

***TCF7L2*: the biggest story in diabetes genetics since HLA?**

E. Zeggini · M. I. McCarthy

Published online: 10 November 2006
© Springer-Verlag 2006

Much has been made over the past decade of the potential for genetics to advance our understanding of the pathogenesis of type 2 diabetes and to ‘revolutionise’ management of this condition [1]. Others have argued that these claims are premature [2]; indeed, some have questioned the contribution of genetic predisposition to the pathogenesis of common forms of type 2 diabetes [3].

In the case of relatively uncommon monogenic and syndromic forms of diabetes, such as maturity onset diabetes of the young (MODY) and neonatal diabetes, identification of rare causal mutations has delivered both knowledge and clinical translation [4, 5]. In contrast, progress in unravelling the genetic architecture of more typical, common, multifactorial type 2 diabetes has been painfully slow [6]. The reasons have been well-rehearsed [7]. The complex web of susceptibility factors—genetic, environmental, social—that contributes to individual risk of developing type 2 diabetes means that most predisposing genetic variants will have only a modest marginal impact on disease risk. The majority of genetic studies performed to date have simply had insufficient power to uncover these reliably [7]. The few type 2 diabetes-susceptibility variants convincingly demonstrated—notably the P12A variant in *PPARG* and E23K in *KCNJ11* [8, 9]—have only modest

effects on disease risk (odds ratios ~1.2), far too small to offer (either individually or in combination) clinically useful predictive testing. Since these variants lie within genes whose products are already known to be therapeutic targets, these particular discoveries have also had limited capacity to deliver novel pathophysiological insights. Among those working on the genetics of type 2 diabetes, there was growing apprehension that these two genes might be providing a representative view of the genetic architecture of type 2 diabetes.

However, recent revelations concerning a novel type 2 diabetes-susceptibility gene (encoding the transcription factor, *TCF7L2* [‘7-like 2’]) show that this is definitely not the case. As two papers in this issue of *Diabetologia* demonstrate [10, 11], common variants in this gene have a marked and reproducible effect on type 2 diabetes risk, identifying sizeable groups of individuals who differ up to twofold in their risk of developing type 2 diabetes, purely as a result of variation at a single nucleotide position within *TCF7L2*.

These studies in Dutch [10] and Indian [11] samples are the latest in a series of reports confirming the powerful effect of *TCF7L2* variation on type 2 diabetes-risk which have followed the initial publication from Iceland in early 2006 [12]. Researchers at Decode Genetics seeking the cause of a previously-identified linkage signal on chromosome 10q [13] found strong associations between type 2 diabetes status and *TCF7L2* variants that they were able to replicate in samples from the USA and Denmark. The effect size in this initial report appeared substantial (each additional copy of the risk allele was associated with an odds ratio of ~1.5), and the strength of the association was impressive ($p \sim 10^{-18}$ overall).

Since the initial report in early 2006, the freezers of diabetes researchers worldwide have been raided and many tens of thousands of samples typed for these same variants.

E. Zeggini · M. I. McCarthy (✉)
Diabetes Research Laboratories, Oxford Centre for Diabetes,
Endocrinology and Metabolism, University of Oxford,
Churchill Hospital,
Old Road, Headington,
Oxford OX3 7LJ, UK
e-mail: mark.mccarthy@drl.ox.ac.uk

E. Zeggini · M. I. McCarthy
Wellcome Trust Centre for Human Genetics,
Roosevelt Drive, Headington,
Oxford OX3 7BN, UK

These studies have, without exception, confirmed those original findings. In UK samples for example, the same *TCF7L2* susceptibility variants were associated with a per-allele odds ratio of ~1.4 [14]. As in the original report, there was clear evidence of a gene dosage effect, such that the 10% of individuals with two copies of the susceptibility allele were at almost twice the risk of developing diabetes as those with none. In participants from the Diabetes Prevention Program, the same *TCF7L2* variants were associated with increased rates of progression from IGT to diabetes (with a hazard ratio of 1.55 between homozygote groups) [15]. Further replications have appeared from analyses in subjects of Amish [16], Finnish [17], French [18] and US [19, 20] origin. Little wonder that one colleague was moved to describe *TCF7L2* as ‘the biggest story in diabetes genetics since HLA’. The speed of confirmation and reproducibility of the findings has certainly been unprecedented.

All very well, you may say, that must be great for the geneticists, but what does all of this mean for our understanding of diabetes? And what difference will this make to the clinical management of this condition? In truth, it is far too early to offer an authoritative answer to such questions, but here are three immediate lessons.

