

# Genome-wide association studies and type 2 diabetes

Eleanor Wheeler and Inês Barroso

## Abstract

In recent years, the search for genetic determinants of type 2 diabetes (T2D) has changed dramatically. Although linkage and small-scale candidate gene studies were highly successful in the identification of genes, which, when mutated, caused monogenic forms of T2D, they were largely unsuccessful when applied to the more common forms of the disease. To date, these approaches have only identified two loci (*PPARG*, *KCNJ11*) robustly implicated in T2D susceptibility. The ability to perform large-scale association analysis, including genome-wide association studies (GWAS) in many thousands of samples from different populations, and subsequently, the shift to form large international collaborations to perform meta-analyses across many studies has taken the number of independent loci showing genome-wide significant associations with T2D to 44. This number includes six loci identified initially through the analysis of quantitative glycaemic phenotypes, illustrating the usefulness of this approach both to identify new disease genes and gain insight into the mechanisms leading to disease. Combined, these loci still only account for ~10% of the observed familial clustering in Europeans, leaving much of the variance unexplained. In this review, we will describe what GWAS have taught us about the genetic basis of T2D and discuss possible next steps to uncover the remaining heritability.

**Keywords:** type 2 diabetes; complex disease; genome-wide association study; meta-analysis; single nucleotide polymorphism; copy number variant

## INTRODUCTION

Type 2 diabetes (T2D) is a complex disease that is characterized by insulin resistance and impaired  $\beta$  cell function. A major public health concern, the prevalence is increasing at an alarming rate in parallel with rising obesity rates worldwide. The highest incidences of T2D are seen in developing countries where 80% of diabetes deaths occur [1, 2], presumably due to poorer healthcare systems. Worryingly, there is also recent evidence to show that the age of onset has decreased and cases of T2D in adolescents and children have been reported [3]. Notably young-onset T2D is more commonly observed in certain ethnicities (e.g. African, Pima Indians) [4, 5]. Although this rise in diabetes prevalence can be mostly attributed to changes in diet and lifestyle,

there is strong evidence of a genetic basis for T2D [6]. For example, a study in Danish twins estimated the T2D concordance rate in dizygotic twins as 43% compared with 63% in monozygotic twins [7], and the relative risk of T2D for a sibling is approximately four- to six-fold higher than that of the general population [8].

Over the years, linkage and small-scale candidate gene studies were successful in the identification of genes that caused monogenic variants of T2D, such as maturity onset diabetes of the young (MODY), syndromes of insulin resistance and Wolfram syndrome [6, 9]. However, when applied to the more common, complex forms of T2D these methods were largely unsuccessful. Combined analysis of multiple studies eventually yielded a number of

Corresponding author. Inês Barroso. E-mail: [ib1@sanger.ac.uk](mailto:ib1@sanger.ac.uk)

**Eleanor Wheeler** is a staff scientist at the Wellcome Trust Sanger Institute, Cambridge, UK where she works as a statistical geneticist to identify genes that increase risk of obesity, type 2 diabetes and related quantitative traits.

**Inês Barroso** is a senior group leader and currently Joint Acting Head of Human Genetics at the Wellcome Trust Sanger Institute, Cambridge, UK with a joint appointment at the Institute of Metabolic Science, Cambridge, UK. Her research focus is on using genetic and genomic approaches to understand the aetiology of common and rare forms of metabolic disease, with a focus on obesity, type 2 diabetes and related quantitative traits.

chromosomal locations with suggested evidence for linkage; however none have revealed the underlying loci [10, 11]. On the other hand, small-scale candidate gene approaches identified only two loci reproducibly associated with T2D, *PPARG* and *KCNJ11* [12–14] (Figure 1). Large-scale association and targeted evaluation of common variants within genes known to be involved in monogenic forms of diabetes identified an additional two signals, at *WFS1* [15] and *HNF1B* [16] (Figure 1).

