

Genome-Wide Association Studies (GWAS) Deliver Big Breakthroughs in Identifying Type 2 Diabetes Susceptibility Genes in the General Population

Wei Cui, MD, PhD, Cuilin Zhang, MD, PhD, Yiqing Song, MD, ScD

[*NA J Med Sci. 2009;2(1):6-8.*]

A striking epidemic of type 2 diabetes mellitus (T2D) is a major public health problem worldwide. Current medical therapies can mitigate this disorder, but rarely provide a cure, leaving patients exposed to many debilitating and life-threatening complications. Although T2D is a multifactorial disease that involves interactions between many genetic and environmental factors, genetics is believed to play an important part in the development of T2D. Linkage analyses have identified several specific gene loci in a handful of families with diabetes that segregates as an autosomal dominant trait termed maturity-onset diabetes of the young (MODY-1,-2,-3,-4,-5, -6). But for the common (>95%) form of T2D in the general population, much remains unknown about its genetic determinants.

With the rapid advance of high-throughput genotyping technologies, genome-wide association studies (GWAS) have emerged as a comprehensive and powerful approach to identifying genetic variants related to complex diseases. In 2006, researchers from deCODE genetics company in Iceland performed a linkage region-wide scan (a region of chromosome 10) and identified an unexpected gene, TCF7L2 (*transcription factor 7-like 2*) gene associated with T2D risk.¹ The first GWAS paper published in the February 2007 issue of *Nature* confirmed TCF7L2 and reported another two genetic loci related to increased risk of type 2 diabetes in a French population.² Subsequently, many GWAS papers were published almost simultaneously and reported more new genes related to T2D.³ Up to now (January 2009), the preponderance of the evidence has led to the identification of 20 important genetic loci with reproducible associations with T2D (listed in **Table 1** below), although some gene variants

underlying T2D remain to be defined in future studies.

Herein we highlighted some landmark findings over the past two years:

First, GWAS uncovered many novel T2D genes, which may indicate previously unexpected disease pathways. Their novel biology requires further investigation. In particular, the TCF7L2 gene identified by deCODE genetics group is the most striking finding.¹ This gene encodes a transcription factor, which was not suspected of playing any role in the pathogenesis of T2D. Although the precise mechanisms are unknown, TCF7L2 is involved in the WNT signaling pathway and is thus linked to the regulation of pancreatic beta-cell function. The identification of TCF7L2 provided a textbook-example of how genetic mapping without any *priori* hypothesis is a powerful approach to identify underlying complex disease in the general population.

Second, the results from GWAS confirmed the known associations of KCNJ11 (*potassium inwardly-rectifying channel, subfamily 1, member 11 gene*) and PPARG (*peroxisome proliferator-activated receptor gamma gene*), which have been identified by candidate gene approaches, but did not support the hypothesis regarding an etiologic role of CAPN10 (*calpain-10 gene*) in the development of T2D. CAPN10 was the first T2D gene identified by both linkage analyses and positional cloning. However, its role in T2D remains controversial due to inconsistent results from previous association studies.

Third, GWAS findings also provided evidence to support the so-called “common disease-common variants” hypothesis. Common variants in these T2D genes (a minor allele frequency of > 1%) may represent genetic determinants of common form of T2D in the general population. Indeed, the vast majority of common genetic variations (85-95%) were shared between populations living on different continents. The effect sizes of common variants may vary among diverse ethnic groups depending on its linkage disequilibrium with true disease-causing variants or mutations.

Fourth, current evidence indicates that genetic variants for T2D in the exception of TCF7L2 often exhibit modest effects (odds ratio, OR, <1.25) in various populations. To date, variants in TCF7L2 have the largest risk effect (OR, 1.35-1.45). It is possible that false negative results due to small

Wei Cui, MD, PhD

Department of Endocrinology, First Affiliated Hospital, School of Medicine, Xi'an JiaoTong University, Xi'an, China

Cuilin Zhang, MD, PhD

Epidemiology Branch, Division of Epidemiology, Statistics, and Prevention Research, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, USA

Yiqing Song, MD, ScD

Division of Preventive Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

sample sizes partly explain the inconsistency in the results of previous GWAS.

Fifth, many of the newly identified risk variants appear to influence insulin secretion rather than insulin resistance.⁴ Of 20 loci, at least nine gene loci have been associated with impaired beta-cell development and function in the general population. In contrast, there is no strong evidence that any of these loci primarily alter insulin resistance. These findings have implicated the key role of pancreatic beta cell function in the pathogenesis of T2D.

Sixth, GWAS has also identified two novel genes that predispose individuals to obesity and T2D. Common variants in the *FTO* (*the fat mass and obesity associated gene*)⁵ and *MC4R* (*melanocortin 4 receptor*) genes⁶ have been reproducibly associated with fat mass and obesity risk. Approximately 58% of T2D globally can be attributable to overweight and obesity and 90% of T2D cases in Western countries are attributed to weight gain. The common form of obesity in adults is most likely polygenic; however, specific genetic factors, especially those that predispose obese individuals to T2D, remain largely undefined. Thus, such findings are very important and also highlight a critical role of the central nervous system in predisposition to obesity and T2D.