Lesson 1: Type 2 diabetes-susceptibility variants with large effect sizes do exist and can be found

The increase in risk of type 2 diabetes associated with the *TCF7L2* variant alleles so far identified is substantially greater than that associated with variants in the other confirmed type 2 diabetes-susceptibility genes (*PPARG*, *KCNJ11*). The limited data so far available from more comprehensive genotyping efforts across the gene [12, 17, 20] suggest that the SNPs reported in the Icelandic study (and typed in most subsequent studies) have the strongest association with type 2 diabetes. However, it remains entirely possible that there are additional (or alternative) variants within the gene which carry even greater risk. If so, current estimates that the population risk attributed to *TCF7L2* variants lies between 10 and 25% may need to be revised upwards [10–12, 14–20]. Certainly, no functional significance has been attributed to the *TCF7L2* variants implicated so far, all of which are intronic. Evidence that the same associations are apparent in non-European samples [11] increases the chances that these SNPs are aetiological, though studies in populations with more divergent linkage disequilibrium structures (particularly those of African origin) will be most informative in this regard. Further, data from both the Icelandic [12] and UK [14] studies make it clear that the variants concerned are not responsible for the previously reported linkage

signals on chromosome 10 [13, 21]. Either the detection of *TCF7L2* within the region of the linkage signal was serendipitous, or else other (most likely rare, highly penetrant) variants within the gene also contribute to disease risk. Deep resequencing of the gene should help to resolve this question.

From a methodological point of view, identification of a type 2 diabetes-susceptibility variant with such a considerable effect size has several important ramifications. First, it speaks to the value of adequate sample size and replication as mechanisms for delivering robust findings [22]. Unusually, the effect size at *TCF7L2* is such that the association signal can be detected with only a few hundred samples: ensuring adequate power becomes even more essential when seeking to validate variants with lesser effect sizes. We can expect these to be far more numerous than variants like *TCF7L2* [7]. Second, the identification of *TCF7L2* is testament to the validity of susceptibility-gene discovery through indirect linkage disequilibrium-mapping approaches [23]. And third, evidence that such powerful effects exist augurs well for the early success of ongoing efforts to exploit genome-wide association mapping as a tool for further gene discovery efforts. The multiple-testing penalty implicit in such studies means that only effect sizes on this scale are likely to emerge from initial rounds of analysis within individual genome-wide association data sets [23].

Lesson 2: Since we know so little about *TCF7L2*, novel insights into diabetes pathogenesis are guaranteed

At this early stage, precious little is known about the normal role of *TCF7L2*, or how dysfunction predisposes to diabetes. The gene seems to be widely expressed [18] and the transcription factor product is known to be involved in the Wnt signalling cascade. Current evidence strongly supports the idea that the predominant effect of *TCF7L2* dysfunction on type 2 diabetes development is mediated through impairment of insulin secretion [11, 15–17, 20], a finding that would be consistent, for example, with the known effects of other (non-homologous) TCF genes (*TCF1* [also known as *HNF1A*] and *TCF2* [also known as *HNF1B*]) that are causal for MODY [4]. Whether or not the reduction in insulin secretion reflects a defect in the enteroinsular axis (early speculation—as yet unsupported by experimental data—pointed to impaired release of glucagon-like peptide-1 (GLP1) from enteroendocrine cells [12]), reduced beta cell mass (for example, through impaired pancreatic development), or intrinsic beta cell dysfunction, remains, as yet, unanswered. One intriguing observation reported across several data sets [12, 15, 18] is that in cases (but not controls), the type 2 diabetes-

susceptibility variants in *TCF7L2* are associated with reduced BMI. Whether this reflects an ascertainment effect (i.e. the reduction in beta-cell function is so profound that carriers tend to develop type 2 diabetes at a lower BMI than otherwise) or genuine biology (e.g. the chronic effects of beta cell insufficiency, given insulin's anabolic role) is another interesting question. Finally, the critical role of *TCF7L2* in glucose homeostasis points out the need for improved understanding of the role played by Wnt signalling in metabolic processes, and exploration of the possibilities for novel therapeutic modalities. *TCF7L2* variants provide an experiment of nature that will allow these and other vital questions about diabetes pathogenesis to be addressed.

In addition to the two studies referred to earlier [10, 11], this issue of *Diabetologia* contains two letters which address the potential contribution of *TCF7L2* to the development of other forms of diabetes [24, 25]. Cauchi and colleagues [24] asked whether rare mutations within *TCF7L2* might be contributing to the causation of monogenic forms of diabetes such as MODY and neonatal diabetes. Growing evidence of overlap in the genes involved in monogenic and multifactorial forms of diabetes [6] made this an excellent proposition. However, despite careful resequencing of all potential exons and intron–exon junctions of *TCF7L2* in 28 subjects with neonatal diabetes, and 17 with MODY (in whom other known causal genes had been excluded) no such mutations were uncovered. A second study, limited to neonatal diabetes alone, reached similar conclusions [14].