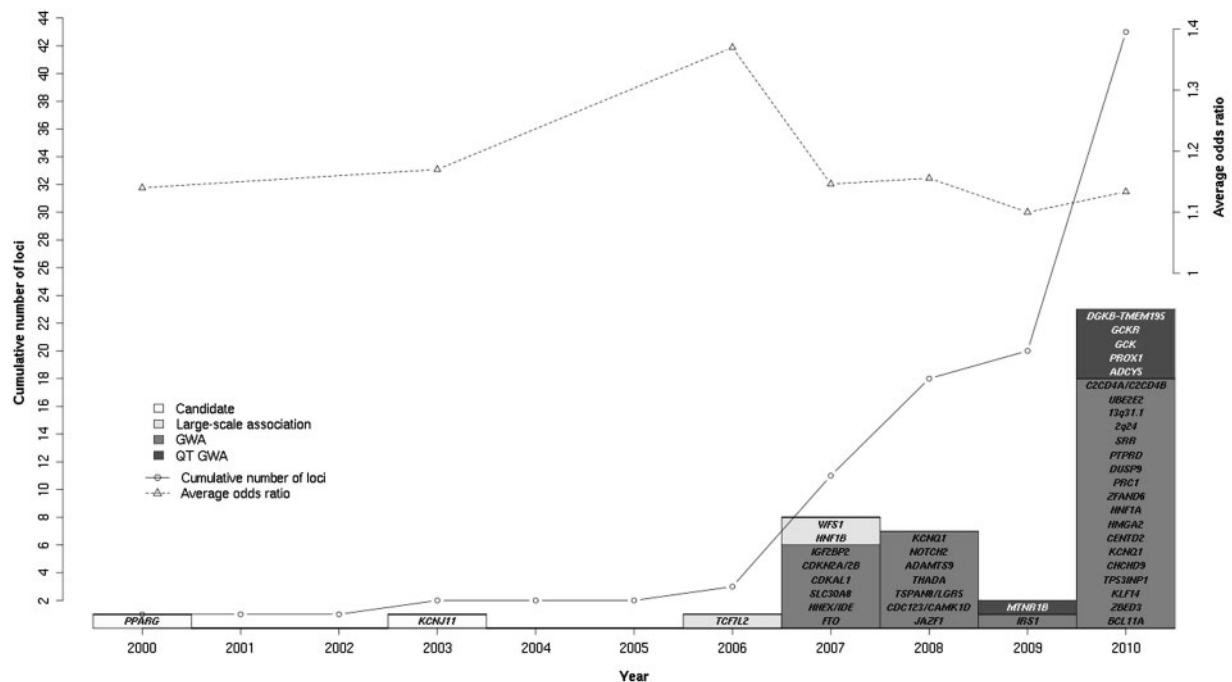
The signal with the strongest effect for T2D was identified early in 2006 through the identification of a suggestive linkage signal at 10q. Researchers at deCODE subsequently fine-mapped the region and discovered the robust association of the *TCF7L2* (transcription factor 7-like 2) gene with T2D [17] (Figure 1). Despite this successful finding, the common variants in *TCF7L2* did not explain the original signal on chromosome 10, suggesting that additional, undiscovered, and rare pathogenic variants may still account for the original linkage signal. *TCF7L2* was not a known biological candidate at the time of its discovery, its finding illustrated the power of an unbiased or ‘hypothesis free’ approach

to reveal important novel insights regarding disease aetiology.

Since linkage scans were shown to have less power than association studies to detect the small genetic effects expected for loci involved in complex diseases [18], the ability to scan the entire genome for association in such an unbiased way was a very attractive prospect. In addition, there was the practical advantage that large samples of unrelated cases and controls could be collected much more easily than family-based samples.

## GENOME-WIDE ASSOCIATION STUDIES

There were several key developments that made genome-wide association studies (GWAS) possible. The first was the completion of the human genome sequence which in turn led to detailed maps of common single nucleotide polymorphisms (SNPs) and patterns of linkage disequilibrium (LD—statistical dependence of alleles at two genetic loci) produced by The International HapMap Consortium (HapMap) [19], and later maps of common copy



**Figure 1:** A timeline illustrating the identification and method through which currently known T2D susceptibility variants have been identified. The figure clearly illustrates the marked increase in the number of loci identified through GWAS and the subsequent meta-analyses. Average ORs for T2D have remained fairly constant over recent years and it is likely that there are many additional signals of similar effect and frequency to be discovered (OR estimates from [35, 37–39, 47, 54, 76]). *KCNQ1* (2008) and *PTPRD*, *SRR*, *I3q31.I*, *UBE2E2*, *C2CD4A/C2CD4B* (2010) were identified in samples of Asian ancestry and ORs included were those reported for those populations.

number variants (CNVs) [20, 21]. Information on the extent of LD was used to show that in European populations, approximately 500 000 SNPs could be used to serve as proxies for ~80% of common variants [22], prompting commercial companies to design high-throughput high-accuracy genotyping arrays. The collection of large, well-phenotyped samples from cohort and case-control studies combined with the advances in genotyping technology meant that it was then both technically and financially possible to genotype many thousands/millions of variants in large numbers of samples. Finally, there were methodological advances, including statistical tools such as PLINK [23] that were made freely available, facilitating the design, analysis and interpretation of the large amounts of data being produced. When performing such large numbers of association tests, the importance of stringent significance thresholds was recognized. Empirical estimation of 1 million independent tests across common variants in the genome of Europeans led to  $P = 5 \times 10^{-8}$  becoming the generally accepted threshold below which genome-wide significance was declared [24, 25]. Suggestive signals identified in an initial discovery sample set are then routinely validated using a staged design which incorporates a second, sufficiently powered, sample set. Finally, it is important to note that when describing an association, it has become standard practice to refer to the identified signal by the closest gene(s) name(s); but this does not necessarily mean that the gene itself is causal.

