Type 2 diabetes: new genes, new understanding

Inga Prokopenko1,2, Mark I. McCarthy1,2,3 and Cecilia M. Lindgren1,2

Over the past two years, there has been a spectacular change in the capacity to identify common genetic variants that contribute to predisposition to complex multifactorial phenotypes such as type 2 diabetes (T2D). The principal advance has been the ability to undertake surveys of genome-wide association in large study samples. Through these and related efforts, ~20 common variants are now robustly implicated in T2D susceptibility. Current developments, for example in high-throughput resequencing, should help to provide a more comprehensive view of T2D susceptibility in the near future. Although additional investigation is needed to define the causal variants within these novel T2D-susceptibility regions, to understand disease mechanisms and to effect clinical translation, these findings are already highlighting the predominant contribution of defects in pancreatic β-cell function to the development of T2D.

A leap forward in complex trait genetics
Type 2 diabetes (T2D) is a common, chronic, complex disorder of rapidly growing global importance. It accounts for >95% of diabetes worldwide and is characterized by concomitant defects in both insulin secretion (from the β-cells in the pancreatic islets) and insulin action (in fat, muscle, liver and elsewhere), the latter being typically associated with obesity [1] (Figure 1). Although strong evidence for familial clustering points to the contribution of genetic mechanisms, recent rapid changes in diabetes incidence (which has, in the United States, doubled in two decades) and prevalence (which is well above 10% in some societies) indicate that environmental and lifestyle factors are also of major relevance [1,2].

Despite strenuous efforts over the past two decades to identify the genetic variants that contribute to individual differences in predisposition to T2D, the story until recently was characterized by slow progress and limited success. The advent of genome-wide association (GWA) analysis has transformed the potential for researchers to uncover variants influencing complex, common phenotypes – including T2D – and has resulted in the identification of a growing number of trait susceptibility loci [3–8].

Until 2006, the main approaches used to track down common variants influencing common dichotomous traits, such as T2D, involved either hypothesis-free genome-wide linkage mapping in families with multiply affected subjects or association studies within ‘candidate’ genes using case-control samples or parent–offspring trios. The former suffered from being seriously underpowered for sensible susceptibility models, because linkage is best placed to detect variants with high penetrance. As far as we can tell, common variants with high penetrance do not contribute substantially to risk of common forms of T2D and few, if any, robust signals have emerged from such efforts [9]. The latter candidate-gene association approach has historically been compromised by difficulties associated with choosing credible gene candidates. Selection was typically based on some particular hypothesis about the biological mechanisms that are putatively involved in T2D pathogenesis but, because the function of much of the genome is poorly characterized, it remains almost impossible to make fully informed decisions. In addition, all too often these candidate-gene studies were conducted in sample sets that were far too small to offer confident detection of variants with the kinds of effect sizes that are now known to be realistic. Only rarely could the findings of one study be replicated in another [10–12]. With hindsight, it is easy to appreciate why these approaches yielded so few examples of genuine diabetes-susceptibility variants.

In fact, only two of the many candidate-gene associations claimed for T2D have stood the test of time. The Pro12Ala variant in the peroxisome proliferator-activated receptor gamma (PPARG) gene (encoding the thiazolidinedione class of drugs used to treat T2D) [11] and the Glu23Lys variant in KCNJ11 (the potassium inwardly rectifying channel, subfamily J, member 11, which encodes part of the target for another class of diabetes drug, the sulphonylureas) [12] are both common polymorphisms shown in multiple studies to influence risk of T2D. Their effect sizes are only modest, each copy of the susceptibility allele increasing risk of disease by 15–20%. Interestingly, rare mutations in both KCNJ11 and PPARG are also known to be causal for certain rare monogenic syndromes (neonatal diabetes and lipodystrophies) characterized by severe metabolic disturbance of β-cell function and insulin resistance, respectively [13,14].

The capacity to undertake efficient, accurate association analysis on a far larger scale than previously (culminating in GWA analysis) has transformed this picture. As we
describe in this review, such studies have led to confident identification of many novel T2D-susceptibility signals, thereby exposing some of the key mechanisms and pathways involved in the pathogenesis of T2D. Here, we also discuss some of the challenges that will need to be faced, if these findings are to be translated into advances in the clinical management of those with diabetes.