Field and colleagues [25] used the substantial resources available to the Cambridge group to determine whether or not the *TCF7L2* variants implicated in type 2 diabetes had any discernible impact on risk of developing type 1 diabetes. Their study, involving almost 14,000 subjects, conclusively demonstrates that they do not. This finding has implications for theories (such as the accelerator hypothesis) that propose a common aetiological basis for both type 1 and type 2 diabetes [26].

Lesson 3: The odds for successful translation of genetic information into clinical management have shortened considerably

Though it is easy to regard *TCF7L2* as having a 'large' effect because of its detectability in quite modest sample sizes, this does not translate into immediate clinical utility. A genetic test based around *TCF7L2* alone would lack sufficient sensitivity and specificity to be useful for risk prediction, in unselected populations at least [27]. But how might *TCF7L2* variants fare in combination with other genetic and clinical markers of prediction? These are early

days, but there are cautious grounds for optimism. Recent theoretical studies have emphasised that as few as 20 susceptibility variants on the scale of those in *TCF7L2*, *PPARG* and *KCNJ11* may suffice to explain as much as 50% of the burden of disease [28]. Empirical data gathered on more than 6,000 UK samples typed for the susceptibility variants in these three genes have recently shown that it is possible, using genetic information alone, to identify reasonably sized subgroups of individuals who differ more than threefold in diabetes risk [29].

The trade-off between sensitivity and specificity for such predictive tests can best be summarised using a so-called receiver-operating characteristic (ROC) curve. The area under such a curve is widely used as an index of the discriminative accuracy of a diagnostic test [30]. Practically speaking, this index denotes the frequency with which the test of interest correctly assigns diagnostic status when presented with two individuals, one of whom has, and other does not have, the disease of interest. The measure therefore ranges from 50% (no discrimination whatsoever; a random call) to 100% (perfect discrimination). In UK samples, the variants in *PPARG*, *KCNJ11* and *TCF7L2* are jointly capable of generating a discriminative value of approximately 58% [29]. Though short of the levels usually considered consistent with clinical utility (at least 75%), this value is the product of the fruits harvested from the very small proportion of the genome (a few percent) so far subjected to detailed examination in adequately powered data sets.

Over the next year, large-scale genome-wide association scan data being generated by several groups worldwide will extend that coverage (for common SNPs at least) to close to 70%. We will know, fairly soon, how many other '*TCF7L2*s' there are likely to be. These first, close-to-comprehensive views of the association landscape will help to settle the extent to which such common susceptibility variant information can be applied to clinically useful risk prediction. In parallel, studies that address the consequences of common variation for other aspects of clinical management, such as the prediction of complication risk and the response to prophylactic and/or therapeutic interventions [15], will establish the potential for other routes to clinical translation.

In recent years, research into the genetic basis of complex traits had been growing increasingly introspective, concerned with the limitations of then-current experimental approaches and the bewildering morass of conflicting and inconsistent results. The rapid elevation of *TCF7L2* from mere candidate to presidential status (a rise enabled by the availability of large sample sets from diverse populations and rapid, accurate genotyping), has meant that years of debate about the validity of *TCF7L2* as a diabetes-susceptibility gene have been avoided. As a result, the full

panoply of biochemical, genomic, epidemiological and other tools is now being deployed to define the pathogenic mechanisms through which it operates and to explore the translational potential. The hope is that the current round of genome-wide association studies will deliver other new diabetes-susceptibility genes for researchers to play with in the years to come.

Acknowledgement E. Zeggini is a Wellcome Trust Research Career Development Fellow.

