

EXPERT REVIEW ON PRECISION MEDICINE

The value of genetic risk scores in precision medicine for diabetes

1. Introduction

The notion that patterns of genetic variation at loci implicated in disease risk can offer clinically-useful stratification of disease aetiology has driven efforts to develop personalised approaches for complex traits [1]. The ability to measure individual predisposition will, it is anticipated, improve the precision with which we predict disease progression, complication risk, and response to preventative or therapeutic interventions.

However, despite substantial advances over the past decade that have led to identification of thousands of robustly-associated risk variants across a range of common, complex diseases, there are few examples of the clinical use of such information. In this review, we describe some of the reasons for this apparent lack of translational success, but offer an optimistic view of the potential for that assessment to change. We focus in particular on type 2 diabetes (T2D), a major cause of morbidity and mortality globally, and simultaneously both a leading contributor to health costs, and a source of massive unmet clinical need.

2. (Almost) defeated by complexity?

As of mid-2018, the number of loci robustly implicated in T2D-risk has climbed beyond 400 [2]. Most reflect common variants identified through genome-wide association studies (GWAS), the largest now involving >1M individuals. In non-isolate populations, these variants have, at best, modest effects: the largest, such as variants near *TCF7L2* (in Europeans) and *PAX4* (in East Asians) have allelic odds-ratios between 1.3 and 1.5, but a typical figure for novel signals emerging from recent rounds of GWAS is nearer 1.05. Of late, the combination of exome sequencing, targeted array genotyping, and the application to genome-wide genotyping data of denser imputation reference panels, has increased the yield of risk-variants with lower allele frequencies [2,3,4]. Given the constraints associated with detection at genome-wide levels of significance, these lower-frequency variants tend to have larger effects than the common variant signals, including a few variants in the sub-1% range with allelic odds-ratios exceeding 2.0 [2].

Nevertheless, multiple lines of evidence indicate that the bulk of the inherited predisposition to T2D-risk is attributable to common variants [4]. Most of the (so-called) “missing” heritability (that is, the apparent gap between the genetic contribution to T2D-risk collectively attributable to signals reaching genome-wide significance, and the estimates of overall heritability derived from twin or family data) can be ascribed to a long “polygenic” tail of (mostly common) variants of individually miniscule effects.

This genetic architecture indicates a broad spread of risk across many hundreds, if not thousands, of loci. It implies that, even if sample sizes for genetic discovery increase by an order of magnitude or two, much of the genetic contribution to risk will not be quantifiable as loci reaching unequivocal levels of genome-wide significance. When combined with the knowledge that a major part of individual predisposition comes from non-genetic factors (lifestyle, behaviour, environment), the prospects for clinically-useful genetic risk prediction can appear poor.

3. Building a genetic risk score

Early analyses certainly bore this out. Lango and colleagues established that the 18 T2D-risk variants identified from early rounds of GWAS, collectively generated a C-statistic (the area-under-the-curve [AUC] for a plot of sensitivity against specificity) of only 60%, higher than random expectation (50%) but well short of the 80% most would consider necessary for clinically-useful prediction [5]. This compared unfavourably with the C-statistic (78%) achieved using more readily-available risk factors such as age, gender and BMI. Adding genetic information to these risk factors had little impact. Similar analyses were repeated as more risk-variants were discovered, but the C-statistic remained stubbornly close to 60% [6,7]. Researchers sensibly shied away from making claims about the value of genetics to support individual prediction, favoring instead a narrative emphasising the undoubted value of genetic discovery for mechanistic inference and target validation.

That reticence is now changing, for two main reasons. First, the most recent GWAS meta-analysis for T2D, involving close to 900,000 Europeans, has moved the needle substantially in terms of predictive power [2]. As well as identifying over 400 T2D-risk signals, this analysis captured close to 20% of overall risk liability across the full GWAS dataset. This equates to around half the median estimates of heritability from twin and family studies. An optimised polygenic risk score (PRS) comprising ~130,000 variants, when applied to UK Biobank, generated an AUC in the mid-60s, a significant improvement over previous estimates, and not dissimilar to that generated in the same sample using age and BMI. The second shift in perspective follows pioneering work on the use of genetic risk scores for hyperlipidemia and coronary artery disease [8]. When individuals within UK Biobank are stratified according to their optimised T2D PRS, those in the top 2.5% have a T2D prevalence close to ten times that of those in the lowest 2.5%. Assuming extrapolation from UK Biobank participants to the wider population, that equates to ~1M individuals in the UK who, on the basis of quantifiable genetic risk alone, have a lifetime risk of T2D approaching 50% [2].

