Strapline: Genetics of T2DM in 2016

Biological and translational insights from T2DM genetics

Mark I. McCarthy

Type 2 diabetes mellitus (T2DM) is a major global health challenge. Development of more effective strategies for prevention and therapy depends on an improved understanding of its pathogenetic mechanisms. 2016 ends a period during which large-scale discovery of risk alleles for T2DM became routine and heralds a shift in research focus towards their exploitation to fuel mechanistic insights.

Over the past decade, the dominant approach to genetic discovery for complex traits has involved genome-wide surveys of (mostly) common variant sites through array-based genome-wide association studies (GWAS). Successive GWAS, conducted in ever larger, more diverse, data sets have identified >100 loci that are robustly associated with type 2 diabetes mellitus (T2DM). The risk variants at these loci typically have modest effects on T2DM predisposition, and collectively account for only 10% of overall disease risk, far short of the proportion that twin and family studies indicate are attributable to overall genetic variation[1].

A variety of explanations have been proffered for this disparity (often described as 'missing heritability'). One of these explanations implicates risk alleles of lower frequency — these alleles are mostly invisible in existing GWAS as these have primarily interrogated common variants, but can be identified through the more comprehensive variant evaluation provided by next-generation sequencing.

The manuscript by Fuchsberger and colleagues[2] describes the first large-scale use of sequencing to explore risk of T2DM, combining whole-genome sequencing data from 2,657 controls and T2DM subjects with whole-exome sequencing data from a further 12,940 individuals. These data were extended, capturing a subset of the sequence-detected variants, through genotyping in 111,548 additional individuals. This study identified many statistically significant association signals for T2DM; however, almost all were common, and many lay within existing GWAS signals. Where the risk variants mapped to protein-coding sequence, these discoveries provided valuable clues regarding which regional transcript was driving GWAS signals hitherto detected through non-coding (regulatory) variation.

Evidence supporting a contribution of lower frequency variants to the risk of T2DM was scant. The only compelling signal involved an excess of rare loss-of-function alleles within the set of genes previously implicated in monogenic diabetes mellitus. The paucity of low frequency association signals suggests little or no selective pressure for T2DM risk alleles during human prehistory. This is consistent with a limited impact of T2D risk alleles on evolutionary fitness, reflecting the recent emergence of this disease, and its predominant impact on those of post-reproductive age. Evidence that most variation in individual risk of T2DM is found in common variants, shared amongst many of us, limits the potential to reassign individuals with T2DM into discrete disease subtypes.

Although rare variants are unlikely to explain much of the risk variance, their identification and characterization remains a valid and important objective. Rare alleles detectable in genome-wide

studies will, for reasons of power, generally have large phenotypic impact, and, as decades of Mendelian disease research have shown, such discoveries often lead to clear and rapid mechanistic insights. Whilst there are few examples of high-impact common variants in outbred populations, such variants can rise to higher frequency in isolated populations, through chance (genetic drift) or local adaptation. The clearest example relates to a variant in *TBC1D4* detected specifically in the Greenlandic Inuit population; in homozygous carriers, this variant is associated with a 10-fold increase in risk of T2DM compared to those with other genotypes[3]. In this population, the allele has risen to a frequency of 17%, probably reflecting historical adaptation for a diet unusually rich in marine fat.

A further example of a high-impact, geographically-restricted variant emerged in 2016. Minster and colleagues conducted a GWAS for BMI in people from Samoa, a population at high risk of obesity and T2DM[4]. The key finding was a replicated association signal localized to a coding variation in *CREBRF*, a gene implicated in fat deposition. The BMI-raising allele has a frequency of 27% in people from Samoa, but is essentially absent outside the Pacific. This frequency difference might reflect selection for a 'thrifty' allele which, by promoting efficient fat storage, enhanced the survival prospects of carriers during the perilous transoceanic migrations that led to the peopling of the Pacific islands.