References

- Bell JI (2003) The double helix in clinical practice. *Nature* 421:414–416
- Holtzman NA, Marteau TM (2000) Will genetics revolutionize medicine? *N Engl J Med* 343:141–144
- Phillips DI, Hales CN, Barker DJ (1993) Can twin studies assess the genetic component in type 2 (non-insulin-dependent) diabetes mellitus? *Diabetologia* 36:471–472
- Owen K, Hattersley AT (2001) Maturity-onset diabetes of the young: from clinical description to molecular genetic characterization. *Best Pract Res Clin Endocrinol Metab* 15:309–323
- Pearson ER, Flechtner I, Njolstad PR et al (2006) Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. *N Engl J Med* 355:467–477
- McCarthy MI (2004) Progress in defining the molecular basis of type 2 diabetes mellitus through susceptibility-gene identification. *Hum Mol Genet* 13:R33–R41
- Hattersley AT, McCarthy MI (2005) A question of standards: what makes a good genetic association study? *Lancet* 366:1315–1323
- Altshuler D, Hirschhorn JN, Klannemark M et al (2000) The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* 26:76–80
- Gloyn AL, Weedon MN, Owen KR et al (2003) Large scale association studies of variants in genes encoding the pancreatic beta-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
- van Vliet-Ostapchouk JV, Shiri-Sverdlov R, Zernakova A et al (2007) Association of variants of transcription factor 7-like 2 (*TCF7L2*) with susceptibility to type 2 diabetes in the Dutch Breda cohort. *Diabetologia*. DOI 10.1007/s00125-006-0477-z
- Chandak GR, Janipalli CS, Bhaskar S et al (2007) Common variants in the *TCF7L2* gene are strongly associated with type 2 diabetes mellitus in the Indian population. *Diabetologia*. DOI 10.1007/s00125-006-0502-2
- Grant SF, Thorleifsson G, Reynisdottir I et al (2006). Variant of transcription factor 7-like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. *Nat Genet* 38:320–323
- Reynisdottir I, Thorleifsson G, Benediktsson R et al (2003) Localisation of a susceptibility gene for type 2 diabetes to chromosome 5q34–q35.2. *Am J Hum Genet* 73:323–335
- Groves CJ, Zeggini E, Minton J et al (2006) Association analysis of 6736 UK subjects provides replication and confirms *TCF7L2* as a type 2 diabetes susceptibility gene with a substantial effect on individual risk. *Diabetes* 55:2640–2644
- Florez JC, Jablonski KA, Bayley N et al (2006) *TCF7L2* polymorphisms and progression to diabetes in the Diabetes Prevention Program. *N Engl J Med* 355:241–250
- Damcott CM, Pollin TI, Reinhard LJ et al (2006) Polymorphisms in the transcription factor 7-like 2 (*TCF7L2*) gene are associated with type 2 diabetes in the Amish: replication and evidence for a role in both insulin secretion and insulin resistance. *Diabetes* 55:2654–2659
- Scott LJ, Bonnycastle C, Willer CJ et al (2006) Association of transcription factor 7-like 2 (*TCF7L2*) variants with type 2 diabetes in a Finnish sample. *Diabetes* 55:2649–2653
- Cauchi S, Meyre D, Dina C et al (2006) Transcription factor *TCF7L2* genetic study in the French population: expression in human β -cells and adipose tissue and strong association with type 2 diabetes. *Diabetes* 55:2903–2908
- Zhang C, Qi L, Hunter DJ et al (2006) Variant of transcription factor 7-like 2 (*TCF7L2*) gene and the risk of type 2 diabetes in large cohorts of U.S. women and men. *Diabetes* 55:2645–2668
- Saxena R, Gianniny L, Burt NP et al (2006) Common single nucleotide polymorphisms in *TCF7L2* are reproducibly associated with type 2 diabetes and reduce the insulin response to glucose in nondiabetic individuals. *Diabetes* 55:2890–2895
- Wiltshire S, Hattersley AT, Hitman GA et al (2001) A genome-wide scan for loci predisposing to type 2 diabetes in a U.K. population (the Diabetes UK Warren 2 Repository): analysis of 573 pedigrees provides independent replication of a susceptibility locus on chromosome 1q. *Am J Hum Genet* 69:553–569
- McCarthy MI, Groop P-H, Hansen T (2005) Making the right associations. *Diabetologia* 48:1241–1243
- Wang WYS, Barratt BJ, Clayton DG, Todd JA (2005) Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet* 6:109–118
- Cauchi S, Vaxillaire M, Choquet H et al (2007) No major contribution of *TCF7L2* sequence variants to maturity onset of diabetes of the young (MODY) or neonatal diabetes mellitus in French white subjects. *Diabetologia*. DOI 10.1007/s00125-006-0505-z
- Field SF, Howson JMM, Smyth DJ, Walker NM, Dunger DB, Todd JA (2007) Analysis of the type 2 diabetes gene, *TCF7L2*, in 13,795 type 1 diabetes cases and control subjects. *Diabetologia*. DOI 10.1007/s00125-006-0506-y
- Wilkin TJ (2001) The accelerator hypothesis: weight gain as the missing link between type I and type II diabetes. *Diabetologia* 44:914–922
- Janssens ACJW, Gwinn M, Valdez R, Venkat Narayan KM, Khoury MJ (2006) Predictive genetic testing for type 2 diabetes. *Br Med J* 333:509–510
- Yang Q, Khoury MJ, Friedman JM, Little J, Flanders WD (2005) How many genes underlie the occurrence of common complex diseases in the population? *Int J Epidemiol* 34:1129–1137
- Weedon MN, McCarthy MI, Hitman G et al (2006) Combining information from multiple common susceptibility polymorphisms increases the predictive power of genetic information: a study of replicated type 2 diabetes variants. *PLoS Med* 3:e374
- Janssens ACJW, Pardo MC, Steyerberg EW, van Duijn CM (2004) Revisiting the clinical validity of multiplex genetic testing in complex diseases. *Am J Hum Genet* 74:585–588