New gene variants alter type 2 diabetes risk predominantly through reduced beta-cell function
John R.B. Perry and Timothy M. Frayling

Introduction
The advent of genome-wide association studies (GWASs) in 2007 has resulted in huge advances in complex trait genetics. Many of the genes now implicated were not expected and so provide completely new insights into common causes of a disease. One of the biggest success stories has been in type 2 diabetes (T2D), where very recent progress has taken the number of confirmed gene loci to 18. One of the reasons behind this success, in a condition that has a considerable environmental component, is that researchers of type 2 diabetes were among the first to combine data from multiple GWAS, providing more power to detect disease loci.

Candidate genes
Until the recent GWASs, the identification of type 2 diabetes genes had been frustratingly slow and difficult. Over the past decade, human geneticists had predominantly relied on two approaches for gene discovery:

- Detailed analyses of genomic regions shared between diabetic relatives more often than expected by chance (linkage approach), and candidate gene studies. Of these efforts, only the candidate-gene approach has lead to any reproducible results.

- The candidate-gene approach provided several important lessons for gene discovery efforts, including the importance of using dense sets of single-nucleotide polymorphisms (SNPs) to capture a large proportion of the common variation across genes. This allows assessment of variation in introns and considerable distance upstream or downstream of a gene, not just the coding or obviously functional regions. Here, data from the HapMap project [1] helped inform which SNPs were redundant due to correlation between common variants. These studies also demonstrated the importance of using several thousand cases and controls together with even larger replication studies to provide robust evidence of association.
Many common diseases, including type 2 diabetes, have several related rare monogenic forms. If mutations in a gene result in diabetes, it suggests the gene product is crucial for normal insulin secretion or sensitivity. These genes are therefore also excellent candidates for a role in common forms of the disease. For type 2 diabetes, there are several single-gene disorders of relevance, including maturity-onset diabetes of the young (MODY) and permanent neonatal diabetes mellitus (PNDM).

Studying common variation in monogenic diabetes genes proved a successful approach. Common variants in or near four genes involved in monogenic diabetes disorders [2–5] predispose to type 2 diabetes; these genes are KCNJ11, PPARG, TCF2/HNF1B and WFS1 [6–9, 10–14]. A common variant in a fifth monogenic gene, TCF1, encoding HNF1A, is also a risk factor for type 2 diabetes in the Oji-Cree population, but this variant is only found in this population of native Canadians [15, 16].

**The noncandidate-gene-based approach to T2D gene discovery**

The identification of common risk variants in TCF7L2 represented an important milestone in T2D genetics. TCF7L2 was the first T2D gene locus to be identified without any prior knowledge of its function.

TCF7L2 was identified in 2006 by Decode genetics [17]. Following up a region on chromosome 10q identified through linkage analysis, Decode genotyped over 200 polymorphisms across the region in 1185 cases and 931 Icelandic controls. They identified a number of variants at the TCF7L2 locus in strong linkage disequilibrium with each other, with the most associated SNP present in nearly 28% of chromosomes from individuals without diabetes and nearly 36% of chromosomes from individuals with T2D. This finding has now been replicated in many studies, including those of non-European population groups [18–20].

To date, variants in TCF7L2 have the largest risk effect of any of the known T2D susceptibility loci – the 7% of European individuals carrying two copies of the risk allele are almost twice as likely to get diabetes than the 55% of individuals carrying two protective alleles. TCF7L2 is a component of the WNT signalling pathway, a target of the HHEX gene that is also implicated in T2D. TCF7L2 is a transcription factor that acts as a nuclear receptor for CTNNB1 (beta-catenin) [21]. WNT signalling plays a crucial role in cell proliferation, motility and normal embryogenesis, and has been shown to regulate myogenesis and adipogenesis [22, 23]. Tight regulation of WNT signalling is also required for normal pancreatic development during embryonic growth [24].

**New genotyping technology results in the identification of six new type 2 diabetes gene loci**

Six new gene loci were identified in early 2007 through independent GWASs. This took the number of disease susceptibility variants to 11, including the four loci discovered by candidate-gene approaches and TCF7L2. The start of 2008 saw a further increase, with an additional seven genes replicated through combined GWASs (see Table 1 and Fig. 1) [25].

This success was the result of new technology combined with large DNA resources. Genome-wide association scans allow hundreds of thousands of SNPs, spread across the entire genome, to be analysed in a single experiment. New technology in the form of DNA chips allows researchers to capture a large proportion of genomic variation, in thousands of DNA samples. Statistically significant allele frequency differences between cases and controls can then highlight potential polymorphisms, and ultimately genes, which are implicated in disease. This hypothesis-free approach provides the opportunity to identify completely unexpected genes, broadening our understanding of disease mechanisms.

