

Prediction of type 2 diabetes: the dawn of polygenetic testing for complex disease

J. B. Meigs

Received: 8 January 2009 / Accepted: 26 January 2009 / Published online: 12 February 2009
© Springer-Verlag 2009

Keywords Clinical · Complex disease · Genetics · Genomics · Prediction · Public health · Type 2 diabetes

Abbreviation

SNP Single nucleotide polymorphism

Advances in the search for the genetic determinants of type 2 diabetes have led, in recent years, to the discovery of dozens of risk loci, with many more expected to be found [1–3]. Each individual locus has a relatively small effect, increasing the risk of diabetes by ~5–40% per risk allele. Although the risk alleles themselves are common, few individuals (~10%) are homozygous at a given locus, no locus accounts for more than a fraction of a per cent of the variance in the risk of diabetes, and only a few loci have a proven function in diabetes pathogenesis [4]. It seems apparent that revelation of novel pathways determining energy metabolism will come from this era of diabetes gene discovery. However, how to apply this new knowledge to improve clinical and public health is less clear. Approaches to genetic testing and counselling for diseases with multiple common loci that exert a modest effect are not well established. The influence of genetic risk knowledge on personal health behaviours for chronic disease prevention

or treatment is not known. Meanwhile, there is growing enthusiasm for the use of new genetic knowledge in disease screening and diagnostic testing [5]. In contrast to uncertainties about prevention and personalisation, the value of current genetic knowledge for prediction of diabetes risk is becoming fairly clear.

In this issue of *Diabetologia*, Lin et al. report on a study in which they tested the ability of multiple novel type 2 diabetes risk loci to identify prevalent diabetes in over 5,000 participants of the population-based CoLaus Study [6]. They calculated an additive, weighted genetic score based on 15 single nucleotide polymorphisms (SNPs), using the lower limit of the 95% CI of the reported effect size on diabetes risk as the weight for each SNP. They tested the main effect of the genotype score on diabetes risk, and then the marginal effect after accounting for common type 2 diabetes risk factors, including age, BMI, family history of diabetes, WHR and triacylglycerol/HDL-cholesterol ratio. After calculating the weighted score, they rescaled it for interpretation as odds of diabetes per risk allele. The 356 people with diabetes in the sample had a mean weighted genotype score of 15.2, whereas those without diabetes had a score of 14.3. This is a minute difference in allele count: on average, people with diabetes had just one more risk allele out of the 15 tested compared with those without diabetes. Nonetheless, the risk factor-adjusted relative odds of diabetes in the fifth of the group with the highest genotype score was 2.7-fold higher than that in the fifth of the group with the lowest genotype score. Thus, it is shown that common disease risk alleles can be summed, with people at the highest end of the genetic risk distribution having a markedly elevated individual risk, especially in contrast to the low relative risk associated with individual loci. However, at the group level, the genotype

J. B. Meigs
Department of Medicine, Massachusetts General Hospital,
Boston, MA, USA

J. B. Meigs (✉)
General Medicine Division, Massachusetts General Hospital,
50 Staniford Street, 9th Floor,
Boston, MA 02114, USA
e-mail: jmeigs@partners.org

score was of very limited value for discrimination of diabetes risk once common risk factors were accounted for. This scenario addresses the situation: At an adult check-up, after the usual things have been measured, would diabetes genetic testing add additional risk prognostic information? The area under the receiver operating characteristic curve (or *c* statistic) was 0.86 for a prediction model that considered common risk factors alone, and 0.87 for one that considered the weighted genotype score in addition to risk factors. Thus, although these *c* statistics were significantly different ($p=0.002$), the results indicate that one would be correct 86% of the time when choosing the higher risk individual from a pair at risk of diabetes on the basis of phenotypic risk factors; confidence in correct discrimination would increase by only 1% with the addition of genetic risk information. Lin et al. conclude that, in their population, the genetic score provides (some) additional information over clinical predictors of prevalent diabetes, but that the clinical benefit of this genetic information is probably limited.