The first GWAS for T2D [26] provided the necessary evidence that GWAS would work for complex disease. They enriched for a genetic effect, and increased their power, by selecting non-obese, earlier onset cases with a family history of T2D. With a relatively modest sample size (694 cases and 645 controls), their design enabled them to identify the positive control at *TCF7L2*, as well as a non-synonymous SNP in the zinc transporter *SLC30A8* and variants in *HHEX* (Figure 1). This study was followed closely by a wave of genome-wide scans from deCODE Genetics [27], the Wellcome Trust Case Control Consortium (WTCCC) [28], Diabetes Genetics Initiative (DGI) [29] and the Finland-US investigation of NIDDM genetics (FUSION) [30]. In addition to the previously identified signals, these investigators also found novel associations at *CDKAL1*, *IGF2BP2* and *CDKN2A/B* (Figure 1). Interestingly, controls were matched to cases in the

deCODE, DGI and FUSION studies whereas in the WTCCC, a common set of controls were used for the seven UK wide case cohorts [31]. Although from an epidemiological perspective, this was not the ideal design, it led to the finding of *FTO*, shown to be associated with T2D through its effect on BMI [32] (Figure 1).

It was clear from the signals identified at that stage that the effect sizes for any additional loci were likely to be small and larger sample sizes would be required to identify additional novel loci meeting the stringent significance thresholds required. The Diabetes and Genetics Replication and Meta-Analysis (DIAGRAM) consortium was thus formed to carry out meta-analyses of three of the previously published studies; WTCCC, DGI and FUSION. This international collaborative effort identified six new loci *JAZF1*, *CDC123-CAMK1D*, *TSPAN8-LGR5*, *THADA*, *ADAMTS9* and *NOTCH2* [33] (Figure 1). The DIAGRAM study marked an important progression from analysis at the cohort level, to form large international consortia and meta-analyse summary level data, therefore increasing the discovery sample size and the power to detect association. Critical to the success of this approach was a clearly defined phenotype across cohorts and a stringent analysis plan with clear quality control (QC) criteria. Novel imputation approaches used local LD patterns from HapMap to infer genotypes at untyped SNPs based on the directly typed SNP genotypes [34]. This development enabled results from different genotyping platforms to be combined, such that ~2.2 million SNPs across the genome could be tested.

Up until that point, efforts to identify T2D susceptibility loci had focused on populations of European descent. The identification of *KCNQ1* in East-Asian samples [35, 36] illustrated how studies from populations of different ancestry might uncover additional T2D loci. Despite the much larger discovery sample sizes, *KCNQ1* was previously unidentified in samples of European descent, yet differences in risk-allele frequency meant that the East-Asian studies were better powered to detect an effect. Recent studies in populations of Asian ancestry have also found novel associations reaching genome-wide significance at *PTPRD*, *SRR*, 13q13.1, *UBE2E2* and *CDC4A-CDC4B* [37–39] (Figure 1), although some may require further replication in additional populations.

In parallel with these ongoing efforts, a complementary approach was to explore, not the clinical

diabetes endpoint, but rather the quantitative intermediate phenotypes. Associations with fasting glucose for variants in *GCK* [40–42], *G6PC2* [40, 41] and *GCKR* [29, 43, 44] were some of the first to be reported.

Subsequently, the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) was established to conduct large-scale meta-analyses of GWAS for glycaemic phenotypes (e.g. fasting glucose, fasting insulin, 2h post-challenge glucose levels) in persons without diabetes. The aim was to understand the genetic determinants of the trait variation in the physiological range, and to identify key pathways leading to pathological glucose levels and T2D risk. The first novel association with fasting glucose found through the MAGIC effort [45], and at the same time in French non-diabetic subjects [46], was at *MTNR1B* where the glucose-raising allele was also associated with an increased risk of T2D (Figure 1). An extension of this approach [47], replicated the signals at *GCK*, *G6PC2* and *GCKR* and identified nine novel associations for fasting glucose (*ADCY5*, *MADD*, *ADRA2A*, *CRY2*, *FADS1*, *GLIS3*, *SLC2A2*, *PROX1* and *C2CD4B*) and one influencing fasting insulin and HOMA-IR (near *IGF1*). By evaluating their impact on T2D risk, the MAGIC investigators found novel associations with T2D at an additional five loci (*ADCY5*, *PROX1*, *GCK*, *GCKR* and *DGKB-TMEM195*) as well as replicating the known signals at *TCF7L2* and *SLC30A8* (Figure 1).

An accompanying study by the MAGIC investigators examined glucose levels 2h after an oral glucose challenge (2h glucose), a quantitative measure also used in the diagnosis of T2D [48]. They reported genome-wide associations with five loci: of which two had been found previously (*GCKR*, *TCF7L2*) and three were novel (*GIPR*, *ADCY5*, *VPS13C*). *ADCY5* was shown to be associated with both T2D risk and fasting glucose, and it was suggested that *GIPR* might play a role in T2D pathophysiology. More precisely, *GIPR* is the receptor for GIP, one of the two gastrointestinal hormones that stimulate insulin response after an oral glucose challenge (incretins). In individuals with T2D, GIP action is reduced, whereas its secretion does not seem to be altered. There is also increasing evidence to support an important role for *GIPR* signalling in the promotion of  $\beta$ -cell function and survival [49]. *GIPR* is therefore a biologically plausible candidate for mediating insulin secretion after oral glucose

challenge and an ideal therapeutic target for T2D. Physiological characterization of loci identified from both 2h glucose and the fasting glycaemic traits using measures of insulin processing, secretion and sensitivity showed that the loci influence glycaemic regulation through diverse pathways [50].