The two main commercial genome-wide genotyping platforms have utilized different methods to design SNP arrays. SNPs on the Illumina panel are selected to capture optimal genomic variation based on HapMap linkage disequilibrium data. Although the genome contains millions of SNP sites, the correlation of SNPs due to linkage disequilibrium allows a reasonable proportion of these to be ‘tagged’ by a modestly sized subset of these SNPs. Alternatively, SNPs on the Affymetrix chips were selected without any consideration for linkage disequilibrium – and so were randomly distributed across the genome. Although the Affymetrix approach would give a reduction in variation coverage, it benefits from not being population specific and also has an inbuilt redundancy (due to linkage disequilibrium) to help compensate for genotyping failures. The limitation of both chips, however, is that the arrays are not exhaustive, with poor coverage of both rare SNPs [minor allele frequency (MAF) <5%] and structural variation such as copy number variants (CNVs).

On a genotyping platform comprising nearly 500000 SNPs, an entire spectrum of ‘significant’ P values will occur through chance. For example, by performing 500000 statistical tests, we would expect a P value of $2 \times 10^{-6}$ (1/500000) to occur once by chance. In addition, the prior probability of any variant being a risk factor for diabetes is very small. These statistical arguments mean that P values lower than approximately $5 \times 10^{-7}$ are required to provide robust statistical evidence of association.
Results from the first T2D GWASs were published in early 2007, providing six replicated new susceptibility loci [26*,27*,28*,29,30,31*]. Six studies, describing five separate T2D scans, genotyped a combined total of 18,000 individuals of Northern European ancestry. Large additional replication cohorts were an essential part of each study design, totalling over 55,000 individuals across the five studies.

These efforts were swiftly followed by researchers combining information from three of these GWA scans (WTCCC, DG1 and FUSION), combining information in a meta-analysis of over 10,000 samples [32]. This meta-analysis used imputed data to test approximately 2.2 million SNPs for disease association. Imputation algorithms allow the prediction of untyped SNP genotypes based on patterns of haplotype variation from the HapMap dataset [33,34].

<table>
<thead>
<tr>
<th>Example variant</th>
<th>Closest gene</th>
<th>Mode of identification</th>
<th>Current evidence (P value)*</th>
<th>Predicted disease mechanism</th>
<th>Odds ratio (per allele)*</th>
<th>RAF (UK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs5215 (E23K)</td>
<td>KCNJ11</td>
<td>Candidate</td>
<td>5 x 10^-11</td>
<td>Beta-cell dysfunction</td>
<td>1.14 (1.10–1.19)</td>
<td>0.35</td>
</tr>
<tr>
<td>rs1111875</td>
<td>HHEXIDE</td>
<td>Genome wide</td>
<td>7 x 10^-17</td>
<td>Beta-cell dysfunction</td>
<td>1.15 (1.10–1.19)</td>
<td>0.65</td>
</tr>
<tr>
<td>rs13266834</td>
<td>SLC30A8</td>
<td>Genome wide</td>
<td>1 x 10^-19</td>
<td>Beta-cell dysfunction</td>
<td>1.15 (1.12–1.19)</td>
<td>0.69</td>
</tr>
<tr>
<td>rs10946398</td>
<td>CDKAL1</td>
<td>Genome wide</td>
<td>2 x 10^-18</td>
<td>Beta-cell dysfunction</td>
<td>1.14 (1.11–1.17)</td>
<td>0.32</td>
</tr>
<tr>
<td>rs7901895</td>
<td>TCF7L2</td>
<td>Region wide</td>
<td>1 x 10^-48</td>
<td>Beta-cell dysfunction</td>
<td>1.37 (1.31–1.43)</td>
<td>0.31</td>
</tr>
<tr>
<td>rs10811661</td>
<td>CDKN2A/2B</td>
<td>Genome wide</td>
<td>8 x 10^-15</td>
<td>Beta-cell dysfunction</td>
<td>1.20 (1.14–1.25)</td>
<td>0.83</td>
</tr>
<tr>
<td>rs4402960</td>
<td>IGFBP2</td>
<td>Genome wide</td>
<td>9 x 10^-16</td>
<td>Beta-cell dysfunction</td>
<td>1.14 (1.11–1.18)</td>
<td>0.32</td>
</tr>
<tr>
<td>rs8050136</td>
<td>FTO</td>
<td>Genome wide</td>
<td>1 x 10^-12</td>
<td>Altered BMI</td>
<td>1.17 (1.12–1.22)</td>
<td>0.4</td>
</tr>
<tr>
<td>rs17782313</td>
<td>MC4R</td>
<td>Genome wide</td>
<td>0.003</td>
<td>Altered BMI</td>
<td>1.09 (1.03–1.15)</td>
<td>0.24</td>
</tr>
<tr>
<td>rs1801282 (P12A)</td>
<td>PPARG</td>
<td>Candidate</td>
<td>2 x 10^-6</td>
<td>Unknown</td>
<td>1.14 (1.08–1.20)</td>
<td>0.87</td>
</tr>
<tr>
<td>rs4430796</td>
<td>TCF2</td>
<td>Candidate</td>
<td>8 x 10^-10</td>
<td>Unknown</td>
<td>1.10 (1.07–1.14)</td>
<td>0.47</td>
</tr>
<tr>
<td>rs10010131</td>
<td>WFS1</td>
<td>Candidate</td>
<td>4 x 10^-10</td>
<td>Unknown</td>
<td>1.12 (1.09–1.15)</td>
<td>0.60</td>
</tr>
<tr>
<td>rs864745</td>
<td>JAZF1</td>
<td>Genome wide</td>
<td>5 x 10^-14</td>
<td>Unknown</td>
<td>1.10 (1.07–1.13)</td>
<td>0.50</td>
</tr>
<tr>
<td>rs12779790</td>
<td>CDC123/CAMK1D</td>
<td>Genome wide</td>
<td>1 x 10^-10</td>
<td>Unknown</td>
<td>1.11 (1.07–1.14)</td>
<td>0.18</td>
</tr>
<tr>
<td>rs7961581</td>
<td>TSPAN8/RGR5</td>
<td>Genome wide</td>
<td>1 x 10^-9</td>
<td>Unknown</td>
<td>1.09 (1.06–1.12)</td>
<td>0.27</td>
</tr>
<tr>
<td>rs7578597</td>
<td>THADA</td>
<td>Genome wide</td>
<td>1 x 10^-9</td>
<td>Unknown</td>
<td>1.15 (1.10–1.20)</td>
<td>0.90</td>
</tr>
<tr>
<td>rs4607103</td>
<td>ADAMTS9</td>
<td>Genome wide</td>
<td>1 x 10^-8</td>
<td>Unknown</td>
<td>1.08 (1.06–1.12)</td>
<td>0.76</td>
</tr>
<tr>
<td>rs10923931</td>
<td>NOTCH2</td>
<td>Genome wide</td>
<td>4 x 10^-8</td>
<td>Unknown</td>
<td>1.13 (1.08–1.17)</td>
<td>0.10</td>
</tr>
</tbody>
</table>

