, rs4760790</td>
<td>TSPAN8-LGR5</td>
<td>Tetraspanin 8–leucine-rich repeat-containing G protein-coupled receptor 5</td>
<td>1.11</td>
<td>MA</td>
<td>Voight et al. (41), Zeggini et al. (53)</td>
</tr>
<tr>
<td>rs10010131, rs6446482, rs1801214</td>
<td>WFS1</td>
<td>Wolfram syndrome 1 (wolframin)</td>
<td>1.11–1.13</td>
<td>CS</td>
<td>Minton et al. (35), Sandhu et al. (36), Voight et al. (41)</td>
</tr>
<tr>
<td>rs4457053</td>
<td>ZBED3</td>
<td>Zinc finger, BED-type containing 3</td>
<td>1.08–1.16</td>
<td>MA</td>
<td>Voight et al. (41)</td>
</tr>
<tr>
<td>rs11634397</td>
<td>ZFAND6</td>
<td>Zinc finger, AN1-type domain 6</td>
<td>1.06–1.11</td>
<td>MA</td>
<td>Voight et al. (41)</td>
</tr>
</tbody>
</table>

* MA, metaanalysis; LA, linkage analysis; CS, candidate study.

The PPARG gene, encoding the nuclear receptor PPAR-γ, was the first gene reproducibly associated with T2D (21). The PPAR-γ receptor is a molecular target for thiazolidinedione compounds, a class of insulin-sensitizing drugs used to treat T2D, a characteristic that makes this a very strong candidate gene. The gene expressed in adipose tissue has an extra exon B and a substitution of a proline for alanine at position 12 of this protein, which is seen in
about 15% of the European population and has been shown to be associated with increased transcriptional activity, increased insulin sensitivity, and protection against T2D (21). Although the initial report was followed by a number of negative studies, in a family-based study, with the use of a transmission disequilibrium test, we observed excess transmission of the Pro allele to affected offspring (22). It soon became clear that most negative studies had been underpowered, and that combining the data from all published studies in a metaanalysis yielded strong support for association between the Pro12Ala variant and T2D (22–24).

**IRS1** encodes a protein that is phosphorylated by insulin receptor tyrosine kinase and is essential to insulin function. The association between a Gly972Arg polymorphism and T2D was reported in 1993 (25) but could not be replicated consistently by subsequent studies. It was, however, found to be strongly associated with T2D in a recent GWAS that also showed an association with reduced basal concentrations of IRS1 protein and decreased insulin induction of IRS1-associated phosphatidylinositol-3-OH kinase activity in human skeletal muscle biopsies (26).

**Common Variants in Genes Causing Monogenic Disorders Increase Risk of T2D**

**KCNJ11** encodes a protein, Kir6.2, that together with SUR1 (sulfonylurea receptor 1) forms an ATP-sensitive potassium channel that regulates membrane potential, and thereby glucose-dependent insulin secretion, in pancreatic β-cells. Activating mutations in this gene cause severe neonatal diabetes (27). A Glu23Lys polymorphism has been associated with modest impairment of insulin secretion and T2D (28).

Large-scale studies and metaanalyses have confirmed the association and shown that the lysine variant increases activation of the channel by 2-fold, resulting in a 1.15 times higher risk of developing T2D (29–31). The association has also been confirmed in GWAS (32–34).

**WFS1** encodes Wolframin, a protein that is defective in individuals suffering from the Wolfram syndrome, characterized by diabetes insipidus, juvenile diabetes, optic atrophy, and deafness. WFS1 was first suggested to be associated with T2D in a small family-based association study (35). The gene was later included as a candidate in a study of 1536 SNPs in 84 candidate genes. Of these 84 genes only WFS1 was associated with T2D and could be replicated in 9533 cases and 11 389 controls (36). The association has also been confirmed in a large metaanalysis comprising 14 000 cases and 16 000 controls (37).

Mutations in the HNF1A gene cause the most common form of monogenic diabetes, maturity onset diabetes of the young 3 (MODY3). It was long debated whether more common variants in the HNF1A gene could contribute to T2D (38, 39), but several pieces of evidence suggest that the variants can increase risk of T2D: common variants result in decreased transcriptional activity (39), they have been shown to predict future risk of T2D in large prospective studies (40), and they have been associated with T2D in a recent large GWAS metaanalysis (the Diabetes Genetics, Replication and Metaanalysis (DIAGRAM)-plus, consisting of the DIAGRAM consortium complemented with additional GWAS cohorts from Diabetes Epidemiology: Collaborative Analysis of Diagnostic Criteria in Europe (deCODE) genetics, the Diabetes Gene Discovery Group, the Cooperative Health Research in the Region of Augsburg (KORA) group, the Rotterdam study, and the European Special Population Research Network (EUROSPAN)) (41).

