Sideways Glance: Genome wide association studies for type 2 diabetes mellitus

Yula Sambuy

Published online: 8 November 2007
© Springer-Verlag 2007

Abbreviations

BMI	Body mass index
CDK5	Cyclin-dependent kinase 5
CDKAL1	Cyclin-dependent kinase 5-regulatory subunit associated protein 1-like 1
CDKN2A/B	Cyclin-dependent kinase-inhibitor A and B
FTO	Fat mass and obesity associated
GWAS	Genome wide association studies
HHEX	Homeobox, hematopoietically expressed
IDE	Insulin-degrading enzyme
IGF2BP2	Insulin-like growth factor 2 mRNA binding protein 2
KCNJ11	Potassium rectifying channel, sub-family J member 11
OR	Odd risk
PPARG	Peroxisome proliferator-activated receptor-gamma
SNP	Single nucleotide polymorphism
TCF7L2	Transcription factor 2, hepatic
T2DM	Type 2 diabetes mellitus

This month's issue of Sideways Glance is devoted to the discussion of some recently published reports of genome wide association studies (GWAS) aimed at identifying the genetic factors involved in the risk of developing type 2 diabetes mellitus (T2DM), which is increasing dramatically throughout the world. But is the enthusiasm raised by these new studies justified by the results? The open questions are: to what extent can these studies contribute to clarifying

the still obscure areas of the pathophysiology of the disease, and how soon will we be able to identify potentially at risk patients and to suggest personalized diets or novel treatments to reduce their risk of developing T2DM.

Genetic predisposition to diabetes mellitus type 2: will large collaborative efforts be able to overcome the geneticist's nightmare?

T2DM affects more than 170 million people worldwide and its prevalence is increasing rapidly in the developed world and in developing countries [4]. The disease has a strong environmental component, as factors like diet and lifestyle can greatly influence the development of the disease. Obesity, defined as body mass index (BMI) greater than 30 kg/m², increases the risk for T2DM. In addition, there is a significant genetic component involved, with siblings of affected individuals having a three to fourfold higher risk of developing the disease. T2DM is therefore considered a typical polygenic, multifactorial disease [5].

Thirty years ago, James V. Neel labeled T2DM as “the geneticist's nightmare” predicting that the identification of genetic factors that predispose to the disease would be very challenging [10]. Until recently, his prediction has proved to be true. In the past 10 years, the search by different approaches for genes linked to an increased probability of developing T2DM has produced a list of at least 52 genes that have good or suggestive evidence of contributing to the disease [7]. However, until recently linkage fine-mapping and candidate-genes studies produced strong evidence for the existence of common variants in only three loci conferring increased T2DM risk. These are the Pro12Ala variant in peroxisome proliferator-activated receptor-

Y. Sambuy (✉)
INRAN, Rome, Italy
e-mail: sayula@tiscali.it

gamma (*PPARG*), the E23K variant in the potassium rectifying channel, sub-family J, member 11 (*KCNJ11*) and common variants in the transcription factor 2, hepatic (*TCF7L2*). Rare, but severe mutations in any of these genes cause monogenic forms of diabetes [11]. In addition, the protein coded by *KCNJ11* is a component of potassium channel in pancreatic beta-cells and is a target for the sulfonyl-urea class of anti-diabetic drugs. Similarly, the target of the class of insulin-sensitizing drug thiazolidinedione is the transcription factor encoded by *PPARG* that is involved in adipocyte differentiation [5].

Some new insight into the genetic components implicated in the development of T2DM has recently emerged in the reports from five GWAS, published between February and June 2007, which identified six additional and confirmed the three loci that had previously been reported, with 12 different single nucleotide polymorphisms (SNPs) associated in various extents with the disease [13, 14, 17, 21, 23]. GWAS differ from previous approaches, as these studies are unbiased with regard to presumed functions and locations of causal variants. The first study, conducted on a French case–control cohort of European descent reported that four relatively common variants with modest effect (heterozygous relative risk = 1.15–1.65) contributed a significant part towards T2DM risk. These variants were localized in the *TCF7L2* gene, in a gene coding for a Zn transporter in pancreatic beta-cells (*SLC30A8*), and two SNPs were located in a linkage disequilibrium block containing genes of potential biological significance for diabetes, namely the homeobox hematopoietically expressed (*HHEX*) and the insulin-degrading enzyme (*IDE*) [17].