These meta-analyses for continuous metabolic traits in healthy individuals identified additional T2D loci and provided valuable insight into the mechanisms underlying the clinical progression to diabetes. The analysis of T2D as a clinical phenotype has yielded many new signals which might not have been found if not for the hypothesis-free approach of GWAS, such as the potential importance of genes involved in cell-cycle regulation. On the other hand, the analysis of quantitative traits has identified some more ‘obvious’ candidate genes, for example, *GCK* and *GCKR* involved in glucose sensing and *SLC2A2* (*GLUT2*) involved in glucose transport. One important observation was that not all loci raising fasting glucose levels are associated with T2D. For example, *ADCY5* and *MADD* both have similar allele frequencies and effect sizes on fasting glucose; while *ADCY5* is strongly associated with T2D [odds ratio (OR) = 1.12,  $P = 5.5 \times 10^{-21}$ ], *MADD* is not (OR = 1.01,  $P = 0.3$ ) [47]. This finding suggests that the molecular mechanism by which the fasting glucose is raised is important in determining progression to diabetes. Indeed, a recent study suggests that variants in *G6PC2* may raise an individual’s glucostat set point and up-regulate fasting glucose, yet additional compromise of the pancreatic  $\beta$  cell function would be required to progress to T2D [40]. The MAGIC results also highlighted the potential role of the circadian system (*MTNR1B* and *CRY2*) in influencing fasting glucose levels. Evidence from the mouse has previously shown this pathway to be implicated in glucose homeostasis [51] and this will be an interesting angle for further exploration.

The first locus associated with insulin resistance and hyperinsulinaemia was *IRS1* [52] (Figure 1). Overall, only two loci for fasting insulin and HOMA-IR were identified, compared with 14 associations with fasting glucose. A similar observation was made with the emerging T2D loci. The initial T2D GWAS results mostly identified loci impacting  $\beta$  cell function, consistent with the idea that common T2D risk alleles primarily act through  $\beta$  cell dysfunction [53]. Reasons that might explain why few insulin resistance-associated loci have been discovered through GWAS approaches include: variants that

increase insulin resistance being less common in the population and/or with smaller effect sizes making them harder to detect; study design (including ascertainment criteria); and lower heritability of insulin resistance traits (which may also be poorly correlated with insulin resistance at the tissue or molecular level). Alternatively, there may simply be fewer variants that affect insulin resistance [53].

The largest meta-analysis for T2D so far, termed the DIAGRAM+ meta-analysis has recently been carried out by the DIAGRAM consortium [54]. They combined genome-wide association data for 8130 T2D cases and 38 987 controls from eight cohorts of European descent, giving an effective sample size of twice that of the original DIAGRAM analysis described above [33]. SNPs from 23 autosomal regions and one signal on the X chromosome were followed up in a second stage. Overall, 14 signals achieved genome-wide significance, of which, two had been previously reported (*MTNR1B*, *IRS1*). Novel associations were identified at *BCL11A*, *ZBED3*, *KLF14*, *TP53INP1*, *CHCHD9*, *KCNQ1*, *CENTD2*, *HMG2*, *HNF1A*, *ZFAND6*, *PRC1* and *DUSP9* (Figure 1), where the signal at *KCNQ1* is independent from the previously identified signal in East-Asian samples [35, 36].

Although there were examples of T2D risk variants that exert their primary effects on insulin action, overall, the results from the DIAGRAM+ meta-analysis results also supported the observation that the contribution to T2D is through  $\beta$  cell dysfunction. Their results also highlighted an interesting overlap between loci predisposing to T2D and other common traits. In fact eight of the newly discovered autosomal loci (near *ADCY5*, *BCL11A*, *ZBED3*, *KLF14*, *CHCHD9*, *HMG2*, *HNF1A* and *PRC1*) also exhibit independent associations ( $P < 1 \times 10^{-6}$ ) with phenotypes other than T2D, such as adult height (*HMG2*) [55], basal cell carcinoma (*KLF14*) [56], low-density lipoprotein (LDL) cholesterol (*HNF1A*) [57] and more recently, birth weight (*ADCY5*) [58]. *KCNQ1* and *KLF14* have also recently been shown to have parental-origin-specific associations [59]. This study also found a novel parent-of-origin specific effect at 11p15 which confers risk when paternally inherited, but is protective when transmitted maternally.

