Genetic risk scores ascertained in early adulthood and the prediction of type 2 diabetes later in life

P. W. Franks

Received: 6 July 2012 / Accepted: 16 July 2012 / Published online: 11 August 2012
© Springer-Verlag 2012

Abstract It is hoped that information garnered from studies on population genetics will one day be translated into a form in which it meaningfully improves the prediction, prevention or treatment of type 2 diabetes. Type 2 diabetes genetics researchers have made extraordinary progress in identifying common genetic variants that are associated with type 2 diabetes, which has shed light on the biological pathways in which molecular defects that cause the disease likely reside. However, the expectation that genetic discoveries will aid the prevention or treatment of type 2 diabetes has not, so far, been fulfilled. In a paper published in this edition of the journal, Vassy and colleagues (DOI: 10.1007/s00125-012-2637-7) test the hypothesis that the predictive accuracy of established genetic risk markers for type 2 diabetes varies by age, with the predictive accuracy being greatest in younger cohorts. The authors found no substantive support for this hypothesis. However, a number of interesting questions are raised by their study concerning why risk alleles for a given genotype may differ in younger and older cohorts and why prospective cohort studies may yield results that are inconsistent with those derived from cross-sectional studies; this commentary discusses these points.

P. W. Franks (✉)
Department of Clinical Sciences, Genetic and Molecular
Epidemiology Unit, Lund University Diabetes Center,
Skåne University Hospital Malmö, CRC, Entr 72,
Building 60, Level 12,
205 02 Malmö, Sweden
e-mail: paul.franks@med.lu.se

P. W. Franks
Department of Nutrition, Harvard School of Public Health,
Boston, MA, USA

P. W. Franks
Genetic Epidemiology & Clinical Research Group,
Department of Public Health & Clinical Medicine,
Section for Medicine, Umeå University,
Umeå, Sweden

Keywords Gene×environment interaction · Genetic association study · Genetic risk score · Predictive accuracy · Risk prediction · Young onset type 2 diabetes

Abbreviation

CARDIA Coronary Artery Risk Development in Young Adults

To date, more than 100 common genetic variants have been reliably associated with type 2 diabetes and its quantitative metabolic traits [1, 2]. The majority of these variants were discovered using hypothesis-free methods such as genome-wide association studies (GWAS) and other targeted, massively parallel genotyping technologies such as Metabochip. To justify the extraordinarily high costs of genetic association studies, which now runs into billions of euros [3], population geneticists have argued that the discoveries emerging from these studies would (1) elucidate the biological basis of complex disease; (2) improve targeted prevention and treatment of disease; and (3) improve the predictive accuracy of conventional risk prediction algorithms. Whilst genetic association studies have indeed opened up many interesting leads concerning the biology of type 2 diabetes, they have done little to improve the prediction, prevention or treatment of the disease.

Findings from several prospective cohort studies [4–6] and clinical trials [7] of older adults have been published that document the predictive accuracy of type 2 diabetes genetic risk scores and compared these with conventional and complex non-genetic risk prediction algorithms. Disappointingly, these studies show that common genetic variants do very little to improve the predictive accuracy of non-genetic risk scores comprised of variables such as age, BMI, sex and diabetes family history. However, some prospective cohort studies indicate that as time to event increases, so too does the predictive accuracy of genetic risk scores, whilst the accuracy of non-genetic clinical risk algorithms decreases [4–6]. These

findings are reinforced by studies in Pima Indians [8] and Finns [9], which show that type 2 diabetes and related metabolic traits have a substantially larger heritable component at younger than at older ages. Some have reasoned that if the relationship between age and genetic risk for type 2 diabetes is linear, obtaining genetic information early in life might help guide preventive interventions, long before the clinical signs and symptoms of diabetes emerge.

In this issue of *Diabetologia* [10], Vassy and coworkers hypothesised that genetic risk scores ascertained early in life would have greater predictive accuracy for type 2 diabetes than when modelled in older adult cohorts. The authors used data from the Coronary Artery Risk Development in Young Adults (CARDIA) [11] study to examine whether 38 reliably associated common genetic variants, derived mainly from cross-sectional adult case-control studies [1], predicted the onset of type 2 diabetes (see Fig. 1). The CARDIA study is comprised of 5,115 black and white adolescents and young adults (18–30 years at baseline) from the USA. Baseline assessments were performed from 1985 to 1986, and participants were followed, on average, for nearly 24 years, with 215 cases of diabetes detected during this time. When modelled in aggregate, each unit (allele) of the genetic risk score corresponded with a 9% (95% CI 5%, 13%) increased hazard of incident diabetes on average. There was no material increase in the C statistic when the genetic risk score

was added to conventional clinical risk prediction models, and there was little gain in net reclassification (maximum net reclassification index of 0.285). The conclusion from these results is that a genetic risk score informed by existing data on common variant associations from adult cross-sectional studies does not meaningfully improve the predictive accuracy of conventional, non-genetic risk prediction algorithms for incident type 2 diabetes in early adulthood. Of note, the authors have conducted a second study in which they tested the same hypothesis in an even younger cohort (14 years old at baseline) from the Bogalusa Heart Study [12], which showed similar results, reinforcing the veracity of these null findings. However, before writing off the possibility that type 2 diabetes genetic risk scores derived in early life are of clinical value, we should consider whether methodological differences between the study by Vassy et al and those alluded to above might explain the disappointing results.