Intriguingly, the BMI-raising allele in *CREBRF* was associated with reduced, rather than increased, risk of T2DM. Such variants, which 'row against the epidemiological current', can provide instructive insights into the mechanistic relationships between otherwise highly correlated phenotypes. The probable explanation here is that risk of T2DM reflects not so much total adiposity, but the distribution of that fat. Efficient storage of excess calories in safe adipose sites, promoted by the *CREBRF* allele, lowers the risk of T2DM because it reduces the opportunities for fat deposition in ectopic sites, such as the liver, that is more metabolically harmful. A recent study of insulin-resistance associated alleles in European populations powerfully supports this inference[5].

The major motivation behind genetic discovery for complex traits is the value of risk variants to reveal mechanisms central to disease pathogenesis (see **figure**). Securing these mechanistic inferences has, until very recently, been a major challenge for T2DM. As with other common diseases, most GWAS signals map to regulatory sequences with few clues to the effector genes through which those variants act. In T2DM, many GWAS loci have predominant effects on insulin secretion, focusing attention on the pancreatic islets; because of their relative inaccessibility, the epigenomic and transcriptomic characteristics of pancreatic islets have been poorly captured by projects such as ENCODE and GTEx.

However, the use of sequence-based omics technologies to generate regulatory and expression maps for the human islet has demonstrated that T2DM-risk variants are preferentially located within active islet enhancers. One T2DM GWAS region where this information has generated valuable insights lies next to the gene for *MTNR1B*, which encodes one of the receptors for melatonin. The probable causal variant disrupts a NEUROD1 binding site in an islet enhancer upstream of the gene, and variation at this site influences *MTNR1B* expression[6], highlighting some of the key machinery which connects the GWAS-detected variant to its influence on T2DM predisposition.

The research described by Tuomi and colleagues puts further flesh on the bones of this story[7], demonstrating that melatonin inhibits cAMP generation in islets and blocks insulin release. In a genotype-based recall study, this group showed that the capacity of melatonin to reduce insulin secretion was enhanced in homozygous carriers of the T2DM risk allele. The implication, that excessive melatonin action is detrimental to islet function, seems to be consistent with the idea that

nocturnal secretion of melatonin has the beneficial physiological effect of dialling down insulin production at night, when food intake is reduced.

However, the story is undoubtedly more complex. Loss-of-function coding variants in *MTRN1B* have been associated with increased, rather than reduced, risk of T2DM and some epidemiological studies suggest a similar directional relationship[8]. It seems likely that these contrasting views reflect differences between the impact of an islet-specific GWAS variant and coding variants with wider tissue effects. Thus, whilst the *MTNR1B* example illustrates the growing range of data that can support the assignment of function to T2DM-associated genetic variants, it also highlights the biological complexity that must be addressed.

Mechanistic inference provides a crucial first step towards translation, defining novel therapeutic targets and highlighting potential biomarkers. However, genetics offers additional clinical opportunities. Geneticcs can help to design more effective preventative strategies, for example, through use of Mendelian randomization approaches to identify the specific components of the Western lifestyle that have the strongest influence on risk of T2DM. In this regard, interest is growing in the role of the gut microbiome (and of dietary composition) in promoting obesity and T2DM.

The picture remains confused. Associations between obesity and T2DM and gut microbiome content and diversity are easy to detect, but it is far from clear whether such differences are causal or reactive; for example, some of the differences in the microbiome apparently associated with T2DM, are, in fact, the consequence of metformin treatment. The work of Pedersen and colleagues provides an important contribution to this story[9], connecting insulin resistance, raised levels of circulating branched-chain amino acids (BCAAs) and a microbiome content that promotes BCAA synthesis. With parallel efforts to identify genetic variants that influence microbiome content and diversity[10], it is now becoming possible to establish (utilizing the power of genetic variation to orientate causal networks) the mechanistic relationships between microbiome content, BCAA levels and the risk of T2DM.

