Brendan J. Keating[1,2,3]

Advances in Risk Prediction of Type 2 Diabetes: Integrating Genetic Scores With Framingham Risk Models

Type 2 diabetes (T2D) is a worldwide epidemic affecting 8.3% of the current global adult population (1), with a growing prevalence in both developed and developing nations. It is estimated that of 29.1 million U.S. individuals with T2D, 8.1 million (27.8%) are undiagnosed—up from 7.0 million (27.1%) in 2010. T2D is strongly heritable, with estimates ranging from 30% to 70% (2). As elevated blood glucose can precede the diagnosis of T2D by up to 10 years, irreversible T2D-related complications including neuropathies and renal disease often begin to develop before diagnosis. Lean offspring of T2D patients are more likely to be insulin resistant than lean offspring of nondiabetic subjects, which appears to be caused in part by inherited mitochondrial defects that result in reduced glucose and fatty acid oxidation efficiency, leading to lipotoxicity and fat accumulation inside of muscle cells (3). With the rapid advancement of high-throughput genotyping technologies and the aggregation and meta-analysis of large-scale genome-wide association studies (GWAS) and gene-centric T2D studies (4–6), >70 loci have been identified, although only ~10% of the genetic variance of T2D risk is explained by these signals. Over the last five years, the field has progressed toward integrating genetic risk scores (GRS) with conventional T2D risk scores such as the Framingham Offspring Study risk score (FORS) (7). If sufficient improvement in risk prediction can be attained with integration of GRS into conventional risk models, there may be clinical utility in earlier identification of higher-risk individuals, particularly those individuals with lower BMI or who lack the traditional nongenetic risk factors.

In 2008, the first T2D GRS analysis in the Framingham Offspring Study (FOS) used an 18-SNP T2D risk score in 2,377 T2D-free participants followed for 28 years, of whom 225 (9.5%) developed T2D. They found a modest but significant 12% relative risk increase of T2D incidence per risk allele. Irrespective of FORS though, individuals possessing the highest GRS (>21 genotype score) compared with those with the lowest GRS (<15 genotype score) had increased T2D risk (odds ratio [OR] 2.6) (8). Concurrently, a Scandinavian study took a similar approach using a 16-SNP risk model and similarly found that the GRS had a slight improvement for incident T2D prediction when compared with a score of clinical risk factors alone (9). A 2011 study by de Miguel-Yanes et al. (10) used longer follow-up periods of incident T2D outcomes in FOS, where 446 individuals (12.8%) developed T2D over a period of 34 years, and used a 40-SNP risk score to show utility for T2D risk prediction in those aged <50 years but not in older individuals. Vassy et al. (11) recently reanalyzed the same FOS data from the de Miguel-Yanes 2011 study (10), together with a smaller data set from younger Coronary Artery Risk Development in Young Adults (CARDIA) subjects, but using an updated 62-SNP model, also showed a modest increase in the added value of T2D incident prediction. This suggests that GRS may be more useful in younger individuals, fitting the biological model that diseases that affect young individuals are more likely to have stronger genetic determinants.

In this issue of *Diabetes*, a new article by Talmud et al. (12) employed a weighted T2D GRS using 65 variants derived from recent large-scale T2D association meta-analyses to examine the impact on T2D risk assessment in seven prospective studies from the UK-based UCLEB (University College London-London School of Hygiene and Tropical Medicine-Edinburgh-Bristol) Consortium of prospective studies. Of 13,294 individuals who were initially free from T2D, 804 developed T2D over a median of 10 years' follow-up. Talmud et al. (12)

[1]Division of Genetics, Children's Hospital of Philadelphia, Philadelphia, PA
[2]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA
[3]Division of Transplantation, Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

Corresponding author: Brendan J. Keating, bkeating@mail.med.upenn.edu.

COMMENTARY

compared the performance of the GRS with the phenotype-based FORS risk model, and then the two models in combination, and investigated whether risk prediction, inclusive of GRS, differs by age and by BMI.