The first moves towards large-scale association mapping

The earliest indication that the ‘hypothesis-free’ association approach to gene identification might succeed for T2D came from the discovery that variants within the transcription factor 7-like 2 (TCF7L2) gene had a substantial effect on T2D susceptibility [15]. TCF7L2 encodes a transcription factor that is active in the Wnt-signalling pathway and that had no ‘track-record’ as a candidate for T2D; indeed, this susceptibility effect was detected through a search for microsatellite associations across a large region of chromosome 10 that had been previously implicated in T2D susceptibility by linkage [16]. Subsequent fine-mapping efforts localized the likely causal variant(s) to an intron within TCF7L2 [15,17]. The fact that this signal was found within a region of apparent T2D linkage seems to have been serendipitous, because none of these variants within TCF7L2 are capable of explaining the linkage effect [15,17]. Across a swathe of replication studies [3–7,18], it has become clear that TCF7L2 variants have a substantially stronger effect on T2D risk than those in PPARG and KCNJ11, with a per-allele odds ratio of ~1.4 (Table 1; Figure 2). As a result, the 10% of Europeans that are homozygous for the risk allele have approximately twice the odds of developing T2D as those carrying no copies [15,18]. The evidence implicating variants within TCF7L2 in T2D susceptibility has naturally prompted efforts to understand the mechanisms involved. Current evidence indicates that alteration of TCF7L2 expression or function disrupts pancreatic islet function, possibly through dysregulation of proglucagon gene expression.

Table 1. T2D-susceptibility loci for which there is genome-wide significant evidence for association

<table>
<thead>
<tr>
<th>Locus (nearest genes)</th>
<th>Year association</th>
<th>Approach</th>
<th>Probable mechanism</th>
<th>Index variant</th>
<th>Effect sizea</th>
<th>Risk-allele frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPARG</td>
<td>2000</td>
<td>Candidate</td>
<td>Insulin action</td>
<td>rs1801282</td>
<td>1.14</td>
<td>0.87</td>
</tr>
<tr>
<td>KCNJ11</td>
<td>2003</td>
<td>Candidate</td>
<td>β-cell dysfunction</td>
<td>rs5215</td>
<td>1.35</td>
<td></td>
</tr>
<tr>
<td>TCF7L2</td>
<td>2006</td>
<td>Large-scale association</td>
<td>β-cell dysfunction</td>
<td>rs7801685</td>
<td>1.37</td>
<td>0.31</td>
</tr>
<tr>
<td>FTO</td>
<td>2007</td>
<td>GWA</td>
<td>Altered BMI</td>
<td>rs8050136</td>
<td>1.17</td>
<td>0.40</td>
</tr>
<tr>
<td>HEX/IDE</td>
<td>2007</td>
<td>GWA</td>
<td>β-cell dysfunction</td>
<td>rs1111875</td>
<td>1.65</td>
<td></td>
</tr>
<tr>
<td>SLC30A8</td>
<td>2007</td>
<td>GWA</td>
<td>β-cell dysfunction</td>
<td>rs1326663</td>
<td>1.69</td>
<td></td>
</tr>
<tr>
<td>CDKAL1</td>
<td>2007</td>
<td>GWA</td>
<td>β-cell dysfunction</td>
<td>rs10946398</td>
<td>1.32</td>
<td></td>
</tr>
<tr>
<td>CDKN2A/2B</td>
<td>2007</td>
<td>GWA</td>
<td>β-cell dysfunction</td>
<td>rs10811661</td>
<td>1.20</td>
<td>0.83</td>
</tr>
<tr>
<td>IGFBP2</td>
<td>2007</td>
<td>GWA</td>
<td>β-cell dysfunction</td>
<td>rs4402960</td>
<td>1.14</td>
<td>0.32</td>
</tr>
<tr>
<td>HNF1B</td>
<td>2007</td>
<td>Large-scale association</td>
<td>β-cell dysfunction</td>
<td>rs4430796</td>
<td>1.10</td>
<td>0.47</td>
</tr>
<tr>
<td>WFS1</td>
<td>2007</td>
<td>Large-scale association</td>
<td>Unknown</td>
<td>rs1001031</td>
<td>1.12</td>
<td>0.60</td>
</tr>
<tr>
<td>JAZF1</td>
<td>2008</td>
<td>GWA</td>
<td>β-cell dysfunction</td>
<td>rs864745</td>
<td>1.10</td>
<td>0.50</td>
</tr>
<tr>
<td>CDC123/CAMK1D</td>
<td>2008</td>
<td>GWA</td>
<td>Unknown</td>
<td>rs12779790</td>
<td>1.11</td>
<td>0.18</td>
</tr>
<tr>
<td>TSPAN8/LGR5</td>
<td>2008</td>
<td>GWA</td>
<td>Unknown</td>
<td>rs7961581</td>
<td>1.09</td>
<td>0.27</td>
</tr>
<tr>
<td>THADA</td>
<td>2008</td>
<td>GWA</td>
<td>Unknown</td>
<td>rs7578579</td>
<td>1.15</td>
<td>0.90</td>
</tr>
<tr>
<td>ADAMTS9</td>
<td>2008</td>
<td>GWA</td>
<td>Unknown</td>
<td>rs4607103</td>
<td>1.09</td>
<td>0.76</td>
</tr>
<tr>
<td>NOTCH2</td>
<td>2008</td>
<td>GWA</td>
<td>Unknown</td>
<td>rs10923931</td>
<td>1.13</td>
<td>0.10</td>
</tr>
<tr>
<td>KCNJ1</td>
<td>2008</td>
<td>GWA</td>
<td>β-cell dysfunction</td>
<td>rs2237892</td>
<td>1.29</td>
<td></td>
</tr>
</tbody>
</table>