These associations were subsequently confirmed in three reports simultaneously published in Science in April 2007 [13, 14, 23]. Although several GWAS had been performed on T2DM in recent years, these latter studies are particularly important because of: the unprecedented joint collaborative effort to combine findings and to perform replication and meta-analysis; the large number of cases examined (each had 1,900 or more cases and controls with a combined number of 14,586 patients and 17,968 controls); and the common European ancestry of all subjects (reviewed in [23]). Despite some differences in the selection of phenotypes, there was remarkable consistency in the genes identified in these studies as linked to T2DM. In addition to replicating positive associations for *TCF7L2*, *KCNJ11*, *PPARG*, *HHEX-IDE* and *SLC30A8*, new variants were found in an intron (non-coding, spacer, section of a gene) of cyclin-dependent kinase 5 (CDK5)-regulatory subunit associated protein 1-like 1 (*CDKAL1*), in an intron of insulin-like growth factor binding protein 2 (*IGF2BP2*), in non-coding regions near the genes for cyclin-dependent kinase-inhibitor A and B (*CDKN2A/B*) on chromosome 9, and in the fat mass and obesity associated (*FTO*) region.

The physiological significance of these genes in the etiology of T2DM is largely unknown, and, in some instances, there is the additional problem of dealing with variants in non-coding regions. For *KCNJ11* and *PPARG*, the physiological link is easier to determine since they play a central role in monogenic forms of diabetes and their products are direct target of antidiabetic drugs.

The only variant with an obvious functional significance is probably the non-synonymous allele (changing an arginine to a tryptophan) in *SLC30A8*, coding for a Zn transporter abundantly expressed in the insulin-producing beta cells of the pancreas. This transporter has been associated with Zn loading into insulin-secreting vesicles, where insulin is stored as a hexamer linked to two Zn ions before secretion [3]. Variations in this transporter activity may thus affect insulin stability, storage or secretion.

IGF2BP2 belongs to a family of three mRNA binding proteins with affinity for key elements in the untranslated region of insulin-like growth factor *IGF-2* transcripts. *IGF-2* is a key growth and insulin-signaling protein and is also expressed in pancreatic islets.

For *CDKN2A/B*, the most strongly associated SNPs reside downstream of the genes in a small region with no characterized genes. *CDKN2A/B* code for proteins involved in beta-cells replication, and there is evidence that they may play a role in the proliferation of islets [22].

CDKAL1 mRNA is highly expressed in human pancreatic islets and in skeletal muscles [23]. Although the function of the gene product has not been characterized yet, protein homology to another CDK5 regulatory protein indicates a role in CDK5 inhibition. CDK5 is important in the glucotoxic loss of beta-cell function. Also, the insulin response of homozygotes for the *CDKAL1* variant was approximately 20% lower than for heterozygotes or non-carriers, suggesting that this variant confers the risk of T2DM through reduced insulin secretion [18].

Variants in the *HHEX-IDE* region on chromosome 10 may increase susceptibility to T2DM through the key role of *HHEX* in pancreatic development, as indicated by the absence of ventral pancreas in *HHEX* knockout mice. In addition, the *IDE* locus, coding for an insulin degrading enzyme, may also be involved. Supporting evidence for the role of *CDKAL1* and *HHEX/IDE* in diabetes came from a study showing an association of the variant alleles with decreased pancreatic beta-cell function, including decreased glucose sensitivity that relates insulin secretion to plasma glucose concentration [9].

One of the strongest T2DM risk-association in all the GWAS studies was found for common variants in *TCF7L2*, a gene coding for a transcription factor that is part of the WNT signaling pathway involved in the regulation of myogenesis and angiogenesis, but also critical for the embryonic development of pancreatic islets [19]. Recently,

it has been shown that the variant allele results in over-expression of *TCF7L2* in pancreatic beta-cells, reducing insulin secretion in response to a variety of stimuli [6, 8]. The odd ratios (OR, is an estimate of the relative risk, with values >1.0 indicating a positive and <1.0 a negative association, conferred by each additional risk allele carried at each locus) calculated in the pooled studies for the T allele in the snp7903146 of *TCF7L2* was 1.37 (1.31–1.43) [13]. This variant resides in an intron of the gene. Other variants at this locus also confer increased risk for T2DM, although the specific genetic defect that results in impaired insulin secretion in carriers has not been identified yet. Alternatively, other genes in the region may contribute to T2DM susceptibility. Associations between the T variant of *TCF7L2* and T2DM have been consistently confirmed in geographically, ethnically, and environmentally diverse populations (references in [19], without evidence of heterogeneity across ethnic groups [2].

