

Testing the Gene or Testing a Variant?

The Case of TCF7L2

Mark O. Goodarzi^{1,2,3} and Jerome I. Rotter^{2,3,4,5}

Given that susceptibility to type 2 diabetes appears in large measure due to genetic makeup, investigators have spent years of effort trying to identify genes that influence type 2 diabetes risk. An important goal of gene identification is to improve our understanding of the pathophysiology of diabetes, leading to new measures of diagnosis, prevention, and treatment. A diabetes gene is considered identified when variants in that gene (more specifically, variation in DNA sequence between individuals) are found to be associated with type 2 diabetes and/or its pathophysiologic abnormalities such as insulin resistance or secretion.

TCF7L2 (transcription factor 7-like 2) was identified as a gene for type 2 diabetes in 2006 (1). Notably, its effect on diabetes (relative risk 1.5–1.6) is substantially larger than previously established diabetes genes (e.g., peroxisome proliferator-activated receptor γ [PPARG] and β -cell inwardly rectifying K^+ channel KIR6.2 [KCNJ11], relative risk ~ 1.2 each). Most studies of TCF7L2 have focused on the genetic variants that were implicated in the original report (1), largely ignoring the remainder of the gene. In this issue of *Diabetes*, investigators took the alternative approach of examining variants across the entire gene, which allowed them to discover a completely novel variant in TCF7L2 that affects diabetes risk (2). We believe this approach to genetic association studies is of great merit, as described below.

In the report that first identified TCF7L2 as a diabetes gene, a microsatellite (DG10S478) was highly associated with type 2 diabetes (in three Caucasian cohorts), as were five single nucleotide polymorphisms (SNPs) in linkage disequilibrium (LD). Of those five SNPs, two (rs12255372 and rs7903146) were identified as most strongly associated with type 2 diabetes; subsequent reports determined SNP rs7903146 had the greatest effect (3,4). Deep resequencing

of exons and the surrounding region has not identified any variants with stronger effect (4). Over 50 articles subsequently described the role of TCF7L2 in multiple cohorts, firmly establishing the gene's place among diabetes genes (5). Quantitative phenotypic associations (3,6,7) and its role in proglucagon gene expression (8) suggest that TCF7L2 modulates diabetes risk via impaired insulin secretion. Further support that TCF7L2 influences insulin secretion comes from an article recently published in *Diabetes*, wherein genetic variation in TCF7L2 was shown to influence efficacy of sulfonylureas (agents that promote insulin secretion) but not efficacy of metformin (insulin sensitizer) (9).

The majority of the subsequent studies followed the lead of the original study, genotyping either the one or two most associated SNPs, or all five, with or without the microsatellite DG10S478. These variants are in a haplotype block that encompasses DG10S478 and includes part of intron 3, all of exon 4, and part of intron 4 (1). In the few instances where more extensive SNP genotyping was performed, it was usually centered around DG10S478 (3,10,11). As a result, SNP rs7903146 (specifically the T-allele) has been firmly established as increasing risk of type 2 diabetes in multiple European populations and in Caucasians, in general, and including diverse groups such as Mexicans, Amish, Indian Asians, and Moroccans (5). Some studies of individuals of African descent were not able to document association of TCF7L2 with diabetes (12,13), while others, including a study in this issue of *Diabetes*, were positive for rs7903146 (4,14). Unlike in Caucasians, the two most associated SNPs exhibit weak LD in Africans, allowing determination of the greater role of rs7903146 than rs12255372 in diabetes susceptibility. The functional role that this intronic SNP may play is still unknown. Furthermore, it does not explain the linkage signal that originally led investigators to chromosome 10q (1), raising the possibility that other variants in the region (in TCF7L2 or other genes) may predispose to type 2 diabetes. Other plausible candidate genes do exist on chromosome 10q (15,16).

The near-exclusive focus on rs7903146 and SNPs in LD with it has probably delayed the discovery of other variants in TCF7L2 that may affect risk of type 2 diabetes. In Asian populations, the frequencies of SNPs rs7903146 and rs12255372 are quite low. Nevertheless, association with these SNPs was identified in two large Japanese cohorts (>1,000 cases each), where the minor allele frequency (MAF) ranged from 3 to 5% (17,18). The large size of these cohorts provided adequate power for discovery despite the rarity of the variants examined. This was fortuitous; had the allele frequencies or sample sizes been slightly lower, false-negative studies would have been the likely result, even though the gene and those variants are producing risk in these populations.

From the ¹Division of Endocrinology, Diabetes and Metabolism, Cedars-Sinai Medical Center, Los Angeles, California; the ²Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California; the ³Department of Medicine, University of California, Los Angeles, Los Angeles, California; the ⁴Department of Pediatrics, University of California, Los Angeles, Los Angeles, California; and the ⁵Department of Human Genetics, University of California, Los Angeles, Los Angeles, California.

Address correspondence and reprint requests to Mark O. Goodarzi, MD, PhD, Cedars-Sinai Medical Center, Division of Endocrinology, Diabetes and Metabolism, 8700 Beverly Blvd., Becker B-131, Los Angeles, CA 90048. E-mail: mark.goodarzi@cshs.org.