*aAbbreviations: ADAMTS9, ADAM metalloepitidase with thrombospondin type 1 motif 9; CAMK1D, calcium/calcmodulin-dependent protein kinase 1D; CDC123, cell division cycle 123 homologue (Saccharomyces cerevisiae); CDKAL1, CDK5 regulatory subunit-associated protein-1-like 1; CDKN2B, cyclin-dependent kinase inhibitor 2B; FTO, fat mass and obesity associated; HHEX, haematopoietically expressed homeobox; HNF1B, hepatocyte nuclear factor 1 homeobox B; IDE, insulin degrading enzyme; IGF2BP2, insulin-like growth factor 2 mRNA binding protein 2; JAZF1, juxtaposed with another zinc finger gene 1; KCNJ11, potassium inwardly rectifying channel, subfamily J, member 11; KCNQ1, potassium voltage-gated channel, KQT-like subfamily, member 1; LGR5, leucine-rich repeat-containing G-protein coupled; NOTCH2, Notch homologue 2 (Drosophila); PPARG, peroxisome proliferator-activated receptor gamma; SLC30A8, solute carrier family 30 (zinc transporter), member B; TCF7L2, transcription factor 7 like 2; THADA, thyroid adenoma associated; TSPAN8, tetraspanin 8; WFS1, Wolfram syndrome1.

*bEstimates of effect size (given as per-allele odds ratios, i.e. the increase in odds of diabetes per copy of the risk allele) and risk-allele frequencies are all reported for European-descent populations based on available data (Figure 2).
leading to reduced insulin secretion and enhanced risk of T2D [19].

**GWA scans identify novel loci influencing T2D susceptibility**

The solution to previous disappointing efforts to map T2D-susceptibility loci had been obvious for some time, but it is only in the past two years that it has become possible, from both technical and economic perspectives, to undertake hypothesis-free GWA testing in samples of sufficient size to generate convincing results. The advent of the GWA approach was the result of a ‘perfect storm’ involving at least three components. The first was delivery of a catalogue of patterns of human genome-sequence variation through the efforts of the International HapMap Consortium (http://www.hapmap.org) [20]. One of the important messages generated by the HapMap was that, in non-African-descent populations at least, extensive correlations between neighbouring single nucleotide polymorphisms (SNPs; i.e. linkage disequilibrium) could constrain the number of independent genetic tests required to survey the genome, such that ~80% of all common variation could be sampled with as few as 500 000 carefully chosen SNPs [21,22]. At the same time, new genotyping methods were developed that addressed the technical challenges required to perform massively parallel SNP-typing at high accuracy and low cost [23].

The third crucial change was the dawning realization by investigators that earlier estimates of locus effect sizes had been wildly optimistic and that success would require analyses of far larger sample sizes than previously considered, an objective that has catalysed large-scale international collaboration [10].

So far, results from a total of 12 GWA scans for T2D have been published (Table 2). Six of these represent ‘high-density’ scans (i.e. at least 300 000 SNPs, offering genome-wide coverage >65%), all in samples of Northern European descent [3–8]. These studies included >19 000 individuals (7142 T2D patients and 11 996 controls) ranging in sample size from 997 to 6674 (Table 2). Two recent studies in East Asian subjects were on a smaller scale (featuring between 82 000 and 207 000 typed SNPs in a few hundred cases only), but large-scale replication did result in at least one confirmed T2D signal (see later) [24,25]. The other four studies featured a wider array of ethnic groups (including Native American [26], Hispanic [27] and populations of European descent [28,29]) but were less extensive with respect to both sample size and SNP density (all used the Affymetrix 100K array) and will not be considered further here.