Could characteristics that vary between the cohorts studied here and previously explain the incompatible results?

The Vassy et al study focused on a cohort of white and black participants from the USA, born 1955–1973. The Lyssenko et al study focused on two cohorts of white participants from Sweden ($n=16,061$) and Finland ($n=2,770$) born in the first

Fig. 1 Comparison of odds ratios (blue bars) from published type 2 diabetes genetic association studies and hazard ratios (red bars) from Vassy et al [9] for 38 established type 2 diabetes loci. ^aIndicates nominally statistically significant effect and contrasting risk allele; ^bdata only available in white participants

half of the 20th century [5]. DNA was obtained at a follow-up exam in the Swedish cohort; importantly, DNA was unavailable in those who had died between baseline and follow-up, which may have affected the observed genotype frequencies. The studies from Meigs et al [4] and de Miguel-Yanes et al [6] were conducted in the Framingham Offspring Study, a cohort of predominantly white participants born 1930–1940 in Massachusetts, USA. Differences in birth era, cultural factors and genetic substructure of the cohorts described above might influence the comparability of results, owing to gene environment interactions or differences in population substructure. One interesting aspect of the study by Vassy et al is that although some of the individual SNP association tests are likely to be underpowered (and thus yield false-negative results), the risk alleles for other SNPs contrast those reported in published studies comprised primarily of older adults (i.e. the diabetes risk raising allele reported previously appears to lower diabetes risk in the study by Vassy et al). Some of these differences will be attributable to random variation, but others highlight the possibility of cohort-dependent effects, which may include gene $\times$ age interactions. For example, the diabetes risk alleles at *PROX1* rs340874, *HCCA2* rs2334499 and *KCNJ11* rs5219 are associated with lower diabetes risk in the published literature [1]; because these effects were nominally statistically significant in the Vassy et al study, the risk allele differences for these loci are unlikely to be due to random variation and may represent genuine cohort-specific effects.

Importantly, before concluding that gene $\times$ age interactions do indeed exist for *PROX1*, *HCCA2* and *KCNJ11*, formal statistical tests of interaction would need to confirm this in one or more adequately powered and appropriately designed prospective cohort studies. Moreover, observations of gene $\times$ environment interactions in epidemiological studies provide little or no evidence of causal effects, as methodological factors such as heteroscedasticity, confounding, reverse causality, and the scale on which data are presented are frequent non-causal explanations for statistically significant findings on gene $\times$ environment interactions, even for interaction effects that have been replicated in independent cohorts. Thus, whilst observations of interactions may generate intriguing hypotheses, experimental studies are often necessary to test those hypotheses and congruent biological explanations should always be pursued.

Interestingly, existing evidence suggests that two of the loci outlined above may have age-dependent effects on the risk of diabetes per se. *PROX1*, for example, is a co-repressor of *HNF4A*, and mutations in the latter gene cause MODY type 1 [13]. Similarly, mutations in *KCNJ11*, one of the first type 2 diabetes loci to be discovered, have long been associated with permanent neonatal diabetes [14] and very recently with MODY [15].

Should we expect the genetic risk factors for younger and older onset type 2 diabetes to be the same?

By virtue of the study design, participants with diabetes in the Vassy et al cohort were considerably younger at diagnosis (mean age 44.1 years) than those in previous prospective cohort studies and cross-sectional case–control studies of type 2 diabetes (~65 years). Type 2 diabetes is diagnosed by excluding other known causes of chronically elevated blood glucose concentrations, which probably captures a range of underlying pathologies all lumped together under a single diagnosis; the molecular defects that cause each specific pathology may differ in frequency by diabetes age of onset. Thus, the genetic markers of these molecular defects may also differ in younger compared with older diabetic cohorts. This is important, as the genetic markers studied by Vassy et al were discovered primarily through cross-sectional case–control studies of older adults.

Should we expect genetic markers of type 2 diabetes discovered in cross-sectional studies to predict incident diabetes?