Finally, a GWAS carried out in 5643 discovery and 84 605 validation samples of European origin identified an association with T2D at 2q24, where the less common alleles are associated with reduced

T2D risk and are also nominally associated with lower fasting glucose and HOMA-IR [60]. This region encompasses two genes *RMBS1* and *ITGB6*, with the most associated SNP lying within intron 3 of *RMBS1*.

## WHAT HAVE WE LEARNED FROM GWAS?

To date, 44 T2D susceptibility variants have been identified. Figure 1 illustrates the timeline and method by which these variants were discovered, showing the marked increase owed to the advent of GWAS. Together, the loci explain  $\sim 10\%$  of observed familial clustering in Europeans [54]. This is likely to reflect the fact that, despite the large sample sizes currently available, studies are still underpowered to discover genetic effects with small effect sizes. Of the 12 novel loci identified in the recent DIAGRAM+ analysis, statistical power to detect association at  $P = 1 \times 10^{-5}$  (based on the Stage 1 sample sizes, Stage 2 ORs, risk allele frequencies in HapMap CEU, assuming an additive model and 8% population prevalence) only reached 80% for one signal. This observation suggests that there are likely to be many additional signals of similar effect and frequency that will be found through ongoing iterations of the genome-wide meta-analysis approach. This principle, has recently been illustrated in a meta-analysis of variants contributing to height [61], and is shown in Figure 1, whereby the average T2D OR of newly discovered T2D risk loci has remained fairly constant in recent years.

The MAGIC GWAS of quantitative traits not only highlighted novel loci influencing disease risk and potential pathways for therapeutic intervention [62], but they have begun to elucidate the underlying physiology. For example, the fasting glucose-raising allele in *GCKR* associated with an increased risk of T2D, is also associated with reduced insulin sensitivity [50].

Methods to predict future disease onset, by incorporating associated gene variants into a risk score [63, 64], or machine-learning approaches [65] have found that common genetic variants offer only a marginal improvement on disease risk prediction compared with common epidemiological risk factors alone [66]. This is likely to be due to the small proportion of genetic heritability explained by the variants identified so far. Incorporation of additional variants as they are discovered and gene-gene and/or

gene–environment interactions might improve the predictive power of these approaches.

As described elsewhere [67], it is also important to remember that the effect sizes found thus far are not a reflection of their biological or clinical significance. Although their predictive value may be small, they might point to important biological pathways which could be targeted for therapeutic intervention. Two existing examples of where this is the case are *PPARG* and *KCNJ11*, both of which, had they not already been known, would have been found using the genome-wide approaches. An example arising from GWAS is *SLC30A8*, for which the biology and potential pathways for therapeutic intervention, are starting to be understood [68]. Studies investigating the impact of genetic variation on response to T2D therapy have shown that *TCF7L2* genotype influences the efficiency of the response to sulphonylureas in those with common forms of T2D [69, 70].

## WHAT NEXT?

To identify new susceptibility loci, there is likely to be another wave of GWAS, including data from populations of different ancestry, most notably populations from Asia, further increasing the sample size and improving power. Studies so far have primarily focused on European populations, however it has been shown that genetic variation is greatest in populations of African ancestry [71]. Studies in these populations and other high risk ethnicities (Arabic, native Indian) will potentially enable us to identify novel associations with T2D, and to narrow down the existing signals.

Alternative GWAS strategies which enrich for genetic effects to improve power (for example, by selecting cases with more severe, earlier onset disease and family history of T2D), and original GWAS study designs (such as response to an anti-diabetic treatment or T2D in the presence of extreme obesity) might help to uncover new loci and/or important pathways. In addition to GWAS, complementary epigenomic approaches such as exploring DNA methylation stratified according to disease susceptibility haplotypes [72] and methods which integrate functional genomics data, from model organisms for example [73], may provide further insight into T2D aetiology.

Structural variants, or CNVs were originally thought to potentially account for a greater

proportion of the missing heritability, however for common CNVs this no longer seems probable [21] and the possible contribution of less common and/or rare variants has received much attention. It is thought that though less common they could contribute substantially to the missing heritability, by having large effect sizes [24]. These variants are not common enough to be detected by the current genotyping arrays, and denser genotyping arrays incorporating information from the 1000 Genomes Project ([www.1000genomes.org](http://www.1000genomes.org); such as the recent HumanOmni2.5-Quad BeadChip from Illumina, or the Axiom Genome-Wide Array Plates from Affymetrix) or through direct sequencing (either whole exome, or whole genome) will be required for the detection of these variants. New analysis methods will be required to analyse the rare variant data being produced since the analysis of each variant individually will be underpowered [74, 75].

Genotype imputation has become an invaluable tool to increase the number of SNPs that can be tested for association, thereby improving the power of a study and enabling results across genotyping platforms to be combined to facilitate meta-analysis. However, it has also been shown that the error rate of imputation increases as the minor allele frequency decreases [34]. Although strategies such as using a combination of haplotype reference panels from different populations might improve imputation accuracy for these rarer variants, validation by *de novo* genotyping may still be required.