As has been reviewed elsewhere [30], the first wave of studies led to identification and mutual replication of six entirely novel T2D-susceptibility loci and confirmed the three previously established signals at PPARG, KCNJ11
and TCF7L2 [3–8]. Each of these six loci has subsequently been replicated in studies involving individuals of both European [31] and Asian [32–35] origin. Apart from variants within the fat mass and obesity associated (FTO) gene [36], which exert their T2D effect through a primary impact on body mass index (BMI), the remaining signals (involving SNPs within or adjacent to hematopoietically expressed homeobox [HHEX]/insulin degrading enzyme [IDE], CDK5 regulatory subunit associated protein 1-like 1 [CDKAL1], insulin-like growth factor 2 mRNA binding protein 2 [IGF2BP2], cyclin-dependent kinase inhibitors 2a and 2b [CDKN2A and CDKN2B] and solute carrier family 30, member 8 [SLC30A8]) exert their primary effect on insulin secretion [31,37,38]. Other reported signals, such as those in exostoses (multiple) 2 (EXT2) and LOC387761 [3] have not been widely substantiated on replication and are less likely to represent genuine signals. Although it has become convention, largely as a matter of convenience, to label these susceptibility loci according to the most credible regional candidate (e.g., FTO), it is vital to remember that these assignments do not imply a confirmed mechanistic link. What these GWA studies deliver are association signals, and extensive fine-mapping and functional studies are required before the causal variants are identified and the molecular and cellular mechanisms through which they function are precisely defined [39].

These GWA studies, as well as detecting new loci, provided the first ‘genome-wide’ perspective of the landscape of T2D susceptibility and thereby enabled clearer ‘bench-marking’ of other claimed T2D-susceptibility effects for which the accumulated evidence from candidate-gene studies remained somewhat equivocal [40]. Examples include variants in the genes encoding calpain-10 (CAPN10; thought to be involved in β-cell function), insulin (INS; an obvious candidate) and PC-1 (ENPP1; the product of which is known to modulate insulin-receptor function). None of these genes has featured prominently in GWA analyses to date and, although this does not necessarily exclude a contribution to T2D predisposition, it indicates that the main effects attributable to these variants are small and/or subject to substantial modification by genetic background or environmental exposures. Either way, it seems likely that exhorbitantly large sample sets will be required before such signals can attain the standard of proof now available for the loci described in Table 1.

### Meta-analysis identifies six further T2D-susceptibility loci

Efforts to find additional T2D-susceptibility loci have to contend with the modest effect sizes anticipated and the stringent significance thresholds required when many hundreds of thousands of SNPs are tested in parallel [39]. The obvious solution is to increase sample size and the most effective strategy for this involves combining existing GWA data through meta-analysis. The Diabetes Replication And Meta-analysis (DIAGRAM) consortium took such an approach [41], integrating data from three previously published GWA scans [4,5,7], thereby doubling the sample size compared to the largest of the individual studies (Table 2) to ~4500 cases and 5500 controls. The consortium also used novel imputation approaches [42] to infer genotypes at additional SNPs that were not directly typed on the commercial arrays used for the original GWA scans, thereby extending the analysis to a total of ~2.2 million SNPs across the genome. (Imputation approaches use information on haplotype structure available from a densely typed reference sample [such as HapMap] to fill-in genotypes at that subset of untyped SNPs for which confident assignment is possible.) In this study, 69 signals showing the strongest associations in the GWA meta-analysis were genotyped in an initial replication set of 22 426 individuals and the top 11 signals (representing ten independent loci) emerging from this second analysis were then evaluated in ~57 000 further subjects.