The clinical value of such data remains a matter of considerable debate. It is, of course, true that for most in the population, with PRS values in the middle of the distribution, such a measure of overall genetic risk has limited clinical utility. There remain unresolved questions over the generalisability of such data (to those of different ethnic backgrounds, for example), and the extent to which genetic risk simply recaptures information much more easily-accessed through BMI, family history and ethnic background. However, at the same time, there is a powerful argument that, were suitable (GWAS or sequence) data already freely and universally accessible within medical records, there would be many for whom this would contribute to risk prediction for T2D (and for many other diseases, besides). This would be particularly true for those who, decades ahead of T2D presentation, were found to be at the upper end of the T2D-risk distribution (including some with quasi-monogenic penetrance), for whom targeted lifestyle intervention could be transformative.

4. Deconstructing genetic risk

The comments above focused on the clinical value of risk scores that capture overall risk of T2D. However, many of the decisions faced by patients and their physicians do not concern diabetes risk *per se*: as we have seen, there are already well-established risk factors for T2D-risk which require nothing more sophisticated than a calendar, a weighing-scale and a tape measure. Those other decisions reflect the clinical heterogeneity that is evident amongst those with adult-onset diabetes: they relate to disease progression, complication risk, and the risks and benefits of different treatment options. Could genetic risk scores be useful in these situations?

There is increasing evidence that they could. A genetic risk score composed of type 1 diabetes risk-variants facilitates diagnosis of the subset of those with late-onset diabetes (often assumed to have type 2) who instead have a type-1 like autoimmune process that presages rapid decline of islet function and early recourse to insulin replacement [9]. The 400 signals influencing T2D-risk provide compelling evidence of multiple mechanisms through which individual predisposition to T2D can be nudged up or down: some variants influence T2D-risk through increasing BMI, others through reducing insulin sensitivity, and yet others through adverse impact on incretin action, or insulin secretion [10,11].

A recently-described model of diabetes predisposition focuses attention on these intermediary processes which collectively contribute to T2D-risk [12]. According to this “palette” model, each is considered under multifactorial (genetic and non-genetic) control. Individual risk-profiles across the set of processes influence not only T2D-risk but also the phenotype and natural history of any diabetes that results. Recent studies have demonstrated that one can, using patterns of genetic association across diabetes-related quantitative traits and relationships to tissue-specific regulatory annotations, define clusters of T2D-risk loci with similar mechanisms of action [3,10], and, from these, generate “partial” or “process-specific” GRS (pGRS) that capture the genetic contribution to each intermediary process [10,13,14].

As expected, given their derivation, these deconstructed risk score profiles capture diverse aspects of the clinical heterogeneity within diabetes: individuals with high pGRS scores for insulin action variants, for example, variants are hyperinsulinemic and hypertriglyceridemic. But, crucially, these profiles are associated with differences in clinical outcome that were not “baked into” their derivation, such as differences in risk of diabetic nephropathy or coronary artery disease [13,14].

5. Next steps

There is much to be done to extend these findings. The clusters underlying the pGRS structure require better definition, and characterisation of their mechanistic basis. The genetic contribution to individual predisposition needs to be combined with information on the external (e.g. lifestyle) and internal (e.g. microbiome) environment, and with real-time measures of the clinical state (including, potentially, process-specific biomarkers) to provide an integrated description of individual trajectories along these different physiological axes. Then, we can explore how these profiles relate to the clinical outcomes that we seek to mitigate, and their value in optimising use of the therapeutic and preventative options at our disposal. There are reasons to be optimistic that these approaches will deliver clinically-useful advances in the personalisation of diabetes management in the coming decade.

