Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

Progress in defining the genetic contribution to type 2 diabetes susceptibility

Andrew P Morris

Candidate gene, genome-wide association, exome array and sequencing studies have identified more than 140 loci associated with type 2 diabetes (T2D) susceptibility. In this review, progress in understanding the genetic architecture of T2D susceptibility across diverse populations and in localising potential causal variants for the disease through fine-mapping studies is discussed. The additional insights gained from these genetic studies into novel molecular mechanisms and pathophysiology underlying T2D susceptibility are described, and the prospects for future genomic investigations of the disease are considered.

Address
Department of Biostatistics, University of Liverpool, Liverpool, UK

Corresponding author: Morris, Andrew P (apmorris@liverpool.ac.uk)

Current Opinion in Genetics & Development 2018, **50**:41–51

This review comes from a themed issue on **Molecular and genetic basis of disease**

Edited by **Ines Barroso** and **Jose Florez**

https://doi.org/10.1016/j.gde.2018.02.003

Introduction

Type 2 diabetes (T2D) mellitus is a common complex disease that currently affects more than 400 million people across the globe. The disease is characterised by insulin resistance and beta cell dysfunction [1], and can lead to a range of microvascular and macrovascular complications [2], increased risk of cardiovascular disease [3] and reduced life-expectancy [4]. Whilst T2D most often occurs as a downstream consequence of obesity, the disease is observed to run in families, with a sibling relative risk of ~2.8 [5] and estimated heritability of 30–70% [6]. Improved awareness of the genetic contribution to T2D susceptibility has thus been considered an important avenue to understanding the underlying pathophysiology of the disease that will help to address the limitations of currently available preventive and therapeutic options.

The earliest genetic investigations of T2D susceptibility focussed on small-scale studies of candidate genes, often identified through family-based linkage analyses or relevant biological insight from monogenic forms of the disease. The results of these studies were often difficult to validate, and success was limited to the identification of variants in a handful of loci including *PPARG* [7], *KCNJ11* [8], *TCF7L2* [9], and *WFS1* [10]. By contrast, the first well-designed genome-wide association studies (GWAS) of T2D susceptibility made use of a 'hypothesis-free' approach that interrogates (many) hundreds of thousands of common single nucleotide polymorphisms (SNPs) across the genome, typically defined to have minor allele frequency of at least 5%. These GWAS substantially increased the number of confirmed disease loci, including *CDKAL1*, *CDKN2A-B*, *HHEX-IDE*, *IGF2BP2*, *SLC30A8*, and *FTO* (primary effect on risk via obesity) [11–16]. However, it is important to note that GWAS loci have historically been named by the gene that maps closest to the lead SNP (i.e. that with strongest association signal), but that this does not imply causality.

Early endeavours to understand the genetic contribution to T2D susceptibility were enhanced through international collaborative efforts that increased power to detect disease loci through meta-analysis of GWAS for the disease and related glycemic traits by the DIAGRAM Consortium and MAGIC Investigators [17–19] in European ancestry populations. Such meta-analyses typically involve the provision of association summary statistics for each SNP (such as allelic odds-ratios and confidence intervals, p-values, and allele frequencies) to a 'central analysis team' across GWAS contributing to the consortium, without the need to exchange individual-level genotype and phenotype data. These efforts expanded the genetic landscape of T2D susceptibility to more than 40 loci at genome-wide significance ($p < 5 \times 10^{-8}$), but also revealed important novel biology, with effects of associated SNPs on both beta-cell function and insulin action, and enrichment for association signals in/near genes involved in cell cycle regulation.

Despite the success of early GWAS meta-analyses in identifying novel loci contributing to T2D susceptibility, challenges remain in the utility of the findings of these studies for clinical translation through informing development of novel therapeutics for the disease and enabling implementation of personalised medicine strategies [20]. First, association signals at GWAS loci explained relatively little (~10%) of the heritability of T2D

susceptibility, inhibiting personalised risk prediction. Second, association signals typically extended over large genomic intervals because of linkage disequilibrium between common SNPs in European ancestry populations, limiting the resolution of fine-mapping efforts to localise the causal variant(s) at each locus, and identify the effector transcript(s) through which their effects are mediated. In this article, progress in addressing these challenges will be reviewed, focussing on larger GWAS meta-analyses of common SNPs across diverse populations, and whole-genome and whole-exome sequencing efforts to access lower frequency genetic variants.