### Table 2. Overview of GWA scans for type 2 diabetes

<table>
<thead>
<tr>
<th>Study</th>
<th>Number of cases</th>
<th>Number of controls</th>
<th>Sample source</th>
<th>Genotyping array</th>
<th>T2D phenotype</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wellcome Trust Case Control Consortium</td>
<td>1924</td>
<td>2938</td>
<td>UK</td>
<td>Affymetrix 500K</td>
<td>Enrichment for family history of T2D, AAO &lt; 65 years</td>
<td>[7]</td>
</tr>
<tr>
<td>Diabetes Genetics Initiative</td>
<td>1464</td>
<td>1467</td>
<td>Finland, Sweden</td>
<td>Affymetrix 500K</td>
<td>Partial enrichment for family history and lean T2D</td>
<td>[4]</td>
</tr>
<tr>
<td>deCODE Genetics</td>
<td>1399</td>
<td>5275</td>
<td>Iceland</td>
<td>Illumina 300K</td>
<td>No specific enrichment for family history, young AAO or BMI</td>
<td>[6]</td>
</tr>
<tr>
<td>Finland-US Investigation of NIDDM</td>
<td>1161</td>
<td>1174</td>
<td>Finland</td>
<td>Illumina 300K</td>
<td>Partial enrichment for family history</td>
<td>[5]</td>
</tr>
<tr>
<td>Diabetes Gene Discovery Group</td>
<td>694</td>
<td>645</td>
<td>France</td>
<td>Illumina 300K + 100K</td>
<td>Family history of T2D, AAO &lt; 45 years, BMI &lt; 30 kg/m2</td>
<td>[3]</td>
</tr>
<tr>
<td>DiaGen</td>
<td>500</td>
<td>497</td>
<td>East Finland, Germany, UK, Aashkenazi</td>
<td>Affymetrix 100K</td>
<td>Enrichment for family history, AAO &lt; 60 years</td>
<td>[8]</td>
</tr>
<tr>
<td>Pima</td>
<td>300</td>
<td>334</td>
<td>Pima Indians</td>
<td>Affymetrix 100K</td>
<td>AAO &lt; 25 years, enrichment for family history</td>
<td>[26]</td>
</tr>
<tr>
<td>Starr County, Texas</td>
<td>281</td>
<td>280</td>
<td>Mexican Americans</td>
<td>Affymetrix 100K</td>
<td>Controlled for admixture</td>
<td>[27]</td>
</tr>
<tr>
<td>BioBank Japan</td>
<td>194</td>
<td>1556</td>
<td>Japanese</td>
<td>Custom set of ~268K SNPs</td>
<td>Enrichment for T2D cases with retinopathy</td>
<td>[25]</td>
</tr>
<tr>
<td>Japanese multi-disease collaborative genome scan</td>
<td>187</td>
<td>1504</td>
<td>Japanese</td>
<td>JSNP Genome Scan 100K SNPs</td>
<td>T2D cases</td>
<td>[24]</td>
</tr>
<tr>
<td>Old Order Amish</td>
<td>124</td>
<td>295</td>
<td>Amish</td>
<td>Affymetrix 100K</td>
<td>Enrichment for family history</td>
<td>[28]</td>
</tr>
<tr>
<td>Framingham Health Study</td>
<td>91</td>
<td>1087</td>
<td>Massachusetts</td>
<td>Affymetrix 100K</td>
<td>Incident T2D cases</td>
<td>[29]</td>
</tr>
</tbody>
</table>

*Abbreviations: AAO, age at onset.

*Projects are listed in descending order of case sample size.

---

**Note:** The table provides an overview of GWA scans for type 2 diabetes, summarizing the number of cases and controls, sample sources, genotyping arrays, and T2D phenotype enrichment strategies. The table also highlights the significance of genetic studies in understanding T2D susceptibility, emphasizing the role of meta-analysis in identifying additional loci, and the importance of replication in validating these associations.
After integrating data from all study subjects, six signals reached combined levels of significance that are highly unlikely to have occurred by chance (P ≤ 5 × 10−8) [41] (Table 1; Figure 2). Only one of these signals contained a gene for which there was reasonable biological candidacy: Notch homologue 2, Drosophila (NOTCH2) is known to be involved in pancreatic development. For the others, mapping within or adjacent to ADAM metallopeptidase with thrombospondin domain type 1 motif 9 (ADAMTS9), calcium/calmodulin-dependent protein kinase 1D (CAMK1D), juxtaposed with another zinc finger gene 1 (JAZF1), tetraspanin 8 (TSPAN8)/leucine-rich repeat-containing G-protein coupled (LGR5) and thyroid adenoma associated (THADA), the mechanisms involved remain unclear.

Further signals from large-scale candidate-pathway studies

At the same time as these GWA efforts, which had added a total of 12 loci to the list, two further loci emerged from large-scale candidate-pathway studies initiated in the pre-GWA era. The first of these focused on genes that were thought to be likely to modulate β-cell function and high-lighted variants in the Wolfram syndrome 1 (WFS1) gene [43]. Wolfram syndrome is a rare monogenic syndrome, which features optic atrophy, diabetes insipidus and deafness as well as diabetes mellitus, for which rare mutations in the WFS1 gene are causal [43]. Although the evidence for association in the initial report fell just short of strict significance thresholds at the genome-wide level (the best SNP generated a P-value of 1.4 × 10−7), replication studies have confirmed this association [44]. The second study examined variants in genes that were causally implicated in monogenic forms of diabetes and generated compelling evidence that common variants in the gene encoding hepatocyte nuclear factor 1-β (HNF1B, also known as TCF2, a transcription factor implicated in pancreatic islet development and function) were associated with T2D [45], an effect independently confirmed when the same gene emerged from a GWA analysis for prostate-cancer-susceptibility loci [46].