HNF1B, also known as transcription factor 2, is another transcription factor that can heterodimerize with HNF1A. Mutations in the HNF1B gene cause MODY5, which in affected individuals is associated with early-onset diabetes and also urogenital and renal pathology. The HNF1B gene is important for pancreatic islet development and function and was first identified as a T2D gene in a study in which common variants in known MODY genes were tested for association with T2D (38). A second variant of this gene was found to be associated with T2D in a GWAS (42).

**Whole-Genome Association Studies**

Rapid improvement in high-throughput technology for SNP genotyping, which has allowed simultaneous genotyping of hundreds of thousands of SNPs, has opened new possibilities for association studies. The HapMap project provided another important tool, and demonstrated that genotyping of approximately 500 000 SNPs is enough to cover about 75% of the common variants (minor allele frequency >5%) in the genome.

An important breakthrough in our understanding of the genetics of T2D occurred in 2007 with the publication of the results of several GWAS in which DNA chips were used with >500 000 SNPs to perform genotyping in large numbers patients with T2D as well as healthy controls (32–34, 43, 44).

The results of the first GWAS on T2D were published in February 2007 by Sladek et al. (43). This GWAS conducted in 661 cases and 614 controls from France and led to the identification of 2 new diabetes loci: hematopoietically expressed homeobox (HHEX) and solute carrier family 30 (zinc transporter), member
A few months later this study was followed by another 4 studies, all performed in European populations with a case-control setup, including the Wellcome Trust Case Control Consortium study in 1924 cases and 2938 controls from the UK (45); the Diabetes Genetics Initiative study in 1464 cases and 1467 controls from Sweden and Finland (32); the FUSION (Finland–United States Investigation of NIDDM [non–insulin-dependent diabetes mellitus]) genetics study in 1161 cases and 1174 controls from Finland (33); and a study by Steinthorsdottir et al. in 1399 cases and 5275 controls from the Icelandic population (44).

The first 3 studies shared results before publication and considered only positive results that were seen and replicated in all 4 studies. Two new loci were identified, cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)/cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4) (CDKN2A/CDKN2B) and insulin-like growth factor 2 mRNA binding protein 2 (IGFBP2) and several of the previously known loci were confirmed. CDK5 regulatory subunit associated protein 1-like 1 (CDKAL1) was independently identified as a new T2D locus in all 4 studies.

About the same time, the fat mass and obesity associated (FTO) gene was identified as a major susceptibility locus for obesity, and therefore indirectly also for T2D (46, 47). Because the effect of FTO on diabetes is through obesity, this gene was not detected in the GWAS of T2D that were matched on BMI.

Results of the first GWAS on T2D in non-European populations were published in 2008; both of these studies used multistage approaches (48, 49). Yasuda et al. first genotyped 100 000 SNPs in 187 T2D cases and 2 different control populations, each including 752 individuals, thereby identifying 2800 candidate SNPs for follow-up in replication cohorts. In total 19 930 individuals of Asian and European descent were genotyped for the strongest signals. Unoki et al. genotyped 268 068 SNPs in 194 cases and 1558 controls, and then the most highly associated SNPs in another 4924 individuals with T2D and 2618 controls, all of Japanese origin (48). In both studies the potassium voltage-gated channel, KQT-like subfamily, member 1 (KCNQ1) was identified, a new T2D gene that encodes the pore-forming subunit of the $I_{Ks}$ channel (or voltage-gated potassium channel), and results of both studies showed that KCNQ1 is expressed in the pancreas. The risk allele of the SNP rs2237895 has also been found to be associated with reduced insulin secretion in Swedish and Finnish cross-sectional and prospective cohorts, conferring a T2D risk of 1.23 in Swedes and Finns (50) compared to 1.31 in Asians (48).

Another study in an Asian population was performed by Tsai et al. in 2010 (51). These authors conducted a 2-stage GWAS, including 2798 cases and 2367 controls from a Han Chinese population, and identified 2 novel T2D loci: rs17584499 in, protein tyrosine phosphatase, receptor type, D (PTPRD) and rs391300 in serine racemase (SRR).