Other confirmed genetics variants associated with T2DM increased risk by only 10–20% when in the heterozygous state. Carrying two copies of the variant allele (homozygous) or carrying multiple variant alleles has been shown to significantly increase the risk of disease. *T2CF7L2* variant homozygotes have a twofold increased risk [19]. It has been calculated that an adverse collection of multiple SNPs can identify a subset of individuals, who are at approximately fourfold increased risk for T2DM [1]. Another study examined the risk associated with carrying an increasing number of the variant alleles for the previously identified genes *T2CF7L2*, *PPARG* and *KCNJ*. Applying a multiplicative model for the best fit of data, they calculated that each additional risk allele increased the odds of developing T2DM by 1.28-fold, and that homozygous carriers for risk alleles at all three loci had an OR of 2.84, almost a threefold increase in the risk for the disease [20]. This level of risk appears higher than reported in one of the recent collaborative GWAS, where the combined contribution of eight variant alleles to the risk for T2DM was reported to be only 2.3%, although it was not specified how this figure was obtained [13]. It would be interesting to extend the original study on cumulative risk [20] to the additional variant alleles later identified and also to determine possible epistatic (gene–gene) interactions. Furthermore, the identification of a larger number of novel genes involved in the pathophysiology of T2DM would increase our power of prediction [1].

The fat mass and obesity associated (*FTO*) gene was found to confer a moderate risk of developing T2DM in association with increased BMI. This association was in fact attenuated when adjusting for BMI and waist circumference [23] and it was not detected in one study, probably because the cases and controls were matched for BMI [13]. This points to the importance of the criteria for

subjects selection in GWAS. In the French study, for example, obese patients were excluded in order to diminish phenotypic heterogeneity, at the expense of the chance of detecting potential loci (such as *FTO*) conferring risk through effects on insulin response in the presence of obesity [17].

Methodological aspects

The success of GWAS in detecting new associations and potential risk factors for any particular disease or condition depends greatly on the experimental design, on careful selection of the populations, on large number of cases and on collaborative analytical approaches. Meta-analysis is a method that combines the results of a number of surveys and of replication studies on the most promising variants. It investigates the underlying processes and has become standard practice for publications of GWAS that search for common genetic variants regulating complex traits and disease risk.

The factors contributing to the success of these studies have been discussed with special reference to the recent GWAS directed at identifying the genetic factors of T2DM [1]. Among the factors to consider in large GWAS involving hundreds of thousands of statistical tests is that modest levels of bias can overwhelm a small number of true results. Therefore, it is of foremost importance to adopt adequate statistical strategies, to search for systematic bias from unrecognized population structure, the analytical approach and genotyping artefacts [13].

But some criticisms have also been raised against GWAS [15, 16]. The GWAS for T2DM have resulted in a small number of genetic associations (overall 5 GWAS on T2DM have resulted in 9–10 confirmed risk loci), often associated with a modest risk, within a context of “complex” multifactorial diseases. Moreover, since GWAS are very expensive and often considered useful to “generate hypothesis” on the pathophysiology of the disease, their cost effectiveness has been questioned. Another methodological criticism that has been raised is the reduced power of replications in which the sample size is significantly decreased compared to the first screen. Finally, failure to replicate across populations may not necessarily imply that the original finding was a false positive; rather, it could indicate a population-specific risk. Modes of selection of predisposing SNPs on the basis of the most extreme *P* values (as often done in GWAS) has also been criticized, as *P* values indicate the probability of finding such an event, not its biological importance. As an example of this, *PPARG* would probably have been discarded in the later GWAS if its association with T2DM had not been previously shown by candidate-gene studies [15, 16].