The results of Lin et al. are remarkably consistent with those of the few extant studies of genetic prediction of type 2 diabetes, conducted either in case–control samples or by longitudinal follow-up in population-based samples [7–12]. The data are clear that combinations of risk alleles can be used to identify a gradient of genetic risk of diabetes. The approach provides a neat and simple initial solution to the problem of accounting for multiple weak risk loci in genetic screening for complex disorders. Although per-allele effects are small for each of the known risk loci, people with more risk alleles are at greater risk than those with fewer, with an increased relative risk of diabetes of about 12–15% per risk allele. This risk effect size is consistent across studies, apparently no matter how many, or which, loci are considered, or whether the resultant genotype score is weighted or not. Furthermore, while groups with apparently dramatically increased genetic risk can be defined, these groups are relatively uncommon. The great majority of European ancestry population samples will be at average or low risk of type 2 diabetes on the basis of their genotype, even if some individuals in these groups have an elevated personal risk on the basis of their phenotype. In addition, the marginal ability of genotype scores to discriminate risk is minuscule, with minimal impact on discrimination after consideration of even a few common phenotypic risk factors.

Although the data reported by Lin et al. [6] primarily confirm other data, the paper is engagingly written, the methods are valid and clearly presented, and the paper offers an excellent précis on the current state of genetic prediction for type 2 diabetes. The one putative innovation of this study is the use of the weighted score over a simple, unweighted allele-counting approach. However, the advantage of this approach is uncertain: the odds ratio for

prevalent diabetes per weighted risk allele was 1.15 (95% CI 1.10–1.20), whereas it was 1.13 (95% CI 1.08–1.19) with use of an unweighted score. The *c* statistic was 0.57 for the unweighted score and 0.59 for the weighted score. The authors argue that the weighting approach improves discrimination. It does, but barely perceptibly. The weighted score is more difficult to calculate, explain and interpret than a simple allele-counting genotype score. Whether the subtle difference provided by weighting the score matters when the score is divided into low, medium and high genotype score groups was not addressed. It is likely that reducing the score to these simpler terms, for instance, to convey risk to patients in a genetic counselling framework, would eliminate any marginal benefit of a more sophisticated genotype score approach. The one other study to examine the performance of a weighted genotype score did not find much of an advantage of the approach over a simple risk allele-counting method [12]. However, further work to define a ‘best’ genetic model is warranted. Additional SNPs are rapidly becoming known and can be added to extend the gradient of genetic risk. A broader gradient can more readily be used to estimate an optimal cut-off point defining a high-risk genotype score, and individuals at the high end of the distribution will no doubt have quite a high personal genetic risk of diabetes. Calibration of models may be further improved by using different approaches to weighting individual SNPs, modelling each allele according to its apparent strongest mode of transmission (the genotype score assumes an additive model; at some loci, a recessive model may more accurately reflect the locus-specific effect), or accounting for locus–locus and locus–environment interactions. Furthermore, as discovery sample sizes increase and/or resequencing efforts uncover rarer variants with stronger effects than currently seen for common variants, their addition to a genotype score could substantially improve risk discrimination. Also, most current risk loci are considered as such on the basis of association data; it is unclear how many of these are the actual causal, functional variant. Loci that are not causal and that have limited correlation with the causal locus will have weaker risk effects than true causal loci or variants in high linkage disequilibrium with causal loci. Inclusion of true causal loci or their perfect proxies in a risk score may further strengthen the ability of genetics to predict diabetes. Finally, although marked ancestry-specific risk allele frequency differences across populations [13] probably require population-specific SNP sets or weights, the approach is readily adaptable to any population in which common disease risk alleles have been identified.