In 2009 Kong et al. performed a genomewide association study in the Icelandic population by analyzing parental-origin–specific effects, thereby identifying a new candidate SNP close to YY1 associated protein 1 (YY1API) (52). The marker rs2334499 was only weakly associated with T2D in a standard case-control analysis. However, when parental origin was taken into account the paternally inherited allele increased risk of T2D with genomewide significance. Interestingly, the maternally inherited allele showed nominally significant evidence of a protective effect on T2D.

Large-Scale Metaanalysis Studies

The first GWAS metaanalysis of T2D was performed by the DIAGRAM consortium in 2008 (53). The authors combined analysis from 3 GWAS, the Diabetes Genetics Initiative, FUSION, and Wellcome Trust Case Control Consortium, and extended SNP coverage by imputing untyped SNPs using a 3 stage approach. In the first stage 2 202 892 SNPs were tested for association followed by 69 SNPs tested in replication cohorts in stage 2, and finally 11 SNPs were followed up in the deCODE, KORA, Danish, HUNT (Nord-Trøndelag Health Study), NHS (National Health Service), GEM (Genetics of Energy Metabolism) Consortium [CCC (Cambridgeshire Case-Control), EPIC (European Prospective Investigation into Cancer), ADDITION (Anglo-Danish-Dutch Study of Intensive Treatment in People with Screen Detected Diabetes in Primary Care)/Ely, Norfolk)] and METSIM (Metabolic Syndrome in Men) study samples (effective sample size, 53 975) (53). The DIAGRAM consortium could thereby identify 6 novel loci, including JAZF zinc finger 1 (JAZF1), cell division cycle 123 homolog–calcium/calmodulin-dependent protein kinase ID (CDCl23-CAMKID), tetraspanin 8–leucine-rich repeat-containing G protein-coupled receptor 5 (TSPAN8-LGR5), thyroid adenoma associated (THADA), ADAM metallopeptidase with thrombospondin type 1 motif, 9 (ADAMTS9), and notch 2 (NOTCH2) (Table 1).

Qi et al. identified a T2D-associated variant in the RNA binding motif, single stranded interacting protein 1 (RBMS1) gene on chromosome 2 in a nested case-control study that combined 2 prospective studies in the discovery phase and 11 studies in the validation
phase, reaching a total number of 13,462 cases and 76,787 controls (54).

In June 2010 the DIAGRAM consortium was complemented with additional GWAS cohorts from deCODE genetics, the Diabetes Gene Discovery group, the KORA group, the Rotterdam study, and EUROSPAN to form DIAGRAM plus, with a higher effective sample size (n/\text{H11005} 22,044) (41).

A total of 2,426,886 SNPs were tested in stage 1, and those showing an association of \( P < 10^{-8} \) were selected for phase 2, in which in silico data from 3 GWAS (Atherosclerosis Risk in Communities study, Nurses’ Health study, and Framingham Heart study) not included in the primary metaanalysis, for a maximum possible stage 2 sample size of 34,412 cases and 59,925 controls, all individuals of European descent. Fourteen loci reached genome wide significance: 2 previously identified genes, [melatonin receptor 1B (\text{MTNR1B}) and \text{IRS1}] and 12 novel loci [\text{B-cell CLL/lymphoma 11A (zinc finger protein) (BCL11A)}; zinc finger, BED-type containing 3 (\text{ZBED3}); Kruppel-like factor 14 (\text{KLF14}); tumor protein p53 inducible nuclear protein 1 (\text{TP53INP1}); transducin-like enhancer of split 4 (\text{TLE4}); coiled-coil-helix-coiled-coil-helix domain containing 9 (\text{CHCHD9}); \text{KCNO1}; ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 1 (\text{ARAP1}); high mobility group AT-hook 2 (\text{HMGA2}); \text{HNF1A}; zinc finger, AN1-type domain 6 (\text{ZFAND6}); protein regulator of cytokinesis 1 (\text{PRC1}); and dual specificity phosphatase 9 (\text{DUSP9})] (Table 1).

\text{MTNR1B} had previously been identified by the Meta-analysis of Glucose and Insulin-Related Traits (MAGIC) consortium as a fasting-glucose locus but was also found to affect T2D (41).

