COMMENTARY

Genomic medicine at the heart of diabetes management

Mark I. McCarthy [1,2,3,4]

Abstract Individual predisposition to type 2 diabetes is influenced by the combined effect of a constellation of genetic variants and a multitude of environmental exposures. Identification of the specific genetic variants involved, and the mechanisms through which they operate, provides a powerful approach for delivering biological insights that can drive translational benefit, one that is already widely exploited in the personalised management of monogenic and syndromic forms of diabetes. This commentary develops the argument that equivalent translational advances for more common forms of diabetes are unlikely to result solely from the ability to define more complete individual inventories of genetic risk and environmental exposure. They will also require identification of complex molecular signatures able to provide integrative, empirical, longitudinal readouts of disease progression. These signatures will track causal mechanisms and capture an individual's position within a complex spectrum of pathophysiological processes, thereby supporting personalised approaches to intervention and treatment. This is one of a series of commentaries under the banner '50 years forward', giving personal opinions on future perspectives in diabetes, to celebrate the 50th anniversary of *Diabetologia* (1965–2015).

Abbreviations
GLP-1   Glucagon-like peptide-1
GWAS   Genome-wide association studies

✉ Mark I. McCarthy
mark.mccarthy@drl.ox.ac.uk

1   Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Old Road, Oxford OX3 7LJ, UK

2   Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK

3   Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, UK

4   Department of Genetic Medicine and Development, University of Geneva Medical School, 1 Rue Michel-Servet, Geneva 1211, Switzerland

An individual's risk of developing type 2 diabetes reflects a mix of genetic predisposition and environmental influence. The evidence suggests heterogeneous aetiology and complex pathophysiology. The rising global prevalence of type 2 diabetes highlights the limitations of current preventative strategies, and high complication rates attest to the deficiencies of available treatments.

For many diseases, most obviously for rare diseases and increasingly for many cancers, human genetics has provided transformative insights into the biology of disease and defined entirely novel translational opportunities. In contrast, despite robust identification of scores of genetic loci influencing risk of type 2 diabetes [1, 2], the impact of these discoveries on clinical management has so far been negligible. The story is similar for most other common chronic diseases. Most of the large-scale efforts to accelerate the implementation of genomic medicine into clinical care (such as the UK's plan to sequence 100,000 genomes) are focused on rare diseases and cancer, and barely recognise the potential value for common chronic diseases. Some have concluded that the role of

🖄 Springer

genetics as an engine for clinical advances for common diseases has been massively oversold.

Such sentiments combine natural impatience with a failure to appreciate the true timescales for translation: most of this science is, after all, less than a decade old. But they do raise a series of questions that need to be addressed: what will it take to make genomic medicine a reality for common complex diseases, what would it look like, and what impact will it have over the next 50 years?

A sober assessment of the genetics of type 2 diabetes

Recent advances in unravelling the genetics of type 2 diabetes and other common complex diseases date back to 2007. The implementation of genome-wide association studies (GWAS), and consequent development of the scientific and anthropological framework that allowed large-scale data aggregation, have generated a catalogue of over 100 loci with robust associations to type 2 diabetes [1, 2]. Due to the technology used, most of the risk alleles discovered to date have been common. They have also been found to be widely represented across human populations. Both observations indicate that these type 2 diabetes risk alleles arose long ago in human prehistory, and suggest a degree of tolerance to their evolutionary impact that is consistent with their modest effects on glucose homeostasis. Cumulatively, these 100 loci explain around 10% of the variation in type 2 diabetes predisposition [1, 2].

Most estimates of the heritability of type 2 diabetes (a rather more ephemeral measure than most acknowledge) are much higher than this (perhaps 30–60%), prompting much speculation regarding the nature and extent of the component of the genetic variance in risk that remains undiscovered. Increasingly, that speculation can be supplanted with empirical insights. There is growing evidence that much, perhaps most, of the undiscovered genetic variance resides in a long tail of common variants of ever more modest effects [2]. Some of the remaining variance is likely to be distributed across rarer alleles, which are not surveyed by common variant focused GWAS but are now increasingly accessible through sequencing; amongst these rarer, younger, variants will be a subset with relatively large effects on individual risk. As far as we can tell from available data, little of the undiscovered variance resides in the interaction between genetic variants themselves or between genes and the environment [3].

The emerging picture of the genetic basis of type 2 diabetes is one of dazzling complexity. Where it has been possible to transmute those signals into biological inference, the data point to the involvement of multiple different aetiological pathways, reflecting the diverse mechanisms through which the complex machinery of glucose homeostasis can be perturbed.