Common variants with modest effect sizes

For the T2D-susceptibility variants so far discovered (Table 1), the risk-allele frequency in Europeans ranges from approximately 10% (NOTCH2) to 90% (THADA, PPARG). This allele-frequency spectrum is, in part, the direct consequence of experimental designs (such as the use of commodity genotyping arrays) that have focused on the detection of common variants: these have low power to detect associations caused by rare causal alleles [47]. Apart from TCF7L2, per-allele effect sizes are, at best, modest (mostly in the 1.1–1.2 range), which explains the massive sample sizes required for robust inference [41]. However, for most of these loci, the causal variant(s) have yet to be identified with any certainty. Although it is likely, for reasons of power, that the causal variants in each of these regions have risk-allele frequencies similar to those of the variants showing maximal association, this is not inevitable. For example, an association detected on genome-wide scanning based on a case-control allele frequency difference of 35% versus 32% could, in principle, reflect linkage disequilibrium with a causal allele present in 4% of cases and only 1% of controls. Such a variant would have an effect size much larger than that of the `index' variant through which the signal had been originally detected.

Ongoing efforts to expand existing meta-analyses of European-descent populations [41] and to extend the number of SNPs taken forward for large-scale replication should uncover additional T2D-susceptibility loci. At the same time, GWA studies conducted in other major ethnic groups should provide complementary views of the susceptibility landscape of T2D. Indeed, the first GWA scans for T2D reported in East Asian subjects [24,25] (Table 2) have recently demonstrated how studies in non-European descent populations can reveal novel susceptibility loci. In these East Asian scans, variants in the potassium voltage-gated channel, KQT-like subfamily, member 1 gene (KCNQ1) have been shown to influence diabetes risk.

Replication studies conducted in samples of East Asian and South Asian origin indicate that the variants found in European-descent populations tend to exert similar effects on T2D susceptibility, although differences in allele frequency can lead to differences in detectability and population-level effect [32–35]. For example, the minor allele frequency at T2D-associated TCF7L2 variants is far lower in East Asians, thereby rendering the association with T2D much less visible [34]. The situation at KCNQ1 is reversed in that, although the effect size in European-descent samples is not that different from that observed in East Asians, substantial differences in risk-allele frequency mean that the association was far easier to detect in studies of Japanese, Korean and Chinese individuals. (It is not yet clear whether these differences in allele frequency are simply the result of chance genetic drift or constitute evidence for selection.) It is likely, therefore, that GWA studies performed in non-European descent populations will reveal additional loci which, by virtue of differences in allele frequency, haplotype structure, genetic background or relevant environmental exposure, would be extremely hard to find in European populations, however large the study.

Understanding function

Gene-discovery efforts are primarily motivated by the expectation that, by finding robust genotype–phenotype associations, we will generate novel insights into disease mechanisms and new avenues for clinical translation. However, the task of moving from association signal to complete mechanistic understanding will often be laborious and early inferences can be misleading. We should expect that, on occasion, genes other than the most obvious regional candidates will turn out to be responsible for the susceptibility effect. Nevertheless, for some of the loci identified by GWA, a compelling functional candidate is clear, the best example being SLC30A8 [3]. This gene encodes an islet-specific zinc transporter, which is implicated in the maintenance of insulin granule function [48]. At this locus, one of the variants displaying the strongest association is a nonsynonymous SNP (Arg325Trp), which might well represent the causal polymorphism.
At other loci, more than one excellent biological candidate maps into the region of the association signal. On chromosome 10q, for example, the causal variants most probably lie, on the basis of the patterns of flanking recombination hot-spots, within a 400-kb region that contains the coding regions of three genes, two of which (HHEX, which encodes a homeobox protein involved in pancreatic development and IDE, which is implicated in rodent models of diabetes) have strong biological claims for a role in T2D pathogenesis [49–51].