The most recent MAGIC effort involved metaanalysis of 21 GWAS cohort data for 46,186 nondiabetic individuals of European decent, including fasting-glucose information (55). An initial 25 lead SNPs were followed up in 76,558 additional study participants, and 9 new loci were found to be associated (\( P < 5 \times 10^{-8} \)) with fasting glucose (and/or HOMA-B) [adenylate cyclase 5 (\text{ADCY5}); MAP-kinase activating death domain (\text{MADD}); cryptochrome 2 (photolyase-like) (\text{CRY2}); adrenergic, alpha-2A-, receptor (\text{ADRA2A}); fatty acid desaturase 1 (\text{FADS1}); prospero homeobox 1 (\text{PROX1}); solute carrier family 2 (facilitated glucose transporter), member 2 (\text{SLC2A2}); GLIS family zinc finger 3 (\text{GLIS3}); and C2 calcium-dependent domain containing 4B (\text{C2CD4B})] and 1 was found to be associated with fasting insulin and HOMA-IR concentrations [insulin-like growth factor 1 (\text{IGF1})]. Five of these loci [glucokinase (hexokinase 4) (\text{GCK}); diacylglycerol kinase, \( \beta \)-90kDa–transmembrane protein 195 (\text{DGKB-TMEM195}); glucokinase (hexokinase 4) regulator, beta 90kDa–transmembrane protein 195 (\text{DGKB-TMEM195}); glucokinase (hexokinase 4) regulator (\text{GCKR}); \text{ADCY5}; \text{PROX1}] were also associated with T2D risk when tested in 40,655 diabetic cases and 87,022 controls (Table 2).

**FUNCTIONS OF ASSOCIATED GENES**

The loci recently identified by GWAS and metaanalysis can be subgrouped on the basis of their association with phenotypes with a key role in T2D etiology.

**INSULIN SECRETION**

\text{MTNR1B} has been reported to be associated with fasting glucose and T2D risk itself in 3 publications based on GWAS (56–58). Melatonin works as a chronobiotic factor, adjusting the timing of the biological clock (59, 60). Its receptors are present in the pancreas, and melatonin is proposed to contribute to the nocturnal lowering of insulin in humans. Recently Lyssenko et al. have shown that the \text{MTNR1B} risk genotype is associ-
ated with impaired early insulin release to both oral and intravenous glucose and that insulin secretion deteriorates over time in the risk allele carriers (56).

Mulder et al. proposed a mechanism of altered melatonin signaling in T2D: the risk allele increases the mRNA concentration of MTNR1B in β-cells, thereby decreasing the cAMP concentrations via inhibitory G proteins. In addition, the insulinoceptive pressure (e.g., that exerted by GLP-1) may be diminished, resulting in a decrease in insulin secretion, contributing to increased fasting plasma glucose, lower circulating insulin concentrations, and increased risk of T2D (61).

Lyssenko et al. reported that genetic variants in the IGF2BP2, SLC30A8, and CDKN2A/CDKN2B gene/loci resulted in impaired β-cell function over time in a prospective study (62). Grarup et al. reported that variants near the JAZF1, CDC123/CAMK1D, and TSPAN8 loci are associated with impaired glucose-stimulated insulin secretion (63).

The CDKAL1 gene is expressed both in skeletal muscle and pancreatic islets but has a largely unknown function (33, 34). However, homozygous risk allele carriers were found to have 20% reduced insulin response, suggesting that the risk to T2D is conferred through low insulin secretion.

The recently identified genes from the MAGIC consortium were replicated in a white cohort (European and US; n = 29 084) by Ingelsson et al. (64), who found that fasting-glucose–increasing alleles of the MADD, gastric inhibitory polypeptide receptor (GIPR), GCK, FADS, DGKB, PROX1, TCF7L2, SLC30A8, and C2CD4B were associated with either abnormal insulin processing or secretion, whereas GCKR and IGF1 were associated with oral glucose tolerance test–based disposition indices and β-cell function (64). The DIAGRAM plus consortium observed that the T2D risk alleles at PPARG, FTO, IRS1, and the newly discovered gene KLF14, were associated with higher fasting insulin, with primary effect on insulin action, whereas 3 loci (TCF7L2, ARAPI, and CDKAL1) were found to be associated with reduced fasting insulin suggestive of β-cell dysfunction (41).