Other aspects that have been overlooked in large GWAS on T2DM relate to environmental effects such as diet, physical activity, and stresses, which may affect gene expression. For example, fish oil may stimulate PPARG in much the same fashion as the thiazolidinedione class of drugs; however, studies on the interaction of the PPARG variant with dietary components have not been performed. The spectacular rise in the incidence of diabetes among Pima Indians and other populations as they adopt Western diets and lifestyles dramatically demonstrates the key role of the environment [12]. Consequently, it could be expected that the effect of a common gene variant among populations that have very different diets and exercise habits might be totally different, thus explaining some instances of lack of replication. [4]. Another variable that influences the statistical and real association of an SNP with a disease or response to a diet is epigenetic interaction. Epigenesis is the study of heritable changes in gene function that occur without a change in the DNA sequence, such as DNA methylation and chromatin remodeling. Both mechanisms can affect gene expression by altering the accessibility of DNA to regulatory proteins or complexes such as transcription factors, and they can be influenced by certain nutrients and by overall caloric intake. Thus, it can be expected that long-term exposure to certain diets could produce permanent epigenetic changes in the genome [7].

Conclusions

Recent large collaborative studies to clarify the genetics of T2DM have identified variants in nine gene areas that are associated with a moderately increased risk of developing the disease. Further studies may identify more of these variants and ultimately improve the possibility of predicting disease risk in healthy subjects. Search for the patho-physiological role of these variants has not been easy, although evidence is emerging for their involvement in either pancreatic development or in the control of insulin secretion. The elucidation of novel pathways involved to the etiology of T2DM may contribute to improved prevention and treatment of the disease. The influence of environmental factors such as lifestyle and diet must not be overlooked, and future studies should be especially focused on the interactions between dietary factors and the genetic variants involved in T2DM risk. In the light of the recent investigative efforts, the genetics of T2DM is probably no longer “the geneticist’s nightmare”, but it certainly remains an intriguing puzzle that is yet to be solved.

References

1. Amos CI (2007) Successful design and conduct of genome-wide association studies. *Hum Mol Genet* 16:R220–R225
2. Cauchi S, El Achhab Y, Choquet H et al (2007) TCF7L2 is reproducibly associated with type 2 diabetes in various ethnic groups: a global meta-analysis. *J Mol Med* 85:777–782
3. Chimienti F, Devergnas S, Pattou F et al (2006) In vivo expression and functional characterization of the zinc transporter ZnT8 in glucose-induced insulin secretion. *J Cell Sci* 119:4199–4206
4. Das SK, Elbein SC (2006) The genetic basis of type 2 diabetes. *Cellscience* 2:100–131
5. Frayling TM (2007) Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nat Rev Genet* 8:657–662
6. Hattersley AT (2007) Prime suspect: the TCF7L2 gene and type 2 diabetes risk. *J Clin Invest* 117:2077–2079
7. Kaput J, Noble J, Hatipoglu B et al (2007) Application of nutrigenomic concepts to Type 2 diabetes mellitus. *Nutr Metab Cardiovasc Dis* 17:89–103
8. Lyssenko V, Lupi R, Marchetti P et al (2007) Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. *J Clin Invest* 117:2155–2163
9. McCarthy MI, Zeggini E (2006) Genetics of type 2 diabetes. *Curr Diab Rep* 6:147–154
10. Neel JV (1976) In: W Creutzfeldt J Kobberling JV Neel (eds) The genetics of diabetes mellitus. Springer, Berlin, pp 1–11
11. Owen KR, McCarthy MI (2007) Genetics of type 2 diabetes. *Curr Opin Genet Dev* 17:239–244
12. Pavkov ME, Hanson RL, Knowler WC et al (2007) Changing patterns of type 2 diabetes incidence among Pima Indians. *Diabetes Care* 30:1758–1763
13. Saxena R, Voight BF, Lyssenko V et al (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316:1331–1336
14. Scott LJ, Mohlke KL, Bonnycastle LL et al (2007a) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316:1341–1345
15. Scott MW, Canter JA, Crawford DC et al (2007b) Letter: problems with genome-wide association studies. *Science* 316:1841–1842
16. Shriner D, Vaughan LK, Padilla MA et al (2007) Letter: problems with genome-wide association studies. *Science* 316:1840–1841
17. Sladek R, Rocheleau G, Rung J et al (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445:881–885
18. Steinthorsdottir V, Thorleifsson G, Reynisdottir I et al (2007) A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet* 39:770–775
19. Weedon MN (2007) The importance of TCF7L2. *Diabet Med* 24:1062–1066
20. Weedon MN, McCarthy MI, Hitman G et al (2006) Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. *PLoS Med* 3:e374
21. Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661–678
22. Zeggini E (2007) A new era for type 2 diabetes genetics. *Diabet Med* 24:1181–1186
23. Zeggini E, Weedon MN, Lindgren CM et al. (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 316:1336–1341