New biology, new interventions

The management of type 2 diabetes 50 years from now, will, one can only hope, be almost unrecognisable to the patients and physicians of today, reliant as they are on a limited range of preventative and/or therapeutic options that all too rarely support a sustained return to normal metabolic homeostasis. The focus will be on preserving normal glucose homeostasis in those demonstrated to be at greatest risk, rather than dealing retrospectively with the consequences of its loss. The interventions available may be surgical or cell-based, but, equally, are likely to involve a new suite of drugs, engineered for efficacy and safety. Where intervention is required in those who have escaped prevention and developed disease, the approaches used will directly modify the disease process (to restore beta cell mass and/or function, and/or to correct the primary defects in insulin sensitivity) rather than, as now, attempting merely to control the most egregious manifestations of disease.

Human genetics will have taken centre stage. The risk variants revealed by population-level sequencing will have highlighted targets and pathways causally implicated in the development of human diabetes [4]. Drugs directed towards human validated targets will bring the promise of clinical efficacy, rescuing pharma from the death spiral of reliance on drugs found, far too late, to be capable of treating 'diabetes' only in mice or cells. At the same time, characterisation of the wider impact of diabetes risk variants on health and disease, providing advance warning of the side-effect profile of cognate interventions, will prioritise drugs that are not only safe, but provide collateral health benefits.

The rudiments of this approach are already visible. The genes encoding the targets of sulfonylureas and thiazolidinediones harbour variants associated with type 2 diabetes [1, 2]. Variants in and around the genes encoding the receptors for glucose-dependent insulinotropic polypeptide and glucagon-like peptide-1 (GLP-1), which are associated with diabetes risk or related glycaemic traits, represent a natural genetic experiment that prefigures the clinical value of the therapeutic manipulation of this pathway using GLP-1 receptor agonists and dipeptidyl peptidase-4 (DPP-4) inhibitors [2, 5]. A similar story holds for the targets of statins and ezetimibe, both of which appear amongst the GWAS signals for lipid traits. These examples provide evidence of the overlap between genetic causation and therapeutic benefit. They also lay bare the fallacy of arguments concerning the relevance (or otherwise) of the modest effect sizes revealed by GWAS: therapeutic modulation of the same targets is typically capable of much more dramatic effects.

Having said that, the most direct route to novel targets lies in the detection, through sequencing, of rare, large-effect alleles with dramatic phenotypic effects. The poster child of this approach is *PCSK9*. It is little more than a decade since it was shown that loss-of-function variants in this gene have

profound effects on lipid levels and coronary disease risk, and several potent proprotein convertase subtilisin/kexin type 9 (PCSK9) inhibitors are now the verge of clinical introduction [6]. Evidence that individuals with only one functional copy of the *SLC30A8* gene have a 70% reduction in diabetes risk offers similar potential [7].

The numerical relationship between the per-base mutation rate ($\sim 1 \times 10^{-8}$ per base per generation) and the size of the human population ($\sim 1.5 \times 10^{10}$ haploid genomes) means that most mutations compatible with survival are present in hundreds of living individuals. The phenotypic consequences of losing one (or both) copies of a given gene are typically being played out in the medical histories of thousands of our fellow humans. Harnessing that information is one of the great challenges of our age, but success will bring exquisite biological insights and many novel interventional targets.

From genetics to prediction

One can view the delivery of mechanistic insights leading to novel interventional targets as knowledge gained through the joint analysis of many human sequences. However, there is a second, parallel imperative, which is focused on understanding how analysis of a single human sequence (yours, mine, your child's) can provide clinically useful information, be that through stratification of future disease risk, definition of disease subtype or prediction of response to a diversity of potential interventions.

Neonatal diabetes provides an example of what is possible. The combination of an extreme phenotype (diabetes diagnosed in the first 6 months of life) and penetrant mutations in a well-understood gene (*KCNJ11*, encoding the beta-cell's K_{ATP} channel) has led directly to genetically-driven individualised therapy (in this case, high-dose sulfonylureas) [8].

Type 2 diabetes could hardly be more different. As we have seen, known risk variants account for only 10% of variation in risk, and, in tests of diagnostic or predictive accuracy, genetic tests are comfortably outperformed by classical risk factors (age, BMI, ethnicity) [9]. There is some evidence that the relative performance of genetic prediction improves when performed early in life, or in lower risk individuals, but it still falls some way short of that required for clinical utility.

It is easy to assume that this simply reflects the limited coverage of existing discovery efforts and that further expansion of these efforts will identify additional variants which will solve the problem. However, there are two main reasons why the quality of risk prediction that is possible using genetics alone will continue to be constrained, even in the face of a more complete inventory of predisposing variants.