The ADRA2A locus is associated with impaired insulin granule docking and reduced β-cell exocytosis in congenic GK (Goto-Kakizaki) rats (65). Rosengren et al. reported that human carriers of the ADRA2A risk variant (rs553668) have reduced fasting insulin and decreased insulin secretion (65) as a consequence of increased expression of the ADRA2 receptor in pancreatic islets. It is well known that epinephrine excess can suppress insulin secretion and cause diabetes. Pretreatment of human islets with the ADRA2A-blocking agent yohimbine normalized insulin secretion in risk genotype carriers.

**INSULIN RESISTANCE**

Insulin resistance is a condition in which peripheral tissues fail to respond adequately to insulin. The healthy β-cell can compensate by producing more insulin; in genetically predisposed individuals β-cell may not be able to do so, and diabetes may develop.

GCKR, first identified in the Diabetes Genetics Initiative study (32) along with IGF1 variants (rs780094 and rs35767), was found in the MAGIC study to be associated with fasting insulin and higher insulin resistance (HOMA-IR) (55). In a metaanalysis study by Orho-Melander et al. (66) with more than 45 000 study participants, the GCKR locus (rs780094) was found to show opposite associations with fasting plasma glucose and triglyceride concentrations along with its association with C-reactive protein concentrations. IGF1 codes for insulin-like growth factor 1, and a null mutation for IGF1 has previously been correlated with improper glucose homeostasis with resistance to insulin and high concentrations of circulating insulin (67). Recently the DIAGRAM-plus consortium found an association of novel T2D loci KLF14 (rs972283 G allele) and insulin resistance (HOMA-IR) in addition to 2 previously known T2D obesity loci, FTO and PPARG (41). KLF14 (a widely expressed, intronless member of the Kruppel-like family of transcription factors) is maternally expressed, and the variant at this locus appears to have a primary effect on insulin action (68).

**OBESITY**

Obesity is an important predictor and cause of T2D and cardiovascular disease. Genes increasing susceptibility to obesity are thus important candidates for T2D risk as well. Increased free fatty acid concentrations are often seen in obese individuals, in whom they cause defective glucose metabolism through insulin resistance development (69–71). Free fatty acids also have a direct lipotoxic effect on β-cells by lipid accumulation, leading to reduced insulin secretion and apoptosis (72).

Obesity shows a heritability ranging from 55%–85% for BMI (73). T2D GWAS detected strong associations of common SNPs at the FTO locus and T2D risk (33); the risk of T2D could completely be explained by the effect on BMI (33, 47). Subsequently, new GWAS have identified many novel obesity-associated loci in both children and adults. Six of these loci [FTO; melanocortin 4 receptor (MC4R); transmembrane protein 18 (TMEM18); glucosamine-6-phosphate deaminase 2 (GNPDA2); SH2B adaptor protein 1 (SH2B1); and neuronal growth regulator (NEGR1)] are also associated with T2D risk (Table 3) (47, 74–78).

The FTO gene confers risk for T2D in Europeans, with each A allele increasing BMI by approximately 0.4
kg/m² (47). However, results have been variable for replication in other ethnic populations such as Hispanics (79), Asians, Oceanics (80), and blacks (81). The FTO gene encodes for a protein 2-oxoglutarate-dependent nucleic acid demethylase involved in fatty acid metabolism, DNA repair, and posttranslational modifications (82). Studies in mice suggest that Fto might affect neuropeptide Y expression in the hypothalamus, which in turn is known to impact feeding behavior (83). Fto–knock-out mice have a reduced size and body weight with improved insulin sensitivity and increased adrenaline concentrations in blood, suggesting that their energy expenditure occurs in the presence of increased sympathetic activation so they are lean even after hyperphagia (84). Olszewski et al. have reported that high body weight in mice downregulates the Fto gene in hypothalamus, which suggests that the Fto gene affects energy metabolism and expenditure and not feeding (85).

The second most important obesity gene, MC4R, is expressed in the hypothalamus and regulates energy expenditure, insulin sensitivity, and energy intake (86). Leptin, an anorexogenic adipokine, stimulates the production of proopiomelanocortin products, which bind to MC4R, resulting in decreased food intake and increased energy expenditure (86). The MC4R variant also confers increased T2D risk in Europeans (87, 88).

