Genetics Factors Contributing to Type 2 Diabetes across Ethnicities

Steven C. Elbein, M.D.

Abstract

Type 2 diabetes mellitus (T2DM) is among the many common diseases with a strong genetic component, but until recently, the variants causing this disease remained largely undiscovered. With the ability to interrogate most of the variation in the genome, the number of genetic variants has grown from 2 to 19 genes, many with multiple variants. An additional three genes are associated primarily with fasting glucose rather than T2DM. Despite the plethora of new markers, the individual effect is uniformly small, and the cumulative effect explains little of the genetic risk for T2DM. Furthermore, the success is largely restricted to European populations. Despite success in mapping genes in Asian populations, success in United States minorities, particularly African Americans, has been limited. The genetic findings highlight the role of the β cell in diabetes pathogenesis, but much remains to be discovered before genetic prediction and individualized medicine can become a reality for this disease.

J Diabetes Sci Technol 2009;3(4):685-689

Introduction

Type 2 diabetes mellitus (T2DM) has long been viewed as a disease with a substantial genetic contribution. This view is supported by studies that predate the understanding of T2DM as distinct from type 1 (autoimmune) diabetes mellitus and the current joint epidemics of diabetes and obesity. Among data supporting a strong heritable component are the marked differences in T2DM prevalence across populations. This prevalence ranges from the high-risk Pima and South Sea Island populations, where it now exceeds 50%, to relatively low-risk European populations, where it had been closer to 5% but is now rising, and intermediate-risk United States minority populations of African and Hispanic ancestry, where the prevalence now approaches 20%.[1,2] Additional support derives from the strong familial aggregation, the high risk of offspring of either one (38% lifetime prevalence) with diabetes, and the estimated sibling relative risk of 3.5.[3] Family history has been noted to double the risk of diabetes, equal to the risk of obesity, which is also heritable. Obesity and family history quadruple the diabetes risk.

These considerations, along with the rapidly developing field of complex disease genetics, have suggested that genetic approaches to unravel the pathophysiology of T2DM would complement physiologic attempts in humans and animals. The physiologic picture is itself complicated, with good evidence for the involvement of multiple organ systems, including adipose, muscle, pancreatic islets, immunity, central nervous system, and liver.

Thus a complicated genetic picture can be anticipated for a disease that represents a maladaptation of redundant pathways that are essential to survival.

Along with other similarly heritable human complex diseases, such as hypertension, obesity, type 1 diabetes, and many others, several genetic approaches have been tried to unravel T2DM with varying degrees of success. Initial approaches focused on the comparison of single nucleotide polymorphism (SNP) and structural polymorphism frequencies between diabetes patients and controls in a small number of candidate genes such as insulin, the insulin receptor, the *GLUT4* glucose transporter, the insulin signaling pathway (*IRS1*), and control of glucose-stimulated insulin secretion (glucokinase). These early studies, which focused almost entirely on coding variation, were largely viewed as nonreproducible and relatively uninformative for common forms of T2DM.[4] Nonetheless, candidate gene studies identified two genes now considered widely replicated: *PPARγ* and the β-cell potassium channel gene, *KCNJ11*.[4]

With the availability of hypothesis-free methods such as linkage studies in families and affected sibling pairs using simple tandem repeat markers, T2DM genetics entered a new era that promised the identification of new pathways not identified by studies of pathophysiology. A plethora of linkage studies across populations identified some reproducible signals, as noted later, but few signals that reached levels of genome-wide significance observed in monogenic diseases, and no linkage peaks that could be explained by single major genetic risk factors. In view of recent genome-wide association studies, many have written off the linkage approach as fundamentally under-powered and thus flawed. Based on the clear limitations of alternative approaches (genome-wide association and current limited deep resequencing studies of candidate genes), it could be argued that the interpretation of these results remains uncertain.

Since 2007, the T2DM genetic field has come full circle, returning to the association approach [genome-wide association (GWA)] but using a hypothesis-free method made possible by the technological explosions of highly multiplexed, chip-based genotyping, a relatively complete catalog of common human single nucleotide sequence variants, and an understanding of the linkage disequilibrium architecture of major human populations made possible by the HapMap project.[5] These studies have provided a logarithmic growth in the number of polymorphisms contributing to human disease. However, studies that in the design phase seemed well powered to

detect markers not detected in linkage studies have in fact been unable to convincingly demonstrate most markers without joining forces to accrue large populations never previously envisioned. Despite the numerical success, the explosion of new studies in 2007 and 2008 has raised fundamental questions about our understanding of the genetic architecture of complex diseases, including T2DM. Hence, in early 2009, our ability to use genetic markers to identify disease subsets, to predict future diabetes risk, or to identify those who might benefit from lifestyle or pharmaceutical intervention remains elusive.