A recent and impressive display of collaboration between meta-analysis consortia for metabolic and atherosclerotic/cardiovascular diseases and traits has been in the design and manufacture of the Cardio-Metabochip. A custom Illumina iSelect genotyping array designed to test ~200,000 SNPs identified through genome-wide meta-analyses including T2D and glycaemic traits. The SNPs have been selected to provide fast replication of existing association signals and for both locus and signal fine-mapping through the inclusion of known 1000 genomes variants. These efforts are still underway, and we are eagerly anticipating the results.

Of course, an association at a SNP does not necessarily mean the nucleotide change at the associated SNP leads to biological consequences. The associated SNP is likely to be highly correlated to a number of potentially functional variants, possibly with larger effects, in a region of linkage disequilibrium. Additional re-sequencing and functional

studies will be required to identify the genomic regions harbouring the causal variants, and to subsequently define how those variants impact disease risk.

## CONCLUSIONS

Over a relatively short time, GWAS have changed the landscape of diabetes genetics. Although the majority of variants discovered so far have been common with small effects, we have uncovered many new loci, leading to exciting insights into the genetic architecture of T2D. As we start to delve deeper into the previously uncaptured rare variation through whole exome and genome sequencing, we expect to uncover many new findings. Next steps will be to explore the associated regions to identify causal variants, and once identified, to understand the underlying mechanism through which the causal variants impact diabetes. Ultimately, the aim is to translate the knowledge we have gained into individualized preventive and therapeutic medicine. There is still a large amount of work to be done, but significant progress has already been made and the pace is sure to continue.

### Key Points

- GWAS have dramatically increased the number of known T2D susceptibility loci.
- To date, 44 T2D susceptibility variants have been identified, explaining ~10% of observed familial clustering in Europeans.
- Meta-analysis of genome-wide association scans has improved the power to detect statistically significant associations with T2D.
- The analysis of related quantitative traits has uncovered new loci associated with T2D and potential pathways for therapeutic intervention.
- Not all loci raising fasting glucose levels are associated with T2D, suggesting that the mechanism by which fasting glucose levels are raised might be important.

### Acknowledgements

I.B. acknowledges funding from the Wellcome Trust (grant 077016/Z/05/Z), United Kingdom NIHR Cambridge Biomedical Research Centre, the MRC Centre for Obesity and Related Metabolic Diseases and BDA 08/0003704.

### References

1. Shaw JE, Sicree RA, Zimmet PZ. Global estimates of the prevalence of diabetes for 2010 and 2030. *Diabetes Res Clin Pract* 2010;**87**:4–14.
2. WHO. *Fact sheet N° 312*. <http://www.who.int/mediacentre/factsheets/fs312/en/index.html>.
3. Alberti G, Zimmet P, Shaw J, *et al.* Type 2 diabetes in the young: the evolving epidemic: the international diabetes federation consensus workshop. *Diabetes Care* 2004;**27**: 1798–1811.
4. Punnose J, Agarwal MM, Bin-Uthman S. Type 2 diabetes mellitus among children and adolescents in Al-Ain: a case series. *East Mediterr Health J* 2005;**11**:788–97.
5. Reinehr T. Clinical presentation of type 2 diabetes mellitus in children and adolescents. *Int J Obes* 2005;**29**(Suppl 2):S105–10.
6. Barroso I. Genetics of type 2 diabetes. *Diabet Med* 2005;**22**: 517–35.
7. Poulsen P, Kyvik KO, Vaag A, *et al.* Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance—a population-based twin study. *Diabetologia* 1999;**42**:139–45.
8. Florez JC, Hirschhorn J, Altshuler D. The inherited basis of diabetes mellitus: implications for the genetic analysis of complex traits. *Annu Rev Genomics Hum Genet* 2003;**4**: 257–91.
9. Vaxillaire M, Froguel P. Monogenic diabetes in the young, pharmacogenetics and relevance to multifactorial forms of type 2 diabetes. *Endocr Rev* 2008;**29**:254–64.
10. Huang QY, Cheng MR, Ji SL. Linkage and association studies of the susceptibility genes for type 2 diabetes. *Yi Chuan Xue Bao* 2006;**33**:573–89.
11. Guan W, Pluzhnikov A, Cox NJ, *et al.* Meta-analysis of 23 type 2 diabetes linkage studies from the International Type 2 Diabetes Linkage Analysis Consortium. *Hum Hered* 2008;**66**:35–49.
12. Altshuler D, Hirschhorn JN, Klannemark M, *et al.* The common PPAR $\gamma$  Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* 2000;**26**:76–80.
13. Gloyn AL, Weedon MN, Owen KR, *et al.* Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (*KCNJ11*) and SUR1 (*ABCC8*) confirm that the *KCNJ11* E23K variant is associated with type 2 diabetes. *Diabetes* 2003;**52**: 568–72.
14. Hani EH, Boutin P, Durand E, *et al.* Missense mutations in the pancreatic islet beta cell inwardly rectifying K<sup>+</sup> channel gene (*KIR6.2/BIR*): a meta-analysis suggests a role in the polygenic basis of Type II diabetes mellitus in Caucasians. *Diabetologia* 1998;**41**:1511–5.
15. Sandhu MS, Weedon MN, Fawcett KA, *et al.* Common variants in *WFS1* confer risk of type 2 diabetes. *Nat Genet* 2007;**39**:951–3.
16. Winckler W, Weedon MN, Graham RR, *et al.* Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes. *Diabetes* 2007;**56**:685–93.
17. Grant SF, Thorleifsson G, Reynisdottir I, *et al.* Variant of transcription factor 7-like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. *Nat Genet* 2006;**38**:320–3.
18. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996;**273**:1516–7.
19. International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005;**437**:1299–320.