In a recent GWAS, Thorleifsson et al. identified additional loci for obesity (74); this study was followed by a metaanalysis confirming the results in Europeans (89). In this investigation, 2 loci each at chromosome 6p21 and 12q13 were found associated with T2D, rs2844479 [close to allograft inflammatory factor 1 (AIF1) and natural cytotoxicity triggering receptor 3 (NCR3)] and rs7138803 [close to BCDIN3 domain containing (BCDIN3D) and Fas apoptotic inhibitory molecule 2 (FAIM2)] even after adjustment for BMI. Chambers et al. also conducted a GWAS on waist circumference and insulin-resistance traits and found the MC4R locus to be associated with HOMA-IR in Asian Indians and Europeans, whereas association with T2D was observed only in Europeans (88). Risk-allele frequencies of the newly identified obesity susceptibility loci vary between 27% and 91% in European populations, suggesting that although they have small effect sizes a substantial part of the risk for T2D development could be attributable to obesity alleles (76).

### The Missing Heritability of T2D

Although a considerable number of variants have been found to affect the risk of T2D, these variants still account for only a small proportion of the total heritability. Most of the identified variants have very modest effect sizes in the range of 1.1 to 1.3. Some of these effect sizes...
sides may even be overestimated owing to winner’s curse, i.e., serendipitously large effects in the discovery population leading to their discovery; however, others will prove to merely be proxies of variants with larger effects. The DIAGRAM-plus consortium estimated that the presented >30 T2D-associated SNPs could explain only about 10%–15% of the sibling relative risk of about 3 (3, 41).

The key question is thus, what can explain the missing heritability? One possibility is of course that the heritability estimates are wrong. This error would not be surprising for T2D, for which the familial risk of 3 has been derived from family studies, but most GWAS studies have not requested that study participants have another family member with T2D. Other possible explanations include a much larger number of variants of even smaller effects that remain to be identified, low frequency (minor allele frequency 0.5%–5%), and rare (<0.5%) variants with larger effects, as well as structural variants (such as copy-number variants and copy-neutral variants) that are poorly captured by existing arrays. Many of these effects could possibly be identified by the currently used strategies by increasing genotyping coverage and sample sizes even further. However, it is likely that many variants escape detection owing to mechanisms such as parental-origin–specific effects, gene–gene interactions, gene–environment interactions, and epigenetic effects, among others. The hunt for these variants will benefit from the development of new detection strategies and methods. In the years to come, next-generation sequencing in families in which the heritability has been estimated should be able to answer some of these questions.

Rare Variants

So far the GWAS performed have been designed to find relatively common variants, typically focusing on variants with allele frequencies more than 5%, owing to the low statistical power to detect associations with more rare alleles. It is plausible that some of the genetic variance is explained by relatively rare variants, some of which could have larger effect sizes than the common variants identified so far. A substantial part of the missing heritability could be due to variants with large or intermediate effect sizes and relatively low frequencies that are likely to have escaped detection by current methods, having too low penetrance to allow linkage analysis and being too rare to detect in GWAS. The strategies to identify such variants largely depend on the frequency of the variants. Some variants in the range of 1% to 5% should be possible to identify by simply increasing genotype density and cohort sizes. An important tool to achieve this is the 1000 Genomes Project, which is extending the catalogue of known human variants to frequencies close to or below 1%. However, detection of many of these rare or intermediate variants will require next-generation sequencing rather than traditional GWAS or genotyping (90).

Variants that are relatively rare in the European population could also be identified by studying other ethnic groups where the variant is more common. For example, variants in the KCNQ1 gene were first identified in Asians where the minor allele frequency of the key SNPs (rs2237892, rs2237897, rs2074196) was 30%–40%, which is much higher than in Europeans with a frequency <10% (48, 49).