Additional T2D susceptibility loci identified through GWAS

Since 2010, the most widespread approach to the discovery of additional loci contributing to T2D susceptibility has been through GWAS. These efforts have been bolstered by improved efficiency of GWAS genotyping technology, enabling interrogation of larger numbers of SNPs that better cover common genetic variation across populations in increased sample sizes. Furthermore, methodological innovations, such as imputation [21], which enables prediction of genotypes at SNPs not typed on GWAS arrays, but which are present in high-density reference panels of whole-genome sequence data, such as those from the 1000 Genomes Project [22] and Haplotype Reference Consortium [23], allow association testing at millions of variants across the genome (Table 1).

As described above, the largest GWAS meta-analyses for T2D susceptibility, to date, have been undertaken by the DIAGRAM Consortium in populations of European ancestry [24,25•], predominantly because of existing infrastructure, sample availability, and relatively poor coverage of common genetic variation in other major ethnic groups by early genotyping arrays [26]. However, European ancestry populations do not fully characterise T2D risk variants in other ethnic groups, and GWAS meta-analyses have also been undertaken in East Asians [27,28], South Asians [29], Hispanics/Latinos [30,31], and African Americans [32]. These studies have highlighted overlap in T2D susceptibility loci between ethnicities and provided evidence that genetic risk scores derived from European ancestry GWAS are transferrable to other population groups [33]. Investigations across ethnic groups have revealed minimal evidence of heterogeneity in allelic odds-ratios on T2D susceptibility at lead SNPs, despite substantial differences in allele frequencies [34,35•]. T2D risk alleles across the genome also demonstrate consistent directions of effect across diverse populations at loci with only nominal association with disease susceptibility [34]. These observations imply that common causal variants at many T2D susceptibility loci are shared across ancestry groups, and arose prior to human population migration out of Africa, prompting trans-ethnic GWAS meta-analyses to maximise sample size and power for additional discoveries [34,36•]. On the other hand, lower-frequency variation is more likely to be ethnic- or even population-specific, highlighting the important of undertaking genetic investigations of T2D susceptibility across diverse ancestries. For example, GWAS undertaken in the isolated Greenlandic population highlighted a missense variant in *TBC1D4* that confers muscle insulin resistance and T2D risk, which is rare or monomorphic in other ancestry groups [37]. Furthermore, differences in LD structure between common variants across ethnicities was beneficial for localising causal variants underlying association signals, as described below.

The genetic architecture of type 2 diabetes susceptibility

GWAS have continued to be extremely successful in identifying loci contributing to T2D susceptibility. Association signals at the majority of these loci are defined by common lead SNPs with modest effects on the disease (allelic odds-ratios less than 1.5). As a consequence, these loci jointly explain no more than ~20% of the heritability of the disease, and thus have limited predictive power for future disease in unaffected individuals when compared with traditional clinical and lifestyle risk factors [38]. However, genetic risk scores that are derived from SNPs, including those with nominal associations with T2D susceptibility that map outside of loci attaining genome-wide significance, do show promise for disease prediction, after accounting for body-mass index, age and sex [39].

GWAS are not designed to capture rare genetic variation (typically defined to have minor allele frequency less than 0.5%), even after supplementation with imputation. There has therefore been considerable support for the notion that much of the 'missing heritability' of complex human traits, including T2D susceptibility, could be explained by lower frequency variation, which can best be assayed through sequencing. To investigate this hypothesis, the GoT2D/T2D-GENES Consortium undertook the largest sequencing study (whole-genome and whole-exome) of T2D susceptibility, to date, in more than 15,000 individuals of diverse ancestry [40••]. A coding variant (*PAX4* p.Arg192His) attained genome-wide significance at the *GCC1* locus that was common in East Asian ancestry populations (minor allele frequency ~10%), but virtually absent from other ethnic groups. These data provided no evidence to support the 'synthetic association' hypothesis [41], under which common GWAS SNP signals can be explained by rare variants that are not interrogated through genotyping and imputation. Modelling of disease architecture demonstrated that: (i) the whole-genome association data were consistent with a 'common polygenic' model in which large numbers of common variants of modest effect explain about 75% of T2D heritability; and (ii) across the exome, the overall contribution of rare and low-frequency coding