Lin et al. confirm that diverse diabetes risk loci can be combined into a genotype score that is strongly associated with risk of type 2 diabetes. However, this is merely the first step in the clinical translation of genetic prediction of diabetes

to improved health. The next step is to demonstrate that this information can beneficially alter patient risk perception, health behaviour and clinical outcomes. Explaining a genotype score to a patient at risk of diabetes requires that the patient be able to interpret probabilistic information concerning disease risk, even if the score is simplified into categories of low, average and high risk. Physicians, too, will need to learn how to think of genetic risk in a probabilistic framework that is different from that used to consider the risk of classical Mendelian disorders. The need to understand probabilistic risk is true for all forms of genetic screening and counselling, but in the case of type 2 diabetes, the consequent action based on discovery of a high-risk genotype is not discrete but a lifetime of changed behaviour. In the case of breast cancer genetic screening, for instance, the genetic test result leads to a discrete set of strategic choices for breast cancer prevention. In the case of type 2 diabetes genetic screening, a high-risk patient must make new decisions about diet, exercise and weight management on a daily basis over the course of many subsequent years to prevent the disease. It is not known whether people will make a sufficiently strong connection between their own genetic risk and the daily lifestyle choices required to prevent diabetes. Furthermore, individuals with relatively less healthy lifestyle behaviours might be more susceptible to genetic risk than those with healthier behaviours [14]. Whether the genotype score would have beneficial value in specific subgroups at elevated risk on the basis of poor health habits remains to be tested. At least as important, however, is that the majority of people screened for diabetes genetic risk will be at average or low risk on the basis of genotype. Whether people at high phenotypic risk but who know that they have low genetic risk would have diminished personal risk perception or unchanged unhealthy behaviours is not known. Perhaps the most important lesson from the paper of Lin et al. [6] is that phenotypic risk factors for type 2 diabetes are far superior to currently known genetic risk factors as harbingers of future disease risk. Phenotypic risk factors are, by and large, also the actionable risk factors for prevention. Lin et al add to the evidence that, despite rapid growth in our knowledge of the genetic underpinnings of type 2 diabetes, as of 2009, identification of adverse phenotypic characteristics remains the cornerstone of type 2 diabetes prediction. Whether genetic risk information can be leveraged to aid current diabetes prevention and treatment efforts is an exciting frontier of inquiry that arises from this prediction research.

Acknowledgements Supported by the National Institute for Diabetes and Digestive and Kidney Diseases R01 DK078616 and K24 DK080140.

Duality of interest J. B. Meigs currently has research grants from GlaxoSmithKline and sanofi-aventis, and serves on consultancy boards for GlaxoSmithKline, sanofi-aventis, Interleukin Genetics, Kalypsis and Outcomes Sciences.

References

1. McCarthy MI, Abecasis GR, Cardon LR et al (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 9:356–369
2. Wray NR, Goddard ME, Visscher PM (2007) Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res* 17:1520–1528
3. Altshuler D, Daly MJ, Lander ES (2008) Genetic mapping in human disease. *Science* 322:881–888
4. McCarthy MI, Hattersley AT (2008) Learning from molecular genetics: novel insights arising from the definition of genes for monogenic and type 2 diabetes. *Diabetes* 57:2889–2898
5. Janssens AC, Gwinn M, Bradley LA, Oostra BA, van Duijn CM, Khoury MJ (2008) A critical appraisal of the scientific basis of commercial genomic profiles used to assess health risks and personalize health interventions. *Am J Hum Genet* 82:593–599
6. Lin X, Song K, Lim N et al (2009) Risk prediction of prevalent diabetes in a Swiss population using a weighted genetic score—the CoLaus Study. *Diabetologia*. doi:10.1007/s00125-008-1254-y
7. 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
8. Vaxillaire M, Veslot J, Dina C et al (2008) Impact of common type 2 diabetes risk polymorphisms in the DESIR prospective study. *Diabetes* 57:244–254
9. Lango H, Palmer CN, Morris AD et al (2008) Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk. *Diabetes* 57:3129–3135
10. van Hoek M, Dehghan A, Witteman JC et al (2008) Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study. *Diabetes* 57:3122–3128
11. Lyssenko V, Jonsson A, Almgren P (2008) Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med* 359:2220–2232
12. Meigs JB, Shrader P, Sullivan LM et al (2008) Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med* 359:2208–2219
13. McCarthy MI (2008) Casting a wider net for diabetes susceptibility genes. *Nat Genet* 40:1039–1040
14. Florez JC, Jablonski KA, Bayley N et al, for the Diabetes Prevention Program Research Group (2006) *TCF7L2* polymorphisms and progression to diabetes in the Diabetes Prevention Program. *N Engl J Med* 355:241–250