The first of these is obvious. Genome sequence variation represents only one of several contributors to an individual's risk of developing a late-onset disease such as type 2 diabetes.

Full specification of that risk requires that information on genetic predisposition be integrated with all the other factors that impinge on metabolic performance over a lifetime, including early life events, mediated through epigenetic modifications, and the constellation of environmental and lifestyle experiences from infancy through to senescence. Not to mention the play of chance, mediated, for example, through somatic mutation [10] (see Fig. 1a).

The second reason is more subtle. Even if proves possible, through massive studies, to enumerate the extended list of variants that explains most of the currently hidden genetic variance, this information will not automatically translate into vastly improved prediction of individual genetic risk. Effect sizes at many of these variants will simply be too poorly specified to be informative [11].

Turn the view around, however, and a different perspective emerges. The costs of genome sequencing are falling fast, and the value of these data is rapidly rising. In a generation from now, when it costs less than a course of antibiotics or an outpatient appointment, the value of depositing an individual's genome sequence within their medical record will no longer be a matter of anguished debate. Once universal genome-wide medical sequencing is in place—justified primarily by the obvious benefits for the management of rare diseases and cancer—the proposition for common, chronic disease changes. The question instead becomes: what additional information is needed to interpret, to refine and to augment the clinical value of these data for a disease like type 2 diabetes?

In 50 years, we will have implemented, for many diseases, strategies for measuring individual risk that integrate genetic information from a multiplicity of risk variants with information from the diversity of environmental impacts on risk. These strategies will go well beyond better measures of the risk factors themselves, achieved through advances in human genetics and in the use of wearable sensors and hand-held devices. The critical step will be the use of biomarkers that directly capture disease progression and pathophysiological profile, and which can, through periodic measurement, provide readouts of fluctuating risk over the course of a lifetime. The perfect biomarker would thereby allow imperfect measures of genetic potential (as revealed by genome sequence) and risk factor exposure to be recalibrated based on an individual's empirical trajectory as they travel through molecular 'space' [12]. These trajectories would be judged not so much against population norms, but against one's own personal history (i.e. using the 'historical self' as control), in much the same way that the authorities monitor athletes for evidence of doping. To be informative of risk throughout life, both before and after intervention, such biomarkers should represent direct readouts of causal pathways (see Fig. 1b).

This aspiration to define integrative, dynamic, causal biomarkers may seem fanciful, but an example from current
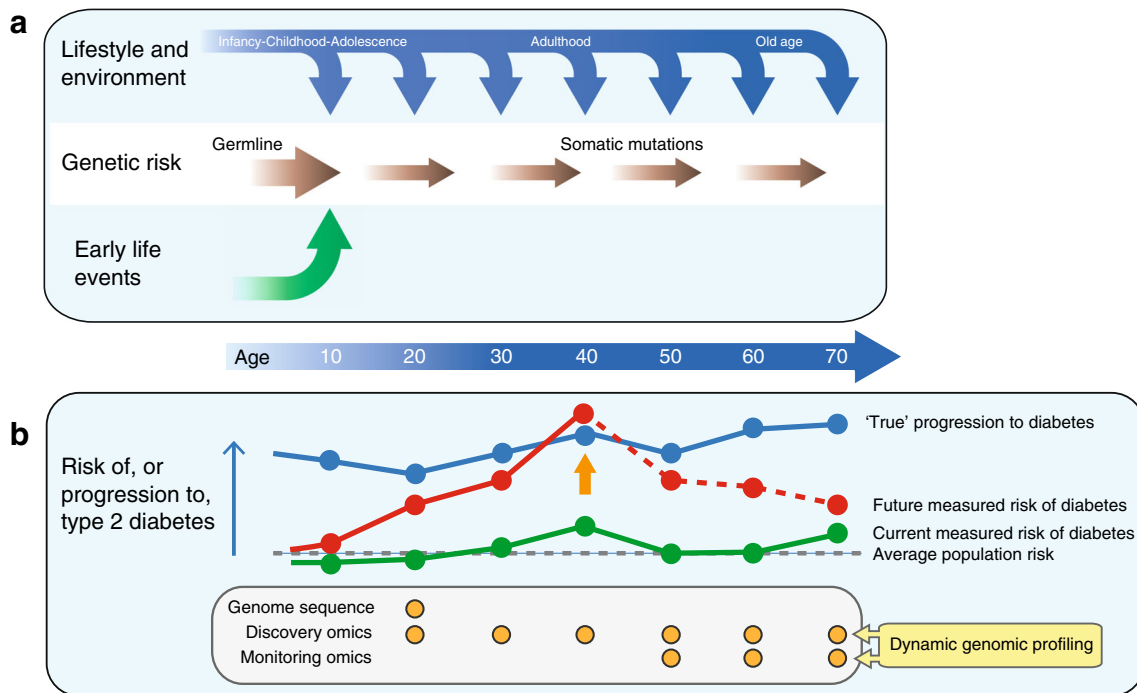
Fig. 1 (**a**) Progress of an individual from health to disease is dependent on the dynamic interplay of genetic, epigenetic and environmental factors during life. Whole genome sequence information gathered in early life captures only a small component of risk. (**b**) Consider the progress of an individual from health to type 2 diabetes. Based on perfect information, this theoretical individual's 'true' trajectory follows the blue line. With current knowledge (e.g. GWAS data, limited clinical and biochemical data), the 'visible' trajectory for that same individual tracks this only poorly (green). With more complete information (genome sequence at age 20 years plus dynamic genomic profiling for a robust molecular signature every 10 years), tracking of disease progression is much improved (red), prompting (in this example) intervention at age ~40 years (orange arrow) that is successful at reducing that individual's risk of type 2 diabetes