The field has advanced beyond recognition in the decade since GWAS arrived. Much remains to be discovered, but the combination of comprehensive elucidation of risk variants, harnessed to genomic information from relevant tissues and physiological studies in humans and in animal models, sets the scene for important translational advances in the coming years.

Mark I. McCarthy is at Oxford Centre for Diabetes, Endocrinology and Metabolism, Churchill Hospital, University of Oxford, Old Road, Headington, Oxford, UK, OX3 7LJ and Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Headington, Oxford, UK, OX3 7BN.

mark.mccarthy@drl.ox.ac.uk

1. Morris, A.P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet.* **44**:981-990 [2012]
2. Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature*. **536**, 41-47 [2016]
3. Moltke, I. *et al.* A common Greenlandic *TBC1D4* variant confers muscle insulin resistance and type 2 diabetes. *Nature*. **512**, 190-193 [2014]
4. Minster, R.L. *et al*. A thrifty variant in CREBRF strongly influences body mass index in Samoans. *Nat Genet.* **48**, 1049-1054 [2016]
5. Lotta, L.A. Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nat Genet*. Nov 14 [2016]
6. Gaulton, K.J. *et al.* Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat Genet*. **47**, 1415-1425 [2015]

7. Tuomi, T. *et al.* Increased melatonin signaling is a risk factor for type 2 diabetes. *Cell Metab.* **23,** 1067-1077 [2016]
8. Bonnefond, A., Karamitri, A., Jockers, R. & Froguel, P. The Difficult Journey from Genome-wide Association Studies to Pathophysiology: The Melatonin Receptor 1B (MT2) Paradigm. *Cell Metab.* **24**:345-347 [2016]
9. Pedersen, H.K. *et al.* Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature*. **535**, 376-381 [2016]
10. Bonder, M.J. *et al.* The effect of host genetics on the gut microbiome. *Nat Genet*. **48**, 1407-1412 [2016].

Competing interest statement

M.McC. is a member of Advisory Boards at NovoNordisk and Pfizer. He has received honoraria from EliLilly, NovoNordisk and Pfizer, and research funding (via the Innovative Medicines Initiative and the Accelerating Medicines Partnership) from Abbvie, Astra Zeneca, Boehringer Ingelheim, Eli Lilly, Janssen, Merck, NovoNordisk, Pfizer, Roche, Sanofi-Aventis, Servier and Takeda.

Figure Legends

Figure 1 | The identification of DNA sequence variants associated with type 2 diabetes mellitus (through a combination of genotyping and sequencing) drives mechanistic insights that can reveal the molecular and physiological basis of disease. In turn, those insights provide a foundation for a range of translational opportunities that might support novel preventative and therapeutic strategies.

Key advances

- **Disease-associated genetic variation can provide critical insights into disease pathogenesis[2,4].**
- **Most of the variation in inherited predisposition to common diseases such as type 2 diabetes mellitus depends on common, shared variants, which mostly have modest effects; however, high-impact variants can be found in isolated populations[2,4].**
- **The intersection of genetic and genomic information from relevant tissues is driving the characterization of pathogenetic mechanisms[2,4,7].**
- **Genetic information is increasingly informing translational efforts, including the capacity to evaluate the contribution of external processes to disease prevalence[9].**

Pullquotes

- *Most variation in individual risk of T2DM is found in common variants, which many of us share.*
- *The field has advanced beyond recognition in the decade since GWAS arrived.*

Author biography: Prof Mark McCarthy is Robert Turner Professor of Diabetic Medicine at the University of Oxford, and Honorary Consultant Physician at the Oxford University Hospitals Trust. He is Wellcome Trust Senior Investigator. His primary interests lie in the genetics of type 2 diabetes and

related phenotypes and in the use of genetic discoveries to fuel biological insights and translational advances.

Copyright assignment (SENT)