Received for publication 5 July 2007 and accepted in revised form 6 July 2007.

LD, linkage disequilibrium; MAF, minor allele frequency; SNP, single nucleotide polymorphism.

DOI: 10.2337/db07-0923

© 2007 by the American Diabetes Association.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

In the study of Chinese Han (760 case and 760 control subjects) reported in this issue (2), if the investigators had followed the usual strategy of looking only at the original associated SNPs, they would have found no associations. In Chinese, the rarity of rs7903146 (MAF 2–2.5%) would require 1,700 cases to detect an association with an effect size similar to that in Caucasians (2). Fortunately, these investigators tested the entire TCF7L2 gene rather than only particular variants of prior interest. In addition to genotyping SNPs previously associated with type 2 diabetes, they used information from HapMap (19) to select a set of 13 SNPs to capture the majority of SNPs in the gene with MAF >20%. The “classic” SNPs showed no association; however, SNP rs290487 at the 3′ end of the gene, as well as haplotypes including this SNP, were associated with type 2 diabetes. The comprehensive approach of this study led to this identification of a TCF7L2 SNP not in LD with rs7903146 that may affect type 2 diabetes risk. This demonstrates the utility of the global gene approach.

Given the narrow focus of prior studies of TCF7L2, no other study to date has examined this recently identified 3′ variant. Those groups with large diabetic cohorts should go back and examine this variant and its neighbors for a role in type 2 diabetes, which is feasible given that rs290487 occurs with appreciable frequency in all HapMap populations except Yorubans. HapMap also reveals that this SNP is not well captured by other SNPs (at $r^2 > 0.8$), indicating that its effect is likely to be missed if the SNP itself is not genotyped. The only other study to take a global approach to TCF7L2 (challenging with this large [~216 kb] gene) was carried out in Mexican Americans (20). This study did not genotype rs290487 but did genotype rs290483, also in the 3′ end, with negative results. Notably, one of the recent genome-wide association studies for type 2 diabetes found that rs290483 was associated with type 2 diabetes (a finding eclipsed by the stronger association of rs7903146 LD group SNPs) (21). This latter observation and the data of Chang et al. (2) suggest the potential functional importance of variation in the 3′ end of the gene, which may influence alternative splicing of TCF7L2, in which there exist many alternative splice sites (22). Clearly this end of the gene deserves the kind of attention that the upstream region of rs7903149 has received.

When follow-up gene marker studies focus only on the original associated variants, they amount only to replication studies that will not result in novel discoveries unless they examine a related but different phenotype, e.g., response to treatment (9), or extend the result to additional populations (14). To test the entire gene, comprehensive genotyping such as that carried out by Chang et al. (2) is necessary. This is particularly important if the replication cohort is not of the same ethnic group as the original cohort. Comprehensive global genotyping leads to a complete understanding of the SNP frequencies and haplotype structure, which may account for differences in association results. Given the availability of the HapMap database, and the falling price of genotyping, investigators now more readily have the option to test the entire gene (or region) rather than only a particular variant.