Lessons from Monogenic Diabetes

One avenue of success in unraveling complex disease genes has been the identification of families with early onset disease.[6] Disease mapping through linkage in autosomal dominant, early onset, noninsulin-dependent diabetes [maturity onset diabetes of the young (MODY)] resulted in the identification of six genes with a surprising convergence on pathways involving pancreatic β-cell transcription factors (hepatocyte nuclear factors *HNF1α*, *4α*, *1β*, insulin promoter factor 1, and *NEUROD*) and pathways involved in glucose-stimulated insulin secretion (glucokinase). Of these six loci, functional variants in *HNF1α*, *GCK*, and *HNF4α* are sufficiently common to be worth clinically screening for in appropriate families. However, MODY represents less than 1% of all T2DM, and although several of these pathways reappear in common T2DM, screening of MODY genes in typical T2DM has failed to identify variants of a comparable effect.[7] As noted later, noncoding variants in or near *HNF1β* and *HNF4α* may contribute to typical T2DM, but for T2DM, as with other adult-onset diseases, early onset Mendelian forms of the disease are largely distinct and rare and provide limited insight into the common, complex forms of the disease.

Linkage Approaches to Type 2 Diabetes Mellitus

Genes identified for typical T2DM to date have too small an effect to generate a linkage signal, particularly with the sample sizes used in most published linkage studies. Initial family-based approaches were replaced with nonparametric, sib-pair approaches, but without added success. The recognition that, under the common disease/common variant hypothesis, single variants were unlikely to generate a linkage signal caused many investigators to discount all linkage findings as spurious. Nonetheless, several genes and regions are difficult to discount. The region of chromosome 1q (q21–q23)

was identified across ethnic groups and in multiple populations and encompasses over 400 expressed genes, including many strong candidates.[8] Studies to date have failed to identify any common variants that can explain the linkage findings, suggesting that rare variants or structural variation may account for the signals. Linkage in Mexican Americans on chromosome 2q identified the calpain 10 (*CAPN10*) gene,[9] which, although not apparent in large GWA studies, remains a likely contributor. Well-replicated linkage on chromosome 20q resulted in the identification of noncoding variants in the *HNF4α* gene, which have been replicated in some studies,[10] but likewise does not appear in GWA studies. Other regions remain with no explanation, and GWA and linkage approaches show minimal overlap. Given the clear limitations of GWA to unravel the genetic risk of T2DM, a reevaluation of these regions under models other than common disease/common variant is needed and might identify structural variants or uncommon SNPs with larger effect size that would be useful in personalized medicine approaches.

Successes and Limitations of Genome-Wide Association Studies

Prior to 2007, candidate gene and linkage approaches had identified three consistently replicated genes: the nonsynonymous (ns) SNP P12A in the thiazolidinedione target *PPARG*, the nsSNP E23K in the sulfonylurea target *KCNJ11*, and noncoding SNPs in intron 3 of a novel transcription factor identified first by linkage, *TCF7L2*. In 2007 and 2008, at least 14 GWA studies of T2DM have been published, ranging in size from under 1000 subjects to over 6000 and from under 100,000 SNPs to almost 400,000. These are nicely reviewed elsewhere.[11] Several salient conclusions can be drawn from these analyses. First, the only nsSNPs are in genes *SLC30A8* and *THADA*; all others are noncoding and many are far from any known gene. Second, the largest effect size (*TCF7L2*) has an odds ratio (OR) under 1.4; most are close to or under 1.1 and thus have little individual predictive value. Of the 20 genes that show convincing replication for an association with T2DM or fasting plasma glucose, 12 appear likely to alter insulin secretion or are known β-cell transcription factors. Several genes appear likely to alter cell cycle and may also alter β-cell mass, although many GWA genes are widely expressed, including adipose and muscle, and likely have other functions. Surprisingly, among primarily obesity genes, only *FTO*, which appears to alter energy metabolism[12] and obesity, is also associated with T2DM. Other obesity genes have not been associated with T2DM risk.

To date, genes for insulin action are elusive. Finally, GWA studies to date are restricted to European populations with the exception of the *KCNQ1* gene, which is a major risk factor in Asians and relatively rare in Europeans. Genome-wide association variants in *TCF7L2*, *CDKAL1*, *SLC30A8*, *IGF2BP2*, *HHEX*, and *CDKN2A/2B* have been replicated in Asian (Chinese, Japanese, and Korean) populations,[13–15] but with the exception of *TCF7L2* and *FTO*, support for other European GWA genes in United States minority populations is lacking.[16] In part, this lack of replication reflects inadequate examination. For example, GWA associations identify groups of SNPs in linkage disequilibrium (haplotype blocks) and not specific SNPs or genes. Because this block structure will differ across ethnic groups, an associated SNP in European populations may not identify an associated haplotype block in African-derived populations. To date, studies have tested European SNPs and appear to assume comparable linkage disequilibrium and allele frequencies. Where presumably functional variants have been identified as nsSNPs (*PPARG*, *KCNJ11*, and *SLC30A8*), the variants are nearly monomorphic in African American populations in contrast to European populations. Hence, insight into the genetic architecture of T2DM in high-risk United States minority populations is currently lacking.