20. Redon R, Ishikawa S, Fitch KR, *et al.* Global variation in copy number in the human genome. *Nature* 2006;**444**:444–54.
21. Conrad DF, Pinto D, Redon R, *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* 2010;**464**:704–12.
22. Kruglyak L. The road to genome-wide association studies. *Nat Rev Genet* 2008;**9**:314–8.
23. Purcell S, Neale B, Todd-Brown K, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;**81**:559–75.
24. McCarthy MI, Abecasis GR, Cardon LR, *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008;**9**:356–69.
25. Pe'er I, Yelensky R, Altshuler D, *et al.* Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* 2008;**32**:381–5.
26. Sladek R, Rocheleau G, Rung J, *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007;**445**:881–5.
27. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, *et al.* A variant in *CDKAL1* influences insulin response and risk of type 2 diabetes. *Nat Genet* 2007;**39**:770–5.
28. Zeggini E, Weedon MN, Lindgren CM, *et al.* Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 2007;**316**:1336–41.
29. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT LU, and Novartis Institutes of BioMedical Research. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007;**316**:1331–6.
30. Scott LJ, Mohlke KL, Bonnycastle LL, *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007;**316**:1341–5.
31. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;**447**:661–78.
32. Frayling TM, Timpson NJ, Weedon MN, *et al.* A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 2007;**316**:889–94.
33. Zeggini E, Scott LJ, Saxena R, *et al.* Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 2008;**40**:638–45.
34. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet* 2010;**11**:499–511.
35. Yasuda K, Miyake K, Horikawa Y, *et al.* Variants in *KCNQ1* are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet* 2008;**40**:1092–7.
36. Unoki H, Takahashi A, Kawaguchi T, *et al.* SNPs in *KCNQ1* are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat Genet* 2008;**40**:1098–102.
37. Shu XO, Long J, Cai Q, *et al.* Identification of new genetic risk variants for type 2 diabetes. *PLoS Genet* 2010;**6**:e1001127.
38. Tsai FJ, Yang CF, Chen CC, *et al.* A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese. *PLoS Genet* 2010;**6**:e1000847.
39. Yamauchi T, Hara K, Maeda S, *et al.* A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at *UBE2E2* and *C2CD4A-C2CD4B*. *Nat Genet* 2010;**42**:864–8.
40. Bouatia-Naji N, Rocheleau G, Van Lommel L, *et al.* A polymorphism within the *G6PC2* gene is associated with fasting plasma glucose levels. *Science* 2008;**320**:1085–8.
41. Chen WM, Erdos MR, Jackson AU, *et al.* Variations in the *G6PC2/ABCB11* genomic region are associated with fasting glucose levels. *J Clin Invest* 2008;**118**:2620–8.
42. Weedon MN, Clark VJ, Qian Y, *et al.* A common haplotype of the glucokinase gene alters fasting glucose and birth weight: association in six studies and population-genetics analyses. *Am J Hum Genet* 2006;**79**:991–1001.
43. Orho-Melander M, Melander O, Guiducci C, *et al.* Common missense variant in the glucokinase regulatory protein gene is associated with increased plasma triglyceride and C-reactive protein but lower fasting glucose concentrations. *Diabetes* 2008;**57**:3112–21.
44. Vaxillaire M, Cavalcanti-Proenca C, Dechaume A, *et al.* The common P446L polymorphism in *GCKR* inversely modulates fasting glucose and triglyceride levels and reduces type 2 diabetes risk in the DESIR prospective general French population. *Diabetes* 2008;**57**:2253–7.
45. Prokopenko I, Langenberg C, Florez JC, *et al.* Variants in *MTNR1B* influence fasting glucose levels. *Nat Genet* 2009;**41**:77–81.
46. Bouatia-Naji N, Bonnefond A, Cavalcanti-Proenca C, *et al.* A variant near *MTNR1B* is associated with increased fasting plasma glucose levels and type 2 diabetes risk. *Nat Genet* 2009;**41**:89–94.
47. Dupuis J, Langenberg C, Prokopenko I, *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 2010;**42**:105–16.
48. Saxena R, Hivert MF, Langenberg C, *et al.* Genetic variation in *GIPR* influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet* 2010;**42**:142–8.
49. Widenmaier SB, Kim SJ, Yang GK, *et al.* A GIP receptor agonist exhibits beta-cell anti-apoptotic actions in rat models of diabetes resulting in improved beta-cell function and glycemic control. *PLoS ONE* 2010;**5**:e9590.
50. Ingelsson E, Langenberg C, Hivert MF, *et al.* Detailed physiologic characterization reveals diverse mechanisms for novel genetic Loci regulating glucose and insulin metabolism in humans. *Diabetes* 2010;**59**:1266–75.
51. Rudic RD, McNamara P, Curtis AM, *et al.* *BMAL1* and *CLOCK*, two essential components of the circadian clock, are involved in glucose homeostasis. *PLoS Biol* 2004;**2**:e377.
52. Rung J, Cauchi S, Albrechtsen A, *et al.* Genetic variant near *IRS1* is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat Genet* 2009;**41**:1110–5.
53. Florez JC. Newly identified loci highlight beta cell dysfunction as a key cause of type 2 diabetes: where are the insulin resistance genes? *Diabetologia* 2008;**51**:1100–10.
54. Voight BF, Scott LJ, Steinthorsdottir V, *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 2010;**42**:579–89.
55. Weedon MN, Lettre G, Freathy RM, *et al.* A common variant of *HMG2* is associated with adult and childhood height in the general population. *Nat Genet* 2007;**39**:1245–50.