ACKNOWLEDGEMENTS

See title page document.

REFERENCES

1. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nature Reviews Genetics*. 2018 2018/05/22. doi: 10.1038/s41576-018-0018-x.
2. Mahajan A, Taliun D, Thurner M, et al. Fine-mapping of an expanded set of type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps [10.1101/245506]. *bioRxiv*. 2018.
The largest genome-wide association study for T2D to date involving ~900,000 subjects, with substantial improvements in risk-variant detection, fine-mapping of causal variants, and risk-score definition
3. Mahajan A, Wessel J, Willems SM, et al. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. *Nature Genetics*. 2018 2018/04/01;50(4):559-571. doi: 10.1038/s41588-018-0084-1.
4. Fuchsberger C, Flannick J, Teslovich TM, et al. The genetic architecture of type 2 diabetes. *Nature*. 2016 Aug 04;536(7614):41-7. doi: 10.1038/nature18642. PubMed PMID: 27398621; PubMed Central PMCID: PMC5034897. eng.
An integrated analysis of whole genome, whole exome and GWAS data regarding T2D-risk, which defined the contribution of common and low-frequency variants to disease predisposition
5. Lango H, Palmer CN, Morris AD, et al. Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk. *Diabetes*. 2008 Nov;57(11):3129-35. doi: 10.2337/db08-0504. PubMed PMID: 18591388; PubMed Central PMCID: PMC5034897. eng.
6. van Hoek M, Dehghan A, Witteman JC, et al. Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study. *Diabetes*. 2008 Nov;57(11):3122-8. doi: 10.2337/db08-0425. PubMed PMID: 18694974; PubMed Central PMCID: PMC5034897. eng.
7. Meigs JB, Shrader P, Sullivan LM, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. *The New England journal of medicine*. 2008 Nov 20;359(21):2208-19. doi: 10.1056/NEJMoa0804742. PubMed PMID: 19020323; PubMed Central PMCID: PMC5034897. eng.
8. Khera AV, Emdin CA, Drake I, et al. Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. *The New England Journal of Medicine*. 2016 Dec 15;375(24):2349-2358. doi: 10.1056/NEJMoa1605086. PubMed PMID: 27959714; PubMed Central PMCID: PMC5034897. eng.
9. Oram RA, Patel K, Hill A, et al. A Type 1 Diabetes Genetic Risk Score Can Aid Discrimination Between Type 1 and Type 2 Diabetes in Young Adults. *Diabetes Care*. 2016 Mar;39(3):337-44. doi: 10.2337/dc15-1111. PubMed PMID: 26577414; PubMed Central PMCID: PMC5034897. eng.

10. Scott RA, Scott LJ, Magi R, et al. An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes*. 2017 Nov;66(11):2888-2902. doi: 10.2337/db16-1253. PubMed PMID: 28566273; PubMed Central PMCID: PMC5652602. eng.
11. Franks PW, McCarthy MI. Exposing the exposures responsible for type 2 diabetes and obesity. *Science (New York, NY)*. 2016 Oct 7;354(6308):69-73. doi: 10.1126/science.aaf5094. PubMed PMID: 27846494; eng.
12. McCarthy MI. Painting a new picture of personalised medicine for diabetes. *Diabetologia*. 2017 May;60(5):793-799. doi: 10.1007/s00125-017-4210-x. PubMed PMID: 28175964; eng.
This short paper describes some of the thinking behind the "palette" model of T2D predisposition
13. van Zuydam NR, Ahlqvist E, Sandholm N, et al. A Genome-Wide Association Study of Diabetic Kidney Disease in Subjects With Type 2 Diabetes. *Diabetes*. 2018 Apr 27. doi: 10.2337/db17-0914. PubMed PMID: 29703844; eng.
14. Udler MS, Kim J, von Grotthuss M, et al. Clustering of Type 2 Diabetes Genetic Loci by Multi-Trait Associations Identifies Disease Mechanisms and Subtypes [10.1101/319509]. *bioRxiv*. 2018.
One of the first applications of the partial GRS approach to T2D, demonstrating relationships between individual profiles and complication risk