Implications for Personalized Medicine

Despite the much reduced cost per genotype with current technological advances, the approximately 20 well-replicated susceptibility variants derived from GWA studies have come at high costs. Current estimates suggest that approximately 60,000 samples will be required to have adequate power to detect additional susceptibility genes. Such studies will drive those costs much higher at a time of shrinking research funds and a global recession. What has this technological tour de force taught us in the clinical arena? Four studies have sought to incorporate combinations of markers, most using a relatively simplistic approach of adding the number of risk alleles without consideration of differences in per-allele risk, gene–gene interactions, or gene–environment interactions.[17–20] Although each study used somewhat different combination, conclusions were similar. When those in the highest-risk category (20–30 risk alleles, 2–10% of the population) were compared to those with the least number of risk alleles (under 10, 2–10% of the population), ORs ranged from 4 to 10 and were uniformly highly statistically significant. The cross-sectional study of Cauchi and colleagues[21] suggested an impressive area under the receiver operating curve of 0.86 for up to 30

risk alleles. That optimistic view is not supported by three prospective studies that examined the contribution of genetic factors to traditional risk factors (age, gender, body mass index), particularly when those risk factors included measures such as fasting glucose. Instead, these studies suggested minimal additional contributions to the receiver operating curves from genetic tests of the order of 2%.[18–20] In the Framingham Offspring Study, the only potential benefit for genetic markers was in the very young who were otherwise at low risk for T2DM. Perhaps not surprisingly, these conclusions are not very different from similar studies of Lyssenko and associates[22] that predated the GWA explosion and used a small number of genes, including *TCF7L2*.

Conclusions

The marketing of genetic testing to the public and to medical professionals has been announced by vendors such as Decode Genetics and 23andMe. However, even using all current markers, which these companies do not propose, little additional information is provided beyond traditional risk factors such as age, obesity measures, waist circumference, family history of diabetes, and fasting glucose. Furthermore, the current risk factors account for only a small amount of the increased risk captured by a family history of diabetes. Certainly, shared environment might explain some of this discrepancy, but most likely, a majority of T2DM susceptibility genes remain to be identified. If all genes have an OR of 1.2 or less, which under the common disease/common variant hypothesis appears probable, then the total number of genes required to account for the estimated heritability is between 200 and 800, consistent with under 10% of the risk now explained. Alternatively, risk may be determined primarily by rare variants of larger effect (OR of 2–10 rather than 1.05–1.2). The latter would require far fewer genes and thus may be more amenable to use in personalizing medicine, whereas more complicated risk models combining very large numbers of variants will be required to determine the pathways that predominate in individuals with T2DM under the common variant model. More research and further technological advances will both be required, likely using high-throughput sequencing in large numbers of individuals from a variety of populations, before the genetic architecture of T2DM can be understood well enough for application in the clinical setting. Even then, the possibility remains that complex diseases are intractable to personalized medicine as currently envisioned.

References:

1. Elbein SC. Evaluation of polymorphisms known to contribute to risk for diabetes in African and African-American populations. Curr Opin Clin Nutr Metab Care. 2007;10(4):415-9.

2. Haffner SM. Epidemiology of type 2 diabetes: risk factors. Diabetes Care. 1998;21 Suppl 3:C3-6.

3. Rich SS. Mapping genes in diabetes. Genetic epidemiological perspective. Diabetes. 1990;39(1):1315-9.

4. Das SK, Elbein SC. The genetic basis of type 2 diabetes. Cellscience. 2006;2(4):100-31.

5. McVean G, Spencer CC, Chaix R. Perspectives on human genetic variation from the HapMap project. PLoS Genet. 2005;1(4):e54.

6. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC. Linkage of early-onset familial breast cancer to chromosome 17q21. Science. 1990;250(4988):1684-9.

7. Winckler W, Weedon MN, Graham RR, McCarroll SA, Purcell S, Almgren P, Tuomi T, Gaudet D, Boström KB, Walker M, Hitman G, Hattersley AT, McCarthy MI, Ardlie KG, Hirschhorn JN, Daly MJ, Frayling TM, Group L, Altshuler D. Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes. Diabetes. 2007;56(3):685-93.

8. Das SK, Elbein SC. The search for type 2 diabetes susceptibility loci: the chromosome 1q story. Curr Diab Rep. 2007;7(2):154-64.