When the top versus bottom quintiles of genetic score were compared, an OR of 2.7 was observed for T2D development in this data set. When the authors used a 10% false-positive rate, they found that the genetic score alone detected 19.9% of incident cases, while using the FORS alone detected 30.7% and, in combination, detection of 37.3% of cases was observed. The areas under the receiver operator curves (A_{ROC}) were 0.60, 0.75, and 0.76, respectively, showing only a marginal improvement in the discrimination after the addition of GRS to phenotypic information. They also used a net reclassification improvement (NRI), a metric that quantifies the extent to which the combined scores move individuals to risk categories that more accurately assess the manifestation of their disease status (13). They observed that the addition of the GRS to the FORS did result in an NRI of 8.1% in this UCLEB data set. The authors also examined whether genetics plays a larger role in incidence T2D risk in different tertiles of BMI. They showed that individuals in the lowest BMI tertile (<24.5 kg/m^2) had an NRI of 27.6% compared with those in the top tertile, with an NRI of 2.6%, showing greater utility in the NRI reclassification for prediction using GRS plus FORS within leaner individuals. Interestingly, there was no differential effect by age categories.

A major strength of the current study is the use of larger independent studies other than the relatively smaller FOS, which was used in three of the previous studies (8,10,11), and FORS; thus, these earlier results cannot be considered independent of each other. The 2008 Scandinavian T2D GRS study (9), while well powered with 2,201 incident T2D cases, used only 16 SNPs in the GRS, limiting the extent of genetic contribution to the scores examined versus the updated 65-SNP panel in the Talmud study (12). The Talmud et al. study also excluded latent autoimmune diabetes in adults cases, as in many genetic studies ~10% of T2D "cases" are actually cases of latent autoimmune diabetes in adults and thus misclassified. A weakness of the study is that the length of follow-up is not as long as many of the previous studies (see Table 1), although this will likely be updated over the next number of decades.

The two major challenges of genetic and phenotypic data-set integration for T2D risk prediction include an incomplete knowledge of the genetic variance and phenotypic heterogeneity. While most of the common genetic variants having associations with the strongest main effects has likely been unveiled through GWAS and second-generation sequencing, additional portions of the genetic variance may be explained by a number of factors including rarer higher-penetrant variants, epigenetics, gene environment, gene-gene interactions, and sex-specific genetic signals (5). Major challenges in accruing sufficient incident

Table 1—Major GRS studies performed for T2D incidence prediction

Author, year (ref. no.)	SNP in GRS	Cohort(s)	Follow-up period (years)	Incident T2D cases, n (%)	GRS improvement	Significant improvement in AROC/C-statistic	Improvement in NRI for FORS
Meigs, 2008 (8)	18	FOS	28	225 (9.5)	Yes: per SD $P < 0.001$	C-statistic = 0.534 for FORS, 0.581 for FORS + GRS ($P = 0.001$)	Yes: up to 4%
Lyssenko, 2008 (9)	16	MPP and Botnia Study	23.5 (median)	2,201 (11.7)	Yes	Modest increase (0.74–0.75, $P = 1.0 \times 10^{-4}$)	Yes: 9% in MPP ($P = 2.5 \times 10^{-5}$) and 20% in Botnia ($P = 0.05$)
Talmud, 2010 (18)	20	Whitehall II	10	302 (5.5)	Yes	No	No
de Miguel-Yanes, 2011 (10)	40	FOS	34	446 (12.8)	Yes: in <50-year-olds only ($P = 0.02$)	No: C-statistic increase from 0.908 to 0.911 ($P = 0.3$)	Yes: 10.2% ($P < 0.001$)
Vassy, 2014 (11)	62	FOS	25.6	446 (12.8)	Yes	C-statistic improved in FOS ($P < 0.001$) but not in CARDIA	Weak increase
	62	CARDIA	24.2	97 (5.9)			
Talmud, 2015 (12)	65	UCLEB	10 (median)	804 (6.0)	Yes: from 30.7 (FORS only) to 37.3% (10% FPR)	Modest increase	Yes: (8.1%), $P = 3.3 \times 10^{-7}$

FPR, false-positive rate; MPP, Malmö Preventive Project, Sweden.

cases and samples to discover such effects remain. Improvements to the portion of genetic variance explained could arise through use of weaker sub–genome-wide signals through the techniques such as genome-wide complex trait analysis (GCTA) (14), as demonstrated in recent publications in T2D (5) and height (15). The polygenic nature of T2D is complex with loci impacting gluconeogenesis, glucose transport, insulin homeostasis, and satiety, subsequently impacting the manifestation, progression, severity, and even the treatment of T2D (16). Phenotypic complexity includes differing mechanisms leading to T2D as well as cases that remain undiagnosed. The heritability of maturity-onset diabetes of the young is typically monogenic in nature with approximately a dozen genes implicated, primarily impacting glucose homeostasis. Risk variants in these same maturity-onset diabetes of the young genes are also enriched in T2D cases, providing another example of common variants in genes involved in Mendelian disease that also contribute to risk of common disease (17).