a Approximate P values and odds ratios calculated by meta-analysis of individual study odds ratios from [23,24,27] except for the signals for HHEX-IDE, which also includes data from [28], CDKAL1, which also includes data from [26*], SLC30A8, which includes data from five GWASs listed above, TCF2, which is based on data from [8,11] and WFS1, which is based on data from [14]. JAZF1, CDC123, TSPAN8, THADA, ADAMTS8 and NOTCH2 data from [28*], MC4R data from [42]. ADAMTS9, ADAM metallopeptidase with thrombospondin type 1 motif 9. BMI, body mass index; CAMK1D, calcium/calmodulin-dependent protein kinase 1D; CDC123, cell division cycle 123 homologue (Saccharomyces cerevisiae); CDKAL1, CDK5 regulatory subunit-associated protein 1 like 1; CDKN2, cyclin-dependent kinase inhibitor 2A; FTO, fat mass and obesity associated; HHEX, haematopoietically expressed homeobox; 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 1; LGR5, leucine-rich repeat-containing G protein coupled; MC4R, melanocortin 4 receptor; NOTCH2, Notch homologue 2 (Drosophila); PPARG, peroxisome proliferator-activated receptor-γ gene; RAF, allele frequency; SLC30A8, solute carrier family 30 (zinc transporter), member B; TCF2, transcription factor 2, hepatic; LF-B3, variant hepatic nuclear factor; TCF7L2, transcription factor 7 like 2 (T-cell specific, HMG box); THADA, thyroid adenoma associated; TSPAN8, tetraspanin 8; WFS1, Wolfram syndrome 1. Adapted from Frayling [25].

The type 2 diabetes gene map: 2008 update Perry and Frayling 373

Seven of the 18 common T2D variants alter disease risk through altered beta-cell function

There is increasing evidence to suggest the genetic component to T2D risk acts through beta-cell dysfunction – seven of the gene loci have been associated with impaired insulin secretion in the general population, but there is little if any evidence that any of the loci alter insulin resistance (see Fig. 2). The common variants in the TCF7L2 gene predispose to type 2 diabetes by reducing beta-cell function and insulin secretion. Multiple studies of individuals without diabetes show that measures of first-phase insulin release are reduced in individuals carrying T2D risk alleles, but there is no evidence for a role in insulin resistance [35–38]. TCF7L2 type 2 diabetes risk alleles also influence birth weight [39*]. Each copy of the diabetes risk allele in mothers is associated with a nearly 30 g increase in offspring birth weight. The mechanism behind this is likely to be reduced maternal insulin secretion causing elevated blood glucose, which, in turn, increases foetal insulin levels, an important foetal growth factor.