At other signals, the links between association and function remain less obvious, but more intriguing. The chromosome 9p21 T2D association signal maps to a ~15-kb interval that lies ~200 kb from the nearest protein-coding genes, CDKN2A and CDKN2B. These genes encode cyclin-dependent kinase inhibitors, primarily known for their role in the development of cancers [52]. However, in rodent models at least, overexpression of the CDKN2A homologue recapitulates the T2D phenotype [53], raising the suspicion that remote regulatory effects on CDKN2A expression, possibly mediated through ANRIL, a non-coding RNA which maps to the region of maximal association, could be responsible [54]. If so, this would be consistent with the observation that several of the newly found T2D-susceptibility loci implicate genes involved in cell-cycle regulation, and that there might be some overlap in the mechanisms influencing cancer and diabetes predisposition.

Lessons about diabetes
Several other interesting features emerge from these new findings. First, analyses of the associated variants in healthy, non-diabetic populations has demonstrated that, for most, the primary effect on T2D susceptibility is mediated through deleterious effects on insulin secretion, rather than insulin action [31,37,38]. This finding addresses a long-standing debate about the relative importance of inherited defects in insulin secretion as opposed to action in the pathogenesis of T2D and focuses attention on mechanisms involved in the regulation of islet β-cell mass and function.

The second interesting observation relates to the frequency with which the genes and variants implicated in T2D-susceptibility are also associated with other traits. For example, a different set of variants at the 9p21 region influences predisposition to coronary artery disease and to pathological dilatation of the major blood vessels (‘arterial aneurysms’) [55–58]. Variants at HNF1B (TCF2) and JAZF1 have clear effects on susceptibility to prostate cancer [46,59] and CDKAL1 has recently been revealed as a susceptibility locus for Crohn’s disease [60]. These examples of pleiotropy point towards previously unsuspected overlap in pathogenetic pathways.

Finally, several of the loci implicated in common forms of T2D are also known to harbour rare mutations that are causal for rare, monogenic forms of diabetes. This is true of PPARG [13] and KCNJ11 [14], and for HNF1B (TCF2) [61] and WFS1 [62]. Of course, having shown that one type of variant in a given gene is capable of disturbing glucose metabolism, it is much more likely that variants from other parts of the allele-frequency spectrum (if they exist) will have similar phenotypic consequences. Further resequencing of the newly identified T2D-susceptibility loci is likely to reveal additional instances of lower-frequency, higher-penetrance alleles and might contribute to more effective functional analyses of these regions.

Individual prediction
In addition to delivering novel biological insights into disease pathogenesis, there is growing interest in the potential for the growing numbers of susceptibility variants to open the door to the provision of individualized medical care contingent upon the results of personalized genetic profiling [63].

At this stage, however, the prospects for individual prediction seem limited. In a recent study, Lango and colleagues addressed the extent to which the known T2D-susceptibility variants might, in combination, provide quantification of individual T2D risk [64]. It is certainly possible to identify small numbers of individuals who, because they have inherited extremely large or extremely small numbers of susceptibility alleles, differ substantially with regard to risk of T2D. With the present collection of variants, this amounts to an approximately fourfold difference in risk between those individuals in the top and bottom 1% of the distribution of risk-allele counts.

When viewed from a more population-based perspective, the situation is less compelling. The discriminative accuracy achieved with the current set of T2D risk variants (0.60) compares unfavourably with the value of 0.78 possible through use of age, BMI (a standardised measure of obesity obtained by dividing weight in kg by the square of the height in metres) and gender alone [64].

For the time being then, these ‘traditional’ risk factors (e.g. age and BMI) provide far more accurate risk-stratification than do the common variants so far identified. It remains unclear whether providing individuals with information based on their personal genetic profiling would actually translate into health benefits through, for example, improved motivation to make lifestyle changes or altered therapeutic interventions.

Next steps
GWA studies have certainly kick-started complex trait genetics but much work remains to be done before a full mechanistic picture emerges (Figure 3). In T2D, the scans performed so far have concentrated on European populations and have only been designed to detect those variants represented on, or tagged by, the sites on the commodity genotyping arrays. Only a small proportion (<10%) of the overall variance in T2D predisposition has been described.

Clearly, identification of the causal variants responsible for the association signals uncovered will provide valuable clues to understanding disease predisposition. This is likely to require detailed re-sequencing of the associated regions followed by exhaustive fine-mapping in multiple ethnic groups. Even then, linkage disequilibrium might make it impossible to limit the causal candidates to a single variant and functional studies will be required for definitive assignment of causation.
In parallel, there are efforts to extend the range of genomic variation encompassed by genome-wide approaches. Structural (or copy number) variants (CNVs) are prime candidates here, given growing evidence of their overall contribution to human genomic diversity and of their phenotypic consequences [65–69]. Armed with more complete inventories of the sites of common CNVs and new tools that will enable genome-wide CNV genotyping (a far more demanding task than CNV discovery), the coming months should see the first reports evaluating the part played by CNVs in T2D predisposition.