9. Hanis CL, Boerwinkle E, Chakraborty R, Ellsworth DL, Concannon P, Stirling B, Morrison VA, Wapelhorst B, Spielman RS, Gogolin-Ewens KJ, Shephard JM, Williams SR, Risch N, Hinds D, Iwasaki N, Ogata M, Ommoi Y, Petzold C, Rietzsch H, Schröeder H-E, Schulze J, Cox NJ, Menzel S, Boriraj VV, Chen X, Lim LR, Linder T, Mereu LE, Wang YQ, Xiang K, Yamagata K, Yang Y, Bell GI. A genome-wide search for human non-insulin-dependent (type 2) diabetes genes reveals a major susceptibility locus on chromosome 2. Nat Genet. 1996;13(2):161-6.

10. Weedon MN, Owen KR, Shields B, Hitman G, Walker M, McCarthy MI, Love-Gregory LD, Permutt MA, Hattersley AT, Frayling TM. Common variants of the hepatocyte nuclear factor-4alpha P2 promoter are associated with type 2 diabetes in the U.K. population. Diabetes. 2004;53(11):3002-6.

11. Frayling TM. Genome-wide association studies provide new insights into type 2 diabetes aetiology. Nat Rev Genet. 2007;8(9):657-62.

12. Do R, Bailey SD, Desbiens K, Belisle A, Montpetit A, Bouchard C, Pérusse L, Vohl MC, Engert JC. Genetic variants of FTO influence adiposity, insulin sensitivity, leptin levels, and resting metabolic rate in the Quebec Family Study. Diabetes. 2008;57(4):1147-50.

13. Ng MC, Park KS, Oh B, Tam CH, Cho YM, Shin HD, Lam VK, Ma RC, So WY, Cho YS, Kim HL, Lee HK, Chan JC, Cho NH. Implication of genetic variants near TCF7L2, SLC30A8, HHEX, CDKAL1, CDKN2A/B, IGF2BP2, and FTO in type 2 diabetes and obesity in 6,719 Asians. Diabetes. 2008;57(8):2226-33.

14. Omori S, Tanaka Y, Takahashi A, Hirose H, Kashiwagi A, Kaku K, Kawamori R, Nakamura Y, Maeda S. Association of CDKAL1, IGF2BP2, CDKN2A/B, HHEX, SLC30A8, and KCNJ11 with susceptibility to type 2 diabetes in a Japanese population. Diabetes. 2008;57(3):791-5.

15. Wu Y, Li H, Loos RJ, Yu Z, Ye X, Chen L, Pan A, Hu FB, Lin X. Common variants in CDKAL1, CDKN2A/B, IGF2BP2, SLC30A8, and HHEX/IDE genes are associated with type 2 diabetes and impaired fasting glucose in a Chinese Han population. Diabetes. 2008;57(10):2834-42.

16. Lewis JP, Palmer ND, Hicks PJ, Sale MM, Langefeld CD, Freedman BI, Divers J, Bowden DW. Association analysis in African Americans of European-derived type 2 diabetes single nucleotide polymorphisms from whole-genome association studies. Diabetes. 2008;57(8):2220-5.

17. Cauchi S, Proença C, Choquet H, Gaget S, De Graeve F, Marre M, Balkau B, Tichet J, Meyre D, Vaxillaire M, Froguel P; D.E.S.I.R. Study Group. Analysis of novel risk loci for type 2 diabetes in a general French population: the D.E.S.I.R. study. J Mol Med. 2008;86(3):341-8.

18. Lango H; UK Type 2 Diabetes Genetics Consortium, Palmer CN, Morris AD, Zeggini E, Hattersley AT, McCarthy MI, Frayling TM, Weedon MN. Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk. Diabetes. 2008;57(11):3129-35.

19. Van Hoek M, Dehghan A, Witteman JC, van Duijn CM, Uitterlinden AG, Oostra BA, Hofman A, Sijbrands EJ, Janssens AC. Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study. Diabetes. 2008;57(11):3122-8.

20. Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, Manning AK, Florez JC, Wilson PW, D'Agostino RB Sr, Cupples LA. Genotype score in addition to common risk factors for prediction of type 2 diabetes. N Engl J Med. 2008;359(21):2208-19.

21. Cauchi S, Meyre D, Durand E, Proença C, Marre M, Hadjadj S, Choquet H, De Graeve F, Gaget S, Allegaert F, Delplanque J, Permutt MA, Wasson J, Blech I, Charpentier G, Balkau B, Vergnaud AC, Czernichow S, Patsch W, Chikri M, Glaser B, Sladek R, Froguel P. Post genome-wide association studies of novel genes associated with type 2 diabetes show gene-gene interaction and high predictive value. PLoS ONE. 2008;3(5):e2031.

22. Lyssenko V, Almgren P, Anevski D, Orho-Melander M, Sjögren M, Saloranta C, Tuomi T, Groop L; Botnia Study Group. Genetic prediction of future type 2 diabetes. PLOS Med. 2005;2(12):e345.