Risk alleles at the CDKAL1, SLC30A8, HHEX-IDE, CDKN2A/B, KCNJ11 and IGF2BP2 loci all alter insulin secretion [10,30,40*,41*,42,43]. For each loci, there are several studies comprising a few hundred to 2000 individuals with detailed oral glucose tolerance test (OGTT) or detailed physiological clamp studies. Some studies have shown insulin-secretion defects at one gene and not others, but overall a consistent picture is emerging that these loci alter diabetes risk by reducing the insulin-secretory capacity of the beta-cells. There is no evidence...
that any of these loci alter insulin resistance and for the
six most recently described loci there are no data yet to
address the question of whether insulin secretion or
insulin resistance is the primary causal pathway.

Two of the 18 loci mediate their effect on T2D risk by
altering BMI. Common variants in the
\textit{FTO} and \textit{MC4R}
genes represent the only polymorphisms reproducibly
associated with altered BMI and obesity risk.

\textit{FTO} was the first of these gene loci to be discovered [44*].
A common variant in intron 1 of the \textit{FTO} gene was the
second strongest locus associated with type 2 diabetes after
\textit{TCF7L2} in the UK T2D GWAS. Subsequent analyses later
confirmed the locus-altered BMI, which, in turn, conferred
diabetes risk. Using 30,000 individuals, Frayling \textit{et al.} [44*]
reported that 16% of the European population carrying two
copies of the T2D risk allele were approximately 2–3 kg
heavier than the 35% with no risk alleles. \textit{FTO} genotypes
alter metabolic traits inline with the effect on BMI, indi-
cating that \textit{FTO} variants do not independently alter other
metabolic traits [45]. Early evidence suggests that the gene
may influence appetite [46]; however, more functional
research is required to confirm this.

Common variants near the \textit{MC4R} gene were also recently
found to be associated with fat mass and obesity risk [47].
\textit{MC4R} variants were identified during a BMI GWAS on
approximately 17,000 individuals, which ranked variants in
\textit{FTO} at the top. The common variant associated with
BMI is 188 kb from the \textit{MC4R} gene but is a compelling
biological candidate because rare mutations are the com-
monest cause of monogenic obesity in humans [48].
A genetic overlap between type 2 diabetes and prostate cancer

Several epidemiological studies have highlighted an inverse relationship between T2D and prostate cancer risk [49]. One of the most interesting findings of recent GWASs is that common variation in the T2D susceptibility genes TCF2 and JAZF1 also influence prostate cancer risk [9, 50]. For TCF2, the risk allele of the T2D susceptibility variant has a protective effect on prostate cancer susceptibility, whereas for JAZF1 there are two different variants predisposing to the two diseases, which are not correlated with each other. The mechanisms by which these genes influence disease are unknown; however, it is possible that they influence cell-cycle control. It is possible that on the one hand a particular allele predisposes to increased cell proliferation, while the alternative allele could influence cell senescence or death, potentially affecting beta-cell mass and function. This theory might also explain the number of cell-cycle genes that have been implicated in T2D susceptibility.

Next steps

Having identified a handful of loci responsible for T2D susceptibility, the next stage is to identify the mechanisms by which these genes act on disease risk. Further physiological studies, alongside research by clinical, cellular, animal and molecular biologists will hopefully lead to further advances in our understanding of the disease. Alongside these efforts, there still remain many more susceptibility loci to find, as only a small fraction of heritability can be explained by the known variants.

Further meta-analysis of studies and analysing additional samples will raise the number of disease loci and our
knowledge of disease. For the majority of T2D loci, it is unknown where the causal variant lies and, in some cases, it is unclear which gene might be implicated (e.g., HHEX/IDE locus). Deep sequencing and additional genotyping will therefore be required to pinpoint functional variant(s) within these genes.

The analysis of additional types of genomic variation, in particular CNVs will also play an important role in future T2D genetics studies. Known CNVs cover more than 15% of the human genome and are often not well correlated with SNPs. CNVs have been shown to associate with various human traits and diseases, including SLE and HIV-1/AIDS susceptibility [51,52]; however, any potential role in T2D is still unknown.

Conclusion
Type 2 diabetes GWASs over the last 2 years have accelerated our understanding of disease aetiology. There are now 18 gene loci reproducibly associated with T2D, identified through large collaborative efforts. These studies have highlighted genes with no previous biological knowledge, implicating new pathways in T2D genetics. There is now a growing amount of evidence to suggest the genetic component to T2D risk acts through beta-cell dysfunction rather than insulin action.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:
• of special interest
•• of outstanding interest

Additional references related to this topic can also be found in the Current World Literature section in this issue (pp. 524–525).

17 The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3000 shared controls. Nature 2007; 447:661–678.