New high-throughput resequencing approaches should also enable investigators to obtain a more comprehensive view of susceptibility effects across the allele-frequency spectrum. Traditional linkage approaches are well-suited to finding rare, highly penetrant variants and have been successful in identifying mutations causal for highly familial forms of diabetes, such as maturity onset diabetes of the young. As has been described earlier, the GWA approach is a powerful tool for detecting common variant effects, which are, in the case of T2D at least, almost exclusively of low penetrance. Neither approach would do well at detecting variants with characteristics that lie between these extremes, with penetrance too modest to be detected by linkage and frequency too low to be captured by current GWA reagents [41]. An improved inventory of such variants is emerging from both local and global resequencing efforts (such as the 1000 Genomes Project; http://www.1000genomes.org) and the large sample sizes now available will enable properly powered targeted association studies to be performed. If these efforts identify low-frequency, intermediate-penetrance variants contributing to T2D predisposition, these are likely to be valuable tools for genomic profiling and individual prediction [39].

Finally, we can expect to see advances in the understanding of genome function to be increasingly integrated into genetic studies. One obvious example of this relates to the growing importance attributed to the regulatory role of microRNAs (miRNAs) [68]. miRNAs are small, ~22-base pair non-coding RNAs that bind target mRNAs and modify their expression by translational repression or target degradation. Because each miRNA is thought to regulate many target mRNAs, a picture is emerging of miRNAs as ‘molecular switches’ that are able to activate (or deactivate) broad programs of cellular response.

Recently, several specific miRNAs have been found to have important roles in regulation of glucose and lipid
metabolism. In humans, there is evidence that miRNA-1 and miRNA-133 are instrumental in the normal proliferation and differentiation of skeletal muscle tissue [69] and, in murine models, miRNA-9 and miRNA-375 have been shown to regulate insulin secretion [70,71]. miRNA-124a2 has also been implicated in pancreatic β-cell development and function, probably through a binding of FOXA2 leading to a downstream effect on FOXA2 target genes including PDX1, KCNJ11 and ABCC8 [72]. Other examples of key miRNA regulators include miRNA-143 (adipocyte differentiation) [73] and miRNA-122 (hepatic lipid metabolism) [74]. Together, these and other studies provide a strong basis for regarding miRNAs as key functional candidates relevant to the pathogenesis of T2D, and ongoing and future studies (for example, examining the contribution of genome sequence variation in miRNA coding and target sequences to T2D predisposition) should help to establish the extent to which this hypothesis is true.

Conclusion
The ability to undertake genome-wide surveys of the relationship between sequence variation and common human phenotypes has provided a powerful new tool for geneticists. It is clear that substantial advances in the understanding and management of major diseases, including T2D, will flow from wider application of these approaches and clinical translation of the findings.

Acknowledgements
The preparation of this manuscript was supported by funding from the European Commission (ENGAGE consortium: HEALTH-F4–2007–2011); Medical Research Council (G0601261); Oxford NIHR Biomedical Unit; and European Commission (ENGAGE consortium: HEALTH-F4–2007–2011). Other examples of key miRNA regulators include miRNA-143 (adipocyte differentiation) [73] and miRNA-122 (hepatic lipid metabolism) [74]. Together, these and other studies provide a strong basis for regarding miRNAs as key functional candidates relevant to the pathogenesis of T2D, and ongoing and future studies (for example, examining the contribution of genome sequence variation in miRNA coding and target sequences to T2D predisposition) should help to establish the extent to which this hypothesis is true.

References
12. Gloyn, A.L. et al. (2003) Large scale association studies of variants in genes encoding the pancreatic β-cell K-ATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with increased risk of type 2 diabetes. Diabetes 52, 568–572
Review


44 Franks, P.W. et al. (2008) Replication of the association between variants in WFS1 and risk of type 2 diabetes in European populations. Diabetologia 51, 458–463


61 Maestro, M.A. et al. (2007) Distinct roles of HNF1β, HNF1α and HNF4α in regulating pancreas development, beta-cell function and growth. Endocr. Dev. 12, 33–45


64 Lango, H. et al. (2008) Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk. Diabetes, DOI: 10.2337/db08-0504