56. Stacey SN, Sulem P, Masson G, *et al.* New common variants affecting susceptibility to basal cell carcinoma. *Nat Genet* 2009;**41**:909–14.
57. Kathiresan S, Willer CJ, Peloso GM, *et al.* Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* 2009;**41**:56–65.
58. Freathy RM, Mook-Kanamori DO, Sovio U, *et al.* Variants in *ADCY5* and near *CCNL1* are associated with fetal growth and birth weight. *Nat Genet* 2010;**42**:430–5.
59. Kong A, Steinthorsdottir V, Masson G, *et al.* Parental origin of sequence variants associated with complex diseases. *Nature* 2009;**462**:868–74.
60. Qi L, Cornelis MC, Kraft P, *et al.* Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. *Hum Mol Genet* 2010;**19**:2706–15.
61. Lango Allen H, Estrada K, Lettre G, *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 2010;**467**:832–8.
62. Lyssenko V, Nagorny CL, Erdos MR, *et al.* Common variant in *MTNR1B* associated with increased risk of type 2 diabetes and impaired early insulin secretion. *Nat Genet* 2009;**41**:82–8.
63. Lyssenko V, Jonsson A, Almgren P, *et al.* Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med* 2008;**359**:2220–32.
64. Meigs JB, Shrader P, Sullivan LM, *et al.* Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med* 2008;**359**:2208–19.
65. Wei Z, Wang K, Qu HQ, *et al.* From disease association to risk assessment: an optimistic view from genome-wide association studies on type 1 diabetes. *PLoS Genet* 2009;**5**: e1000678.
66. Liu S, Song Y. Building genetic scores to predict risk of complex diseases in humans: is it possible? *Diabetes* 2010;**59**: 2729–31.
67. Florez JC. Clinical review: the genetics of type 2 diabetes: a realistic appraisal in 2008. *J Clin Endocrinol Metab* 2008;**93**: 4633–42.
68. Nicolson TJ, Bellomo EA, Wijesekara N, *et al.* Insulin storage and glucose homeostasis in mice null for the granule zinc transporter ZnT8 and studies of the type 2 diabetes-associated variants. *Diabetes* 2009;**58**:2070–83.
69. Pearson ER, Donnelly LA, Kimber C, *et al.* Variation in *TCF7L2* influences therapeutic response to sulfonylureas: a GoDARTs study. *Diabetes* 2007;**56**:2178–82.
70. Schroner Z, Javorsky M, Tkacova R, *et al.* Effect of sulphonylurea treatment on glycaemic control is related to *TCF7L2* genotype in patients with type 2 diabetes. *Diabetes Obes Metab* 2011;**13**:89–91.
71. Frazer KA, Ballinger DG, Cox DR, *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007;**449**:851–61.
72. Bell CG, Finer S, Lindgren CM, *et al.* Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus. *PLoS ONE* 2010;**5**:e14040.
73. Guan Y, Ackert-Bicknell CL, Kell B, *et al.* Functional genomics complements quantitative genetics in identifying disease-gene associations. *PLoS Comput Biol* 2010;**6**: e1000991.
74. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* 2008;**83**:311–21.
75. Liu DJ, Leal SM. A novel adaptive method for the analysis of next-generation sequencing data to detect complex trait associations with rare variants due to gene main effects and interactions. *PLoS Genet* 2010;**6**:e1001156.
76. Prokopenko I, McCarthy MI, Lindgren CM. Type 2 diabetes: new genes, new understanding. *Trends Genet* 2008;**24**: 613–21.