Finally, the most robustly associated genetic variants were found in non-coding regions thought to have no function. The fact that these variants are not necessarily in the coding regions for genes may indicate some unknown mechanisms underlying genome biology and gene regulation. Future experimental studies are required to investigate the mechanisms by which genetic variants affect mRNA expression, splicing, and degradation and the consequences on protein translation, stability, activity, and tissue-specific expression patterns.

Although the field of T2D genetics is moving rapidly, these are still many important questions that require answers before we can add such genetic information into routine clinical examination. First, it remains to be determined in large populations whether the identification of T2D common genetic variants could improve our ability to identify high-risk patients in both primary and secondary prevention. The assessment of T2D genetic alleles so far identified by GWASs together appeared to have minimal effect on the improvement of T2D prediction in three prospective cohorts as reported by two recently published papers in *New England Journal of Medicine*.^{7,8} Given limited data, the debate remains over the predictive value of genetic information by

genotyping all T2D susceptibility variants in predicting risk of T2D beyond traditional diabetes risk profiles. Second, such compelling evidence from GWAS contributes to our understanding of the pathophysiology of T2D. Since T2D is caused by both environmental and genetic factors, a careful investigation of gene-environment interactions in future well-designed and large-scale studies is required for a better understanding of the etiology of diabetes. Also, the interrelationships between T2D susceptibility loci and intermediate phenotypes such as insulin resistance and insulin secretion are less well studied. Additionally, there is a notable lack of data addressing whether these T2D susceptibility genes could contribute to other diabetes-related abnormalities. For example, individuals with some genetic variants may be at increased risk of developing essential hypertension, nonalcoholic fatty liver diseases, polycystic ovary syndrome, certain forms of cancer, and sleep apnea, which share some pathophysiological traits that also underlie T2D (e.g., insulin resistance and/or compensatory hyperinsulinemia). Finally, the new discoveries of genetic factors for T2D may shed light on pathways involved in T2D and identify new targets for therapeutic interventions.

In summary, GWAS data identified several novel candidate genes for T2D without any biologically-focused hypothesis. Understanding the mechanisms by which genetic risks influence diabetes pathogenesis will lead to new strategies for prevention and treatment.

References

1. 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(3):320-323.
2. Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature.* 2007;445(7130):881-885.
3. Groop L, Lyssenko V. Genes and type 2 diabetes mellitus. *Curr Diab Rep.* 2008;8(3):192-197.
4. 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(1):82-88.
5. 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(5826):889-894.
6. Loos RJ, Lindgren CM, Li S, et al. Common variants near *MC4R* are associated with fat mass, weight and risk of obesity. *Nat Genet.* 2008;40(6):768-775.
7. 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(21):2220-2232.
8. 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(21):2208-2219.

Table 1. Candidate gene approach and GWAS approach.

Candidate gene approach	KCNJ11;	Beta-cell dysfunction
	PPARG;	Insulin sensitivity
	TCF2; WFS1	Unknown function
GWAS approach	HHEX/IDE; SLC30A8; CDKAL1; TCF7L2; CDKN2A/2B; IGF2BP2; KCNQ1;MTNR1B	Beta-cell dysfunction
	FTO; MC4R;	Obesity

Abbreviations in the table:

KCNJ11, potassium inwardly-rectifying channel, subfamily 1, member 11; *PPARG*, peroxisome proliferator-activated receptor gamma gene; *TCF2*, transcription factor 2; *WFS1*, Wolfram syndrome 1; *HHEX/IDE*, hematopoietically expressed homeobox/insulin-degrading enzyme; *SLC30A8*, solute carrier family 30 (zinc transporter), member 8; *CDKAL1* (cyclin-dependent kinase 5 regulatory subunit associated protein 1-like 1), *TF7L2*, transcription factor 7-like 2 (T-cell specific, HMG-box); *CDKN2A/CDKN2B* (cyclin-dependent kinase inhibitor 2A /2B); *IGF2BP2*, insulin-like growth factor 2 mRNA-binding protein 2; *KCNQ1*, potassium voltage-gated channel, KQT-like subfamily, member 1; *MTNR1B*, melatonin receptor 1B; *FTO*, fat mass and obesity-associated; *MC4R*, melanocortin 4 receptor; *JAZF1*, juxtaposed with another zinc finger gene 1; *CDC123/CAMK1D*, cell division cycle 123 homologue/calcium/calmodulin-dependent protein kinase 1D; *TSPAN8/LGR5*, tetraspanin 8/leucine-rich repeat containing G protein coupled; *THADA*, thyroid adenoma associated; *ADAMTS9*, A disintegrin-like and metalloproteinase with thrombospondin type 1 motif 9; *NOTCH2*, Notch homologue 2

	JAZF1; CDC123/CAMK1D; TSPAN8/LGR5; THADA; ADAMTS9; NOTCH2.	Unknown function
--	--	------------------