Gene–Gene and Gene–Environment Interactions

Gene–gene interactions with nonadditive effects are another potential source of missing heritability. How much such epistatic effects contribute to the genetic variation in the predisposition to T2D is unknown, and so far there is little evidence in human studies (91–93) that epistatic changes contribute to disease pathology. In contrast, epistatic effects seem to be very common in animal models, in which they can be identified by using selective breeding strategies. For example, Shao et al. compared 41 quantitative traits (14 of them metabolic) that differed significantly between the parental strains in chromosome substitution strains and found that for 40 of 41 traits the cumulative phenotypic effect of all chromosomes was greater than the parental difference. This finding suggests that nonlinear effects is a rule rather than an exception (if all effects were additive the sum of effects would equal the parental difference) (94). Studies of gene–gene interactions in outbred populations are difficult because the number of tests required usually exceeds our calculation and power capacity (95). Therefore, a common strategy to limit the number of tests is to restrict the analysis to loci with significant main effects (45, 96) However, it is well known from studies in plants and animals that epistatic effects are often detected in the absence of main effects (97, 98). An illustrative study was performed by Carlborg et al., who mapped quantitative trait loci affecting body size in chickens in a cross between 2 lines that had been selectively bred for 40 generations, resulting in a 6-fold difference in body weight. In spite of the large genetic variance, only 1 significant locus, with a minor effect, was identified in standard quantitative-trait loci analysis. However, when epistatic interactions were considered, an additional 5 loci turned out to have significant effects and together the 6 genes formed an interaction network accounting for 45% of the difference between the 2 parental lines (97). Similar networks have been shown to determine obesity in mice (99).
There are no reasons to assume that the genetics of human metabolism will be less complex.

Gene–environment interactions are another potential source of missing heritability that could have effects similar to those of gene–gene interactions, but these are also poorly investigated in humans owing to the fact that the potentially relevant environmental factors are huge in number and by their nature difficult to measure in a standardized fashion to allow analysis in adequately large sample sizes.

**Epigenetics**

Epigenetics has been defined as heritable changes in gene function that occur without a change in the nucleotide sequence. Epigenetic modifications can be passed from one cell generation to the next (mitotic inheritance) or between generations of individuals (meiotic inheritance). Although meiotic inheritance is well established in plants, there is only limited information about the inheritance of epigenetic traits between generations in animals and humans so far (100, 101). These types of epigenetic effects could explain some of the missing “genetic” variance component of complex diseases. Epigenetic effects can also occur during life, stochastically or in response to environmental stimuli, thereby influencing the effects of genetic variants and thus acting as a mechanism of gene–environment interaction.

Both DNA methylation and histone modifications can change the response of our genome to the environment during life. DNA methylation often results in decreased expression of a gene, which in its extreme form becomes completely suppressed (imprinted). Post-translational modifications on the N-terminal histone tails in the chromatin also play an essential role in regulation of gene expression and function. Modification of the chromatin structure can allow access or prevent access of proteins to binding with a transcription factor, which has been shown for TCF7L2 (102).

In the pathogenesis of T2D, special attention has been given to the role of intrauterine DNA methylation and imprinting for the programming of diabetogenic effects later in life (103). An interesting study by Dabelda et al. has shown that intrauterine exposure to diabetes increases the risk of diabetes and obesity in the offspring compared to siblings who were born before their mothers’ onset of diabetes (104). However, the exact mechanism for this maternal effect remains to be determined.

Intrauterine growth retardation has also been associated with increased DNA methylation of the Pdx1 promoter in islets from experimental animals (105). A few additional epigenetic studies have been carried out in target tissues from individuals with T2D (106–108). One such study describes hypermethylation of the PPRG1A gene promoter in islets from T2D patients, resulting in decreased gene expression, and glucose-stimulated insulin secretion (106). Methylation of the insulin gene promoter also seems to affect insulin secretion (109). In addition, some studies have suggested a role for epigenetic regulation of genes involved in energy metabolism, appetite control, and β-cell function, i.e., peroxisome proliferator-activated receptor alpha (PPARA) (110), leptin (LEP) (111), and pancreatic and duodenal homeobox 1 (PDX1) (105).

**Prediction of Disease**

Several studies have investigated the predictive value of genetic markers in comparison with clinical risk factors with similar results (62, 112–114). The largest study was performed by our own group in 2 Scandinavian cohorts. Inclusion of genetic information from 16 genotyped risk variants increased the area under the ROC curve from 0.74, for a model including only clinical risk factors, to 0.75. Notably the difference in predictive value increased the earlier the genetic analyses were performed (62). The use of genetic markers in diagnosis of T2D is thus not very useful so far, although identification of new variants with larger effects could change this rapidly. It should also be noted that the effect size of a locus is not at all correlated with the potential for the involved pathway as a target for clinical therapy.

**Conclusions**

It is evident that existing genetic markers explain only a modest (<15%) part of the heritability of T2D. The question to be answered by future studies is whether the missing heritability will be accounted for by the identification of an increasing number of common variants, as shown for Crohn disease (explaining over 40% of heritability) and genetic variants that determine lipid concentrations (>35%), or whether finding the answer will require the use of next-generation sequencing to identify rare variants with stronger effects.