medical practice may help. The use of cholesterol levels as a biomarker of coronary disease risk is predicated on precisely the same requirements.

If the goal is to find a 'cholesterol' for type 2 diabetes, then where should we look? The burgeoning range of genomic phenotypes allows us to search beyond classical biomarkers and admits the possibility of complex molecular signatures of risk, based around networks of genetic, transcriptomic, proteomic and metabolomic features. Access to large, well-characterised cohorts, with longitudinal sampling and linkage to electronic health records, will generate rich datasets that can be mined to identify them.

There is no guarantee that such a biomarker exists for type 2 diabetes, or for any other particular chronic disease for that matter. The marked pathophysiological heterogeneity of type 2 diabetes may represent an obstacle that proves impossible to dissect. A more optimistic view is that the longitudinal integration of diverse data types may actually empower efforts to shatter the type 2 diabetes monolith and expose the distinct processes that contribute to individual risk of disease, each of which may have its own molecular signature. This hypothesised reclassification of type 2 diabetes is often simplistically described in terms of defining discrete subtypes (type 2a, 2b, 2c, etc.). It seems more likely, given the genetic architecture and pervasiveness of relevant exposures, that

each individual has his or her own pathophysiological palette of type 2 diabetes risk. If so, the aim should be to describe what contribution each of these processes is, at any given time, making to the evolution of their disease, and to tailor their management accordingly.

Diabetes care in 2065

If all this comes to pass, the focus of diabetes care will have shifted beyond recognition. The disease may not have been eliminated, even in wealthy countries, but the emphasis, and the funds, will have migrated from treatment to prevention. Reliable, actionable, real-time information on type 2 diabetes risk and subtype profile will be available through the integration of baseline genome sequence information and enhanced measures of exposures, with complex molecular biomarkers gathered as part of a universal programme of genomic disease prediction and surveillance. Such information will guide individualised preventative strategies that draw upon a far wider range of effective interventions. This will include smarter drugs, optimised for efficacy and safety, which can be targeted to individuals on the basis of their component pathophysiological profile. If so, genomic medicine, already

a reality for monogenic forms of diabetes, will have proven equally transformative for more complex forms of the disease.

References

1. Mahajan A, Go MJ, Zhang W et al (2014) Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet 46:234–244

2. Morris AP, Voight BF, Teslovich TM et al (2012) Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 44:981–990

3. Langenberg C, Sharp SJ, Franks PW et al (2014) Gene-lifestyle interaction and type 2 diabetes: the EPIC interact case-cohort study. PLoS Med 11:e1001647

4. Plenge RM, Scolnick EM, Altshuler D (2013) Validating therapeutic targets through human genetics. Nat Rev Drug Discov 12:581–594

5. Wessel J, Chy AY, Willems SW et al (2015) Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. Nat Commun 6:5897

6. Cohen JC, Boerwinkle E, Mosley TH Jr, Hobbs HH (2006) Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N Engl J Med 354:1264–1272

7. Flannick J, Thorleifsson G, Beer NL et al (2014) Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. Nat Genet 46:357–363

8. Gloyn AL, Pearson ER, Antcliff J 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

9. 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

10. Jaiswal S, Fontanillas P, Flannick J et al (2014) Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med 371:2488–2498

11. Visscher PM, Yang J, Goddard ME (2010) A commentary on 'common SNPs explain a large proportion of the heritability for human height' by Yang et al. (2010). Twin Res Hum Genet 13:517–524

12. Chen R, Mias GI, Li-Pook-Than J et al (2012) Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell 148:1293–1307