Almost without exception, cross-sectional studies have been used to discover novel loci for type 2 diabetes [1]. Whilst cross-sectional case–control studies are generally more accessible than prospective studies, making them attractive for large-scale investigations, one should not expect both study designs to yield identical results. Indeed, recent results from the European Network for Genetic and Genomic Epidemiology (ENGAGE) Consortium support this notion [16]. Reasons for differing results by study design may include that estimates of risk and discriminative accuracy in cross-sectional studies can be considerably higher than in prospective studies; for example, in the former, cases and controls are sometimes selected so as to maximise power to detect differences in genotype frequencies between the two groups, which may also inflate risk estimates (a source of spectrum bias). Another often overlooked issue is that with type 2 diabetes and other fairly prevalent diseases, odds ratios derived from case–control studies that are not appropriately matched are prone to be larger than risk estimates such as rate ratios or incidence density ratios derived from prospective cohort studies [17, 18]. Thus, for a variety of methodological reasons, the often-cited odds ratios for type 2 diabetes loci from consortia such as Diabetes Genetics Replication and Meta-analysis Consortium (DIAGRAM) [1] should not automatically be considered analogous to hazard rate ratios derived from prospective studies.

Whilst the study by Vassy et al will be interpreted by many as adding another nail to the coffin of genetic risk prediction models for type 2 diabetes, it is important to

understand why their results do not fit with the expectation that genetic risk models perform better at younger ages. It is also important to bear in mind that the genetic risk prediction models for type 2 diabetes examined to date have focused on common gene variants, and we simply cannot conclude at this stage whether rare variants will or will not be clinically useful for prediction. We should also consider that, if the risk alleles for specific loci truly vary by age (the risk alleles for 11 of the 38 SNPs studied in the paper by Vassy et al contrast those reported in the published literature, see Fig. 1), genetic risk algorithms derived in adulthood will be inappropriate for use in younger populations, and algorithms that are specific to this younger age group, where effect alleles are coded appropriately, will be required. The study by Vassy et al provides useful information in this regard, which may inform the development of risk prediction algorithms for future studies of young onset type 2 diabetes.

Acknowledgements I thank W. C. Knowler and B. Grice from the Phoenix Epidemiology and Clinical Research Branch of the NIDDK for their helpful comments pertaining to this manuscript.

Funding The author is supported by grants from the Swedish Diabetes Association, Novo Nordisk, the Swedish Heart-Lund Foundation, the European Union, the National Institutes of Health, and Exodiab.

Duality of interest The author declares that there is no duality of interest associated with this manuscript.

Contribution statement PWF was the sole contributor to this paper.

References

- Voight BF, Scott LJ, Steinthorsdottir V et al (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 42:579–589
- Scott RA, Lagou V, Welch RP, et al. Large-scale association study using the Metabochip array reveals new loci influencing glycemic traits and provides insight into the underlying biological pathways. *Nat Genet* (in press)
- Gitlin JM. Calculating the economic impact of the Human Genome Project. Available from <http://www.genome.gov/27544383>. Accessed 1 July 2012
- 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
- Lyssenko V, Jonsson A, Almgren P et al (2008) Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med* 359:2220–2232
- de Miguel-Yanes JM, Shrader P, Pencina MJ et al (2011) 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* 34:121–125
- Hivert MF, Jablonski KA, Perreault L et al (2011) Updated genetic score based on 34 confirmed type 2 diabetes loci is associated with diabetes incidence and regression to normoglycemia in the diabetes prevention program. *Diabetes* 60:1340–1348
- Hanson RL, Knowler WC (1998) Analytic strategies to detect linkage to a common disorder with genetically determined age of onset: diabetes mellitus in Pima Indians. *Genet Epidemiol* 15:299–315
- Almgren P, Lehtovirta M, Isomaa B et al (2011) Heritability and familiarity of type 2 diabetes and related quantitative traits in the Botnia Study. *Diabetologia* 54:2811–2819
- Vassy JL, Durant NH, Kabagambe EK et al (2012) A genotype risk score predicts type 2 diabetes from young adulthood: the CARDIA study. *Diabetologia*. doi:10.1007/s00125-012-2637-7
- Friedman GD, Cutter GR, Donahue RP et al (1988) CARDIA: study design, recruitment, and some characteristics of the examined subjects. *J Clin Epidemiol* 41:1105–1116
- Vassy JL, Mahapatra PD, Meigs JB, et al (2012) Genotype predicts type 2 diabetes in adulthood in a biracial adolescent population. *Circulation* 125 (abstract)
- Yamagata K, Oda N, Kaisaki PJ et al (1996) Mutations in the hepatocyte nuclear factor-1 α gene in maturity-onset diabetes of the young (MODY3). *Nature* 384:455–458
- Gloyn AL, Pearson ER, Antcliff JF et al (2004) Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes. *N Engl J Med* 350:1838–1849
- Bonnefond A, Philippe J, Durand E et al (2012) Whole-exome sequencing and high throughput genotyping identified *KCNJ11* as the thirteenth MODY gene. *PLoS One* 7:e37423
- Lyssenko V (2012) Genetics—clinical implications. Genetic prediction models for diabetes risk. Abstracts of the 72nd Scientific Sessions of the American Diabetes Association, 8–12 June 2012. American Diabetes Association, Philadelphia, p 11 (abstract)
- Zhang J, Yu KF (1998) What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *JAMA* 280:1690–1691
- Miettinen O (1976) Estimability and estimation in case-referent studies. *Am J Epidemiol* 103:226–235