REFERENCES

- Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, Helgason A, Stefansson H, Emilsson V, Helgadóttir A, Styrkarsdóttir U, Magnusson KP, Walters GB, Palsdóttir E, Jonsdóttir T, Gudmundsdóttir T, Gylfason A, Saemundsdóttir J, Wilensky RL, Reilly MP, Rader DJ, Bagger Y, Christiansen C, Gudnason V, Sigurdsson G, Thorsteinsdóttir U, Gulcher JR, Kong A, Stefansson K: Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* 38:320–323, 2006
- Chang YC, Chang TJ, Jiang YD, Kuo SS, Lee KC, Chiu KC, Chuang LM: Association study of the genetic polymorphisms of the transcription factor 7-like 2 (TCF7L2) gene and type 2 diabetes in the Chinese population. *Diabetes* 56:2631–2637, 2007
- Saxena R, Gianniny L, Burt NP, Lyssenko V, Giuducchi C, Sjogren M, Florez JC, Almgren P, Isomaa B, Orho-Melander M, Lindblad U, Daly MJ, Tuomi T, Hirschhorn JN, Ardlie KG, Groop LC, Altshuler D: 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, 2006
- Helgason A, Palsson S, Thorleifsson G, Grant SF, Emilsson V, Gunnarsdóttir S, Adeyemo A, Chen Y, Chen G, Reynisdottir I, Benediktsson R, Hinney A, Hansen T, Andersen G, Borch-Johnsen K, Jorgensen T, Schafer H, Faruque M, Doumatey A, Zhou J, Wilensky RL, Reilly MP, Rader DJ, Bagger Y, Christiansen C, Sigurdsson G, Hebebrand J, Pedersen O, Thorsteinsdóttir U, Gulcher JR, Kong A, Rotimi C, Stefansson K: Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution. *Nat Genet* 39:218–225, 2007
- Florez JC: The new type 2 diabetes gene TCF7L2. *Curr Opin Clin Nutr Metab Care* 10:391–396, 2007
- Damcott CM, Pollin TI, Reinhart LJ, Ott SH, Shen H, Silver KD, Mitchell BD, Shuldiner AR: 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, 2006
- Florez JC, Jablonski KA, Bayley N, Pollin TI, de Bakker PI, Shuldiner AR, Knowler WC, Nathan DM, Altshuler D: TCF7L2 polymorphisms and progression to diabetes in the Diabetes Prevention Program. *N Engl J Med* 355:241–250, 2006
- Yi F, Brubaker PL, Jin T: TCF-4 mediates cell type-specific regulation of proglucagon gene expression by beta-catenin and glycogen synthase kinase-3beta. *J Biol Chem* 280:1457–1464, 2005
- Pearson ER, Donnelly LA, Kimber C, Whitley A, Doney AS, McCarthy MI, Hattersley AT, Morris AD, Palmer CN: Variation in TCF7L2 influences therapeutic response to sulfonylureas: a GoDARTS study. *Diabetes* 56: 2178–2182, 2007
- Watanabe RM, Allayee H, Xiang AH, Trigo E, Hartiala J, Lawrence JM, Buchanan TA: Transcription factor 7-like 2 (TCF7L2) is associated with gestational diabetes mellitus and interacts with adiposity to alter insulin secretion in Mexican Americans. *Diabetes* 56:1481–1485, 2007
- Scott LJ, Bonycastle LL, Willer CJ, Sprau AG, Jackson AU, Narisu N, Duren WL, Chines PS, Stringham HM, Erdos MR, Valle TT, Tuomilehto J, Bergman RN, Mohlke KL, Collins FS, Boehnke M: Association of transcription factor 7-like 2 (TCF7L2) variants with type 2 diabetes in a Finnish sample. *Diabetes* 55:2649–2653, 2006
- Humphries SE, Gable D, Cooper JA, Ireland H, Stephens JW, Hurel SJ, Li KW, Palmieri J, Miller MA, Cappuccio FP, Elkeles R, Godsland I, Miller GJ, Talmud PJ: Common variants in the TCF7L2 gene and predisposition to type 2 diabetes in UK European Whites, Indian Asians and Afro-Caribbean men and women. *J Mol Med* 84:1–10, 2006
- Elbein SC, Chu WS, Das SK, Yao-Borengasser A, Hasstedt SJ, Wang H, Rasouli N, Kern PA: Transcription factor 7-like 2 polymorphisms and type 2 diabetes, glucose homeostasis traits and gene expression in US participants of European and African descent. *Diabetologia* 50:1621–1630, 2007
- Sale MM, Smith SG, Mychaleckyj JC, Keene KL, Langefeld CD, Leak TS, Hicks PJ, Bowden DW, Rich SS, Freedman BI: Variants of the transcription factor 7-like 2 (TCF7L2) gene are associated with type 2 diabetes in an African-American population enriched for nephropathy. *Diabetes* 56:2638–2642, 2007
- Goodarzi MO, Lehman DM, Taylor KD, Guo X, Cui J, Quinones MJ, Clee SM, Yandell BS, Blangero J, Hsueh WA, Attie AD, Stern MP, Rotter JI: SORCSI: a novel human type 2 diabetes susceptibility gene suggested by the mouse. *Diabetes* 56:1922–1929, 2007
- Lehman DM, Fu DJ, Freeman AB, Hunt KJ, Leach RJ, Johnson-Pais T, Hamlington J, Dyer TD, Arya R, Abboud H, Goring HH, Duggirala R, Blangero J, Konrad RJ, Stern MP: A single nucleotide polymorphism in MGEA5 encoding O-GlcNAc-selective N-acetyl-β-D glucosaminidase is associated with type 2 diabetes in Mexican Americans. *Diabetes* 54:1214–1221, 2005
- Hayashi T, Iwamoto Y, Kaku K, Hirose H, Maeda S: Replication study for the association of TCF7L2 with susceptibility to type 2 diabetes in a Japanese population. *Diabetologia* 50:980–984, 2007
- Horikoshi M, Hara K, Ito C, Nagai R, Froguel P, Kadowaki T: A genetic

- variation of the transcription factor 7-like 2 gene is associated with risk of type 2 diabetes in the Japanese population. *Diabetologia* 50:747–751, 2007
19. The International HapMap Consortium: The International HapMap Project. *Nature* 426:789–796, 2003
 20. Lehman DM, Hunt KJ, Leach RJ, Hamlington J, Arya R, Abboud HE, Duggirala R, Blangero J, Goring HH, Stern MP: Haplotypes of transcription factor 7-like 2 (TCF7L2) gene and its upstream region are associated with type 2 diabetes and age of onset in Mexican Americans. *Diabetes* 56:389–393, 2007
 21. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P: A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445:881–885, 2007
 22. Duval A, Rolland S, Tubacher E, Bui H, Thomas G, Hamelin R: The human T-cell transcription factor-4 gene: structure, extensive characterization of alternative splicings, and mutational analysis in colorectal cancer cell lines. *Cancer Res* 60:3872–3879, 2000