Even as we struggle with the clinical meaningfulness, cost-benefit metrics, and justifications of preemptive genetic testing, there are major implementation challenges ahead. These include implementing clinical genotyping/sequencing; Consent Education, Regulation and Consultation (CERC) between the patient, health care system, and physicians; and return of results using clinical decision support models. Implementation of solutions to these problems are proceeding on a large scale for a number of diseases across 10 major U.S. hospital networks/institutes by the Electronic Medical Records and Genomics (eMERGE) Network. eMERGE is using discovery GWAS and electronic medical records data from 60,000 individuals taken from the biorepositories of several hundreds of thousands patients. Genetic and environmental outcomes across diabetic comorbidities have already been examined, with a major focus on integrating clinically guided dosing of drug therapies using pharmacogenomic approaches (19,20). While T2D GRS will likely improve incrementally, the clinical utility remains to be determined at national scales, although it is likely that benefits will be reaped by at-risk populations such as lean individuals with T2D who may not present to primary care with later stages of disease manifestation.

References

1. International Diabetes Federation. IDF Diabetes Atlas, 5th Edition. Available from http://www.idf.org/diabetesatlas/5e/. Accessed 19 February 2012

2. Newman B, Selby JV, King MC, Slemenda C, Fabsitz R, Friedman GD. Concordance for type 2 (non-insulin-dependent) diabetes mellitus in male twins. Diabetologia 1987;30:763–768

3. Petersen KF, Dufour S, Befroy D, Garcia R, Shulman GI. Impaired mitochondrial activity in the insulin-resistant offspring of patients with type 2 diabetes. N Engl J Med 2004;350:664–671

4. Voight BF, Scott LJ, Steinthorsdottir V, et al.; MAGIC investigators; GIANT Consortium. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 2010;42:579–589

5. Morris AP, Voight BF, Teslovich TM, et al.; Wellcome Trust Case Control Consortium; Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) Investigators; Genetic Investigation of ANthropometric Traits (GIANT) Consortium; Asian Genetic Epidemiology Network–Type 2 Diabetes (AGEN-T2D) Consortium; South Asian Type 2 Diabetes (SAT2D) Consortium; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 2012;44:981–990

6. Saxena R, Elbers CC, Guo Y, et al.; Look AHEAD Research Group; DIAGRAM consortium. Large-scale gene-centric meta-analysis across 39 studies identifies type 2 diabetes loci. Am J Hum Genet 2012;90:410–425

7. Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB Sr. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. Arch Intern Med 2007;167:1068–1074

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;359:2208–2219

9. 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;359:2220–2232

10. de Miguel-Yanes JM, Shrader P, Pencina MJ, et al.; MAGIC Investigators; DIAGRAM+ Investigators. Genetic risk reclassification for type 2 diabetes by age below or above 50 years using 40 type 2 diabetes risk single nucleotide polymorphisms. Diabetes Care 2011;34:121–125

11. Vassy JL, Hivert MF, Porneala B, et al. Polygenic type 2 diabetes prediction at the limit of common variant detection. Diabetes 2014;63:2172–2182

12. Talmud PJ, Cooper JA, Morris RW, et al.; UCLEB Consortium. Sixty-five common genetic variants and prediction of type 2 diabetes. Diabetes 2015;64:1830–1840

13. Pencina MJ, D'Agostino RB Sr, Demler OV. Novel metrics for evaluating improvement in discrimination: net reclassification and integrated discrimination improvement for normal variables and nested models. Stat Med 2012;31:101–113

14. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. Am J Hum Genet 2011;88:294–305

15. Wood AR, Esko T, Yang J, et al.; Electronic Medical Records and Genomics (eMEMERGEGE) Consortium; MIGen Consortium; PAGEGE Consortium; LifeLines Cohort Study. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet 2014;46:1173–1186

16. Pearson ER, Donnelly LA, Kimber C, et al. Variation in TCF7L2 influences therapeutic response to sulfonylureas: a GoDARTs study. Diabetes 2007;56:2178–2182

17. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature 2010;466:707–713

18. Talmud PJ, Hingorani AD, Cooper JA, et al. Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. BMJ 2010;340:b4838

19. Gottesman O, Kuivaniemi H, Tromp G, et al.; eMERGE Network. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet Med 2013;15:761–771

20. Hall MA, Dudek SM, Goodloe R, et al. Environment-wide association study (EWAS) for type 2 diabetes in the Marshfield Personalized Medicine Research Project Biobank. Pac Symp Biocomput 2014;200–211